Genetic and Chemometric Properties of Maltese Olive Oils: Developing an Analytical Profile for Protected Determination of Origin Certification.

Frederick Lia

Doctor of Philosophy in Chemistry

at the University of Malta

Supervisor: Prof. Claude Farrugia and Dr. Marion Zammit Mangion

Department of Chemistry, University of Malta

May 2018

Declaration of Authenticity

I, the undersigned, declare that this dissertation is my own original work, except as acknowledged in the text, and was carried out under the supervision Prof. Claude Farrugia and Dr Marion Zammit Mangion. I have clearly indicated where I used the published and the unpublished work of others and I have provided the sources of such work. I have acknowledged all mains sources of help, and where the work was done jointly with others, I have specified what I have contributed and what has been contributed by others. Any conclusions, suggestions or assumptions are mine unless otherwise stated or attributed.

Signature _____

FREDERICK LIA

The undersigned confirms that this dissertation has been undertaken under my supervision and that I approve of its submission for final assessment by the Board of Examiners.

Signature _____

Prof. Claude Farrugia

Date _____

Signature _____

Dr. Marion Zammit Mangion

Date _____

The undersigned confirms that all the changes and corrections as required by the Board of Examiners have been included in a satisfactory manner.

Signature _____

FREDERICK LIA

Date _____

Abbreviations

ABTS	2, 2′-Azino-bis (3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt
AFLP	Amplified fragment length polymorphism
PNN	Artificial Neural Network
CDA	Canonical Discriminate Analysis
cDNA	Chloroplast Deoxyribonucleic acid
CV	Cross-Validation
DA	Discriminate Analysis
DI-MS	Direct Infusion Mass Spectrometry
DNA	Deoxyribonucleic acid
DPPH	2, 2-Diphenyl-1-picrylhydrazyl
EEC	European Economic Community
EEM	Excitation-Emission Matrix
ESI	Electrospray Ionisation
EVOO	Extra virgin olive oil
FFA	Free Fatty Acids
FRAP	Ferric Reducing Activity Potential
FTIR	Fourier Transform Infrared Spectroscopy
HPLC	High Performance Liquid Chromatography
mDNA	Mitochondrial Deoxyribonucleic acid
NMR	Nuclear Magnetic Resonance
NOS	Nitrous Oxide Scavenging
OO	Olive Oil
p-AV	p-anisidine Value
PC	Principal Component
PCA	Principal Component Analysis
PCR	Polymerase chain reaction
PLS	Partial Least-Squares
PV	Peroxide Value
RFLP	Restriction Fragment Length Polymorphism
RMSE	Root Mean Square Error
SCAR	Sequence Characterized Amplified Region
SEEFS	Synchronous Excitation-Emission Fluorescence Spectroscopy
SLC-DA	Stepwise Linear Canonical Discriminate Analysis
SNP	Single-Nucleotide Polymorphism
SSR	Simple Sequence Repeat
TAC	Total Antioxidant Capacity
TdPC	Total diphenolic Content
TFC	Total Flavonoid Content
TPC	Total Phenolic Content
UNK	Unknown
UPLC	Ultra Performance Liquid Chromatography
UV	Ultraviolet
VIP	Variable Importance Plot
p-HPEA	(4-(2-Hydroxyethyl) phenol), tyrosol
3, 4 DHPEA	(3, 4-Dihydroxyphenyl) ethanol, hydroxytyrosol,
p-HPEA- EDA	(4-(2-Hydroxyethyl) phenol) dialdehydic form of elenolic acid
3, 4-DHPEA- EDA	(3, 4-Dihydroxyphenyl) ethanol, dialdehydic form of elenolic acid

Glossary

Analysis of Variance (ANOVA) - A collection of statistical models used to analyze differences among group means. Throughout this work, Kruskal-Wallis ANOVA was used. It is a non-parametric method that does not rely on the assumption of normality.

Cluster analysis – The task of grouping a set of objects in groups, wherein objects in one group are more similar to each other than to objects in other groups.

Correlation Matrix – A symmetrical matrix (M = M': where the variable represented in column *i* is also represented in row *j*), showing the correlations between all the variables. This means that the leading diagonal of the matrix consists of entries equal to 1.

Covariance – A measure of how much two variables change together. A variable which tends to increase as another variable increase is said to have a positive covariance with the other value, while a variable which tends to decrease as another variable increase is said to have a negative covariance with the other variable.

Covariance Matrix – A matrix similar to the **correlation matrix** listing covariances between variables instead of correlations. The leading diagonal of this matrix will contain the variances for each variable.

Deresolve – A signal processing technique that changes the apparent resolution of the instrument, from a high resolution to a lower resolution. It is sometimes used for noise reduction. This function uses a **triangle kernel filter** for smoothing, which in essence convolves spectra in order to estimate their lower resolution equivalent with the original number of variables. The triangular kernel can be regarded as a moving average filter which is weighted more towards the central point than the adjacent ones. In this study, two chPNNels were used for convolution in order to preserve as much information from the spectrum as possible.

Derivative transformation – A signal processing technique often applied to spectra to resolve overlapped bands and to correct for baseline effects. The first derivative of a spectrum is a measure of the slope of the spectral curve at every point, which is independent of baseline offsets, first derivative transformation are often hard to interpret as the peaks in raw spectra tend to become a zero-crossing point upon derivative transformation. On the other hand, the second derivative transformation measures the change in the slope of the curve thereby removing both the baseline offset and the baseline slope of a spectrum. **Gap-Segment** or **Savitzky-Golay** methods are also employed in order to avoid the problem of noise enhancement by applying a degree of smoothing. This is done by using information from a localized segment rather than from adjacent data points of the spectrum to calculate the derivative.

Detrending - A signal processing transformation that removes non-linear trends in spectroscopic data by calculating the baseline function as the least squares fit of a polynomial to the sample spectrum. A second order detrending

transformation is used in order to remove offset, slope and curvature baseline effects. Detrending is often used in conjunction with **SNV** to reduce multicollinearity in spectra, whilst retaining the shape of the spectra.

Excluded Rows Validation - A supervised validation method where a number of samples (usually 1/3) are excluded to be used as the **test set** while the model is built on the remaining samples in the **training set**.

Explained Variance – An output I PLAS-DA statistical analysis. It is the percentage of variation explained by the model in X (**predictor matrix**) or Y (**response matrix**) when considering the number of factors used in the model. A high explained variance is generally desired from PLS models especially in the case of explained Y variation as this indicates that the model is explaining the prediction well. A low explained X variation is undesirable as it indicates that only a small part of the variation in the predictors is explaining the data and thus shows that a large amount of redundant data is present in the model.

Feedforward artificial neural networks (FF-PNN) – neural networks consist of a set of interconnected nodes analogous to a network of neurons in a brain. Feed-forward neural networks consist of neurons which only transfer information in one direction, from the input layer to a number of hidden layers and finally to an output layer. In 'scikit-learn' (Pedregosa *et al.*, 2011) neural networks are implemented in the **Multi-layer Perceptron** (MLP) Module which learns a function $f(\cdot) : \mathrm{R}^m \rightarrow \mathrm{R}^o$ where m is the number of dimensions for input and o is the number of dimensions for the output.

The input layer consists of a set of nodes containing the input features, the input neurons are connected to the hidden layer whereas each neuron in the hidden layer transforms the values from the previous layer with a weighted linear summation $w_1x_1 + w_2x_2 + \ldots + w_mx_m$ which is followed by a non-linear activation function in the case of non-linear models. The hidden layer is then connected to the output layer which transforms the values into the classification or prediction.

The MLP classifier takes a number of parameters including; **Hidden layer sizes –** the number of nodes in the hidden layer, **solver** – the solver (minimiser) for weight optimisation, in this study, the **LBFGS** (Limited-memory Broyden–Fletcher–Goldfarb–ShPNNo) solver was used as it is good for relatively smaller data sets (n<1000) and generally more robust than conjugate gradient methods albeit more computationally intensive, **learning rate –** 'invscaling' was used which gradually decreases the learning rate with every iteration, **tolerance –** a low tolerance was used in order to ensure that the model converges with a small RMSE, **maximum iterations -** a large number of iterations (5000) was selected to ensure that the model converges at a low tolerance, **activation –** the activation function 'identity' was used for linear models as this function does not apply any non-linear transformation to the hidden layer , and 'relu' was used for non-linear models as this function returns $f(x) = \max(0, x)$, r**andom state –** a random seed which is used to start the model at different states in order to check if the model reached a suitable minimum.

***F*-ratio –** In SLC-DA statistical analysis, the statistical significance of the discrimination between the groups. The larger the ***F*-ratio** the higher the significance.

Gaussian filter – A signal processing technique similar to a **moving average filte**r. However, the averaging function is determined by a Gaussian function.

Hierarchical Cluster Analysis (HCA) – Hierarchical Cluster Analysis is a method of **cluster analysis** which seeks to build a hierarchy of clusters. This study only employs Ward's method for HCA where the criterion for clustering is based on minimizing the error sum of squares.

K-Fold cross-validation (CV) – A validation method similar to the **LOOCV** but instead of leaving a sample out, the sample set is split into groups determined by the number of folds. A K-fold validation will produce k subsamples, the statistical model is built for k times with each of the k subsamples used once as a **test set** while the remaining samples are used as the **training set**.

Linear Discriminant Analysis (LDA) – A supervised classification technique that uses a linear combination of variables in order to characterize two or more classes by linearly transforming *n*-samples into an *m*-dimensional space ($m<n$). Samples from the same class will lie close to each other while samples belonging to different classes will be far apart. The LDA function within the 'scikit-learn' toolbox (Pedregosa *et al.*, 2011), models the class distribution of the data after which predictions are obtained using Bayes' rule. However, going into the mathematical model of how this is performed is outside the scope of this work.

Leave one out Cross-validation (LOOCV) – A validation procedure that successively leaves each sample out of the **training set** and then uses that model to predict the excluded sample. This method is used to assess the predicting ability of the model especially in cases where the number of samples is small. It is not suited for datasets with a large number of samples as a statistical model must be constructed for each sample present.

***p*-value –** In SLC-DA statistical analysis, the probability of *F,* tests if the inclusion of a variable has no effect on the group, a low *p*-value denotes that the inclusion of the variable is significant.

PARAFAC – A multiway case of PCA. While PCA models consist of components which describe the original data structure using one score vector and one loading vector (equation 2), a PARAFAC model uses components with one score vector and two loading vectors (equation 1).

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk} \quad \textbf{- Equation 1.}$$

Where x_{ijk} is one element of the data, a_{if} and b_{jf} are loadings while c_{kf} are the scores, e_{ijk} is the residual matrix and F is the number of components. The scalar notation is used in this case since linear algebra cannot be used to describe multi-way arrays easily. It is worth mentioning that in the multiway analysis it is a common practice that scores and loadings are not distinguished and are treated equally numerically.

PARAFAC uses an ALS (alternating least squares) in order to find a unique solution and minimize the residual error.

Principal Components Analysis (PCA) - An unsupervised multivariate analysis technique that is used to elucidate the relationships between samples having several or more variables. PCA highlights the differences between the analysed samples and thus provides useful information on the data under analysis. This makes PCA a useful tool in chemometrics through which visually uninterpretable differences in spectroscopic measurements can be emphasised, aiding in the interpretation of a large number of variables present in spectroscopic measurements. Briefly, through PCA the original set of multidimensional data is orthogonally transformed into a set of linearly uncorrelated variables referred to as **Principal Components (PCs)** such that the greatest variance lies in the first component while subsequent PC's have a smaller variance than the preceding PC. Most variance within the data is usually explained within the first few PCs and usually, these components are used in order to build a model as shown in equation 2 using linear algebra:

$X = TP' + E$ **- Equation 2.**

Where X is the data, T is the scores matrix, P' is the transposed loadings matrix and E are the **residuals** or the error matrix. The combination of T and P' is the part of the data explained by the principal components thus T and P will be $n \times m$ matrices with n being equal to the number of PCs and m being equal to the number of samples in matrix T and the number of variables in matrix P. Matrix E will contain the data which is not explained by the latent variables (PCs).

S**cores** describe the data structure and are used to show sample similarities or differences along a component; in essence, the score reflects the sample location (coordinate) on a PC. Similar scores for different samples show that they are similar to a PC while different scores mean that the samples are different with regards to that PC. Consequently, scores from two different principal components are usually plotted in bi-plots in order to give a visual representation of the variation of the data in a score plot.

On the other hand, **loadings** describe the influence each individual variable has on a principal component. Loadings with small values for a given PC mean that they are not well accounted for by that PC. For loadings with positive values, it follows that samples with positive scores will have a higher value than the average value for a given variable, while samples with negative scores will have a lower value, the opposite applies for loadings with negative values.

Median filter - A signal processing technique that replaces each data point by the median of its neighbours.

Moving-average filter - A signal processing technique that replaces the value at each data point with the average of its nearest neighbours with the number of nearest neighbours being determined by the gap size.

Multiplicative Scatter Resolution (MSC) –A signal processing method developed to deal with multiplicative scattering in reflectance spectroscopy, though it can also be used to treat similar effects such as path length variations, offset shifts and interference. In this case, the correction is performed by generating two

coefficients *a* and *b* through a regression for each spectrum onto the average spectrum of the samples. The coefficient *a* is the intercept of the regression line and *b* is the slope; these coefficients are then used to correct the values of each sample.

Orthogonal Signal Correction (OSC) – A signal processing technique, used as a transformation method prior to PLS regression, which removes irrelevant variances from the predictors which are not correlated to the response (Fearn, 1999).

Partial Least Squares Discriminant Analysis (PLS-DA) – Partial Least Squares Discriminant Analysis models both the X- and Y- matrices simultaneously in order to determine the **Factors** (latent variables) in X that will best predict the latent variables in Y. In essence, PLS-DA is used to correlate the information in X to the information in Y by maximising the covariance between them whilst minimising the residual error.

Assuming that Y-Matrix can be predicted using equation 3 below where *U* is the score matrix, *Q'* is the transposed loading matrix and *F* are the residuals. The X-Matrix can be modelled using equation 2.

$$Y = UQ' + F \quad \textbf{- Equation 3.}$$

The PLS model will be developed in such a way as to maximize the covariance between the scores in X (*T*) and the scores in Y (*U*) for each factor as shown in equation 4.

$$\begin{aligned} u_1 &= r_1 t_1 \\ u_2 &= r_2 t_2 \\ &\vdots \\ u_n &= r_n t_n \end{aligned} \quad \textbf{- Equation 4.}$$

Where u_n is the scores for each extracted factor in the X-matrix, t_n are the scores for each extracted factor in the Y-matrix and r_n are some constants.

This means that scores in *T* can be used to predict the scores in *U* and subsequently the Y-matrix from the loadings *Q'*.

Predictors – used to describe the X-matrix or the matrix containing the spectral measurements.

Random Holdback – a validation method similar to the **Excluded Rows** but the **test set** is built through a random selection of samples rather than a supervised one.

Response - refers to the Y-matrix which in this study always refers to the classification system.

Savitzky-Golay – A signal processing method in which a smoothing algorithm fits a polynomial of a specified order to the data points, and the averaged value is then predicted from this equation. This method is thus useful for removing noise without a great distortion in the signal.

Standard Normal Variate (SNV) – A signal processing technique in which a transformation is applied to spectroscopic data to remove scatter effects by centring and scaling each individual spectrum. The mean and standard

deviation of all data points in a spectrum are calculated after which the transformed spectrum is calculated by subtracting the mean from every data point in the spectrum and dividing by the standard deviation.

Stepwise linear canonical discriminant analysis (SLC-DA) – A linear discriminant method that is used for classification and data reduction. In this study discriminant analysis using a linear method was applied. This works by calculating distances from group means as the Mahalanobis distance whereby a common covariance matrix is used for all groups.

A **stepwise** analysis allows for manual selection of variables used to build the linear model up to a maximum number entries (n-1) where n is the number of samples in the sample set. This enables the user to select **variables** with large ***F*-ratios** and small ***p*-values** in order to build a model on the most discriminant **variables.** With the addition of a new variable, the values for F and p will change for the both the included and excluded variables with the termination step being when the F-ratio and p-value approach a value of 0 and 1 respectively.

Canonical discriminant analysis generates components referred to as canonical discriminant functions that are described by scores and loadings. Similarly to PCA, the first two canonical functions provide the maximum separation amongst different groups.

Sample - represents a row and is equivalent to describing the whole spectrum of a single sample of honey or a series of concentrations of different phenolic compounds for one sample.

Supervised Learning – multivariate methods where the outputs of the training samples are known prior to modelling and are thus used to train the model to give desired outputs.

Support Vector Machines (SVM) – a set of supervised learning methods used for classification or regression with the advantages of being effective in high dimensional spaces (and in cases where the number of variables is larger than the number of samples), computationally efficient and versatile. SVM attempts to map samples as points in space so that separate categories are divided by a clear gap which is as wide as possible. New samples can then be classified into categories depending on which side of the gap they fall. **A Linear kernel SVM** model was used as implemented within the 'scikit-learn' toolbox (Pedregosa *et al.*, 2011) which attempts to find a hyperplane which separates the points in multidimensional space into separate classes. The value for ***C*** or **penalty parameter** will determine the size of the hyperplane margin with a larger value for the C parameter giving a smaller-margin hyperplane margin which increases the probability of all the training points being correctly classified but reduces the margin of separation between the classes.

Test set – sample set used for validation of the statistical model built on the **training set.**

Training set – sample set used to build the statistical model, also referred to as **calibration set**.

Unsupervised Learning – methods where no output data is given prior to modelling, thus the model itself attempts to find structure or relationships within the dataset provided.

Variable – used to describe a column or a single wavelength in the case of spectroscopic data. In PCA and PLS-DA modelling each variable is equivalent to a dimension in an *n*-dimensional space where *n* is the number of variables.

Variable importance in projection (VIP) – The VIP score is a measure of a variable's importance in the PLS-DA model. It represents the contribution of a variable to the PLS model and is determined through a weighted sum of the squared correlations between the model components and the original variable. A value of less than 0.8 is typically considered to be a small VIP and thus a candidate for deletion from the model. (Wold, 1995; Eriksson *et al.*, 2006)

Wilks' Lambda (λ_w) – A parameter used in SLC-DA to test whether there is difference between the means of identical groups on a combination of independent variables. This parameter tests the null hypothesis which states that the means of all independent variables are equal across the groups being tested. In order to achieve good predictability, the means of the groups must have different values thus variables which minimize the value of Wilks' lambda (equation 5) should be included in the model.

$$\lambda_w = \frac{S_{intra}}{S_{intra} + S_{inter}}$$ **- Equation 5**

Where S_{intra} is the sum of squares of data points belonging to one category, and $S_{intra} + S_{inter}$ is the total sum of squares (Vandeginste *et al.,* 1998). Values of λ_w approaching zero are obtained in cases where the categories are well resolved, while values of λ_w close to 1 signify overlapping categories (Lerma-Garcia *et al.,* 2007).

Abstract

The determination of the singularity of the Maltese olive oil provides an opportunity for local producers to purpose a PDO certification. This objective of this study was to identify chemical and genetic parameters which enable the discrimination of Maltese olive oils.

Prior to the identification of markers, a preliminary survey of quality was conducted whereby it was shown that monocultivar olive oils grown locally were of sufficiently high quality to be classified as extra virgin olive oils. The minor constituents present in EVOOs were subjected to a more detailed study as chemical markers. Phenolic compounds were extracted and quantified using microtiter assays. The antioxidant activity of these extracts was determined using different redox-based assays including FRAP, ACC and CUPRAC assays whilst the radical scavenging activity towards different radicals, including DPPH, ABTS and NO, was determined. It was shown that EVOOs derived from the indigenous cultivars had a significantly lower TPC and TFC whilst no significant difference was observed in the TdPC when compared to EVOOs from other Mediterranean countries. The low TPC and TFC reflected the significantly lower antioxidant and radical scavenging activity of the indigenous cultivars. Whilst microtiter plate assays showed that the indigenous cultivars had a lower phenolic content, application of HPLC for phenolic profiling revealed that both the indigenous and locally grown foreign cultivars had a significantly higher content of six phenolic compounds, namely p-coumaric acid, tyrosol acetate, 3,4 DHPEA-EDA, p-HPEA-EDA and two unidentified compounds. Elemental characterisation via the application of semi-quantitative XRF analysis enabled discrimination not only between the geographical origin of the EVOOs but also between EVOOs and other refined seed oils. Application of classical statistics and chemometrics on the phenolic profiling showed that geographical discrimination between the different samples was possible. The application of chemometric techniques namely PCA, PLS-DA, PNN, SLC-DA on data derived from different chemical techniques namely fluorescence (SEEF and EEM), FTIR, NMR and direct infusion mass spectrometry not only enabled the complete discrimination of the different EVOOs but enabled the identification of markers which had the highest discriminatory power. In the case of SEEF spectroscopy, compounds having emissions in the 380-480 nm and 650-700 nm ranges were found to have the greatest discriminating power, as confirmed through the use of 3-way chemometric analysis (PARAFAC and N-PLS) on the EEM offering comparable results to those obtained using SEEF. For NMR, ^{1}H chemical shifts in the 3.5-4.5, 6.7-7.0 and 9.1-9.5 ppm had the most discriminate power and were attributed to the presence of glycerides, terpenic, phenolic and carbonyl containing compounds. In the case of DI-MS under positive ESI, it was found that the data obtained was highly redundant, nonetheless minor phenolic compounds identified through their m/z values were found to be the most discriminate. FTIR spectroscopy was found to be one of the most effective chemical fingerprinting methodologies owing to its simplicity, with compounds having functional groups which vibrated at 500-1500 cm^{-1} being the most discriminate.

Application of genetic analysis on the three indigenous cultivars using SSR markers revealed that cultivar discrimination was possible via the use of two most informative markers namely DCA-3 and GAPU 101. Analysis of these markers on genomic DNA showed that the 'Bidni' cultivar consisted of a homozygous population whereby the individual trees are clones of each other showing a very similar allelic pattern to olive cultivars cultivated in northern Tunisia. In the case of the 'Malti' cultivar, the population was highly heterozygous, showing very similar allelic patterns to Spanish, Greek and Southern Italian cultivars, suggesting a multi-cultivar population coined under the same nomenclature. In the case of 'Bajda' at the two loci studied this could not be distinguished from another Leucocarpic cultivar grown in southern Italy baring the name of Morachia, Cannellina, Bianca, and Chiarita. From the results obtained it was concluded that the authenticity of the Maltese EVOO can be defined using both genetics analysis and chemical fingerprinting methods.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to both my supervisors Prof. Claude Farrugia and Dr. Marion Żammit Mangion, for thier constant help and support. I would also like to thank Prof. Roboert Borg for the training and help regarding during the NMR experiments. Together with Mr.Joseph Grech for his help during the XRF and not to forget the head of department Prof Emanuel Sinagra for his endless fight to get the department more fund and new state of the art equipment.

I would also like to thank the Board of MGSS Scholarship Scheme, for permitted this project. I would also like to thank all the staff of the Department of Chemistry who helped me throughout this research project and Mr Sam Cremona for providing the majority of the olive samples used in this project. A great appreciate goes to both Prof. Antonella Pasqualone, and Prof. Cinzia Montemurro for their hospitality and help during my short visit to the University of Bari.

My grateful thanks goes to my family and friends, who offered endless support and encouragement.

Table of Content

List of Figures

List of tables

Introduction

In recent years the olive cultivation industry in the Maltese Islands has re-emerged, potentially allowing the creation of a niche market for high-quality olive oils produced by the Maltese agribusiness sector. At present the majority of the cultivated trees used for oil production are imported, since they are associated with better oil yields, placing the relatively unexploited and yet uncatalogued native olive trees at risk. There are three major identified olive cultivars within the Maltese islands which are thought to be native, namely the 'Bidni', 'Bajda' and 'Malti' (Borg, 1922). Whilst the 'Bidni' and 'Bajda' are monocultivars the 'Malti' is thought to be made up of a number of ancient varieties which are geographically isolated from each other (Mazzitelli *et al.*, 2015). In this study a variety of olive oils selected from different areas around the Maltese islands and countries around the Mediterranean were studied. The first part of this work focused on the development of chemometric methods capable of assessing olive oil quality and the possibilities of developing methods which allow the determination of the authenticity of the Maltese olive oil. For comparison oils from geographically neighbouring regions were also studied. The second part of this study focused on establishing the genetic profile of the olive oils and hence the cultivars from which they are derived. Fresh samples of cold pressed olive oil and leaf samples were collected from different geographical areas of the Maltese islands, and included the monocultivar 'Bidni' and 'Bajda' olive cultivars as well as a representative of the 'Malti' cultivar. The ultimate aim of the study was that through the use of both chemical and genetic analysis enough evidence would be collected to prove the singularity of the Maltese olive oil.

The determination of the singularity of the Maltese olive oil provides the opportunity for local producers to pursue the PDO certification, which will improve the status of Maltese olive oil as the raw material will become more valued. There is also a significant economic benefit for the producers, as well as for the Maltese economy, in promoting the unique qualities and flavours of the product and therefore in protecting the very notion of what defines that product. This certification would protect and guarantee a premium price for authentic products, and eliminate unfair competition and the misleading of consumers by non-genuine products. Apart from obtaining the PDO certification this study involved the development of methods which

can be used by private and governmental facilities for the monitoring and analysis of the quality of Maltese olive oil as well as other autochthonous products (Sacco *et al.,* 2000). This will ensure food traceability which is important as it confirms the local origin of the olive cultivar. Fraudulent behaviour such as intentional blending with other oils of cheaper quality will also be reduced, therefore safeguarding and reassuring the public about the quality and safety of their product.

1.1 Aims and objectives

The objective of this study was to identify and define the uniqueness of Maltese olive oil and the cultivars from which it is derived, using both genetic analysis and chemometric methods. This led to the development of fast and sensitive methods for the characterization and authentication of Maltese olive oils, in relation to its quality, genetic variety, chemical composition and geographical origin.

The specific objectives of this study were as follows:

Chemical analysis objectives

1. Collection of monocultivar extra virgin olive oils from the Maltese islands and different countries across the Mediterranean.

2. Determination of the quality of the EVOOs derived from the Maltese islands via the assessment of the following criteria:
    1. Free fatty acid content
    2. Iodine value
    3. Oxidation parameters: peroxide value, p-Anisidine value and TOTOX
    4. Spectroscopic parameters: K_{232} K_{266} K_{270} K_{274} ΔK

3. Assessment of the variation in the microconstituents of olive oils derived from different geographical origins , including the determination of the following parameters:
    1. Chlorophyll and carotenoid content
    2. Quantification of the different phenolic classes found in olive oils via microtiter assays, including the total phenolic, total flavonoid and o-diphenolic content

3. Qualitative analysis of micro elemental constituents present in different olive oils and seed oils via the application of X-Ray fluorescence (XRF)

4. Application of metabolic profiling and fingerprinting in conjunction with chemometric models, in order to develop methods that enable the discrimination of EVOOs from different geographical origins and thus determine the authenticity of the Maltese EVOOs. This approach would be applied to results obtained from the following methods:
    1. Phenolic profiling via high liquid chromatography
    2. Synchronous fluorescence spectra
    3. 3D Excitation and emission spectrofluorometric spectra
    4. Direct infusion mass spectrometry
    5. Infrared spectroscopy
    6. Nuclear Magnetic Resonance

5. Determination of the antioxidant activity of complete EVOOs and their phenolic constituents through the application of non-cellular models based on the following assays
    1. DPPH, ABTS and nitrous oxide radical scavenging activity
    2. Ferric, cupric and molybdate reducing activity
    3. Correlate the observed antioxidant activity with the amounts of phenolic compounds present within the oil.
6. Adulteration studies using different seed oils and Maltese EVOOs
7. Determination of the oxidative stability of Maltese EVOOs, through the application of a forced thermal degradation assay.

Biomolecular analysis objectives

1. Development and optimization of methods allowing extraction of good quality DNA from both EVOOs and leaves of their corresponding cultivars, focusing on the indigenous Maltese cultivars.
2. Amplification of the extracted DNA with an array of SSR primers, with the aim of determining the most consistent and informative amplification pattern.
3. Assessment of the genetic similarity between the different cultivars.

1.2 Historical overview of Olive and Olive Oil

Olive oil derived from the olive (*Olea europaea L*. subsp. *europaea* var. *europaea*) is the main edible vegetable oil which symbolizes the Mediterranean countries. The oil is obtained by crushing and pressing the fruits of the domesticated olive tree. Domestication of the tree is thought to have occurred approximately 6,000 years ago in the eastern Mediterranean area. The fossil record of olive trees traces back to the tertiary period as the presence of fossilised leaves of ancestral olive trees were found in Italy (Vossen, 2007). The presence of olive stones dating back to the palaeolithic period (Bronze Age) (35,000– 8,000 B.C.) near ancient human settlements found in southern Europe confirms that at the time humans already exploited this plant as an edible source (Schafer-Schuchardt, 1988). However, the precise origin of the domesticated Mediterranean olive tree is still not clear. Phylogeographical and phylogenetic reconstructions of the diffusion of the olive trees suggest that it most probably originated in the regions of Persia and Mesopotamia (Boskou, 1996) from the domestication of the wild olive tree (*O.europaea* var. *sylvestris*) and then diffused to the present Syria, Lebanon and Israel (Zeder *et al.,* 2006; Besnard *et al.,* 2001; Breton *et al.,* 2006).

Expansion of the olive tree's habitat continued with the immigration of the Phoenician colonies that advanced in the Mediterranean from east to west, from the regions of Lebanon to the coasts of Egypt and the island of Crete spreading the olive tree with their colonization. By the fourth millennium BC, Phoenician colonies colonised Libya, Greece, Sicily and southern Italy, further spreading colonisation of the olive tree (Harwood and Aparicio 2000; Vossen, 2007). At the end of the Roman era, both the olive tree cultivation and the primordial technology for olive oil production had spread to most parts of the Mediterranean, with the exception of parts of Spain, France and North Africa (Grigg, 2001). It was first introduced in France around 600 B.C (Luchetti, 2002; Grigg, 2001). Recent studies suggest that the olive trees reached the Iberian Peninsula 700 years ago (Arnan *et al.,* 2012) rather than in the second millennium BC, implying that the Arabs were the first to introduce the production of olive oil in Spain. Evidence of this may be that the Spanish word for oil - *aceite* - comes from the Arabic *alzat*, or olive juice. The expansion of the olive tree in the Americas was undertaken by the Spanish conquistadors who planted the first

trees in the Caribbean, and afterwards in the American continent (Vossen, 2007; Kapellakis *et al*., 2008).

For the Maltese islands, there are no specific records which date the introduction of olive tree cultivation and olive oil extraction. Analysis of burnt wood remains found in the prehistoric temple of Skorba (limits of Mgarr) by Dr C. R. Metcalfe of the Royal Botanical Gardens provided information of the flora growing in the Maltese islands between 5000-4300 B.C. This analysis revealed the presence of olive tree remains, which according to Metcalfe originated from a canopy which covered the temple. On further analysis, it was shown that these remains were in fact derived from cultivated trees rather than wild olive trees and dated back to the copper age (3150-2500 B.C) (Trump, 1966). Palynological analysis of pollen grains found in clay layers of a water aquifer in the limits of Luqa also revealed the presence of pollen derived from *Olea* genera however it was not clear whether the pollen grains were derived from wild or cultivated trees.

Ancient records by Diodorus suggest that the Phoenicians designated Malta as a place to settle in order to extend their trade towards the western part of the Mediterranian (Bonanno, 1991). Furthermore, it is thought that the settlers aided in the development of new agriculture lines, in particular, olive cultivation, nevertheless the evidence for this is still not substantial. The archaeological records of the Roman period provide the most conclusive evidence regarding olive oil production in the Maltese islands. Several *trapeta* - huge stone mills in which olives were crushed to an oily paste were - discovered in Roman villas at San Pawl Milqi (Cassar, 2015). It is thought that a large number of olive trees were found in certain locations of the Maltese islands such as Żebbuġ (both Malta and Gozo), Żejtun and Birżebbuġa, all names synonymous with the growth and production of olive trees. The names of these locations were coined by the Arabs, nonetheless during this period the importance of olive cultivation came to a halt as the introduction of cotton and citrus plants took place. The majority of the olive trees were cut down in order to make room for cotton and citrus production which were deemed to be more commercially viable. This fragmentation of olive groves caused the formation of secluded old olive groves scattered along the Maltese islands.

During the 1990's the olive oil industry in Maltese islands regained attention with the importation of international cultivars. Nowadays most of the olive groves found in the Maltese islands are made up of cultivars originating from Spain and Italy, with the most common including Frantoio, Carolea, Pendolino and Bella di Spagna. In November 2006 the PRIMO project (Project for the Revival of the Indigenous Maltese Olive) was launched, with the aim of reviving and conserving the indigenous olive oil cultivars found in the Maltese islands.

1.3 Regulations safeguarding the origins of Olive Oils

Due to the autochthonous nature and the symbolic meaning of the olive tree and its oil, the establishment of specific production protection systems came into action. In 1992 the European Commission introduced two types of accreditation namely the protected designation of origin (PDO) and protected geographical indication of origin (PGI) (EEC Regulation No. 2082/92 and later No. 510/06). These systems were designed to protect the typical characteristics and to authenticate food products, with the aim of discouraging competition from similar replacement products. This was done by the introduction of rigorous regulations such as those imposed by the EEC Regulation (No. 510/2006) regarding labelling, production and commercialisation of olive oil. According to the EEC regulation Reg. (510/06) for a product to be conferred PDO status the entire production cycle, from raw material to finished product including processing and packaging, must be carried out in one given territory. The amalgamation of different factors including the starting raw materials (olive oil varieties), environmental physiognomies, location and skilfulness of the producer, makes the product unique and not reproducible elsewhere. The product obtained bears distinctive and exclusive characteristics such as defined chemical composition and distinct organoleptic parameters.

The difference between PDO and PGI is that the latter has less stringent requirements but still demands that the marketed product needs be produced in the geographical region which defines its origin. The geographical linke to the originating country must occur in at least one stage of production, processing or preparation. Other lesser-known designations defined in European Union Law for the protection of regional foods include traditional speciality guaranteed (TSG) and geographical indications (GIs). According to the European Council Regulation (EC), No 509/2006

a product established under TSG is defined as an edible agricultural product which is produced according to a traditional production method. Compared to GI, products can be both agricultural and non-agricultural named after the producing region and own a self-assured reputation or qualities derived from their place of origin.

Products having an authentication certificate are highly appreciated as they bear the assurance of quality. The determination of the origin and the authenticity of olive oils has been studied extensively in the past few years using both biological and a variety of physicochemical techniques in conjunction with chemometric studies. From a chemical perspective, these studies can be classified into two main categories (Dupuy *et al.,* 2005). In one category the samples are previously subjected to a chemical treatments in order to determine and quantify the individual chemical constituents such as fatty acids, triacylglycerols (Bucci *et al.,* 2002; Ollivier *et al.,* 2003), sterols (Leardi and Paganuzzi, 1987), phenolic compounds (Faouzia *et al.,* 2008; Talhaoui *et al.,* 2015; Gong *et al.,* 2003) aliphatic alcohols (Rigane *et al.,* 2011) and inorganic multi-elemental composition (Beltrán *et al.,* 2011). The other category is based on sample preservation, where a small set of samples is analysed without extensive preliminary treatment, using methods that include ^{1}H ^{13}C ^{31}P NMR analysis (Alonso-Salces *et al.,* 2010); Fourier transform infrared spectroscopy (FTIR) (De Luca *et al.,* 2011; Bendini *et al.,* 2007), Near infra-red (Galtier *et al.,* 2007); Synchronous excitation-emission fluorescence spectroscopy (SEEFS) (Dupuy *et al.,* 2005) and direct infusion electrospray ionization mass spectrometry (Lerma-García *et al.,* 2008).

Since the chemical composition of olive oil is highly complex it is invariably affected by a multitude of factors including the variety from which it originates, the terroir, fruit quality, degree of ripeness and extraction technology (Solinas *et al.,* 1987; Solinas, 1987; Patumi *et al.,* 1992; Fontanazza *et al.,* 1993). Although the use of complex chemometric models is helpful in determining the authenticity however it requires specific chemical markers which are less likely to decompose over time.

The use of DNA analysis provides an invariant traceability to the producing cultivar, independently of the place of origin, leading to specific identity or authenticity determination. The specificity of this technique does not highlight differences which the cultivar acquires by time from the environment and thus although the foreign cultivars which are grown in Malta produce oil of different

chemical composition, the inherited genetic material would be invariably the same to that found in other countries. The most commonly employed genetic techniques involve the isolation of DNA from olive oil, followed by polymerase chain reaction (PCR) amplification, aiming for specific biomolecular markers. A large number of techniques have been developed throughout the years these include; Random Amplified Polymorphic DNA (RAPD), Simple Sequence Repeat Polymorphism (SSR), Cleavable Amplified Polymorphic Sequences (CAPS), Amplified Fragment Length Polymorphism (AFLP), Sequence Characterized Amplified Region (SCAR) and Inter- Repeat Amplification (IRA), (Dietrich, Weber, Nickerson and Kwok, 1999).

Locally the extraction of olive oils is carried out using a continuous extraction system. The first step involves the washing and leaf removal, this step is vital for both the end product and the machinery as it reduced the chance of vegetable or non-vegetable parts that could be harmful to the machinery or contaminate the product. This operation is carried out by specialised machinery equipped with powerful blowers, which removes leaves and twigs, and a washing tank, with forced water circulation in which olives are washed. The second step involves the crushing of the olives into a paste, this is carried out using a metallic crusher, which consists of a metallic body, that rotates at high speed and throwing the olives against a fixed metal grating. Once the olive paste has formed it is homogenized into a mixing vessel. Mixing is vital, as it promotes contact between oil droplets allowing the separation of a continuous oily phase, this process is known as malaxation. Mixing is done for 30 to 60 minutes depending on the amount of water present in the flesh of the olives. Water is sometimes added in order to soften the mixture favouring phase separation. The temperature of mixing is another important parameter which determines both the organoleptic parameters and the time of mixing. The higher the temperature (40 ºC), the longer the time of mixing and the higher the total phenolic content (Di Giovacchino *et al.,*2002; Inarejos Garcia *et al.,*2009). The last stage of the extraction process involves the use of a horizontal centrifuge. The horizontal centrifugal force separates oil from other liquid and solid phases of olive paste, based on the weight of the fractions.

1.4 Definitions of olive oil

According to the international olive council (IOC), olive oil is solely obtained from the fruit of the olive tree (*Olea europaea* L.) and excludes oils obtained via solvent extraction or re-esterification processes. The early EEC Regulation 136/66 in 1966 established a set of rules which define the organization olive oils. Articles 35 and 36 of the legislation define the different olive oils according to their free acid content, expressed as oleic acid w/w and organoleptic parameters.

Virgin olive oils

- Extra virgin olive oil is distinguished by perfect sensory characteristics and the content of free fatty acids does not exceed 1 g/100 g.
- Fino olive oil has perfect sensory characteristics, however it has a higher free fatty acid content, which does not exceed 1.5 g/100 g.
- Ordinary olive oil has some sensorial imperfections that are still acceptable and a free fatty acid content which does not exceed 3.3 g/100 g.
- Lampante olive oil has unpleasant organoleptic parameters either attributed to bad fruit or careless processing, with a free fatty acidity higher than 3.3 g/100 g. Lampante olive oils are not fit for human consumption and are subjected to refining and sold as a component in refined seed oils. Refining involves heat treatment of oil, filtration, and other chemical processes including bleaching. The refining process removes any unpleasant taste and smells from the oil, making it suitable for blending with virgin olive oil in order to impart some flavour and aroma.

Refined olive oils

- Refined olive oil is obtained by subjecting lampante virgin olive oil to various refining methods which do not alter the initial glyceridic structure.
- Pure olive oil is obtained by a combination of refined olive oil and virgin olive oil for consumption.

1.5 Chemical constituents of olive oil

The chemical composition of olive oil can be broadly divided into two, saponifiable and unsaponifiable fractions which are dependent on the behaviour of their respective chemical constituents in the presence of a strong alkaline solution and heating. The saponifiable fraction constitutes 98% to 99% of the total weight, and as the name implies it is composed of compounds which form soaps under the aforementioned conditions. These include free fatty acids and their corresponding glycerol esterified compounds, triglycerides, diglycerides and monoglycerides. The remaining unsaponifiable fraction (1-2%) is composed of microconstituents which do not form soaps. Although this fraction is very small it is the most diverse and imparts the beneficial nutritional aspect of olive oil. Furthermore, this fraction is also vital from an analytical point of view as it offers authentication markers, and determines the olive oil's shelf life. This fraction contains mostly sterols, fat-soluble vitamins, waxes, aliphatic alcohols, aromatic compounds and phenolic antioxidants.

1.5.1 Major constituents of olive oil

1.5.1.1 Triglycerides and Fatty acids

The beneficial aspect of olive oil is also derived from the prevalent presence of the monounsaturated oleic acid (C18:1) which confers the ability to prevent cardiovascular diseases (Keys, 1980). The fatty acid fraction is composed mainly of monounsaturated fatty acids (MUFA) and a relatively low level of polyunsaturated fatty acids (PUFA). Triglycerides, which account for 98% to 99%, consist of one molecule of glycerol on which are esterified up to three fatty acid chains (saturated and unsaturated).

Due to the difference in the biosynthetic pathways within the plant kingdom, triglycerides and their corresponding fatty acids provide a means to detect adulterations with olive oils. Within olive oil the presence of free unesterified fatty acid is normally absent, thus their presence is indicative of adulteration with oils from different seeds. The amended EC Regulation 2568/91 states that in order for an olive oil to classify as virgin the relative percentages of myristic, linolenic, arachidic, eicosenoic, behenic and lignoceric acids need to be measured. The typical fatty acid composition of olive oil ranges from 7.5 to 20.0% palmitic acid, 0.5 to 5.0% stearic

acid, 0.3 to 3.5% palmitoleic acid, 55.0 to 83.0% oleic acid, 3.5 to 21.0% linoleic acid, 0.0 to 1.5% linolenic acid, 0.0 to 0.8% arachidic acid, 0.0 to 0.2% behenic acid, and 0.0 to 1.0% lignoceric acid (Montedoro *et al.,* 2003).The fatty acid composition may differ from sample to sample, depending on the zone of production, latitude, climate, variety, and the fruit stage of maturity at the time of harvest.

Olive oil has a fatty acid composition similar to that of high-oleic seed oils such as sunflower, hazelnut and almond. However the total MUFA content is of much higher (70–80 g/100 g) than the other vegetable oils, such as canola (59 g/100 g), peanut (46 g/100 g), sunflower (32 g/100 g), corn (29 g/100 g), almond (28 g/100 g), or hazelnut (62 g/100 g) (Nicklas *et al.,* 2004). Yorulmaz *et al.,* (2013) showed that stearic acid, linolenic acid, and monounsaturated/polyunsaturated fatty acid ratio decreased while linoleic acid content increased with the maturation. Furthermore as shown by Diaz *et al.,* 2005 and Aranda *et al.,* 2004 the application of chemometric techniques based on triglyceride profiling can be used to distinguish between different olive oils derived from different cultivars.

1.5.2 Minor constituents

1.5.2.1 Phenolic Compounds

Phenolic compounds are a class of secondary metabolites found in plants distributed non-uniformly at both the cellular and subcellular level. They occur as both soluble and insoluble compounds. The majority of insoluble phenolics are the components of cell walls, while soluble phenolics are compartmentalized within plant cell vacuoles (Pridham, 1960). Phenolic compounds are transferred from the fruit to the oil during processing and thus analysis of olive oil phenolic compounds reflects those found in the drupes. The presence of phenolic compounds in olive are related to the organoleptic characteristics of the oil (Gutiérrez-Rosales *et al*., 2003). Their strong antioxidant activity protects the oil from auto-oxidation (Pellegrini *et al.,* 2001; Mateos *et al.,* 2005).

There are six major classes of phenolic compounds found in olive oil, these are phenolic acids, simple phenols, complex oleuropein derivatives, flavonoids, lignans and hydroxyl-isochromans. The individual composition of phenolic compounds found

in olives is affected by a number of factors including a combination of agronomic, climate, degree of ripeness, extraction technology and conditions of storage (Forcadell *et al.,* 1987; Saitta *et al.,* 2002; Pinelli *et al.,* 2003 and Van der Sluis, 2005).

Phenolic acids are aromatic secondary plant metabolites, with a widespread occurrence throughout the plant kingdom. As the name implies these phenols possess carboxylic acid functionality. There are two main distinct groups of phenolic acids, classified according to the constitutive carbon frameworks, namely the hydroxycinnamic and hydroxybenzoic structures. A number of hydroxybenzoic acids have been identified throughout the years; these include: gallic, protocatechuic, *p*-hydroxybenzoic, vanillic, and syringic acid. Similarly the hydroxycinnamic acids identified include *p*- and *o*-coumaric, caffeic, ferulic and cinnamic acid (Buiarelli, *et al.,* 2004; Cartoni, *et al.,* 2000; Carrasco Pancorbo *et al.,* 2004 Tsimidou 1998, 1999; Morales and Tsimidou 2000; Servili *et al.,* 2004).

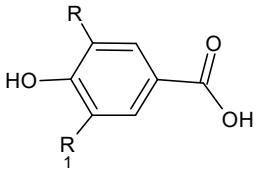
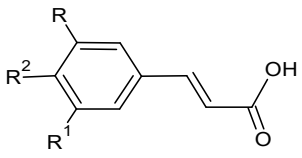
Class	Structure	Phenolic acid
Hydroxybenzoic acid		R_1=R=H➔*p*-Hydroxybenzoic acid R_1=H, R=OH➔Protocatechuic acid R_1=R=OH➔Gallic acid R_1=H, R=OCH_3➔Vanillic acid R_1=R=OCH_3➔Syringic acid
Hydroxycinnamic acid		R^1=R=H, R^2= OH➔ *p*-Coumaric acid R^1=R^2=H, R= OH➔ *o*-coumaric acid R_1=H, R^2=R=OH➔ Caffeic acid R_1=H, R^2= OH, R=OCH_3➔ Ferulic acid

Figure 1.1: Major classes of phenolic acids present in olive oil.

Simple phenolic alcohols are also known as phenylethanoids. These compounds have a simple phenolic basic structure which bears an aliphatic alcohol

moiety on the phenolic ring. A number of simple phenolic alcohols have been identified and to date most of them are structurally related to tyrosol (4-(2-Hydroxyethyl)phenol) *p*-HPEA. These include: tyrosol acetate; (3,4-dihydroxyphenyl) ethanol, hydroxytyrosol, 3,4 DHPEA; hydroxytyrosol acetate (Montedoro *et al.,* 1992, Tsimidou *et al.,* 1992; Angerosa *et al.,* 1995, Cortesi *et al.,* 1995, Pinelli *et al.,* 2003, Artajo *et al.,* 2007); and (3,4-dihydroxyphenyl) ethanol-glucosides (Bianco *et al.,* 1998). Both tyrosol and hydroxytyrosol are considered as the main simple phenols which are found in olive oil. Similar to other phenolic compounds the concentration of these compounds within the fruit depends on genetic factors, such as the producing cultivar, but also other factors, including fruit maturity and agropedoclimatic conditions (González-Rodríguez *et al*., 2003 and 2004).

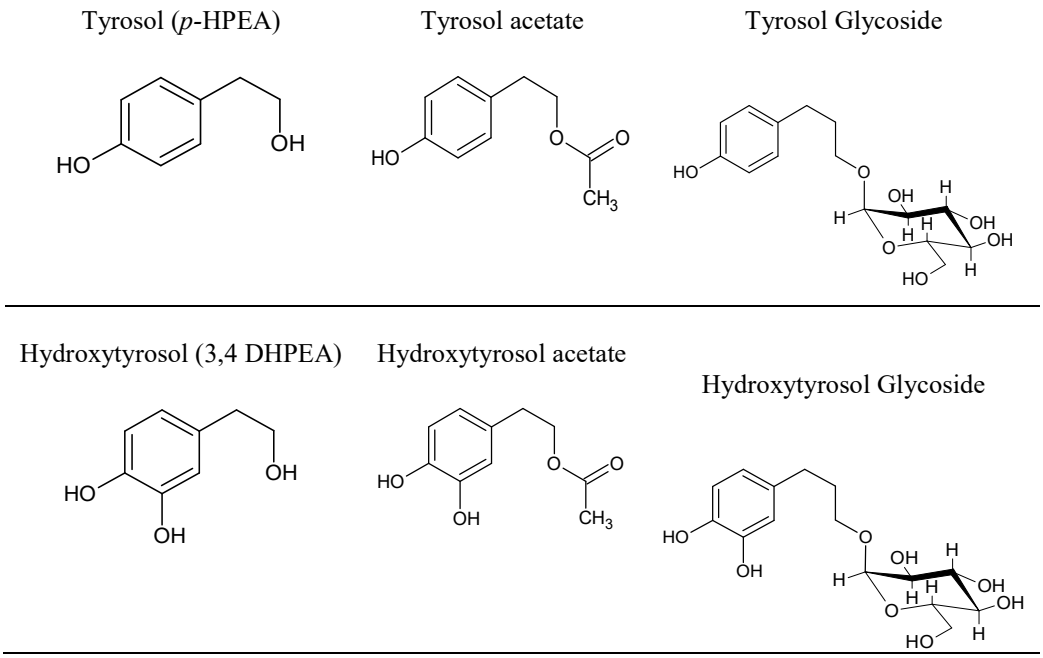


Figure 1.2: Simple phenols present in olive oil and their acetate and glycoside derivatives

Flavonoids are compounds that share the typical 15 carbon skeleton (C_6-C_3-C_6). These can be further subdivided into four major subclasses depending on the oxidation state of the pyran ring. Flavonoids are largely planar molecules and their structural variation derives in part from the pattern of modification by hydroxylation, methoxylation, prenylation, or glycosylation. The most commonly reported flavonoid compounds found in olive oil are luteolin and apigenin and their corresponding 7-monosubstituted glycosides (Rovellini *et al.,* 1997; Vázquez-Roncero *et al.,* 1976;

Carrasco Pancorbo *et al.,* 2004). These belong to the flavone class of the flavonoids and are characterised by unsaturated C-rings that connect the A and B rings in a single conjugated system. Apart from flavones, to date only one flavanonol, taxifolin, was also identified in Spanish olive oil (Carrasco Pancorbo *et al.,* 2004) and Algerian olive oil (Douzane *et al.,* 2013). Catechin (Gilani *et al*., 2006), quercetin (Lamzira *et al*., 2014) and rutin (Dağdelen *et al*., 2013 and Silva *et al*., 2006) have not yet been found in olive oil however these compounds were found within the actual olive fruit, whilst diosmetin was only tentatively identified within the leaves (Peralbo-Molina *et al.,* 2012).

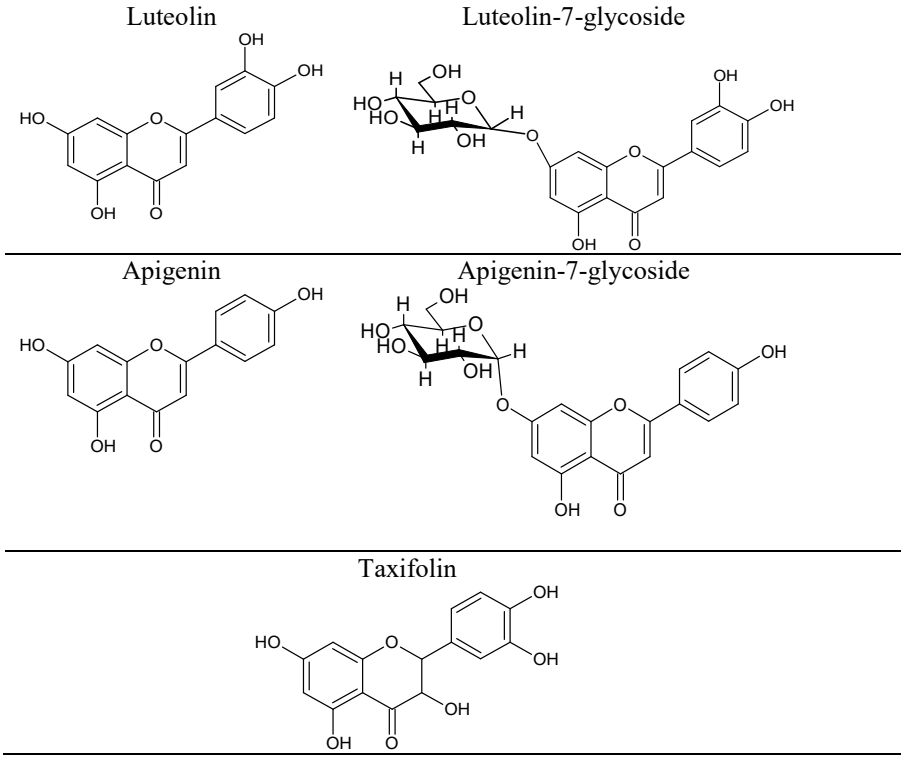


Figure 1.3 Flavonoid compounds and glycoside derivatives found in olive oil

Complex oleuropein derivatives also known as secoiridoids are the most diverse and abundant class of phenolic compounds found in olive oil and contribute to the bitter taste of olive oil. Secoiridoids are a type of monoterpenes derived from geraniol. These are characterized by the presence of either eleanolic acid or eleanolic acid derivatives in their molecular structure (Garrido Fernández *et al.,* 1997). The most commonly known secoiridoids are oleuropein and ligstroside which are most

commonly found conjugated in a dialdehydic form of elenolic acid connected to hydroxytyrosol or tyrosol and oleuropein aglycone. These compounds were first identified by Montedoro *et al*. (1992), who also assigned their chemical structure (Montedoro *et al,*. 1993) on the basis of IR, NMR and UV analysis. Later these structures were confirmed by other authors (Angerosa *et al.,* 1995) by the use of GC-MS, prior to subjecting them to derivatisation processes. Recent studies revealed that both oleuropein and ligstroside aglycones are present in olive oil however they are found in significantly lower quantities than their conjugated derivatives (Owen *et al.,* 2000 and Perri *et al.,* 1999).

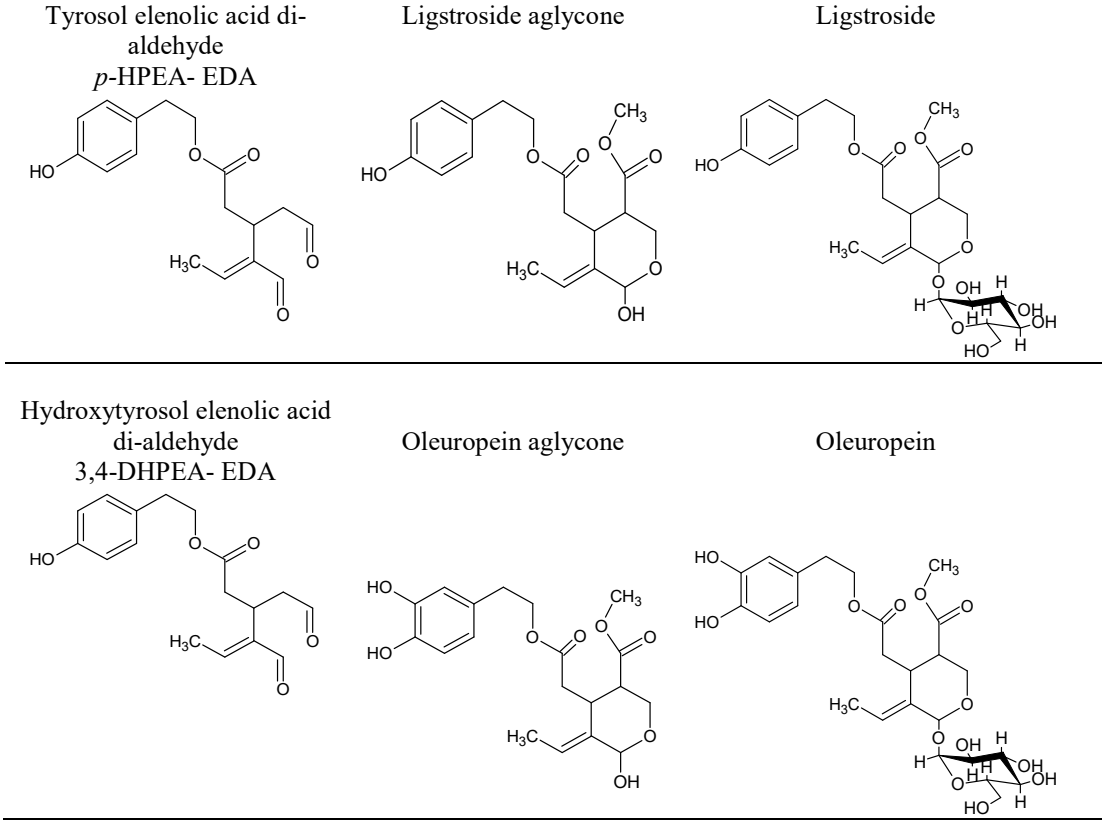
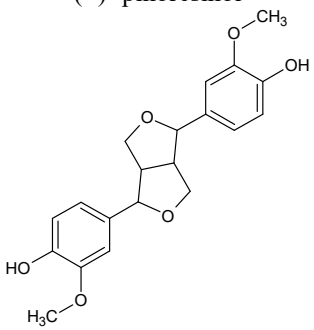


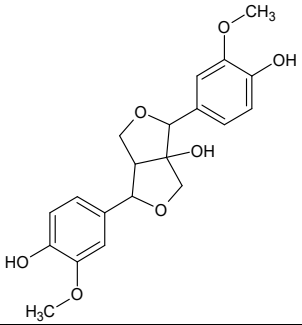
Figure 1.4: Simple and complex secoiridoid compounds found in olive oil

Lignans are polyphenolic substances derived from phenylalanine and the dimerization of substituted cinnamic alcohols with the aid of oxidative enzymes known as dirigent proteins. The most ubiquitous lignans found in olive oil are (+)-1-acetoxypinoresinol, (+) - pinoresinol and (+)-1-hydroxypinoresinol. These were identified and characterised by Owen *et al.,* (2000) and Brenes *et al.,* (2000). The latter suggested that the presence of 1-acetoxypinoresinol can be used as an authentication marker for the Spanish cultivar Picual (Brenes *et al.,* 2002).

(+)- pinoresinol



(+)-1-hydroxypinoresinol



(+)-1-acetoxypinoresinol

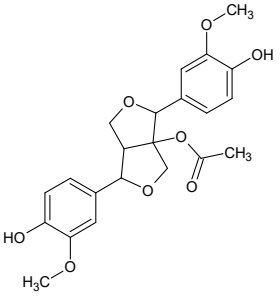
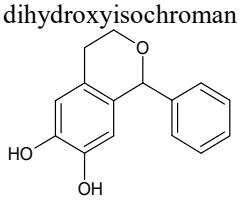


Figure 1.5: Lignans present in olive oil

The most recently discovered class of phenolic compounds found in olive oil are isochromans and their hydroxyl derivatives. These compounds are 3,4-dihydro-^{1}H-benzo[c]pyran derivatives which are present in nature most of the time forming part of fused intricate ring systems (Peng, Lu, and Ralph, 1999). These compounds are formed during the oil preparation especially during the process of kneading of the olive paste, which causes uncontrolled enzymatic hydrolysis. The presence of glycosidases and esterases present within the olive paste degrade secoiridoid favours the release of simple phenols and isochroman derivatives (Bianco *et al.,* 2001). Bianco *et al.,* identified two major isochromans namely 1-phenyl-6, 7-dihydroxy-isochroman and 1-(39-methoxy-49-hydroxy) phenyl-6, 7-dihydroxy-isochroman.

1-Phenyl-6,7-dihydroxyisochroman



1-(39-methoxy-49-hydroxy) phenyl-6, 7-dihydroxy-isochroman.

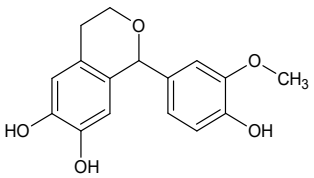


Figure 1.6: Isochromans present in olive oil

1.5.2.2 Antioxidant activity of phenolic compounds

Polyphenols are an important group of natural compounds, which are produced in the secondary metabolism of many plants in nature. The best-known characteristic of polyphenols is certainly their antioxidant capacity. This antioxidant capacity is attributed to their ability to donate a hydrogen or transfer an electron and/or to delocalize the unpaired electron within the aromatic structure (Bors *et al*., 1990). Polyphenols act as antioxidant compounds, as they are able to break free alkylperoxyl radicals produced by lipid oxidation chain reactions, themselves forming stable derivatives. The antioxidant activity of the phenolic fraction of olive has been extensively studied. The majority of the published studies focus on the relationship between the phenolic constituents and their effect on the stability of olive oil over time using two main accelerated oxidation methods, AOM (Active Oxygen Method) and Rancimat (Brenes, *et al.,* 2000; Baldioli *et al.,* 1996; Pirisi *et al.,* 2000), which provide insights about the shelf-life of the product. The relationship between the amount of phenolic compounds, the antioxidant activity and the stability of olive oil was confirmed from a number of studies (Lonso *et al.,* 2003; Lee *et al.,* 2007; Franco *et al.,* 2014; Servili, *et al.,* 2009).

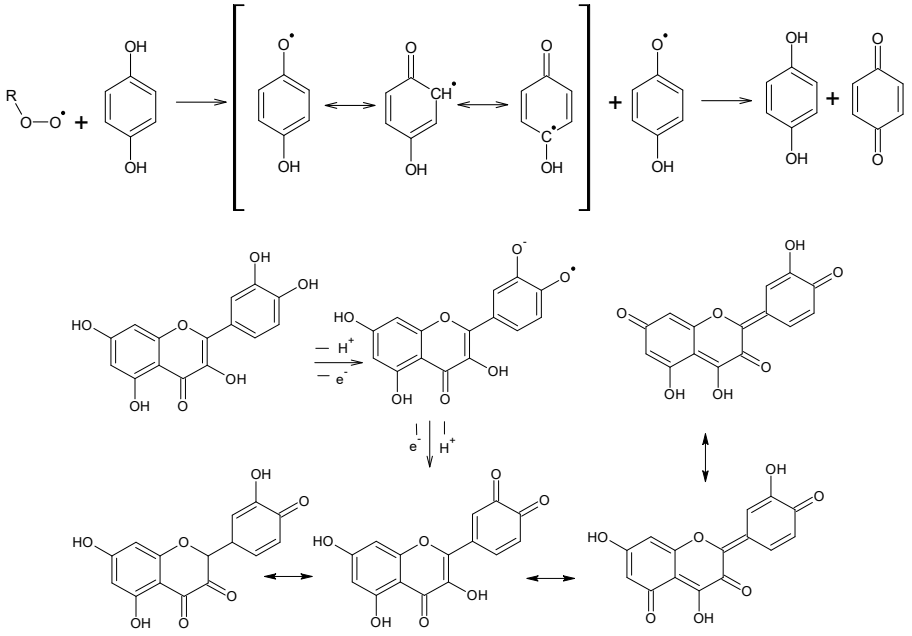
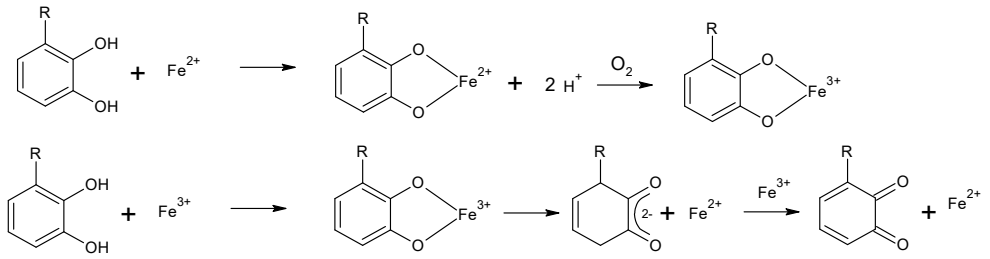


Figure 1.7: Free radical stabilisation mechanism carried out by phenolic compounds (Top) phenolic acids (Bottom) flavonoid compounds.

Apart from scavenging free radicals the phenolic compounds are able to act as antioxidants by chelating metal ions such as iron and copper which act as catalysts, aiding in the formation of oxygen radicals via the Fenton reaction (Yamamoto and Niki 1988; Yamauchi *et al.,* 1988; Benjelloun *et al.,* 1991; Mei *et al.,* 1998). Although the phenolic compounds are capable of chelating the ferrous and coppers metal ions, the strong reducing activity of phenolic compounds can ultimately cause the reduction of the oxidized counterparts of these ions, causing the formation of more catalytic metal ions; this effect is known as pro-oxidation (Brune *et al.,* 1 991; Dieana *et al.,* 1995; Solinas *et al.,* 1996; Moran *et al.,* 1997). Studies showed that lipid peroxidation is accelerated in the presence of both ferric ions and reducing antioxidants such as α-tocopherol and ascorbate which reduce Fe^{3+} to the Fe^{2+} state catalysing hydroperoxide decomposition (Yamamoto and Niki, 1988). Studies done by Keceli and Gordon (2002) showed that olive phenolic compounds are no exception. It was demonstrated that in the absence of ferric metal ions, the phenolic compounds displayed antioxidant activity preventing the degradation of α-tocopherol stripped sunflower oil, however, in the presence of ferric ions the phenolic extract showed a pro-oxidative effect accelerating the decomposition.



Fe^{2+}or Cu^{+} + H_2O_2 + H^{+} ➔ Fe^{3+}or Cu^{2+} + $OH^{\bullet}$ + H_2O

Figure 1.8: The prooxidant mechanism of phenolic compounds in the presence of metal ions.

1.5.2.3 Sterols

There are four main classes of sterols which are found in olive oil: 4-α-desmethylsterols, 4-α-methylsterols, 4, 4-dimethyl sterols or triterpene alcohols and triterpene dialcohols. Although they share same basic steroidal carbon skeleton, they differ in the level of unsaturation of the B ring and the presence of side groups on the cyclopentane C ring.

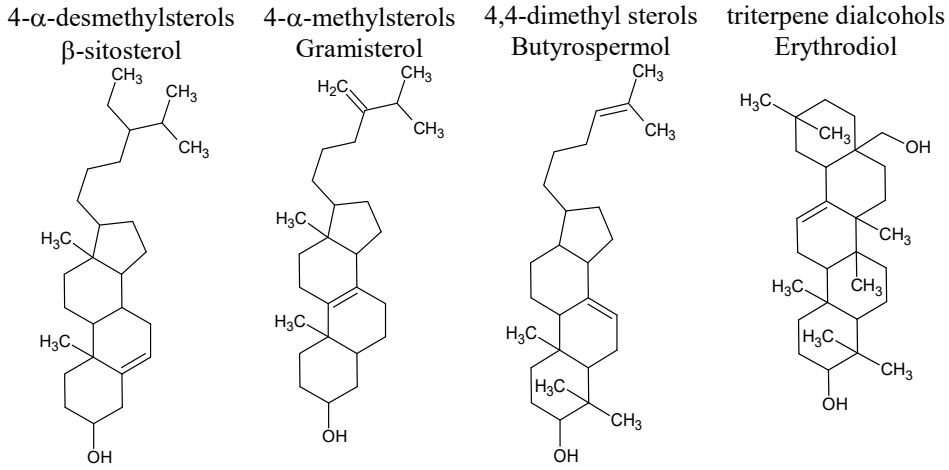


Figure 1.9: Four main classes of sterols present in olive oil 4-α-desmethylsterols (β-sitosterol), 4-α-methylsterols (Gramisterol), 4, 4-dimethyl sterols (Butyrospermol) and triterpene dialcohols (erythrodiol)

4-α-desmethylsterols are the most abundant sterols present in olive oil and their content ranges from 100 to 200 mg/g. The main 4-α-desmethylsterols in olive oil are β-sitosterol (75–90 %), Δ^5-avenasterol (5–36 %) and campesterol (3 %) (Boskou 2002; Jiménez de Blas and del Valle González, 1996; Itoh *et al.,* 1981)). The presence of 4-α-desmethylsterols provides the olive oil with a higher oxidative capacity when compared to other seed oils. It was shown that olive oil was the most stable and resistant to oxidation during deep frying (Blekas and Boskou, 1999; Bastida and Sanchez-Muniz, 2001; and Naz *et al.,* 2005). This was attributed to the presence of Δ^5-avenasterol, its low iodine value and the presence of phenolic compounds. The presence of ethylidene side chain in Δ^5-avenasterol retards the oxidative polymerization in triacylglycerols when subjected to heat treatment (Blekas and Boskou, 1999).

The most abundant 4-α-methylsterols found in olive oil are obtusifoliol, cycloeucalenol, gramisterol and citrostadienol. These act as intermediate compounds in the biosynthesis of 4-α-desmethylsterols. Their isolation and quantification proved to be very difficult and thus their content in olive oil can only be approximated. Boskou, 1996 approximated that their content ranged from 20–70 mg/100 g based on gas chromatography against internal standards combined with thin layer chromatography.

The most abundant 4,4-dimethyl sterols or triterpene alcohols present in olive oil are α- and β-amyrin, cycloartenol, butyrospermol, 24-ethylenecycloartanol, taraxerol, dammaradienol and 24-methylene-24-dihydroparkeol (Boskou, 2002;

Kiritsakis *et al.,*2003; Paganuzzi 1982) whose content ranges between 100 and 150 mg/100 g oil (Kiosseoglou *et al*. 1987). Due to their high complexity, it has been proposed that the triterpene alcohol or dialcohols composition (Eisner, 1965; Spencer, 1979) could be used as tools for the detection and identification of pomace oil in commercial olive oils. The major triterpene dialcohols present in olive oil are erythrodiol and uvaol. Their total content in olive oil ranges from 1 to 20 mg/g, however in refined olive oil their content increases significantly up 280 mg/100 g. Therefore their quantification can be used to detect the presence of refined oil in virgin olive oil (Boskou, 2002; Mariani *et al*., 1987)

1.5.2.4 Chlorophylls and Carotenoids

Chlorophyll pigments are responsible for the green hue of olive oil. There are a number of different chlorophyll pigments present in olive oil, depending on the age of the oil. The freshly pressed olive oil contains a significantly higher amount of chlorophyll a, however its content is significantly reduced via the deterioration to pheophytin a via removal of the Mg^{2+} ion, and the latter is the major chlorophyll pigment which is found in packaged olive oil. The first high performance liquid chromatographic studies carried out by Minguez-Mosquera *et al.,* (1992) and Gandul-Rojas and Minguez-Mosquera (1996) on monocultivar olive oil from Spain revealed the presence of chlorophyll a, chlorophyll b, pheophytin a and pheophytin b, however, studies done by Psomiadou and Tsimidou, 2001 on monocultivar olive oils grown in Greece found no chlorophyll a and only traces of chlorophyll b and pheophytin b. Cerretani *et al*, 2008 showed that chlorophyll composition did not depend on the olive cultivar or growing region. However, a number of studies have shown that via the application of chemometric techniques to HPLC data it was possible to distinguish between oils derived from different cultivars (Cichelli and Pertesana, 2004; Giuffrida *et al.,*2007).

The presence of chlorophyll pigments in olive oil has both an antioxidant effect and a prooxidant effect depending on the presence of light. In the presence of light the chlorophyll pigments have been shown to have a prooxidant effect (Endo *et al.,* 1984; Wanasundara *et al.,* 1993). This is attributed to the fact that chlorophylls and their degradation products pheophytins and pheophorbides, act as sensitizers to produce oxygen radicals in the presence of light and atmospheric oxygen. The

presence of these compounds increases the rate of oil oxidation (Whang and Peng, 1988). According to Endo *et al.* (1984), the photosensitizing activity of pheophytins is higher than that of chlorophylls and lower than that of pheophorbides.

The main carotenoids present in olive oil are β-carotene and lutein and these compounds impart a yellow hue to the oil. Other minor oxygenated carotenoids (xanthophylls) have also been reported. These include neoxanthin, violaxanthin, anteraxanthin and β-cryptoxanthin (Gandul-Rojas and Minguez-Mosquera 1996; Roca and Minguez-Mosquera, 2001). Similar to the chlorophyll composition, carotenoid composition is dependent on the fruit variety, the growing region, the degree of fruit ripeness, the extraction process, and the storage conditions of the oil (Serani and Piacenti 1992; Minguez-Mosquera *et al.,* 1992). The chlorophyll and the carotenoid concentration in olive fruits decreases with ripening whilst the synthesis of anthocyanins increases, thus olive oil derived from fruits which have a low index of maturation have a higher concentration of pigments. Moreover, oils obtained from a continuous system technology have a higher pigment content than those produced by the traditional pressure system.

Giuffrida *et al.,* (2007) proposed that the ratio lutein/β-carotene could be useful to differentiate oils derived from monocultivars. Roca *et al*., (2003) have shown that the chlorophylls/carotenoids ratio remained stable for one year of storage at 15 ºC in the dark and thus proposed that this ratio could be used as an authenticity parameter. Similar to chlorophylls, β-carotene can also act as an antioxidant and as a prooxidant. The antioxidant capability can be attributed to its ability in blocking singlet oxygen, inhibiting the oxidation of lipids (Beltra´n *et al.,* 2005; Hrncirik and Fritsche, 2005). However, in the presence of high concentrations of oxygen, carotenoids act as prooxidants via the generation of carotenoid radicals which react with atmospheric oxygen to generate carotenoid-peroxyl radicals. The latter compounds act as pro-oxidants by promoting oxidation of unsaturated lipids (Stahl and Sies, 1992).

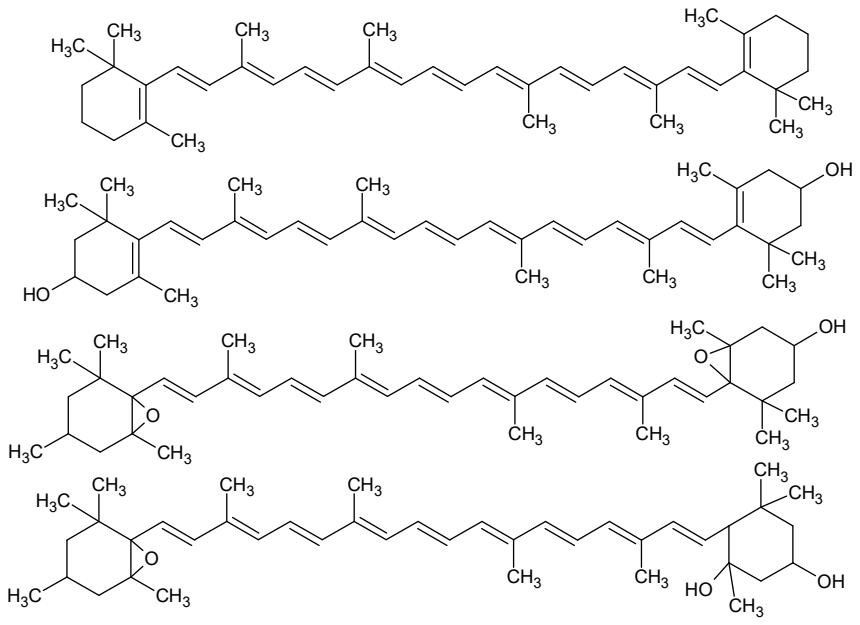


Figure 1.10: Carotenoids and their oxygenated derivatives xanthophyll present in olive oil (Top ➔Bottom) β- carotene, lutein, violaxanthin and neoxanthin

1.5.2.5 Hydrocarbons

The most abundant and important hydrocarbon present in olive oil is squalene. This is a highly unsaturated aliphatic triterpenic hydrocarbon which can constitute up to 90% of the hydrocarbon fraction of olive oil and up to 50% of the total unsaponifiable fraction (Aguilera *et al.,* 2005). Squalene is the most important antioxidant present in unsaponifiable fraction of olive oil (Manzi *et al.,* 1998; Mateos *et al.,* 2003). Apart from the presence of phenolic compounds present in olive oil, the significantly higher quantities of squalene have been also implicated in its major health benefits (Gapor and Rahman 2000). Its chemopreventive activity has been reported by a number of studies (Rao *et al.,* 1998; Smith *et al.,* 1998; Harold 2006, Gaforio *et al.,* 2015). The antioxidant activity of squalene has also been shown in the oil itself, however, it degrades faster than tocopherols on storage (Manzi *et al*. 1998) and thus according to Psomiadou and Tsimidou (1999), it only plays a limited role in olive oil stability.

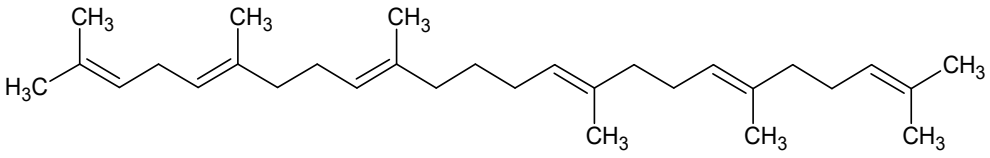


Figure 1.11: Squalene a naturally 30-carbon hydrocarbon compound present in olive oil

1.5.2.6 Waxes

Wax esters are formed by the esterification of high molecular mass alcohols with fatty acids. The most commonly found waxes in oil are C_{40}, C_{42}, C_{44} and C_{46} and these are mainly derived from seeds rather than from the fleshy mesocarp. The determination of wax content by gas chromatography (Pérez- Camino Cert, 1999; Ranalli *et al.,* 1999) enables the detection of lampante olive oil and seed oils adulteration due to the use of organic solvents. The presence of several types of esters (saturated and unsaturated, straight-chain, even-numbered esters) and also benzyl alcohol, phytyl and geranylgeranyl esters makes the wax fraction highly complex (Reiter and Lorbeer 2001).

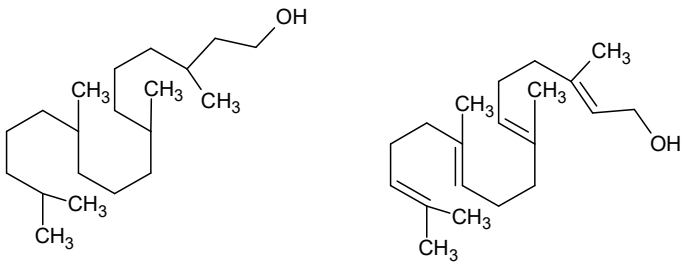


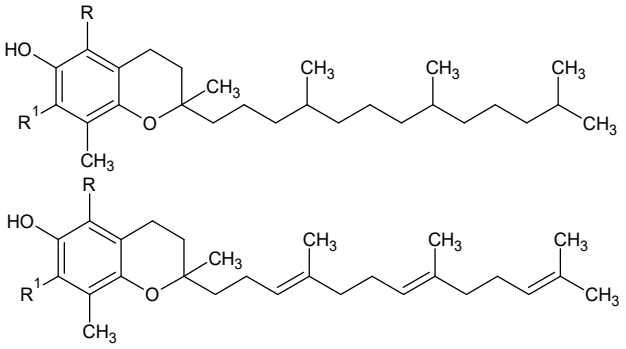
Figure 1.12: (Left) Saturated and (Right) unsaturated esters, responsible for the diversity of waxes

1.5.2.7 Tocopherols

Tocopherols are another important class of antioxidants which are found in olive oil. These compounds are collectively known as Vitamin E. There are four classes of tocopherols (α, β, γ and δ) based on the different side groups present on the chromanol ring and their corresponding unsaturated isomers known as tocotrienol. α-tocopherol constitutes up to 95% of the total tocopherol/tocotrienol content present in the virgin olive oil.

The biological activity of tocopherols has been demonstrated by a number of studies and ranges from chemoprevention (Campbell *et al.,* 2003) and prevention of

cerebral infarction (Mishima, *et al.,* 2003) to cholesterol synthesis and absorption (Theriault *et al.,* 1999). Due to their structural similarity the development of new analytical techniques which enable the separation and quantification has been accomplished by a number of researchers using techniques including HPLC (Gimeno *et al.,*2000) RP-HPLC (Gliszczyńska-świgło *et al.,*2007) UPLC (Cunha *et al* 2006) and GC-MS (Parcerisa, *et al.,*2000; Melchert *et al.,*2000). However, GC is normally disregarded due to the nonvolatile nature of these compounds, requiring derivatization prior to the quantification step. Similar to the phenolic compounds, the tocopherol and the tocotrienol profiles vary depending on cultivar, fruit ripening stage, edaphic and climatic conditions, and olive-growing techniques (Beltran *et al*., 2005; Benito *et al.,*2010 Nieves Franco *et al.,*2013).



	R	R_1
α	CH_3	CH_3
β	CH_3	H
γ	H	CH_3
δ	H	H

Figure 1.13: Tocopherols (Top) and unsaturated isomers Tocotrienols (Bottom) and there corresponding class based on the side chain functional group (Right)

1.5.2.8 Volatile compounds and aromas

Unlike other vegetable oils, olive oil possesses a unique characteristic - the presence of volatile and non-volatile flavours. This distinctive sensorial property can be attributed to a complex mixture of more than one hundred volatile compounds (Morales *et al.,* 1994; Vichi *et al.,* 2003) formed during the extraction process.

During the crushing step, the plant tissue is perturbed causing the release of enzymes, namely lipoxygenases. These enzymes cause the degradation of hydroxyperoxides formed during the oxidation of polyunsaturated fatty acids, causing the release of aldehydic compounds. The aldehydes which are formed are reduced to

their corresponding alcohols by the presence of reducing enzymes including dehydrogenases and transferases, and the ultimate products are scented hexyl esters (Angerosa *et al.,* 1999 and Angerosa, 2002). The volatile fraction of olive oil consists of a mixture of hydrocarbons, alcohols, aldehydes, esters, phenols, phenol derivatives, oxygenated terpenes and furan derivatives (Boskou, 1996; Reiners and Grosch, 1998; Morales and Aparicio, 1999; Morales and Tsimidou, 2000).

Analysis of the volatile profile of olive oil has been widely achieved using GC-MS by previous extraction and pre-concentration of these compounds (Cavalli *et al.,* 2003; Zunin *et al.,* 2004; Vichi *et al.,* 2007). The newly developed headspace solid-phase microextraction (HS-SPME) coupled with gas chromatography proved to quite reliable. Pizarro *et al.,* 2010 showed volatile profiles obtained using HS-SPME were able to discriminate between Spanish olive oils derived from different geographical locations. Dabbou *et al.,* 2011 was able to discriminate olive oils derived from cultivated olive trees and oleasters grown in Tunisia on the bases of volatile profiles.

1.5.3 Application of Chemometric techniques for the identification of origin

The determination of the origin and the authenticity of olive oils has been studied extensively in the past few years using extremely varied physical-chemical techniques in conjunction with chemometric methods. These studies can be classified into three main categories.

In target-based analysis, a specific metabolite or a class of metabolites is targeted. The metabolites need to be selectively extracted and isolated from the matrix and other metabolites present within, with the ultimate aim of concentrating the selected metabolites and minimising interference from other compounds. From a chemical perspective these studies require the need of separation based analytical techniques including HPLC (Cichelli and Pertesana, 2004) and UPLC (Herrero *et al*., 2011) and GC (Ranalli *et al.,* 1999), coupled with a detector suitable for the quantification and identification of the compounds of interest including UV (Cichelli and Pertesana, 2004), NIR (Galtier *et al.,* 2007), NMR (Alonso-Salces *et al.,* 2010) and MS (Suárez *et al*., 2008). Typical targeted compounds include fatty acids, triacylglycerols (Bucci *et al.,* 2002; Ollivier *et al.,* 2003), sterols (Leardi and Paganuzzi 1987) phenolic compounds (Faouzia *et al.,* 2008; Talhaoui *et al.,* 2015; Gong *et al.,* 2003) aliphatic alcohols (Rigane *et al.,* 2011) and inorganic multi-elemental composition (Beltrán *et al.,* 2011).

Similar to target based analysis, metabolic profiling requires the identification and quantification of a number of selected metabolites, belonging to various classes of compounds. However unlike target based analysis the metabolites are analysed without a separation procedure (Vigli *et al*., 2003; Goodacre *et al*., 2002 and Garcia-Gonzalez *et al*., 2004).

The third analytical category is known as metabolic fingerprinting (Fiehn, 2002). In this category, studies are performed without any preliminary identification of individual metabolites and rather than attempting to decipher the individual metabolites, the whole interactions among the different unknown metabolites is taken into consideration (Nicholson, 2008). Any fingerprinting technique can produce a spectrum that can be considered the whole trace of the product.

Chemometrics involves the application of multivariate data analysis and mathematical methods to extract relevant information from chemical data (Gasteiger and Engel, 2003). Multivariate analysis enables the comparison of x number of X (X_x) variables with a y number of Y variables (Y_y), compared to univariate statistics which compares a single independent x variable with a single dependent y variable. The application of multivariate analysis is advantageous when compared to the traditional univariate analysis as it gains a new and higher quality in data evaluation. Chemometrics vary from the standard statistical analysis in that in most cases the number of variables is often much higher than the number of objects, since the variables tend to be highly correlated. This causes traditional statistical methods to fail on using this kind of data. This form of data is often referred to the 'small N large P problem'. It is typical of recent data analysis areas such as bioinformatics and genomics. Several approaches have been developed to overcome the multicollinearity problem and these include the elimination of some predictors through the use of stepwise methods, and the use of ridge regression, such as principal component regression (PCR), both of which will be presented at a later stage in this study. The use of specialised methods capable of handling large amounts of data have been developed in the field of chemometrics. These include the use partial least-squares (PLS) regression, artificial neural networks (ANN), stepwise linear canonical discriminant analysis (SLC-DA) and soft independent modelling by class analogy (SIMCA).

The use of multivariate techniques can be divided into two, supervised and unsupervised techniques. The supervised techniques involve the use of a subset of the data which is used for learning whereby the algorithm is fitted in order to satisfy the response, while the remaining subset of data is used for the model evaluation. Unsupervised techniques identify the natural clustering pattern on the basis of similarities among the samples without prior knowledge of the classification. The most common method of unsupervised pattern recognition are cluster analysis (CA), principal coordinate analysis (PCoA) and PCA which are recognised as very powerful statistical tools for extracting information within a dataset (Forina, Oliveri, Lanteri, and Casale, 2008; Poulli, Mousdis, and Georgiou, 2005).

PCA is a multivariate projection method, generally used in chemometrics to compress large dimensional data into a smaller-dimensional space with the smallest loss of information by extracting systematic variations in a data set (Aguado *et al.,*

2008). The purpose of PCA is to represent as much of the variation as possible in the first few axes. To do this the data is first centred so that the majority of the variables have a mean of zero. The data is then rotated in order to obtain as much variation as possible on the first axis whilst the second axis contains as much of the remaining variation. Plotting the first two axes will give a plot which would represent the maximum variation in two dimensions PC1 and PC2.

Discriminate analysis (DA) is a statistical technique used to classify observations into mutually exclusive and exhaustive groups based on a set of measurable parameters. The discriminate analysis is a term used to refer to multivariate statistical methods that are used to classify observations into a known number of groups. These methods involve a two-step process: the first step involves the discrimination between the different groups present through a linear transformation of the samples into a new dimensional space, and in so doing samples belonging to the same *class* are close together but samples from different classes are far apart from each other. The discriminate model that is built is then followed by a classification step. Examples of DA method are Fisher Linear Discriminant analysis and canonical discriminant analysis. The problems which are associated with these forms of discriminate analysis is that they require a larger sample size than the number of predictors thus these cannot be used for data sets in which $n<p$. However, the application of stepwise methods enables the use of variables which are only significantly affecting the discrimination. Discriminant Analysis is considered supervised learning since it explicitly makes use of the groupings in developing the functions; unlike PCA and PCoA the group membership must already be known prior to initial analysis. Discriminant analysis can also be used in order to determine the variables which are responsible for the group differences.

The main aim of PLS regression is to construct an adequate mathematical model which explains or describes the relationships that may exist between two or more explanatory variables and a response variable. The model that is built would enable the estimation of unknown values of the response. PLS regression analysis is used to extract latent factors that account for much of the apparent factor (Zheng and Lu, 2011). Similar to PCA. PLS can be thought of as a dimension reduction technique which is applied to the data prior to conducting classical multivariate techniques such as multiple linear regression and discriminant analysis. The multitude of signals

derived from each overlapping variables can be analysed using powerful multicomponent analysis such as PLS regression (Fuller and Griffiths, 1978; Haaland and Thomas, 1988). This technique allows a sophisticated statistical approach which is able to fit a mathematical algorithm using the full or partial spectral region rather than distinct wavelengths. The mathematical algorithm is fitted based on the ability to correlate the chemical response variable data to a property matrix of interest. At the same time, the mathematical algorithm accounts for all other significant spectral factors that perturb the data, for example spectral noise (Liang and Kvalheim 1994).

Artifical neural networks are non-linear mathematical algorithms which are fitted on a subset of data with the aim to transform the input information into the output one. Throughout the training process each variable is weighed in order to produce the correct output value which is as close as possible to the target value (Kruzlicova *et al*., 2009). Compared to other linear models of regression like PLS, the main advantage of ANN is that it allows the high performance modelling, of nonlinear sensor responses, very much related to human pattern recognition (Cetó *et al.,* 2013).

Cluster analysis is a multivariate analysis which, unlike DA, attempts to identify determined groups of observations, without any previous information about any observation or grouping membership, and generally not even the amount of clusters is known (Everitt, 1974; Gordon, 1999; Kaufmann and Rousseeuw, 1990; Massart and Kaufmann, 1983; Ripley, 1996). Cluster analysis can be also be employed in order to confirm the groups which were employed during the DA. The most commonly used clustering method is the hierarchical method, whereby the observations are partitioned in hierarchal fashion in the form of a dendrogram. The optimum number of clusters and the relationships between each observation are manually determined by the observer.

1.5.4 Biomolecular and genetic studies used to distinguish *Olea europaea* cultivars.

Over the centuries, through vegetative propagation, humans have selected olive trees based on the quality and quantity of the olive. Although this procedure should conserve the genetic identity of the parent stock, this did not avoid the problematic nature of crossing between the newly introduced cultivars and the local germplasm. The probability of somatic mutation and inherited genetic variability makes cultivar identification challenging, as described by several authors (Angiolillo *et al.,* 1999; Bautista *et al.,* 2003; Belaj *et al.,* 2002, 2003, 2004, 2006; Cordeiro *et al.,* 2008; Gemas *et al.,* 2000; Gomes *et al.,* 2008, 2009; Martins-Lopes *et al.,* 2007, 2009; Sefc *et al.,* 2000). Since olive trees are wind pollinated self-incompatible flowers, pollination can only occur by means of cross-pollination. This allogamous type of pollination increases the genetic variability between and within the cultivars (Mekuria *et al.,* 1999; Ouazzani *et al.,* 1996; Zohary, 1994). It is believed that a small percentage of progeny arose from self-pollination, even when a cultivar is considered to be self-incompatible, however, seed genotyping showed that they were products of cross-pollination in almost all cases (Diaz *et al.,* 2006).

Olea europaea subsp. *europaea* is present in two forms, namely the wild oleaster (var. *sylvestris*) which is haploid, and the cultivated (var. *europaea*), which is diploid, (Breviglieri and Battaglia 1954) with a genome size range between 2.90 pg/2C and 3.07 pg/2C, with 1C = 1,400-1,500 Mbp (Loureiro *et al.,* 2007). These values correspond to ≈ 3120 Mbp because 1pg is ≈978 Mbp based on the conversion factor proposed by Dolezel *et al*. (2003). The wild oleaster is thought to be indigenous to the Mediterranean basin (Green, 2002). Although the oleaster form is is similar to the cultivated form, it differs by having spinescent juvenile shoots, smaller fruits with a higher stone/mesocarp ratio, and relatively low oil content (Zohary and Spiegel-Roy, 1975; Lumaret *et al*., 2004). Cross-pollination between the two varieties is also possible and may produce fertile offsprings, providing access to an enormous pool of genetic variability (Muleo *et al.,* 2012). The application of biomolecular techniques, namely genetic techniques, seems to be the only possibility for identifying the cultivar and the oil deriving from it (Busconi *et al.,* 2003). Molecular biology research on olives can be mainly divided into two major disciplines: genome and genetic diversity

studies, or gene characterization and functional genomics (Banilas and Hatzopoulos, 2013).

1.5.4.1 Phenotypic and morphological analysis

Morphological, phenotypic agronomic characteristics have been widely used for distinguishing olive cultivars (Barranco and Rallo, 1985, Cantini *et al.,* 1999 and Barranco *et al.,* 2000; Barranco and Rallo, 2000 and Del Rio, 1994). Rugini and Lavee, (1992) managed to describe 2600 cultivars of *Olea europaea* using morphological analysis, however many of them might be synonyms, homonyms, ecotypes or the result of crosses between olive cultivars (Barranco *et al.,*2000). Bartolini *et al*. (1998) and Barranco *et al*. (2000), stated that biometric indexes used for classification should always be accompanied by a detailed morphological description of the organs (inflorescence, leaf, fruit, and stone) of olive varieties following the International Union for the Protection of New Varieties of Plants method. The recognition of olive cultivars based on morphological and phenotypic characteristics is problematic, especially when the trees are still juvenile.

The application of morphological analysis for cultivar identification is very laborious. Furthermore, the presence of environmentally dependent polymorphs makes identification more complicated (Mohan *et al.,* 1997; Tanksley and Orton, 1983; Caraffa *et al.,* 2002). Researchers have shown that different cultivars are morphologically variable based on geographical locations and under different pedo-climatic conditions (Grati *et al.,* 2002, Guerfel *et al.,* 2009 and Youssefi *et al.,* 2011). Although nowadays it seems that the use of morphological classification is becoming less important, as it has been replaced by more rigorous and less time-consuming biomolecular techniques, its importance is still significant. Al-Ruqaie *et al.,* (2016) employed comparative morphological characteristics in order to distinguish between eight olive cultivars grown in Saudi Arabia, in conjugation with newly advanced pattern recognition statistical software. The advanced multivariate analysis software revealed that biometrics of leaves, fruits and seeds can be used to distinguish between the different varieties (Al-Ruqaie *et al.,* 2016).

1.5.4.2 Isozymes and allozymes analysis

Isozymes are multiple forms of the same enzyme that differ only in the amino acid sequence but still retain the same chemical catalytic activity. These are coded by the same genes which diverged over the time. Allozymes are enzymes which are derived from different alleles of the same gene. Although isozymes and allozymes are not the same the two words are most of the times used interchangeably. They are usually separated by means of electrophoresis, with minimal preparation. Since enzymes are coded by genes, the analysis of these enzymes provides insights of gene expression. Compared to morphological studies enzymatic expression studies permit unequivocal identification of nearly all genotypes growing in different geographic locations (Wilson *et al.,* 1977; Kimura, 1983; De Vienne, 1984).

Various enzymatic expression studies have been carried out in order to distinguish between the cultivated *Olea europaea* cultivars (Ouazzani *et al*, 1993; Ouzzani *et al*, 1995; Lumaret *et al*, 1997). Enzyme polymorphism was also used in order to distinguish between wild and cultivated olives collected in several sites of the Mediterranean Basin, in order to assess the effect of olive domestication (Ouazzani 1994, Ouazzani *et al*., 1993, 1995, 1996). Pontikis *et al*. (1980) found high isoenzymatic variability in olive pollen samples of 27 cultivars, mostly of Greek origin. All cultivars could be identified with only two out of the sixteen enzyme systems used. In a similar study using pollen, isoenzymatic variability carried out by Trujillo and Rallo (1995) showed that a combination of different enzymes could be used to distinguish and identify 85% of the cultivars. The remainder were identified through morphological analysis. Relatively recent studies conducted by Lumaret and Ouazzani (2004) have shown that through the study of leaf allozymes, it was possible to discern two major oleaster populations, the eastern oleaster populations which gave rise to cultivated olive clones in the Mediterranean Basin, and the western populations that are related to the wild Canarian populations. Furthermore, this study also highlights the significantly lower heterozygosity in cultivated olive trees than in oleasters, indicative of an intensive selection process. The disadvantage associated with the use of enzyme analysis is that the banding patterns obtained are highly variable and dependent on the electrophoresis gel conditions (Trujillo and Rallo, 1995). Furthermore, these enzymes are unable to detect the genetic changes produced

by somatic mutations (DeWald *et al*, 1988; Tao and Sugiura, 1987; Weeden and Lamb, 1985).

1.5.4.3 DNA analysis

Although the enzyme profiling and morphological analysis offer a rudimental form of cultivar identification, the analysis of DNA markers is by far the most robust, as it is usually unaffected by environmental and developmental factors (Fabbri *et al.,*1995). However, selecting the right DNA marker is essential and depends upon the ultimate objective of the study. In general, the DNA markers chosen must have certain characteristics. In order for the study to be successful these include being highly polymorphic, evenly and commonly distributed throughout the genome, stable over cohorts, simple, quick and inexpensive, and requiring only small amounts of DNA (Agarwal *et al.,*2008; Hatzopoulos *et al.,* 2002). The application of genetic molecular markers has provided great insights into the geographical locations of early olive domestication and later genotype diffusion (Besnard *et al.,* 2013).

1.5.4.4 Non- PCR based techniques

1.5.4.4.1 RFLP

Restriction fragment length polymorphism (RFLP) analysis is based on the analysis of banding patterns derived from cleaved DNA using different restriction enzymes, which cleave DNA at a certain base pair sequence. The resulting fragments are then hybridized using specific probes (Mohan *et al.,* 1997). The use of RFLP on both the genomic and chloroplast DNA has been widely employed in order to assess the genetic linkage between the cultivated olive trees and the wild oleaster (Amane *et al.,* 1999; Lumaret *et al.,* 2000). Amane *et al.,* (1999) assessed the variation and relationship between old olive trees and oleasters, cultivated around the Mediterranean basin using RFLP from cytoplasmic DNA. Later, a similar approach was used to study chloroplast DNA variation in wild and cultivated Moroccan cultivars (Amane *et al.,* 2000).

A number of different variants of RFLP analysis have been employed throughout the years including the use of other less common sources of DNA. Besnard, Baradat and Bervillé (2001) employed the use of mitochondrial DNA in order to assess different mitotypes found in the Mediterranean. In a similar study, Rugini *et al.,* (2011)

showed the existence of three distinct mitotypes present in 37 different Italian cultivars. The use of mitochondrial DNA also provides insights into the maternal lineages involved in the evolution of olives. RFLP analysis using ribosomal rRNA has also been carried out, however the problem with using rRNA is that low gene diversity was observed (Besnard *et al.,* 2001). However, in the same study, this was overcome by using the Inter Gene Spacer (IGS), which has been reported to be the most variable rRNA (Rogers and Bendich 1987)

With the advance of PCR based methodologies, RFLP analysis markers are now not very widely used. This is mainly attributed to its time consuming laborious process and the use of radioactive and toxic reagents. Furthermore, in order to obtain consistent results, the method relies on a large quantity of high-quality DNA. The advances in molecular markers which do not require *a priori* sequence information for probe generation are ensuring the replacement of RFLP methods (Agarwal *et al.,* 2008).

1.5.4.5 PCR Based techniques

1.5.4.5.1 RAPD

The random amplified polymorphic DNA (RAPD) method was first described by Williams *et al*. (1990) and Welsh and McClelland (1990). It is based on the amplification of random DNA segments, by means of an arbitrary known nucleotide sequences which enable the detection of polymorphisms and the genetic map construction. The diverse DNA fragmentation pattern arises from the annealing of primers to multiple sites in different regions across the genome, generating numerous amplified products that are composed of repetitive DNA sequences (Paran and Michelmore, 1993). The first application of RAPD to olive tree germplasm was by Fabbri *et al*. (1995) who showed a high degree of polymorphism present in the germplasm. Furthermore, they managed to fully discriminate between 17 cultivars using very few primers. A large number of following studies carried out on different cultivars present in different countries also highlighted the high genetic diversity present in the olive tree genome or managed to fully discriminate between cultivars using RAPD in Iran (Shahriari *et al.,*2008), Spain (Belaj *et al.,*2002), Portugal (Cordeiro *et al.,*2008; Gemas *et al.,*2004; Martins-Lopes *et al.,*2007) Turkey (Wiesman *et al.,*1998; Çelikkol Akçay *et al.,*2014, Sesli and Yegenoglu, 2009, 2010)

Greek (Nikoloudakis *et al*., 2003) Italian (Besnard *et al.,* 2001) and Malta (Mazzitelli *et al*., 2015). Owing to its high discriminatory power, RAPDs were shown to detect intra-cultivar variability between olives as shown by Belaj *et al.,*(2004). On the contrary, in studies carried out by Çelikkol Akçay *et al.,*(2014) on Turkish cultivars, Gemlik showed minimal intra-cultivar variation.

Application of RAPD has also been extended in the determination of the geographical origin, Gemas *et al.,* (2000) showed that cultivars grown in Portugal were found to be genetically divergent according to their geographic location of origin. Banilas *et al.,* 2003 showed that clones of Ladolia, the main Cypriot cultivar, were highly morphologically and genetically diverse, suggesting that this cultivar originated from genetically distant landraces. The use of RAPD markers is not limited to the cultivated olive tree but was also applied to oleasters, Sesli and Yegenoglu 2010 showed a very high genetic diversity in oleasters grown in Turkey. Bronzini de Caraffa *et al.,* 2002 studied both cultivated varieties and the wild populations from Corsica and Sardinia. RAPD analysis revealed the existence of a genetic divergence between the oleasters and the cultivated varieties suggesting that the Corsican varieties were probably selected from local wild forms, contrary to the Sardinian varieties.

The use of RAPD is advantageous over non-PCR techniques as it is a low cost, time efficient technique which enables the researcher to access genetic variability with low amounts of DNA, even if not of good quality, without the need of a previous DNA sequence. However, the use of RAPD technology has been criticized a lot due to it low inter-laboratory reproducibility.

1.5.4.5.2 AFLP

The principal of amplified fragment length polymorphism (AFLP) is based on the amplification of DNA fragments produced by restriction enzymes (Vos *et al.,* 1995). AFLP employs the use of restriction enzymes in order to generate fragments allowing the simultaneous sampling of multiple loci distributed throughout a genome. The use of AFLPs has proven quite successful in the assessing the genetic diversity of the olive trees. Results obtained by a number of studies suggest that this technique is equally or more efficient than RAPD in intravarietal genotype discrimination (Belaj *et al*., 2003; Ercisli *et al.,* 2009; Grati-Kamoun *et al.,* 2006; Montemurro *et al.,* 2005). The first AFLP study on the genus *Olea* was carried out by Angiolillo *et al*. (1999),

who studied the genetic variations within and among populations of cultivated olives from different locations of the Mediterranean basin. The result showed a clear distinction between the cultivated and the oleaster germplasms. This study also showed that whilst all locations studied had similar levels of genetic variability, the cultivars from Sicily clustered in the same group which contained cultivars from Israel and Syria. In another study, Baldoni *et al.,* (2006) showed that oleasters and cultivated trees on the islands of Sicily and Sardinia were highly different, suggesting that all cultivars were introduced into these regions from the outside. The study also showed that cultivars found in mainland Umbria did not show large differences between the cultivated and wild types suggesting that these cultivars have originated either by selection from local oleasters or by a direct introduction from other regions. Similar results were obtained by Albertini *et al.,*(2011), who, using AFLP markers, showed that old varieties cultivated in the Abruzzo cluster together, suggesting that the oleaster population spread in central Italy via seed propagation.

AFLP has also been employed in the traceability of olive oil, Busconi *et al.,*(2003) reported that AFLP fingerprinting of olive oil is partially superimposable and in agreement with the cultivar from which the oil was derived. However, there are few studies which claimed non-concordant genetic profiles of olive oil and fruit (Doveri *et al.,* 2006; Pafundo *et al.,* 2005). According to Pafundo *et al.,*(2005) the results obtained from AFLP analysis from olive oil are not reliable due to low, degraded DNA content in olive oil and the presence of phenolic compounds and polysaccharides, which can inhibit the activity of both restriction enzymes and DNA polymerases. Doveri *et al.,* (2006) also suggest that the non-concordance is also due to the presence of paternal DNA from the embryos making such analyses difficult.

1.5.4.5.3 SSR

The simple sequence repeats (SSR) technique was described by Tautz *et al,* in 1986 and by Litt and Luty in 1989. This method depends on short repeating stretches of DNA, occurring in the genomes of higher organisms (Rafalski and Tingey, 1993; Wu and Tanksley, 1993). The application of SSR markers marked a new era in assessing the genetic diversity of olive trees. A large number of studies involving the use of SSR markers took over size-based molecular markers. Unlike RAPD and AFLP, microsatellites are codominant allowing the identification of alleles at a single locus.

These markers are highly reproducible within and among different laboratories and require a very small amount of DNA (Katsiotis *et al*. 1998; Cipriani *et al*. 2002). The main drawback of using SSR markers is that a previous DNA sequence is required for primer design, however knowledge about the sequence will enable development of a genetic map, application in linkage analysis, and fingerprinting studies (Bracci *et al.,* 2009; Cipriani *et al.,* 2002; De la Rosa *et al*., 2004; Gomes *et al.,* 2009; Karp *et al.,* 1996; Muzzalupo *et al.,* 2009; Rallo *et al.,* 2002; Sefc *et al.,* 2000)

A number of researchers managed to isolate sequences of DNA in order to develop successful primers for the SSR analysis of olives. Some of the most commonly used SSR primers include DCA (Sefc *et al.,* 2000) GAPU (Carriero *et al.,* 2002) UDO (Cipriani *et al.,* 2002) and EMO (de la Rosa *et al.,* 2002). These primers have proven quite successful in cultivar identification and genetic diversity studies which are useful for other oleaceous genera. The high discriminatory power of these markers was best illustrated by Sarri *et al.,* 2006, whereby a combination of three markers was used to differentiate 118 cultivars from different countries. Roubos *et al.,* (2010) showed that the application of SSR markers to Greek olive germplasms did not yield the same clustering results as those obtained when using morphological parameters. However, the dendrogram obtained displayed similarities with the dendrogram reported by Owen *et al*., 2005, based on AFLP. The comparative study made by Belaj *et al.,* (2003) showed that SSR markers had the highest index of polymorphism when compared to RAPD and AFLP, however, the expected heterozygosity and discriminatory power were comparable. Baldoni *et al.,* 2009 used, 37 SSR loci, in order to analyse 21 olive cultivars. The aim of their study was to develop a list of recommended markers for the creation of a universal molecular database of olive cultivars and protocols for olive.

The use of SSR markers was also used on chloroplast DNA, Hannachi *et al,* (2010) used 7 previously developed SSR markers in order to probe both the nuclear DNA and chloroplast DNA in Tunisian cultivated and oleaster populations. Nuclear genomic DNA analysis showed five mixed clusters of cultivar and oleaster trees suggesting a close relationship, and one oleaster single cluster. Chlorotype SSR markers revealed that olive trees present in Tunisia had three separate origins depending on the chlorotype.

The application of SSR markers was also extended to determine the botanical origin of the olive oil. Pasqualone *et al.,* (2004) studied the potential of SSRs in determining the cultivar origin of different olive oils. From the seven primer pairs used, six yielded amplified fragments of dissimilar lengths suggesting that SSR markers can be potentially employed in order to determine the cultivar from which the olive oil was derived. Studies done by Breton *et al*. (2004) involved mixing olive oils derived from different cultivars and later assess the efficiency of SSR markers in order to distinguish and identify the cultivar origins. The results obtained were partially successful as SSR identification of the major cultivar was possible, however, this did not apply to the minor ones. Other studies showed discrepancies in both the size and the number of the alleles which are present in olive oil when compared to the originating cultivar (Alba *et al.,* 2009; Ayed *et al.,* 2009; Vietina *et al*. 2011). This was attributed to the low DNA quality and the presence of extra genes derived from pollen.

1.5.4.5.4 SCAR

Sequence characterized amplified region (SCAR) involves sequencing of RAPD or AFLP fragments for further definition and design of more specific primers (Naqvi and Chatoo, 1996; Adam-Blondon *et al.,* 1998; Negi *et al.,* 2000). The application of PCR products makes this technique highly reproducible under different conditions (Lawson *et al.,* 1998). SCARs have been employed in olive trees for germplasm evaluation and mapping (Bautista *et al.,*2003; Busconi *et al.,*2006; Hernández *et al.,*2001), and for analysis of olive oil traceability (Pafundo *et al.,*2007). Hernández *et al*. (2001) were the first to apply SCAR markers to olive trees and were able to provide the first sequence of the intermediate RAPD region. Bautista *et al*. (2003) were able to identify 22 olive- tree cultivars using only 10 different, specific, repeatable markers. They demonstrated that SCAR markers are a simple, cheap, and reliable procedure for the identification of geographically related olive cultivars. Busconi *et al.* (2006) applied SCAR markers in order to evaluate the olive germplasm of forty different olive cultivars from Liguria and were able to fully distinguish 26 different cultivars, whilst the remaining clustered in one group, indicating a very high genetic similarity.

1.5.4.5.5 SNP

Single nucleotide polymorphisms are a marker system based on detection of variations in the genome at a single nucleotide base level. Such variations are present in large abundance in the genomes of higher organisms including plants. Such fine variation can differentiate between individuals within the same population (Agarwal *et al.,* 2008). This technique is still in its infancy as the olive genome is still unknown, making primer design rather difficult. The first study involving the application of SNP's was carried out by Reale *et al.,* (2006), who were able to fully discriminate 77% of the cultivars collected from different Mediterranean regions. However, the same cultivar derived from different sources was revealed as identical, demonstrating the utility of these markers as tools for resolving nomenclature issues, not for individual identification. Reale *et al.,* (2006), developed SNP markers using the sequence-related amplification polymorphism (SRAP) method whereby primers were constructed with the aid of olive gene sequences available in the GenBank database. Hakim *et al.* (2010) identified nine SNPs loci by partial sequencing of two genes in sixteen Tunisian cultivars and proposed the combination of both SNPs and SSR in order to increase the discriminatory power for complete cultivar identification. Recent studies conducted by Biton *et al.* (2015) applied the use of next-generation sequencing technology and managed to identify 145,974 SNP loci. From these, they employed 138 SNPs to analyze the genetic relationships between 119 cultivars, which constituted most of the Israeli olive germplasm collection. Furthermore, it was shown the clustering pattern was very similar between SSRs and SNPs suggesting that SSR markers are also very reliable.

1.5.4.5.6 Chlorophyll analysis

Apart from genomic DNA marker analysis, chloroplast DNA (cpDNA) has also been studied in a variety of olive trees. The use of cpDNA has been shown to be a powerful tool for phylogenetic reconstruction at both inter- and intra-species levels (Palmer 1987). Amane *et al.,* (1999) showed that the use of RFLP's from cpDNA enables the determination of the mode of inheritance of cpDNA in *O. europaea*. Results showed that in cpDNA there are three different sites corresponding to five distinct chlorotypes. Chlorotype I was predominant in both oleasters and cultivated olive trees, further confirming the relationship between the oleasters and cultivated

olives. Chlorotypes II, III and IV were found exclusively in secluded oleaster forests while chlorotype V contained three mutations located in discrete parts of the cpDNA. Besnard and Bervillé (2002) and Lumaret *et al.,* (2000) successfully managed to distinguish between the different North African subspecies *Olea europaea, maroccana, guanchica* and *laperrinei* using chloroplast DNA. Besnard and Bervillé (2002) concluded that amplification of RFLP markers revealed more polymorphisms in the cpDNA than the classical RFLP method of genomic DNA. Hannachi *et al.,* (2010) applied SSR markers to cpDNA for olives, revealing the presence of five clusters. In this study, cultivated and oleaster olive trees clustered together indicating a close relationship between the two. The chlorotype SSR markers revealed three probable olive origins: a CE chlorotype originating from the East of the Mediterranean basin, the haplotype CCK originating from the west coast of northern Africa and the COM chlorotype originating from West Mediterranean.

Methodology

2.1 Biomolecular analysis

2.1.1 Selection of Olive Varieties and Plant Leaf Collection

Plant leaf tissue was collected from three different classes of olive cultivars found in the Maltese islands; the indigenous class was composed of three different cultivars namely 'Bidni', 'Malti' and 'Bajda', the foreign locally grown class composed of three different cultivars which are commonly present in the Maltese islands that included Carolea, Frantoio and Pendolino. The 'Bidni' and the 'Bajda' cultivars are morphologically distinct and were identified onsite. The foreign cultivars were identified through there importation certificate. In the case of 'Malti' complete identification could not be carried as the cultivar is not yet defined. The last class contained the wild olive trees, the *O.europaea* var. *sylvestris* identified by its spinaceous shoots.

Two different kinds of 'Bidni' samples were collected, those which were derived from young grafts cultivated in a private estate in Wardija and very old trees from a secluded location in 'Bidni'ja from which the grafts were derived. The 'Malti' samples were derived from all over the island due to its wide distribution and the inherited possibility of the existence of morphotypes. Old 'Malti' olive trees present in Sant Anton, Lija, Mellieha, Mdina, and Wardija were also collected. In the case of ''Bajda',' one individual was obtained from a private estate in Wardija. Wild olive trees which were previously morphologically identified were collected from two locations, one of the specimens was collected from the University of Malta and three other specimens were obtained from the garigue area found in Manikata.

Where applicable, young leaves found near the apical buds were collected, stored in labelled plastic bags and placed over ice during transportation. Once in the lab, these were stored at -80ºC. Young leaves were preferred over older leaves due to their low concentration of polyphenolic compounds and polysaccharides, allowing better quality DNA.

Figure 2.1: Map of sites where plant samples were obtained.

2.1.2 Extraction of genomic DNA.

DNA extraction was performed according to the method described by Doyle and Doyle (1990) with minor modifications. Approximately 0.5 g of each leaf tissue sample were macerated in a mixture of acetone and dry ice and pulverized with the aid of a mortar and pestle. The resulting mixture was warmed in order to remove the solid dry ice and acetone. 5 ml of extraction buffer containing 100 mM Tris-HCl, 50 mM EDTA, 2 % CTAB (w/v) PVP, 2 % (w/v) and Tween 20, 0.2 % (v/v) was added.

The crushed leaves were further macerated in the presence of the extraction buffer and later transferred to 1.5 ml polypropylene tubes and incubated at 75 ° C for 1 hour in a water bath. The mixture was shaken periodically, opening the caps in order to prevent the build-up of pressure. The samples were then centrifuged at 13,000 rpm for 10 minutes at 30ºC in an Eppendorf Centrifuge. The supernatant was removed by the aid of a sterilized Pasture pipette and added to an equal volume of phenol:

chloroform: isoamyl alcohol in a ratio of 25:24: 1(Sigma) and the resulting mixture was shaken vigorously. The mixture was then centrifuged for 10 minutes at 13,000 rpm at 25ºC. The aqueous supernatant was transferred to a new tube and the extraction was repeated once again. The obtained supernatant was transferred to another tube and an equal volume of chloroform: isoamyl alcohol (24: 1) (Sigma), was added, shaken and centrifuged at 13,000 rpm for 15 minutes. The supernatant was collected and transferred to 1.5 ml tubes followed by 600 µL of cold isopropanol and 80 µL ammonium acetate (7.5 M) and agitated slightly.

The samples were incubated overnight at -20 ° C, and then centrifuged at 13,000 rpm for 15 minutes and 4ºC, the DNA pellet obtained was washed three times with 70% ethanol and dried at room temperature. Depending on the size, the pellet resuspended in 50-200 µL TE buffer consisting of 10mM Tris-HCl (pH 8.0), 1 mM EDTA.

2.1.3 Quantitative and Qualitative Analysis

The DNA concentration and quality were analysed with Thermo Scientific NanoDrop™ 2000. Compared to the conventional use of spectrophotometers the NanoDrop™ 2000 requires minute quantities (<1 µl) of the undiluted sample, allowing conservation of the extracted genomic DNA.

DNA concentration and purity assessments are based on the UV absorbance at three different wavelength maxima namely 280 nm, 260 nm and 230 nm. The maxima observed at 280 nm is attributed to the presence of proteins. The unsaturated aromatic ring bearing amino acids absorb at 280 nm, these include cysteine, tyrosine and tryptophan with an increasing molar absorption coefficients in that order (Gill and Von Hippel, 1989). The presence of aromatic heterocyclic ring structures present in DNA corresponding nucleotide bases, contribute to the absorbance at 260 nm (Heaton and Keer, 2008). The presence of phenolic compounds, polysaccharides derived from the plant material, together with thiocyanate and EDTA derived from the extraction buffer contribute for the UV absorbance maxima observed 230 nm (Arif *et al.,*2010).

Ratios of these maxima provide insights into the purity and quantity of DNA. Thus, the 260/280 nm gives an indication of the purity of DNA and RNA. In general, a ratio of ~1.8 is generally accepted as an almost pure form of DNA. The presence of

proteins present in the sample which absorb at 280 nm effectivity lower the 260/280 ratio. The 260/230 nm ratio is the second measure of DNA purity, it effectively measures the amount of the DNA relative to other contaminants. The higher the ratio, the purer is the DNA, typical values between 1.8 and 2.2.

2.1.3.1 Determination of DNA Concentration and Purity using NanoDrop

The pedestal of the Thermo Scientific NanoDrop ™ 2000 was cleaned by the addition of 1 µL of distilled water, the arm was lowered gently and both the arm and the pedestal were wiped with lens tissue. The TE buffer used for the resuspension of the DNA pellet was used to blank the Thermo Scientific NanoDrop ™ 2000 spectrophotometer. 1 µL of DNA sample was pipetted onto the pedestal, and the arm was gently lowered. The 260:280 ratio, 260:230 ratio and DNA concentration, were recorded. The pedestal and the arm of the pipette were cleaned between each sample.

2.1.4 Marker Based DNA Fingerprinting using SSR markers

The use of DNA molecular markers provides insights about the genetic relationships that different cultivars present in Maltese islands share. The use of SSR-PCR analysis involves the amplification of fragments containing a complementary sequence to that of the primer. The SSR primer is composed short nucleotide repeating sequence which binds to flanking region of a highly variable region of the genomic DNA. Once paired with the primer-DNA double strand is amplified. On the basis of the number of repeats that are amplified, fragments of different sizes that can be separated, visualized and sized. Analysis of these fragments reveals the presence of a unique set of amplicons which are unique to each cultivar enabling discrimination.

The use of molecular markers is based on the amplification polymerase chain reaction (PCR). This reaction amplifies and elongates complementary primer-DNA regions via a thermal resistant DNA polymerase enzymes in the presence of deoxynucleoside triphosphates (dNTPs) which assist the *de novo* synthesis of DNA. The PCR reaction is based on three steps. The denaturation where the double-stranded DNA template is separated into two strands. During the Annealing stage, complementary primers add across to the DNA template. This stage is highly temperature dependent as each primer has its own temperature of Annealing

temperature depending on its size and guanidine cytosine content. The last stage is the extension or elongation step, the polymerase enzyme attaches to the newly formed primer-DNA double strand and elongates the DNA region, flanked by the primer sequence.

PCR amplification is dependent on a number of different factors which affect its performance. One of the most important constituents of the PCR mixture in the presence of $MgCl_2$. This salt has a concentration-dependent dual effect in the PCR. At the correct concentration, it promotes DNA/DNA interactions and aids in the formation of DNA/dNTPs complexes. At a low concentration of Mg^{2+} , the primer-DNA interactions are reduced, preventing Annealing to the target DNA, while at a high concentration Mg^{2+} the base pairing of the primer sequence will be strengthened so much that it will fail to denature completely at 94°C, resulting in nonspecific primer-dimer amplification rather than DNA-primer amplification (Williams, 1989).

2.1.4.1 Analysis with microsatellite markers

For the molecular characterization of olive trees cultivated in the Maltese islands, a pool of 18 established SSR markers was used (Sefc *et al.,*2000; Carriero *et al.,*2002; Cipriani *et al.,*2002 de la Rosa et al., 2002). The markers were selected on the basis of their ability to discriminate, clarity of signal, number of alleles, heterozygosity and segregation independence (Sarri *et al.,* 2006; Baldoni *et al.,* 2009).

The amplification reactions were performed in a final volume of 25 µl containing 0.75 µl of 50mM of $MgCl_2$, 1.5 µl of 5 pmol of forward and reverse primer, 2.5 µl of 10X buffer, 8 µl of 2.5mM dNTPs, 0.5 µl of 5 U/µl Taq and 1 µl of 20 ng/µl of genomic DNA was added and the mixture was topped up with distilled water. Table 2.2 shows the thermocycler (Eppendorf ® Mastercycler Personal) program employed for the amplification procedure. Table 2.3 shows the SSR primers which were used in this study.

Table 2.2: PCR Amplification cycle program

PCR Amplification cycles		
No. of cycles	Amplification phase	Temperature and Time
1	Initial Denaturation	95°C 5min
35	Denaturation	95°C 30sec
	Amplification	(*) 30sec
	Extension	72°C 30sec
1	Final Extension	72°C 30-60min

(*) Temperature of Annealing depends on the primer to be used

Table 2.3 Primer sequences repeating motif and their corresponding Annealing temperature

Primer	Sequence 5' ➔ 3'	Motif	Tm °C	Reference
DCA-3	CCCAAGCGGAGGTGTATATTGTTAC TGCTTTTGTCGTGTTTGAGATGTTG	$(GA)_{19}$	59	Sefc *et al.,* 2000
DCA-4	CTTAACTTTGTGCTTCTCCATATCC AGTGACAAAAGCAAAAGACTAAAGC	$(GA)_{16}$	54	
DCA-9	AATCAAAGTCTTCCTTCTCATTTCG GATCCTTCCAAAAGTATAACCTCTC	$(GA)_{23}$	52	
DCA-16	TTAGGTGGGATTCTGTAGATGGTTG TTTTAGGTGAGTTCATAGAATTAGC	$(GT)_{13}(GA)_{29}$	57	
GAPU-101	CATGAAAGGAGGGGGACATA GGCACTTGTTGTGCAGATTG	$(GA)_8$ $(G)_3$ $(AG)_3$	53	Carriero et al., 2002
GAPU-103	TGAATTTAACTTTAAACCCACACA GCATCGCTCGATTTTTATCC	$(TC)_{26}$	52	
GAPU-59	CCCTGCTTTGGTCTTGCTAA CAAAGGTGCACTTTCTCTCG	$(CT)_9$	52	Chafari *et al.,*2008
GAPU-71A	GATCATTTAAAATATTAGAGAGAGAGA TCCATCCATGCTGAACTT	$(AG)_{10}$	55	
GAPU-71B	GATCAAAGGAAGAAGGGGATAAA ACAACAAATCCGTACGCTTG	$(AG)_6$ $(AAG)_8$	56.5	
UDO-99-12	TCACCATTCTTAACTTCACACCA TCAAGCAATTCCACGCTATG	$(GT)_{10}$	56	Cipriani *et al.,*2002
UDO-99-19	TCCCTTGTAGCCTCGTCTTG GGCCTGATCATCGATACCTC	$(GT)_{20}(AT)_5$	56	
UDO-99-39	AATTACCATGGGCAGAGGAG CCCCAAAAGCTCCATTATTGT	$(AT)_5$ $(GT)_{11}$	56	
UDO-99-43	TCGGCTTTACAACCCATTTC TGCCAATTATGGGGCTAACT	$(GT)_{12}$	53	
UDO-99-24	GGATTTATTAAAAGCAAAACATACAAA CAATAACAAATGAGCATGATAAGACA	$(CA)_{11}(TA)_2$ $(CA)_4$	51	
EMO-30	GTCTCTGCCCAACAATG CATACATGAGTGTGTGTG	$(AC)_8$	51	de la Rosa *et al.,*2002)
EMO-03	GGTGTAGCCCAAGCCCTTAT TGCATGACCGTGGTGTAAGT	$(A)_{15}(CA)_7$	60	

The PCR amplicons were then analyzed by agarose gel electrophoresis at 60 V in a 2 % (w/v) 25 cm by 10 agarose gel prepared as described in section 2.1.5.1. A 100 bp ladder (Solis Biodyne®) was used as a size standard. The gel was then stained as described in section 2.1.5.3 and observed under a UV transilluminator [UVP]. A photographic record was made using a Canon PC 1130 camera and preliminary analysed using Gel Analyzer Software 2010 developed by Istvan Lazar and Dr Istvan Lazar.

2.1.5 Determination of SSR amplicons using Gel Electrophoresis

Qualitative analysis of SSR amplicons was carried out using gel electrophoresis in order to provide an indication of primer amplification. Gel electrophoresis allowed the identification of the primers which gave a discriminatory and consistent amplification in the majority of samples. The bands produced were analysed using Gel Analyzer Software 2010 and the size of the amplicons was determined. The size of the alleles was placed into a contingency table and analysed using statistical software Past 3.06 (Øyvind Hammer, University of Oslo).

2.1.5.1 Molecular Marker Selection

The preliminary size of the PCR amplicons was estimated in terms of base pairs by comparing the distance migrated of the products to that of the 100 bp DNA ladder. DNA markers which gave consistent amplicon sizes were used as molecular markers. The distances moved by the PCR products were measured using Gel Analyzer Software 2010, using the 100bp ladder as a reference. Samples which gave the most consistent amplification were selected and sized using a more accurate capillary electrophoresis method performed using the Applied Biosystems 3130 Genetic Analyzer at MLS BIODNA.

Apart from their reproducibility, primers were also selected in terms of their heterozygosity and the ability to amplify different cultivars. From the set of 18 primers, only six primers showed these possible characteristics. Bands appearing in the negative control were not considered.

2.1.5.2 Cluster analysis generated using SSR data

Once the correct size of each amplicon was determined, the presence-absence of alleles present within each cultivar for each particular primer was noted in a two-dimensional contingency table. A dendrogram was constructed in order to visualize the relatedness between different samples. The unweighted pair-group method using arithmetic averages (UPMGA) was used to build a dendrogram using the Jaccard coefficient.

2.1.5.3 Assessing Diversity

2.1.5.3.1 Allele Frequency

Allele frequency is defined as the total number of one particular allele divided by the by the total number of alleles scored at that particular microsatellite marker. This is calculated for each locus.

$$Allele\ frequency = \frac{2Nxx + Nxy}{2N}$$

Where Nxx is the number of homozygotes for allele X (XX), and Nxy is the number of heterozygotes containing the allele X (Y can be any other allele). N = the total number of samples. Allele Frequency can also be determined simply by direct count of the proportion of different alleles.

2.1.5.3.2 Heterozygosity

Heterozygosity (H) is a measure of the genetic diversity within a population. This can be defined as the mean percentage of individuals heterozygous per locus (Crow and Kimura, 1970). High values of heterozygosity suggest a lot of genetic variabilities whilst low heterozygosity is synonymous with low genetic variability. There are two measures of heterozygosity, the observed heterozygosity defined as the number of individuals heterozygous per locus. The expected heterozygosity (also called gene diversity) is calculated from individual allele frequencies (Nei, 1987) which is defined as the probability that at a single locus any two alleles chosen at random are different from each other.

$$H_o = \frac{No.of\ Heterozygous}{N}$$

Where Ho is the observed heterozygosity, i.e. the proportion of N samples that are heterozygous at a given locus.

$$H_e = 1 - \sum p_i^2$$

Where He is the expected heterozygosity, i.e. the proportion of heterozygosity expected under random mating and pi is the allele frequency of the ith allele.

2.1.5.3.3 Fixation Index

The Fixation Index F_i (also called the Inbreeding Coefficient) exhibits values ranging from -1 to +1. Values close to zero are expected under random mating, while substantial positive values indicate inbreeding or undetected null alleles. Negative values indicate an excess of heterozygosity, due to negative assortative mating, or selection for heterozygotes.

$$F_i = \frac{H_e - H_o}{H_e}$$

2.1.5.3.4 Resolving Power

The resolving power (RP) of each primer is the sum of the individual fragment informativeness that can be represented on a 0–1 scale.

$$RP = \Sigma Ib,$$

In which I_b represents fragment informativeness.

$$I_b = 1 - [2 \times (0.5 - p)]$$

In which p is the proportion of the samples that contain the fragment.

2.1.5.3.5 Polymorphism information content

The Polymorphism information content (PIC) value for each SSR primer was calculated according to the formula

$$PIC_i = 2f_i \frac{1 - f_i}{N}$$

In which fi is the frequency of the marker fragments that were present, and 1 - fi is the frequency of the marker fragments that were absent and N total number of alleles. PIC was averaged over all fragments for each primer.

2.2 Chemical Analysis

2.2.1 Sampling and olive oil extraction

Samples derived from fresh pressed olive oil cultivar were obtained directly from the various olive presses found around the island. Coupage and monocultivar olive oil samples were obtained and stored at 4ºC prior usage. Olive oils derived from the Maltese islands were classified into two, the local indigenous which included 'Bidni', 'Malti' and 'Bajda' and the foreign locally grown cultivars. Olive oils were collected from 2013 until the end of this study. Monocultivar samples from neighbouring countries including Italy, Greece, France, Spain, Tunisia and Turkey were purchased as references samples for a total of 70 samples (35 Maltese and 35 Foreign).

Locally the extraction of olive oils is carried out using a continuous extraction system. The first step involves the washing and leaf removal, this step is vital for both the end product and the machinery as it reduced the chance of vegetable or non-vegetable parts that could be harmful to the machinery or contaminate the product. This operation is carried out by specialised machinery equipped with powerful blowers, which removes leaves and twigs, and a washing tank, with forced water circulation in which olives are washed. The second step involves the crushing of the olives into a paste, this is carried out using a metallic crusher, which consists of a metallic body, that rotates at high speed and throwing the olives against a fixed metal grating. Once the olive paste has formed it is homogenized into a mixing vessel. Mixing is vital, as it promotes contact between oil droplets allowing the separation of a continuous oily phase, this process is known as malaxation. Mixing is done for 30min to 60min depending on the amount of water present in the flesh of the olives. Water is sometimes added in order to soften the mixture favouring phase separation. The temperature of mixing is another important parameter which determines both the organoleptic parameters and the time of mixing. The higher the temperature (40ºC) the longer the time of mixing and the higher is total phenolic content (Di Giovacchino *et al.,*2002; Inarejos Garcia *et al.,*2009). The last stage of the extraction process involves the use of a horizontal centrifuge. The horizontal centrifugal force separates oil from other liquid and solid phases of olive paste, based on the weight of the fractions.

Table 2.6: The major cultivars used in this experiment and their country of origin.

Maltese[1] Indigenous	Foreign Locally Grown	Italian	Spanish	French	Greece
'Bidni'[2]	Frantoio	Frantoio, Leccino, Pendolino, Lecco del Corno	Arbequina	Picholine	Koroneiki
'Malti'[3]	Pendolino	Carolea, Cannava, Perzana, Ottobratica Tortiglione, Croatina, Moraiolo, Cerasuola Oleovastro, Gentile di Chieti	Hojiblanca	Grossane	
'Bajda'[4]	Carolea		Picual	Burguette	
	Picholine				

1. Due to the fact, that very few studies have been carried out on the Maltese olive oil and olive trees it is important to note, that the subsequent cultivar have not yet been established with a PDO or PGI certification.
2. The 'Bidni' cultivar is characterised by a small an ovoid-globose drupe; it is 8 - 9 mm in diameter and 10 - 15 mm in length which turns purple during maturity.
3. 'Malti' cultivar is characterised by at least five different morphotypes and it's still unknown whether each one of them is a different cultivar or else a single cultivar with different morphologies.
4. 'Bajda' cultivar is a distinct cultivar which bears white olive fruits is found in very small regions in the Maltese islands but never in the wild. There are two hypothesis regarding the origins of this cultivar in the Maltese islands. The first one is that it is a genetic mutant of a Maltese local cultivar, the other one is that it was imported by one of the previous conquistador as an ornamental plant.

2.2.2 Determination of olive oil quality parameters

The main quality parameters which are used both for the classification of different olive oils as well offering an insight about the quality of the EVOO include peroxide and p-Anisidine value, K_{232}, K_{270}, ΔK and free acidity. These quality parameters are determined using standard laboratory methods, which although not excessively sophisticated, they are required large amounts of solvents and starting material, as well as time-consuming and destructive sample preparation. The methods for the determination of olive oil quality parameters were obtained from the European Community methods (E. C. Regulation N. ° 2568/91) with minor modifications.

2.2.2.1 Determination of free fatty acids

Free acidity with olive oil is caused by the activity of lipolytic enzymes which attack triglycerides causing the formation of free fatty acids which increase the overall acidity of the oil. These enzymes are most active in freshly pressed olive oil, where the presences of the small amounts of water (approximately 0.5 %) increase the chances of hydrolysis of these triglycerides. The reaction is accelerated under storage temperatures higher than 18–20 °C (Salvador *et al.,* 2001).

The content of free fatty acids is expressed as the amount of potassium hydroxide or sodium hydroxide (mg) required to neutralize the free fatty acids in 1g of fat/oil. A sample (approximately 2g) was dissolved in a previously neutralized, equal part mixture of diethyl ether and 95 % ethanol (v/v). The resulting mixture was titrated against a previously standardized solution of ethanolic potassium hydroxide 0.01 mol/l using phenolphthalein as an indicator.

$$\text{Acidity expressed as \% of oleic acid} = \frac{VxcxM}{10xm}$$

V = the volume of titrated potassium hydroxide solution used, in millilitres;
c = the exact concentration in mol/L of the titrated solution of potassium hydroxide used;
M = the molar weight in grams per mole of the acid used to express the result (= 282);
m = the weight in grams of the sample.

2.2.2.2 Determination of peroxide value

The formation of lipid peroxides is the first stage of oil oxidation. The degradation of lipids process often called the PUFA cascade (Feussner and Wasternack 2002; Wang and Hammond 2010). It occurs by means of different enzymes mainly which convert unsaturated fatty acids into lipid peroxides. Lipid peroxides can also form via lipid autoxidation under incorrect storage conditions. The presence of light, temperature, metals, pigments, and microorganism accelerate the rate of lipid oxidation (Frankel, 1989, 2005).

The peroxide value is defined as the quantity of those substances namely hydroperoxides which are produced from the primary oxidation of fatty acids which

are able to oxidize potassium iodide under the operating conditions described. The peroxide value is expressed in terms of milliequivalents of active oxygen per kilogram. Briefly, 5-10 g of oil was dissolved in 10 ml of chloroform followed by 15 ml of acetic acid and 1 ml of saturated potassium iodide solution (140 g/100 mL). The flask was closed and the mixture was shaken for 1 minute and left for exactly 5 minutes away from light at a temperature of 25 °C. 75 ml of distilled water were added in order to quench the reaction and the resulting solution was titrated against a previously standardized solution of 0,002 mol/L sodium thiosulfate solution using starch as indicator.

$$PV \text{ expressed as mEq of } O_2 = \frac{VxTx1000}{m}$$

V = the number of ml of the standardized sodium thiosulfate;

T = the exact morality of the sodium thiosulfate solution used;

m = the weight in g, of the test portion.

2.2.2.3 Determination of *p*-anisidine value

The anisidine value is defined as 100 times the optical density measured at 350 nm in a 1cm cell of a solution containing 1 g oil in 100 ml of a mixture of solvent and reagent. This method determines the amount of aldehyde (mainly 2-alkenals and 2, 4-alkadienals) present in olive oil, formed from the decomposition of lipid peroxides during the secondary stages of oxidation. Under acidic conditions, *p*-anisidine (4-methoxyaniline) is able to react with aldehydes and ketones in order to form yellow compounds having an absorption maximum at 350 nm.

The method measures the colour change produced by the reaction of *p*-methoxyaniline (anisidine) with the aldehydic compounds found in the oil (Doleschall *et al.,* 2002). This is done via the Mannich reaction which comprises the amino alkylation of an acidic proton located next to a carbonyl functional group, resulting in the final product β-aminocarbonyl compound also known as a Mannich base. This reaction occurs under acidic conditions since the acid is needed to activate the carbonyl group of the aldehyde. The Mannich base has a yellowish colour with a λ_{max} at 350 nm (Doleschall *et al.,* 2002). The colour is quantified and converted to *p*-AnV.

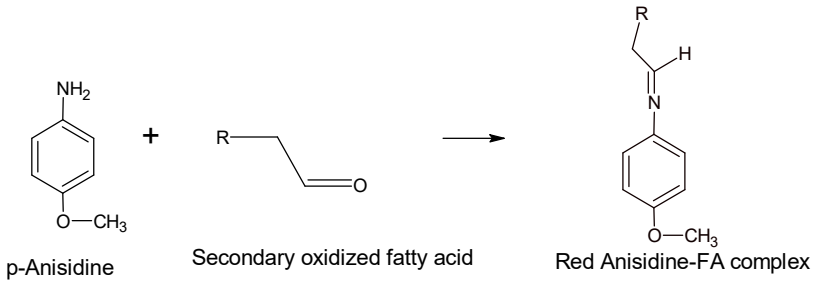


Figure 2.1: The reaction of *p*-anisidine with aldehydes formed from the secondary oxidation of fatty acids

The *p*-anisidine value was determined by dissolving 2.5 g of oil in 25 ml of iso-octane. 5ml of the resulting solution was aspirated and added to1ml of previously prepared *p*-anisidine solution (0.25% v/v in acetic acid). The solution was vortexed for 1 minute and left to stand in the dark for exactly 10 minutes. The spectrophotometer was blanked with a solution containing 5 ml of iso-octane and 1 ml of *p*-anisidine. The absorbance of the sample was measured at 350 nm followed by the absorption of the sample without *p*-anisidine. The *p*-anisidine value was calculated using the formula below

$$p - anisidine\ value\ = \frac{25\ (1.2E_b - E_a)}{w}$$

E_a = net absorbance of the oil solution

E_b= net absorbance of the oil *p*-anisidine solution

W= weight of the sample used

2.2.2.4 Determination of Iodine value

The determination of the iodine value is based on the addition of iodine to the double bonds of unsaturated fatty acids. This value give a measure of the total amount of unsaturation present within the olive oil, expressed as g I_2 consumed by 100 g of oil. Approximately 1g of oil was weight and transferred to a 500 ml conical flask followed by 20 ml of equal volume mixture of cyclohexane and acetic acid. Once dissolved 25 ml of Wijs reagent was added, the flask was mixed and left in the dark for exactly 1hour after which 20 ml of saturated KI and 150 ml of water were added. The resulting solution was titrated against 0.1M sodium thiosulfate using starch as indicator.

$$I_2\ value\ expressed\ as\ \frac{gI_2}{100g} = \frac{12.69x0.1(T_B - T_A)}{W}$$

T_a = Titer value (ml) of the oil solution

T_b= Titer value (ml) of the blank

W = weight (g) of the sample used

2.2.2.5 Spectrophotometric analysis

Spectrophotometric analysis gives an indication of the state of oxidation by detecting specific oxidized compounds, some generated from secondary oxidation, and also detect possible adulteration with refined oils. The oxidation of polyunsaturated fatty acids is accompanied by an increase in the UV absorption of the products. In fact, fatty acids containing methylene-interrupted dienes and trienes show a shift in their double-bond position during oxidation due to isomerisation and conjugate formation. The resulting conjugated dienes exhibit an intense absorption at 232 nm while conjugated trienes absorb at 270 nm.

The oil solution was diluted by a factor of 10 or 100 using hexane; an aliquot of the resulting solution was withdrawn and placed in a quartz cuvette. The absorbance at 274,270,266 and 230 nm was determined using a UV/Vis spectrophotometer. The absorbance obtained was standardized to their K value using the formula below.

$$K\lambda = \frac{Abs(\lambda)}{DXL}$$

Abs (λ) = is the absorbance of the diluted oil at a specific wavelength

D= is the dilution factor

L= is the cuvette path length

Since the products of oxidation of dienes (aldehydes and ketones) also absorb at wavelengths of 270,266 and 274 nm, these are taken into account with the use of ΔK coefficient defined by the following formula.

$$Delta\ K\ = K_{270} - \frac{K_{266} + K_{274}}{2}$$

2.2.3 Identification and quantification of phenolic compounds

Phenolic compounds characterise one of the families of antioxidants and are relatively abundant in EVOO (Visioli and Galli, 1998). There are a variety of different phenolic classes found in EVOO, the most characteristic are: phenols (hydroxytyrosol), phenolic secoiridoids (oleuropein aglycone), lignans (pinoresinol), flavonoids (luteolin and apigenin) and hydroxyisochromans (hydroxyisochromans 1-phenyl-6,7-dihydroxyisochroman) (Carrasco-Pancorbo *et al.,* 2005; Owen, *et al.,*2000; Bianco *et al.,*2001). The level and composition of the different classes of phenolic compounds are important in the evaluation of the quality of EVOO as they confer protection against oxidation (Servili and Montedoro, 2002). The phenolic profile of different olive oils is dependent on a number of different factors namely; the producing cultivar, storage, season, pedoclimatic conditions, maturation and extraction technique (Tovar *et al.,* 2001, Romero *et al*., 2002, Garcia *et al*., 2003, Vinha *et al*., 2005, Cerretani *et al*., 2004).

There is various method for the extraction of phenolic compounds present in olive oil but namely two procedures dominate the existing literature namely those based on liquid/liquid partitioning extraction (Montedoro *et al.,*1992) and solid phase extraction (SPE) (Pirisi *et al.,*2000). In the majority of cases, methanol is used as the main extracting solvent. In the case of liquid/liquid extraction, the phenolic fraction can be isolated from olive oil by using only methanol (Owen *et al.,* 2000) or with various levels of methanol/water ranging from 0% and 40% water (Montedoro *et al.,* 1992 Carrasco-Pancorbo *et al.,* 2005). According to Montedoro *et al.,*1992 the use of 80:20 (v/v) methanol/water was reported as the most efficient extraction solvent and it is used in the official method for the determination of diphenols in olive oil. However, Angerosa, *et al.,* (1995) reported incomplete recovery of some components and the formation of considerable emulsions between the oil and the methanol-water layer and suggested methanol 100% as an extraction solvent.

2.2.3.1 Extraction of phenolic compounds using solid phase extraction

In this study, phenolic compounds in olive oil were extracted by mean of solid-phase extraction (SPE) rather than the conventional liquid-liquid extraction. The application of solid phase extraction has numerous advantages over liquid-liquid extraction including; greater selectivity; high recoveries; good reproducibility; less labours and low solvent and analyte volumes are required. Polar stationary phases were used for the isolation of phenolic compounds from oil which is a nonpolar matrix. The polar stationary phase is able to retain polar constituents present in olive oil while the fatty acids and other non-polar constituents pass through without being retrained. A diol bonded stationary phase was used in this experiments due to its negligible activity on labile esters (Perez-Camino *et al.,* 1996 and Mateos *et al.,* 2001). Prior extraction the cartridge was conditioned by eluting first, with methanol (20 mL) and then with hexane (20 mL), and the oil solution was then applied to the SPE column. 7.5 g sample of virgin olive oil were weighed and dissolved in 20 mL of hexane.

1000mg diol-bonded phase cartridges (Analytical Columns SolGel-1ms™, England) were used for the extraction of phenolic compounds. The sample was passed through the column under a vacuum of not more than 15 inHg. The column was washed with hexane (2 X 6 mL) and twice with hexane/ethyl acetate (85:15, v/v; 8mL) which were run through the cartridge and discarded. The passage of the diluted olive oil - hexane solution through a diol cartridge retained the polar compounds on the solid phase. Hexane washing eliminated hydrocarbons, waxes, tocopherols, and triacylglycerols. Subsequent washing with hexane/ethyl acetate (85:15, v/v) removed the major part of the oxidized triacylglycerols, sterols, and diacylglycerols. The phenolic fraction was eluted with methanol (20 mL) filtered through a 0.45-µm nylon filter and concentrated to half of the initial volume with a rotary evaporator at 35 ºC under vacuum and then to dryness using nitrogen evaporator at 45 ºC under a slow stream of nitrogen, yielding a yellow solid residue. The phenolic residue was reconstituted in 1.5 ml methanol/acetonitrile (1:1 v/v).

2.2.3.1.1 Quantification of total phenolic content using microtiter plating

Folin-Ciocalteu colourimetric method (Singleton *et al.,*1999) was used in order to determine the total phenolic content present in the hydroalcoholic extracts derived from extra virgin olive oil against a standard calibration curve made with gallic acid (Sigma Aldrich) (R^2=0.997). Folin-Ciocalteu (FC) reagent is formed from a mixture of phosphotungstic acid,$H_3PW_{12}O_{40}$, and phosphomolybdic acid, $H_3PMo_{12}O_{40}$, which, after oxidation of phenols, is reduced to a mixture of blue oxides of tungsten, W_8O_{23}, and molybdenum, Mo_8O_{23}.

The concentrated extracts derived from SPE were diluted by a factor of 10 using methanol/acetonitrile (1:1 v/v). 20 µL of the resulting solution was oxidized with 100 µL of Folin-Ciocalteu reagent followed by neutralization by the addition of Na_2CO_3 (80 µL, 7.5%) in 96-well microtiter plate. After incubation at room temperature for 2 hours in dark, the absorbance at 600 nm was recorded using a microtiter plate reader (SPECTROstar ^{Nano}). TPC was expressed as mg gallic acid equivalents per gram of olive oil (mg GAE g^{-1}).

2.2.3.1.2 Quantification of total *o*-diphenolic compounds using microtiter plating

The Arnow's colourimetric method (Mateos *et al.,* 2001) was used for the determination of *o*-diphenolic compounds present in the hydroalcoholic extracts derived from extra virgin olive oil against a standard calibration curve made with pyrocatechol (Sigma Aldrich) (R^2=0.997). The assay is based on the *in situ* formation of nitrous acid, obtained by dissolving sodium nitrite in a diluted mineral acid, which instantaneously gives a red or pink colour, changing slowly to purple or blue in the presence of *o*-diphenolic compounds (Procter and Paessler 1901). Apart from the Arnow's assay, there are, two other closely related methods which are used for the determination of *o*-diphenolic compounds, the Bate-Smith and the Hoepfner assay, however, the addition of sodium molybdate in the Arnow's reagent is preferred as it prevents the decomposition of nitrous acid (Arnow, 1937). The mechanism of the reaction is not yet clear, but the *o*-diphenolic compounds are fairly easily oxidized in the presence of nitrous acid to form a red stable quinone chromophore.

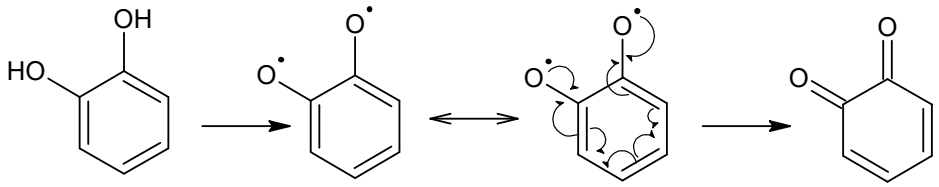


Figure 2.3: The oxidation of *o*-diphenolic compounds in the presence of the nitrous acid is thought to proceed via a radical intermediate to yield the red coloured quinone chromophore.

20 µL of the diluted phenolic extract was added to 20 µL of 1M HCl in a 96-well microtiter plate. The resulting mixture was briefly mixed followed by the addition of 20 µL of Arnow's reagent, previously prepared by dissolving 10g of sodium nitrite and 10 g of sodium molybdate dihydrate in 100ml ethanol/water (1:1). The plate was shaken vigorously and after 15 min, 80 µL of water and 40 µL of 1M NaOH were added and the absorbance was measured at 370 nm was recorded using a microtiter plate reader (SPECTROstar Nano). TdPC was expressed as mg pyrocatechol equivalents per gram of olive oil (mg PyCE g^{-1}).

2.2.3.1.3 Quantification of total flavonoid compounds

Due to their structure, flavonoids can easily chelate metal ions and form complex compounds. It seems that metal-flavonoid complexation reactions are particularly appropriate for analytical objectives as the formed complexes bear exceptional spectrophotometric characteristics. The respective colourimetric protocols, mainly those involving Al (III), are simple, rapid, and inexpensive and have met wide applications even as official methods. The principal of aluminium chloride colourimetric method is that aluminium chloride forms acid stable complexes. Al (III) is likely to bind with 1, 2 or 3 molecules of bidentate ligands to form 1:1, 1:2, and less often 1:3, complex compounds. In the case of flavonoids, those that can act as ligands bear 5 or 3-hydroxy-4-keto and/ or *o*-dihydroxy group (Mabry *et al.,* 1970).

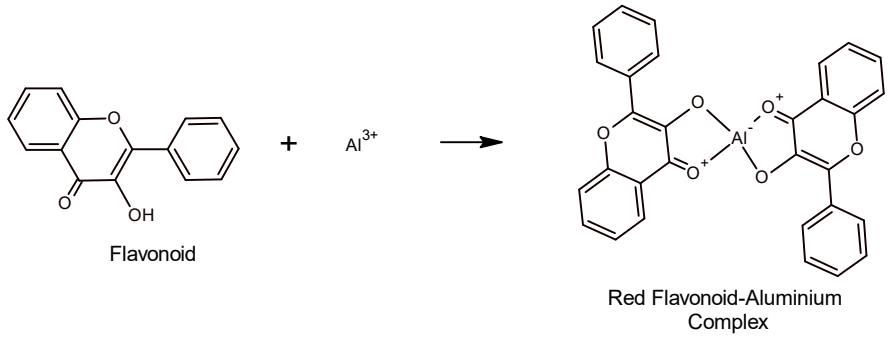


Figure 2.4: Complexation reaction involving flavonoids and aluminium ions. The degree of coordination is dependent on the flavonoids' side groups; the larger the groups the more sterically hindered the complex is and thus a smaller coordination number is achieved.

The flavonoid content present in the hydroalcoholic extracts derived from extra virgin olive oil was determined using the aluminium chloride colourimetric method. This method relies on the formation of a red complex between the flavonoids and the aluminium (III) ions in the presence of sodium nitrite. This complexation reaction provides a bathochromic displacement and a hyperchromatic effect.

25 µL of the diluted extract was mixed with 7.5 µL of 10 % aluminium chloride, 7.5 µL of 7% w/v sodium nitrite and 80 µL of distilled water, and left at room temperature for 30 min. After which, 100 µL of 1M NaOH solution was added. The plate was shaken vigorously and the absorbance of the reaction was recorded at 415 nm. The calibration curve was prepared by using catechin (Sigma Aldrich) (R^2=0.997). TFC was expressed as mg catechin equivalents per gram of olive oil (mg CE g^{-1}).

2.2.3.1.4 Determination of phenolic profiles present in extra virgin olive oils using high performance liquid chromatography.

Although a large number of HPLC studies have focused on the analysis of the phenolic compounds in virgin olive oil, this fraction is still a complex mixture and its chemical nature has not been completely elucidated (Montedoro *et al.,* 1992 and 1993; Brenes *et al.,* 1999; Owen *et al.,* 2000; Mateos *et al.,* 2001). Furthermore a large number of studies have focused on the relationship between the phenolic content and the oxidative stability of the oil (Papadopoulos *et al.,*1991; Tsimidou *et al.,* 1992; Baldioli *et al.,* 1996), the influence of the extraction system (De Stefano *et al.,* 1999; Garcia *et al.,* 2001; Vierhuis *et al.,* 2001) maturation (Monteleone *et al.,* 1995; Rotondi *et al.,*2004; Amiot *et al.,* 1989) and climate (Vinha *et al.,* 2005; Hashemi *et al.,* 2010) on the phenolic composition of the oil and olive drupes. Although it has been suggested by most of the previous studies there are very few studies which deeply study the effect of the olive cultivar on the composition of the phenolic profile and the use of this variable as a tool for classification and identification of olive oil varieties (Oliveras-Lopez *et al.,* 2007; Gomez-Rico *et al.,* 2008; Talhaoui *et al.,* 2015).

Phenolic separation is mainly achieved using reverse phase HPLC, whereby a modified hydrophobic silica stationary phase is used. The stationary phase involved range from C_8-C_{30} depending on the hydrocarbon chain length which is attached to the silica stationary phase. The mobile phases used in reverse phase HPLC are polar aqueous solution and buffers with at various concentrations of acidity. In RP-HPLC, higher polarity phenolic compounds will be eluted earlier than the one with lower polarity. The most commonly employed column for the separation of different classes of phenolics and their glycosides is C_{18}-bonded silica column ranging from 100 to 300 mm in length and with an internal diameter of 2–4.6 mm (Merken and Beecher, 2000; Stalikas, 2010; Tsao and Deng, 2004; Tulipani *et al.,*2012). The use of C_8 is less common and is almost exclusively used for the separation of phenolic acids only. The column temperature is usually kept at either room temperature or at higher temperatures up to 40°C. The increase in column temperature increases the repeatability of elution times and resolution and reduces back pressure at high flow rates (Merken and Beecher, 2000; Tsao and Deng, 2004). Gradient elution is usually preferred over isocratic elution, gradient elution involves the use of two different solvent mixtures; an organic solvent such as methanol or acetonitrile, pure or acidified

and aqueous solvent phase which is most of the times acidified with either organic acids namely, formic and acetic acid, or inorganic acids mainly phosphoric acid. Although a large number of studies have been carried out throughout the years complete identification and quantification of certain phenolic compounds namely secoiridoids has still not been fully achieved, the main reason being is that these compounds are not commercially available and alternative methods have been proposed for their quantification and identification Mateos *et al.,* 2001.

The HPLC system was made up of a Waters 717 plus Autosampler, a Waters 600 pump, a Waters column heater module and a Waters 996 photodiode array detector managed by Empower software (Waters Inc., Milford, MA). The stationary phase was a Symmetry® C18 analytical column (250 × 4.6 mm i.d.) with a particle size of 5 µm (Woodford, Santry, Dublin).

The mobile phases for chromatographic analysis were previously degassed and constituted of (A) water: acetic acid (98: 2, v/v) and (B) methanol: acetonitrile (1: 1, v/v) at a constant flow rate of 1 mL/min. The column temperature was set to 35ºC whilst the sample chamber was set to 12ºC to prevent phenolic degradation. The gradient elution program of solvent was as follows:

Table 2.4: Gradient elution solvent program.

Time (min)	% Solvent A	% Solvent B
0 – 30	80	20
30 – 45	70	30
45 – 55	50	50
55 – 65	40	60
65 – 75	0	100
75 – 79*	80	20

* Represent the post equilibration phase

Detection was performed simultaneously at 280 nm to measure phenols (except ferulic acid), cinnamic acid, and lignans, at 320 nm for flavones and ferulic acid and at 520 nm for any anthocyanin compounds. Phenolic compound concentrations were expressed as mg/ kg of syringic acid. The identification of some phenolic components

was carried out comparing the peak retention times with those obtained by injection of pure standards (hydroxytyrosol and tyrosol, purchased at Extrasynthese, Genay Cedex, France; vanillin, purchased at Sigma-Aldrich, Milan, Italy). The identification of the other phenolics (3, 4-DHPEA-EDA, *p*-HPEA-EDA, pinoresinol, 3, 4-DHPEA-EA, *p*-HPEA-EA was made on the basis of studies found in the literature (Brenes and others 2000; Gómez-Alonso and others 2002, 2007; Morello and others 2004; Gómez-Rico and others 2008).

2.2.4 Estimation of Chlorophyll and carotenoid content in olive oil using spectrophotometry.

The presence of pigments in olive oil is highly dependent on a number of different factors similar to the phenolic composition, these include genetic factors (olive variety), the index of maturity during harvest, environmental conditions, extraction and storage conditions. During the extraction and storage, there is an inevitable loss of oil pigments, mainly chlorophylls which to their structure, tend to undergo chemical reactions which cause the loss of the central Mg^{2+} ion and the formation of pheophytin, in the presence of acids. Chlorophylls pigments impart a greenish hue to freshly pressed olive oil, however, the most abundant is pheophytin a, which is more stable than the corresponding chlorophyll a. Carotenoid pigments namely lutein and b-carotene are responsible for the yellow colouration olive oils. In this part of the experiment, pigments were only determined by the use of spectrophotometric techniques.

Pigments were determined using the method of Minguez-Mosquera *et al.,* (1991). 1.5g of olive oil were accurately weighed and dissolved in cyclohexane up to a final volume of 5 mL. Chlorophyll content was calculated from the absorption value of the olive oil solution at 670 nm and specific coefficient for pheophytin a, $E_o = 613$ using the equation

$$C = \frac{(A_{670} \times 10^6)}{613 \times 100 \times d}$$

where “C” represents the content of chlorophyll pigments expressed in mg/kg of pheophytin a (ppm), “A” stands for the absorbance at the respective wavelength (nm) and “d” represents the thickness of the spectrophotometer cell (1 cm).

Cyclohexane (UV–Vis Spectroscopy) was purchased from Scharlau Chemie, Spain. A lambda 25 UV/VIS spectrometer (Perkin Elmer) was used.

Carotenoid content was calculated from the absorption value of the olive oil solution at 470 nm and specific coefficient for lutein, E_o = 2000 using the equation

$$C = \frac{A_{470} \times 10^6}{2000 \times 100 \times d}$$

Where "C" represents the content of carotenoid pigments expressed in mg/kg of lutein (ppm)

With respect to the other phenolic compounds anthocyanins pigments tend to be present in very small amounts in order for them to be determined by the pH differential protocol (Lee *et al.,* 2005), thus an indirect approach was used. The total anthocyanins concentrations were calculated by measuring the absorbance of undiluted oil fraction at 535 nm which is the absorption maxima of most anthocyanins pigments. The absorbance value is then converted to mg/kg by using the average extinction coefficient for anthocyanin pigments (Francis, 1982).

$$Total\ anthocyanin\ content\ mgKg^{-1} = \frac{Abs_{\ 535}\ x\ 1000}{AvE_{535}\ \frac{1cm}{10}}$$

Abs_{530}= the total optical density at 535 nm of the phenolic fraction

AvE_{535}= is the average coefficient for the total anthocyanins pigments when 1cm cuvette and 1 % (10 mg/ml) standards are used which is equal to 982.

2.2.5 Determination of the antioxidant activity of phenolic compounds from extra virgin olive oils

Huang *et al.,* 2005 distinguished between the antioxidant activity and the antioxidant capacity, whereby the antioxidant activity of a particular compound is governed by the reaction conditions that must be taken into account, whilst on the other hand the capacity does not take into account the conditions and it is used for comparisons when the experimental conditions are different. The antioxidant capacity assays can be subdivided into two, those which involve the hydrogen atom transfer reaction-based assays and, those involving single electron transfer reaction-based assays. Electron transfer reaction is less time consuming but nonetheless they provide comparable results. In this section, a number of electron-based assays will be used in order to assess the antioxidant activity of the olive oil phenolic compounds.

Several publications have determined the resistance to oxidation of olive oil is closely related to its total polyphenol content (Hrncirik and Fritsche 2005; Matos *et al.,* 2007). Phenolic compounds have the ability to interrupt the initiation and propagation stages of the oxidative chain reaction since they react with lipid radicals to form more stable products (Frankel, 1998). Measuring the intrinsic antioxidant capacity of the olive oil together with the antioxidant capacity of its minor constituents give an insight into the stability towards oxidation and ultimately shelf-life.

2.2.5.1 Determination of the total antioxidant capacity

Phosphomolybdenum complex assay for the antioxidant activity is based on the reduction of molybdate (VI) to molybdate (V) by the sample and subsequent formation of a green phosphate/Mo (V) complex in acidic pH (Abdel-Hameed, 2009). This reaction is also the bases the for the determination of the total phenolic content by the use of FC reagent, which consists of a mixture of heteropoly phosphotungstates-molybdates, however, it is believed that the molybdenum is responsible for the formation of the blue complex rather than the tungsten.

The total antioxidant capacity assay of samples was carried out by the phosphomolybdenum method according to Umamaheswari and Chatterjee, 2007. This assay was modified to work with 96- well plates. 20 µl of diluted extract was shaken with 310 µl of reagent solution (0.6M sulfuric acid, 28 mM sodium phosphate and 4 mM ammonium molybdate). The plates were covered and incubated in an oven set at

65 °C for 120 min. After the samples were cooled, the absorbance of the mixture was measured at 765 nm. The antioxidant capacity was estimated and expressed as ascorbic acid equivalents using a previously prepared calibration.

2.2.5.2 Determination of the total reducing power

The reducing capacity is a reflection of the total antioxidant activity, compounds with a high reducing power are cable of donating electrons in order to stabilise radical intermediates. Total reducing power of the phenolic extracts were determined using the ferric reducing properties of phenolic compounds. Antioxidant compounds which possess a reduction potential, react with potassium ferricyanide (Fe^{3+}) to form potassium ferrocyanide (Fe^{2+}), which in turn reacts with ferric chloride to form a ferric ferrous complex that has an absorption maximum at 700 nm.

100 µl of the diluted phenolic stock solution was added into each well followed by the addition of 100 µl 1.0M hydrochloric acid, 20 µl 1% sodium dodecyl sulphate and 30 µl 1% potassium ferricyanide. The mixture was incubated at 50 °C for 20 min. After which 20 µl of 90% TCA (trichloroacetic acid) was added followed, 30 µl of 0.1 % ferric chloride was added to each solution and the absorbance was read at 750 nm using microplate reader (Micro Quant, Biotek Instruments). The antioxidant activity of the sample was determined using the following equation

$$Antioxidant\ Activity(\%) = \frac{Abs_{sample}\ x\ Abs_{Control}}{Abs_{Sample}}\ X100$$

2.2.5.3 Radical scavenging activity using DPPH assay

The radical scavenging activity of both the phenolic extracts and the complete olive oil was measured using the DPPH assay. DPPH is one of the few stable and commercially available organic nitrogen radicals. It is a well-known radical trap ("scavenger") for other radicals. DPPH radicals have a deep-violet colour with a strong absorption band centred at about 520 nm, becoming colourless or pale yellow when neutralized. Therefore, the rate of reduction of DPPH radicals is used as an indicator of the reaction progress. This property allows visual monitoring of the reaction, and changes in the initial number of radicals can be accounted by measuring the change in optical absorption at 520 nm.

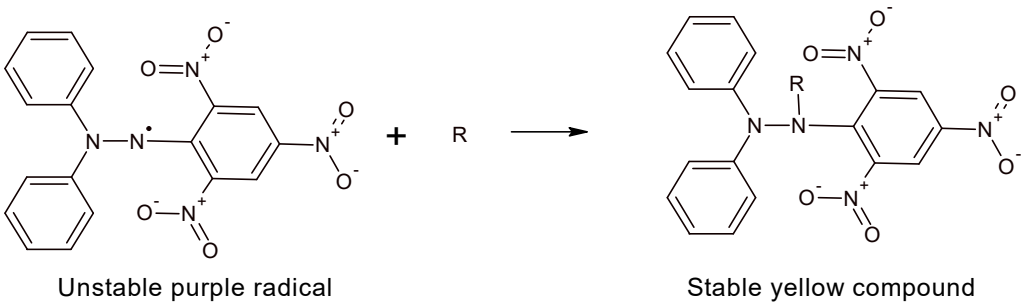


Figure 2.5: The colour change on stabilisation of DPPH radicals makes it useful for monitoring the reaction progress.

In order to determine the radical scavenging activity of phenolic compounds derived from olive oil, a stock solution of DPPH was daily prepared by dissolving 5 mg of DPPH in 2 ml of methanol (MeOH) and the solution was kept in the dark at 4 ºC. A stock solution of olive oil phenolic compounds adjusted to be equivalent to 500 mg GAE g^{-1} was prepared by diluting the crude phenolic extract in absolute methanol.

25 µl of the phenolic stock solution was added to the well and it was further down diluted to the lowest concentration (7.8 µg/ml) by performing a serial two-fold dilution in a 96-well microtiter plate. A row of negative DPPH control was added in the same 96-well microtiter plates by adding 100 µl of MeOH into each well. Gallic acid, caffeic acid catechin, pyrocatechin, quercetin, vitamin C and tocopherol were used as positive controls. Then, 100 µl of methanolic DPPH was added into each well and reaction was allowed to proceed for 30 minutes in dark. The absorbance was measured at 517 nm by a microplate reader (Micro Quant, Biotek Instruments).

For kinetics studies, the rate of disappears of DPPH radicals was determined by measuring the absorbance of the reaction at regular time intervals (2minutes for 10 minutes and every 5 minutes for 20 minutes). The DPPH scavenging effect (%) was calculated using the following formula:

$$\%\ Radical\ Scavenging = \frac{Abs_{DPPH\ in\ methanol} - Abs_{\ DPPH\ with\ sample}}{Abs_{DPPH\ in\ methanol}}\ x100$$

The efficient concentration at different times EC_{50}, t (mg extract mg^{-1} DPPH) was the amount of the extracts in relation to the amount of initial DPPH, which was calculated using the following equation:

$$EC_{50\,t} = \frac{IC_{50,t}}{[DPPH]_{t=0}}$$

$IC_{50,\,t}$ = inhibitory concentration at different times, defined as the concentration of extract (mg mL^{-1}) required to scavenge 50 % of DPPH

$[DPPH]_{t=0}$ = initial concentration of DPPH (mg mL^{-1}).

The intrinsic radical scavenging activity of olive oil was carried out using a 50 µl sample which was two-fold serially diluted in isoamyl alcohol in a 96 well plate, followed by the addition of 100 µL of DPPH (4 mg) dissolved in (50 ml) isoamyl alcohol.

2.2.5.4 Determination of ABTS Radical Cation Decolourization Assay

ABTS assay measures the relative ability of antioxidant compounds to scavenge ABTS radical cations which are generated in the aqueous phase, the antioxidant activity is most of the time compared and expressed in terms of Trolox (water-soluble vitamin E analogue) standard. The radical cations are generated by the reaction of the strong water-soluble oxidizing agent (potassium permanganate or potassium persulfate) with the ABTS chloride salt. The oxidation of the ABTS salts results in the formation of a blue-green ABTS radical cation. Antioxidant compounds are able to reduce back the radical cation causing the suppression of its characteristic longwave (734 nm) absorption maxima, enabling the visual monitoring of the reaction.

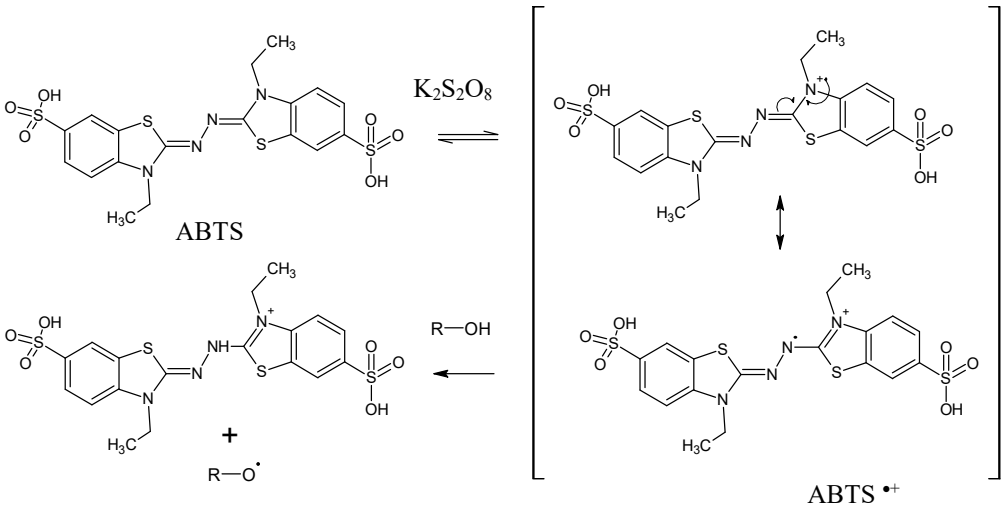


Figure 2.6: Oxidation of ABTS by potassium persulfate to generate stable radical cations and its reaction with antiradical compound (ROH)

The ABTS method is more flexible than the DPPH method as it can be used at different pH levels (unlike DPPH, which is sensitive to acidic pH) and thus is useful when studying the effect of pH on the antioxidant activity of various compounds under different pH conditions. Additionally, ABTS is soluble in aqueous and organic solvents and is thus useful in assessing antioxidant activity of samples in different media and is most commonly used in simulated serum ionic potential solution (pH 7.4 phosphate buffer containing 150 mM NaCl) (PBS).

The $ABTS^{\bullet+}$ was produced by the reaction of 7 mM stock solution of 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) with 2.45 mM potassium persulfate and allowing the mixture to stand in dark at room temperature for 12 h before use. The concentration of the blue-green ABTS radical solution was adjusted with methanol to an absorbance of 0.700 (0.020± mean ±SD) at 734 nm. To 280 µL of this solution was added 20 µL of sample or solvent in a 96-well plate. For phenolic extracts, the stock solution of (500 µg/ml) was added into the well and it was further down diluted to the lowest concentration (7.8 µg/ml) by performing a serial two-fold dilution. The mixture was incubated for 5 min at 30 °C, and the absorbance at 734 nm was measured with a microplate reader. The percentage inhibition of $ABTS^{\bullet+}$ was calculated as follows;

$$\%\, ABTS^{\cdot+}\ inhibition\ =\ 100\ -\ 100\left(\frac{Abs_{sample}}{Abs_{control}}\right)$$

Similar to the DPPH the efficient concentration at different times EC_{50}, t (mg extract mg^{-1} ABTS) was calculated using the following equation:

$$EC_{50\,t} = \frac{IC_{50,t}}{[ABTS^{\cdot+}]_{t=0}}$$

2.2.5.5 Determination of the intrinsic antioxidant activity of olive oil using ABTS assay.

Since ABTS is soluble in various organic solvents it also investigated the use of ABTS radical cations for measuring the intrinsic antioxidant capacity of olive oils using microtiter plates. First, the solubility of both potassium persulfate and ABTS in different solvents which need to be strong enough to dissolve the reagents but not the Poly (methyl methacrylate (PMMA) plates (Table 2.5) was assessed.

Table 2.5: Preliminary solubility study for the determination of the intrinsic antioxidant activity of olive oil against ABTS. The DMSO: Isoamyl alcohol solvent system was the most appropriate for studying the intrinsic antioxidant capacity of olive oils.

Solvent	Solubility of $K_2S_2O_8$	Solubility of ABTS	Solubility of olive oil
Methanol: Acetonitrile 1:1	-	+	-
Methanol	-	+	-
Ethanol	-	+	-
Chloroform *	-/+	+	+
Chloroform: Isoamyl alcohol (1:1)	-/+	+	+
Isoamyl alcohol	-	+	+
Dimethyl sulfoxide (DMSO)	-/++	+	-/+
DMSO: Isoamyl alcohol (3:1)	+	+	+

*Dissolved the PMMA plates.

2.45 mM potassium persulfate (26.5 mg) was dissolved in 10ml of DMSO while 7 mM stock solution of ABTS was prepared by dissolving 143.9 mg of ABTS in 10 ml of DMSO, once dissolved the two solutions were mixed followed by the addition of 10 ml of DMSO and 10 ml of isoamyl alcohol. The mixture was gently heated in a microwave oven at 500 W for 1minute and left to stand for 12 hours in the dark. The concentration of the blue-green ABTS radical solution was adjusted with isoamyl alcohol to an absorbance of 0.7 (±0.020) at 734 nm. The adjusted solution was subjected to stability testing by recording the absorbance at 734 nm over a period of 24 hr to assess whether degradation of the radical cation occurs.50 µL of olive oil was twofold serial diluted in isoamyl alcohol in a 96 well plate, followed by the addition of 200 µL of the previously adjusted ABTS solution. The results obtained were expressed as % $ABTS^{\bullet+}$ scavenging and EC_{50} in µL of olive oil and compared with those obtained by measuring the intrinsic antioxidant activity using DPPH assay.

2.2.5.6 Determination of NO radical scavenging activity using Griess Reagent System

The nitrous oxide scavenging activity of olive phenolic was assessed using the Griess Illosvoy reaction (Garratt, 1964). The Griess Reagent System is based on the chemical reaction between sulfanilamide and N-1-napthylethylenediamine dihydrochloride (NED) under acidic (phosphoric acid) conditions. In the presence of hydrogen ions, nitrite forms nitrous acid, which reacts with sulfanilamide to produce a diazonium ion which reacts with NED to form a pink azo-chromophore which absorbs at 543 nm (Green *et al.*, 1982).

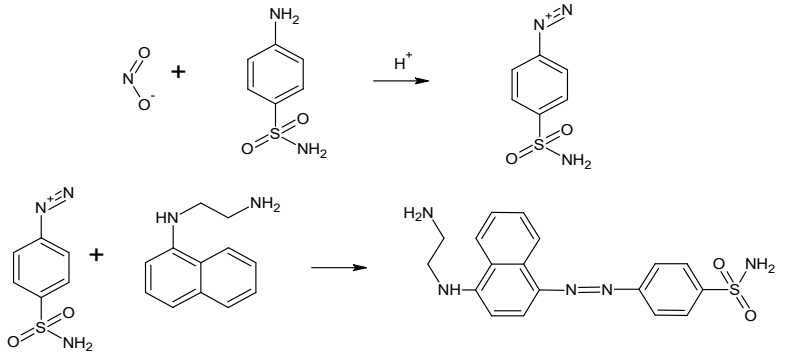


Figure 2.7: Reaction between nitrite ions and sulfanilamide in the presence of an acid form a diazonium salt intermediate that subsequently reacts with NED to form an azo dye.

NO-scavenging activity of each the available standards and phenolic extracts was determined by the method of Tsai *et al.,* (2007) with minor modifications. Sixty microliters of two-fold dilution sample were mixed with 60 µl of 10 mM sodium nitroprusside in phosphate buffered saline (PBS) into a 96-well flat-bottomed plate and incubated under light at room temperature for 150 min. Finally, an equal volume of Griess reagent prepared in-house by 0.2% naphthylethylenediamine dihydrochloride (NEDD) w/v, and 2% w/v sulfanilamide in 5% phosphoric acid v/v was added into each well and the absorbance at 550 nm was measured with a microplate reader.

The percentage NO-scavenging activity was calculated as follows;

$$\%\ \text{NO scavenging} = 100 - 100\left(\frac{Abs_{sample}}{Abs_{control}}\right)$$

2.2.6 MIR spectra accusation

Spectra were acquired in the mid-infrared range at room temperature without any further sample pre-treatment step, through the use of an attenuated total reflection (ATR) cell made of a ZnSe crystal (10 reflections at 45°angles; PerkinElmer Inc., San Jose, CA) using a PerkinElmer 1600 Series FT-IR. A volume of 10 μL of oil was deposited on the crystal and, using the press tower of the ATR set, at the constant height so that the layer of oil was uniform throughout the cell. Spectra were then acquired between 4000 and 630 cm^{-1}, collecting 90 scans at a nominal resolution of 2 cm^{-1}. A background spectrum was recorded prior each sample analysis. The crystal was cleaned after each analysis, using first hexane, followed by chloroform and wiped dry using lens paper tissues. The spectra were exported as an ASCII file using the instrumental software Spectrum (PerkinElmer Inc., Waltham, MA) and imported directly into The Unscrambler X 10.3 (CAMO Software Oslo, Norway) for all subsequent mathematical data processing.

During the data analytical stage, 11 different spectroscopic signal processing techniques were evaluated and compared: ATR correction, 5 point smoothing, subtraction of a linear baseline, multiplicative scatter correction (MSC), orthogonal signal correction (OSC), Standard Normal Variate (SNV), Savitzky–Golay 1^{st} degree derivatization, first and second derivative. Furthermore, flat line regions of the signal (3200–4000 cm^{-1}) were eliminated prior to the statistical analysis. The effect of the different spectral transformations on the final classification outcomes was compared to those obtained without any signal processing.

2.2.7 ESI-DI-MS spectral acquisition

Electrospray ionization (ESI) was used in the determination of the metabolic fingerprint of different EVOOs.

All oils were analyzed by diluting them 1500-fold in a mixture of 85:15 propanol: 40 mM methanolic KOH and filtered through a 0.2 µm PVDF micron filter. The samples were loaded into a 250 µL volume HP series syringe and introduced directly into a Waters ACQUITY® TQD tandem quadrupole mass spectrometer equipped with a Z-spray electrospray ionization (ESI) source (Waters, Milford, MA, USA) operating in positive mode. The sample was infused at a rate of 50 µL/min for 10 minutes. Nitrogen was used as the nebulizing and desolvation gas. The MS conditions were: capillary potential 3.0 kV, extractor cone voltage 30V, RF lens voltage 0.1V, source temperature 130 °C, desolvation temperature 300 °C, desolvation gas flow rate 500 L/h. Data was collected by use of MassLynx 4.0 software resident in a personal computer. All samples were injected at least three times, and each time the data were integrated at 10 min and averaged.

2.2.8 ASAP-Q-TOF-MS analysis

Samples were directly introduced into the ASAP-Q TOF MS Xevo G2 QTOF (Waters Corporation, Manchester, UK) by dipping a solid glass capillary in the liquid samples. The samples wetted the exterior of the glass capillary. Two dips were used for each analysis. Nitrogen was used as a desolvation gas at 450 l h-1flow and no cone flow was needed for this technique. Optimization of key ion source parameters, corona current (µA), sample cone (V) and desolvation gas temperature (◦C) were carried out using 283 m/z as the reference peak which corresponded to free oleic acid (M-H) +•. Optimization of these parameters was done in order to prevent extensive fragmentation of molecular ions thus maximizing number of intact molecular ions reaching the MS detector. Ambient ionization MS tend to be classified as soft ionization techniques, thus providing a spectra rich in molecular ions with minimal fragmentation. The obtained spectra were formed by the protonated molecular ion (M-H) +•. The voltage of the sampling cone was varied from 20 V to 80 V and the voltage of the extraction cone was fixed at 0.1 V. Target samples were analyzed in continuous mode (3 minutes with a cone voltage (22V) and desolvation gas temperature ramp (350◦C). Atmospheric Pressure Ionization (API) in positive polarity was selected, source

temperature was 120◦C. Each sample was analyzed in triplicate. A blank sample was also analyzed under the same experimental conditions.

2.2.9 ^{1}H NMR spectra accusation

For NMR analysis, 20 μL of EVOO was placed in 5 mm NMR tubes and dissolved in 700 μL chloroform-d followed by the addition of 20 μL of DMSO-d and vortexed for 20 s. The analysis was performed on a model AVANCE III 600 NMR spectrometer equipped with a 5mm 1H/D-BB probehead with z-gradient, automated tuning and matching accessory and a BTO-2000 accessory for temperature control (Bruker BioSpin GmbH, Rheinstetten, Germany). Samples were measured at 300.0 K after a 5 min resting period for temperature equilibration. NMR spectra were acquired using Topspin 2.1 (Bruker). Automated tuning and matching, locking and shimming using the standard Bruker routines, ATMA (automatic tuning and matching in automatic mode), LOCK (frequency-field lock to offset the effect of the natural drift of the NMR's magnetic field B0) and TopShim, were used to optimize the NMR conditions.

The conventionally used zg30 1H NMR method was selected for analysis of the EVOO owing its ability to produce sharp signals based on proton spin-spin coupling. Coupling occurs due to the interaction between protons connected together through chemical bonds. The different chemical environment of the proton influence shielding and deshielding which results in variation in the chemical shifts. Protons within the same chemical environment can couple with one another to form multiple signals. The integrated peak intensity is related to the number of hydrogen influence that particular peak. Apart from the zg30, EVOOs were also analysised using a second NMR method NOESY also known as Nuclear Overhauser Effect SpectroscopY (NOESY). NOESY detects protons that are physically close to each other in space even if they are not chemically bonded. Detection of such protons is through the usual pulsed detector probe with the addition of a decoupler probe delivering continuous wave irradiation at the resonance frequency of the irradiated nucleus. The NOE signal strength decreases rapidly with increase in distance between the atoms. Since the observed enhancement is usually of less than 20%, thus to increase the sensitivity a conventional ^{1}H spectrum is obtained by irradiation to a blank region of the spectrum and then subtracting this from the irradiated spectrum leaving only the enhanced absorption.

zg1H standard single pulse experiment zg30 was performed using the following parameters; RD – P (90°)-acquisition of the free induction decay (FID). The nonselective 90 °hard pulse P (90°) was adjusted to 10 μs. The relaxation delay (RD), and acquisition time (AQ) were set to 4 s, and ~3.27 s, respectively, resulting in a total recycle time of ~ 6.66 s. FIDs were collected into time domain (TD) = 65536 (65 k) at a resolution of 0.305Hz complex data points by setting: dummy scans (DS) = 4, number of scans (NS) = 100, spectral width (SW) =19.999 ppm and a receiver gain (RG) = 40.3.

NOESY 1D ^{1}H experiment was carried out using a standard presaturation noesypr1d NMR pulse sequence using the following parameters; RD – P (90°)-acquisition of the free induction decay (FID). The nonselective 90° hard pulse P (90°) was adjusted to 10 μs. The relaxation delay (RD), and acquisition time (AQ) were set to 4 s, and ~ 2.04s, respectively, resulting in a total recycle time of ~ 6.66 s. FIDs were collected into time domain (TD) = 32768 (32 k) at a resolution of 0.489 Hz complex data points by setting: dummy scans (DS) = 4, number of scans (NS) = 100, spectral width (SW) = 16.02 ppm and a receiver gain (RG) = 71.8.

Prior to Fourier transformation, the free induction decays (FIDs) were zero-filled to 64k and a 0.3 Hz line-broadening factor was applied. The chemical shifts are expressed in d scale (ppm), referenced to the residual signal of chloroform (7.24 ppm) (Hoffman, 2006). The region of the NMR spectra studied comprised from 0 ppm to 10 ppm. The spectra were phased- and baseline-corrected manually, binned with 0.01 ppm- wide buckets. The spectra were exported as an ASCII file using the Topspin 3.5 (TopSpin™ version 5) and imported directly into The Unscrambler X 10.3 (CAMO Software Oslo, Norway) for all subsequent mathematical data processing.

2.2.10 Synchronous fluorescence spectra (SEEF) accusation

The olive oils were and diluted 50% in spectrophotometrically pure 2,2,4-trimethylpentane (Sigma- Aldrich) prior spectrofluorometric analysis. A three-dimensional fluorescence spectrum (3D-FS) made up of excitation-emission matrix (EEM) was obtained for each sample using a Jasco FP-8300 fluorescence spectrophotometer, with both the excitation and the emission bandwidths set at 5 nm for a measurement range between 210 to 750 nm. The acquisition interval and the integration time were maintained at 0.5 nm and 10 ms, respectively, with a scan speed

of 5000 nm·min^{-1}. The oil samples were examined by means of right-angle geometry. SEEFS were acquired by simultaneous scanning of the excitation and the emission monochromators, with a constant distance, δλ, of 10, 30, 60, 80, 120 and 185 nm. All analyses were carried out in duplicates, and the results reported as mean values. Fluorescence intensities were plotted as a function of the excitation wavelength.

2.2.11 3D Fluorescence data extraction and 3-way chemometric analysis

3D fluorescence spectra were exported to separate .txt tab delimited files using Spectra Manager II (Jasco) after which a Python script opened the delimited .txt files and combined them into a single array to be used by MATLAB R2015b (Simulink, 2015) for multi-way data analysis.

For PARAFAC analysis the data from the excitation range of 220-240 nm was removed from the 3D-FS due to anomalous noise and instrumental artefacts present, furthermore areas with emission wavelengths smaller or equal to excitation wavelengths were set to a value of zero as expected by the laws of physics, namely that a fluorophore cannot emit light of higher energy than the source of excitation. The values along the diagonal where $\lambda_{ex} = \lambda_{em}$ and the values right next to this diagonal were set to missing values. An analysis where the triangular part above the aforementioned diagonal of the EEM was set to missing values representing c. 47% missing values was also carried out however this yielded unsatisfactory results. Furthermore a blank spectrum (iso-octane) was recorded with three accumulations and subtracted from all sample spectra prior to multi-way modelling.

There are two suggested ways on how to handle missing EEM data when using PARAFAC - either inputting missing data as ‘NaN’ which represents a missing value, or else by using a value of ‘0’. The two approaches were tested in the case where there was missing EEM data, corresponding to regions where $\lambda ex \geq \lambda em$ and where Raman scattering was expected (the region right next to and including the diagonal where $\lambda ex = \lambda em$). Substituting missing EEM data with ‘NaN’ led to inconsistent models which converged only after a large number of iterations. On the other hand more consistent and quickly converging models were obtained when the missing data with the

exception of the Raman scattering range was filled with zeros (Christensen *et al*., 2005).

PARAFAC and N-PLS modelling was performed using the 'N-way' MATLAB toolbox from Eigenvector. No data pre-processing treatments were applied to the input data array and non-negative constraints were applied to the PARAFAC model in all modes as no negative spectra or concentrations are expected. The convergence criteria were set to a minimum tolerance of 1×10-10 and a maximum analysis time of 1 hour using SVD for model initialization. For each PARAFAC model built, it was determined that the convergence criteria with respect to tolerance were met. The optimum number of components was determined by building 10 PARAFAC models each having a different number of components (1-10) and the optimum model was determined using split-half analysis. Each PARAFAC model was replicated ten-fold, in order to ascertain true convergence. PLS-DA analysis was also performed on scores from mode 1 of the PARAFAC model with the optimum number (4) of components using Unscrambler (SAS). For N-PLS analysis the data was prepared in a similar fashion as for PARAFAC however only the region between 270-510 nm (excitation) and 290-575 nm (emission) was used. The optimum number of components was determined by building 15 N-PLS models each having a different number of components (1-15) and the optimum model was determined using the prediction accuracy and RMSE error of calibration and validation models. Validation of the model was carried out using Venetian blinds cross-validation which selects every sth sample from the data by making s data splits such that all samples are left out exactly once (s=3). Furthermore the classification of geographical origin was also predicted for a number of test samples with a similar identity to the ones used for FTIR Analysis. Mean centring was applied in the first mode of the multi-way array which corresponds to the sample mode prior to N-PLS analysis and all missing values were set to have a 0 value.

2.2.12 Seed oil Adulteration experiment

Fresh extra virgin olive oil (EVOO), which had been previously analysed for its peroxide value, free acidity, K_{232} and K_{270} in order to establish its quality, was obtained from a small-scale local press. Samples of six commonly used adulterants, namely sunflower (SFO), peanut (PO), soya bean (SBO), linseed (LSO), corn (CO) and rapeseed (RSO) oil, were purchased from local supermarkets. Solutions of olive oil, pure, and adulterated with increasing levels of seed oil up to a composition of pure seed oil, were prepared and diluted 50% in spectrophotometrically pure 2,2,4-trimethylpentane (Sigma- Aldrich). A three-dimensional (3D) matrix excitation-emission matrix (EEM) was obtained for each sample using a Jasco FP-8300 fluorescence spectrophotometer, with both the excitation and the emission bandwidths set at 5 nm for a measurement range between 210 to 750 nm. The acquisition interval and the integration time were maintained at 0.5 nm and 10 ms, respectively, with a scan speed of 5000 $nm \cdot min^{-1}$. The oil samples were examined by means of right-angle geometry. Synchronous fluorescence spectra were acquired by simultaneous scanning of the excitation and the emission monochromators, with a constant distance, $\Delta\lambda$, of 24 nm. All analyses were carried out in duplicates, and the results reported as mean values. Fluorescence intensities were plotted as a function of the excitation wavelength.

2.2.13 Thermal oxidation Studies.

Oil samples (14 mL) were poured into 15 mL amber glass bottles and capped tightly. A small headspace was used in order to account for the volumetric expansion of the oil during the heating process, preventing excessive build up of pressure. 12 amber bottles for each cultivar were prepared and placed in a small oven at 56°C. The oven window was covered with aluminium foil in order to further mitigate light exposure. The study took place over a period of 12 weeks whereby one sample for each cultivar was withdrawn from the oven on a weekly basis and frozen until analyzed.

2.2.14 Application of chemometrics models for the geographical discrimination of origin.

Traditionally, analytical methodology involving the analysis of a complex mixture compounds involved the identification of each and every individual compound followed by the quantification of the predefined specific chemical marker. However, this methodology has a number of disadvantages, as it involves the identification of a sufficiently adequate number chemical marker form a very complex signal, which would later enable the correct of identification of the sample. Modern techniques in the field of chemometrics offer another solution to the problem, by attaining a more holistic view, through the analysis of the complete set of unidentified and unquantified markers a "fingerprint" of the material under investigation is obtained. This allows the determination geographical origin without the prior need to identify and quantify the specific markers since the specificity lies within the complexity of the signal obtained. Taking a more practical scenario from the field of chromatography, the identification of a particular mixture of compounds can be tackled by considering the chromatographic profiles as continuous and non-specific signals, avoiding a priori identification of compounds. As a result, one analysis of one single sample can easily yield megabytes of data.

Chemometrics can be defined as a chemical discipline that employs the use of statistical and mathematical methods, designed to select an optimum number of variables in order to provide maximum chemical information with the aid of multivariate analysis. The use of multivariate analysis provides a more realistic approach for data analysis and thus gains a new and higher quality of data evaluation. There are two different approaches for multivariate analysis one which is model driven and one which is data driven. The model-driven statistical approach is generally preferred by statisticians whereby the data is seen as a realization of random variables with an underlying statistical model. The data-driven approach is on the other hand preferred by the chemometricians whereby the multivariate statistical tools are seen as algorithms that can be applied to the data available. This data-driven approach is adopted due to the type of data which is handled, whereby the number of variables are much higher than the number of observations, making the application of standard

classical statistics to fail. Figure 2.8 shows a typical workflow which is carried out by chemometricians in order to identify the origin of different samples.

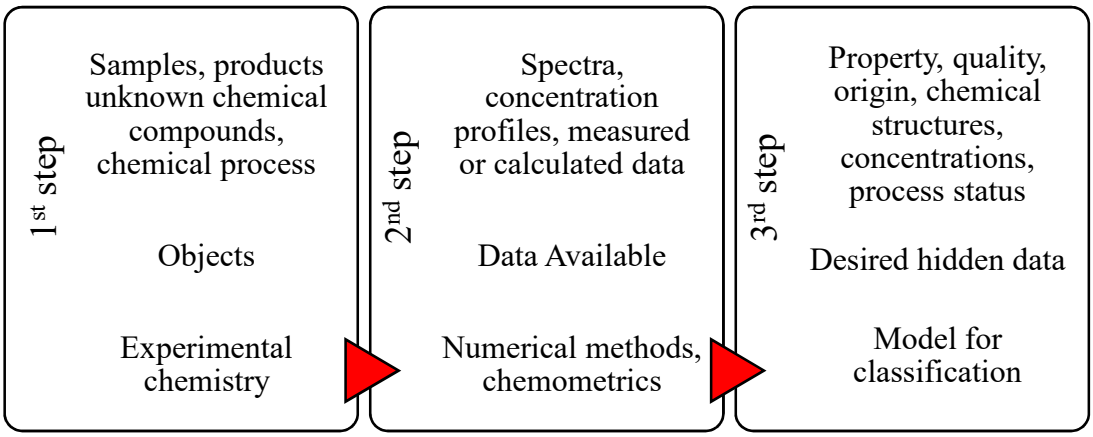


Figure 2.8 Flow diagram showing the major steps which enabled the use of chemometrics for the determination of geographical origin

The aim of the present study lies in identifying a series of chemical markers by using data obtained from different analytical techniques including spectroscopic, spectrophotometric and chromatographic data in order to discriminate geographical origin of olive oils. This was achieved through the development of a strategy based on the analysis of the available data using multivariate analysis, in order to work directly with the whole raw data. The raw data was treated in different ways depending on the number of variables which contained.

2.2.14.1 Preliminary data treatment

In order to preliminary reduce the number of variables in order to allow for downstream multivariate data processing, in some cases the data was trimmed whereby the uninformative parts of the spectra or chromatograms were removed. Other preliminary data reduction involved the maximum or average integration at certain intervals whereby only maximum or average values were obtained at constant intervals (Binning). Baseline removal was also employed in order to further reduce the number of variables. In case that the number of variables was still large and the computing power was not enough in order to allow for multivariate statistics to be carried out the data was bucketed. Data bucketing involves subdividing the number of variables into different groups and subjecting each individual bucket to multivariate statistics. This allowed a more focused variable extraction within each bucket. Once that all the buckets were treated and the variables which showed the highest discriminatory power were obtained the multivariate methods were repeated once again this time using all the selected variables from each bucket. The bucketed data does not require to be the same size and thus buckets containing a different number of variables were sometimes employed depending on the nature of the data.

2.2.14.2 Multivariate and Univariate normality testing

Normality testing is one of the most neglected tests found in the scientific research a review paper published by Curran-Everett and Benos in 2004 stated that at least 50% of the published literature had one statistical error. The most commonly used statistical procedures which are used in any form of research include correlation, regression, t-tests, and analysis of variance. All the aforementioned test are parametric tests which are based on the assumption that the data follow a normal distribution or a Gaussian distribution. In order to draw accurate and reliable conclusions for the statistical tests, normality and other assumptions need to be taken into consideration (Field, 2009). However, with large sample sizes (> 30 or 40), any violation of normality should not cause major problems (Pallant, 2007). This implies that parametric procedures can be employed even when the data are not normally distributed (Elliott 2007). This done according to the according to the central limit theorem, which states that if the sample data are approximately normal then the

sampling distribution will so be normal furthermore the sampling distribution for a large sample size (>30) tends to be normal regardless of the shape of the data (Field, 2009). Although normality tends to be overlooked in this study even though a large sample size (≈ 70 samples) both univariate and multivariate normality tests were carried out. In this study of univariate normality, testing was carried out using Shapiro-Wilk test (SPSS Inc). In the case of, where the null hypothesis was rejected (p-value < 0.05) non-parametric tests were carried out.

Similar to univariate analysis some multivariate statistical methods, such as linear discriminant analysis (Venables and Ripley 2002), and principal component analysis (Husson *et al*., 2014) require multivariate normality (MVN) assumption. The performances of these methods dramatically decrease if they are applied to datasets which are non-multivariate normal. Thus, prior conducting this statistical analysis an MVN testing is usually carried out in order to determine the suitability of the data for a particular test. According to the review by Mecklin and Mundfrom (2005), more than fifty statistical methods are available for testing MVN. Mecklin and Mundfrom (2005) stated that there is no single test which excels in both type I (incorrect rejection of the null hypothesis) and type II (incorrect acceptance of the null hypothesis) error. The authors suggested using Henze-Zirkler's and Royston's tests among others for assessing MVN because of their good type I error control and power. Moreover, they suggested that the use of Mardia's test should be used in order to diagnose the reason for deviation from multivariate normality. In this study, Henze-Zirkler's and Royston's test were used in order to detect any deviations from normality. This was done through the use of web-friendly interface created by Korkmaz, Goksuluk and Zararsiz (2014) available at http://www.biosoft.hacettepe.edu.tr/MVN/.

In the case of MVN, this was restricted linear discriminant analysis (LDA) since it is not a strict requirement for PCA since it was only used for exploratory purposes and is not a p-value driven technique, similarly MVN testing was not used for PLS regression since it is not a strict requirement (Sawatsky *et al*., 2015). In case of LDA, when the data was multivariate normal a Bayes LDA or QDA was carried out, if on the other hand that data was not multivariate normal a Fisher LDA was used. Such an approach was gave very good results when it came to the analysis of the PARAFAC results, however when it came to the other methods, the Fisher LDA

models obtained for non-normally distributed data were highly over fitted. Such models, were able to predict the geographical origin of EVOOs for the training set but failed completely when it came to the testing set (predictability of the model) thus Bayes LDA was carried out since it proved better results even if the data was not normally distributed. Appendix 19 shows some of the results obtained which support the aforementioned statement.

2.2.14.3 Multivariate statistical techniques

A number of different multivariate techniques have been used in this study, these can be broadly be divided into two supervised and unsupervised statistical techniques. Unsupervised statistical techniques are able to extract regularities and identify the natural clustering pattern on the basis of similarities among the samples directly from the input data without referring to classes known in advance. The most commonly employed unsupervised statistical techniques for pattern recognition and discrimination between samples of different geographical origin include cluster analysis (CA), principal component or coordinate analysis (PCA, PCoA).

2.2.14.3.1 Unsupervised multivariate statistical techniques

2.2.14.3.1.1 Principal component analysis

PCA is primarily a dimension reduction technique method used for extraction of variations in one data set (Kettaneh *et al*., 2005). PCA pursues a linear combination of variables in order to extract the maximum variance from the variables. It then eliminates the extracted variance and pursues a second linear combination which explains the maximum proportion of the remaining variance, and so on. PCA can be used to determine which variables (for spectral data, wavelengths) describe the differences between samples, which of these variables contribute most to an observed difference and which variables are correlated to each other and thus are contributing to the same apparent variation. Although it is not the main aim of PCA it also enables the detection of sample patterns and possible clustering of observation and to detect outliers.

In matrix representation, the model with a given number of components has the following equation:

$$X = TP^T + E$$

where T is the scores matrix, P the loadings matrix and E the error matrix. The combination of both the scores and loadings provides the most informative part of the data. The remaining part which cannot be modelled well E is known as the error or residual.

In order to fully interpret the PCA an inspection of the loadings is usually carried out. Loadings provide information on how the variables vary along a specific model component. In the case of spectroscopy, the loadings need to have a "spectral characteristics" about that show minimal noise characteristics. Given that principal components have an orientation which links the samples to the variables by means of scores and loadings. Loadings can be either negative or positive value; so can scores. In general variables with a very small loading, indicate that these are not accounted by that principal component and thus they are not contributing to the observed variation in the data. On the other hand, if the variables exhibit a positive loading, samples with positive scores have higher than average values for that particular variable and vice versa.

Some of the assumptions which need to be taken into consideration in PCA include the normality of data. Unlike factor analysis, the principal component analysis does not make any assumption about an underlying causal model whereby each observed variable is assumed to be normally distributed. However when the sample size is larger than 25 the model obtained is still robust against violations of normality (Hatcher and Stepanski, 1994). Apart from normality PCA assumes random sampling whereby each observation will contribute one score on each observed variable.

2.2.14.3.1.2 Principle component regression (PCR)

PCR is a two-step process, the first step which consists of a PCA on the x-variable set, in order to reduce the reduce dimensionality and extract the set of the uncorrelated scores. In the second step, involves the application of a standard multiple linear regression analysis (MLR) using principal component scores as the x-variable set (Esbensen, 2002). MLR is an extension of a univariate regression with the difference being that in MLR one y-variable is regressed against several x-variables through the use of least squares fitting. The critical drawback of this method is that all x-variables must be linearly independent, that is no significant X-variable collinearity

is allowed, the use of principal component scores avoids this problem when PCR is applied. Furthermore the number of X-variables must be smaller than the number of samples this is further avoided through the application of PCA which is in itself a dimensional reduction technique, outlier can also affect the accuracy of MLR (Esbensen, 2002; Naes *et al*., 2002) which were previously detected and removed through the application of PCA.

MLR can be expressed in the following matrix form

$$y = Xb + f$$

The overall aim is to determine the vector of regression coefficients b that minimizes f, the error term, this is done through the application of the least squares method. The assumptions of PCR are the same as those used in regular multiple regression: linearity, constant variance (no outliers), and independence. Since PC regression does not provide confidence limits, normality need not be assumed.

2.2.14.3.1.3 Soft independent modelling of class analogies (SIMCA)

In SIMCA each class model is defined on the basis of a principal component model which is previously built, in order to reduce the dimensionality of the data but at the same time retaining the maximum information defining the class. For instance, the model of class "Maltese" can be assumed to be described by a principal components model, according to:

$$X_{Maltese} = T_A P_A^T + E$$

Where similar to the equation used in PCA $X_{Maltese}$ is the submatrix of the original data set generated by means of selecting only the samples from Maltese, T_A and P_A are the matrices containing the first A scores and loading vectors, respectively, and E are the residuals. The same process repeated for the "Foreign" EVOOs. Prior developing a SIMCA model an overall PCA was carried out in order determine if the data under study exhibits any tendency to cluster by the classes furthermore it also enabled the identification of outliers, which were removed prior further analysis. The application of PCA on the individual categories enables the determination of the directions of maximum variance in the class space. Consequently, no attempt is made separate

classes, on the opposite of, partial least squares discriminant analysis (PLS-DA), which directly models the classes on the basis of the descriptors (Marengo *et al*., 2006).

In the SIMCA model, the score matrix T_A and the residual matrix E are used to compute the probability distributions for the distances within the model space (T^2 statistics) and for the orthogonal distance to the model space (Q statistics), respectively. Threshold values for the probability distributions of the distances within the model space and for the orthogonal distance to the model space is selected as a percentile of the distributions (usually 95%, however, depending on the sensitivity required this can be decreased to 75%). It is done by computing the T^2 and Q ratio for each sample which is defined as a distance to the class model in terms of reduced variables.

$$d_i = \sqrt{(T^2_{red,i})^2 - (Q_{red,i})^2} = \sqrt{(\frac{T_i^2}{T_{lim}^2})^2 + (\frac{Q_i^2}{Q_{lim}})^2}$$

Where T^2_{red} and Q_{red} represent the distances within the reduced model space and the orthogonal distance to the reduced model space respectively, whilst T^2_{lim} and Q_{lim} represents the threshold values for the two statistics corresponding to a selected percentile of the distributions. Commonly, if the reduced distance of a sample (T^2_{red} and Q_{red}) exceeds new classification, the sample is considered as an outlier and rejected by the class model, if the distance is lower than this value, it is accepted and recognized as being part of that class. SIMCA it is defined 'soft' since no hypothesis on the distribution of variables is made, and 'independent' since the classes are modelled one at a time-independent for each other.

One of the major outputs of SIMCA is a Coomans plot which takes the form of a graph where the two axes represent the distance of the samples to each of two class models. The horizontal and vertical lines corresponding to the threshold distances also known as the significance limit can be adjusted up depending on the sensitivity required. These lines divide the Coomans plot into four different regions: the uppermost left and the lowermost right which correspond to unmistakeable acceptance by a single category model, the lowermost left to acceptance by both classes while the uppermost right to rejection by both category models. Under certain condition, a third line which dissects the Coomans plot diagonally is sometimes employed when there is an overlap

between the two classes and the threshold limits fail to separate the two individual classes (Bevilacqua *et al*., 2012 and Vitale *et al*., 2013) . The diagonal line bisecting the plot represents the discriminant classification boundary so that all the samples lying above are classified as being part of one class, while all the samples lying below are predicted to belong to another class.

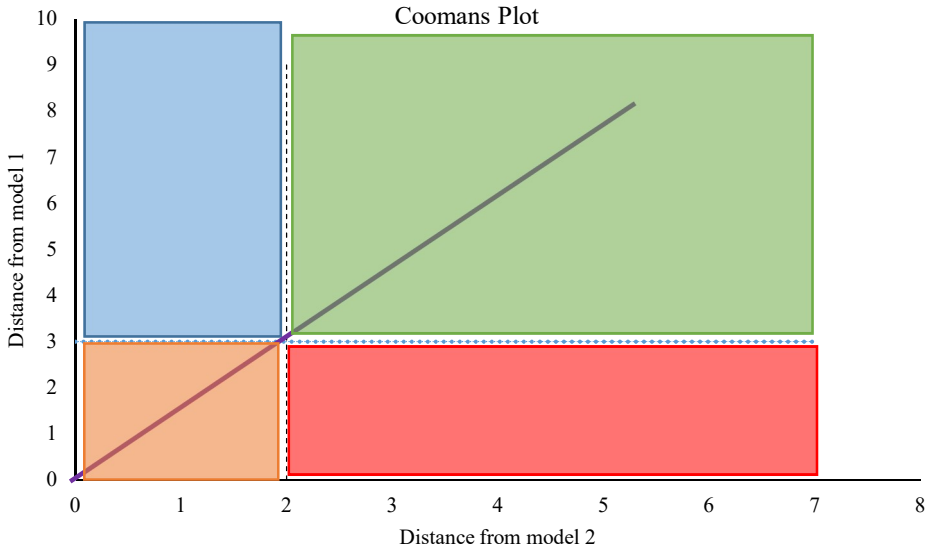


Figure 2.9: Coomans Plot showing the different regions which are employed in the discrimination of observations. Blue and red areas represent the unmistakable acceptance of observations by model 1 and model 2 respectively. The orange represent the area where the samples can be accepted by both model 1 and model 2, whilst the green represent the area where the samples are rejected by both the models. The diagonal dissecting purple line represent a new classification boundary which is sometimes added when the model fails to discriminating between the different classes.

2.2.14.3.2 Supervised multivariate statistical techniques

In the case of principal components analysis, major objective is to dimensionally reduce the data into a series of components constructed for the sole purpose to explain as much of the variation of the observed data as possible, without taking account of the important variable to the response. On the other hand, the application of supervised methods enables the construction of components which are able to model both the X (predictors) and Y (response) matrices in order to maximize the covariance between the responses and the predictors. The application of supervised methods is preferred since there is no theoretical basis why the constructed components explaining large predictor variation only, should also have a predictive ability with respect to the of the response (Nguyen and Rocke, 2002).

A number of supervised statistical techniques have been used in this study these include partial least squares regression analysis (PLS), artificial neural networks (PNN) and stepwise linear canonical discriminate analysis (SLC-DA). Supervised statistical technique both (both classification and regression) attempt to discover the relationship between input independent variables and the target dependent variable. The main objective of supervised techniques is to discover relationships between the two variables by the aid of complex algorithms in order to obtain a model which can predict future observations. These techniques have been successfully been employed to extract variables and identify samples from different geographical origin, studies involving the use of soft independent modelling of class analogy (SIMCA) and support vector machine classification and regression have also been employed.

2.2.14.3.2.1 Partial least squares regression analysis (PLS)

PLS is a supervised learning technique which is based on the relationship between the signal intensity and the geographical origin of the sample (Martens *et al.,* 1979). PLS is used when working with highly dimensional data such as spectra and we want to conduct multiple linear regression with the aim to estimate unknown values of the response y. PLS works by extracting successive linear combinations of the predictors, called factors (also called components, latent vectors, or latent variables), which optimally address one or both of these two goals—explaining response variation and explaining predictor variation. This sophisticated technique allows for the interference and overlapping of the spectral information (Fuller and Griffiths, 1978; Haaland and Thomas, 1988), furthermore, it can take in consideration the full or partial spectral region rather than unique and isolated analytical bands. The algorithm obtained is based on the ability to mathematically correlate spectral data to a property matrix of interest in this case being the geographical origin of the cultivar (Dupuy *et al.,* 2005).

Samples of known origins are used as calibration samples, and then the origins of unknown samples are directly calculated using the resulting equation under the same conditions. In this study the performance of the PLS model was evaluated on using the complete spectral or chromatographic profiles information and also using an adjusted (cropped) PLS analysis whereby regions of the spectrum or chromatogram which had a variable importance factor (VIP) less than 0.8 were eliminated and the performance of the model was evaluated again. For a further interpretation of the models obtained an inspection of the VIP scores was carried out. The VIP of a predictor is a value that expresses the contribution of the individual variable in the definition of the latent vector model. In particular, it is defined according to the formula:

$$VIP_j = \sqrt{N_{vars} \frac{\sum_{k=1}^{F} (b_k^2 \, t_k^T t) \left(\frac{w_{jk}}{W_k}\right)^2}{\sum_{k=1}^{F} (b_k^2 t_k^T \, t)}}$$

where t_k is the vector of sample scores along the k^{th} latent variable, b_k is the coefficient of the k^{th} PLS inner relationship, N_{vars} is the number of experimental variables and w_{jk} and W_k are the weight of the j^{th} variable for the k^{th} latent variable and

the weight vector for the k^{th} latent variable respectively. A VIP score is a measure of a variable's importance in modelling both X and Y. If a variable has a small coefficient and a small VIP, then it is a candidate for deletion from the model (Wold, 1994) a value of 0.8 is generally considered to be a small VIP (Eriksson *et al*, 2006). However Bevilacqua *et al*., 2012 stated that since the average of squared VIP scores equals 1, they employed 'greater than one rule' as a criterion to identify the most significant variables.

The goodness of fit of the PLS regression model obtained was evaluated using % of variability explained in terms of X and Y which give an indication of the portion of the data explained by means of the fitted algorithm. In general, the higher the number latent of variables extracted from the regression analysis obtained the higher is the % variability explained in terms of X and Y. For a specific number of latent variables, the prediction error sum of squares is found by comparing the value obtained from the prediction formula to the observations found in the validation set. The smaller the value of PRESS, the better the prediction for a random effect model, with a value of 0 indicating perfect prediction.

Other statistical details which were taken into account in order to determine the goodness of fit of the model are the van der Voet T^2 and the Prob > van der Voet T^2. Van der Voet T^2 is a statistical test which determines whether the model obtained with a specific number of extracted latent variables differs significantly from an optimum model. The test is constructed on the null hypothesis defined as the squared residuals for both the model obtained and the model have the same distribution. In simpler terms, the null hypothesis states that both models have the same predictive ability. A common practice is to extract the smallest number of latent factors, in order to prevent the model from overfitting, for which the Prob > van der Voet T^2 level exceeds 0.10.

Application of PLS regression for discrimination or classification purposes is possible through the application of PLS-DA whereby a discriminate analysis (LDA or CDA) is carried out on the reduced data matrix. Through the application of The Unscrambler® PLS-DA is not listed as a separate method and employed the same algorithm as the one used during PLSR. Although the latter works with a continous response, the geographical origin of the EVOOs needs to be transformed in such a way

that it appears to be a numerical variable response rather than a categorical one. Although this might seem to be inappropriate it has been employed throughout the literature (Alonso-Salces *et al*., 2010ab; De Luca *et al*., 2011; Karim *et al*., 2015; Rezzi *et al*., 2005; Mannina *et al*., 2010; Longobardi *et al*., 2012 and much more) which employed the same statistical program. The geographical origin is transformed in suitably designed dummy response vectors so that, traditional regression methods can be used also to tackle classification problems. In particular, when dealing with a classification problem involving 2 classes, as in our case, one can build a dummy binary- coded one-dimensional response vector, so that, if a sample belongs to the class "Maltese", it will be described as 0, while if it belongs to class "Foreign" (other origins), it will be coded as 1. Under these assumptions, to compute a classification model corresponds to calculating the regression vector between the data matrix and this dummy vector of responses (Bevilacqua *et al*., 2012).

Since a non-conventional PLS-DA was carried out but rather a PLS regression was carried flowed by a discriminate analysis on the actual output, it is assumed that both the input and output variables (geographical origin) are continues. The actual discriminate part was carried out through the use of 0.5 cutoff point. Although this is not true it needs to be assumed as being as such in order to carry out the regression. Samples having a regression output higher than 1 were classified as foreign EVOOs whilst samples with a negative regression output were classified as 0 (Maltese EVOO). The cutoff point was used to classify the remaining samples, an output smaller than 0.5 were classified as 0, conversely, output at 0.5 or higher was classified as 1. In comparison, to other methods, PLSR requires fewer assumptions, these include the normality of the data. Although not imperative, the data should be relatively normally distributed and prior any form of regression, the data were screened for influential extreme outliers prior to analysis. Outlier analysis was carried out using both PCA, however it is acknowledged that other methods such as Mahalanobis distances and adjusted quantile of the chi-square distribution methods exist and enable a more direct way for the identification of outliers. Prior carrying out PLS regression analysis the data was centred and scaled prior to analysis to ensure that each variable has an equal opportunity to influence the model.

Although it was not the main aim of the study for certain techniques a second classification system was employed in order to determine the complete classification of olive oils depending on the country of origin using the following a five binary coordinate numerical assignment, whereby the origin of the olive oil was defined by five numbers depending on the geographic origin: indigenous Maltese cultivars (0,0,0,0,0); foreign locally grown cultivars (1,0,0,0,0); Italian cultivars (0,1,0,0,0); Spanish cultivars (0,0,1,0,0); French cultivars (0,0,0,1,0) Greece cultivars (0,0,0,0,1).

2.2.14.3.2.2 Predictive Artificial neural networks (PNN)

Artificial neural network (PNN) is a mathematical algorithm with the capability of relating the input and output parameter, learning from examples through iteration without requiring a prior knowledge of the relationships between the process variables (Cevoli *et al*., 2011). The main advantage of a neural network model is that it can efficiently model different response surfaces due to its nonlinearity, allowing a better fit to the data given enough hidden nodes and layer, providing an accurate prediction for kind of data. Unlike other modelling and discriminate methods (SIMCA, LDA, PLS) the main disadvantage of a neural network model is that the results are not easily interpretable, due to presence of intermediate hidden layers which direct path from the X variables to the Y variables, as in the case of regular regression but cannot be fully interpreted.

In this study, TanH activation function was employed as the standard neuron activation function in JMP software. TanH function transforms values to be between -1 and 1, acting like the centred and scaled version of the logistic function. In general, PNN techniques are a family of mathematical models that mimic the human brain function. PNN methods share the same common concepts these include the use of “neurons” (also called “hidden layers”). Each neuron represents a synapse which is controlled by activation functions. Once the threshold of these activation functions is reached propagation of neuron signal to the next layer can occur. A hidden layer is composed of a regression equation which is able to processes the input information into a non-linear output data. The larger the number of hidden layers (neuron) the larger number of non-linear correlations can be treated at once.

There are two types of PNN, the feed-forward networks which are composed of unidirectional connections between network layers, where the connection flow from the input to output direction. On the other hand the feedback PNNs, where the connections among layers occur in both in the forward and backward directions (Zurada, 1992; Zupan and Gasteiger, 1999; Gasteiger and Zupan, 1993). In this study, the use of feed-forward PNN's was used, where the input supplied by the spectral data (factors) in order to develop an algorithm which predicts the country of origin (response).

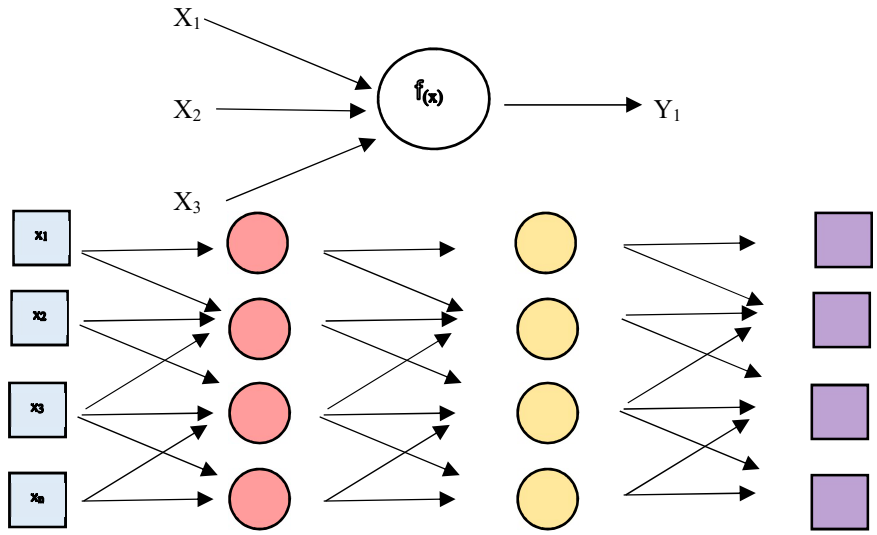


Figure 2.10 Top Scheme of synaptic connections in a node or unit within a multilayer perceptron (MLP). Bottom General Scheme of a MLP with multiple input variables two hidden layers (red and yellow circles) and output response (purple boxes).

Three different cross-validation techniques can be employed in order to prevent model overfitting; the *k*-fold, holdback and excluded rows. For the *k*-fold cross-validation, the original data were randomly divided into *k* equivalent subsamples, and a single subsample retained as the validation data for testing the model, while the remaining $k-1$ subsamples were used as training data. Similarly, holdback validation was carried out by randomly selecting a portion of the data for training whilst 0.33% of the data was used as a holdback portion for testing. Unlike the other cross-validation techniques whereby the testing and training portions were randomly selected, in the case of the excluded row holdback the testing portion of data is chosen by the user, approximately 40% of the observations were excluded, the algorithm was fitted and cross-validated with the excluded samples.

2.2.14.3.2.3 Discriminate analysis.

Discriminant analysis enables the classification and predictions of distinct observations into groups or categories which best describe the membership of the observation based on observed values of several continuous variables. Like the rest of the supervised techniques, the identity of the group membership must be known prior carrying out discriminate analysis. The difference between discriminant analysis and logistic regression is that in logistic regression, the classification variable is random and predicted by the continuous variables. On the other hand discriminant analysis, the classifications are fixed, and the covariates (Y) are realizations of random variables. However, in both cases, the categorical value is predicted by the continuous variables.

Discriminant analysis refers to a multivariate technique for classifying a set of observations into a previously identified set of groups. In the case of PCA whilst the sole scope is to reduce the dimensionality of the data through the use of highly variable predictors. In the case of discriminate analysis is like PCA, reduces data dimensionality but the main goal is to maximize the separability between the two different classes. The discriminate analysis employs the use of predictors in order to create a new axis and projects the observation onto the new axis in such a way to maximize the separation between the different classes. The new axis is built using two criteria which are considered simultaneously, the first is the maximization of the distance between the means of the two groups (μ), the second criteria is the minimization of the variation within each category also known as the 'scatter' (S2). These are considered simultaneously through the use of the equation below;

$$\frac{(\mu_1 - \mu_2)^2}{S_1^2 + S_2^2}$$

Where μ_1 and S_1^2 represent the mean distance and the scatter respectively of one group whilst μ_2 and S_2^2 represent the mean distance and scatter of the other class. In an ideal separation the mean distance between the groups is large whilst the scatter between the observation is minimal. Similar to PCA discriminate analysis rank the new axes which are generate in the order of importance comparable to the principal components in which PC1 and PC2 account for the most variation in the data whilst axis generated in DA are ranked in order to account for the most variation between the

different categories. Furthermore, whilst in PCA an inspection of the loading scores provide information on which predictor is accounting for the majority of the variation, in discriminate analysis predictors which are the most significant are given in terms of correlations.

In this study, two types of discriminate analysis were carried out using two statistical software JMP (SAS) was used for stepwise linear discriminate analysis (SLC-DA) whilst The Unscrambler 10.3 (Camo) was used for the conventional linear discriminate analysis (LDA) on the most in informative variables. The stepwise analysis allows the manual selection of variables, each time that a variable is selected the F ratios and p-values are updated. Selection of predictors to be included in the LDA models was performed by using the JMP stepwise algorithm. According to this algorithm, a predictor is selected when a reduction of the Wilks Lamda (λw) is produced after its inclusion in the model. Variables were selected on the bases of F Ratio and Prob>F given, variables which have the lowest p-value for the Prob>F and the highest F-ratio are the first to be selected. The entrance of a new predictor modifies the significance of those predictors, which are already present in the model and those which are not yet included in the model. The process terminates when there are no predictors entering or being eliminated from the model so that the all the remaining variables will have an F-Ratio and Prob>F equivalent to 0 and 1 respectively. However since SLC-DA was also used to extract the minimum number of the most important variables the model was terminated as soon as the Prob > F reached 0.05 representing the last significant variable to be included in the model at the 95% confidence level.

Similar to a PCA, canonical discriminate analysis is also represented by a biplot where the axes are the first two canonical variables. The first two canonical variables provide the maximum separation among the different groups into a two-dimensional space. Each observation within the biplot is represented in terms of canonical variables and the covariate which contributes more to that specific canonical function. Covariates are represented as rays on the biplot, the direction towards which they point indicates the degree of association of that covariate with the first two canonical variables. A good discriminatory model will classify observations belong to same group-very closely to the mean centroid group. Eigenvalues in canonical discriminate analysis are the product of the between matrix and the inverse of the

within matrix. The larger the eigenvalue the larger is the amount of variance explained by its associated discriminant function. Canonical correlation values are the result of correlation analysis between the covariates and the groups defined by the category. The larger the value of the correlation the more close are the covariates as one set of variables and the indicator variables representing that category.

Unlike other supervised methods LDA comes with a number of assumptions which need to be satisfied these include independent observations and that the predictors need to have a multivariate normal distribution. The latter was checked through the use of Henze-Zirkler's and Royston's multivariate normality tests. Another important assumption in LDA is the within-group covariance should be equal across groups. The latter assumption was checked using Box's M test. Under Box's M test the null hypothesis is that the observed covariance matrices of the dependent variables are equal across groups, thus a p-value smaller than 0.001 indicates a non-homogenous dataset. In the case of normally distributed data set a non-homogenous data set a quadratic discriminate (QDA) analysis was carried out. If on the other hand the data was neither normal nor homogenous a Fisher LDA was used since the departure from normality overrides the need of homogeneity. The last assumption of an LDA is that the group membership is mutually exclusive, that is there is no significant overlap between the different classes as they are truly categorical.

2.2.14.3.2.4 Support vector machines (SVM)

Support vector machines (SVMs) have been developed by Vapnik (Vapnik, 1998) and extensive details about the SVM algorithm is beyond the scope of this study however these can be found in literature (Amendolia *et al*., 2003; Cristianini and Shawe-Taylor, 2000). The basic idea behind SVM is to obtain an optimal hyperplane which enables the separation of linearly separable observations. However, SVM can also be extended to non-linearly separable observations through the use of Kernel functions which modify the 3D space of the hyperplane in order for non-linearly separable observations to become separable. SVM employ the use of support vectors which can be defined as data points which are found very close to hyperplane and thus they are the most difficult to classify but on the other hand they provide a direct bearing to the optimum location of the decision surface. The overall aim of SVM is to maximize the separation margin between the different classes using a small subset of the training samples as support vectors to obtain a decision function.

Without going into extensive detail, in order for SVM to find the optimal hyperplane employs the use of an optimization problem which employs the use of Lagrange multipliers (α) and dot products that transform the problem in a format that can be solved analytically and in a linear fashion. It is the modification of the dot product function which enables it to be replaced with a Kernel type function (Φ) in order to obtain a higher dimensional non-linear function.

In order for SVMs to maximize the margin of separation between the two classes and at the same time reduce the overall error (empirical risk) a new trade-off parameter between the two objectives is introduced (commonly known as the C parameter). If the C parameter is very low then the errors produced during the training stage become less important thus risking the model to become overfitted. Unlike LDA, SVM and the inclusion of any form of kernel implicitly for a non-linear transformation, requires no assumptions regarding the functional form of the transformation that is whether it is normal or homogenous.

2.2.14.3.2.5 Assessment of model performance

The prediction values obtained for the numerical coordinate system were round up to the nearest integer and then compared to the actual previously assigned values. In the case of negative values these were assigned a value of zero whilst those which had a higher value than unity were assigned a value of one, this method for assessing misclassifications. In case of classification methods such as PNN, LDA, SVM and SIMCA there was no need to round up the predicted numerical coordinates and were used without any subsequent modification. The % misclassification was determined accordingly

$$\%\, missclassifiction = \frac{K_m}{K_t}\ x\ 100$$

Where K_m is the number of misclassified observations and K_t is total number of observations used to generate the prediction model. In the case of PLS an inspection of PRESS was also carried out in order to provide an indication about the sensitivity (the variation from their actual true value) of the model under different spectral transformations and variable selection methods.

Results and Discussion

3. Bimolecular analysis

3.1.1 Morphological diversity

During sample collection it was noted that whilst for trees belonging to the 'Bidni' cultivar the overall morphological appearance was constant, in the case of the 'Malti' cultivar/s this was found to be highly varied. Throughout the collection at least five different morphotypes were noted and collected, as illustrated in Figure 3.1. Preliminary morphological analysis showed that the individual morphotypes had pronounced differences in the shape and size of the fruit, which are reflected in their common Maltese nomenclature, nonetheless they are still classified as 'Malti'. For the purpose of this study, it was sufficient to identify the individual morphotypes using the drupe morphology. However it is acknowledged that morphological differences in the inflorescence and in the leaf morphology need to be taken into consideration for a more accurate botanical description. This was however outside the scope of the study.

Figure 3.1: Five different types of morphologies commonly referred to as 'Malti' cultivar/s found across the Maltese islands. Commonly all classified as 'Malti' but are known by different names (Left to right) l-irqiqa, tawwalija, żengulija, mezzana, il-ħoxxna.

The first morphotype, known by the common name '*irqiqa*' (literally translated to "thin one"), had small spherical to slightly elongated fruits ranging from 0.5 - 0.7 cm which turn to dark purple during maturation. The fruits also had a slight asymmetrical shape with a pointed apical tip and rounded base, together with a tenuous nipple. The fruit had a very high endocarp to mesocarp ratio, as the spherical endocarp was very prominent once the fruit was dissected. The presence of a higher stone/mesocarp ratio and relatively low oil content, suggests that this morphotype could either be related to the wild type olive cultivar *Olea europaea* var. *sylvestris* which is also characterised by a similar fruit and endocarp morphology, or is a feral form. Feral forms are olive cultivars that have escaped from orchard cultures and become established in wild eco-systems where they revert back to their wild morphological traits generating a continuity in the morphological variation, making the distinction between the wild and feral communities more difficult (Besnard *et al.,* 2001 Breton *et al.,* 2006).

The second morphotype, known by the common name '*tawwalija*' (translated to "elongated one"). As the name suggests the fruits of this morphotype are elongated, ranging from 1.5-1.8 cm in length and 0.5-0.7 cm in width, and turn purple on maturation. The fruits are asymmetrical, with a pointed apex and a truncated base with a tenuous nipple. The endocarp to mesocarp ratio is slightly smaller than the former morphotype however it is still small when compared to other cultivate olive cultivars. The endocarp is elongated, asymmetrical and pointed at the apex.

The third morphotype, known by the name '*żengulija*' (translated to "ovoid one") is characterised by small ovoid fruits ranging from 1.0-1.5 cm in length and 0.5-0.7 cm in width, and which turn purple on maturation. The fruits are symmetrical with a round apex, a rounded base and an absent nipple. The endocarp to mesocarp ratio is smaller than the first morphotype and resembles more the second morphotype. The endocarp is spherical with a rounded, slightly asymmetrical and with a round apex.

The fourth morphotype reported in this study is known by the name '*mezzana*' (which translates to "middle one"). The morphotype resembles traits derived from both the first and the last morphotypes. The '*mezzana*' fruits have symmetrical, spherical shape with a slightly pointed base, central apex position and a tenuous nipple. The

endocarp to mesocarp ratio is small, similar to other cultivated olive trees. The endocarp is ovoid, slightly asymmetrical and with a round apex.

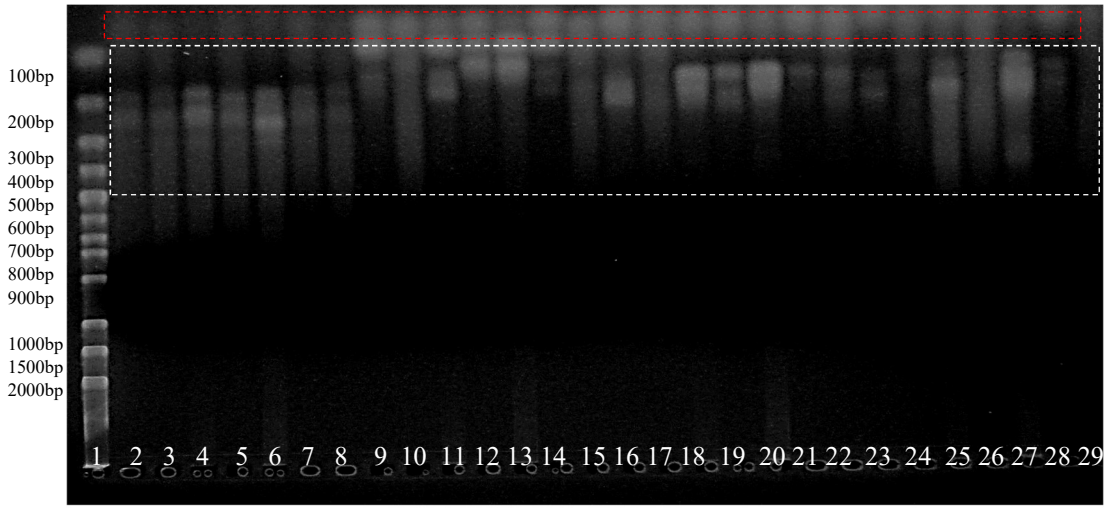
The last morphotype identified in this preliminary study is known by the name '*ħoxxna*' (translated to the "*fat one*"). As the name implies the endocarp to mesocarp ratio is smaller than the first morphotype and it is one of the distinguishing characteristics. The fruit are predominantly spherical, slightly asymmetrical, ranging from 1.5-1.7 cm in diameter and 1.8-2.0 cm in length. The fruits have a slightly pointed base, central apex position and a prominent nipple. The endocarp is ovoid, slightly asymmetrical and with a round apex. Although the application of morphological parameters enabled the distinction between the different 'Malti' cultivar/s, the observed morphotypes can be the result of a spectrum of a single cultivar which is the result of intercrossing between the wild and domesticated *O. europaea*. Furthermore a number of different studies have shown that different cultivars are able to display different morphologies based on geographical locations and under different growth management practices (Grati *et al.,* 2002, 2009; Youssefi *et al*., 2011). In order to determine whether the 'Malti' cultivar/s consist of one cultivar which displays multiple morphologies or else is composed of a number of different cultivars which are conveniently grouped under one cultivar, genetic analysis was carried out in this study and metabolic profiling proposed as a supplementary study that needs to be conducted.

3.1.2 Primer identification

Genomic DNA from the leaves of different cultivars found within the Maltese islands was extracted and subjected to purity testing. Samples were then analysed using a battery of SSR primers in order to identify primers which were able to amplify with the majority of the cultivars and gave the most consistent. The genomic DNA was amplified in the presence of SSR markers and visualised on a 2.5-3% agarose gel. Figures 3.2-3.4 display the most promising SSR primers which were identified. The majority of the SSR primers that were tested were able to amplify easily with the 'Bidni', 'Bajda' and the rest of the Italian cultivars, resulting in sharp, easily distinguishable amplicons. However very few of the primers which were used were able to produce sharp distinct amplicons with both the 'Malti' cultivars and the wild type *Olea europaea* var. *sylvestris* under the same PCR conditions. In order to fully distinguish between all the cultivars which were taken in consideration during this

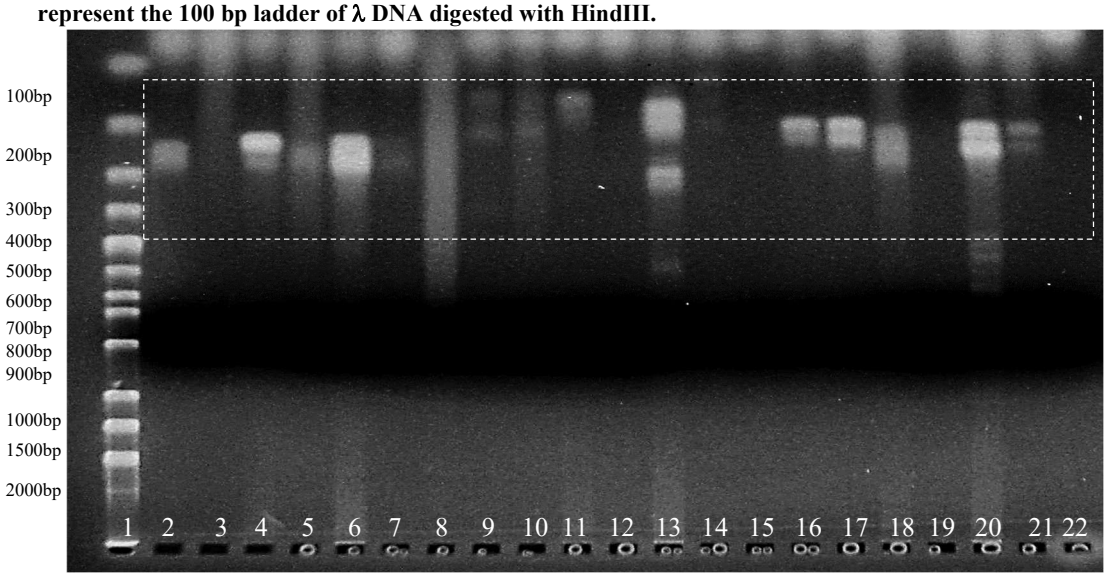
study only two primers, DCA-3 and GAPU-101, provided sufficient results which enabled the inclusion of the 'Malti' cultivar within the study. The recalcitrance of the majority of the 'Malti' cultivar and the wild type to amplify with the tested primers can be due to two main reasons. The first reason may be that, the 'Malti' cultivar had a higher degree of inhibitors present within the extracted genomic DNA as shown by the lower 260/230 ratio, indicating the presence of residual sugars and phenolic compounds. Phenolic compounds and residual polysaccharides are known to inhibit the activity of DNA polymerases and provide irregular PCR amplifications (Testolin and Lain, 2005). Although CTAB and PVPP, which are the most commonly used reagents to remove these contaminants, were used, negligible changes to the purity were seen with the 'Malti' cultivars.

The second possible reason for the reduced amplification may be due to the presence of a different base pair sequence found within the flanking region where the primer attaches. This would prevent the primer from annealing with the genomic DNA, resulting in either inconsistent amplification causing the formation of smears or else no amplification (Figure 3.2). Smears were produced from an erratic, inconsistent binding of the primer to the DNA during the PCR reaction. If this is the case, one possible explanation may be that the 'Malti' and the wild olive trees share a common ancestor, suggesting the 'Malti' cultivar is the product of the domestication of the wild oleaster community found in the Maltese islands. Terral (1997) showed that cultivation and domestication of the oleasters broadly increases drupe size and pruning of oleasters triggers the production of larger fruits whilst the ratio length/diameter of the fruit is retained. This suggested that cultivation causes an increase in the size of the fruit. Considering the five previously identified 'Malti' morphotypes, the increase in the fruit size might the result of the pruning and domestication of one or more wild olive trees present within Maltese islands in ancient times. The domestication technique enabled the formation of a whole spectrum of drupe sizes for the same cultivar which is linked with the wild olive trees. In the absence of an amplification based method that can distinguish between the different morphotypes, genome skimming technology needs to be carried out in the future as it is the most cost-effective technique, and is not dependent on the sequence homology which is currently present.



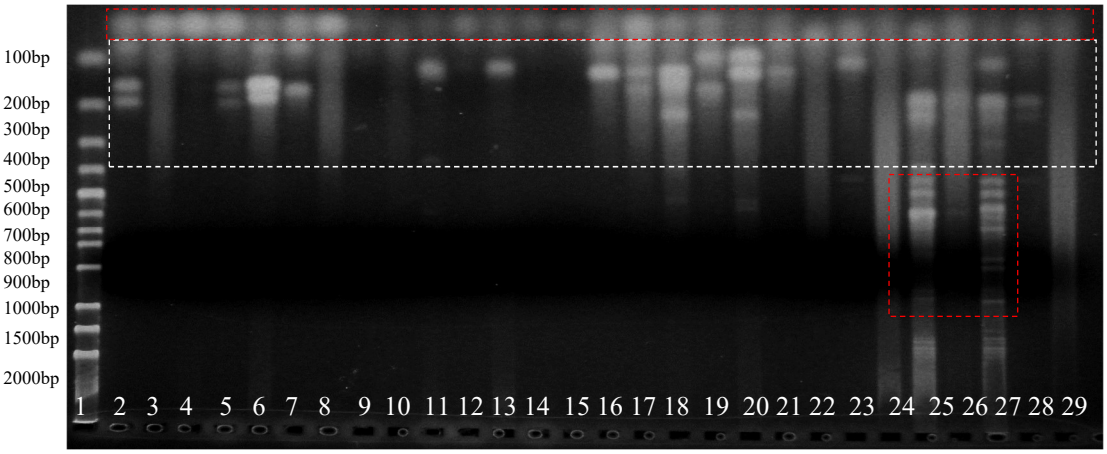
Cultivar	Well number			
	GAPU 101	DCA-4	DCA-9	GAPU 103
'Bajda'	2	9	16	23
Pendolino	3	10	17	24
'Bidni'	4	11	18	25
Frantoio	5	12	19	26
Carolea	6	13	20	27
'Malti'	7	14	21	28
-ve Control	8	15	22	29

Figure 3.2: Gel electrophoresis using 2.5% agarose, showing SSR amplification for seven commonly found cultivars in the Maltese islands for 4 SSR primers, GAPU 101, DCA-4, DCA-9 and GAPU 103. Lane 1 represent the 100 bp ladder of λ DNA digested with HindIII.



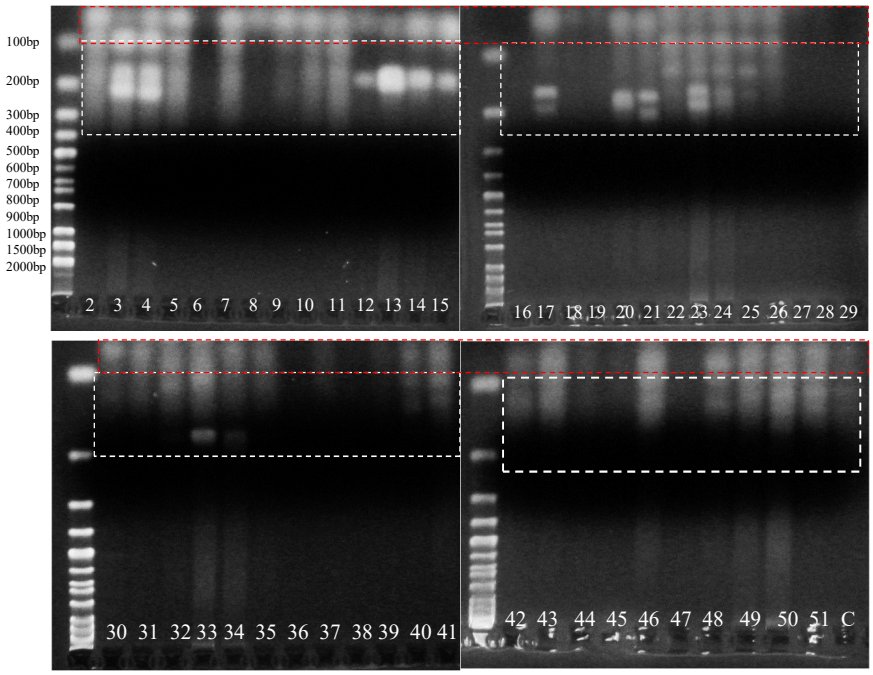
Cultivar	Well number		
	DCA-3	DCA-16	GAPU 59
'Bajda'	2	9	16
Pendolino	3	10	17
'Bidni'	4	11	18
Frantoio	5	12	19
Carolea	6	13	20
'Malti'	7	14	21
-ve Control	8	15	22

Figure3.3: Gel electrophoresis using 2.5% agarose, showing SSR amplicons obtained using three SSR primers DCA-3, DCA-16 and GAPU 59. Lane 1 represent the 100bp ladder of λ DNA digested with HindIII.



Cultivar	Well number			
	UDO-99-12	EMO-30	UDO-99-19	UDO-99-39
'Bajda'	2	9	16	23
Pendolino	3	10	17	24
'Bidni'	4	11	18	25
Frantoio	5	12	19	26
Carolea	6	13	20	27
'Malti'	7	14	21	28
-ve Control	8	15	22	29

Figure 3.4: Gel electrophoresis using 2.5% agarose, showing SSR amplicons obtained using four SSR primers UDO-99-12, EMO-30, UDO-99-19 and UDO-99-39. Lane 1 represent the 100bp ladder of λ DNA digested with HindIII.



Cultivar		Well number			
	DCA-4	DCA-9	UDO-99-24	UDO-99-39	UDO-99-12
'Malti' BZ	2	12	22	32	42
'Malti' BZ2	3	13	23	33	43
'Malti' RBT	4	14	24	34	44
'Malti' Fawwara	5	15	25	35	45
'Malti' Kornja	6	16	26	36	46
'Malti' Wardija	7	17	27	37	47
'Malti' Wardija 1	8	18	28	38	48
'Malti' Wardija 2	9	19	29	39	49
'Malti' Wardija 3	10	20	30	40	50
'Malti' Wardija 4	11	21	31	41	51

Figure 3.5: Gel electrophoresis using 3.5% agarose, showing SSR amplicons obtained using four SSR primers DCA-4, DCA-9, UDO-99-24, UDO-99-39 and UDO-99-12 on different 'Malti' cultivars. Lane 1 represent the 100bp ladder of λ DNA digested with HindIII.

3.1.3 Sizing

The set of 16 SSR markers employed were chosen based on literature data (Sefc *et al.,* 2000; Cipriani *et al.,* 2002; Carriero *et al.,* 2002; De la Rosa *et al.,* 2002)) that allowed amplification of a single locus. Nonetheless, after the initial screening of the different primers, sizing of the amplicons was carried out using Gel Analyser. 15 out of the 31 the samples and two primers were chosen and sized using a more accurate capillary electrophoresis method performed using the Applied Biosystems 3130 Genetic Analyzer at MLS BIODNA. Cultivars showing only one amplified allele per primer pairs were conventionally considered to be homozygous at that locus. One of the main limitations of homozygous cultivars is the underestimation in the genetic diversity, especially if null alleles occur. In a single case, the 'Bajda' cultivar showed three amplified alleles for the GAPU 101 locus; these multi-locus patterns were not considered during the calculation of the observed heterozygosity and expected heterozygosity, as segregation data was not yet available. Similar to what was observed by Rallo *et al*., (2000) and other researchers who worked with olive SSR markers (Pasqualone *et al*., 2004, 2007) a high frequency of multiple unspecific products were found during SSRs amplification. These unspecific products are probably the result of multiple priming sites along the genome, together with primer dimmer amplification which results in small > 100bp fragments observed in all the presented gels (red box in Figure 3.2-3.4). Apart from small fragments, larger fragments (500-700bp) were observed during the amplification with UDO 99-39 and these were attributed to the amplification of two different loci as suggested by Cipriani *et al*. (2002) as a result of genome duplication. According to Poljuha *et al*., (2008) this phenomenon is relatively common in species having a multiple genome origin, although this was not clearly demonstrated in olive and may be due to genome fusion and chromosome duplication events during evolution (Alba *et al*., 2009).

The two SSR primer pairs produced polymorphic and reproducible amplification fragments on all the 15 cultivars, allowing discriminate amongst them. A total of 19 alleles, with an average of 9.5 alleles per primer pair was scored; 7 alleles were determined using DCA-3 whilst GAPU 101 revealed the presence of 12 alleles per locus (Table 3.1). The most common allele (frequency 0.37), which was found at the DCA-3 locus, had a size of 231 bp whilst the least common alleles were 236 bp and 244 bp alleles, which belonged to Pendolino and one 'Malti' sample from Rabat

sample respectively. For the GAPU 101 locus the most common allele (frequency 0.20) was found at 212 bp, whilst a number of unique alleles were found at this locus namely the 218 bp and the 226 bp alleles which both belong to the 'Bajda'cultivar, which is thought to be indigenous to the Maltese islands and the 194bp allele which was found in one 'Malti' sample from Wardija.

Table 3.1 Table showing the size of different alleles and there corresponding frequency, determined at two loci, DCA-3 and GAPU 101

Allele Size bp	Frequency	Allele Frequency
Locus DCA-3		
242	6	0.20
231	11	0.37
236	1	0.03
244	1	0.03
252	4	0.13
248	5	0.17
239	2	0.07
Locus GAPU 101		
212	6	0.20
200	2	0.07
206	2	0.07
267	1	0.03
208	4	0.13
218	1	0.03
222	2	0.07
226	1	0.03
185	4	0.13
190	2	0.07
192	4	0.13
194	1	0.03

Table 3.2 Table showing the number of different alleles, observed and expected heterozygosity and the fixation for the different loci which were studied.

DCA-3		GAPU-101
N_a	7	12
H_o	0.60	1.00
H_e	0.77	0.88
RP	3.06	3.20
PIC	0.73	0.89
FI	0.22	-0.13

Table 3.2 illustrates the different genetic parameters which were determined for the two different loci that were selected. The measure of heterozygosity enables the determination of the genetic diversity within a population. In general, in this study, high values of observed (H_o) and expected heterozygosity (H_e) were recorded for the two SSR markers investigated, with average values consistent to those reported in literature (Ganino *et al*., 2007; Poljuha *et al.,* 2008). For the DCA-3 locus, H_o was lower than H_e whilst for the GAPU 101, H_o was higher than H_e as the cultivars were found to be heterozygotes with respect to the GAPU 101 locus. Thus GAPU 101 revealed a higher genetic variability among the cultivars screened. The high % of heterozygous varieties and the higher values of the H_o compared to H_e were previously described by Diaz *et al.,* (2006). According to these authors the SSRs in the medium to high levels of heterozygosity in olives are the results of out-crossing of species which are clonally propagated. It was observed that for the DCA-3 locus all the 'Bidni' samples which were tested were revealed to be homozygous with respect to the allele found at 231 bp, which was also common to Frantoio, one of the Italian cultivars which was tested, whilst the majority of the 'Malti' cultivars were found to be heterozygous for the DCA-3 locus.

The presence of a high degree of heterozygosity found within the 'Malti' variety supports the hypothesis that this cultivar is derived from the domestication of multiple wild oleaster trees found within the Maltese islands. Mazzitelli *et al*., 2015 postulated that the 'Malti' cultivar was remnant of olive trees that were propagated and domesticated by individual farmers for the production of oil and table use. Furthermore since the majority of the trees are more than 100 years old it is unlikely that genetic exchange with newly imported varieties has occurred, thus retaining the high degree of heterozygosity which reflects their oleaster origins but at the same maintain at the degree of genetic distance from the Italian varieties imported in the past to meet the demand for olive oil production. Contrary to the 'Malti' cultivar, the 'Bidni' cultivar, which is also thought to be indigenous to the Maltese islands, was found to be homozygous with respect to the DCA-3. Furthermore no allelic difference was found for all the samples of 'Bidni' with respect to GAPU 101. This suggests that the original ancient 'Bidni' trees, which are a secluded grove of 26 massive trees, are reminiscent of a larger population which dominated the Maltese islands prior the Arab period. During the Arab period the majority of these ancient trees were cut down for the supply

of wood and to provide space for the cultivation of cotton and citrus. This abrupt change in the population number may have resulted in the formation of a bottleneck. The limited genetic exchange caused the inevitable loss in genetic diversity. Since most olive trees are wind pollinated, this demographic isolation reduced cross-pollination, increasing the rate of inbreeding between the trees and causing an increase in homozygosity. This strengthens the identity of the cultivar but reduces genetic variation putting the population in the risk of extinction. Such a paradox is known as the 'through increased identity by descent'.

Application of the fixation index (FI), also known as the inbreeding coefficient, for all the samples tested enables the extraction of information about selection. A positive value very close to zero was observed indicating that random cross-fertilization was occurring; conversely a negative value close to zero was obtained indicating that the population structure was subjected to an excess of heterozygosity which might be due to the natural selection for heterozygotes. Analysis of the 'Bidni' subpopulation with respect to the DCA-3 showed that all the individuals tested were homozygotes thus the fixation index was equivalent to 1 further confirming inbreeding or undetected null alleles, whilst the 'Malti' subpopulation had a value of -0.08 indicating random cross fertilisation within the subpopulation.

3.1.4 Classification

From the selected and sized SSR markers, a contingency table consisting of presence and absence of the different alleles was constructed. The data was analysed by means of hierarchal cluster analysis using unconstrained Jaccard's genetic distances (Figure 3.6.) The dendrogram was characterised by a wide range of genetic differences among the different genotypes with similarity coefficients ranging from 0.1 to 0.72. The lowest level of similarity (0.1) was observed between the indigenous 'Malti' cultivars and the 'Bidni' cultivars. These results contradict the results which were previously obtained using RAPD analysis whereby it was shown that these so-called 'native' cultivars separated into a single cluster. The largest similarity was found between the individual old 'Bidni' samples derived from a secluded grove of 26 massive trees, presumed to be over 1,500 years old, which were showed to be identical clones of each other. These results suggest that these trees were propagated by means of grafting which preserved their genomic integrity.

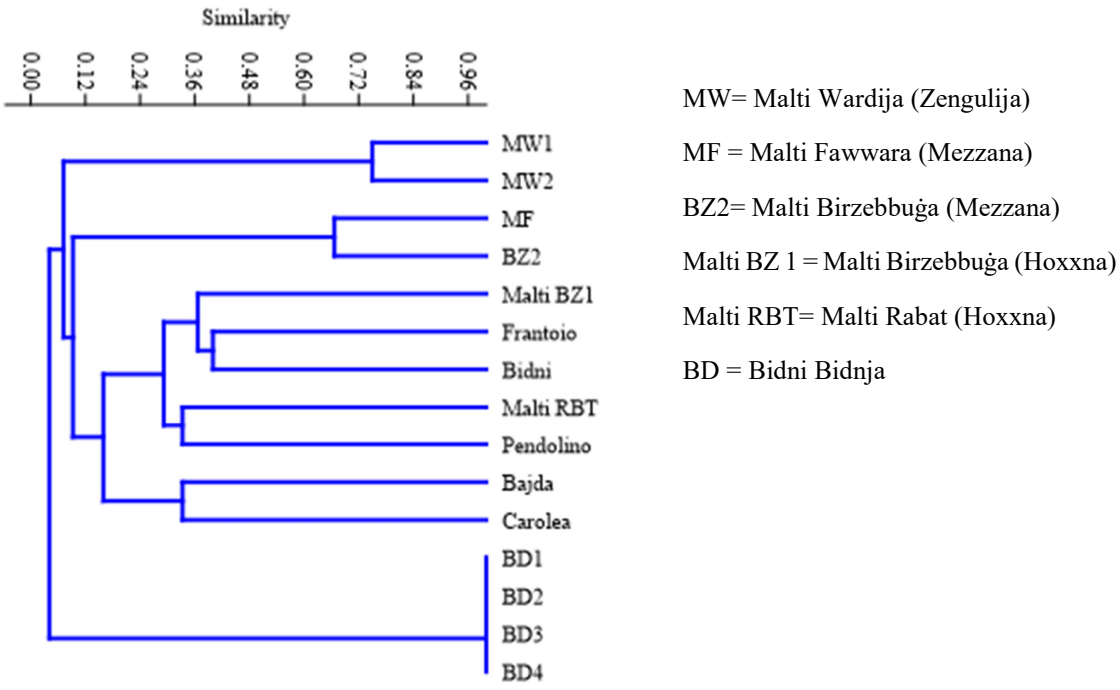


Figure 3.6: Dendrogram based on presence and absence of alleles between the different olive samples, and constructed by hierarchical cluster analysis using Jaccard's unconstrained method.

Similar to results reported by Mazzitelli *et al.* (2015), the greatest similarity was observed between 'Malti' trees derived from the same geographical location. In this study 'Malti' samples derived from Wardija formed a single cluster. There appeared to be a clustering trend according to the drupe size where the 'Malti' from Wardija, known by the common name '*żengulija*', clustered together, as did the 'Malti' from Fawwara and one of the Birżebbuġa samples which belonged to the '*mezzana*' morphotypes, whilst the 'Malti' samples belonging to the '*ħoxxna*' morphotypes clustered with the Italian cultivars. These results may support the hypothesis that the 'Malti' cultivar is in fact a series of clusters of different cultivars which are grouped together under the same common name. Furthermore some of the 'Malti' cultivars, especially those exhibiting an enlarged drupe, might be of a foreign origin.

The 'Bajda' cultivar is an example of a variety exhibiting Leucocarpa. The peculiarity of Leucocarpa is that during ripening stages the drupe, rather than turning purple-black colour due to anthocyanins synthesis, attains a white colour colouration due to the silenced anthocyanin producing genes (Lavee, 1986). Although it is often referred to as a Maltese cultivar, similar to previous results obtained by RAPD analysis, it was shown that this cultivar is highly similar (0.36) to Carolea - a foreign cultivar. This similarity is derived from the shared alleles at the DCA-3 locus with

Carolea 242 bp and 251 bp, whilst differing at the GAPU 101 locus. The latter allele appears to be useful to completely distinguish this cultivar from the rest of the cultivars. SSR analysis conducted by Pasqualone *et al.* (2012) on Leucocarpa cultivars grown in Italy showed the same allele pattern with respect to the DCA-3 locus. These results provide the confirmatory evidence that the 'Bajda' cultivar it not an indigenous species but was imported as an ornamental plant in the past, as originally proposed by Borg in 1922. These results suggest that the 'Bajda' may be an olive cultivar occurring in Southern and central Italy reported as early as at the end of 19th century (Società Botanica Italiana, 1894), known also with the synonyms of Morachia, Cannellina, Bianca, and Chiarita (all meaning "white") (Bartolini, 2008) which is thought to be of Greek origin.

3.1.5 Application of SSR for the classification of the Maltese olive cultivars with respect to other Mediterranean cultivars.

Simple sequence repeats or microsatellites are the most interesting for genotype identification since they are co-dominant, highly informative and reproducible tools. Being highly specific and reproducible, they enabled the comparison of the Maltese cultivars with respect to other Mediterranean cultivars. For this part of the analysis, a small set of cultivars derived from different regions of the Mediterranean were used, for which the allele sizes were obtained from a molecular database (http://www.oleadb.it/) at the two most discriminate loci GAPU 101 and DCA 3. A number of different cultivars were used for comparison: these included Tunisian, Italian (North and South), Greece and Spanish cultivars. The choice of these cultivars was based first of all on the proximity of these cultivars to the Maltese islands (Tunisia and Southern Italy) but also with respect to historical records of past colonies that settled in the Maltese islands (Spanish). The aim of this part of the study was to try and link the existing cultivars found in the Maltese islands and their genetic similarity with other cultivars found in the aforementioned Mediterranean countries, in order to provide insights about their past origin.

Application of UPGMA cluster analysis revealed the existence of two major clusters - one which contained the cultivars of Tunisian (Red) and Greek (Yellow) origin and one which contained Spanish, Northern and Southern Italian cultivars (Blue and Green). The similarity of these clusters can be explained in terms of history. The

close similarity of the greek and Tunisian cultivars is expected since several authors have claimed that the olive was introduced to North West Africa and possibly in Libya and Tunisia by the Phoenicians (14^{th}-12^{th} century BC) during the foundation of Carthage in the 9^{th} century BC. However, it does not exclude the introduction of other cultivars from Greece during the 8^{th}-7^{th} century BC during the foundation of Cyrenaica, which constitutes of modern-day Libya and Tunisia. In the case of the 'Bidni' cultivar, it was found to be very similar to the Tunisian cultivars Chemlali de Sfax, Chemlali de Zarzis and Zalmati, which are the main oil varieties of the Zarzis region which is found in the 200 km away from the Maltese islands. Inspection of the typical morphological characteristics revealed that they are very similar to the 'Bidni' cultivar. These cultivars are produced small round drupes with an average weight of 1.1 g, which are typically very rich in oil with a low pulp/seed ratio Grati Kamoun *et al*. (2001).

In the case of some of the 'Malti' cultivars, including those found in the Wardija, Zebbug, and Lija, these were found to be very similar to the southern Greek cultivars, namely Koroneiki, Mastoiditis, and Kerkiras. Although morphologically they are very different, they still show some degree of similarity to the ones found in the Maltese islands. The 'Malti' from Zebbug tends to be morphologically similar to 'Koroneiki' having a very small weight (0.5 g), with a mastoid shape and ending in a teat with a leaf length of 4.5–5.2 cm and the ratio of length: width of 4.2–5.5:1. Comparison of morphological studies carried out in another study (Aquilina 2017) showed that the length to width ration of the adult leaves of this particular cultivar was 5.35:1, suggesting a close similarity to the quoted values for 'Koroneiki'. In the case of the 'Malti' cultivar from Lija and Wardija, these tended to have larger drupes (1.4–2.2 g) with a cylindroconical (slightly elongated) shape and a prominent teat very similar to 'Mastoidis' and 'Kerkiras'. Leaf morphological characteristics were also similar in the case of 'Mastoidis', with a quoted value of 6–7 cm in length and 1.1 cm width, whilst for 'Kerkiras' a leaf length to width ration of 5:1 is quoted. From our studies it was found that the leaves of the 'Malti' cultivar derived from Waridja had a leaf length of 7.88±0.62 and a width of 1.29±0.32, very similar to 'Mastoidis'. In the case of 'Malti' cultivars obtained from Lija, a leaf length/width ratio of 4.85:1 was obtained which is very similar to that of 'Kerkiras'.

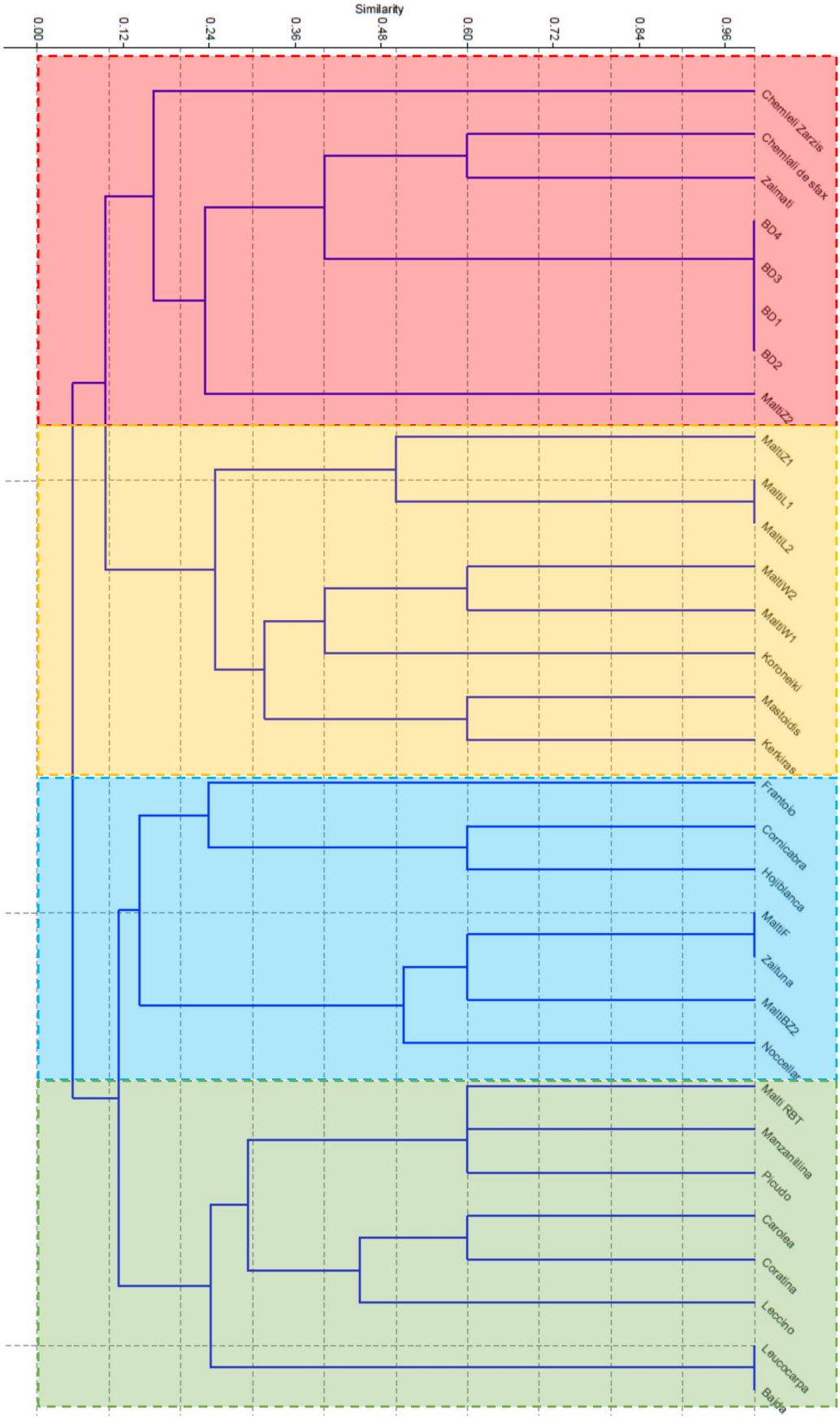


Figure 3.7: Hierarchal cluster analysis using Jaccards method based on the presence, absence data for alleles found in the DCA-3 and GAPU 101 loci, for a selected number of cultivars typically grown in the Mediterranean area.

The second major cluster contained the Spanish cultivars Picudo, Manzanilla, Hojiblanca and Cornicabra, and the Southern Italian cultivars namely Zaituna and Nocellara del Belice together with the rest of the Italian cultivars. The close similarity of the Italian cultivars with respect to the Greek cultivars can be attributed to the spread of Greek colonies, which replaced the 'paleo-Italic' Phoenician colonies. Furthermore it is well documented that the Etruscans, in 7th – 6th century BC, imported olive oil from Greece to the Italian peninsula (Heurgon, 1961). By the 2nd century BC olive growing progressed throughout the Italian peninsula in the regions of Calabria, Basilicata, Campania and Puglia, collectively known as the Magna Greca (Arambarri 1992). With regards to the Spanish cultivars' close similarity to the Italian cultivars it is difficult to evaluate the evolution of these cultivars due to the historical records. Nonetheless, Acerbo (1937) claimed that it was the Phenicians which brought olive trees to the Iberian peninsula, while others tend to believe that it was the Romans (Patac *et al*., 1954; Arambarri 1992) and Arabs (Caruso 1883; De Candolle 1883) which introduced the olive cultivars. From the analysis of the cluster analysis it seems that there is no definitive answer for the introduction of Spanish cultivars. Whilst Hojiblanca and Cornicabra showed very similar allelic patterns with Greek and Southern Italian cultivars suggesting a more likely introduction by Phoenician colonies, Manzanilla and Picudo tended to show allelic patterns similar to Northern Italian cultivars suggesting a more likely Roman introduction.

The 'Malti' cultivars, namely the ones obtained from Fawwara and Birzebbugga, tended to cluster with the Spanish and Southern Italian cultivars. The 'Malti' cultivar obtained from a century-old tree in Fawwara showed an exact allelic pattern to that displayed by the indigenous Sicily Zaituna. As the name implies it is Arabic for Zejtun, meaning olive tree, and suggesting most probably an Arabic influence on this particular cultivar. Morphological analysis of the 'Malti' tree from the Fawwara cultivar revealed a very high similarity to that described in the literature with an elliptic-lanceolate leaf having blade length of 5.5 ± 0.5 and width of 1.3±2 cm, with an ovoid slightly asymmetric fruit without any prominent nipple. The results obtained tend to suggest that the 'Malti' tree from Birzebbugga and Fawwara are in fact cultivars of Southern Italian origin which easy adapted to the Maltese islands, favoured by similar pedo-climatic

conditions. The ‘Malti’ cultivar obtained from another century-old tree in Rabat showed very similar allelic patterns to two Spanish cultivars Manzanilla and Picudo, suggesting a more likely relatively recent Spanish importation during the Spanish colony establishment in the Maltese islands. Presence of Sicilian and Spanish cultivars in the Maltese islands is expected since Malta was part of the Kingdom of Sicily for nearly 440 years. It is during this period that Malta was sold and resold to various feudal lords and barons including the Crown of Castile . Eventually Malta became part of the Spanish empire in 1479.

Similar to what was observed before, the Maltese cultivar ‘Bajda’ was found to be the most dissimilar cultivar from the ‘Bidni’ cultivar as it clustered along the major Italian cultivars namely, Carolea, Coratina and Leccino. Whilst Carolea and Coratina are highly diffused along the Italian peninsula especially in the regions along the lower Adriatic Sea, namely Puglia and Calabria, Leccino is mainly diffused along the Northern regions especially in the regions of Toscana. Nonetheless these three cultivars showed a high similarity at the two loci which were studied, confirming their Italian origin. With regards to the actual ‘Bajda’ cultivar this was found to be very similar with another Leucocarpa cultivar grown in southern Italy. These findings further indicate that this particular cultivar and some of the “‘Malti’” cultivars might not be actually indigenous to Malta, due to the high allelic similarity which they share with both Italian and Spanish cultivars. These findings suggest that these cultivars were most probably introduced within the Maltese islands through the course of history.

In conclusion, with respect to genetic analysis the results obtained are still not definitive as there is a potential application for the analysis of alleles at a number of different loci which were identified but not pursued in this study. However the application of DCA-3 and GAPU 101 provided a preliminary overview of the different cultivars present within the Maltese islands, which respected the historical events which occurred within the Maltese islands. In the case of the 'Bidni' cultivar this was found be highly homozygous and share a very similar allelic pattern with other Tunisian cultivars suggesting a more likely introduction through the establishment of Phoenician and Greek colonies. In the case of the 'Malti' cultivar the results indicate that this is not a single cultivar but a number of different cultivars, some of which are coined under the same name. Some of the 'Malti' cultivars tend to have an allelic pattern very similar to Greek and Tunisian cultivars whilst others share allelic similarity to Southern Italian and Spanish cultivars, the latter suggesting a rather more recent introduction than the former. In the case of the 'Bajda' the similar allelic pattern found with another Leucocarpa cultivar found in Southern Italy suggests that this cultivar is most likely not indigenous to the Maltese islands but rather a case of importation during 1130 to 1816, during which period Malta was part of Italy under the rule of Roger II. Given that during this period Northern Tunisia was also a part of the Kingdom of Sicily, it is also legitimate to assume that the 'Bidni' and some of the 'Malti' cultivars could also be a case of importation however such statement would only hold true if the age of the 'Bidni' trees, claimed to be over 2,000 years old, is similar to the 'Malti' cultivars which share allelic similarities to Greek cultivars, and this still needs to be investigated.

4. Chemical Analysis

4.1.1 Assessment of quality

A number of different quality parameters were determined for monocultivar EVOOs obtained from the Maltese islands and compared to the limits set by European legislation of the European Economic Community (EEC). For all the checked parameters it was shown that the Maltese EVOOs could perfectly classify as EVOOs due to their very low oxidation parameters

Table 4.1: Summary of the quality parameters tested for different olive oils derived from the Maltese islands. Mean values are represented in the table below accompanied by ±1SD n =3.

	FFA	Peroxide ($mEpO_2/Kg$)	*p*-Anisidine	I_2	K_{273}	K_{270}	Δk
'Bidni' 2013	0.42±0.01	10.28±1.23	3.32±0.24	81.01±0.23	1.128±0.198	0.114±0.015	0.001
'Bidni' 2014	0.35±0.04	11.54±1.02	4.54±0.15	80.45±0.12	1.258±0.187	0.089±0.025	0.004
'Bidni' 2015	0.28±0.12	8.94±2.12	3.12±0.22	81.20±0.18	1.894±0.012	0.087±0.025	0.005
'Malti' 2012	0.32±0.05	11.24±0.98	6.15±0.47	82.12±1.11	1.125±0.065	0.098±0.014	0.005
'Malti' 2013	0.48±0.15	10.28±1.23	6.02±0.28	81.56±0.28	1.784±0.087	0.095±0.012	0.002
'Malti' 2014	0.56±0.09	8.94±0.97	3.45±0.47	80.47±0.23	1.894±0.102	0.102±0.026	0.001
'Malti' 2015	0.21±0.04	6.64±0.47	4.52±0.45	81.23±0.58	1.238±0.119	0.115±0.014	0.000
Carolea	0.48±0.11	10.29±1.32	6.10±0.23	82.09±0.56	1.287±0.014	0.084±0.018	0.002
Pendolino	0.45±0.10	11.25±1.47	5.89±0.18	81.23±0.78	1.964±0.108	0.114±0.014	0.001
Panadina	0.68±0.08	12.08±1.62	3.33±0.18	82.03±1.12	1.965±0.141	0.089±0.025	0.002
Frantoio	0.49±0.04	10.56±1.29	5.89±0.26	80.96±0.98	1.994±0.064	0.097±0.012	0.008
'Bajda'	0.89±0.14	17.01±0.89	6.67±0.27	81.25±1.25	2.011±0.051	0.198±0.024	0.009

One of the most important oxidation parameters is the acid, value of which gives an indication of the amount of free fatty acids (FFA) present in the oil. FFAs are formed during both oxidation and thermal degradation of unsaturated fatty acids. These processes cause the hydrolysis and pyrolysis of fatty acids, which result in cleavage of triglycerides causing the release of FFA. Thus, measuring the amount of FFA allows rapid assessment of the degree of degradation of the oil. FFAs accelerate the rate of oil oxidation and thus low levels of FFAs are crucial to prevent the oxidation of fats. The pro-oxidant effect of FFAs is derived from their free carboxylic group,

which speeds up the decomposition rate of hydroperoxides, shifting the equilibrium of the reaction towards the formation of new hydroperoxides (Frega *et al.,* 1999). According to the IOC, extra virgin olive oil is defined as a virgin olive oil which has a free acidity, expressed as oleic acid, of not more than 0.8 gram per 100 gram. From the results obtained for the majority of the monocultivars tested a FFA< 0.8 was obtained with the exception of the 'Bajda' EVOO. This was attributed to the extensive damage to the drupe caused by the *Bactrocera oleae* during the time of study.

The peroxide value PV was determined as an indicator of the initial stages of oxidative change, since in these stages there is formation of hydroperoxides measured by PV. In general, a low PV indicates an oil of a good quality according to the EEC, since for an olive oil to be classified as virgin it must have a PV < 20 mEq O_2/Kg. Although a low PV is desirable, it is important to note that PV decreases as secondary oxidation products appear. Thus oils with significant levels of peroxides may still be odourless if secondary oxidation has not begun. If oxidation is more advanced, the PV may be relatively low but the oil would have become rancid. The primary oxidation involves the production of hydroperoxides and is measured using the PV. However, hydroperoxides are unstable and are susceptible to decomposition, causing the formation of a complex mixture of volatile, non-volatile, and polymeric secondary oxidation products. A major form of secondary oxidation products include aldehydes and ketones, but can also include volatile organic acids, and epoxy compounds. These are responsible for the rancid smell of the oil, and they are measured by the p-AnV. Although this parameter is not specified either by the EEC or by the IOC, the lower the p-AnV, the better the quality of the oil, with a p-AnV of 10 being required for the market. The extent of the damage caused by the *B. oleae* on the 'Bajda' EVOO also reflected in the high PV and p-AnV obtained when compared to the other monocultivar EVOOs.

Other parameters used for the assessment of the quality of the extracted oil were K_{232} and K_{270} which determine the amount of conjugated dienes and trienes respectively. The oxidation of polyunsaturated fatty acids is accompanied by an increase in the UV absorption of the products. In fact fatty acids containing methylene-interrupted dienes and trienes show a shift in their double-bond position during oxidation due to isomerization and conjugate formation. The resulting polyunsaturated

oxidised compounds exhibit an intense absorption at 232 nm for dienes, while conjugated trienes absorb at 270 nm. Since the products of oxidation of dienes are usually aldehydes and ketones, these also tend to absorb at wavelengths of 270, 266 and 274 nm. This fact is taken into account when using ΔK. According to the EEC for an olive oil to classify as virgin K_{232} nm and K_{270} nm should be less than or equal to 2.50 and 0.22 respectively, whilst the ΔK should be equal or less than 0.01. Analysis of the K parameters showed that all the monocultivar EVOOs tested could pass as EVOOs, however if one takes in consideration FFA and the other parameters it is clear that the 'Bajda' olive oil borders the EVOO classification.

5. Elemental analysis via X-ray fluorescence (XRF)

Although the overall quality of olive oils is generally defined by both its organoleptic properties and oxidation parameters defined in Section 5.2.1, the inorganic content of olive oils has a very important role in terms of food safety and shelf-life. Trace heavy metals in vegetable oils are known to have an effect on the rate of oil oxidation. Oxidation of olive oil leads to the development of unfavourable odours and taste causing deterioration of olive oils. The major factors that affect the rate of oxidation are the degree of unsaturation, the amount of oxygen, temperature, light and the presence of metals (mainly transition metals such as Fe and Cu) (Meira *et al.,*2011; Sikwese and Duodu, 2007). Benedet and Shibamoto (2008) showed that trace amounts of Fe, Cu, Cr, Pb and Cd contribute oxidative effects to lipid peroxidation (Benedet and Shibamoto, 2008).The presence of these trace metals enhanced the rate of oxidation of edible oils by increasing the generation of free radicals from fatty acids and hydroperoxides.

The presence of metals in vegetable oils can be attributed to two major uptake pathways which can be endogenous or exogenous. Endogenous pathways, are connected with the plant metabolism whereby inorganic constituents are incorporated into the oil through the natural uptake and preconcentration of the element by the plant. Exogenous pathways are attributed to contamination during the production and the collection of olives and seeds during the oil extraction and treatment processes, (by processing actions such as bleaching, hardening, refining and deodorization) (Leonardis, Macciola and Felice, 2000). They may also arise from systems and materials of packaging and storage (for example, from foreign bodies during harvesting or wear metals in the press) (Coco *et al.,* 2003 and Dantas *et al.,* 2003). The determination and the analysis of trace metals offers a challenge mainly because of the hard organic content of the oil matrix. A number of analytical techniques have been employed for the determination of metal in oils relying on both emission and absorption spectrophotometric techniques as well as electroanalytical techniques. (Hendrikse *et al.*, 1991; Coco *et al.,* 2003; Buldinia Zeiner *et al.,* 2005 Benincasa *et al.,* 2007). Traditional methods for elemental analysis include inductively coupled plasma optical emission spectroscopy (ICP-OES), atomic absorption spectrometry (FAAS/GFAAS), and ion chromatography (IC). The application of XRF has been

shown to be very important for the determination of trace, minor and major elements in a large variety of matrices (Jenkins, 1999). XRF has a number of advantages over other techniques; it uses very small amounts requires minimal sample preparation and offers a non-destructive quantification method. On the other hand, the determination of organic content by XRF is still considered a difficult task since the X-ray cross-sections for light elements are very low (Bortoleto *et al.,* 2005).

In the present work, X-ray scattering processes, resulting from the interaction of the sample with X-ray beams, was used for the discrimination of different vegetable oils using multivariate statistical analysis – specifically, principal component analysis (PCA) and stepwise linear canonical discriminate analysis (SLC-DA). Furthermore the application of XRF for the identification of geographical origin of olive oils was also assessed. The hypothesis behind the application of elemental analysis of olives and its direct reflection of the corresponding geographical origin is based on the different mineral content of soils which are later incorporated into the plant tissues and ultimately in the oils.

5.1 Matrix Correlation analysis

Clustered correlation analysis showed the existence of two major elemental positively correlated clusters found in oils. The first cluster consisted of Si, P, S, Al, Mg and Ca which are most likely to be derived from pedological and biological sources. The second correlation cluster consisted of heavy metal elementals including Fe, Cr, Zn, Cu, Ni and Co (Figure 5.1). The significant positive correlation between each element suggests that they tend to originate from a source, in which the concentration of these elements has a fixed composition. However one must exclude that the plant is simultaneously bioaccumulating these elements at different rates, irrespective of the individual concentration in the soil. Many species of plants have been shown to be capable of taking up heavy metals from soils which are essential for plant growth (Fe, Mn, Zn, Cu, Mg, Mo, and Ni) and other metals with unknown biological function (Cd, Cr, Pb, Co, Ag, Se, Hg) have been also found to accumulate (Cho-Ruk *et al.,* 2006). Only very few bioaccumulation experiments have been performed using *Olea europaea*, mainly because research is focused on the use of non-edible crops for the extraction of heavy metals from contaminated soils. A study conducted by Aghabarati, Hosseini and Maralian (2006) demonstrated that olive trees

irrigated with municipal effluent, increased the levels of Zn, Pb, Cr and Ni in soil and plant, but were below the permissible limits. Similar results were observed by Llorent-Martínez *et al.,* 2011, whereby the positive correlation between the concentration of Cr and Fe was attributed to the fact that the production procedure can involve stainless steel mechanisms containing the two elements, however in the study by Llorent-Martínez *et al.,* 2011 no correlation was seen between the concentration of Cr, Fe with other heavy metal elements present in oils such as Mn and Cu.

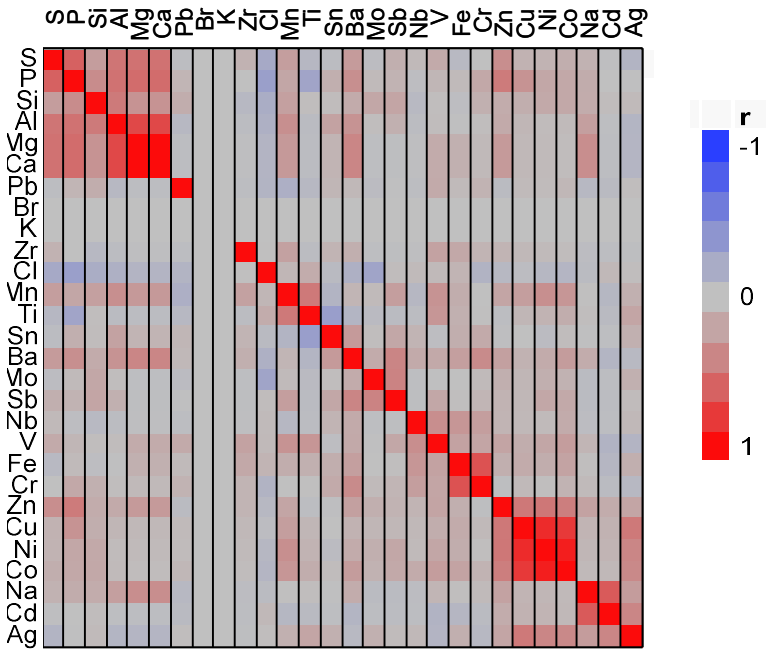


Figure 5.1: Clustered correlation analysis of elemental analysis present in olive oils and other seed oils. Common earth elements; sulphur, phosphorous, silicon, aluminium, magnesium and calcium show a positive correlation with each other. Similarly heavy metals including iron, copper, zinc, chromium, nickel and cobalt tend show a positive correlation with each other

5.2 Determination of geographical origin using elemental analysis.

The determination of geographical origin of extra virgin olive oils in order to assess traceability can be affected by chemical species that are linked to the production area. The hypothesis is that the transfer of elements from the soil to the olive tree and later to the oil is subject to minor variations, and thus the elemental composition of EVOOs is a direct reflection of its geographical origin. The minor variations in the elemental constituents could provide information that can be used for geographical traceability (Benincasa *et al.,* 2007; Zeiner, Juranovic-Cindric and Skevin, 2010). Trace element analysis also plays an important role as a basis for oil adulteration detection and oil quality control (Marfil *et al.,* 2008; Llorent-Martínez, Ortega-

Barrales, Fernándezde Córdova, Domínguez-Vidal, and Ruiz-Medina, 2011). Kruskal Wallis one way ANOVA revealed significant differences in the percentage abundance of the Ba, Cl, Ti, Co, Sn and P across olive oils of different geographical origin and other refined oils used in this study. Figure 5.2 shows the mean percentage abundance of the aforementioned elements for the different olive oils and refined oils. Mann Whitney pairwise comparison showed that % abundance of titanium was significantly higher in refined oils when compared to EVOOs. This was attributed to the extensive treatment and processing to which refined oils are subjected, including bleaching, refining and deodorization. These treatment processes could increase risk of heavy metal incorporation in the oil matrix. In this study no significant difference was observed in the percentage abundance of other heavy metals which are typically used in the seed oils refining processes such as copper and nickel which are typically used as hydrogenation catalysts, (Zeiner *et al.,* 2005) or chromium and iron as potential contaminants of the oil deriving from the processing equipment (Jacobs and Klevay, 1975).

It was also shown that olive oils derived from both indigenous and local cultivars had a significantly higher percentage abundance of barium and phosphorus, this observation suggests that the concentration of these elements can be used as a typical marker for the determination of origin. The presence of a significantly higher barium content in olive oils of Maltese origin can be attributed to pedological effects. Barium is not very mobile in most soil systems, due to the formation of water-insoluble salts and an inability of the barium ion to form soluble complexes with fulvic and humic acids. The presence of organic material, calcium carbonate, chlorides and soil pH affects the availability of barium. In soils with a high organic matter content barium mobility is limited (Bates 1988; Kabata-Pendias and Pendias 1984). High $CaCO_3$ and sulphate salts content also limit mobility, by precipitation of barium carbonate and sulphate (Lagas *et al*. 1984, Bodek *et al*. 1988). However in the presence of chloride ions, barium is more mobile and is more likely to be incorporated into the plants due to the high solubility of barium chloride as compared to other chemical forms of barium (Bates 1988; Lagas *et al*. 1984). The availability of barium in Maltese soils is thus dependent on two opposing effects, the moderately alkaline (pH between 7.3 and 8.5) and the high calcareous content which reduce the mobility of barium whilst the low organic content (MALSIS, 2004) and relatively high concentration of chloride

ions derived from marine origin increase the mobility of barium. Although the presence of calcium carbonate reduces the availability of barium to the plants, its presence fixes the barium content in Maltese soils preventing it from leaching out in solution.

In the study carried out by Camilleri and Vella in 2010 it was shown that the aerial barium concentration in the Maltese islands significantly increases during July and August. This seasonally-dependent emission was attributed to the burning of fireworks during the summer period which also coincides with the olive fruit maturation stage. Camilleri and Vella (2010) also showed that the content of Ba persists even during the latter part of the summer. Barium from fireworks is expected to be particularly bioavailable since it consists of water-soluble species such as $BaCl_2$, BaO, $Ba(OH)_2$ and residual $Ba(NO_3)_2$ (Steinhauser *et al*., 2008).

The significantly higher concentration of phosphorus in olive oils of Maltese origin may be related to a number of different effects mainly the extensive use of fertilizers (Sillanpää and Jansson, 1992) and the high pH of the soil. Similar to barium, the presence of calcium and the moderately alkaline pH across all the Maltese soils prevents phosphorus compounds from leaching. Unlike barium, in the case of phosphorus the incorporation in the plant material is an active process requiring an energy-driven transport mechanism to move P through the membranes into the plant root cells via phosphate transporters. Thus although the phosphorus might be chemically fixed, the plant might still be able to strip phosphorous compounds from the surrounding soil.

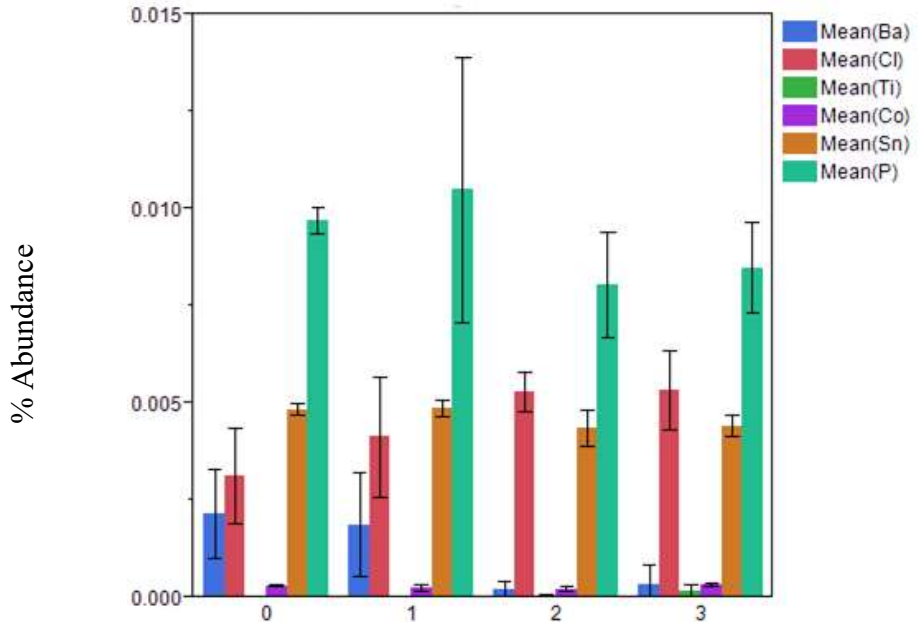


Figure 5.2: % abundance of 6 elements which were found to vary significantly across the four classes of oils. 0= indigenous (n=8), 1= local imported cultivars (n=8), 2= foreign olive oils (n=18) and 3= refined seed oils (n=6).

The application of multivariate analysis showed the spatial distribution for both olive oils derived from different origins and an ability to discriminate between extra virgin olive oils and other refined seed oils. Figure 5.3a shows the results obtained on using stepwise linear canonical discriminate analysis SLC-DA; the model obtained was based on excluding the % abundance K, Ca, Mg and Br whilst retaining all the other elements. The model was able to explain 90.59 % of the observed variance over two canonical functions resulting in an overall 2.17% misclassified observations. The model was able to fully discriminate between olive oils and refined oils, as well as olive oils of different geographical origins. Figure 5.3b shows the major elements which contributed to the observed variation, with Ba, Si, Ag and Cl contributing to the first canonical function and Cu, Zn, Ti and Ni to the second canonical function.

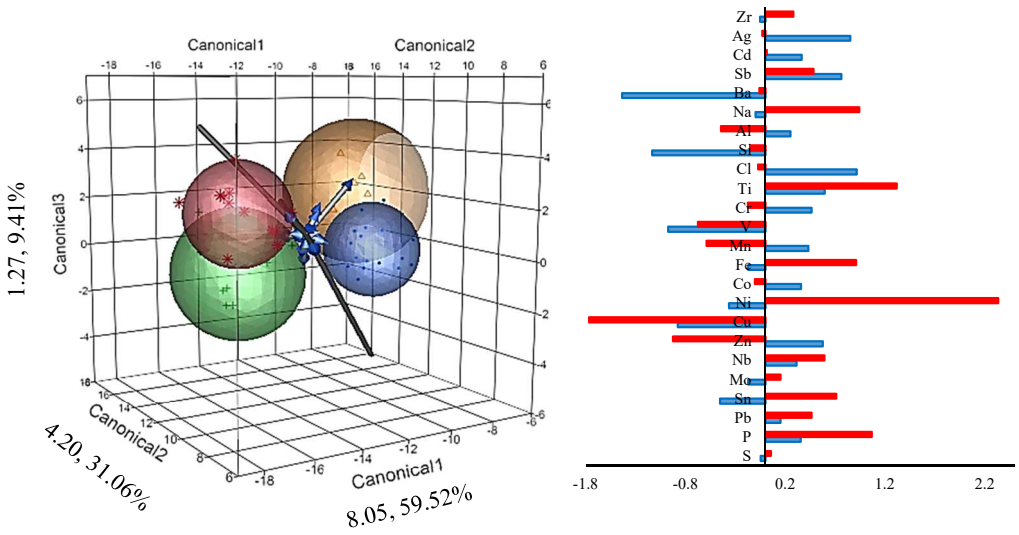


Figure 5.3 3D canonical score plot obtained for the elemental analysis using SLC-DA. Values next to the canonical function represent the eigenvalue and the corresponding % of variation explained. (Right) canonical scores for each element, blue bars correspond to the canonical scores obtained for the 1st canonical function whilst the red bars represent the scores obtained for the 2nd canonical function. Red (*) represents the olive oils derived from indigenous cultivars, Green (+) represents locally derived foreign olive oils, Blue (□) represents foreign olive oils whilst orange (Δ) represents oils derived from other seed oils.

Principal component analysis revealed a similar clustering pattern as that obtained on using SLC-DA. However unlike SLC-DA the % explained variance (28.17% over two principal components) was very low further confirming that the use of supervised techniques is far more informative. Nonetheless on analysis of the principal component scores Figure 5.4 it was observed that heavy metal content namely Co, Ni, Cu and Zn had the highest Eigenvectors in the first principal component indicating that the % abundance of these elements in oils offers the largest explanation of covariance. Similarly Ba and Ag offered the highest % of covariance explained in the 2nd principal component scores.

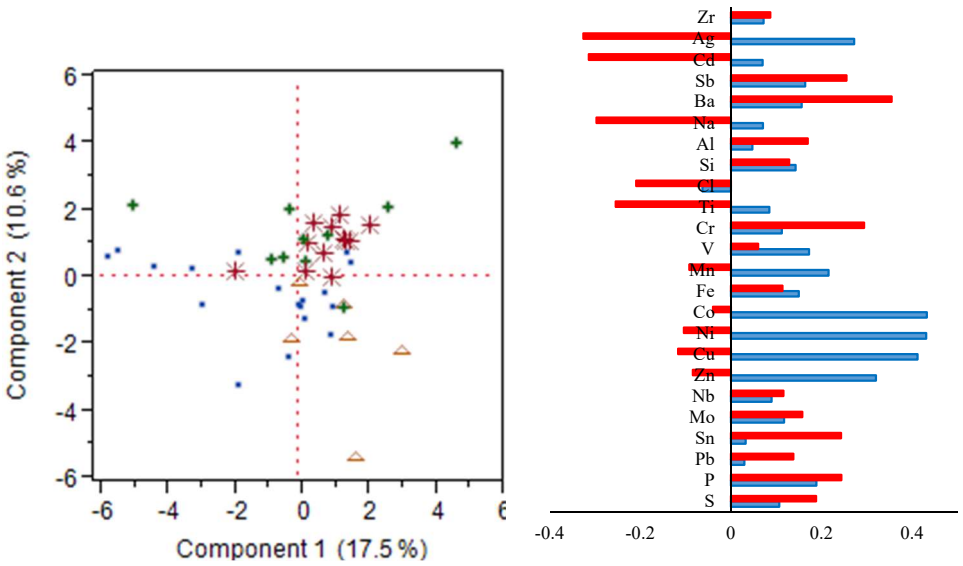


Figure 5.4: Principal component analysis (left) of the elemental analysis of oils derived from different sources and olive oils of different geographical origin. Markers represent the same data presented in the previous figure. (Right) Eigenvector loading scores for each element for the 1st principal component (blue bars) and the 2nd principal component (red bars).

In this study, a multivariate discriminant analysis has been performed to develop a discriminating function which can help predict olive oil origin based on the % abundance of elements. Similar studies conducted by Cabrera-Vique *et al.,* (2012) have shown that elemental analysis of extra virgin olive oil from PDO's of Granada and Jaén had a significantly different concentration in the Ni, Mn, Cu, Cr and Fe content. Furthermore in this study the values of the five trace elements had the highest potential as predictor variables in the model built using multiple linear analysis. The importance of these heavy metal elements in the discrimination of olive oils of different geographical origin is concordant with the results obtained by other authors Benincasa *et al*. (2007) and Zeiner *et al*. (2010). Zeiner *et al*., (2010) reported that the determination of the trace metal pattern of olive oils by atomic spectrometric methods showed that the concentrations of Cu, Mn, and Ni vary significantly by geographical origin of olive oils obtained from Croatia and can be used for geographical characterisation of the oils. Similar to the presented study the concentrations of Fe, Mg, Na, and Ca in the samples, showed no significant differences according to the geographical origin of the oils. Zeiner *et al*., (2010) further reported that higher metal levels in olive oils may be an indication of improper production of extra virgin olive oils and therefore jeopardize the results for a certain geographical region. Benincasa *et al*., (2007) used linear discriminant analysis (LDA) in order to distinguish between olive oils derived from three Italian geographical locations for two different cultivars

Coratina and Carolea. Similar to our study, from the available data it appears that the inter-cultivar variation and olive oils derived from indigenous and imported, locally grown cultivars cannot be distinguished by elemental composition since they are grown under the same geographical conditions. This suggests that the terroir has a large effect on the elemental composition of the olive oil rather than the cultivar itself. Furthermore this also suggests that imported olive cultivars have adapted to the Maltese pedological elemental content.

In this study semi-quantitative techniques were used, whereby the % abundance of each element in the sample was determined in EVOOs. The use of XRF for the determination of the elemental composition of EVOOs proved to be an effective alternative to conventional absorption spectrometry. The additional advantages of XRF lay in the simplicity of the spectra obtained, minimal sample preparation and its non-destructive nature of analysis. In the study a significant positive correlation between elements of pedological origin and heavy metal elements which can be both of botanical origin or derived from anthropological sources, was shown. This study suggest that based on the elemental composition especially the concentration of Ba, P, Co, Ni, Cu and Zn can be employed for the discrimination of olive oils of Maltese origin.

6. Quantification of minor constituents

6.1 Chlorophyll and carotenoid content

The natural colour of virgin olive oil is attributed to the presence of liposoluble chlorophylls and carotenoids present in the fresh fruit. Chlorophyll and carotenoid pigments are highly appreciated as functional components both for their colouring properties, affecting the consumer preference and acceptance together for their health benefits. Carotenoids, impart a yellow colouration to the fruits, vegetables and oils, furthermore they are bioactive compounds, as apart from having a provitamin A function they exhibit strong cellular antioxidant activity, preventing age-related molecular degeneration and cataract formation (Seddon *et al.,* 1994). Similarly, chlorophyll pigments apart from imparting a green colouration, exhibit a series of biological activities in both *in vitro* and *in vivo* animal model assays, ranging from antioxidant to antimutagenic activities, responsible for the prevention of degenerative diseases (Ferruzzi and Blakeslee, 2007).

The analysis of chlorophylls and carotenoid pigments can be divided into two. The first method involves the analysis of pigments within the lipid matrix without the need of prior extraction, the other method is more laborious and targets the profiling of the individual compounds present in the oil by the use of liquid chromatography on a purified pigment fraction. Although the latter method is far more informative, the extraction process invariably causes a considerable loss of pigmentation, particularly within the chlorophyll fraction due to structural transformation caused by the liberation of acids and oxidation (Minguez Mosquera *et al.,* 1990). A number of different methods have been developed in the past in order to assess and quantify the amount of pigments present in olive oil without the need for extraction. Papaseit (1986) implemented the bromothymol blue (BTB) index through the comparison of the oil with visual standards. In this experiment the method employed by Psomiadou and Tsimidou, (2001) was used, where the absorbance of previously diluted sample of olive oil was determined at three wavelengths 670 nm, 630 nm and 710 nm. The carotenoid content was determined by measuring the absorbance at 470 nm as previously determined by Minguez Mosquera *et al.,* (1990).

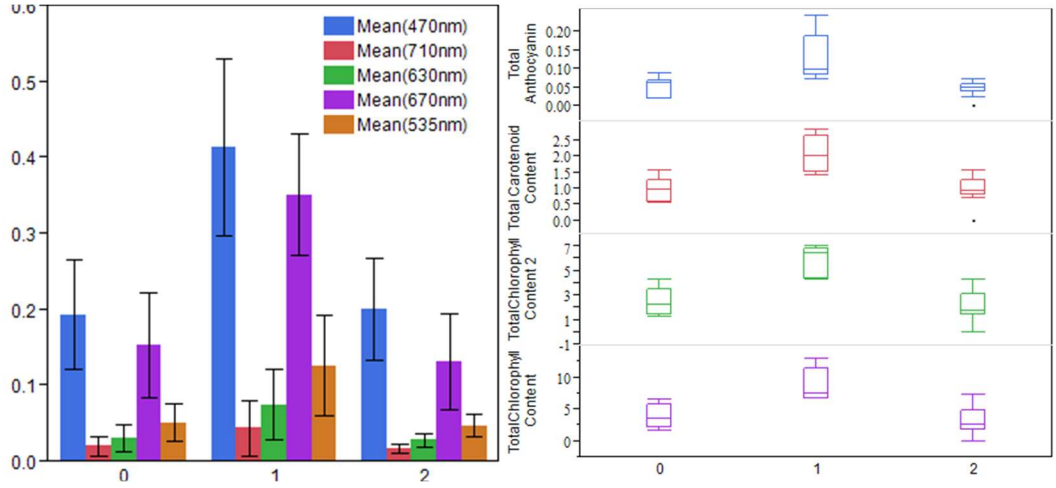


Figure 6.1 : (Left) Bar graph showing the distribution of different pigments present in olive oils of different geographical origin (0 = indigenous cultivars, 1 = locally grown foreign cultivars 2 = foreign cultivars) expressed in terms of the standardised absorption and different wavelengths. Standard error bars represent ± 1SD. (Right) Box plot representing the quantified amounts of different pigments present in olive oils. 1st and 2nd row show the distribution of chlorophyll pigments 3rd and 4th row represent the total carotenoid and anthocyanin pigments present across cultivars of different origins.

Figure 6.1 (left) shows the standardised absorbances at different wavelengths, for olive oils derived from different geographical origins. The quantified chlorophyll, carotenoid and anthocyanin content are shown in Figure 6.1 (right). As expected the presence of anthocyanin pigments was minimal and results were expressed in µg/kg in malvidin equivalents (Francis, 1982). Use of the Kruskal Wallis one way ANOVA showed that olive oils derived from locally grown foreign cultivars had a significantly higher pigment content when compared to olive oils derived from indigenous cultivars (p-value of 0.031 chlorophyll, 0.003 for carotenoid and 0.026 for anthocyanin pigment content) and from olive oils derived from foreign cultivars (p-value of 0.001 chlorophyll, 0.011 for carotenoid and 0.001 for anthocyanin pigment content). However, this might be due to the presence of the Leucocarpa cultivar which was classified as an indigenous cultivar. The absence of chlorophyll pigments within this particular cultivar might have caused an underestimation of the chlorophyll content for either group. In comparison, the pigment content of olive oils derived from indigenous and foreign cultivars did not seem to be significantly different. There are a number of different studies that show quantitatively that the pigment content of olive fruit and virgin oils varies among different olive cultivars (Gandul-Rojas and Minguez-Mosquera, 1996; Psomiadou and Tsimidou, 2001; Roca and Minguez-Mosquera, 2001) and can be used to establish taxonomic affinities and/or differences. Ranalli and Modesti (1998) showed that the effect of genetic and olive variety on olive oil colour is more important than that of the extraction technology used.

The amount of pigments present in olive oils is determined by a number of different factors related to both the amount of pigments present within the fruit, the extraction process and the storage conditions to which the oil was subjected. Degradation during storage might have caused the loss of certain pigments. In the future a forced degradation experiment should be conducted in order to assess the changes of the pigment content. It is well known that these compounds tend to undergo a series of chemical transformation resulting in a discolouration of the olive oil during storage or heat treatment. Carotenoids tend to undergoing trans-cis isomerization and 5,6-epoxide groups to 5,8-furanoxide rearrangement reactions (Mínguez-Mosquera and Jarén-Galán, 1999; Pérez-Gávez, Jarén-Galán, and Mínguez-Mosquera, 2000; Shi and Le Maguer, 2000; Sanchez, Carmona, Ordoudi, Tsimidou, and Alonso, 2008; Zhao, Kim, Pan, and Chung, 2014), while chlorophyll tends to undergo decarbomethoxilation and allomerization of the isocyclic ring (Mínguez-Mosquera, Gandul-Rojas, Gallardo-Guerrero, Roca, and Jarén- Galán, 2007). Aparicio-Ruiz and Gandul-Rojas (2014) showed that in the absence of oxygen degradation of pigments in olive oils undergoes a first-order kinetic mechanism with the carotenoids being more susceptible than chlorophylls to oxidation, which is attributed to the more stable structure of the latter.

6.1.1 Determination of geographical origin and cultivar influence on olive oil pigment content.

The application of pigment content for the determination of geographical origin of olive oils was assessed using SLC-DA, shown in Figure 6.2. It was observed that on the basis of the standardised absorbances a spatial clustering resembling the actual geographical origin was obtained, however the model obtained had a high percentage of misclassification (14.63%), furthermore the Eigenvalues for the 1st and 2nd canonical functions were both less than 1 (0.89 and 0.50 respectively). This suggests that although there is a potential for using the pigment content in order to determine the geographical origin, this part of the study requires further investigation in order to fully optimise the method of classification. Moreover, additional seasonal variation and maturity of the olive fruits at the time of harvest must also be taken in consideration in order to build a more rigorous model.

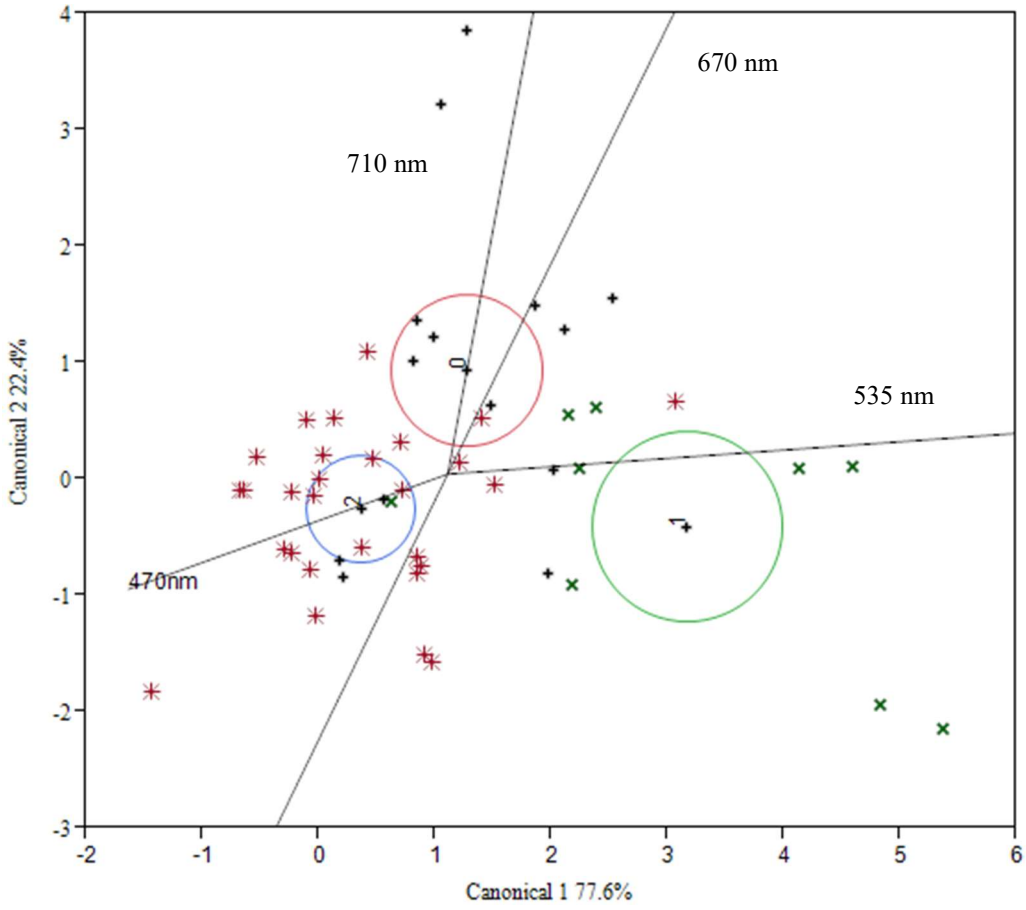
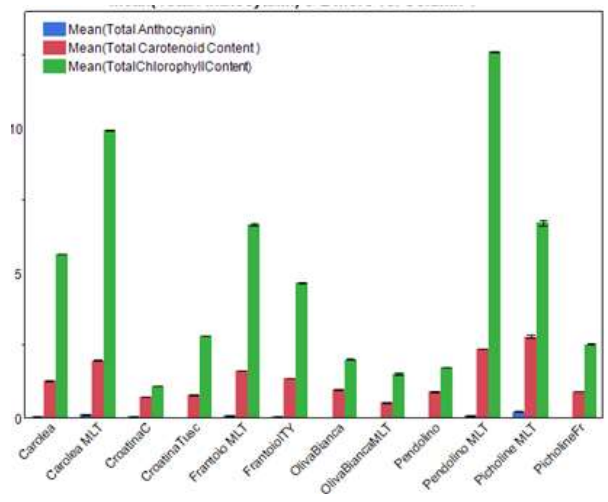


Figure 6.2: CDA biplot show the spatial distribution of olive oil samples derived from different geographical origins (Black (+) represent indigenous cultivars n=12 Green (x) = locally grow cultivars n=8 Red (*) = foreign cultivars n= 24) based on the standardised absorbance of the different pigments.

In order to assess the effect of geographical location on cultivar pigment content, pairwise comparisons between olive oils derived from imported cultivars grown in the Maltese islands and the same cultivar not grown in Malta were carried out using Kruskal Wallis one way ANOVA for all pairwise comparisons. Figure 6.2 summarises the results obtained. It was shown that in general imported cultivars grown in the Maltese islands had no significant difference in the pigment content, for both the chlorophyll and carotenoid content, with the exception of 'Picholine' and 'Pendolino', whereby the olive oils derived from the Maltese islands showed a significantly higher pigment content when compared to olive oils derived from France and Italy respectively. This considerable difference in the pigment content of the two cultivars was attributed to the fact that at the time of harvest the two cultivars had a very low index of maturation. Immature olive fruits have a green colouration attributed to the presence of chlorophyll pigments, as the fruits mature and the photosynthetic activity decreases the chlorophylls disappear. In some olive cultivars the carotenoids, secondary photosynthetic pigments may also disappear at the same time. At the end of

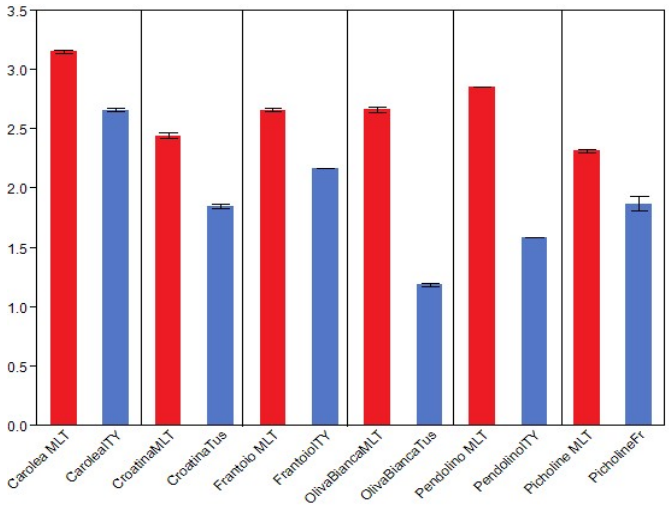
the maturation process, anthocyanin and belatins pigment production takes over giving a violet or purple colour to the olive fruit (Roca and Minguez-Mosquera, 2001). These changes in olive fruit with ripening are directly reflected in the pigment composition of the olive oil as virgin oils. However in certain cultivars the concentration of carotenoid pigments is retained, and furthermore certain cultivars synthesise new carotenoids during maturation. These are known as carotenogenic fruits in which the typical pattern of chloroplast carotenoids does not change during ripening. The qualitative distribution of the chloroplast pigment pattern and the rate of movement or interchange are what distinguish genera and varieties. For example, in olives of Manzanilla, Gordal, and Hojiblanca varieties (Miguez-Mosquera and GarridoFernandez, 1989; Mınguez-Mosquera and Gallardo Guerrero, 1995), the qualitative pattern of chloroplast pigments does not vary with ripening, but in general matches that of the fruits, the final coloration of which is due to the synthesis of compounds of a different nature, such as the anthocyanins, which may even mask the presence of chlorophylls and carotenoids.



Cultivar	Chlorophyll Content	Carotenoid Content	Anthocyanin content
Picholine	0.094*	0.017**	0.001**
Carolea	0.502	0.364	0.364
White Olive	0.403	0.073*	0.473
Pendolino	0.020**	0.017**	0.176
Croatina	0.073*	0.702	0.473
Frantoio	0.550	0.720	0.359

Figure 6.3: (Left) Pairwise comparison of the chlorophyll, carotenoid and anthocyanin content of the different cultivars grown in the Maltese islands and those derived from other Mediterranean countries. (Right) Table illustrating the p-values obtained for each cultivar on using Kruskal Wallis One way ANOVA. (*) indicate a significant difference at the 90% CL whilst (**) indicate a significant difference at the 95% CL.

Although the individual concentrations of chlorophyll and carotenoid pigments did not seem to vary significantly between cultivars which are locally grown and those grown in other countries in the Mediterranean, the chlorophyll: carotenoid ratio was found to be significantly different for the same cultivar grown in different locations. In general it was observed that both imported and indigenous cultivars had a significantly (p-value < 0.001 for both pairwise comparisons) higher chlorophyll: carotenoid ratio (Ch/Cr) when compared to foreign cultivars, however there was no significant (p-value 0.752) difference in the Ch/Cr between the imported locally grown cultivars and the indigenous cultivars. Figure 6.4 shows that the higher Ch/Cr was also consistent on analysing each individual cultivar. This suggests that olive fruits derived from the Maltese islands are subjected to higher light intensities than those grown in other regions of the Mediterranean. This is because unlike chlorophylls, carotenoid compounds are not the primary photosynthetic compounds and their production is upregulated when photosynthesis by chlorophyll is not sufficient.



Cultivar	p value
Carolea	0.125
Croatina	0.089*
Frantoio	0.089*
Oliva Bianca	0.007**
Pendolino	0.002**
Picholine	0.085*

Figure 6.4: (Left) Pairwise comparison of the chlorophyll /carotenoid ratio of the different cultivars grown in the Maltese islands (red bars) and those derived from other Mediterranean countries (Blue). (Right) Table illustrating the p-values obtained for each cultivar on using Kruskal Wallis One way ANOVA. (*) indicate a significant difference at the 90% CL whilst (**) indicate a significant difference at the 95% CL.

The production of chlorophyll and carotenoid pigments is affected by a number of different factors, including light, temperature and water stress. A large number of studies conducted on a number of different higher plants have shown that plants subjected to drought exhibited a decrease in the chlorophyll content and photosynthesis (Kuroda *et al.,* 1990, Sairam, 1994; Kraus *et al.,* 1995; Pastori and Trippi, 1992, Zobayed *et al.,* 2005). Plants grown under drought condition have a lower stomatal conductance in order to conserve water, causing a decrease in CO_2 fixation, resulting in less production and yield of plants. Under drought conditions the decline in the chlorophyll a and b is the result of oxidative damage (Maniannan *et al.,* 2007; Smirnoff, 1995). Apart from water, salinity has also been found to negatively affect the pigment concentration and photosynthesis in higher plants. There is strong evidence that salt affects photosynthetic enzymes, chlorophyll and carotenoids (Stepien and Klobus, 2006).

Light intensity might be also responsible for the variation in pigment content in plants the increased exposure to harmful UV radiation causes degradation in chlorophyll pigment Conversely, lately a number of different studies suggest that exposure to UV-B suppress chlorophyll degrading enzymes such as chlorophyllase and chlorophyll peroxidase and Mg-dechelatase (Aiamla-or *et al.,* 2010; Srilaong *et al.,* 2011; Kaewsuksaeng *et al.,* 2011).

Since the Maltese islands are characterized by very dry summers, drought stress conditions are inevitable, and thus it is expected that olive oils found in the Maltese islands would have a lower chlorophyll and carotenoid content however in this study the contrary was shown. These unexpected results suggest that there are other factors rather than the abiotic stress conditions which are affecting the concentration of these pigments. Furthermore most of the available research present on the effects of stress on pigment content is associated with leaves since they are the major organs responsible for photosynthesis. At the time of study very limited research targeted the effect of stress on pigment concentration in olive fruits. The presence of a thick waxy cuticle, reduced amount of stomatal apertures and their transient existence within the plant suggests that the factors governing the rate of degradation and biosynthesis of these compounds is different from that of the leaves.

Furthermore, presence of antioxidative and secondary metabolites which are produced under levels of abiotic stress conditions could explain the increased or comparable pigment content present in olive oils of Maltese origin. One of such compounds is proline. Research has shown that the proline content increases under drought and thermal stress (Sanchez *et al.,* 1998; Alexieva *et al.,* 2001; Verbruggen and Hermans 2008). Proline does not interfere with normal biochemical reactions but allows the plants to survive under stress (Stewart, 1981). Other compounds which can be produced under stress conditions and enable the plant to survive under drought stress include mannitol, sucrose and raffinose oligosaccharides, and nitrogen-containing compounds, such as amino acids and polyamines (Bohnert *et al.,* 1995). The presence of alternative fluorophores has been shown to present in Maltese olive oils (Section 11.3) and might also contribute to the observed Ch/Cr. However the presence of these compounds as a clear indication of stress has not been determined in olive oils, and needs to be further addressed in future studies.

6.2 Phenolic constituents

6.3 Phenolic constituents

Total phenolic constituents of olive oils derived from different geographical origins were evaluated using microtiter methods. The total phenolic content was determined using Folin-Ciocalteu reagent and gallic acid as standard, an established method for the determination of phenolics described in the literature (Slinkard and Singleton, 1977). The Folin-Ciocalteu reagent is sensitive to reducing compounds such as, but not exclusively, polyphenols, as it also reacts with other reducing substances including sugars, aromatic amines, and organic acids. Thus the presence of these compounds within the hydro-alcoholic extracts might lead to an overestimation of polyphenols. However, this method has the advantage that it is reproducible, quick, inexpensive and particularly helpful for a large number of samples and therefore was adopted as the method of use. The problem with this method is that it can only be performed in the aqueous phase, thus not being applicable for lipophilic compounds/matrices. Thus phenolic compounds present in olive oils required a preliminary extraction process in order to isolate phenolic compound from the lipid matrix. Flavonoid content was determined using the aluminium chloride assay as described by Mabry *et al.,* (1970). The reaction relies on the complexation of flavonoids with an Al^{3+} to give the characteristic red complex, however coordination of flavonoids is dependent on the size, large flavonoids compounds tend to form 1:1 complexes with Al^{3+} ions as higher coordination around the centred aluminium ions are hampered by steric interactions. In the case of ortho diphenolic compounds these were determined using Arnow's reagent as described by Arnow (1937) which relies on the oxidation of ortho-diphenols to a red stable quinone chromophore. The limitation of this assay is the relatively low detection limit as side chains containing a catechol might not be easily oxidisable, resulting in an underestimation of the total amount of ortho diphenolic compounds present in the extracts.

Overall, analysis (Figure 6.5) of the data obtained for different olive oils showed a significant variation in the concentration of phenolic compounds depending on the geographical origin. Use of Kruskal Wallis one way ANOVA showed that olive oils derived from indigenous cultivars had a significantly lower total phenolic compared to olive oils derived from foreign cultivars (p-value of 0.018 TPC and 0.002 for TFC).

In comparison, the total phenolic and flavonoid content of olive oils derived from locally grown imported cultivars and foreign cultivars did not seem to be significantly different (p-value 1.00 for TPC and 0.291 for TFC). The concentration of phenolic compounds present in olive oils is a result of a mix of the multitude of factors that influence both the olive fruit and production process. Cultivar is the most important factor followed by fruit maturity and processing techniques (Kalua *et al*. 2007). The results obtained during this study ranged between 150 – 300 ppm which are in the range to those determined by other studies (García-González *et al.,* 2010; Tura *et al.,* 2007; Uceda *et al.,* 2005; Tous *et al.,* 2005 and Pannelli *et al.,* 2001). These studies showed that olive oils derived from a number of different cultivars produced in various countries around the world had a TPC ranging between 182 to 1,240 ppm. Even though these studies used minor adjustments to the standard method of phenolic compound determination, the general rankings of the varieties by total polyphenol content is consistent. Thus although the results obtained in this study might not be fully replicated by minor adjustment to the methods, the general trend that the indigenous cultivars have a mean lower phenolic content when compared to the mean of other cultivars locally grown and cultivars of different geographical origin would still be consistent.

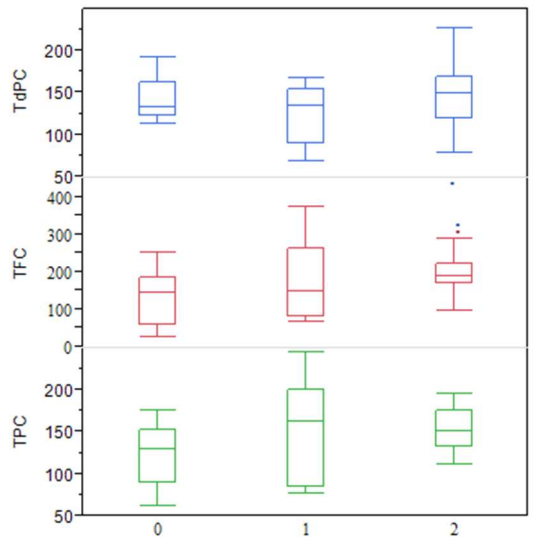


Figure 6.5: Box plot representing the quantified amounts of different phenolic classes present in olive oils. 1st , 2nd and 3rd row show the distribution of the total phenolic content, total flavonoid content and total diphenolic content present in olive oils of different geographical origin (0 = indigenous cultivars n=12, 1 = locally grow cultivars n=8, 2 = foreign cultivars n= 24). Statistical analysis showed that both the TPC and TFC varied significantly form olive oils derived from different geographical origin whilst the o-diphenolic did not show any significant variation across olive oils of different origins (p-value 0.074).

This lower phenolic content can be mainly attributed to the different genotypes that make the Maltese indigenous cultivar unique in terms of their phenolic content. A study conducted by Romani *et al.,*(1999) showed that the phenolic content and the flavour profile of five different olive oils obtained from different Italian cultivars is more dependent on the genotypes rather than on external factors such as climate, irrigation, and harvesting. Analysis of the individual indigenous cultivars found that olive oils derived from the 'Malti' cultivar had the lowest TPC whilst 'Bidni' cultivar had the highest TPC from the indigenous group. Comparison of the TPC, TFC and TdPC obtained from the 'Bidni' EVOOs to the foreign EVOOs showed no significant difference between the two groups (p-value 0.07, 0.102 and 0.407 for TPC, TFC and TdPC respectively). This observation suggests that although classifying EVOOs according to their origin could provide insights for an overall comparison, the huge heterogeneity of the groups, due to the large variation in phenolic content of the cultivars, does not necessarily represent the overall class distribution and individual cultivars need to be compared.

6.3.1 Environmental effects on the total phenolic contents

Although the genotype, might be most determining factor affecting the total phenolic content in olive oils, other factors such as temperature and altitude need to be taken in consideration. It was shown that the warmer the temperature and the higher the altitude at which the olive trees are grown the lower the phenolic content of the oil (Osman *et al.,* 1994). The high temperatures which characterise the Maltese islands during summer might also explain the lower phenolic content found in the indigenous cultivars, however this does not explain why the imported cultivars did not show a phenolic content comparable to the indigenous cultivars, further corroborating that the phenolic content is more determined by the genotype. These observations seem to be in line with the results obtained by Mousa and Gerasopoulos (1996) whereby very little or no differences was observed in olive oils derived from cultivars grown under different temperatures and altitudes. The effect of irrigation and water stress on the phenolic content was determined by Berenguer *et al*. (2006) In their study it was concluded that the olive trees which are cultivated in a high and moderate irrigation regime tend to have a lower phenolic content than those obtained from non-irrigated olive trees. This was attributed to the decrease in both the synthesis of polyphenols and their dilution in within the fruit itself. Similar results were obtained by Gomez-Rico *et al.,* (2007) and Servili *et al.,* (2007). In fact the activity of enzymes responsible for phenolic compound synthesis, such as L-phenylalanine ammonialyase activity is greater under higher water stress conditions (Morello, Romero, Ramo, and Motilva, 2005). Applying this study, to the presented data, it seems that the results obtained are not consistent with the existing literature mainly because most of the olive trees cultivated in the Maltese islands are not irrigated and thus it is expected to have a higher phenolic content than those obtained from other Mediterranean countries. Since the method for determination of the phenolic content is not specific, the possible increase in these stress metabolites cannot be fully accounted for by means of microtiter methods. However, the aforementioned studies just consider the total phenolic content whilst, as will be shown in the latter stages of this study, the complete phenolic profile must be taken in consideration. Since phenolic compounds are secondary metabolites which are produced by the plants in response to herbivory and water stress, there is a possibility that the indigenous cultivars are producing phenolic

compounds which would not react with Folin's reagent and thus are not quantified during this experiment.

The effect of salinity, can also affect the concentration of phenolic compounds in olive oils, although studies conducted by Grattan *et al*. (2006) , Ramos and Santos (2010) and Wiesman *et al*. (2004) showed an increase in the production of antioxidant compounds in olive trees irrigated with saline water. Studies also showed that salinity effectively reduces the phenolic content in plants. Telesiński *et al*. (2008) and Noreen and Ashraf (2009). Telesiński *et al.,* (2008) demonstrated that salinity, apart from causing a decrease in the phenolic content, increases the production of other phenolic compounds such as flavonoids. However the data obtained from the presented study shows that the total flavonoid content of olive oils derived from indigenous cultivars was found to be significantly lower than that of foreign olive oils.

6.4 Antioxidant activity

A number of microtiter assays were carried out in order to determine the antioxidant activity of the olive oil phenolic extracts. These can be broadly be divided into two: the antioxidant capacity assays and radical scavenging activity assays. The chemistry involved in order to distinguish between the two kinds of assays employed in this study can be either reactions which involve hydrogen atom transfer reaction-based assays or single electron transfer reaction-based assays. In this study, the antioxidant activity of phenolic compounds in olive oils was determined by the use of electron transfer -based assays, that include the DPPH, ABTS and NO assays. These involve a redox reaction with the oxidant which is also the probe for monitoring the reaction as an indicator of the reaction endpoint. These antioxidant capacity assays tend to follow the same pattern whereby the probe itself is an oxidant that abstracts an electron from the antioxidant, causing colour changes of the probe, the extent of the colour change is proportional to the antioxidant concentrations. Unlike hydrogen atom based assays, in the assays which were used there is no competitive reaction involved. Furthermore there is no oxygen radical in the assays, and the antioxidant capacity of a sample is taken to be equivalent to its reducing capacity.

Figure 6.6 and Table 6.1 summarise the results obtained from the different antioxidant assays. The results were divided in three classes that reflect the geographical origin of the olive oil. Similar trends observed from the analysis of the

phenolic compound present in olive oil were observed. A Kruskal Wallis one way ANOVA revealed a significant difference in the observed antioxidant activity of phenolic extracts from olive oils derived from different geographical origins.

.

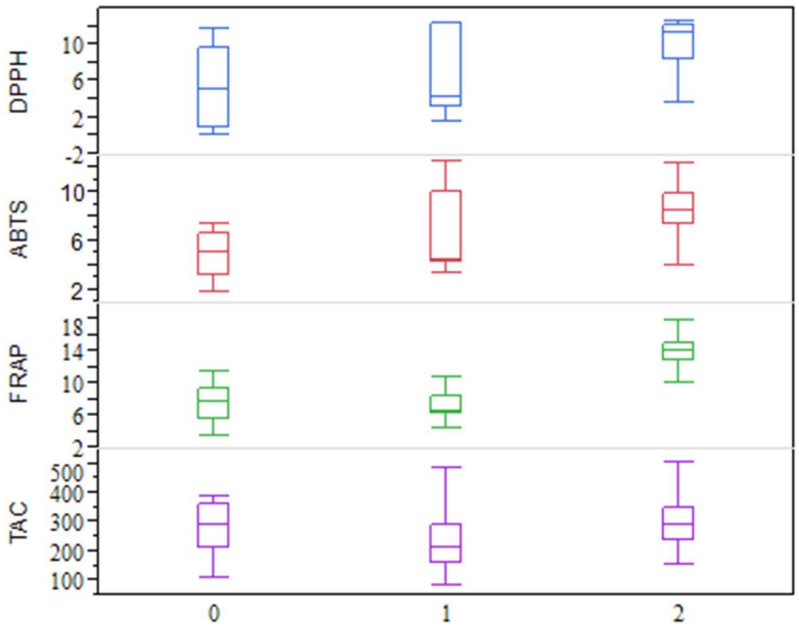


Figure 6.6: Box plot representing the antioxidant activity of the polar phase of olive oils under different assays. 1st row represent the total antioxidant assays carried, 2nd row represent the ferric reducing activity, whilst 3rd and 4th represent the free radical scavenging activity against ABTS and DPPH radicals using phenolic fractions derived from olive oils of different geographical origin (0 = indigenous cultivars, 1 = locally grow cultivars 2 = foreign cultivars).

Table 6.1: The mean antioxidant activity ±1SD on using different assays. Superscript letters in the same column represent statistically distinct homogeneous subsets as determined by Kruskal Wallis ANOVA a 5% confidence level.

	TAC	FRAP	NO	DPPH	ABTS
Indigenous	278.05 ± 94.77^{a,b}	7.19 ± 2.41^{a}	61.01 ± 12.55^{a}	5.39 ± 4.47^{a}	4.85 ± 1.86^{a}
Local	233.21 ± 104.12^{a}	7.02 ± 1.53^{a}	62.62 ± 14.37^{a}	6.72 ± 4.36^{a}	6.61 ± 3.08^{a}
Foreign	307.40 ± 84.17^{b}	13.93 ± 1.55^{b}	67.54 ± 9.30^{a}	10.20 ± 2.46^{b}	8.40 ± 0.92^{b}

Table 6.1 shows pairwise comparisons between each subset; it was found that in general both the indigenous and the locally grown imported cultivars showed a significantly lower antioxidant activity when determined by the different methods. In the case of the total antioxidant activity (TAC) of extracts, this was evaluated by the phosphomolybdenum assay method (Prieto *et al.,* 1999) which is based on the reduction of Mo (VI) to Mo (V) by the sample and the subsequent formation of green phosphate/Mo (V) complex in the acidic environment. This method is quantitative since the total antioxidant activity is expressed as equivalents of ascorbic acid (Prieto

et al., 1999). The ferric reducing activity (FRAP) of the extracts was determined by using the assay described by Yen and Chen adjusted for microtiter plating. The results obtained showed that on comparing the olive oils derived from foreign locally cultivars, these had a significantly lower reducing activity towards Mo (VI) and Fe (III), when compared to olive oils derived from other countries in the Mediterranean.

In the case of Mo (VI) it was observed that whilst the imported olive oil cultivars had a lower reducing activity when compared to foreign cultivars, there was no significant difference in the reducing potential of the indigenous cultivars when compared to the foreign cultivars. This variation in the trends can be explained in terms of the different electrode potentials between the Mo (VI) and Fe (III): the former has an E^o of 0.70 V whilst the latter has an E^o of 0.77V. The slightly higher electrode potential of Fe (III) indicates that the reduction is more easily achieved when compared to Mo (VI), thus phenolic compounds of a lower reducing activity are able to reduce Fe (III) but not Mo (VI). Different classes of phenolic compounds show a structure relationship which affects the antioxidant activity. In the case of phenolic acids the antioxidant activity is affected by the number of hydroxyl groups in their molecules and on the steric effects (Leja *et al.,* 2007). The position of hydroxyl groups, as well as the type of substitution on the aromatic ring, influences the antioxidative activity of these compounds. The antioxidative activity of phenolic acid increases with increased number of hydroxyl groups together with the presence of additional methoxy group in the ring (Cuvelier *et al.,* 1996). Furthermore, cinnamic acid derivatives display better antioxidative properties than benzoic acid derivatives (Siquet *et al.,* 2006). The introduction of an ethylene group between the phenyl ring containing a hydroxy group in the para-position and a carboxyl group, increases reductive properties. In the case of flavonoids, the number and location of the phenolic OH groups affect the reducing potential and the radical scavenging activity, however the presence of a 3', 4'-dihydroxy (catechol structure) and the 3-OH moiety of the C ring are considered the most important factors affecting the reducing capability of flavonoids (Yokozawa *et al.,* 1998; Lien *et al.,* 1999; Heijnen *et al.,* 2001). Whilst the structure-antioxidant activity of phenolic acid and flavonoids have been extensively studied the antioxidant activity of specific hydrophilic compounds of olive oils such as 3,4-DHPEA and *p*-HPEA is very limited namely because of their complexity and standard unavailability, furthermore studies are based on doping refined seed oils with olive phenolic

compounds rather than microtiter assays. Baldioli *et al.,* (1996) showed that 3,4-DHPEA, 3,4-DHPEA-EDA and 3,4-DHPEA-EA possess much higher antioxidant activity than *p*-HPEA and α-tocopherol; this was attributed to the complex structures of secoiridoids and the huge number of derivatives in natural unrefined extracts.

The NO-scavenging activity (NOS) of each of the phenolic extracts was determined by the method of Tsai *et al.,* (2007) with minor modifications. The results obtained showed no significant variation in the radical scavenging activity of nitrous oxide radicals across phenolic extracts derived from olive oils of different geographical origins. These results suggest that the phenolic extracts were equally effective in scavenging the formation of nitrous oxide radicals. This also suggests that the scavenging of nitrous oxide radicals is not dependent on either the concentration or the reducing activity of the different phenolic compounds present in olive oil. Compared to the other radical scavenging activity assays, NOS is carried out in a medium phosphate buffered saline (PBS) whilst both DPPH and ABTS are carried out in polar protic solvents such as methanol. The type of solvent and polarity may affect the single electron transfer and the hydrogen atom transfer, which are key aspects in the measurements of antioxidant capacity (Perez-Jimenez and Saura-Calixto, 2006). Similar to the previous experiments carried out by Fernandez-Pachon *et al.,* (2004), Villano *et al.,* (2005), Perez-Jimenez and Saura-Calixto (2006) showed that the higher the concentration of the polar protic solvents (methanol or ethanol) the higher is the antioxidant activity of the same extract compared to the antioxidant activity displayed if the assay was carried out in water. Strong hydrogen bonding in water may induce dramatic changes in the H-atom donor activities of phenolic antioxidants, reducing the antioxidant capacity of phenolics (Pedrielli, Pedulci, and Skibsted, 2001; Pinelo *et al.,* 2004). The study carried out by Perez-Jimenez and Saura-Calixto (2006) also showed that in the case of highly water soluble phenolic acids and flavonoids, the changes between methanol and water did not show any dramatic changes in the displayed antioxidant activity, suggesting that the solvent affects different compounds in different ways. From the results obtained in this study it can be concluded that changes in solvent polarity had a marked effect on the radical scavenging activity of the extracts. The presence of a mixture of both hydrophilic and relatively lipophilic compounds such as phenolic acids and secoiridoid compounds respectively within the extracts suggests that in aqueous conditions, the combined effect of the more soluble

phenolic compounds had the most prominent effect on radical scavenging towards nitrous oxide radicals. Thus, the observed antioxidant activity was not inclusive for all the different classes of phenolic compounds present within the extracts

The radical scavenging activity of olive oil extracts against DPPH and ABTS was assesed. The DPPH and ABTS assays are simple, but have some disadvantages and is limited in its applications. The reaction mechanism carried out by DPPH and ABTS is different from the hydrogen atom transfer reaction that normally occurs between antioxidants and radicals as the mechanism involves the transfer of a single electron from the stable DPPH radical or ABTS radical cation (which is also the probe for monitoring the reaction) to the oxidant. The model of scavenging the stable DPPH radical and ABTS radical cation is a widely used method to evaluate the free radical scavenging ability of various samples (Ebrahimzadeh *et al.,*2008). In this experiment the total antioxidant activity (Table 6.2) of the phenolic extracts against DPPH radical and ABTS radical cation was expressed as GAE mg/mL. In conjunction the kinetics of the reaction were also studied as shown in Figure 6.7 and in Table 6.2, where one can see the different EC_{50}s obtained.

Similar trends to those observed with the other antioxidant assays were observed, the indigenous and local imported cultivars showed a significantly lower antioxidant capacity when compared to olive oils derived from other countries. Radical-scavenging activities of all extracts were found to increase with increasing concentration. On examining the EC_{50}s obtained for the two radical scavenging assays it was also found that for DPPH assay the indigenous cultivars had a significantly higher EC_{50} when compared to both the imported local cultivars and to foreign cultivars. This observation further confirms the significantly lower antioxidant activity of the indigenous cultivars. DPPH and ABTS radical-scavenging activity was found that usually, higher total phenolic and flavonoid contents lead to better scavenging activity (Ebrahimzadeh *et al.,* 2009). This is line with the results obtained.

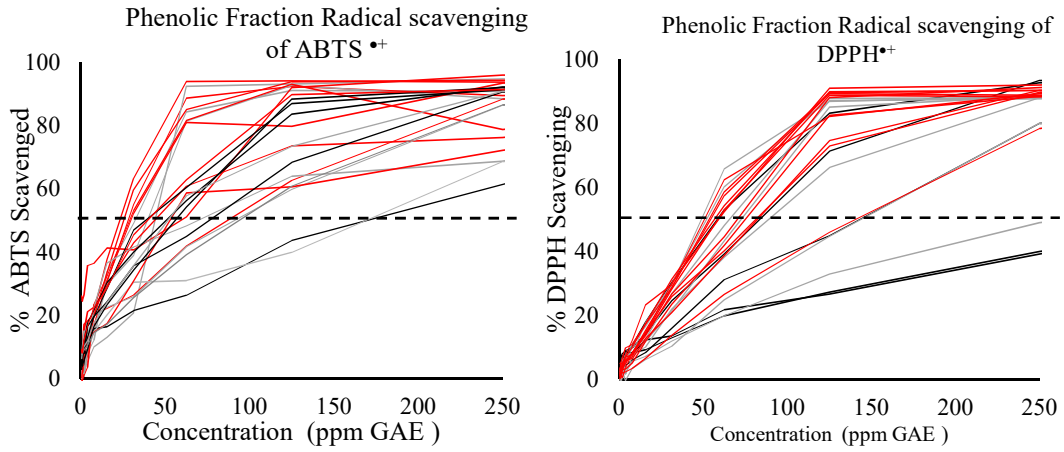


Figure 6.7: Free radical scavenging activity of olive oil phenolic extracts: Graph shows the % inhibition of ABTS (Right) and DPPH (Left) radicals against extract concentrations derived from olives oils of different geographical origins; foreign (Red), imported (Green) and indigenous (Black) cultivars. The dotted line represent 50% inhibition of radicals, the corresponding effective concentration is known as the EC_{50}.

With regards to the EC_{50} for the ABTS radical cations, the trend observed was similar to that observed for the total antioxidant capacity. Whilst the imported cultivar olive oils had a lower EC_{50} activity when compared to foreign cultivars, there was no significant difference in the EC_{50} of the indigenous cultivars when compared to the foreign cultivars. Since for the determination of EC_{50} the same concentrations of phenolic extracts were used, after a stock solution was adjusted to 500 mg/mL GAE, the observed variation between each extract might be dependent on the concentration of different types of flavonoids and phenolic compounds present within the induvial extracts. This hypothesis was later tested when examining the phenolic profile for each cultivar.

Table 6.2: The mean EC_{50} ±1SD of olive oils phenolic extracts from different geographical origins. Superscript letters in the same column represent statistically distinct homogeneous subsets as determined by Kruskal Wallis ANOVA a 95% confidence level.

	EC_{50}_DPPH	EC_{50}_ABTS	EC_{50}_DPPH (Intrinsic)
Indigenous	143.61 ± 76.48^{a}	37.04 ± 12.63^{ab}	0.18 ± 0.11^{a}
Local	87.41 ± 37.92^{b}	31.07 ± 13.87^{a}	0.23 ± 0.04^{a}
Foreign	71.86 ± 21.09^{b}	40.95 ± 11.21^{b}	0.16 ± 0.03^{b}

6.4.1 Intrinsic DPPH radical scavenging activity

Measuring the intrinsic antioxidant activity of olive oil or any other oil presents a number of different challenges, where the main challenge is the solvent of choice. The solvent must be strong enough to dissolve the oil and the stable radical probe, but on the other hand it must not be toxic. It is also ideal that the solvent chosen does not dissolve the microtiter plate which is used for the examination. There are very few studies which target the development of microtiter plate methods in order to determine the intrinsic antioxidant activity of oils as the use of Rancimat methods are generally preferred. Whilst the Rancimat methods provide very good results in terms of the oil oxidative stability they do not provide insights on the actual antioxidant activity of the oils in presence of radicals.

In this part of the study, a variety of different solvents were used in order to assess both the solubility of the oil and the stable radical probe. In previous studies carried out by Lia and Buhagiar (2014) and Ismail *et al.,* (2010) the use of toluenic DPPH rather than methanolic DPPH as previously described by Ramadana and Moersel (2006) was applied in order to determine the intrinsic antioxidant activity of various seed oils, including grapeseed oils and cantaloupe seed oil respectively. Toluene allowed the complete dissolution of the oil while also allowing the oil to come in direct contact with DPPH radicals, however these studies did not employ the use of microtiter assays. In the present study, the use of isoamyl alcohol was employed, as this solvent proved to be most ideal for the determination of the intrinsic antioxidant activity of olive oils. Isoamyl alcohol (3-methyl-1-butanol) allowed the complete dissolution of both the DPPH radicals and the oil, however unlike toluene it did not react with the methacrylate 96- well plates; this allowed the use microtiter plate methods. From the preliminary study carried out on ABTS, results suggest the use of mixture of isoamyl alcohol and dimethyl sulfoxide in order to allow the dissolution of the ABTS radical cation. Unlike DPPH, ABTS radical cation is more polar; furthermore its *in situ* generation employs the use of persulfate, a rather ionic compound. The ideal solvent mixture thus needs be polar enough to dissolve the ionic compound but not so polar that will cause emulsion formation on addition of the oil. In future the use of organic soluble radical generators and the use of external radical generators could be investigated.

A Kruskal Wallis one way ANOVA performed on the data obtained from the EC_{50}s of the different olive oils shown in Table 6.2, corroborated the results obtained in the previous experiments. Olive oils derived from indigenous and locally imported cultivars had a significantly higher EC_{50} when compared to olive oils derived from foreign cultivars whilst no significant difference was observed between the indigenous and the locally derived olive oils. The intrinsic antioxidant activity of the olive oil can be related to a number of different constituents present within the oil such as pigments, sterols, squalene and other minor compounds. The ability of carotenoids to quench reactive singlet oxygen has been linked to the conjugated double bond system, the maximum efficiency being shown by carotenoids with nine or more conjugated double bonds (Foote *et al.,* 1970).The electron rich conjugated double bond structure is primarily responsible for the excellent ability of β-carotene to physically quench radicals without degradation. It is also responsible for the chemical reactivity of β-carotene with free radicals, and for its instability towards oxidation (Britton, 1995; Krinsky, 1994). There are two possible explanations for the antiradical activity mechanism of carotenoids and chlorophylls. The presence of an electron rich conjugated double bond system within the unsaturated molecules allows easy donation of electrons to the DPPH radicals, forming radically stable species. Endo *et al.,* (1985) showed that chlorophyll can deliver hydrogen for DPPH radical reduction, as well as scavenge the fatty radicals formed during lipid oxygenation. In 2005, Lanfer-Marquez *et al*. showed that the metabolic products derived from the degradation of chlorophyll, including phaeophorbide b and pheophytin b, were stronger natural antioxidant compounds than chlorophyll itself, further confirming that carotenoids and chlorophyll and its derivatives show antioxidant properties. However, even though the chlorophyll content increases the radical scavenging activity of DPPH radicals it can also have a pro-oxidant effect on the oil itself causing the oil to become oxidised.

The results obtained from analysing the pigment concentration of olive oils derived from different geographical origin showed that the local and the indigenous cultivars had a higher concentration of pigments when compared to olive oils derived from other geographical locations, thus if it is assumed that the major antioxidant compounds present in olive oils are due to these pigments, olive oils from the Maltese islands would have the higher intrinsic antioxidant activity. However, the opposite was

observed. This unexpected result suggests that the intrinsic antioxidant activity of the olive oils is related to other compounds present in olive oils rather than the pigments.

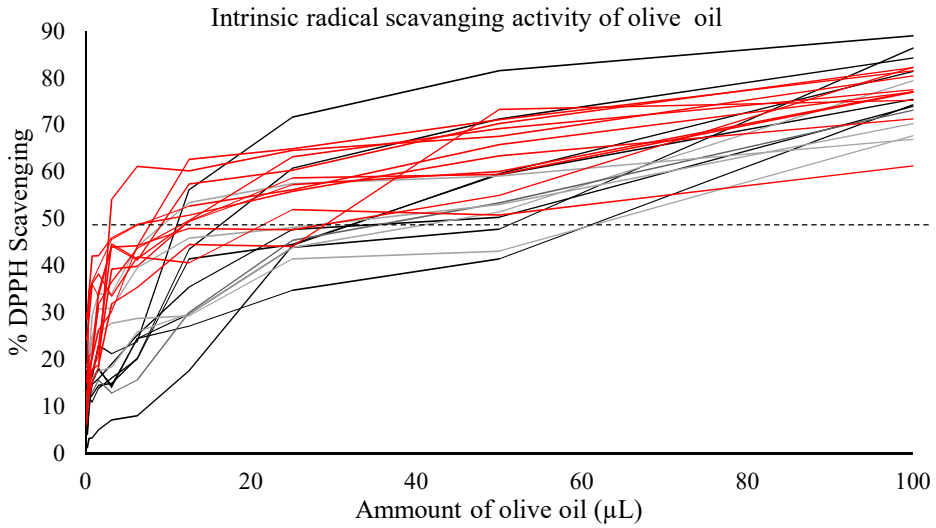


Figure 6.8: intrinsic free radical scavenging activity of olive oils: Graph shows the % inhibition of DPPH radicals against olive oil concentrations different geographical origins; foreign (Red), imported (Green) and indigenous (Black) cultivars. The dotted line represent 50% inhibition of radicals, the corresponding concentration is known as the IC_{50}.

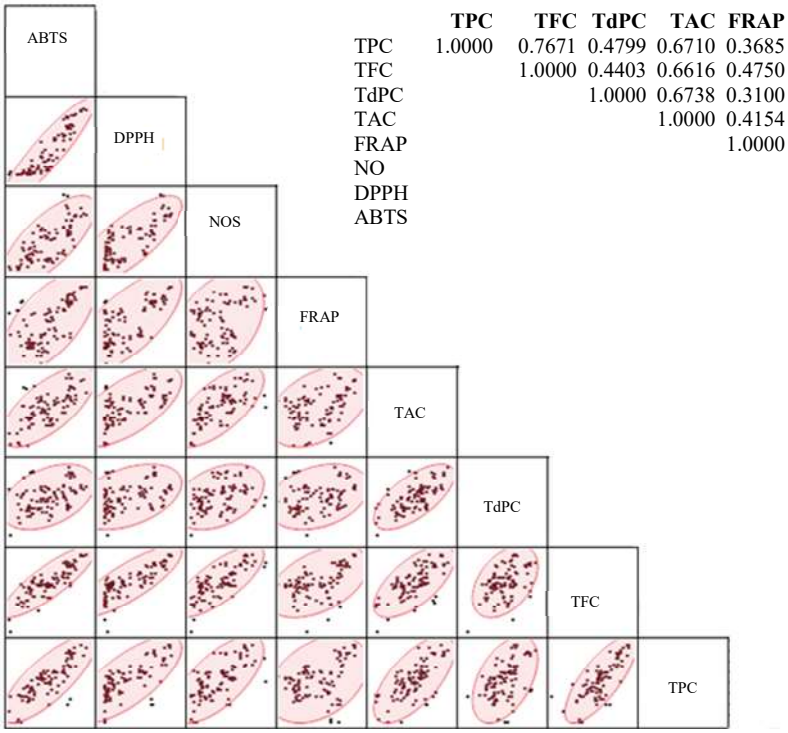
Apart from the presence of carotenoid pigments, phenolic compounds might also contribute to the radical scavenging activity of the oil. It is generally assumed that due to their low solubility, compared to pigments and other lipophilic compounds, the main radical scavengers present within the oil are those which have a higher lipophilic nature however studies done by Schwarz *et al.,*(2000) tend to disagree. Schwarz *et al.,* (2000) introduced the concept of the 'polar paradox'' which states that lipophilic antioxidants are more effective in polar media, while polar antioxidants are more active in lipophilic media, thus although phenolic compounds might be considered as a minor constituents in olive oils their presence can greatly affect both its stability towards oxidation and antioxidant activity.

Given the results obtained from the previous experiments, it was confirmed that the indigenous cultivars had a significantly lower phenolic content when compared to olive oils of different geographical origins. The marginally lower intrinsic radical scavenging activity, can either be attributed to phenolic compounds present in much higher concentration in the foreign cultivars or else to the presence of higher lipophilic antioxidant compounds. Lipophilic phytosterols in the oils can also be considered as a source of antioxidant compounds. The antioxidant activity of phytosterols has been

attributed to the formation of an allylic free radical and its isomerization to other relatively stable free radicals (Warner and Frankel, 1987; Wang *et al.,* 2002). Furthermore the presence of polar lipids, especially phospholipids, which are present in crude, unrefined olive oils are usually considered as free radical scavengers and antioxidant synergists (Hildebrand *et al.,* 1984; Jung *et al.,* 1989). The emulsifying action of phospholipids due to their amphipathic nature can play an important role by increasing the contact between the antioxidant.

6.4.2 Assessing the effect of phenolic content on the antioxidant activity

Figure 6.9 shows the results obtained from the correlation analysis made using nonparametric Spearman's rho in order to assess the relationship between the observed antioxidant activity for different assays and the content of different classes of phenolic constituents present in olive oil. The results obtained show that the observed antioxidant activity determined by different methods showed a correlation with the different phenolic compounds. The table on the right-hand side of the figure shows the correlation values between the different phenolic classes and the antioxidant assays.



	TPC	TFC	TdPC	TAC	FRAP	NO	DPPH	ABTS
TPC	1.0000	0.7671	0.4799	0.6710	0.3685	0.6785	0.7314	0.7842
TFC		1.0000	0.4403	0.6616	0.4750	0.7599	0.8189	0.8528
TdPC			1.0000	0.6738	0.3100	0.3825	0.4182	0.4208
TAC				1.0000	0.4154	0.6782	0.6540	0.6928
FRAP					1.0000	0.3913	0.6974	0.6281
NO						1.0000	0.7725	0.7256
DPPH							1.0000	0.9108
ABTS								1.0000

Figure 6.9: Correlation analysis between the observed total phenolic content (1st row), total flavonoid content (2nd row) and total diphenolic content (3rd row) against the observed antioxidant activity determined using $ABTS^{+\cdot}$ (1st column), DPPH (2nd column), $NO^{\cdot}$ (3rd column), FRAP (4th column) and TAC (5th column). (Right) Table represent the correlation coefficient using Pearson's correlation analysis.

The TAC, NOS, FRAP, ABTS and DPPH methods are mainly single electron transfer assays which operate on the same principle as the determination of total phenolic content by the Folin-Ciocalteu reagent method. The difference in the extent of correlation among the different methods is mainly attributed to the pH of the reaction medium (Ballus *et al.,* 2015). Total phenolic content is measured under basic conditions to allow the phenolic proton to dissociate, leading to a phenolate anion, which is capable of reducing Folin-Ciocalteu reagent. In comparison, both the total *o*-diphenolic and flavonoid content are determined under acidic pH, in order to fully oxidise the diphenolic compounds and allow for the formation of acid stable complexes with aluminium. The NOS, DPPH and ABTS method are performed at neutral pH, whereas the FRAP assay is performed under acidic (pH 3.6) conditions. The effect of pH and medium can explain the difference between the observed absolute values obtained for the different assays since different antioxidant assays operate under different mechanisms. In the case of DPPH the mechanism is thought to involve a

hydrogen atom transfer reaction, but it studies done by Huang *et al.,* (2005) showed that the rate-determining step involves a fast electron transfer process from the phenoxide anions to DPPH. The hydrogen atom abstraction from a phenolic compound by DPPH becomes a negligible reaction path as it occurs very slowly in strong hydrogen-bond-accepting solvents, such as methanol and ethanol (Huang *et al.,* 2005). Thus, it is easy to understand why the results of all of the antioxidant methods displayed the same trends, despite having different absolute values. The differences in the absolute values is dependent on the reaction type, rate of reaction, medium and also because the structures of the compounds and their ease of reduction under a single electron transfer reaction as discussed by Ballus *et al.,* (2015). The application of correlation between the different phenolic classes and each of the antioxidant capacity methods provides a more empirical and holistic evidence rather than comparing each individual assay on its own.

Ballus *et al.,* (2015) presented results similar to this study, whereby the phenolic extracts of different olive oils derived from different cultivars was correlated to different antioxidant assays. Ballus *et al.,* (2015) showed there was a high and significant positive correlation between the total phenolic contents and FRAP (R^2 = 0.8904), ABTS (R^2 = 0.7837) and DPPH (R^2 = 0.7908). In a similar study carried out by Samaniego Sánchez *et al.,* (2007) the correlations of the total phenolic content with the observed antioxidant activity determined by ABTS and DPPH were 0.8586 and 0.7914, respectively. Comparable values were obtained for the same antioxidant assays in this study as a positive significant correlation was conserved for both DPPH (R^2=0.7314) and ABTS (R^2= 0.7842) whilst a lower correlation was obtained for FRAP (R^2=0.3685). Comparing the results obtained to those present in the literature, TPC seems to express a higher correlation with ABTS rather than DPPH; this can be attributed to the different mechanisms under which radical stabilisation takes place. The significantly lower correlation observed in the present study between the TPC and FRAP assay was attributed to the different methodologies which were employed in the determination of ferric reducing potential of the extracts. Results for the correlation analysis of flavonoids and *o*-diphenolic content to antioxidant were not available in the existing literature.

Figure 6.10 shows the correlation analysis obtained on comparing the EC_{50} obtained from examining the intrinsic antioxidant activity of olive oils. The results

obtained showed a lower correlation value with respect to TPC and TFC when compared to those obtained using the hydrophilic extracts. Furthermore, the intrinsic antioxidant activity of olive oils showed no correlation with the *o*-diphenolic content. Similar results were obtained by Minioti and Georgioua (2010) whereby the EC_{50} obtained from DPPH of lipophilic olive oil fraction correlated poorly with the phenolic constituents. The lower correlation between the intrinsic antioxidant activity and the phenolic constituents might be due to a synergistic relationship between the tocopherols and phenolics (Espin *et al.,* 2000).

The results are presented by Ballus *et al.,* 2015 whereby, similar to this study the phenolic extracts of different olive oils derived from different cultivars was correlated to different antioxidant assays. Ballus *et al.,* 2015 showed there was a high and significant positive correlation between the total phenolic contents and FRAP (R^2 = 0.8904), ABTS (R^2 = 0.7837) and DPPH (R^2 = 0.7908), in a similar study carried out by Samaniego Sánchez *et al.,* (2007) the correlation of the total phenolic content with the observed antioxidant activity determined by ABTS and DPPH was 0.8586 and 0.7914. Comparable values were obtained for the same antioxidant assays in this study as the positive significant correlation was conserved for both DPPH (R^2=0.7314) and ABTS (R^2= 0.7842) whilst a lower correlation was obtained for FRAP (R^2=0.3685). Comparing the results obtained to those present in the literature TPC seems to express a higher correlation with ABTS rather than DPPH, this can be attributed to the different mechanisms under which radical stabilisation takes place. The significantly lower correlation observed in the presented study, between TPC and FRAP assay was attribute to the different methodologies which were employed in the determination of ferric reducing potential of the extracts. Results for the correlation analysis of flavonoids and *o*-diphenolic content to antioxidant were not available in the existing literature.

Figure 6.10 shows the correlation analysis obtained on comparing the EC_{50} obtained from examining the intrinsic antioxidant activity of olive oils. The results obtained showed a lower correlation value with respect to TPC and TFC when compared to those obtained using the hydrophilic extracts. Furthermore, the intrinsic antioxidant activity of olive oils showed no correlation with the *o*-diphenolic content.

Similar results were obtained by Minioti and Georgioua 2010 whereby the EC_{50} obtained from DPPH of lipophilic olive oil fraction correlate poorly with the phenolic constituents. The lower correlation between the intrinsic antioxidant activity and the phenolic constituents might be due to a synergistic relationship between the tocopherols and phenolics (Espin *et al.,* 2000).

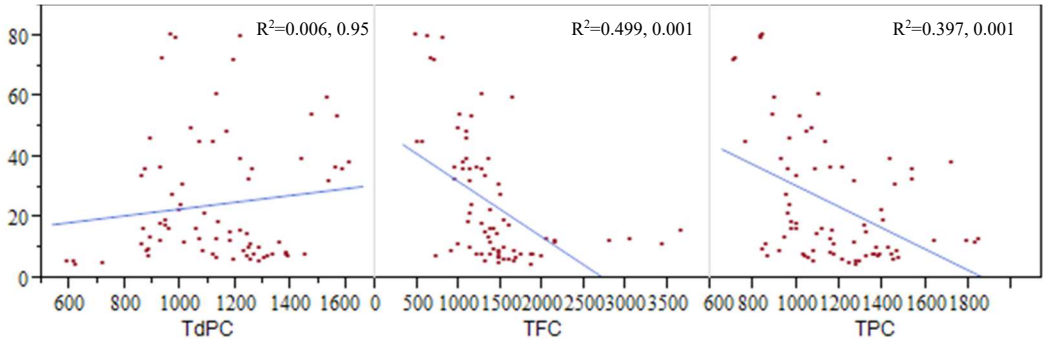


Figure 6.10: Correlation analysis on the intrinsic antioxidant activity of olive oils and the different classes of phenolic compounds present within the hydroalcoholic phase. Figures in each box represent the R^2 value obtained on using Pearson's correlation analysis, followed by the p-value.

It was observed that, irrespective of the method used to determine the antioxidant activity, it is the total flavonoid content of the extract rather than its total phenolic content which is mostly correlated to the antioxidant activity, whilst the total *o*-diphenolic content showed the lowest correlation with the observed antioxidant activity. The higher correlation of antioxidant activity to the flavonoid content was also observed by other authors, who worked with phenolic extracts derived from other botanical origins. Sung and Lee (2010) reported a higher positive correlation between radical scavenging activity and TFC found in phenolic extracts of grape seeds. The higher correlation observed between the TFC and the antioxidant activity, when compared to TPC, is due to the low specificity of the Folin-Ciocalteu method, as the colour reaction can occur with any oxidizable phenolic hydroxy group (Capannesi *et al.,*2000; Hrncirik *et al.,*2004). In the case of flavonoids, the free radical scavenging activity is related to their ability to donate hydrogen atoms to free radicals. As reviewed by Pietta *et al.,* (2000) various authors have examined the radical scavenging activity of flavonoids, and numerous attempts have been made to establish the relationship between flavonoid structure and their radical-scavenging activity. The

radical-scavenging activity of flavonoids is influenced by the molecular structure and the substitution pattern of hydroxyl groups, depending on the flavonoids' ability to donate phenolic hydrogen and on the stability of phenoxyl radicals formed after the donation of hydrogen. The catechol structure is the main target for free radicals due to its ability to donate electrons (van Acke *et al.,* 1996).

From the correlation analysis, it was shown that the phenolic compounds in olive oil are efficient electron donors, regardless of the reaction medium conditions and the compounds to be reduced. Similar correlation values were observed between the total phenolic contents and ABTS and DPPH for EVOO polar extracts in the study of Ballus *et al.,* 2015 and Samaniego Sánchez *et al.,* (2007) suggesting that these methods have a similar extrapolative capacity of olive oil antioxidant activity. The high correlations between the TPC, DPPH and ABTS methods indicate that the total phenolic content can be used as an indicator for olive oil antioxidant activity assessed in the hydrophilic fraction however it seems that the more functionally active flavonoids seems to provide the best indicator for the antioxidant activity.

6.4.3 Assessing the effect of phenolic content and antioxidant activity on the geographic origin of olive oils

In this study the use of phenolic content and the observed antioxidant activity towards different target molecules were used in order to provide information for their potential use as markers to identify the geographical origin of the olive oil. This was assessed using supervised and unsupervised techniques. Figure 3.21 shows the principal component analysis carried out on the results obtained from the aforementioned parameters. The PCA plot was used in order to assess natural variation within the data, and the results obtained showed that the use of only two principal components explained 79.2% of the variance of the data based on the total number of different phenolic constituents and the observed antioxidant activity. The score plots obtained classified the oils into at least 2 distinct areas; the indigenous and the local imported cultivar subset, and the locally imported and the foreign olive oils. This suggests that locally imported cultivars display characteristics which resemble both the indigenous cultivars and the foreign cultivars. These results further highlight the effect of the environment and how it modifies the phenolic content and the antioxidant activity of the oils. The indigenous cultivars which have been in the Maltese islands for a longer period of time display characteristics that implicate that they are fully

adapted to the local conditions, whilst the imported cultivars are still in the process of adapting the local environment.

The EC_{50} obtained against DPPH radicals had the largest influence of the separation of the two subsets in the 1st principal component, which accounted for 66.3% of the observed variation, whilst the total *o*-diphenolic content, FRAP and radical scavenging activity against ABTS radical cations had a large influence on the separation of the two subsets in the 2nd principal which accounted for 12.9% of the observed variation.

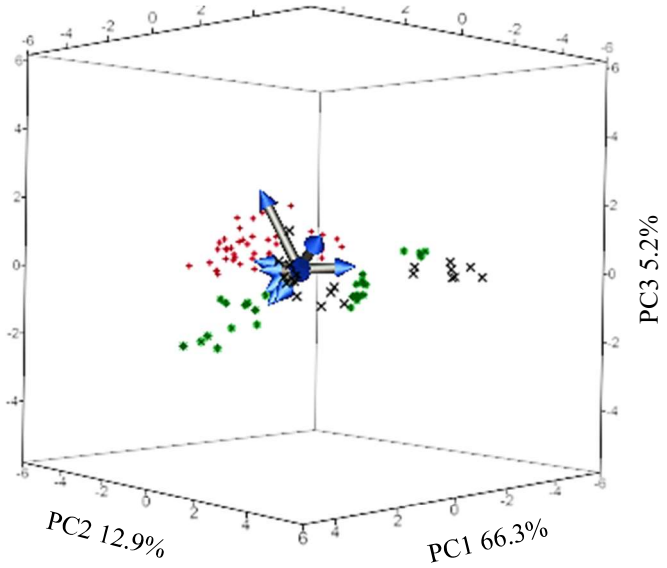


Figure 6.11: Principal component analysis showing clustering of the different monocultivar olive oils based on their phenolic content and observed antioxidant activity.

The use of supervised techniques in order to classify the different geographical origins of olive oils was assessed by the use of canonical discriminate analysis (CDA) (Figure 3.22). The results obtained showed that CDA was able to fully distinguish between the different subsets with a low % misclassifications (2.174%) and furthermore the model was able to fully explain the observed variation by the use of 2 discriminate functions, a 1st discriminate function that explained 81.0 % of variation whilst the 2nd discriminate function explained 19.0% of the variation. X and Y fitting obtained by plotting the scores obtained from the 1st discriminate function showed that the projected formula was able to fully separate the olive oils of foreign origin from those obtained locally whilst by the 2nd discriminate function the indigenous cultivars were fully separated from both the foreign and the imported cultivars. The results

obtained from the FRAP and the EC_{50} against ABTS radical cations had the highest scoring coefficients and were the most discriminate parameters.

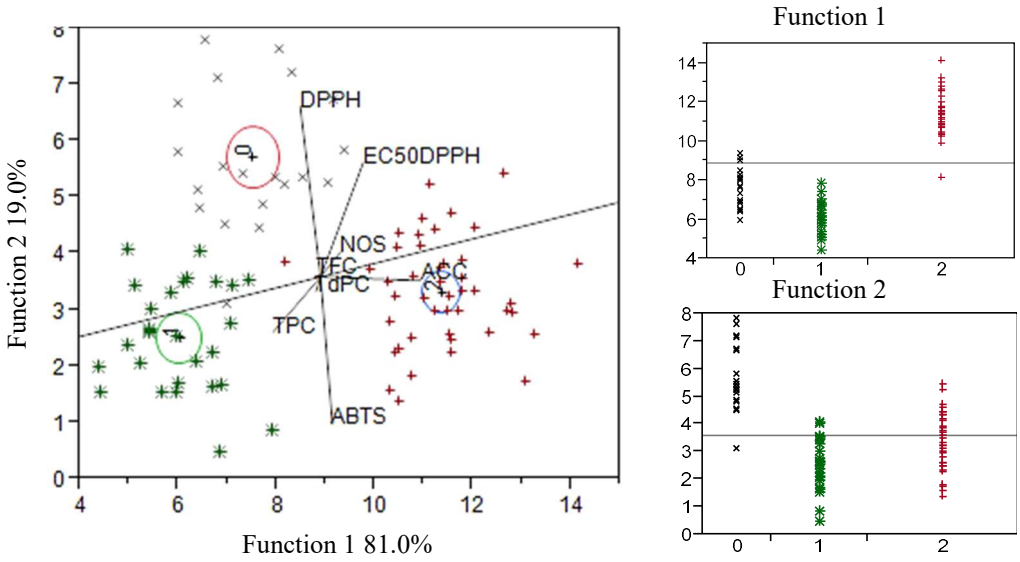


Figure 6.12: Biplot obtained using SLC-DA showing clustering of the different monocultivar olive oils based on their phenolic content and observed antioxidant activity. Black(x) represent the olive oils derived from indigenous cultivars, Green (*) represent locally derived foreign olive oils, Red (+) represent foreign olive oils

7. Application of high performance chromatography for the determination of geographical origin of olive oils

Analysis of the phenolic compounds in virgin olive oil is mainly performed using reverse phase-HPLC, however due to the complexity of the phenolic mixture, the chemical nature has not been completely elucidated. However with advances in the field of liquid chromatography especially with the introduction of ultra-high pressure liquid chromatography, the study of these compounds has registered a significant improvement (Capriotti *et al.,* 2014 Cao *et al.,* 2013; Alarcón Flores *et al.,* 2012). A large number of studies have focused on the relationship between the phenolic content and the oxidative stability of the oil (Tsimidou *et al.,*1999; Montedoro *et al.,*1992; Mateos *et al.,*2001; Papadopoulos *et al.,*1991) and on different factors which may alter the phenolic composition such as the extraction process (De Stefano *et al.,*1999; Garcia *et al.,*2001; Vierhuis *et al.,*2001; Perez *et al.,*2003), the producing cultivars (Aparicio and Luna 2002 ; Luna *et al*., 2006 ; Kalua *et al*., 2005), the environmental factors (Paz Aguilera *et al*., 2005 ; Mousa and Gerasopoulos 1996 ; Ranalli *et al*. 1999 ; Ripa *et al*. 2008 ; Servili *et al*., 2004 ; Tous *et al*., 1997 , 1999 ; Tura *et al*., 2008, Berenguer *et al*.,2006), degree of ripeness (Dag *et al*., 2011 ; Famiani *et al*., 2002 ; Gómez-González *et al*., 2011 ; Lazzez *et al*., 2008 ; Mailer *et al*., 2010 ; Varzakas *et al*., 2010) and storage (Li,X *et al.,*2014; Fregapane *et al.,*2006 and 2013; Migliorini, *et al.,*2013).

Very few studies have been published that go deeply into the effect of the olive variety on the composition of the phenolic profile or the use of chemometrics as a tool to classify olive oil varieties according to their geographical origin, although it has been suggested by a number of studies (Mateos *et al.,* 2001; Garcia *et al.,* 2001). Recent studies have shown that these minor constituents of the olive oil can used in the characterization and authentication with respect to geographical origin and cultivars (Petrakis *et al.,* 2009 ; Lerma-Garcia *et al.,* 2009; Ocakoglu *et al.,* 2009 ; Ouni *et al.,*2011; Alkan *et al.,*2012). These studies showed that with the combination of specific quantified phenolic compounds, different analytical techniques, in conjunction with chemometric analysis, it is possible to obtain a classification system which reflects the geographical origin of the olive oils.

The aim of this study was to establish the use of phenolic profile as a tool for the determination of geographical origin, focusing mainly on the discrimination of

Maltese olive oil from other extra virgin olive oils (EVOO) derived for other countries in the Mediterranean. The classification of olive oil samples according to their phenolic profiles was performed by the application of both supervised and unsupervised techniques namely principal component analysis (PCA), hierarchical cluster analysis (HCA), partial least square regression (PLS) and stepwise linear canonical discriminate analysis (SLC-DA). The findings of this study can provide ways for the varietal authenticity of Maltese olive oil according to their phenolic profiles as the geographical indicators, therefore they can be used in PDO or PGI labelling of Maltese EVOOs. The novelty of this study is the development of a strategy based on the analysis of the whole phenolic profile in collaboration with multivariate analysis. The fingerprint phenolic profile was treated as a continuous form of non-specific variables which change in terms of absorbance and retention time. The identification of phenolic profiled related to the geographical origin of EVOOs through the blind analysis of chromatographic profiles has not been previously reported. The majority of the research published aims to identify the individual peaks and the resulting compounds. The different composition based on the analysis of specific quantifiable peaks is currently used in order to assess differences in the phenolic profiles. The novelty about this study was to develop a technique which is able to operate on the whole phenolic profile, rather than on distinct peaks (compounds) without the need of previous identification of the chemical marker.

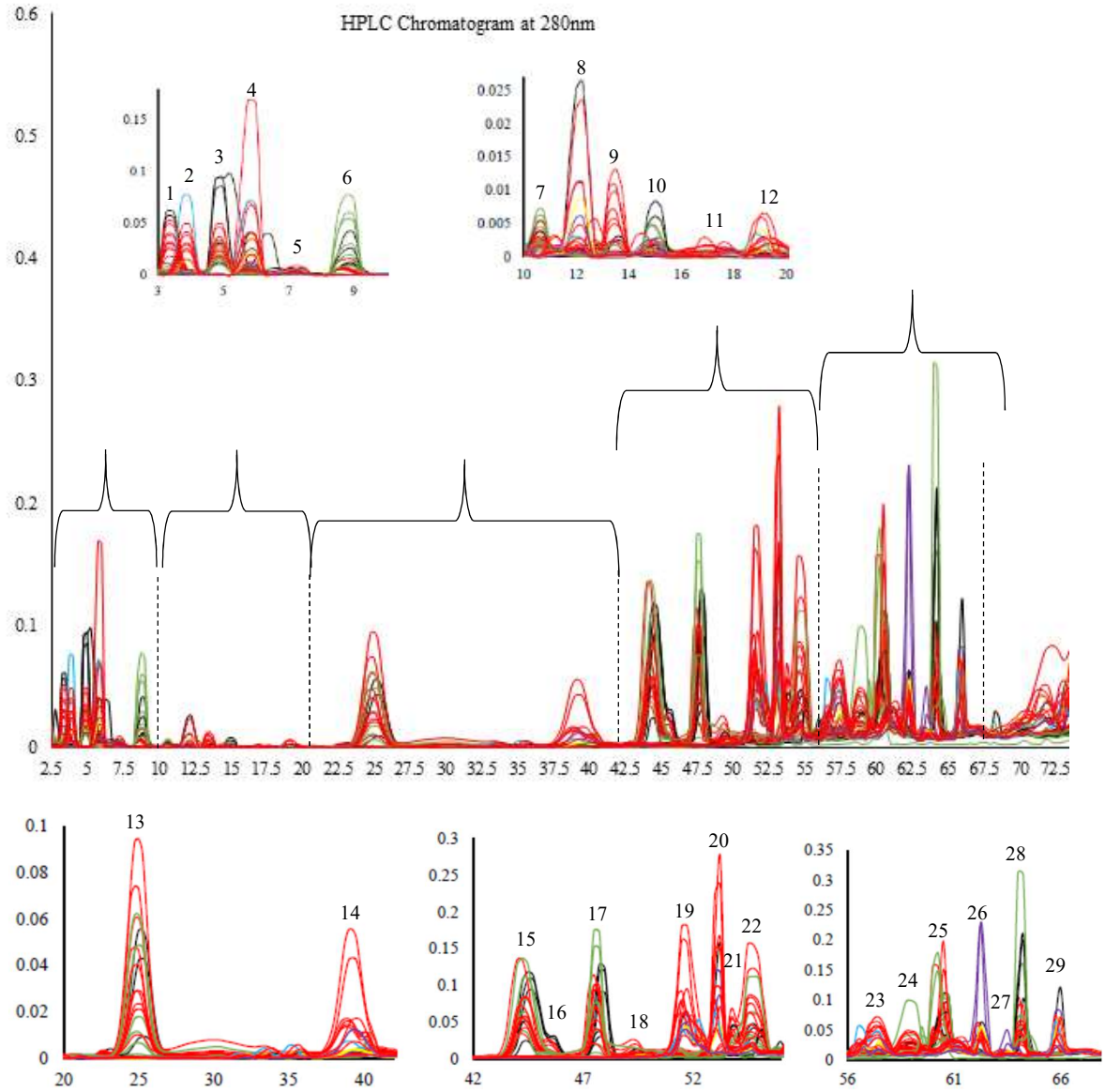
7.1 Application of Univariate Analysis of peaks

In the first part of this study, the use of typical chromatographic data processing was employed. The chromatograms were first integrated using Empower and their corresponding peak area was obtained, each peak area and its corresponding retention time were aligned and subjected to univariate analysis. The purpose of this part of this study was to compare the results obtained from the traditional chromatographic data handling to the proposed chemometric methods.

Figure 7.1 shows the major peaks which were obtained - a total number of 28 peaks. Comparison to the existing literature enabled the identification of more peaks, notwithstanding minor changes to the pH of the mobile phase and to the gradient elution program. The use of a smaller %v/v of acetic acid in solvent A effects the extent of phenolic compound ionisation; under reverse phase conditions, the ionised form is

more polar and thus less well retained. Thus the use of less acidic pH ensured that phenolic compounds were less ionised in order to increase their interaction time with the stationary phase and thus favouring separation. The use of a longer gradient time also improved the separation, however this came at the cost of more solvent and longer analysis times.

Peaks were identified (Figure 7.1) on the bases of their retention times in comparison to the available standards. In the case of unavailable standards peak identification was carried out by comparison of the relative retention time with respect to syringic or hydroxytyrosol to those present within the literature. The use of typical response parameters, elution patterns and absorption parameters were also taken into account in the identification of peaks where standards were not available. However a total number of 9 peaks were still not identified - the future use of UPLC-MS/MS should enable both the identification of the unknown peaks obtained and confirm the identity of peaks which were identified via the relative retention times (Appendix 19 shows the preliminary work which was done). Quantification of the individual peaks is currently carried out and thus statistical analysis presented is given in terms of peak areas. The dialdehydic form of elenolic acid linked to tyrosol (*p*-HPEA-EDA) was generally the main phenolic compound followed by the dialdehydic form of elenolic acid linked to hydroxytyrosol (3, 4-DHPEA-EDA). *p*-HPEA-EDA was found in all samples analysed whilst 3, 4-DHPEA-EDA was not found and this was mainly attributed to its high oxidability and ultimately its degradation. Other derivatives of hydroxytyrosol and tyrosol were found in relatively high concentrations including oleuropein (3, 4-DHPEA-EA) and ligstroside aglycones (*p*-HPEA-EA).



Peak1	3.0-3.45	Gallic Acid/ Quinic Acid
Peak2	4.0-4.27	Hydroxytyrosol
Peak3	4.9-5.06	Protocatechuic acid
Peak4	5.3-5.9	2,4-hydroxyphenlyethanol
Peak5	7.0-7.51	Syringic Acid
Peak6*	8.89-9.0	*p*-coumaric acid
Peak7	10.9-11.5	Vanillin
Peak8	11.8-12.2	*o*-coumaric acid
Peak9	13.1-13.6	Ferulic acid
Peak10*	14.8-15.2	Tyrosol Acetate
Peak12	19.2-19.9	3- methoxycinnamic acid
Peak13	24.9-25.5	Verbascoside
Peak14	39.1-41.0	Oleuropein

Peak15*	42.8-43.1	3,4 DHPEA-EDA
Peak16*	44.7-45.5	Unk1 Hydroxy-decarboxymethyl oleuropein aglycone
Peak17*	47.3-47.8	p-HPEA-EDA
Peak18	48.9-49.6	Cinnamic Acid/Acetoxypinoresinol
Peak19	51.7-52.3	Pinoresinol
Peak20	53.1-53.5	Luteolin
Peak21	53.8-54.1	Unk3 Decarboxymethyl-ligstroside aglycone
Peak22	54.5-55.5	3,4 DHPEA-EA
Peak23	56.7-57.3	Unk4 Hydroxy-decarboxymethyl-ligstroside aglycone
Peak24	57.6-58.3	p-HPEA-EA
Peak25	59.0-61.0	Unk6
Peak26	61.5-62.1	Unk7 10-Hydroxy-oleuropein aglycone
Peak27	62.5-62.8	Unk8
Peak28*	63.4-64.3	Unk9 Aldehydic form of oleuropein aglycone
Peak29	65.5-66.6	Unk10

Figure 7.1: Phenolic chromatogram observed at 280 nm obtained for all the monocultivar EVOOs studied different colours represent olive oils derived from different geographical origin. Black = Indigenous Maltese cultivars; Green = Foreign cultivars which are locally grown; Red = Italian cultivars Blue = Greece origin; Purple= French origin and Yellow = Spanish origin. The peaks are labelled accordingly and identified in the table.

The main simple phenols found in the EVOO were hydroxytyrosol and tyrosol; from all the samples tested the presence of the oxidised corresponding hydroxyphenyl acetic acid was not found. Chlorogenic acid, first reported to occur in leaves of *Olea europaea*, was found for the first time in olives by Ryan *et al*. (2002) and was later confirmed by Vinha *et al.,* 2005 and Dagdelen *et al.,* 2013. Other phenolic acids including *p*-hydroxybenzoic, *o*-coumaric, hydroxy caffeic, *p*-hydroxyphenylacetic, and syringic acids synaptic, were very low and in the majority of cases undetectable which is consistent with reports by other authors for other olive oil varieties (Baldioli *et al.,* 1999, Gómez *et al.,* 2002, Montedoro *et al.,* 1992, Gómez-Alonso *et al.,* 2002). Although these phenolic acids have been found in literature, these were not found in this study mainly due to the different extraction procedure which was carried out. The use of solid phase extraction over the conventional liquid-liquid extraction could have caused the loss of some phenolic acids, especially during the washing step involving the 85:15 hexane/ethyl acetate mix.

In general the concentration of tyrosol was higher than that of hydroxytyrosol and was found in all of the samples analysed. These results are similar to those reported by several authors for other olive oil varieties (Tsimidou *et al.,* 1992; Baldioli *et al.,* 1999; De Stefano *et al.,* 1999 and Gómez *et al.,* 2002). The tyrosol/ hydroxytyrosol ratio was very similar to extra virgin olive oils obtained from other Mediterranean countries. The reason for the wide distribution of these two phenols is attributed to the partial hydrolysis of their derivatives releasing hydroxytyrosol and tyrosol (Montedoro *et al.,* 1992; Cinquanta, *et al.,* 1997). In addition, the higher antioxidant activity of hydroxytyrosol when compared to tyrosol (Papadopoulos *et al.,* 1991) makes it more susceptible to degradation and disappearance.

The presence of pinoresinol and acetoxypinoresinol could not be fully identified furthermore the presence of trans-cinnamic acid was found to coelute with 1-acetoxypinoresinol. According to Owen *et al*. (2000) these compounds are the main components of the phenolic fraction which is derived from the olive seed as they are practically absent from the pulp, leaves, and therefore their presence in the oil must be due to breaking of the pits when the olives are crushed. The presence of these compounds could be potentially be used as an index of the crushing conditions and of the fruit pulp/seed ratio during olive processing as suggested by (Gómez-Alonso *et*

*al.,*2002). With respect to luteolin and apigenin, the flavone compounds were identified in the majority of the samples in substantial amounts.

Table 7.1: Peak areas and their corresponding compounds which were found to vary significantly under ANOVA statistical test across the three classes of EVOOs studied.

	Peak6 *p*-coumaric acid	Peak10 Tyrosol Acetate	Peak15 3,4 DHPEA-EDA	Peak16 Unk1	Peak17 *p*-HPEA-EDA	Peak28 Unk9
Indigenous	$3.81x10^{05a}$	$1.20x10^{05a}$	$2.62x10^{06a}$	$1.81x10^{06a}$	$1.56x10^{06a}$	$1.00x10^{06a}$
Local	$9.80x10^{05b}$	$3.57x10^{04c}$	$4.27x10^{06b}$	$2.12x10^{06b}$	$1.37x10^{05c}$	$6.26x10^{05b}$
Foreign	$9.30x10^{04\ a*}$	$2.95x10^{04c}$	$1.16x10^{06a}$	$8.57x10^{05b*}$	$1.23x10^{05c}$	$2.84x10^{05b}$

*Superscript letters in the same column represent statistically distinct homogeneous subsets as determined by ANOVA post hoc Tukey analysis at a 5% confidence level, same letters followed by an * indicate homogeneous subsets at 10% confidence level*

The application of univariate statistical analysis on the distinct peak areas showed a significant difference in the concentration of individual compounds between the different EVOO's derived from different geographical locations. Table 7.1 illustrates the peaks which showed a significant difference under the application of ANOVA, Tukey post hoc hypothesis testing for the analysis of variance within the between the different groups is also shown. The results obtained showed that EVOOs derived from locally grown cultivars and indigenous cultivars had a significantly higher concentration of *p*-coumaric acid when compared to EVOOs derived from other Mediterranean countries, the observation was significant at the 95% confidence level when comparing local to foreign EVOO and significant at the 90% confidence level when comparing the indigenous to the foreign EVOOs.

In the case of tyrosol acetate EVOOs derived from indigenous cultivars had a significantly higher concentration when compared to EVOOs derived from both foreign and imported cultivars. Similar results were obtained on comparing the content of the unknown peak 1 which was tentatively assigned as hydroxy-decarboxymethyl oleuropein aglycone; unknown peak 9 (tentatively identified as aldehydic form of oleuropein aglycone) and 3, 4-DHPEA- EA.

In the case of *p*-HPEA-EDA it was found that the EVOOs derived from imported cultivars had a significantly higher concentration when compared to both the EVOOs derived from indigenous cultivars and foreign grown cultivars. It is quite difficult to

compare the results of phenolic concentrations provided by different studies owing to the great variety of factors which affect the individual components of phenolic compounds present in EVOO including genetic, pedo-climatic, geographical origin, agronomic and technological factors. El Riachy *et al.,* 2012 showed that the influence of the genotype seems to be greater than that of the ripening index and interaction on both total phenols content and individual phenols content was the main contributor for the spatial separation of the different cultivars, however certain compounds including tyrosol, luteolin and apigenin were not affected by neither the genotype nor the index maturity. Similar results further confirmed that the genotype is the main contributor for the phenolic difference have been observed by Montedoro and Garofolo, 1984 and Brenes *et al.,* 1999. On the other hand Skevin *et al*. (2003) showed that ripeness of olive fruits exerts greater effect than cultivars on the total phenolic profile of olive oils. The differences among the cultivars was significant only at the initial stages of ripening, when the phenolic content was higher, than at the end of the ripening process, when the differences between the cultivars were not significant. This decrease was most likely correlated with the increased activity of hydrolytic enzymes observed during ripening which alters both the total phenolic content and the individual phenol concentration.

The evolution of individual phenols throughout the ripening period of olive fruits has been studied by several authors. The decrease in the 3, 4-DHPEA-EDA is generally accompanied by an increase in hydroxytyrosol content with ripening, which is attributed to gradual hydrolysis of oleuropein first to the aglycone form and then, to hydroxytyrosol (Brenes *et al.,* 1999; Baccouri *et al.,* 2008; Jemai *et al.,* 2009). However, other studies have shown that both 3, 4-DHPEA-EDA and hydroxytyrosol decreased during ripening (Morelló *et al.,* 2004; Damak *et al.,* 2008). This is due to polymerization reactions which transform these two compounds into phenolic oligomers, as described by Cardoso *et al*. (2006). Phenolic acids have also been found to vary with respect to the degree of ripeness with a general decrease in sinapic and *o*-coumaric acids, and an increase in the contents of *p*-coumaric and syringic acids. From the results obtained the high concentration of *p*-coumaric acid present in the indigenous cultivars and the imported cultivars compared to foreign EVOOs might be due to difference in ripening stages of the different olives, however the differences in the ripening stages does not explain the significantly higher concentration of 3,4-

DHPEA-EA in the indigenous cultivars but not imported cultivars. The significantly higher concentrations of 3, 4-DHPEA-EA, hydroxy-decarboxymethyl oleuropein aglycone and aldehydic form of oleuropein aglycone in the indigenous cultivars can only be attributed to a combination of the genotype and pedo-climatic factors.

The significantly higher concentration of 3, 4 DHPEA-EDA, p-HPEA-EDA, secoiridoid compounds present in the indigenous cultivars has very important implications. Secoiridoid compounds are not soluble in oil and, after the process of mechanical extraction, only a small portion is recovered in the oil, and thus it can be assumed that since the oils were extracted by the same mechanical process. The concentration of these compounds within the actual drupe is higher in the indigenous cultivars than both the imported cultivars and foreign grown cultivars. These compounds also contribute to the quality of olive oil influencing the oil oxidative stability (Servili *et al.,* 2004) and the organoleptic properties, being responsible for bitter and pungency sensory. Several authors suggested that secoiridoid derivatives of oleuropein and demethyloleuropein such as 3, 4-DHPEA-EDA and 3, 4-DHPEA-EA are the main contributors to VOO bitterness (García *et al.,* 2001; Kiritsakis, 1998). Tovar *et al*. (2001) moreover showed a strong correlation between the bitter and pungent tasty notes and the *p*-HPEA-ED. Andrewes *et al.,* 2003 concluded that the *p*-HPEA-EDA is the phenolic compound solely responsible for the majority of the burning-pungent tasty note in EVOO.

On the other hand the antioxidant activity determined by Rancimat test, which evaluates the time (hours) of resistance of the oil samples exposed to a stream of dry air at a temperature of 120 °C to oxidation, showed that 3,4-DHPEA, 3,4-DHPEA-EDA and 3,4-DHPEA-EA possess much higher antioxidant activity than *p*-HPEA and α- tocopherol (Baldioli *et al.,*1996). Artajo *et al.,* 2006 showed that 3, 4-dihydroxy and 3, 4, 5-trihydroxy structures linked to an aromatic ring such as oleuropein, 3, 4-DHPEA-EDA, and the methylated form of 3, 4-DHPEA-EA are the main functional groups responsible for the antioxidant activity of these compounds. Although in this study it was shown that the antioxidant activity of the phenolic fractions derived from indigenous and imported EVOOs were significantly lower than the foreign EVOOs , this does not imply that the effective shelf-life of the Maltese EVOO's would be shorter. As previously stated, similar to the determination of phenolic compound the

methodologies employed do not reflect the actual matrix in which the phenolic compounds are found. Furthermore these methods are oversimplified as they tend to oversensitive to certain classes of phenolic compounds. Obied *et al.,*(2007) showed that linear structured phenolic compounds found in olive mill waste, such as hydroxytyrosol acylclodihydroelenolate, are more efficient radical scavengers than 3,4-DHPEA and oleuropein. This was attributed to the linear structure which makes it more accessible to the sterically hindered DPPH radical (Obied *et al.,*2007). Furthermore, even though it was observed that the Maltese indigenous cultivars had a significantly higher concentration of tyrosol acetate, the observed antioxidant activity was still very low. Studies done by Chen *et al.,* 2013 showed that tyrosol acetate had a significantly lower antioxidant activity towards ABTS FRAP and DPPH when compared to tyrosol. The derivatization of tyrosol effectively lowers its antioxidant behaviour however in the same study it was shown that tyrosol acetate had comparable and sometimes even higher antiosteoporotic effects than tyrosol. These results can be explained in terms of compound bioavailability, increased lipophilicity of tyrosol acetate increased its cellular uptake and membrane crossing abilities (Grasso *et al.,* 2007).

7.2 Statistical correlation between phenolic compounds, microtiter assays and antioxidant activity.

Comparison to the results obtained by the microtiter assays using Folin-Ciocalteu reagent, it was found that total phenolic content obtained from these assays does not necessarily reflect the phenolic profile of the EVOOs. It was shown that indigenous cultivars had a significantly higher concentration of certain phenolic compounds when compared to EVOOs derived from imported and foreign cultivars. The determination of phenolic compounds by Folin-Ciocalteu only measures the number of potentially oxidizable phenolic groups. The number of phenolic groups per molecule will vary greatly both within and among different phenolic compound classes, thus this method is subject to both over and underestimation of phenolic classes. The use of Folin-Ciocalteu provides a very useful index for phenolic content, but it would not be expected to correlate with the actual weight of the individual classes of phenolics compounds present.

From the correlation analysis shown in Figure 7.2 it can be seen that total phenolic flavonoid and *o*-diphenolic content showed a significant positive correlation with respect to a number of different distinct peaks namely, peak 23 (Unk 4, Hydroxy-decarboxymethyl-ligstroside aglycone), peak 24 (Unk 5), and peak 20 (*p*-HPEA-EA) however a significantly negative correlation was found with respect to peak 15 (*p*-HPEA-EDA) and peak 28 (Aldehydic form of oleuropein aglycone) for TPC and TdPC and peak 6 (*p*-coumaric acid) with respect to TFC and TdPC. Peak 23 also showed a significantly high positive correlation with the observed total antioxidant activity (R^2= 0.685 p-value <0.05) whilst the ferric reducing activity of the phenolic extracts seems to be more positively correlated with the presence of minor phenolic acids present synaptic acid (R^2= 0.545 p-value < 0.05). The nitrous oxide radical scavenging activity was found to be significantly moderately correlated to the peak 21 (Unk 3 Decarboxymethyl-ligstroside aglycone R^2 = 0.419 p=value < 0.05).

The observed antioxidant activity was also significantly negatively correlated to number of peaks depend on the antioxidant assay used. The total antioxidant activity and ferric reducing activity were found to be negatively correlated to the peak 6 (*p*-coumaric acid) (R^2= -0.5612 and -0.532 respectively and p-value < 0.05 for both), the nitrous oxides scavenging activity was only negatively correlated to peak 10 (tyrosol

acetate R^2= -0.352, p-value < 0.05). These results suggest that the observed antioxidant activity is dependent on the concentrations of different compounds furthermore the results obtained illustrate the existence of both synergistic and antagonistic interactions of phenolic compounds found in olive oil phenolic extracts which are dependent on the antioxidant activity assay.

The combinations of antioxidants can result in larger overall effects compared to the effect expected from a simple addition of the effects of the individual antioxidants entailing synergism (Uri, 1961). Several studies have shown that plant polyphenols have a synergistic effect with other antioxidants present in plant material (Graversen *et al.,* 2008; Miller and Rice-Evans, 1996; Roberts and Gordon, 2003). Several hypotheses have been proposed in order to explain synergistic and antagonistic effects of antioxidant compounds in both their pure state and in extracts. Peyrat Maillard and others (2003) proposed that synergism in a combination of a weak antioxidant and strong antioxidant can occur, whereby the weak antioxidant may restore the strong antioxidant, improving overall radical quenching ability of the combination. Conversely, antagonism may be explained by the regenerating the weak antioxidant from the strong antioxidant, which in turn reduces the radical scavenging activity. Other hypothesises proposed to explain the interactions of antioxidants are based on changes in the rate reaction, polarity, and the effective concentration of the antioxidants at the site of oxidation (Frankel *et al.,* 1994; Koga and Terao, 1995; Cuvelier *et al.,* 2000).

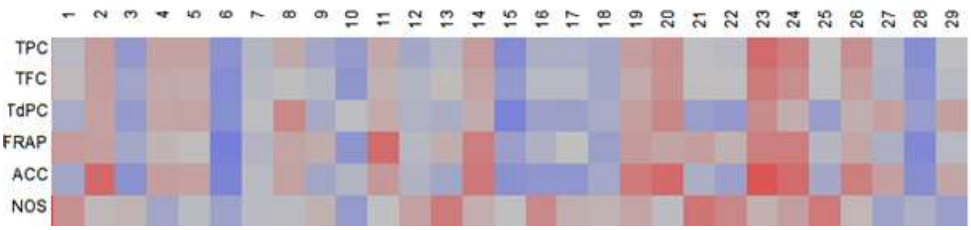


Figure 7.2: Heat map comparing on the Pearson's correlation R^2 value obtained on comparing the individual peaks to the microtiter assays. A positive correlation is shown as red shades, whilst a negative correlation is shown as blue shades, the intensity of the colour represents the strength of the correlation. No correlation is shown as a grey shade.

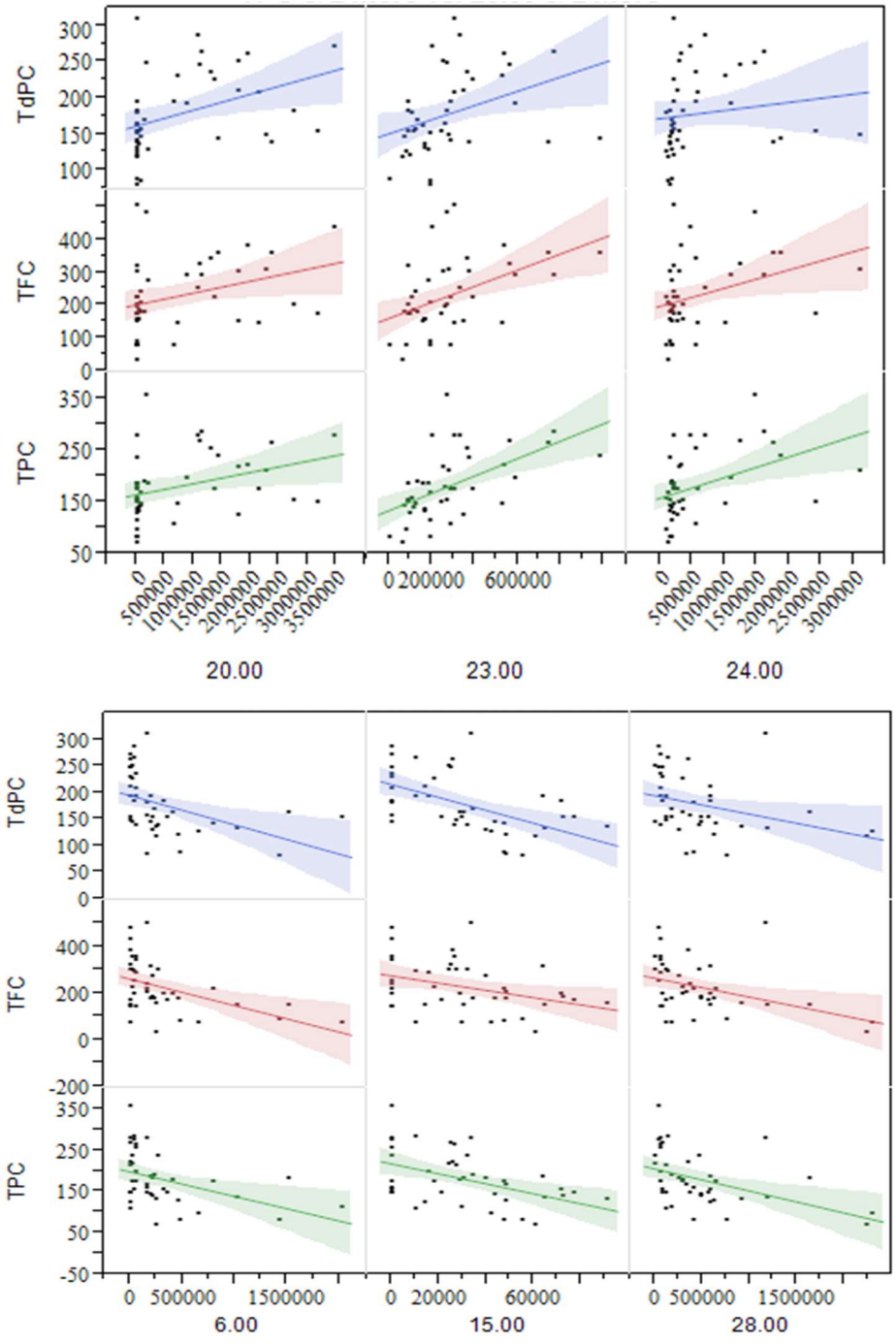


Figure 7.3: Scatter plot showing the most significant correlations observed on comparing the individual peak areas to the observed phenolic content obtained by microtiter plate assays

7.3 Statistical correlation between phenolic compounds

Considering the data set formed by comparing the peak areas of the individual phenolic compounds in the different monovarietal EVOOs used, a Pearson correlational analysis was carried out to find statistical correlations between the individual phenolic compounds. Few studies have focused on correlation analysis among phenolic compounds present in VOO. Pérez *et al.,* 2014 studied the relationship among four groups of phenols in VOOs from Picual/Arbequina crosses. However, this statistical evaluation has not been applied to different monovarietal VOOs. The recent study conducted by Sánchez de Medina *et al.,* 2015 applied the use of correlation analysis for EVOOs phenolic compounds derived from seven Spanish cultivars analysed by LC-MS/MS. In the present work, several correlations were found between pairs of phenols with p-value < 0.05 and correlation coefficient R^2 >0.60, which are indicative of at least a moderate relationship between them.

A significant positive correlation (R^2= 0.632 p-value = 0.02) was found between the amount of gallic acid and protocatechuic acid in the phenolic extracts. These correlations could be tentatively explained according to the main pathways involved in the biosynthesis of phenolic compounds. Multiple studies have suggested the presence of two potential pathways for gallic acid synthesis (Figure 7.4). It could either be formed from phenylalanine or from an initial shikimate intermediate via the 3-dehydroshikimate (3-DHS) (Neish *et al.,* 1964; Kato *et al.,* 1968). However, the carboxylic group of gallic acid has been found to be biosynthetically equivalent to the carboxylic group of shikimate, rather than to the phenylalanine side chain (Werner *et al.,* 2004). This suggests that gallic acid is formed from 3-DHS (Werner *et al.,*2004; Ossipov *et al.,*2003) either by direct dehydrogenation or via protocatechuic acid as an intermediate (Kambourakis *et al.,*2000).

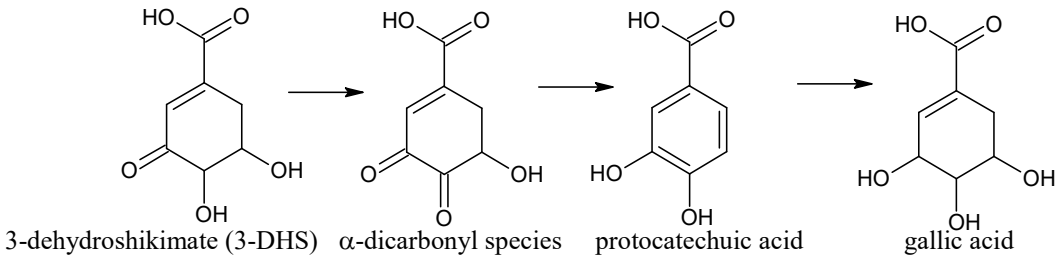


Figure 7.4: Propose biosynthetic pathway linking protocatechuic acid and gallic acid. The biosynthetic pathway illustrated 3-DHS as the primary compound which is oxidised to α-dicarbonyl species which is then reduced back to protocatechuic acid and ultimately to gallic acid.

Furthermore it was observed that protocatechuic acid was significantly positively correlated (R^2= 0.616 p-value 0.03) to 3-Hydroxy-4-methoxycinnamic acid. 3-Hydroxy-4-methoxycinnamic acid can be transformed into hydroxybenzoic acids via a first conversion to the glucose esters which serve as activated intermediates in the side chain shortening reaction (Funk and Brodelius, 1990).

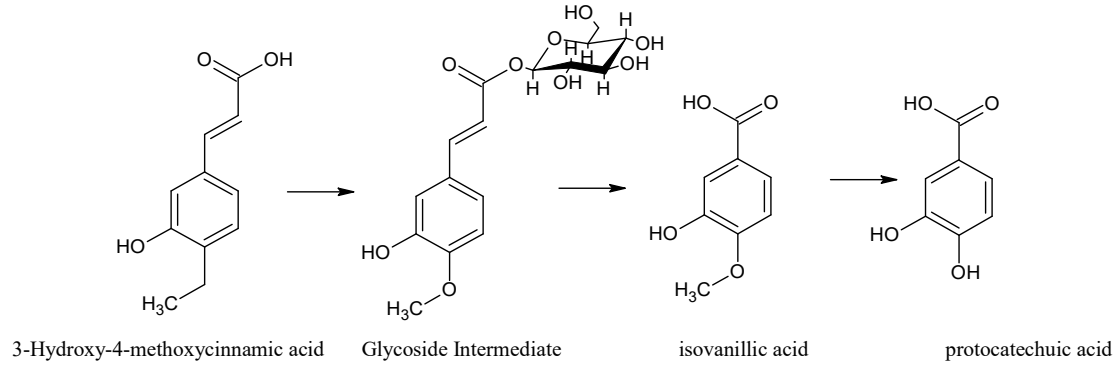


Figure 7.5: Propose biosynthetic pathway linking protocatechuic acid and 3-Hydroxy-4-methoxycinnamic acid. The 3-Hydroxy-4-methoxycinnamic acid undergoes chain shorting reactions via the formation of glycoside, which results in the formation of isovanillic acid intermediate. The hydroxybenzoic acid is tend transformed to protocatechuic acid.

A significantly positive correlation (R^2= 0.602 p-value 0.001) was observed between ferulic acid and the verbascoside content in the EVOOs phenolic fraction. This correlation suggests that the two compounds share a common biosynthetic pathway. Verbascoside is a complex ester composed of hydroxytyrosol, caffeic acid and the sugar alpha-L-rhamnopyranosyl. The presence of the caffeic acid moiety with the molecule suggests that the biosynthesis of verbascoside requires the presence of phenolic acids. However both caffeic and ferulic acid share the same biosynthetic pathway. Caffeic acid is biosynthesized by hydroxylation of the coumaroyl ester of quinic acid, which in turn is formed from the caffeic acid ester of shikimic acid, which converts to chlorogenic acid. The latter acts as the precursor for the formation of ferulic acid catalysed by the enzyme caffeic acid-O-methyltransferase (Boerjan *et al.,* 2003). Figure 7.6 shows the proposed biosynthetic pathway linking the production of ferulic acid and verbascoside.

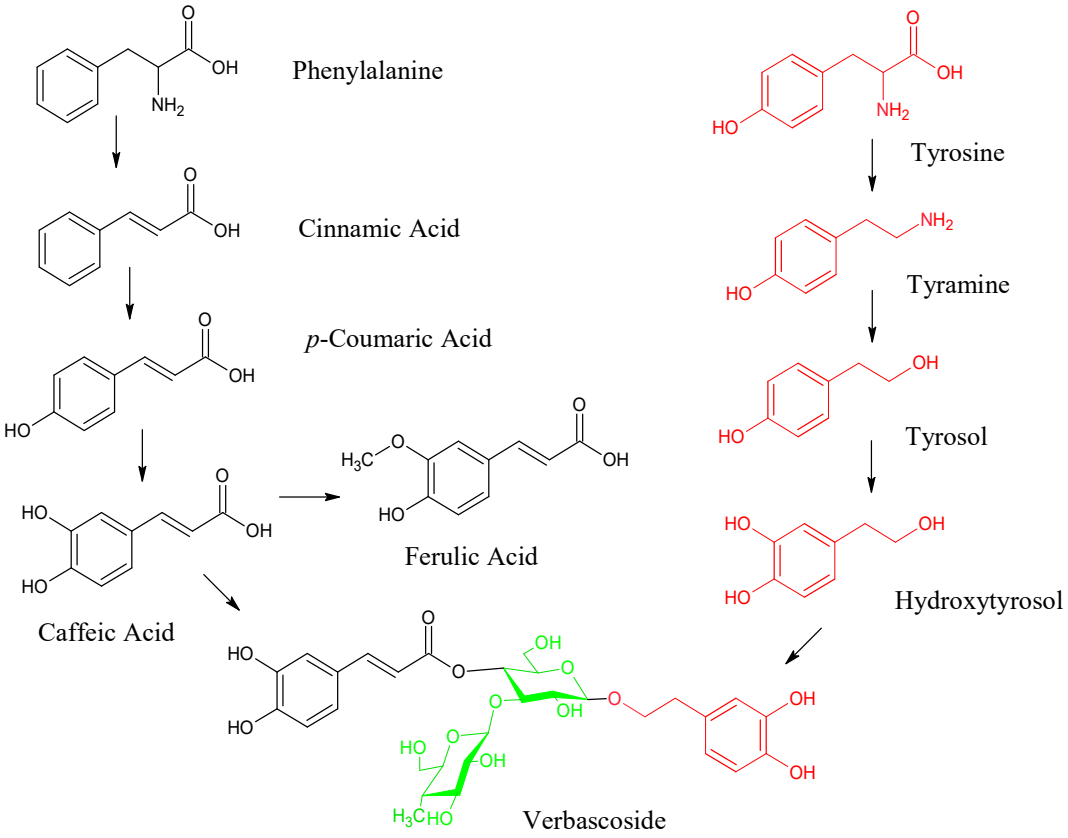


Figure 7.6: Simplified pathway for verbascoside biosynthesis, showing the links between the phenolic acids, simple phenols and verbascoside (modified after Ellis, 1983; Saimaru and Orihara, 2010; Alipieva *et al.*, 2014).

Apart from ferulic acid, verbascoside was found to be positively correlated (R^2= 0.795 p-value < 0.01) to peak 16 which was assigned as hydroxy-decarboxymethyl oleuropein aglycone. Although the two are not structurally related, it is proposed that since the two molecules share the same hydroxytyrosol side group, the two molecules are dependent on the rate of biosynthesis of hydroxytyrosol. In the case of verbascoside the biosynthesis begins with the generation of phenylalanine and tyrosine precursors by the shikimate pathway. The hydroxytyrosol moiety of verbascoside and hydroxy-decarboxymethyl oleuropein aglycone is biosynthesized from tyrosine either through tyramine and/or dopamine (Ellis, 1983). There was no significant correlation between verbascoside and hydroxy-decarboxymethyl oleuropein aglycone to the amount of free hydroxytyrosol present within the EVOOs phenolic extracts. This is because the presence of free hydroxytyrosol is the result of the degradation of a multitude of compounds which contain hydroxytyrosol as a

moiety and no single molecule can be attribute to be the sole source of hydroxytyrosol within the phenolic fraction.

A significant positive correlation was observed in a number of identified secoiridoids, oleuropein and unidentified peak 24, identified as 3, 4 DHPEA-EDA (R^2= 0.668 p-value 0.021), *p*-HPEA-EDA and hydroxy-decarboxymethyl oleuropein aglycone (R^2= 0.614 p-value 0.034), *p*-HPEA-EA and decarboxymethyl-ligstroside aglycone (R^2 = 0.699 p-value < 0.01). The observed correlation can be explained in terms of both the biosynthesis of phenolic compounds within the olive fruit and in terms of their degradative pathways during olive oil production.

The synthesis of oleosides and ligstroside into oleuropein derivatives involve a hydroxylation step (Damtoft, *et al.,* 1996). This process is quite active at the early stages of fruit maturation, which is dominated by high concentration of oleuropein. During the ripening stages, as well as during crushing and malaxation of olive fruits during oil production, the release of β-glucosidases degrades the glycoside to the aglycones, forming ligstroside and oleuropein aglycones (*p*-HPEA-EA and 3,4-DHPEA- EA) (Ryan *et al.,*2002). Demethyloleuropein derived from the esterase activity may also act as a substrate for endogenous β-glucosidases. The aglycones which are formed are then hydrolyzed by action of a decarboxymethylase, which via a ring opening reaction, forms the dialdehydic form of secoiridoids (*p*-HPEA-EDA and 3,4-DHPEA- EDA) or else the corresponding dihydroxy open ring form. In the presence of an aprotic solvent the dialdehydic form of secoiridoids undergoes keto-enol tautomerism, forming the more stable corresponding aldehydic form. These biosynthetic pathways tentatively justify the positive correlation found for the different secoiridoid forms identified in this study. Similar results were observed by Sánchez de Medina *et al.,* (2015) who showed a strong positive correlation between *p*-HPEA-EA and 3,4-DHPEA-EA (R^2= 0.93) as well as between *p*-HPEA-EDA and 3,4- DHPEA-EDA (R^2 =0.89, p-value). The lower correlation values obtained in this study were attributed to the different methodologies.

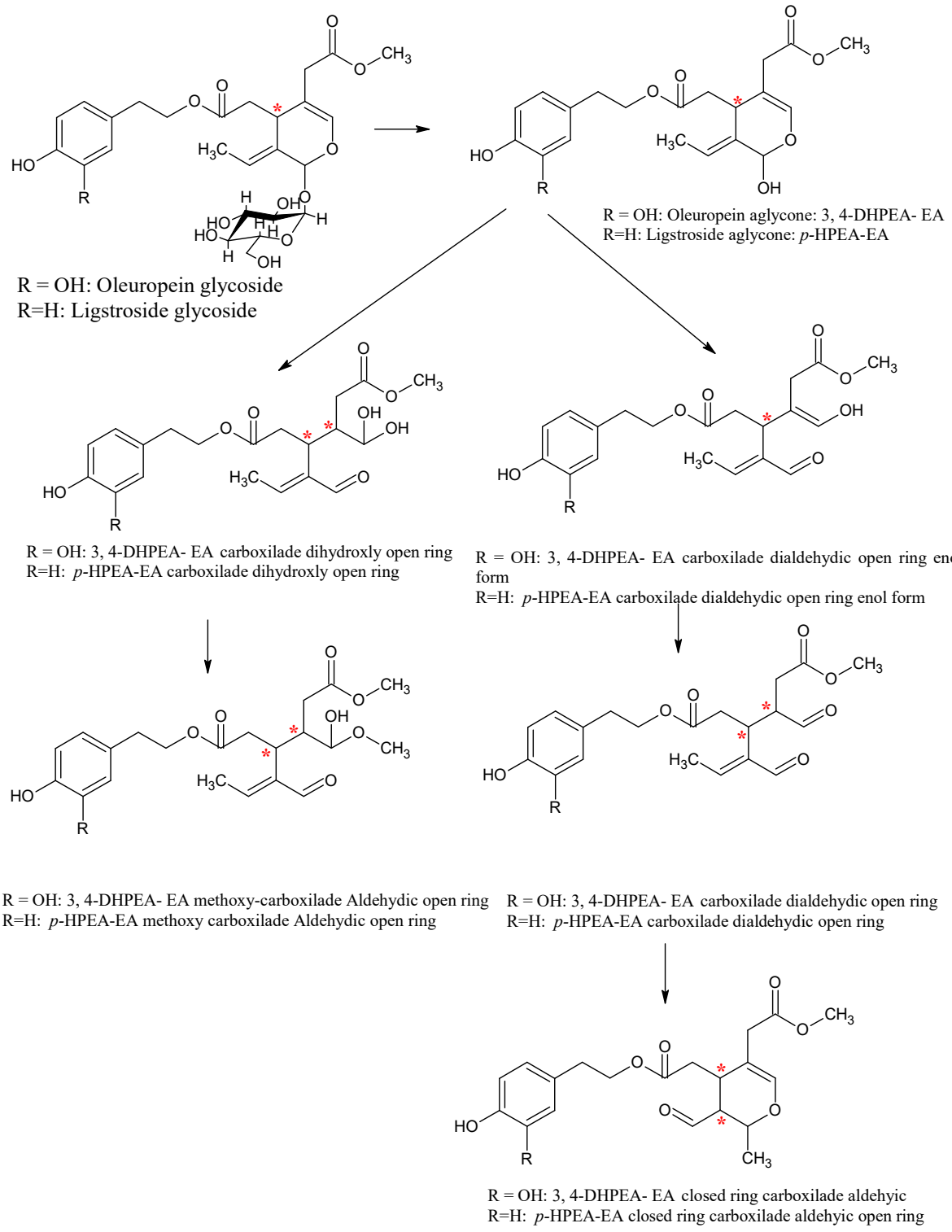


Figure 7.7: Synthesis of oleosides and ligstroside into oleuropein and ligstroside derivatives via enzymatic hydrolysis. (*) indicates the presence of chiral carbons within the molecule.

7.4 Chemometric analysis of Phenolic profiles

Since the proposed strategy was based on the application of blind multivariate analysis of chromatographic profiles, areas of chromatographic peaks were not integrated, however assignment of the peaks to its corresponding available standards and relative retention times to those quoted in the literature was carried out as presented in the sections before.

A minimum of 30 resolved peaks were observed including, hydroxytyrosol, 4-hydroxybenzoic acid, tyrosol, 2,3-dihydroxybenzoic acid, caffeic acid, vanillic acid, vanillin, syringic acid, *p*-coumaric acid, ferulic acid, cinnamic acid, elenolic acid, apigenin and luteolin were identified. It is well known that sensorial perception of olive oils is determined by the phenolic composition. Although this study does not aim to generate a tasting perception panel for the Maltese EVOO's it is known that the 'Bidni' has a more pungent bitter taste whilst the 'Bajda' has a more fruity sweeter taste. According to Kiritsakis, (1998) amongst all phenols, hydroxytyrosol, tyrosol, caffeic acid, coumaric acid, and *p*-hydroxybenzoic acid exhibit the greatest effect on the sensory characteristics of olive oil. Bitterness is closely related to the presence of oleuropein and ligstroside derivatives (Andrewes *et al*. 2003 ; Cerretani *et al*. 2008 ; Gutiérrez-Rosales *et al*. 2003 ; Mateos *et al*. 2004 ; Siliani *et al*. 2006 ; Esti *et al*. 2009). The difference in the sensorial perception of bitterness and astringency is derived from the presence of catecholic groups present within the phenolic molecules and there interaction with the salivary proteins. Nevertheless there are studies (Favati *et al.,* 1995) which suggest that the bitterness and astringency do not always coincide with the total polar phenol content.

Figure 7.9 shows the major identified peaks, however due to the unavailability of standards some of the peaks were not identified. Preliminary assessment of the chromatographic profile already indicated a variation in the amount of different compounds present within EVOOs of the Maltese origin; these are marked with an arrow on Figure 7.9. Although these differences would directly discriminate between the local EVOO's and the foreign it was chosen to subject the data to a subsequent chemometric treatment, in order to attain a more holistic less biased analysis.

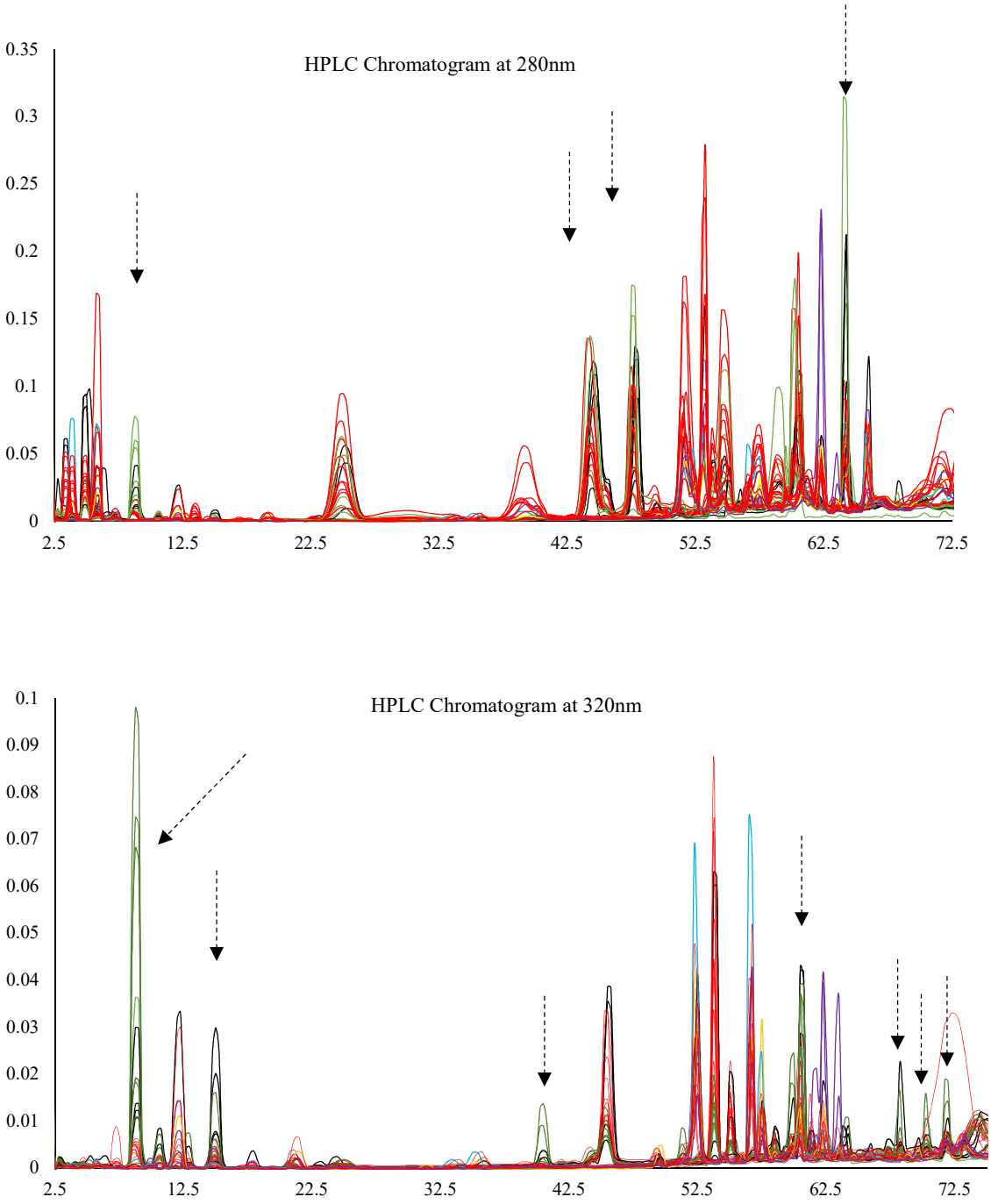


Figure 7.8: HPLC chromatogram of olive oil phenolic fraction observed at 280 nm (Top) and 320 nm (Bottom) different colours represent olive oils derived from different geographical origin. Black = Indigenous Maltese cultivars; Green = Foreign cultivars which are locally grown; Red = Italian cultivars Blue = Greece origin; Purple= French origin and Yellow = Spanish origin

7.4.1 Preliminary data handling

The extracted data for each phenolic profile consisted of 48000 variables. Due to this large number of variables preliminary data treatment was necessary in order to reduce the amount of data, while at the same time retaining the maximum information. The data was subjected to four different kinds of variable reduction techniques. The first one involved the integration of maximum peak heights at specific time intervals across all the cultivars phenolic profile followed by the removal of redundant peaks corresponding to the injection and the post equilibration peaks. This preliminary data treatment reduced the phenolic profile to a more manageable 310 variables. Although this huge change in variable numbers might give the impression of huge data loss, this was not case, in fact this data type reduction technique not only reduced the number of variables and retained the same phenolic profile but also compensated for small changes in the retention times and lowered the baseline which is typically associated with a gradient flow. Thus such data reduction methodology allowed one to reduce the data and synchronise retention times before applying the multivariate chemometric approach.

Data reduction is an essential step in building a pattern recognition technique as it is generally not statistically accepted that the number of variables exceed the number of samples. Since the present work involves a profile study, the number of variables is much larger than the number of objects and therefore, three other different strategies were considered to achieve a reduction in the dimensionality of the data and to remove redundant variables. In the first attempt, SLC-DA was performed in order to select the number of significant most discriminatory variables. A second attempt was done by the application of PLS to select the most suitable variables in order to optimize the discrimination and correct classification. In the third step a combination of the two techniques was employed whereby variables which had both a VIP>0.8 and a p-value < 0.05 in the SLC-DA were used. The data was bucketed into different buckets in order to determine which bucket was the most discriminatory. The phenolic profiles observed at 280 nm and 320 nm were bucketed into five different buckets. The three remaining data reduction techniques were applied on each bucket and then recombined in order to obtain a reduced highly discriminatory phenolic chromatogram.

7.4.2 Phenolic profile preliminary assessment.

Figure 7.10 shows the principal component analysis on the preliminary unbucketed chromatograms obtained at 280 nm and 320 nm. The results obtained showed a spatial clustering resembling the geographical origin of EVOO, the biplot obtained explained 46.7% and 54.3% of the total variability respectively. From the loading score plots, it was shown that retention times between 43-72 min were the most responsible for the observed variation between the phenolic profiles observed at 280 nm. This region of the chromatogram corresponded to peaks related to the least hydrophilic phenolic constituents present in olive oil. These correspond to oleuropein aglycones and ligstroside aglycone and their corresponding derivatives, however due to the complexity and possible isomers, peak identification was not fully possible. In the case of the phenolic chromatograms observed at 320 nm, spatial distribution was dependent on a number of different peaks mainly those obtained at a retention time (R_t) of 9 minutes which had the highest separatory power in the 1st principal component, which explained 39.9% of the total 54.3% variability explained. Peaks obtained between 43-64 minutes seemed to also provide a high degree of variation in the chromatogram observed at 320 nm in both the 1st and the 2nd principal component. The strong peak obtained at R_t 9min was attributed to *p*-coumaric acid; unlike other phenolic acids, which are easily observed at 280 nm, *p*-coumaric acid displays a second absorption maxima 320 nm. This bathochromic shift is attributed to the 3-methoxy and 4-hydroxy group. The compounds present between R_t 43-64 min correspond to flavonoids mainly luteolin, apigenin and their corresponding glycosides, together with lignans, mainly pinoresinol and acetoxypinoresinol, and other oleuropein and ligstroside derivatives including 3, 4 DHPEA-EDA and their corresponding glycoside.

The application of canonical discriminant analysis showed that use of this supervised technique enabled the complete discrimination between the EVOO's of different geographical origin. The X-Y fit on the scores obtained from the 1st canonical function which corresponded to 68.7% and 97.7% of the total variability explained for the phenolic chromatograms observed at 280 and 320 nm respectively, showed that in the former case the EVOOs derived from foreign cultivars were completely discriminated from those cultivars which are locally grown, however at the 1st canonical function the chromatograms obtained at 280 nm did not manage to fully

discriminate between the indigenous and the locally grown cultivars. On the other hand the high % variability explained in the chromatograms observed at 320 nm is reflected in its ability to fully discriminate between the three different classes.

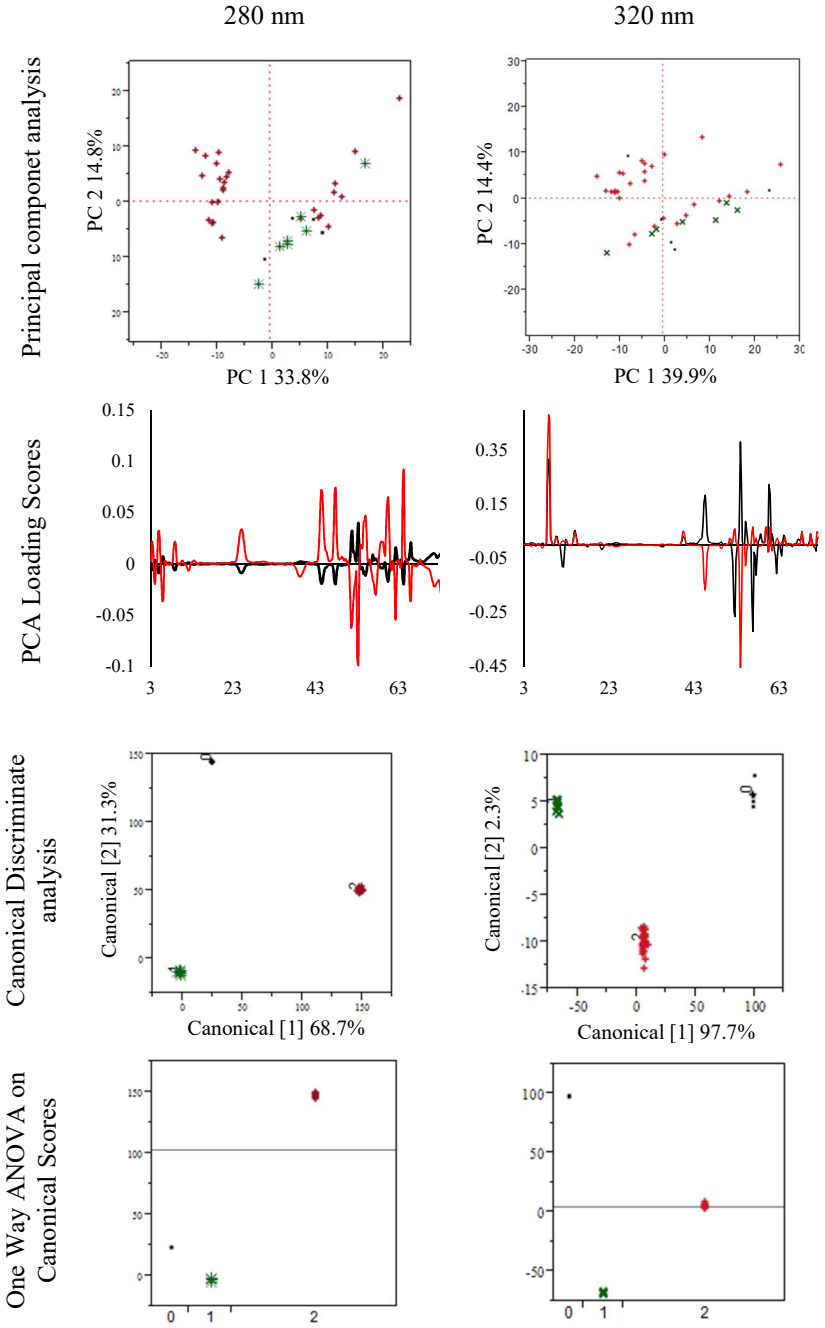


Figure 7.9: Preliminary chemometric analysis of phenolic profiles obtained from HPLC chromatogram at 280 nm (Left column) and 320 nm (Right column) different marker and colours represent olive oils derived from different geographical origin. Black (•) = Indigenous Maltese cultivars; Green (*) = imported cultivars; Red (+) = foreign. Summary of the chemometric analysis include PCA biplot (1st row) PCA variable score plot (2nd row), CDA biplot (3rd row) and One way ANOVA plot on the canonical scores (4th row), highlighting clustering and discrimination of the different monocultivar olive oils based on their phenolic profile.

Figure 7.10 shows the hierarchal cluster analysis done on the chromatograms, without any previous variable selection. Although clusters were observed these did not reflect the geographical origins of the EVOOs and thus a variable selection procedure was carried out. Variable selection procedures enable the determination of the most discriminate variables, those which are able improve the classification which resembles the geographical origin.

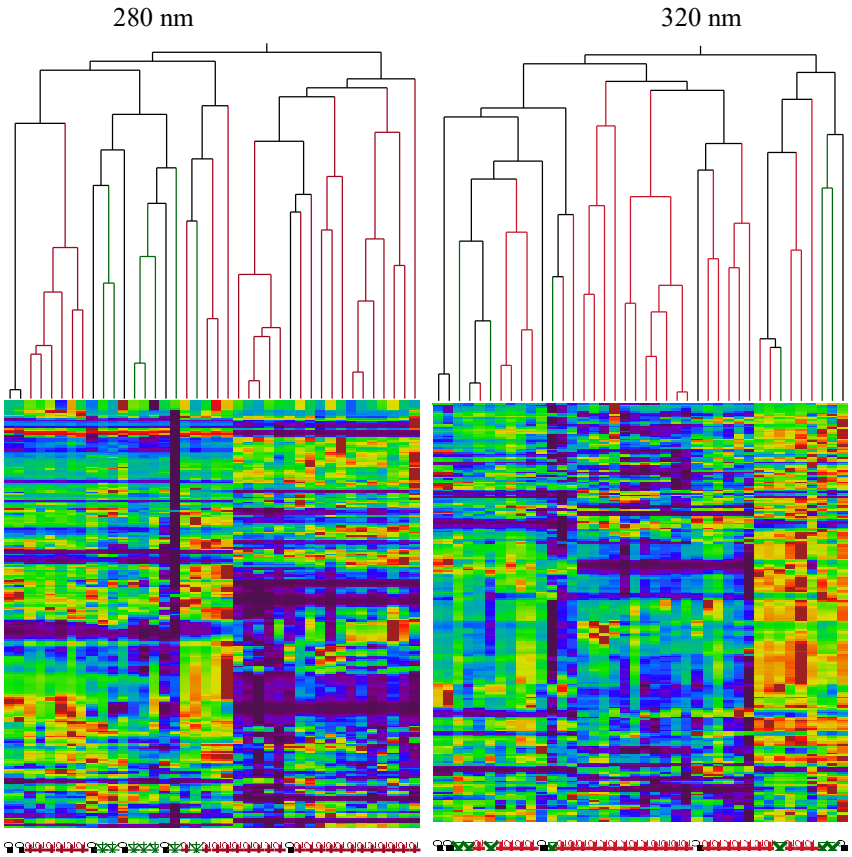


Figure 7.10: Cluster analysis of phenolic profiles obtained from HPLC chromatogram at 280 nm (Left column) and 320 nm (Right column) different colours represent olive oils derived from different geographical origin. Black (●) = Indigenous Maltese cultivars; Green (*) = imported cultivars; Red (+) = foreign. The spectral heatmap represent the intensities at the corresponding wavelength varying across the chromatogram.

Figure 7.11 and Figure 7.12 shows the subdivision of the chromatograms obtained into the different buckets, each bucket was subjected to both supervised and unsupervised chemometric techniques. From the results obtained it was shown that the 3rd bucket which corresponded to peaks observed between 20-42 minutes was the least informative in both the chromatograms observed at 280 nm and 320 nm. The PCA obtained showed very little spatial separation between the different samples furthermore the X-Y fit score plot of the 1st canonical function did not manage to fully separate between the indigenous cultivars and the imported cultivars. In the case of the 5th bucket for the chromatograms observed at 280 nm, the X-Y fit score plot of the 1st canonical function did not manage to fully separate between the imported cultivars and the foreign cultivars however the PCA showed distinct spatial clustering resembling the geographical origin of EVOOs. From the results obtained using both the PCA and CDA, it was shown that 4th 1st and 5th bucket correspondingly seem to be more informative buckets for the clustering and discrimination of samples, based on the phenolic profiles observed at 280 nm. These data buckets were chosen on the bases of the % variability explained in 1st and 2nd PC, 1st canonical function and the low Wilk's lambda value.

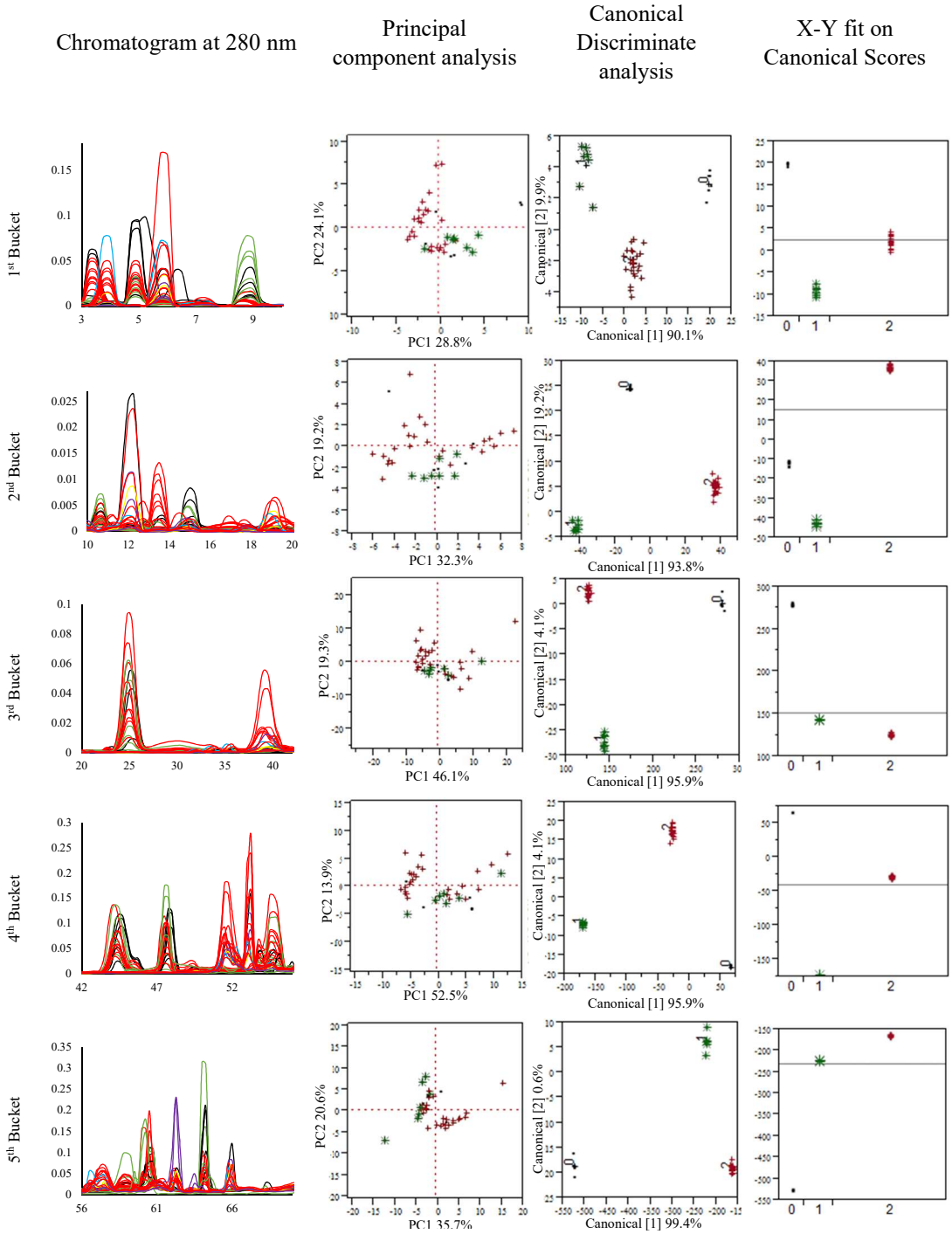


Figure 7.11: The HPLC chromatogram at 280nm bucketed into 5 data bucket (rows), each data bucket was subjected to individual chemometric analysis in order to determine which part of the chromatogram offered the highest discriminatory power. PCA (2nd column), SL-CDA (3rd column) and X-Y Fit on Canonical Scores obtained from the 1st canonical function (4th column) were carried out on the individual buckets. Indigenous Maltese cultivars (Black •); imported cultivars (Green *); Foreign cultivars (Red +).

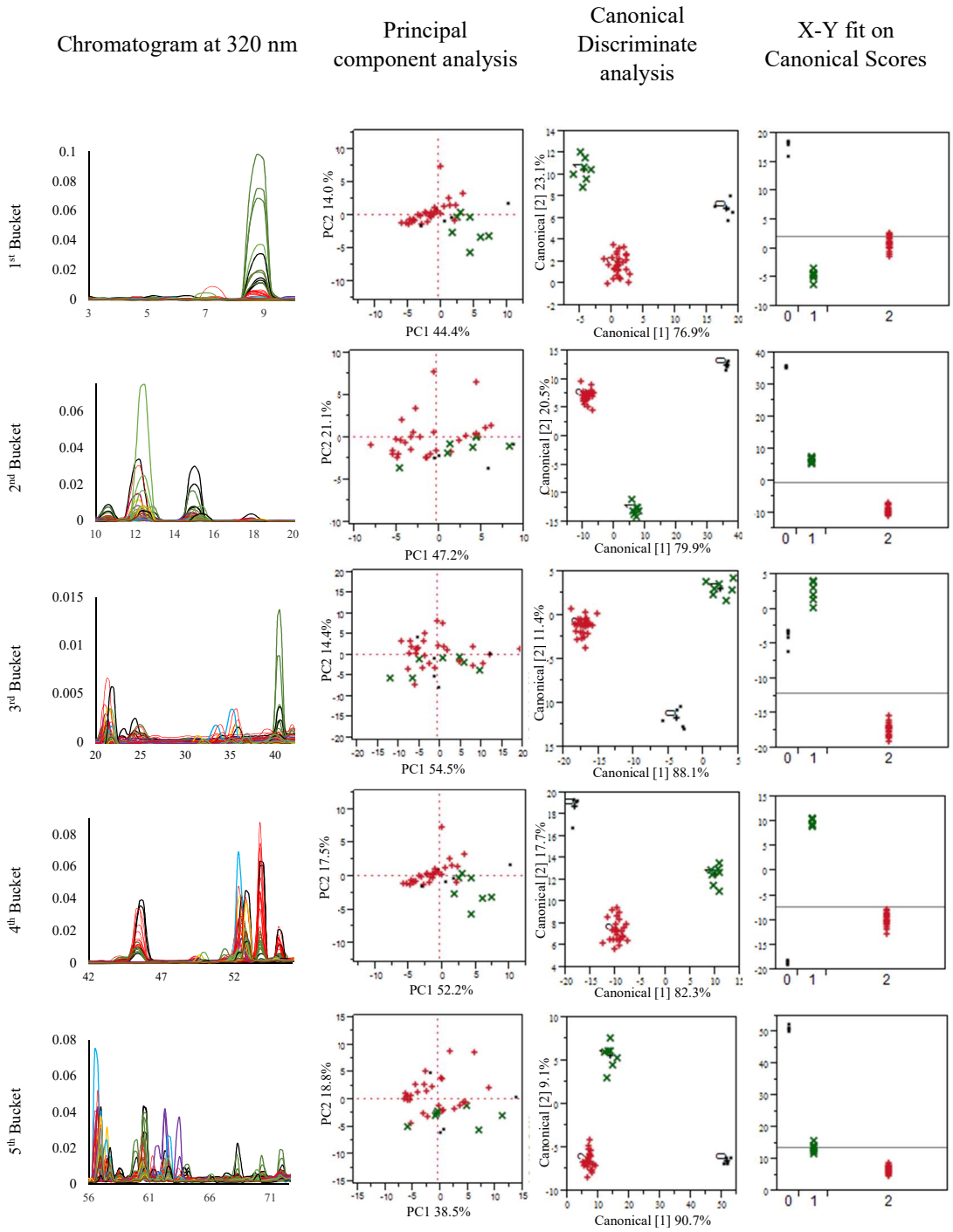


Figure 7.12: The HPLC chromatogram at 280 nm bucketed into 5 data bucket (rows), each data bucket was subjected to individual chemometric analysis in order to determine which part of the chromatogram offered the highest discriminatory power. PCA (2nd column), SLC-DA (3rd column) and X-Y Fit on Canonical Scores obtained from the 1st canonical function (4th column) were carried out on the individual buckets. Indigenous Maltese cultivars (Black •); imported cultivars (Green *); Foreign cultivars (Red +).

Comparing the data buckets of the phenolic profiles observed at 320 nm, all of the five buckets managed to explain a higher % variability in the 1st and 2nd principal component. However once again it was shown that the 3rd bucket was the least informative as very little spatial clustering was observed, and furthermore it was the only data bucket which at 320 nm did not manage to fully discriminate between the local indigenous and imported cultivars at the 1st canonical function. The results suggest that whilst the rest of the data bucket are highly discriminatory in the determination of geographical origin the 3rd bucket contains phenolic compounds which are either present in very low concentration or whose presence does not reflect the geographical location and tend to be more cultivar selective, however such a hypothesis requires further testing.

7.4.3 Application of PLS

The application of PLS analysis was used in order to determine a regression model built using the phenolic chromatograms and the geographical origin of the EVOOs. PLS was applied to the individual buckets and to the whole chromatogram, using two different validation methods. The first validation method, the leave one out, was integrated within the software used, the second cross-validation method was done externally, whereby 40% of the observations were forcefully omitted. The PLS model was built on the remaining 60% of the observations and later cross-validated against the omitted data. The overall % misclassified for both the cross-validation are represented in Table 7.2. In the PLS model, goodness of fit was also assessed in terms of the number of factors extracted, the % of variability explained in terms of X and Y and the predicted root mean square error (PRESS). The ideal model should have a small number of extracted factors, and a high % of variability explained coupled with a low PRESS, however model overfitting needs to be assessed. In this study model overfitting was avoided by the use of two different cross-validation techniques.

Table 7.2: PLS compiled statistical data obtained from the analysis of the whole chromatogram obtained at 280 nm (top) and 320 nm (bottom) using two validation methods; indicating the number of extracted factors in the PLS which had the lowest Root Mean PRESS (predicted residual sum of squares), the % variation and the % misclassified observations.

Internal Validation Whole Data Set at 280nm					
Data Bucket	No. Factors	% X	% Y	PRESS	% Misclassified
1st 3-10 min	2	42.85	58.14	0.902	17.5
2nd 10-20 min	3	49.95	46.99	0.942	20
3rd 20-42 min	7	89.94	68.34	0.895	10
4th 42-56 min	2	65.33	32.7	0.977	30
5th 56-70 min	1	34.05	30.71	0.921	30
6th 70-80 min	2	59.28	43.1	0.873	22.5
Full	2	43.96	55.66	0.817	9.76*
External Validation Whole Data Set at 280nm					
Data Bucket	No. Factors	% X	% Y	PRESS	% Misclassified
1st 3-10 min	2	44.31	64.86	0.81	17.5
2nd 10-20 min	3	38.75	53.14	0.941	20
3rd 20-42 min	4	76.57	57.92	0.898	10
4th 42-56 min	2	65.96	35.13	0.963	30
5th 56-70 min	3	60.39	55.65	0.884	30
6th 70-80 min	8	94.05	84.67	0.717	22.5
Full	4	58.96	84.25	0.766	7.50*

External Validation Whole Data Set at 320nm					
Data Bucket	No. Factors	% X	% Y	PRESS	% Misclassified
1st 3-10 min	1	42.25	49.22	0.788	20
2nd 10-20 min	10	95.41	49.22	0.909	0.0*
3rd 20-42 min	14	98.16	89.16	0.876	0.0*
4th 42-56 min	6	86.55	81.59	0.763	8.9
5th 56-72 min	1	30.22	30.5	0.907	20
Full	5	67.84	86.31	0.741	2.22*
External Validation Whole Data Set at 320nm					
Data Bucket	No. Factors	% X	% Y	PRESS	% Misclassified
1st 3-10 min	1	38.47	53.32	0.753	20
2nd 10-20 min	13	99.11	94.23	1.018	6.67*
3rd 20-42 min	1	43.02	15.86	1.009	22.22
4th 42-56 min	5	84.73	84.86	0.652	6.67*
5th 56-72 min	1	84.72	29.9	0.943	17.77
Full	5	68.08	91.56	0.835	4.44*

Analysis of the results obtained from Table 7.2 show that data buckets obtained from the phenolic chromatogram observed at 280 nm exhibited minor differences in the % of variability explained in terms of X and Y. Furthermore the % classification did not change on changing the cross-validation method. This suggests that the regression models built for each bucket did not show any form of overfitting. Contrary to what was expected it was found that the 3rd bucket had the highest % variability explained and the lowest % misclassification from all the buckets obtained at 280 nm. Similarly the 3rd bucket observed at 320 nm showed 0% misclassification under the internal validation method however, this was a case of model overfitting as % misclassifications drastically increased on using the external cross-validation method. The 2nd and 4th bucket had comparable % misclassifications to those obtained on analysing the full chromatogram at 320 nm under both cross-validation methods with the 2nd bucket being the most promising. However, from the results obtained none of the buckets seemed to be more informative than the whole chromatogram, neither those obtained at 280 and 320 nm, irrelevant of the cross-validation method used. This suggests that data buckets for the analysis of phenolic profiles is not justified, and the whole chromatogram must be taken in consideration. Furthermore this suggests that it is not just one peak or a small number of peaks which are able to discriminate between EVOOs of different geographical origins but it is the multitude of signals obtained

across the whole chromatogram which enables the best classification and provides the possibility of geographical discrimination.

PLS analysis was repeated once again using only variables which had a VIP larger than 0.8. The results obtained in Table 3.8 showed an improvement in the majority of the buckets, in both the % variability explained in terms of X and Y and in the % of misclassification for both the cross-validation methods. The improvement in the regression analysis was also extended to the whole chromatogram analysis obtained at 280 nm. The adjusted PLS analysis did not show any significant improvement when the whole chromatogram obtained at 320 nm was considered. This suggests that in the latter case PLS analysis without variable selection was sufficient in the identification of EVOO's using the phenolic chromatograms obtained at 320 nm, indicating that the chromatograms obtained at 320 nm are the most informative and require minimal variable selection procedures. Analysis of the individual buckets showed that the 5^{th} bucket observed at 280 nm provided the lowest % of misclassification under the two cross-validation methods used. Similarly the 1^{st} and the 4^{th} buckets observed at 320 nm seemed to be most informative in the discrimination of EVOOs. However, data bucketing with regards to chromatographic data was still not justified as none of the buckets achieved a lower misclassification rate than the whole chromatogram.

Figure 7.13 shows the VIP which were selected in the analysis of the adjusted PLS analysis, using these selected variables the PCA showed a slight improvement in the % of variance explained in the 1^{st} and 2^{nd} principal component. This shows that although the number of variables were reduced the PCA did not change as the same spatial distribution of the clusters was obtained. The increased in the % variance further indicates that the removed variables were not significantly affecting the variation of the EVOOs and consequently these variables were ignored as they were redundant.

Table 7.3: PLS compiled statistical data obtained from the analysis using only the VIP>0.8 chromatogram obtained at 280 nm and at 320 nm using two validation methods; indicating the number of extracted factors in the PLS which had the lowest PRESS, % variation and the % misclassified observations.

Internal Validation Using VIP's 280nm					
Data Bucket	No. Factors	% X	% Y	PRESS	% Misclassified
1st 3-10min	2	54.43	56.21	0.849	17.5
2nd 10-20min	3	59.16	49.68	0.877	20
3rd 20-42min	7	94.37	68.1	0.822	7.5
4th 42-56min	2	68.72	35	0.922	25
5th 56-70min	4	65.31	61.38	1.074	7.5*
6th 70-80min	2	71.5	43.98	0.876	22.5
Full	4	60.45	80.21	0.727	2.5
External Validation Using VIP's 280nm					
Data Bucket	No. Factors	% X	% Y	PRESS	% Misclassified
1st 3-10min	2	53.69	63.8	0.824	17.5
2nd 10-20min	3	55.3	51.94	0.921	17.5
3rd 20-42min	7	95.19	70.53	0.905	12.5
4th 42-56min	2	69.53	36.1	0.938	25
5th 56-70min	4	66.06	64.64	0.957	7.5*
6th 70-80min	2	82.39	67.89	0.747	7.5
Full	4	62.52	83.92	0.703	5*

Internal Validation Using VIP's 320nm					
Data Bucket	No. Factors	% X	% Y	PRESS	% Misclassified
1st 3-10min	9	97.16	83.79	0.861	2.22*
2nd 10-20min	9	96.33	81.93	0.866	2.22
3rd 20-42min	2	68.97	43.49	0.871	20
4th 42-56min	6	87.68	78.22	0.779	8.89
5th 56-72min	4	69.68	64.77	0.84	11.11
Full	5	72.78	86.9	0.651	2.22
External Validation Using VIP's 320nm					
Data Bucket	No. Factors	% X	% Y	PRESS	% Misclassified
1st 3-10min	7	94.07	86.12	0.691	6.66*
2nd 10-20min	1	42.13	30.18	1.05	24.44
3rd 20-42min	1	48.51	17.36	1.011	22.22
4th 42-56min	5	84.73	85.12	0.673	8.88*
5th 56-72min	1	38.95	29.68	0.947	17.77
Full	5	73.17	91.41	0.811	4.44*

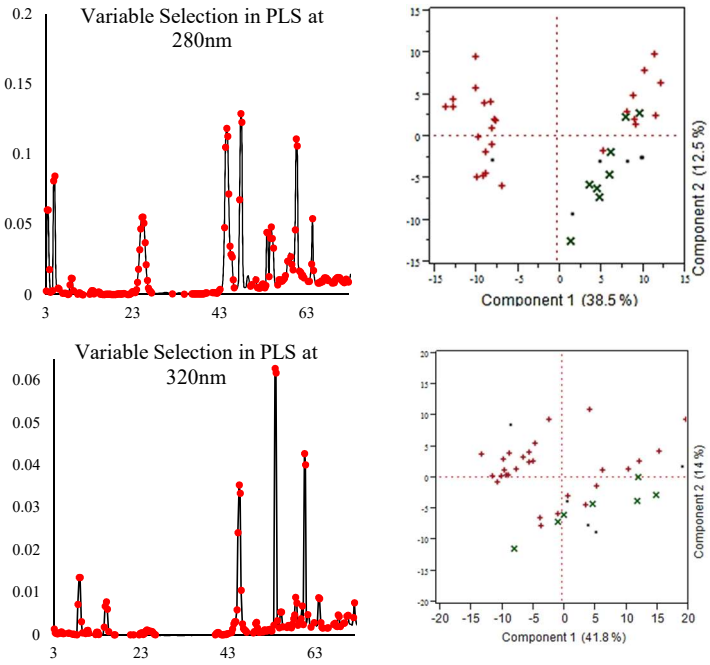


Figure 7.13: (Left) Red points indicate the variables which were had a VIP> 0.8 during the PLS analysis of chromatograms obtained at 280 nm. The selected variables were then subjected to PCA (Right) showing clustering of the different monocultivar olive oils based on the PLS selected variables. Indigenous Maltese cultivars (Black●); imported cultivars (Green *); Foreign cultivars (Red +).

7.4.4 Application of SLC-DA

Since the result obtained from the preliminary variable selection analysis did not justify the use of data buckets for the analysis of phenolic chromatograms, variable selection using SLC-DA was carried out on the full chromatogram.

Figure 7.14 shows the variables which were selected during the SLC-DA. The application of SLC-DA enabled the identification of which Rt were the most discriminate. These Rt were selected on the basis of their ability to minimize the Wilks' lambda, λ_w. In the stepwise algorithm a predictor is selected when the reduction of λ_w produced after its inclusion in the model exceeds the entrance threshold of a test of comparison of variances or F-test. Variables were selected on the bases of F Ratio and Prob>F given, variables which have the lowest p-value for the Prob>F and the highest F-ratio are the first to be selected. The entrance of a new predictor modifies the significance of those predictors, which are already present in the model and those which are not yet included in the model. The process terminates when there are no predictors entering or being eliminated from the model so that the all the remaining variables will have an F Ratio and Prob>F equivalent to 0 and 1 respectively.

Similar to the PLS, VIP scores, the SLC-DA eliminates redundant variables, however PCA analysis revealed a slight reduction in the % of variability explained by the 1st and 2nd principal component. This was attributed to a lower amount of variables which were selected during the SLC-DA causing an inevitable loss in the explained variation. SLC-DA had a pronounced increase in the spatial clustering of samples at 280 nm, but this spatial separation was less prominent in the chromatograms observed at 320 nm. This observation further corroborated the analysis of hierarchal cluster analysis which showed a significant improvement when compared to the cluster analysis obtained without variable selection (Figure 7.11).

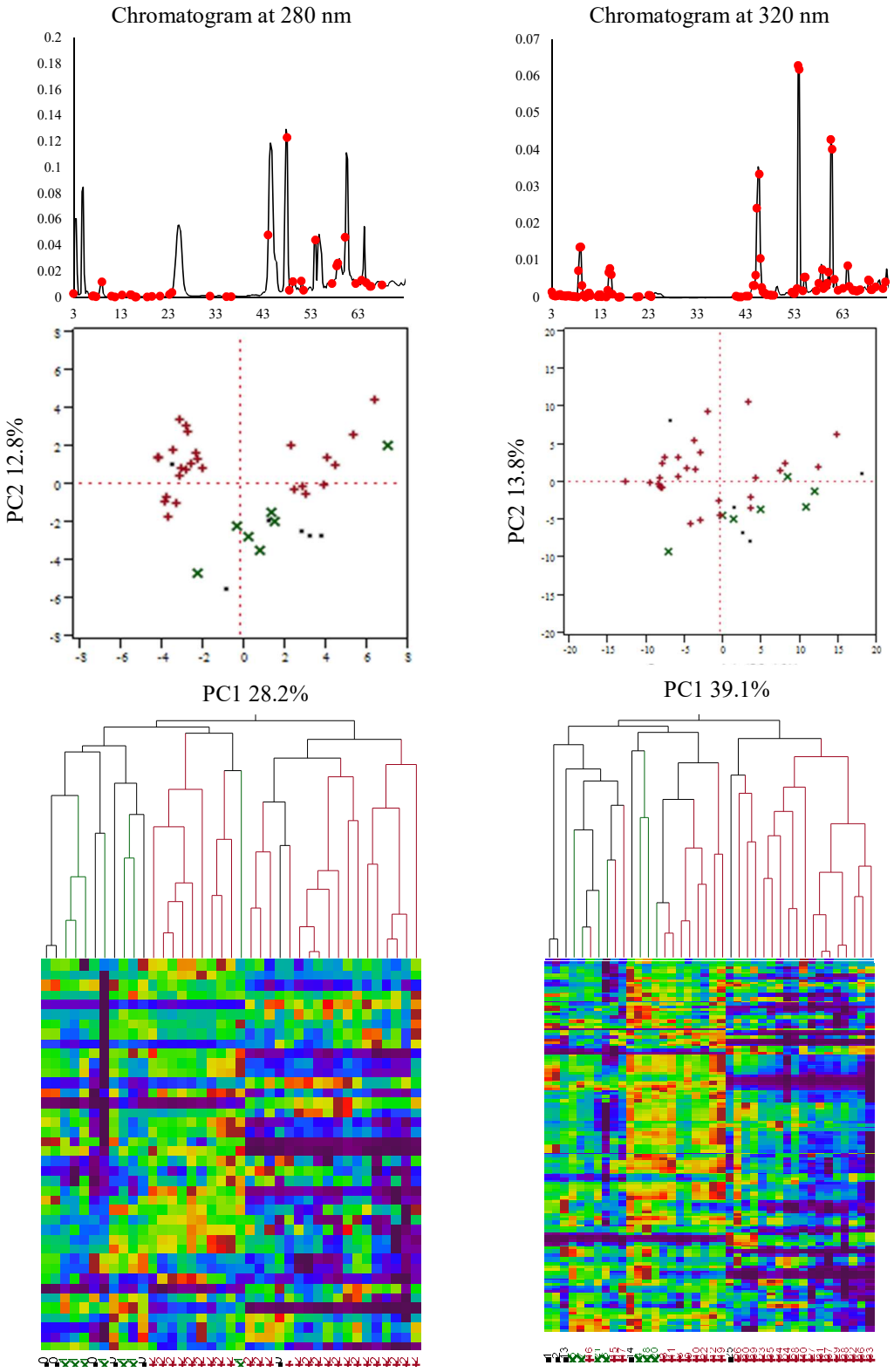


Figure 7.14: Variables selection using SLC-DA (1st row) for the chromatograms obtained at 280 nm (Left) and 320 nm (Right). PCA (2nd row) showing clustering of the different monocultivar olive oils based on the SLC-DA selected variables. The clustering concordant to the origin of olive oils was further confirmed by hierarchical cluster analysis using Ward's minimum variance method

Application of PLS analysis on the variables selected only from the SLC-DA for the phenolic chromatogram observed at 280 nm showed a lower % misclassification, when compared to both the non-pre-treated data set and to the reduced data set containing only VIP>0.8, whilst no significant improvement was found in the phenolic chromatogram observed at 320 nm. These observations further confirm that chromatograms observed at 320 nm are the most informative and require minimal variable selection procedures.

Combining the variable selection process obtained from both the PLS and SLC-DA it was shown that the subsequent application of unsupervised statistical techniques namely PCA and HCA showed a notable improvement both in the spatial separation and clustering resembling the true geographical origin but also in the % of variability explained in terms of 1st and 2nd PC.

In the case of the phenolic chromatograms obtained at 280 nm the PCA managed to explain a total of 54% variability compared to the full chromatogram (46.7%), VIPs (51.0%) only and SLC-DA (41.0%) variables only, furthermore the spatial separation between the different origins was highly improved. However in the case of phenolic chromatograms obtained at 320 nm the application of PLS and SLC-DA common variables did not show any improvement on the % variability explained in the PCA (54.0%) compared to the untreated chromatogram (54.3), VIPs (55.8%) and SLC-DA (52.3%) selected variables only.

HCA on the chromatograms observed at 280 nm showed almost a complete spatial clustering of the EVOO's which resembles their geographical origins. The HCA cluster analysis also shows minor clustering which reflects the cultivar from which the EVOO was derived and the geographical location. These results further confirm that the phenolic profile is the result of a very complex multivariate interaction between genotype and environmental factors (Montedoro and Garofolo, 1984; Lavee and Wodner, 1991). In fact, drastic variability between cultivars was recorded by Baccouri O. *et al.,*2007; Montedoro *et al.,*2007; Tura *et al.,*2007; Baccouri *et al.,*2008; Issaoui *et al.,*2010, similarly by several environmental factors can considerable effect EVOO phenolic profile (Yousfi *et al.,*2006; Baccouri *et al.,*2007; Damak *et al.,*2008;

Gómez-Rico *et al.,*2008). Comparing the cluster analyses for the chromatographic data obtained at 280 nm and 320 nm, the results suggest that compounds which absorb at 280 nm have a greater potential in the discrimination of both the EVOOs producing cultivar and the geographical origin. The application of PLS regression showed no significant improvement in the % misclassification in comparison to the results obtained on using variables selected from PLS and from SLC-DA. This was attributed to the inevitable loss of variables.

Table 7.4: PLS compiled statistical data obtained from the analysis using SLC-DA variables for chromatograms obtained 280 nm and 320 nm using two validation methods; indicating the number of extracted factors which had the lowest PRESS, % variation and the % misclassified observations. The goodness of fit was determine by the van der Voet T^2 and the Prob > van der Voet T^2

	SLC-DA				PLS and SLC-DA			
	280 nm		320 nm		280 nm		320 nm	
	Internal	External	Internal	External	Internal	External	Internal	External
No. Factors	15	15	15	15	3	3	5	3
% X	91.86	93.52	91.86	93.52	63.99	58.59	72.62	57.46
% Y	98.63	99.52	98.63	99.52	74.39	73.79	87.48	78.85
PRESS	0.826	0.993	0.826	0.993	0.810	0.806	0.618	0.698
VVT2	0.00	4.75	0.00	4.75	5.68	0.00	0.00	0.00
p> VVT2	1.00	0.05	1.00	0.05	0.05	1.00	1.00	1.00
% Misclassified	0.0	2.5	0.0	2.5	2.5	5.0	5.0	2.5

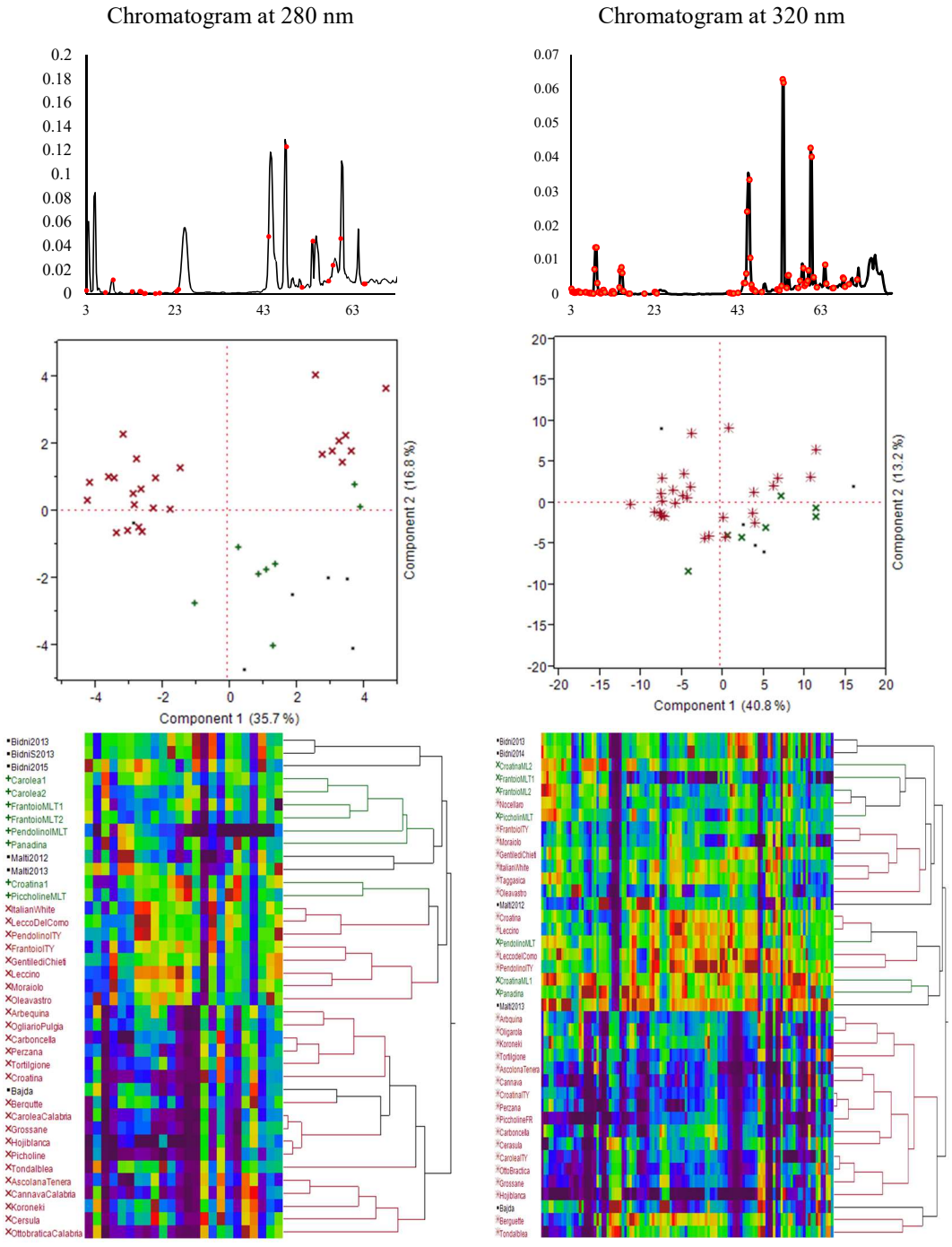


Figure 7.15: Variables selection using common variables selected from both PLS and SLC-DA (1st row) for the chromatograms obtained at 280 nm (Left) and 320 nm (Right). PCA (2nd row) showing clustering of the different monocultivar olive oils based on the SLC-DA selected variables. The clustering concordant to the origin of olive oils was further confirmed by hierarchical cluster analysis using Ward's minimum variance method

8. Application of Fourier Transform Middle Infra-Red Attenuated Total Reflectance (FT-MIR-ATR) analysis for the determination of the geographical origin of olive oil.

Vibrational molecular spectroscopy techniques, including FT-Raman, FT- IR and NIR, are emerging analytical techniques that have shown great potential in the determination of adulterant concentrations (Rohman & Che Man, 2012), fatty acid and triacylglycerols composition (Galtier *et al*., 2008; Inarejos-Carcía *et al*., 2013), oxidised fatty acids (Lerma-García, Simó-Alfonso, Bendini, & Cerretani, 2011), sensory characteristics, phenolic and volatile compounds (Lerma-García *et al*., 2011), geographical and botanical origins (Lerma- Garcia, Ramis-Ramos, Herrero-Martinez, & Simo-Alfonso, 2010) as well as providing insights into the overall quality of olive oil including peroxide value (Bendini, Cerretani, Carrasco-Pancorbo, *et al*., 2007; Bendini, Cerretani, Di virgilio, *et al*., 2007) and acidity (Lerma-García *et al*., 2011). These techniques tend to be used in conjunction with chemometric procedures which enable both the identification and quantification of adulterants and quality parameters, but also enable classification of virgin olive oils according to their geographical and botanical origins (Baeten *et al*., 1996, 1998 ; Marigheto *et al*., 1998 ; Bertran *et al*., 2000 ; Mannina *et al*., 2001 ; Downey *et al*., 2003 ; Marquez *et al*., 2005 ; Yang *et al*., 2005 ; Wang *et al*., 2006). Thus, FTIR spectroscopy combined with chemometric methods offers non-destructive techniques which enable direct and fast determination of different properties of EVOO's in particular when it is used in conjunction with Attenuated Total Reflectance (ATR), which enables direct measurement without any form of sample pre-treatment (Pasquini, 2003).

The aim of this part of the study was to use FTIR–ATR spectroscopy associated with chemometrics in order to differentiate the Maltese EVOO's from other EVOO's derived from other countries within the Mediterranean region, thus developing a quick, easy and cost-saving verification of the origin of EVOOs from the Maltese islands, which is a key aspect of the path for the application of protected designation of origin. In this study, the spectroscopic data were processed both by a discriminant chemometric tools including PLS, SVM, and LDA but also using modelling chemometric tools such as SIMCA and PNN. Moreover, different forms of signal pretreatment were employed in order to enhance the potential of FTIR as a tool for authentication purposes.

8.1 MIR peak identification

The FTIR spectra of the majority of vegetable oils in the MIR region display a high degree of similarity, since the majority of the plant oils are mainly composed of triacylglycerols together with di- and monoacylglycerols, which constitute up to 90–95% by weight of the oils, the remaining 5% include minor components including waxes, sterols, and plant secondary metabolites. The vast majority of the research carried out throughout the years has been focused on the authentication using FTIR and chemometric tools, to detect and quantify refined seed oils, added fraudulently to EVOOs. There is very small amount of research which involves the use of NIRS for the determination of both the botanic origin and the geographical origin of EVOO's.

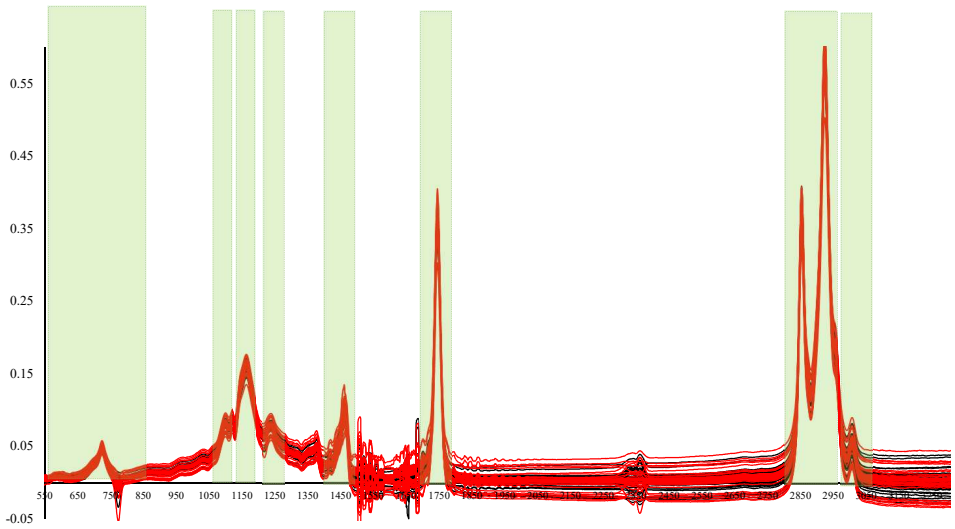


Figure 8.1: The major peaks of interest obtained using FTIR of EVOOs

Figure 8.1 shows a typical EVOO NIR spectrum, the maxima obtained at an absorbance of 3006 $cm^{-1}$is attributed to the stretching vibration of (=C–H) of oleic acid acyl groups and linoleic and linolenic acyl groups. The strong band absorptions observed in the region of 3000-2800 cm^{-1}corresponds to the (-C–H) stretching vibrations of methylene (–CH_2–) and methyl (–CH_3) groups which are observed at frequencies of 2922 and 2853 cm^{-1}, respectively. The bending vibrations of the methylene and methyl groups are observed at 1465 $cm^{-1}$and 1377 cm^{-1},which correspond to C=H scissors deformation vibration and bending vibration of CH_2

groups respectively. The sharp intense peak around 1740 cm^{-1} is attributed to the presence of carbonyl groups which corresponds to the (-C=O) double bond stretching vibration of the ester carbonyl functional group of the triglycerides. The medium intensity peaks observed at 1160.74 cm^{-1} and 1236.86 cm^{-1} is assigned to the vibration of the C-O ester groups and CH_2 groups respectively. Small intensity peaks observed at 1117 cm^{-1} are associated with the stretching vibration of the C-O ester group. The low-intensity peak observed at 722 cm^{-1} correspond to the cis-CH=CH– bending out of plane. (Guillen and Cabo, 1997; Lerma-Garcia *et al*., 2010).

8.3 Application of chemometrics to Middle Infra-Red Spectra

As previously stated the aim of this study was to build reliable classification models for the traceability of EVOOs from Malta by coupling near-infrared spectroscopic techniques and chemometrics. To this purpose, MIR spectra of olive oil samples from the Maltese islands and from other Mediterranean countries were collected as described in Section 2.2.1 and analyzed as described in Section 2.2.6.3. In order to obtain spectral fingerprints corresponding to the origin of the sample, discriminant (PLS, LDA) and modelling (SIMCA, PNN) classification approaches were used and compared.

8.3.1 Unsupervised chemometric techniques – PCA

Different kinds of spectral pretreatments were tested and compared in order to overcome the instrumental limitation and account for scattering and other minor variations which would hinder the performance of the classification models. A total of 12 spectral pretreatment methods were used, in each case, after pretreatment a principal component analysis was carried in order to dimensionally reduce the number of variables into a small set of principal components whilst retaining the information of the larger set. PCA enabled the preliminary identification of which pre-treatment method offered the highest variability and possible clustering. Figure 8.2 shows the different forms of spectral pretreatments employed and the corresponding PCA plot for the first two principal components. From the % variability explained it was found that 5 points smoothing enhanced the variability explained by the 1st principal

component when compared to the ATR correction. This was attributed to the improved signal to noise ratio especially in the 1550-1650 cm^{-1} region, thus this observation justified the use of smoothing prior to other spectral pre-treatment methods. Whilst the other spectral pretreatment methods displayed an improvement in the variability explained after smoothing, quantile normalization (QN), multiplicative scatter correction (MSC), and Standard Normal Variate (SNV) showed a lower % variability when compared to the basic ATR correction and smoothing. In the case of QN, this was expected as, unlike the other pretreatment methods it aims to achieve the same distribution of intensities of all spectra, making it not particularly useful when dealing with spectra of a continuous nature.

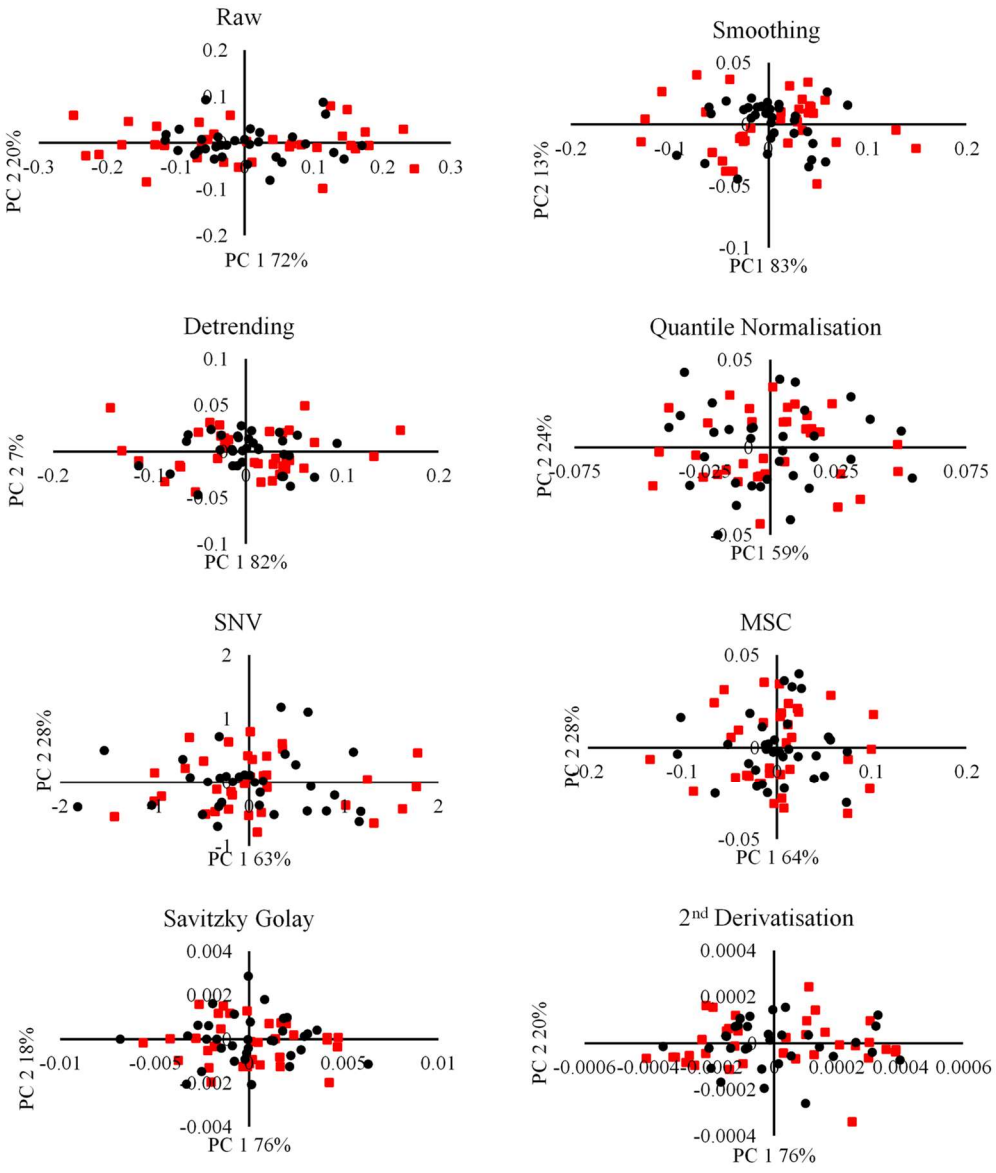


Figure 8.2: FTIR spectral transformations and the corresponding PCA biplot obtained. EVOOs of Maltese origin are represent as black dots whilst non-Maltese EVOOs are represented as red squares.

The next step was to divide the whole dataset into training and test sets (the former to build the model, the latter to validate it). In order to preserve the diversity in the training and test sets and to account for the fact that different pretreatments had to be tested a unique sample splitting scheme was required. The following method was adopted in order to cover as such variation in the two sets and at the same time being able to compare the outcomes after the different pretreatments. The Maltese and the non-Maltese samples were grouped in an ascending way so that the first 35 samples would represent Maltese EVOO's whilst the rest corresponded to non –Maltese

EVOO's. A stratified random sampling method was used in order to exclude 20% of the observations so that they would be retained as the testing set. The remaining 80% of the observation were used as to build the training set. Similar studies using FTIR employ the use of other statistical methods such as the Duplex algorithm as employed by Bevilacqua *et al*. (2012) was used to split the data set. The Duplex algorithm involves the consecutive selection of observations which how the highest dissimilarities according to their Euclidean.

8.3.2 Supervised chemometric techniques – PLS-DA

After splitting the data according to the procedure described above, chemometric classification models were built and tested on all the MIR spectral pretreatment using a PLS regression algorithm using JMP 10 and its inbuilt leave one out cross-validation method (LOOCV). Table 8.1 shows the number of latent variables extracted, the predicted root mean square error and the % variation explained in terms of X and Y for the different spectral pretreatment methods. The % accuracy (correct classification), showed that with the exception of normalization all the other spectral treatments had the same effectiveness in correctly classifying the geographical origin of EVOOs. Another PLS algorithm was used using only 80% of the data and LOOCV, and the obtained regression formula was then used to project the classification of origin on the remaining 20% test subset as shown in Table 8.1. From the results obtained it was observed that the Savitzky Golay, 1st derivative, MSC, OSC, and detrending showed consistently higher explained variation in % Y (the geographical origin), lower PRESS and a higher % accuracy and % predictability. On the other hand normalisation seem to be negatively affecting the amount of variability of the data, in fact from the PCA analysis it was shown that majority of the explained variation was due to the 1st PC and very little variation was explained by the remaining components, suggesting that normalization reducing the overall variation in the data reducing its ability when it comes to the determination of geographical origin of EVOOs.

Table 8.1 PLS-DA analysis using the whole FTIR data. (Top) the results obtained using LOOCV on the training dataset. (Bottom) the results obtained using LOOCV and 20% of the data as the validation group

FTIR Spectroscopy Whole Data internal validation					
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy
Raw	7	97.65	80.10	0.81	100.00
Smoothing	1	99.55	80.34	0.80	98.53
Baseline	15	99.96	91.14	0.74	100.00
Normalized	1	93.84	2.86	1.01	48.53
Quantile normalized	8	89.13	86.42	0.81	98.53
Detrend	15	99.10	99.19	0.54	100.00
Deresolve	15	99.81	90.39	0.77	100.00
SNV	15	99.81	96.14	0.61	100.00
MSC	15	99.84	97.23	0.59	100.00
OSC	15	99.29	98.32	0.62	100.00
SG	3	36.46	88.44	0.68	100.00
1^{st}	3	36.46	88.44	0.68	100.00
2^{nd}	3	35.87	91.27	0.83	100.00

FTIR Spectroscopy Whole Data internal validation						
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy	% Predictability
Raw	5	96.81	78.09	0.79	92.65	100.00
Smoothing	15	99.75	69.35	0.73	91.18	100.00
Baseline	8	99.64	71.49	0.75	89.71	84.62
Normalized	1	96.04	2.48	1.02	50.00	61.54
Quantile normalized	5	83.74	68.99	0.88	85.29	92.31
Detrend	15	99.22	99.62	0.49	94.12	100.00
Deresolve	15	99.8	95.62	0.71	92.65	100.00
SNV	9	99.42	79.77	0.68	91.18	100.00
MSC	15	99.88	98.45	0.65	94.12	100.00
OSC	15	99.26	99.28	0.56	92.65	100.00
SG	9	56.73	100	0.65	97.06	100.00
1^{st}	9	56.73	100	0.65	97.06	100.00
2^{nd}	13	56.39	100	0.8	92.65	100.00

In order to fully interpret the PLS models obtained, an inspection of the VIP scores was used in order to determine which predictors (variables) are mainly influencing the latent vectors obtained. VIP is an index of how much a single variable contributes to the bilinear model and it is scaled in such a way that indices having VIP larger than 0.8 are considered to be significantly contributing to discrimination. VIP scores > 0.8 for the PLS models built on the different pretreated MIR data are reported in the first column of Figure 8.3. As shown in the figure, the VIP > 0.8 identified relevant features in the spectra, particularly, stretching vibration of (=C–H) of acyl groups 3006 cm^{-1}, (-C=O) double bond stretching (around 1700 cm^{-1}) and C-H bending in the fingerprint region (650–750 cm^{-1}) appear to be the regions contributing the most to the bilinear model. Additionally, C-H stretching (2800–3100 cm^{-1}) and C-O single bond stretching (1100 cm^{-1}) also show a VIP value significantly larger than 0.8. The next step was to build another PLS model this time using only variables which had a VIP score > 0.8. Table 8.2 shows the results obtained on using the adjusted PLS model. Comparing the models obtained using variable selection to that previously obtained without any variable selection, no noticeable differences were observed in % accuracy and predictability of the model. Nonetheless a lower PRESS was observed for the majority of the pretreatments. This observation suggests that in the case of ATR-FTIR data no need of extensive variable selection is required to obtain a very good classification method with the use of PLS models.

Table 8.2 PLS-DA analysis using the VIP>0.8 dataset. (Top) the results obtained using LOOCV on the training dataset. (Bottom) the results obtained using LOOCV and 20% of the data as the validation group

FTIR Spectroscopy VIP > 0.8 Data internal validation					
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy
Raw	5	91.26	75.16	0.86	97.06
Smoothing	13	99.07	86.92	0.78	100.00
Baseline	2	98.55	3.77	1	97.06
Normalized	11	99.72	73.56	0.76	57.35
Quantile normalized	8	90.11	86.57	0.74	98.53
Detrend	14	98.67	96.86	0.58	100.00
Deresolve	15	99.49	88.82	0.73	100.00
SNV	15	99.59	93.66	0.59	100.00
MSC	15	99.73	93.27	0.61	100.00
OSC	15	98.43	99.89	0.13	100.00
SG	5	46.68	96.62	0.46	100.00
1st	5	46.68	96.62	0.46	100.00
2nd	8	61.24	99.96	0.53	100.00

FTIR Spectroscopy VIP > 0.8 Data internal validation						
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy	% Predictability
Raw	5	91.79	78.63	0.85	97.06	100.00
Smoothing	15	99.29	95.98	0.72	91.18	100.00
Baseline	1	97.16	2.44	1.01	88.24	84.62
Normalized	5	82.33	72.38	0.79	50.00	61.54
Quantile normalized	10	99.72	72.45	0.76	86.76	92.31
Detrend	14	98.72	98.67	0.46	94.12	100.00
Deresolve	15	99.53	93.69	0.72	92.65	100.00
SNV	13	99.59	90.36	0.66	91.18	100.00
MSC	13	99.67	92.45	0.66	92.65	100.00
OSC	15	97.85	99.92	0.19	100.00	100.00
SG	5	43.85	98.65	0.43	98.53	100.00
1st	5	43.95	98.38	0.43	97.06	100.00
2nd	8	61.47	99.99	0.53	98.53	100.00

8.3.3 Modelling Chemometric techniques – SIMCA

The classification analysis of EVOOs FTIR data was then repeated using a modelling approach based on the SIMCA algorithm. The latter is a class modelling algorithm that allows the analysis one class at a time. For SIMCA analysis two PCA were built for each spectral pretreatment one for the EVOOs of Maltese origin and the other one for the EVOOs of non-Maltese origin. In each case, the optimal number of principal components were chosen in order to obtain optimal model complexity in 10-fold row-wise cross-validation. The use of a two-stage model, one for each category, allowed the comparison between the two categories, thus this allowed us to check whether samples are accepted by one, both or none of the modelled classes. The output of SIMCA analysis was assessed by the use of Coomans plot shown in Figure 8.3.

A Coomans plot takes the form of a graph where the two axes represent the distance of the samples to each of two class models. The horizontal and vertical lines corresponding to the threshold distances also known as the significance limit can be adjusted up depending on the sensitivity required. One of the major disadvantages of using SIMCA is that one has to set a confidence level, α. If the data are normally distributed, α % (e.g. 5%) of objects belonging to the class will be considered as not belonging to it. For this experiment given that the FTIR data was highly similar and no significant difference in variance was observed (PCA analysis revealed no significant clustering resembling the origin of EVOOs) the significance limit was increased up to 25% rather than the default 5%. These lines (blue) of significance divided the Coomans plot into four different regions: the uppermost left and the lowermost right will correspond to unmistakeable acceptance by a single category model, the lowermost left to acceptance by both classes while the uppermost right to rejection by both category models. From a preliminary survey of the Coomans plot outcome at the 25% level of significance, very few samples showed an unambiguous acceptance by a single category. The SIMCA analysis classified the majority of the EVOO's in the lowermost left part suggesting that the samples were accepted by both classes and thus failing to discriminate between the two classes. Thus a diagonal (green) line bisecting the plot corresponding to discriminant classification boundary was built in order to represent a new significant boundary so that all the samples lying above are classified as being Maltese, while samples lying below are predicted as from other origins.

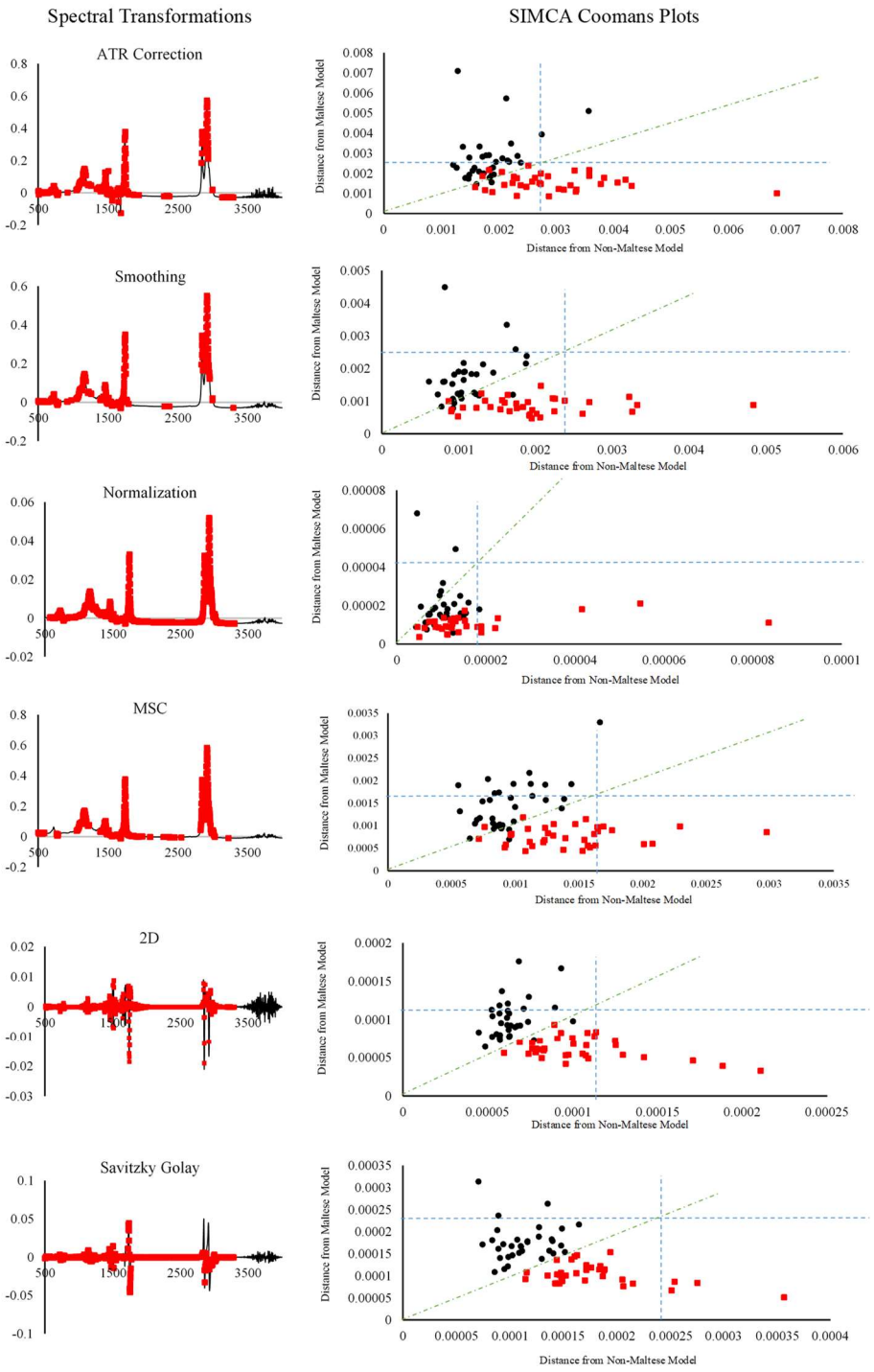


Figure 8.3: (Left) illustrates the variables which had a VIP score > 0.8 (Red dots) for the different spectra pretreatments which were selected for an adjusted PLS-DA and SIMCA. (Right) Coomans plot obtained using SIMCA on the selected variables only. The blue dotted lines represent the 25% confidence level whilst the green dotted represent the discriminant classification boundary decision boundary employed.

Table 8.3: Results obtained for SIMCA modelling for the different spectral pretreatments using the only the variables having a VIP > 0.8. The values recorded in the table represent the % sensitivity of the different models towards the two classes using the discriminant classification boundary decision boundary.

Application of SIMCA on the VIP>0.8 Data Set			
Pre-treatment	%Sensitivity Maltese EVOOs	% Sensitivity Non-Maltese EVOOs	Overall % Accuracy
Raw	93.94	88.57	89.66
Smoothing	72.73	97.14	82.76
Baseline	81.82	97.14	87.93
Detrend	78.79	97.14	86.21
Deresolve	84.85	91.43	86.21
Normalized	18.18	100.00	53.45
Quantile normalized	93.94	88.57	89.66
MSC	90.91	94.29	91.38
SNV	81.82	91.43	84.48
OSC	93.94	91.43	91.38
1^{st}	96.97	100.00	98.28
2^{nd}	93.94	97.14	94.83
SG	100.00	100.00	100.00

The results shown in Table 8.3 summarize the classification obtained on using SIMCA model using the discriminant classification boundary as the major divider as the boundary. Similar to what was observed in the PLS-DA the different pretreatment methods have a great effect on the sensitivity of the model. Spectra pretreated using MSC, 1^{st}, 2^{nd} and Savitzky Golay derivation showed a consistently higher sensitivity towards the Maltese and non-Maltese EVOOs with an overall % accuracy of 91.38, 98.28, 94.83, and 100.00 % respectively. As observed in the PLS analysis spectra, pretreatment using a normalisation procedure showed a very low sensitivity towards the Maltese EVOOs (18.18%) as the majority of the Maltese samples were classified with the foreign EVOOs, thus although a 100% sensitivity is obtained with regards to the non-Maltese samples this value does not take in consideration the Maltese EVOOs classified in the same region, a more realistic approach is to take the overall % accuracy 53.45%. This suggests that the model obtained seemed to be overfitted and thus it is unable to fully discriminate fully the two classes. Comparing the results obtained from the two classification models, it was shown that in cross-validation methods, the sensitivity obtained from SIMCA model was significantly lower for many of the pretreatments considered when compared to PLS-DA counterparts. These results corroborate the result obtained by Bevilacqua *et al*., (2012) whereby it was

shown that SIMCA and PLS-DA models have different sensitives in classifying the Sabine PDO EVOOs, with SIMCA showing a lower sensitivity. The difference between the two model sensitivity was attributed to the fact that the two methods are affected in a different way by the shape of class distributions and the presence of high leverage points, characteristics that, in turn, can be severely affected by the chosen pretreatment (Bevilacqua *et al*., 2012). However unlike the study carried out by Bevilacqua *et al*., (2012), the trend in the different kind of spectral pretreatments was conserved when on a comparison between the SIMCA and PLS-DA.

In order to obtain a more robust method of classification with the use of a smaller number of variables, the VIP data set obtained from the previous analysis was subjected to a stepwise linear canonical discriminate analysis SLC-DA. SLC-DA was performed on the MIR data from all the pretreatment methods in order to extract only a small amount of highly discriminate variables which would enable an easier and faster discrimination between the origins of EVOOs. This strategy involved a substantial reduction of the dimensionality of the data in such a way that only the variables shown in Figure 8.4 were retained. In order to further reduce the number of variables selected from the SLC-DA analysis, a minimum of 14 variables was selected in order to carry out a conventional LDA. During the SLC-DA the variables were chosen by applying a forward stepwise variable selection algorithm using JMP 10 using a Wilks' Lambda as a selection criterion and an F-statistic factor to determine the significance of the changes in Lambda when the influence of a new variable is evaluated. The most significant variables were then extracted and their canonical scoring coefficients were plotted as shown in Figure 8.5. The main advantage of using SLC-DA over the conventional LDA is the ability to perform a feature selection. Regarding this fact, only those variables which helped to improve classification performance were used whereas variables without discriminant information were discarded. Figure 8.4 shows that the variables selected during SLC-DA for all spectral pretreatments were mainly concentrated in the 600-500 cm^{-1} range which corresponds to the C-H bending in the fingerprint region. Shouldering peaks next to the stretching vibration of (=C–H) of acyl groups at 3006 cm^{-1} and (-C=O) double bond stretching (around 1700 cm^{-1}) also appear to be the regions contributing the bilinear model.

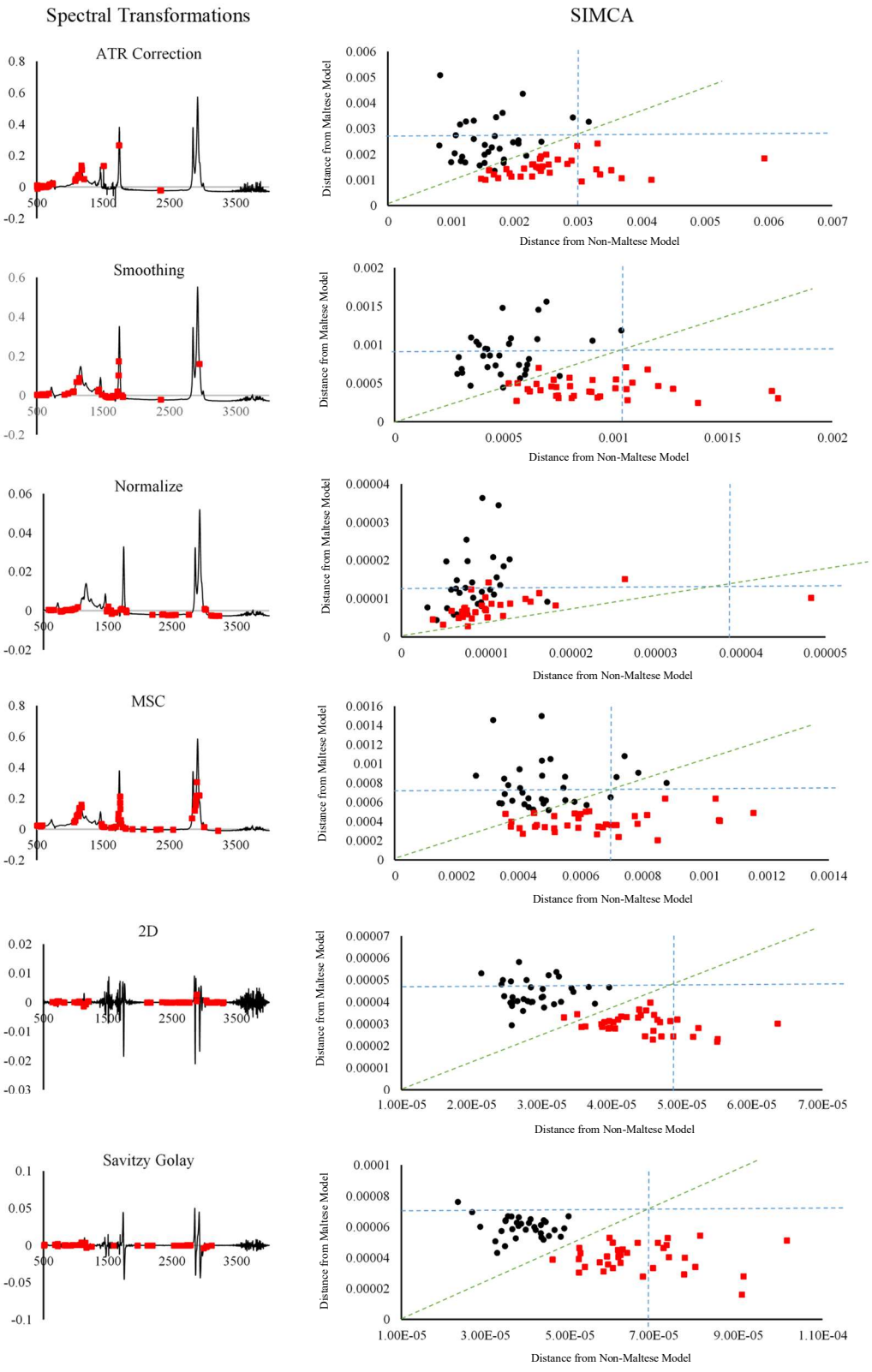


Figure 8.4: (Left) illustrates the variables which were selected in the SLC-DA and had a VIP score > 0.8 (Red dots) for the different spectra pretreatments which were selected for an adjusted PLS-DA and SIMCA. (Right) Coomans plot obtained using SIMCA on the selected variables only. The blue dotted lines represent the 25% confidence level whilst the green dotted represent the discriminant classification boundary decision boundary employed.

Table 8.4: PLS-DA analysis using common variables selected from both SLC-DA and having a VIP>0.8. (Top) the results obtained using LOOCV on the training dataset. (Bottom) the results obtained using LOOCV and 20% of the data as the validation group

FTIR Spectroscopy VIP > 0.8 & SLC-DA Data internal validation					
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy
Raw	10	88.26	86.08	0.89	100.00
Smoothing	15	99.52	90.06	0.63	100.00
Baseline	15	96.17	97.80	0.37	100.00
Normalized	1	89.82	3.14	1.01	52.94
Quantile normalized	14	99.90	81.72	0.67	100.00
Detrend	15	99.42	96.26	0.44	100.00
Deresolve	15	99.75	90.44	0.55	100.00
SNV	15	99.84	93.29	0.57	100.00
MSC	13	99.88	92.13	0.48	100.00
OSC	15	98.84	100.00	0.01	100.00
SG	15	66.76	99.88	0.10	100.00
1st	15	66.76	99.88	0.10	100.00
2nd	15	59.80	99.98	0.05	100.00

FTIR Spectroscopy VIP > 0.8 & SLC-DA Data internal validation						
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy	% Predictability
Raw	4	78.55	66.38	0.99	91.18	84.62
Smoothing	15	99.59	92.92	0.70	95.59	100.00
Baseline	1	91.80	2.46	1.01	91.18	100.00
Normalized	15	96.69	97.83	0.46	52.94	61.54
Quantile normalized	14	99.94	87.47	0.63	98.53	100.00
Detrend	15	99.43	98.05	0.44	98.53	100.00
Deresolve	15	99.78	94.89	0.45	95.59	100.00
SNV	14	99.86	93.55	0.60	97.06	100.00
MSC	15	99.94	96.50	0.46	100.00	100.00
OSC	15	98.88	100.00	0.01	100.00	100.00
SG	15	68.91	99.86	0.15	100.00	100.00
1st	15	68.91	99.86	0.15	100.00	100.00
2nd	15	64.19	99.91	0.16	100.00	100.00

Once the variables were selected for each spectral pretreatment a PLS-DA model was applied in order to determine whether variable selection using the linear method provided ways for a better form of classification. Table 8.4 shows the results obtained from the PLS using the data set composed of variables which had a VIP score > 0.8 and were selected during the SLC-DA analysis. The results show that with the exception of the non-treated raw data and normalization the models obtained from PLS-DA the % accuracy in the training set was in the range of 100-91% whilst 100%

accuracy was obtained in the validation set of corrected, suggesting that a variable selection using the two techniques greatly improves the modelling power. Furthermore, a higher number of latent variables were extracted for the majority of the spectral pretreatments resulting in a higher % of variance explained in terms of X and Y. A marked decrease in the PRESS was also noted with OSC, 1st, 2nd, and Savitzky Golay derivatives having a PRESS lower than 0.2 (Table 8.4), indicating that a very strong and robust classification was obtained even when 20% of the data was omitted and a leave one out cross-validation was applied.

Table 8.5 Results obtained for SIMCA modelling for the different spectral pretreatments using the only the variables selected using SLC-DA and having a VIP > 0.8. The values recorded in the table represent the % sensitivity of the different models towards the two classes using the decision boundary.

Application of SIMCA on the VIP>0.8 & SLC-DA Data Set			
Pre-treatment	% Sensitivity Maltese EVOOs	% Sensitivity Non-Maltese EVOOs	Overall % Accuracy
Raw	96.97	100.00	98.28
Smoothing	96.97	91.43	93.10
Baseline	84.85	91.43	86.21
Detrend	93.94	88.57	89.66
Deresolve	75.76	100.00	86.21
Normalized	18.18	100.00	53.45
Quantile normalized	100.00	94.29	96.55
MSC	90.91	94.29	91.38
SNV	100.00	82.86	89.66
OSC	93.94	94.29	93.10
1st	100.00	100.00	100.00
2nd	100.00	94.29	96.55
SG	100.00	100.00	100.00

The data sets obtained using common variables selected using the SLC-DA algorithm and those having a VIP>0.8 were subjected to SIMCA analysis in order to determine whether the reduction of data will improve the classification models obtained. Table 8.5 shows the results obtained. Although a higher sensitivity was obtained for both classes for the majority of the spectral pretreatments, the results obtained using Savitzky Golay and 1st derivative reached a 100% accuracy and sensitivity towards the two classes. On comparison with the results obtained by using only variables which had a VIP score > 0.8, it was found that spectral pretreatments carried out using ATR correction, 5 point smoothing, baseline correction, SNV and OSC also showed a marked improvement in the classification carried out using SIMCA on variable selected both using SLC-DA and VIP>0.8. These observations

suggest that the two-fold variable selection process employed in this study was not only able to remove large amounts of redundant data but at the same, it preserved the discriminatory power and sensitivity of the variable combinations which are retained.

8.3.4 Supervised chemometric discriminate analysis techniques – LDA

In comparison with SIMCA, LDA avoids the normality problem and confidence interval adjustment making it a more reliable method for classification. The LDA method employs linear decision boundaries, which are defined in order to maximize the ratio of between-class to within-class dispersion (Fisher, 1936). It has been successfully applied to a number of classification problems (Gambarra Neto *et al*., 2009; Gori *et al*., 2012; Riovanto *et al*., 2011; Sinelli *et al*., 2010; Souto *et al*., 2010). When compared with SIMCA and PLS-DA, the LDA method has the disadvantage that the number of training samples must be larger than the number of variables included in the LDA model. In order to fully satisfy this constraint a smaller number of variables were selected based on the standardized scoring coefficients obtained from the SLC-DA. The standardized scoring coefficients of the variable selected during the SLC-DA were obtained and plotted as shown in Figure 8.5. The importance of these coefficients lies in their use to compute canonical scores in terms of the standardized data often referred to as loadings. They are highly informative when it comes to comparing the relative importance in their discriminatory power of the independent variables.

In order to build the LDA, the selected variables obtained in SLC-DA were arranged in ascending order in terms of their scoring coefficients. A smaller set of variables were selected which consisted of 14 variables which corresponded to the 7 most positive and 7 most negative standardized scoring coefficients. An LDA was carried out on the training set using only the small set of variables which were selected. The results obtained for the training samples were visualized on an LDA biplot samples as shown in Figure 8.5 whereby each sample was projected as the scores obtained for the first two discriminate functions.

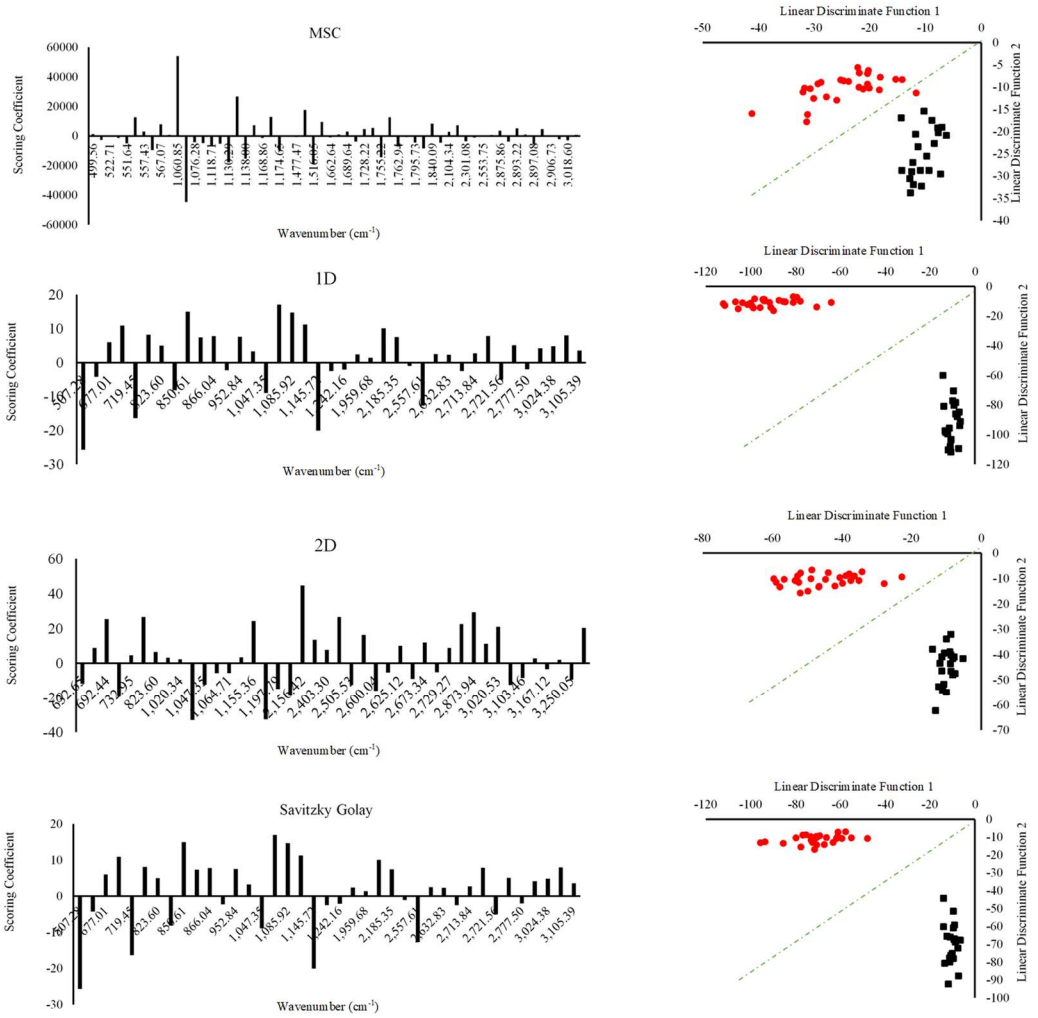


Figure 8.5: (Left) Bar graph showing the standardised scoring coefficients of the variables selected in the SLC-DA for some of the different spectra pretreatments, 14 of which were selected for LDA. Results of the LDA are shown on the right, the green dotted lines represent decision boundary.

From Table 8.6, it can be seen that during the training phase the LDA models obtained for all the pretreatments ranged from 81-100% accuracy. The classification model obtained was then repeated on the testing data set; with the exception of OSC and quantile normalization the validation accuracy ranged from 70-100%. From the results obtained it was shown that 1st, 2nd and Savitzky Golay derivatisation of the spectra had the highest % accuracy in both the training and validation dataset (100% for both), these corroborate the results obtained using PLS-DA and SIMCA. In the case of normalization spectral pretreatment, a higher rate of % accuracy (100% training 95% validation set) was obtained when compared to both the PLS-DA and

SIMCA models previously obtained. In the case of SIMCA pre-treating the spectral data so that to become normally distributed samples belonging to one class will be considered as not belonging to it due to the inherited loss of variation which the normalization procedure causes. The difference also dissent from the fact that in that SIMCA is a soft classification method and thus allows a single object to belong to more than one class, whilst LDA is a more robust form of classification as the objects are either classified in either one or the other class.

8.3.5 Supervised discriminate chemometric techniques – Support Vector Machine (SVM)

The dataset containing only variables which were selected using SLC-DA and having a VIP score > 0.8 were subjected to another classification method, known as support vector machine (SVM). SVM is similar to PLS and can be used for both classification and regression (Christianini and Shawe-Taylor 2000; Vapnik 1995). During this part of the experiment, SVM was used in the context of classification. Classification methods based on SVM employ the use of linear boundaries which are produced between discrete groups in a transformed space. The space in which the groups are projected is in terms of the x-variables which is usually of much higher dimension than the original space generated by the x-variables. The increase in the dimensional space allows the groups to become more linearly separable. The boundary (hyperplane) that separates the two classes projected into this higher dimensional space is known as the maximum margin classifier. Comparison of SVMs with other classification and regression methods has shown that they exhibit mostly good performances, although other methods proved to be very competitive (Meyer *et al*. 2003). Table 8.6 shows the results obtained on using SVMs using different Kernel tricks including radial, polynomial and sigmoidal types however only radial kernel tricks seemed to have any significant use for the FTIR data.

Kernel tricks (functions) were employed in order to account for linearly nonseparable classes in which the groups overlap in the transformed space. The kernel function allows for modifications in hyperplane in order to increase the sensitivity and the configuration of the hyperplane. Unlike PLS-DA and LDA, the SVMs decision boundary is mainly oriented at samples that show no clear classification whether they belong to one or the other class. Although SVM and Kernel functions seem to provide answers for data sets in which the class membership is difficult to obtain, SVMs tend

to be more greatly affected by the presence of outliers in the data set causing instabilities in model generated, especially if data outliers are used as support vectors (Steinwart and Christmann 2008).

Table 8.6: Comparison of the % accuracy and predictability of LDA models generated and SVM using a linear hyperplane and a radial Kernel trick.

Classification FTIR Spectroscopy VIP>0.8 & SLC-DA Dataset						
Pre-treatment	LDA		SVM			
			Kernel Type Linear		Kernel Type Radial	
	% Accuracy	% Predictability	% Accuracy	% Predictability	% Accuracy	% Predictability
Raw	97.92	95.00	100.00	70.00	100.00	80.00
Smoothing	87.50	70.00	100.00	100.00	89.58	80.00
Normalized	100.00	95.00	97.92	90.00	95.83	75.00
Quantile normalized	81.25	65.00	91.67	75.00	97.92	70.00
Baseline	100.00	85.00	100.00	60.00	100.00	65.00
Detrend	91.67	85.00	100.00	70.00	93.75	70.00
Deresolve	100.00	95.00	97.92	65.00	91.67	100.00
SNV	97.92	85.00	100.00	95.00	100.00	60.00
MSC	100.00	95.00	100.00	95.00	95.83	90.00
OSC	83.33	30.00	100.00	80.00	79.17	60.00
Savitzky Golay	100.00	100.00	100.00	100.00	100.00	95.00
1st Derivative	100.00	100.00	100.00	100.00	100.00	95.00
2nd Derivative	100.00	100.00	16.67	35.00	100.00	95.00

Results obtained from SVM classification are presented in Table 8.6. High rates of accuracy and predictability were obtained for the majority of the spectral pretreatments further validating that SVM classification is highly adaptable to the kind of data used. In the case of linear SVM, the best classification was obtained on using 5 point smoothing, 1st and Savitzky Golay derivatization techniques as 100% accuracy and predictability were obtained. Unlike what was observed in PLS-DA, SIMCA and LDA, spectra pretreated with 2nd order derivatisation had the lowest % accuracy in the training set (16.17%) and % predictability (35%) when compared to the rest of the spectral pretreatments under the linear type SVM. Unlike the rest of the other spectral pretreatments, the 2nd order derivatisation showed an increase in accuracy and predictability on the use of a radial type Kernel function. The 2nd order derivatisation reach a 100% accuracy and a 95% predictability on the use of a radial type Kernel function, suggesting that the group projected in the higher dimensional space cannot

be separated using a linear hyperplane but a spherical hyperplane formed from the used of the radial Kernel type function.

8.3.6 Application of hierarchical cluster analysis

Application of hierarchical cluster analysis using Ward's method was used on the variables selected using SLC-DA and VIP>0.8 for Savitzky Golay derivatization which most consistently showed higher rates of % accuracy and % precision. The cluster analysis revealed that very few (4 samples) of the EVOOs of Maltese origin were classified incorrectly. The dendrogram obtained showed the presence of 3 major clusters (marked, Red, Blue and Green), where the red cluster contained exclusively EVOO's of Maltese origin, whilst the Blue and Green clusters contained almost exclusively samples of non-Maltese origin. The overlaid data are the standardised β coefficients obtained from PLS-DA analysis. From the figure it can be seenthat certain regions within the spectra are highly variable between the different classes, and such variables would contribute more (terms of β coefficient magnitude) to the prediction formula. The most informative ranges which enable the highest discrimination between EVOOs of Maltese origin from those of non-Maltese origin included 851-922 cm^{-1} and 1044 – 1088 cm^{-1} which fall in the known region (1,500 –900 cm^{-1}) or "fingerprint region" because the pattern of the bands is particularly characteristic of molecular composition and can be used to identify minor substances.

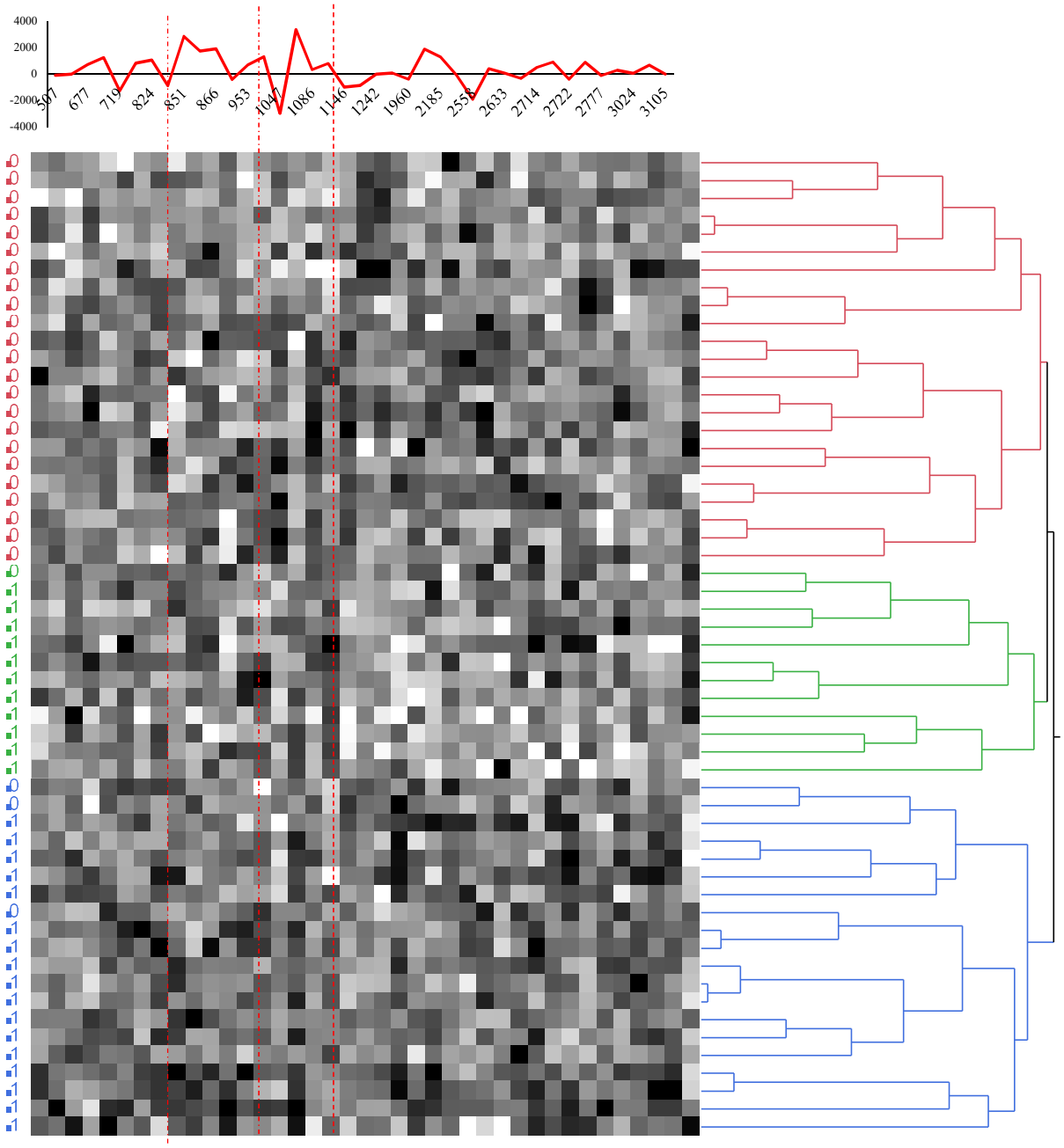


Figure 8.6. Application of hierarchal cluster analysis using Wards methods on the SLC-DA and VIP > 0.8 variables for the Savitzky Golay spectral pretreatment. The dendrogram shows three major clusters (Red) which contained exclusively EVOOs of Maltese origin (0) whilst the Blue and Green cluster contained mainly EVOOs of non-Maltese origin (1). (Top) The β coefficient scores obtained from the PLS-DA analysis of the selected variables.

8.3.7 Whole FTIR modelling using feed-forward predictive artificial neural networks.

The use of feed-forward predictive neural networks on the FTIR data as a method for classification was assessed using three different forms of validation, namely 33.3% of data holdback, CV-10 k-fold and excluded row validation. Artificial neural network (ANN) is a mathematical algorithm with the capability of relating the input and output parameter, learning from examples through iteration without requiring a prior knowledge of the relationships between the process variables (Cevoli *et al*., 2011). The main advantages of ANN are its nonlinearity, allowing better fit to the data; noise insensitivity, providing accurate predictions. The algorithm fitted on the training set was later tested on the validation data and % predictability of the model was obtained. Table 8.7 shows % accuracy and % predictability for the different forms of cross-validation. Similar, to what was observed in the PLS-DA and SIMCA, spectra treated using normalization had the lowest accuracy and predictability. Conversely the results obtained from the Savitzky Golay derivatization had the highest rates of accuracy and predictability, when compared to the other spectral pretreatments over the different forms of cross-validation used.

Comparison to the PLS-DA models obtained without any form of variable selection (Table 8.2), FF-ANN had a lower performance especially when it comes to the testing phase. The lower % precision recorded in the FF-ANN is coherent with a number of other studies which showed that PLS-DA has a higher sensitivity and performance (Khanmohammadi *et al*., 2011; Efstathios *et al*., 2011; Sampson *et al*., 2011) when compared to FF-ANN. FF-ANNs work better if they deal with non-linear dependence between input and output vectors and generally are more efficient in modelling classes separated with non-linear boundaries, however from the experimental data it has been shown that FTIR data for the different EVOO origins attains a more linearly discrimination as shown by SVM, and LDA results. Nonetheless ANN can provide a substantially good corroboration of PLS-DA without the excessive need of variable selection.

Table 8.7. Results summarizing the FF-PNN model performance with no variable selection for three different cross-validation methods.

FF-PNN FTIR Spectroscopy Whole Dataset						
CV Type Pre-treatment	Hold back		K-fold		Excluded Row	
	% Accuracy	% Predictability	% Accuracy	% Predictability	% Accuracy	% Predictability
Raw	100.00	100.00	85.51	92.31	85.51	92.31
Smoothing	97.10	92.31	97.10	92.31	94.20	76.92
Normalized	81.16	76.92	91.30	92.31	53.62	61.54
Quantile normalized	92.75	92.31	100.00	100.00	91.30	92.31
Baseline	63.77	69.23	98.55	100.00	68.12	69.23
Detrend	100.00	100.00	98.55	92.31	98.55	92.31
Deresolve	97.10	97.10	94.20	97.10	97.10	94.20
SNV	94.20	100.00	100.00	100.00	82.61	84.62
MSC	98.55	100.00	100.00	100.00	84.06	76.92
OSC	92.75	61.54	92.75	61.54	95.65	92.31
Savitzky Golay	95.65	92.31	100.00	100.00	91.30	100.00
1st Derivative	97.10	100.00	98.55	92.31	98.55	92.31
2nd Derivative	92.75	92.31	100.00	100.00	97.10	84.62

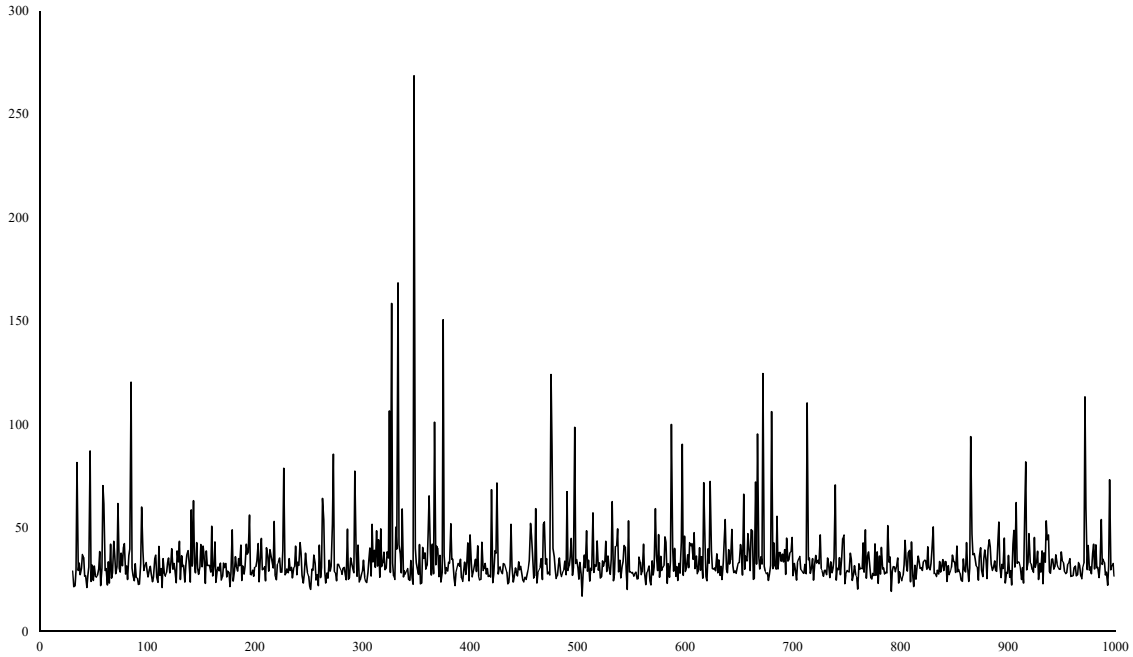
In conclusion, it was shown that FT-MIR-ATR spectra in conjunction with a number of chemometric methods, provided a cheap, fast and reliable way for the determination of the geographical origin of EVOOs, especially when it comes to discrimination of Maltese EVOOs from non-Maltese EVOOs. From the preliminary assessment using only unsupervised PCA models no significant clustering occurred. This was attributed to the high levels of similarity between the two classes of EVOOs studied. Therefore such a method was deemed to be unsatisfactory when it comes to discrimination of geographical origin. Application of supervised methods of classification namely PLS-DA, FF-PNN, LDA and SVM were shown to be highly effective in classifying local and non-local EVOOs samples. The use of the variable selection methods significantly increased the effectiveness of PLS-DA models when compared to no variable selection. FF-ANN, SVM and LDA models were also shown to offer similar classification rates to PLS-DA models and thus corroborated the results obtained from the PLS-DA models and put confidence in the use of FT-MIR-ATR methods in conjunction with spectral transformation for the classification of Maltese and foreign EVOOs samples.

9 Application of direct infusion mass spectrometry for the determination of geographical origin of olive oils

Direct infusion electrospray ionization mass spectrometry (DI-MS) is a fast analytical approach in which complex mixtures of different compounds are resolved into components differing in ion mass. Traditional methods which are used to establish the authenticity of olive oils require the separation and the analysis of a single class of compounds which are treated separately from each other. Typical traditional analysis ranges from fatty acid profiling (Aranda *et al.,* 2004; Bucci *et al.,* 2002; Caponio, Alloggio, and Gomes, 1999; D'Imperio *et al.,* 2007 and many other) to phenolic compound profiling (Caponio *et al.,* 1999; Go´mez-Alonso, Salvador, and Fregapane, 2002; Krichene *et al.,* 2007). All these techniques require the identification and resolution of compounds by the use of classical separation methods like GC and HPLC. Compared to the above methods DI-MS requires little or no sample preparation providing almost instantaneous information about the composition of a given sample, without any preliminary chromatographic separation.

The use of DI-MS has been proposed as a very fast, versatile, reproducible, and sensitive technique, which is capable of ionizing a wide range of molecules, especially polar ones without the need of chemical derivatization or extraction from the matrix (Alves *et al.,* 2010). A number of different studies have been published regarding the use of this methodology for seed oil and olive oil authentication (Alves *et al.,* 2010; Catharino *et al.,* 2005; Lerma-García *et al.,* 2008ab and 2011; Goodacre *et al.,* 2002). The main aim of this part of the study was to investigate the ability of direct infusion electrospray ionization mass spectrometry, without prior chromatographic separation, to produce information-rich and informative mass spectra from extra virgin olive oil and its application in the determination of geographical origin via the application of supervised statistical techniques.

9.1 DI-MS peak identification



$[M-X]^+$	m/z	$[M-X]^+$	m/z
OLL + 3K	994	OLn+K-C_{12}	489
OOO+3K- H_2O	986	OO-C_{12}	474
OLP+2K	972	Ln+K-H_2O/ olivil	374
OOP+2K	937	oleuropein derivative $[M-H]^+$	366
LLL+K	916	ligstroside [M-H-glu]	361
OOO	865	demethyl ligstroside [M-H-glu]	347
OLL-C_{10}	739	oleuropein derivative fragment $[377-CO_2]^+$	332
OLL-C_{12}	712	Olivin [M-H-H_2O-CH_2O]	326
OO+2K-H_2O	680	Verbascoside(M-H-Rham)$^+$	316
OO+2K	672	Ligstroside [[M-H-glc-C_4H_6O]	292
OO+K	667	closed ring carboxilade hidroxilade form –MeOH	272
LnLn+K	654	O-H_2O	263
OO/ verbascoside	623	ligstroside/ oleuropein fragment or methoxyluteolin fragment	226
LnLn	617	open ring decarboxilade aldehydic form	217
LnLn-H_2O	597	p-coumaric acid	194
PP-H_2O+K	587	Caffeic acid	178
oleuropein	539	Oleosides	140
OO-C_9	496	tyrosol fragment/ glycerol	94

Figure 9.1. The major peaks of interest obtained using DI-MS of a typical EVOOs.

Figure 9.1 shows a typical DI-MS spectrum of EVOO in methanolic KOH. A total of 36 peaks were tentatively identified, full identification using MS2 fragmentation studies as a confirmatory method for peak identification was not carried out, however, it is being proposed as a confirmatory study for a more comprehensive method.

The DI-MS spectrum can be simplified into four regions, the 1000-700 m/z range in which different molecular ions corresponding to triglycerides are expected to be found. In this region four different triglycerides were identified on the basis of their molecular weight and these included, oleic acid triglyceride OOO, oleic acid palmitic acid triglyceride OOP, oleic linoleic palmitic acid triglyceride OLP, linoleic acid triglyceride LLL, and oleic linoleic acid triglyceride OLL. Although the molecular ion corresponding to the exact expected molecular ion was observed for OOO the rest of the triglyceride compounds were mainly observed in the combination of potassium ($[LLL+K]^{+}$), derived from the methanolic KOH Furthermore, the presence of multiple potassium ions in conjunction with the triglycerides was also postulated on the basis of the constituents molecular masses, $[OLL + 3K]^{3+}$, $[OLP+2K]^{2+}$ and $[OOP+2K]^{2+}$. These compounds could not be fully identified due to their metastable nature, whereby their existence is attributed to the presence of high degree of unsaturation in the fatty acid chain which stabilises the extra positive charges on the molecular ion. In the case of an oleic linoleic acid (OLL 880 m/z), triglycerides fragments corresponding to the loss of C10 and C12 were observed at 739 m/z and 712 m/z. This was attributed to the combination allylic cleavage without the double bond migration, a typical fragmentation pattern displayed by straight chained alkenes of both linoleic and oleic acid. Although allylic cleavage without the double bond migration is plausible, alkenes can undergo another a major type of fragmentation which leads to alkyl radical loss whereby the double bond migrates before fragmentation thus the observed of ions of this type have little structural value. The combination of the two forms of fragmentation makes the identification of the fragmentation pattern of triglycerides fairly complicated and far beyond the scope of the study. Figure 9.2 shows the fragmentation pattern for linoleic acid in which the loss of C4 and C11 can occur prior to the double bond rearrangement. Figure 9.3 shows the different forms of oleic acid double bond migrations undergone in MS conditions giving rise to different fragmentation patterns.

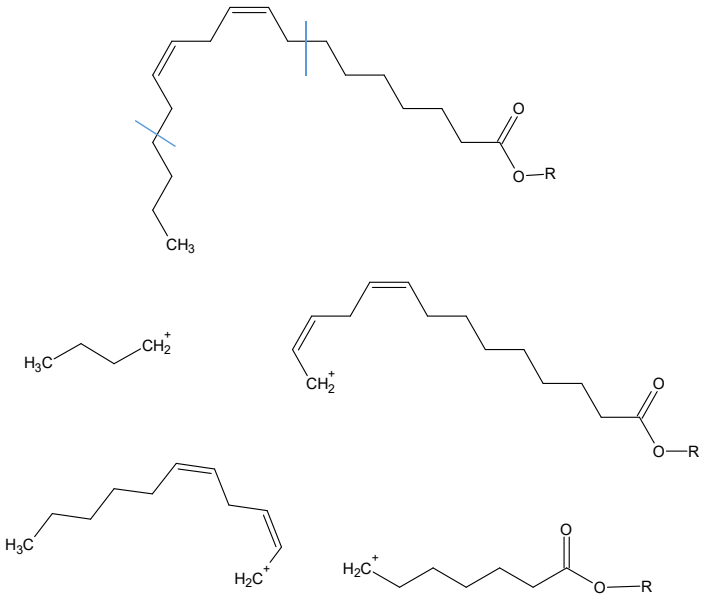


Figure 9.2. Allylic fragmentation pattern of linoleic acid without double bond rearrangement.

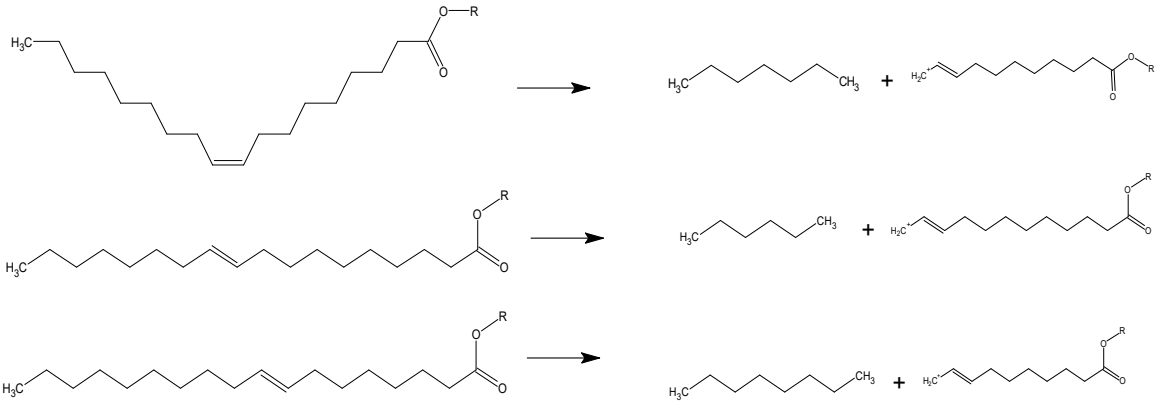


Figure 9.3. Allylic fragmentation pattern of oleic acid with double bond rearrangement.

The second region of the mass spectrum is the diglyceride region which occurs between 680-474 m/z. Similar to what was observed in the triglyceride region, this region was dominated by the diglycerides of oleic and linoleic acid, observed both in their free form (OO and LL) or else conjugated with single or multiple potassium ions. Fragments of the corresponding diglycerides were also observed as C9 and C12 fragments from oleic acid. Furthermore, the combination of potassium ions and loss of water molecules were also observed, however their presence is only tentative. In this region two high molecular weight phenolic compounds were identified, these included verbascoside and oleuropein, a presence which was also identified during the HPLC analysis of the phenolic fraction. Although their abundance was expected to be low the DI-MS was able to identify m/z values corresponding to the minor unsaponifiable fraction of EVOOs. This was further corroborated in the third region (380-220 m/z) of the spectrum which contained almost exclusively phenolic compounds and free fatty acids present in EVOOs. In this region of the spectrum the most abundant m/z value was obtained at 347 m/z which was tentatively identified as linoleic acid monoglyceride derivative, however, the presence of olivil could not be dismissed. Olivil is a dilignols which are exudates of the olive fruit responsible for the bittersweet taste and acid properties. The presence of this compound was further confirmed by the 326 m/z which correspond to olivil fragment after the loss of and CHOH.

Oleuropein and ligstroside derivatives and their corresponding fragments dominated this region. Figure 9.4 shows the major forms of these compounds and fragments observed. The presence of verbascoside previously identified at 623 m/z was further confirmed at 316 m/z which corresponds to the loss of the rhamnose sugar moiety. The presence of flavonoid compounds was not very compelling as only a tentative fragment observed at 226 m/z was observed corresponding to methoxy luteolin. This suggests that under the specified cone voltage flavonoid compounds present in EVOOs, namely luteolin and apigenin could not be fully observed as these compounds were easy fragmented or masked by more prominent stable mass fragments. Minor signals were observed at 299 m/z 285 m/z, and 269n/z which correspond to luteolin, apigenin, and methoxy luteolin. The last region (200-95 m/z) of the MS spectrum included the presence of simple phenolic compounds namely caffeic and p-coumaric acid together with tyrosol and glycerol as fragments from

complex phenolic compounds and glyceride fatty acids respectively. The prominent peak observed at 40 m/z was attributed to potassium which was used during the experiment.

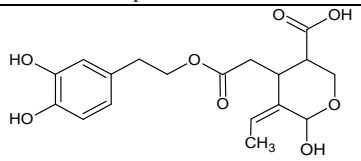
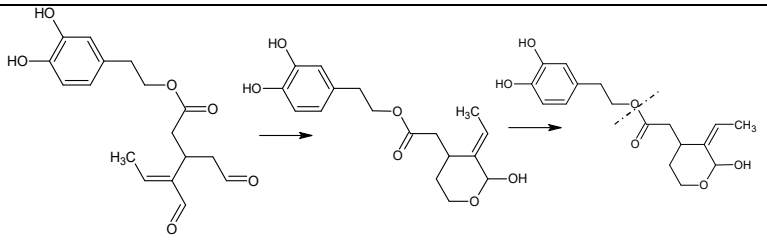
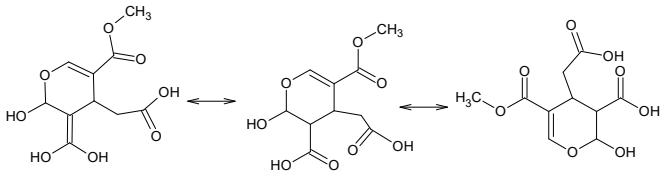
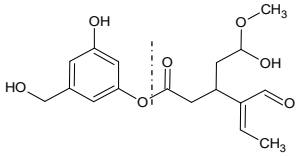
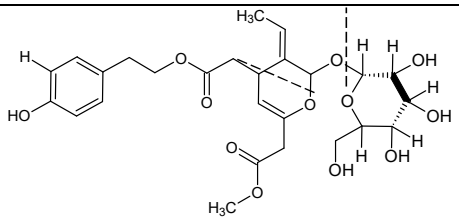
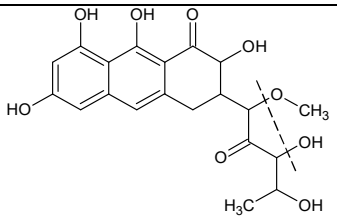
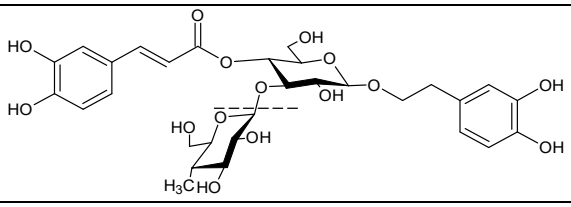
m/z	Name	Compound Structure
347	Oleuropein aglycone demethlyated derivative	
226	After rearrangement of Decarboxylated Form of Ligstroside aglycon	
272	Closed ring carboxilade dihydroxilade form formed after rearrangement reaction	
366 266 217	Open ring Decarboxilade Aldehydic form of Oleuropein	
361 292	Ligstroside [M-H-glu] and Ligstroside [[M-H-glc-C_4H_6O]	
326	Olivin [M-H-H_2O-CH_2O]	
316	Verbascoside (M-H-Rham)$^+$	

Figure 9.4. Phenolic compounds and their corresponding fragments tentatively identified using DI-MS

9.2 Application of chemometrics to Direct infusion mass spectroscopy

As previously stated the aim of this study was to build reliable classification models for the traceability of EVOOs from Malta by coupling electrospray ionization direct infusion mass spectroscopy and chemometrics. To this purpose, DI-MS spectra of olive oil samples from the Maltese islands and from other Mediterranean countries were collected as described in Section 2.2.1 and analyzed as described in Section 2.2.6.3. To identify the spectral fingerprints corresponding to the origin of the sample, discriminant (PLS, LDA) and modelling (SIMCA, PNN) classification approaches were used and compared.

9.2.1 Unsupervised chemometric techniques – PCA

Different kinds of spectral pretreatments were tested and compared in order to overcome the instrumental limitation and account for scattering and other minor variations which would hinder the performance of the classification models. A total of 12 spectral pretreatment methods were used, where in each case, after pretreatment a principal component analysis was carried. Principal component analysis is a variable reduction procedure. Although preliminary variable reduction was carried out in order to reduce the amount of data, PCA enables visualization of data in a smaller number of principal components (artificial variables). Principal components will account for most of the variance in the observed dataset thus avoiding redundancy in original variables. The mass spectra obtained are expected to have a very high amount of redundancy due to the nature of fragmentation as some of the variables would be inheritably correlated with one another. PCA enabled the preliminary identification of which pre-treatment method offered the highest variability and possible clustering. Figure 9.5 shows the different forms of spectral pretreatments employed and the corresponding PCA plot for the first two principal components.

Preliminary assessment of the score plots obtained for the first two principal components and the majority of the spectral pretreatment resulted in a low % variability explained. This was attributed to the presence of a high degree of redundancy in the MS data, furthermore, mixing the variables (m/z) with a different meaning is likely to create blocks of very weak correlations making it difficult for any principal component to capture a large share of the variability. In addition, it was found that the majority of the observations fell into a straight line on the first principal

component suggesting that the first principal component was highly correlated to all the variables in the mass spectrum, failing to identify any variations between actual peaks and the instrumental noise.

From the results (see Figure 9.5) obtained it was found that for the raw data, baseline correct, resolved and orthogonality corrected spectra a higher % variation was explained in the first principal component, furthermore the biplot indicated a sample distribution which seemed to be reflecting the geographical origin, as EVOOs of the Maltese islands tended to cluster separately from those obtained from other Mediterranean countries. Figure 9.5 also shows the loading plots observed for the aforementioned data sets. As expected the loading of the first principal component seemed to be equally affected by the entire spectrum and failed to pick out peaks of relevant importance for the observed variation. However, with the use of higher order components 3-7 which were not plotted in Figure 9.5, a pattern resembling the major peaks obtained in the MS started to emerge, giving a further confirmation that the use of just two components is not enough to describe the variation in the dataset. Data obtained using mass spectrometry is highly dimensional, requiring much more than two or three PCs to get close to capturing most of the variation making PCA unsuitable for such data set.

Similar to what was carried out in the other techniques, the next step was to divide the whole dataset into training and test sets (the former to build the model, the latter to validate it). The Maltese and the non-Maltese samples were grouped in an ascending way so that the first 15 samples would represent Maltese EVOO's whilst the rest correspond to non –Maltese EVOO's. A stratified random sampling method was used in order to exclude 20% of the observation so that they would be retained as the testing set. The remaining 80% of the observation were used as to build the training set.

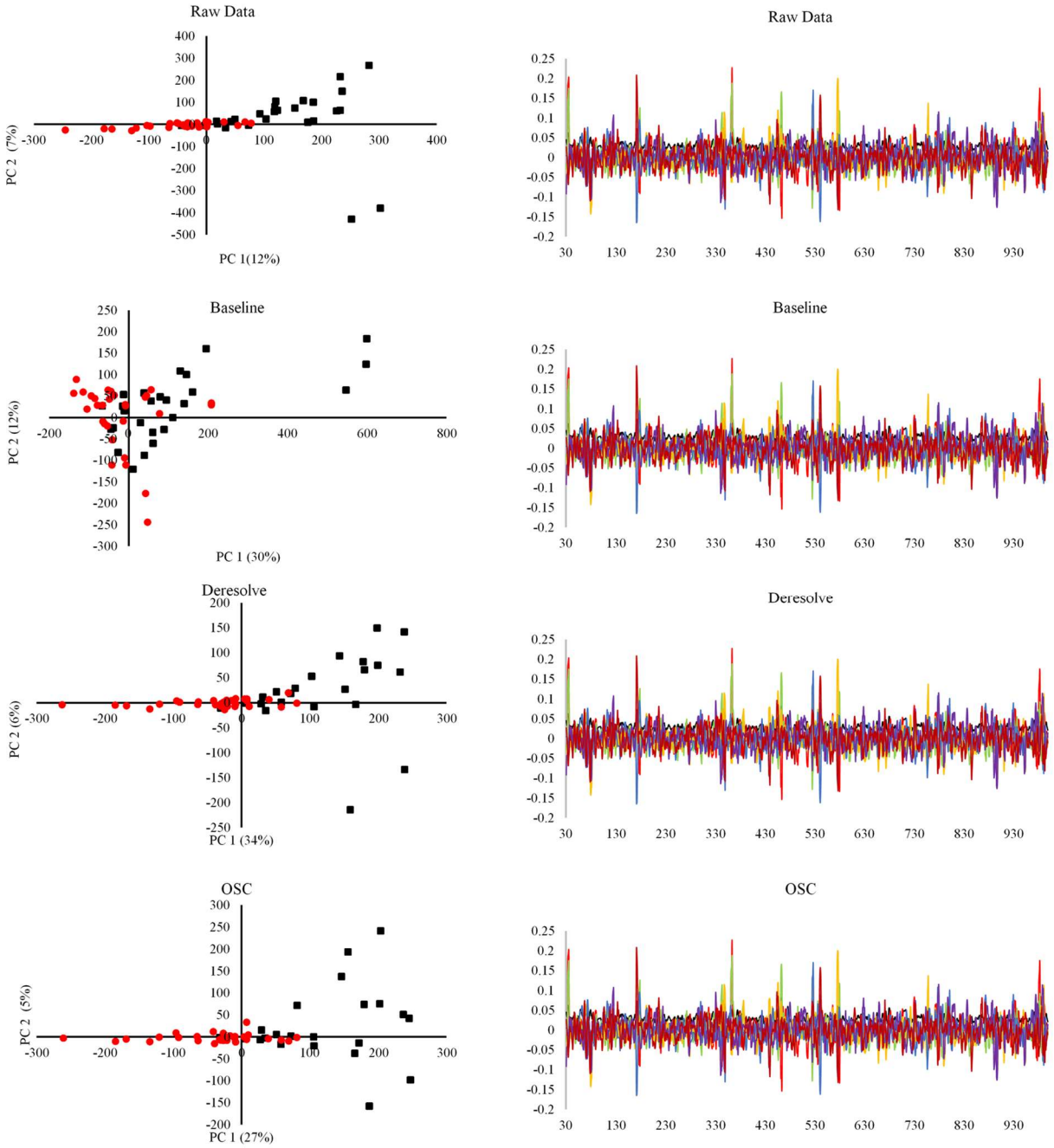


Figure 9.5: DI-MS spectral transformations PCA biplot obtained for EVOOs of Maltese origin (black spectra and black dots) and non-Maltese (red spectra and red squares) and there corresponding loading plots for 7 principle components. 231

9.2.2 Supervised chemometric techniques – PLS-DA

After splitting the data according to the procedure described above, chemometric classification models were built and tested on all the MS spectral pretreatment using a PLS regression algorithm using JMP 10 and its inbuilt leave one out cross-validation method (LOOCV). Partial least squares regression combines features from and generalizes principal component analysis (PCA) and multiple linear regression. Its goal is to predict a set of dependent variables (geographical origin) from a set of independent variables or predictors (m/z). This prediction is achieved by extracting from the predictors a set of orthogonal factors called latent variables which have the best predictive power.

Table 9.1 shows the number of latent variables extracted, the predicted root mean square error and the % variation explained in terms of X and Y for the different spectral pretreatment methods. Unlike what was observed in other spectroscopic methods, with the exception of OSC in all the other spectral pretreatment methods PLS managed to extract only one latent variable. Similar to what was observed in PCA analysis this was attributed to the high level of data collinearity, making the extraction of orthogonal latent variables from the predictors difficult. This is one major limitation of PLS, that occurs when the number of predictors is large compared to the number of observations, having a singular nature, the regression approach tends to fail. This form of data is often referred to the 'small N large P problem.' It is typical of recent data analysis areas such as bioinformatics and genomics. Several approaches have been developed to overcome the multicollinearity problem these included the elimination of some predictors through the use of stepwise methods, and the use of ridge regression such as called principal component regression (PCR), both of which will be presented at a later stage in this study.

Further indications that the PLS models obtained were not suitable to model the entire MS spectrum come from the van der Voet T^2 test. The van der Voet T^2 test enables the determination of whether the PLS model obtained with a specified number of extracted factors differs significantly from a proposed optimum model. The test the null hypothesis states that both models have the same predictive ability. From Table 9.1 it was shown that only models obtained using raw data, deresolve, SNV, MSC and

OSC had a probability > van der Voet T^2 test p-value > 0.05 indicating that there was no significant difference between the model obtained and a proposed optimum model.

Table 9.1 PLS-DA analysis using the whole DI-MS data. (Top) the results obtained using LOOCV on the training dataset. (Bottom) the results obtained using LOOCV and 20% of the data as the validation group.

DI-ES+ Mass Spectrometry Whole Spectrum Internal Validation							
Pre-treatment	Latent Variables	% X	%Y	PRESS	T^2	$p>T^2$	% Accuracy
Raw	1	13.54	55.39	0.84	0.00	1.00	93.48
Normalized	1	3.10	90.06	1.02	2.63	0.10	69.57
Quantile normalized	1	2.54	93.80	1.02	1.41	0.25	100.00
Baseline	1	27.51	20.17	1.06	0.63	0.42	100.00
Detrend	1	3.43	87.71	1.14	3.14	0.09	100.00
Deresolve	1	36.08	41.23	0.84	0.00	1.00	80.43
SNV	5	12.40	100.00	1.02	0.00	1.00	100.00
MSC	2	49.36	37.48	1.01	0.00	1.00	80.43
OSC	15	56.88	100.00	0.57	0.00	1.00	97.83
Savitzky Golay	1	3.62	81.27	14.02	4.92	0.20	100.00
1st Derivative	1	3.51	82.38	1.18	3.51	0.82	97.83
2nd Derivative	1	3.52	85.87	1.11	1.38	0.25	100.00

DI-ES+ Mass Spectrometry Whole Spectrum External Validation								
Pre-treatment	Latent Variables	% X	%Y	PRESS	T^2	$p>T^2$	% Accuracy	% Predictability
Raw	1	13.54	58.35	0.87	0.00	1.00	93.48	90.00*
Normalized	1	3.70	90.63	1.19	4.09	0.04	69.57	50.00
Quantile normalized	1	3.07	93.45	1.12	4.70	0.03	93.48	70.00
Baseline	1	28.31	24.52	1.03	0.00	1.00	95.65	80.00*
Detrend	1	4.01	89.00	1.23	4.53	0.02	93.48	70.00
Deresolve	1	35.85	40.27	0.87	0.00	1.00	82.61	90.00*
SNV	1	3.24	92.04	1.08	1.25	0.28	95.65	80.00*
MSC	1	14.90	33.68	1.03	0.00	1.00	76.09	70.00*
OSC	13	58.17	100.00	0.71	0.00	1.00	95.65	90.00*
Savitzky Golay	1	9.51	82.71	1.16	2.84	0.02	91.30	70.00
1st Derivative	1	4.10	84.51	1.24	4.13	0.02	93.48	80.00
2nd Derivative	1	4.32	88.59	1.08	0.41	0.55	93.48	70.00*

From the chosen PLS models obtained which had a probability, > van der Voet T^2 test p-value > 0.05 the % accuracy (correct classification) obtained in the training test ranges from 76-95% with an overall validation accuracy of 70-90%. Of the selected PLS models, MS spectra pretreated using OSC had the best performance followed by the untreated raw data. This might suggest that rather than improving the model performance the application of spectral pretreatment methods on MS data is, in fact, hindering the performance of PLS. This could be due to the amplification of

multicollinear data during the spectral transformations which aim to increase the signal to ratio.

In order to fully interpret the PLS models obtained, an inspection of the VIP scores was used in order to determine which predictors (variables) are mainly influencing the latent vectors obtained. VIP is an index of how much a single variable contributes to the bilinear model and it is scaled in such a way that indices having VIP larger than 0.8 are considered to be significantly contributing to discrimination. VIP scores > 0.8 for the PLS models built on the different pretreated MS data are reported in the first column of Figure 9.6.

As shown in the figure, the VIP > 0.8 identified relevant features in the spectra, particularly, peaks which corresponded to previously identified compounds, however some noise data was also selected. The next step was to build another PLS model this time using only variables which had a VIP score > 0.8. Table 9.2 shows the results obtained on using the adjusted PLS model. A marked improvement from the previously obtained PLS models was observed indicating that the variable selection procedure removed some of the redundant collinear data, enhancing the performance of the PLS models. The VIP score is a measure of a variable's importance in modelling both response (geographical origin) and the predictors (m/z). If a variable has a small coefficient and a small VIP, then it is a candidate for deletion from the model (Wold,1995). A value of 0.8 is generally considered to be a small VIP (Eriksson *et al*, 2006) thus it is assumed that variables having a value less than 0.8 are not informative and are mainly attributed to noise.

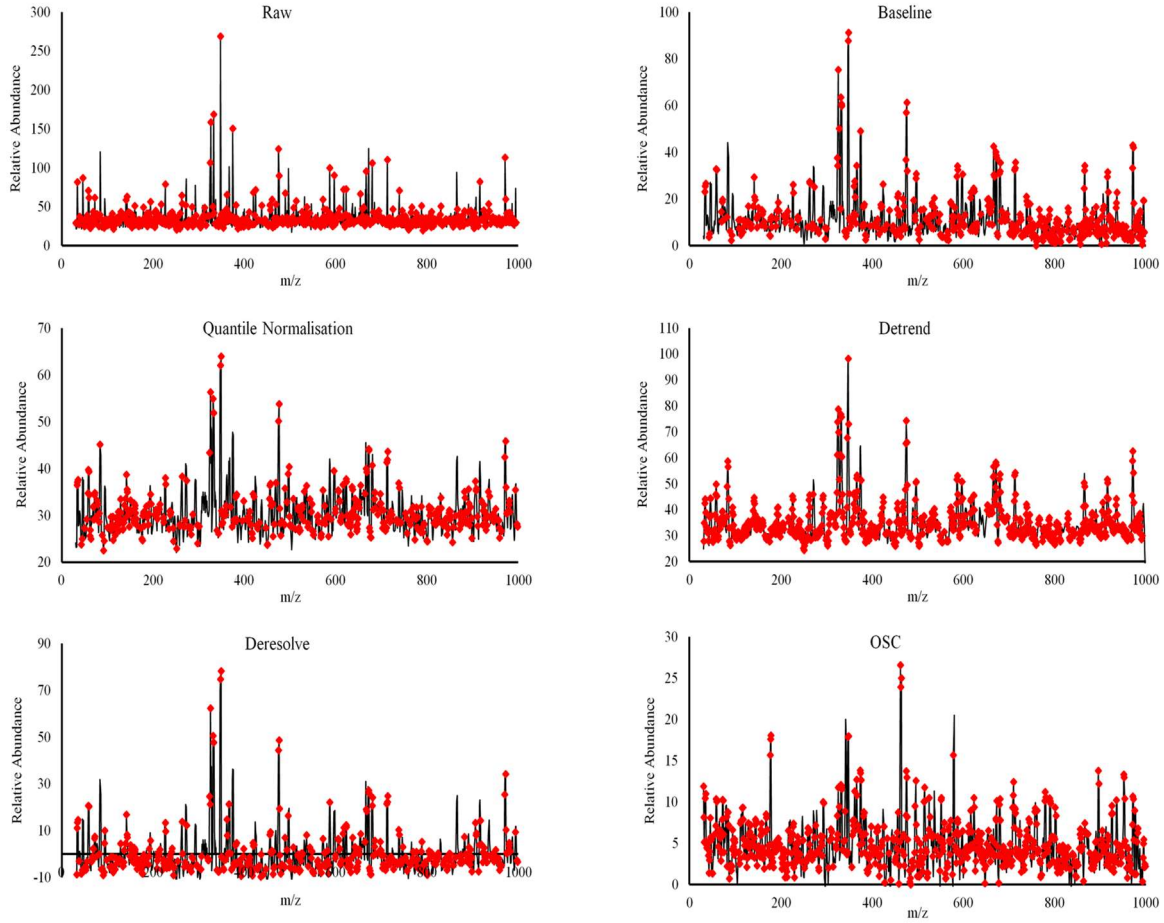


Figure 9.6: The variables which had a VIP score > 0.8 (Red dots) for the different spectral pretreatments which were selected for an adjusted PLS-DA

Table 9.2: PLS-DA analysis using the DI-MS data having a VIP>0.8. (Top) the results obtained using LOOCV on the training dataset. (Bottom) the results obtained using LOOCV and 20% of the data as the validation group

DI-ES+ Mass Spectrometry VIP>0.8 Spectrum Internal Validation					
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy
Raw	1	18.36	61.16	0.71	93.48
Normalized	2	9.36	97.20	0.57	69.57
Quantile normalized	2	7.91	99.06	0.55	100.00
Baseline	1	34.77	21.52	1.00	100.00
Detrend	1	6.39	86.83	0.61	100.00
Deresolve	1	38.96	46.04	0.79	84.78
SNV	4	13.40	99.96	0.53	100.00
MSC	3	63.37	61.56	0.86	86.96
OSC	15	59.34	100.00	0.51	100.00
Savitzky Golay	1	6.86	80.75	0.67	97.83
1st Derivative	1	6.44	81.59	0.70	97.83
2nd Derivative	1	6.80	84.93	0.61	100.00

DI-ES+ Mass Spectrometry VIP>0.8 Spectrum External Validation						
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy	% Predictability
Raw	1	18.29	61.90	0.73	93.48	100.00*
Normalized	2	10.49	97.73	0.65	71.74	60.00
Quantile normalized	1	5.60	92.14	0.63	100.00	100.00
Baseline	1	36.44	24.83	0.99	100.00	100.00
Detrend	2	10.98	97.44	0.67	100.00	100.00
Deresolve	1	38.67	44.62	0.83	84.78	90.00
SNV	11	35.29	100.00	0.63	100.00	100.00
MSC	1	14.84	36.56	0.98	76.09	80.00
OSC	13	60.30	100.00	0.64	100.00	100.00
Savitzky Golay	1	7.87	80.77	0.72	95.65	90.00
1st Derivative	1	6.80	82.47	0.79	97.83	100.00
2nd Derivative	1	7.77	87.05	0.60	95.65	80.00

Comparing the models obtained using variable selection to that previously obtained without any variable selection a noticeable difference was observed in % accuracy and predictability of the model. Furthermore all the models obtained had a probability > van der Voet T^2 test p-value > 0.05 indicating that the models obtained did not vary significantly from an optimal model. Furthermore, a lower PRESS was observed for the majority of the pretreatments. This observation suggests that unlike the case of ATR-FTIR data MS data require an extensive variable selection, is required to obtain a very good classification method with the use of PLS models.

9.2.3 Modelling Chemometric techniques – SIMCA

The classification analysis of EVOOs MS data was then repeated using a modelling approach based on the SIMCA algorithm. The latter is a class modelling algorithm that allows the analyses one class at a time. For SIMCA analysis two PCA were built for each spectral pretreatment one for the EVOOs of Maltese origin and the other one for the EVOOs of non-Maltese origin. In each case, the optimal number of principal components were chosen in order to obtain optimal model complexity in 10-fold row-wise cross-validation. The use of a two-stage model, one for each category, allowed the comparison between the two categories, thus this determined whether samples are accepted by one, both or none of the modelled classes. The output of SIMCA analysis was assessed by the use of Coomans plot shown in Figure 9.7. Similar to the FTIR experiment for this experiment given that the MS data was highly similar the significance limit was increased up to 25% rather than the default 5%.

From a preliminary survey of the Coomans plot outcome at the 25% level of significance, it was shown that the SIMCA models obtained were highly specific to the Maltese EVOOs whilst less specific to the non-Maltese EVOOs. This can be observed for the majority of the spectral pretreatments whereby the majority of the Maltese EVOO samples clustered in the lowest right region whilst the rest of the non-Maltese EVOOs clustered on the lowest side of the Coomans plot indicating that these samples were expected by both the Maltese and non-Maltese EVOO models built. Although in previous chemometric analysis using different spectroscopic methods, a new line bisecting the plot corresponding to discriminant classification boundary was built in order to represent a new significant boundary in order to overcome the ubiquity of the non-Maltese EVOOs this could not be applied to MS data thus the results presented in Table 9.3 are given only in terms of the specificity of the model towards the Maltese EVOOs. The table shows the % sensitivity and % specificity of the models obtained towards the Maltese EVOOs. Sensitivity is the percentage of samples from the modelled class (Maltese EVOOs) that are accepted by the class model, while specificity is the percentage of samples from other classes (Non-Maltese EVOOs) which are rejected by the class model. It was observed that SIMCA models obtained from 2nd order derived spectra and multiple scattering-corrected spectra failed to produce any form of classification, whilst MS raw data, baseline corrected, and resolved spectra obtained 100% sensitivity and 100 % specificity towards the Maltese

EVOOs. These results corroborated the results previously obtained using PLS analysis.

Table 9.3: Results obtained for SIMCA modelling for the different spectral pretreatments using the only the variables having a VIP > 0.8. The values recorded in the table represent the % sensitivity and % specificity of the different models towards the two classes using the 25% decision boundary.

Pre-treatment	% Sensitivity	% Specificity
Raw	100.0	100.0
Normalized	93.3	100.0
Quantile normalized	60.0	100.0
Baseline	100.0	100.0
Detrend	73.3	93.5
Deresolve	100.0	100.0
SNV	60.0	100.0
MSC	13.3	0.0
OSC	80.0	100.0
Savitzky Golay	80.0	100.0
1^{st} Derivative	86.7	87.1
2^{nd} Derivative	0.0	0.0

Figure 9.8 shows the discriminatory power of the variables used for the mass spectra obtained using raw data (black), baseline corrected (red) and detrending function (green). Discriminatory Power computes the standard deviation of residuals for all samples (EVOOs) of a specific class after being fitted to the model of another class and then compares this standard deviation with those calculated for samples fitted to their own class model. The plot shows how much each variable contributes to separating two models. It was shown that certain peaks tend to show a common contribution for the three most sensitive models these include the peaks observed at; 326 m/z, 347 m/z, 374 m/z, 462 m/z, 527 m/z, 575 m/z, and 757 m/z.

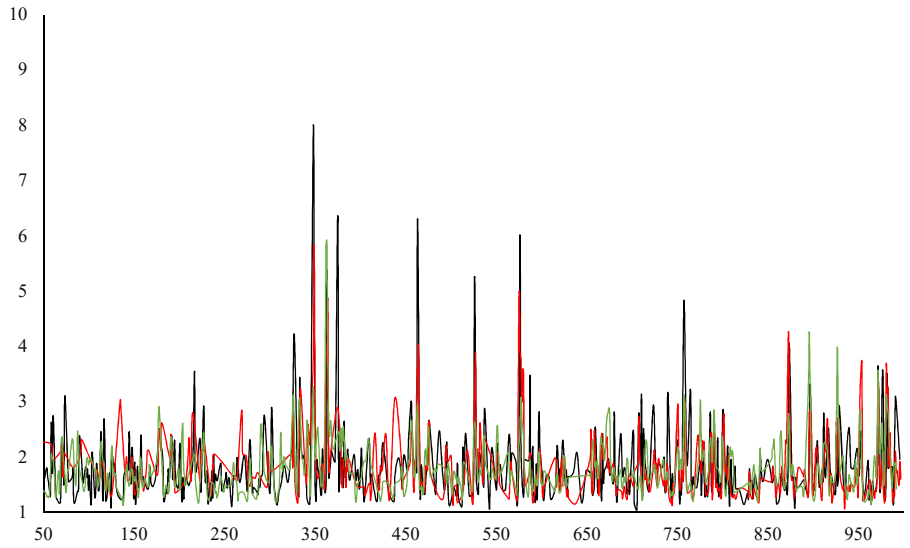


Figure 9.7: The discriminatory power of the DI-MS variables used for the mass spectra obtained using raw data (black), baseline corrected (red) and detrending function (green).

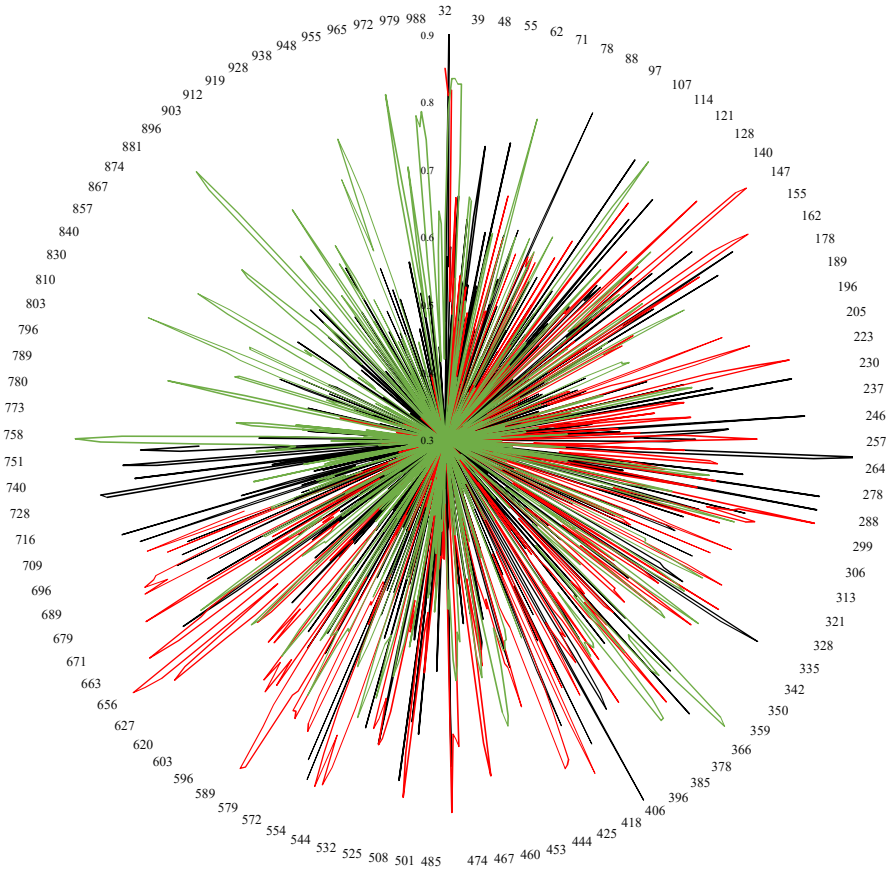


Figure 9.8: The modelling power of DI-MS variables for the Maltese EVOOs class, for the mass spectra obtained using raw data (black), baseline corrected (red) and detrending function (green).

Figure 9.9 shows the modelling power for the Maltese EVOOs class, modelling power is used to assess the ability of a variable to model the data in a specific class in PCA based models. Variables with a modelling power near one are important for the model. From the modelling distribution, it was shown that raw data and detrending function had a higher number of important modelling variables in the 716-998 m/z range the baseline corrected spectra displayed that m/z values with a high modelling power in the range of 425-700 m/z. In the case of raw data the most important modelling variables which enabled the classification of Maltese EVOOs include; 263 m/z, 409 m/z (closed ring carboxilade hydroxilade oleuropein form), 713 m/z, 739 m/z. In the case of baseline corrected spectra the 149 m/z (oleuropein fragment [M-H-glc-$CH_2CHPh(OH)_2$-CH_3COOH-CH_3OH]$^+$), 226 m/z, 292 m/z, 425 m/z ([LP–C_{16}]K^+), 496 m/z, and 654 m/z offered the best modelling power for the Maltese EVOO class. The detrended MS data showed that 366 m/z, 758 m/z, 810 m/z (PPP^+), 896 m/z ([OOP]K^+) and 972 m/z offered the best modelling power.

In order to obtain a more robust method of classification with the use of a smaller number of variables, the VIP data set obtained from the previous analysis was subjected to a stepwise linear canonical discriminate analysis SLC-DA. SLC-DA was performed on the MS data from all the pretreatment methods in order to extract only a small amount of highly discriminate variables which would enable an easier and faster discrimination between the origins of EVOOs. This strategy involved a substantial reduction of the dimensionality of the data in such a way that only the variables shown in Figure 9.10 were retained. In order to further reduce the number of variables selected from the SLC-DA analysis, a minimum of 14 variables was selected in order to carry out a conventional LDA. During the SLC-DA the variables chosen by applying a forward stepwise variable selection algorithm using JMP 10 using a Wilks' Lambda as a selection criterion and an F-statistic factor to determine the significance of the changes in Lambda when the influence of a new variable is evaluated. The most significant variables were then extracted and their canonical scoring coefficients were plotted as shown in Figure 9.11. The main advantage of using SLC-DA over the convention LDA is the ability to perform a feature selection. Regarding this fact, only those variables which helped to improve classification performance were used whereas variables without discriminant information were discarded. The PCA, SIMCA and PLS analysis were carried out on the reduced data, as shown in Appendix 15.

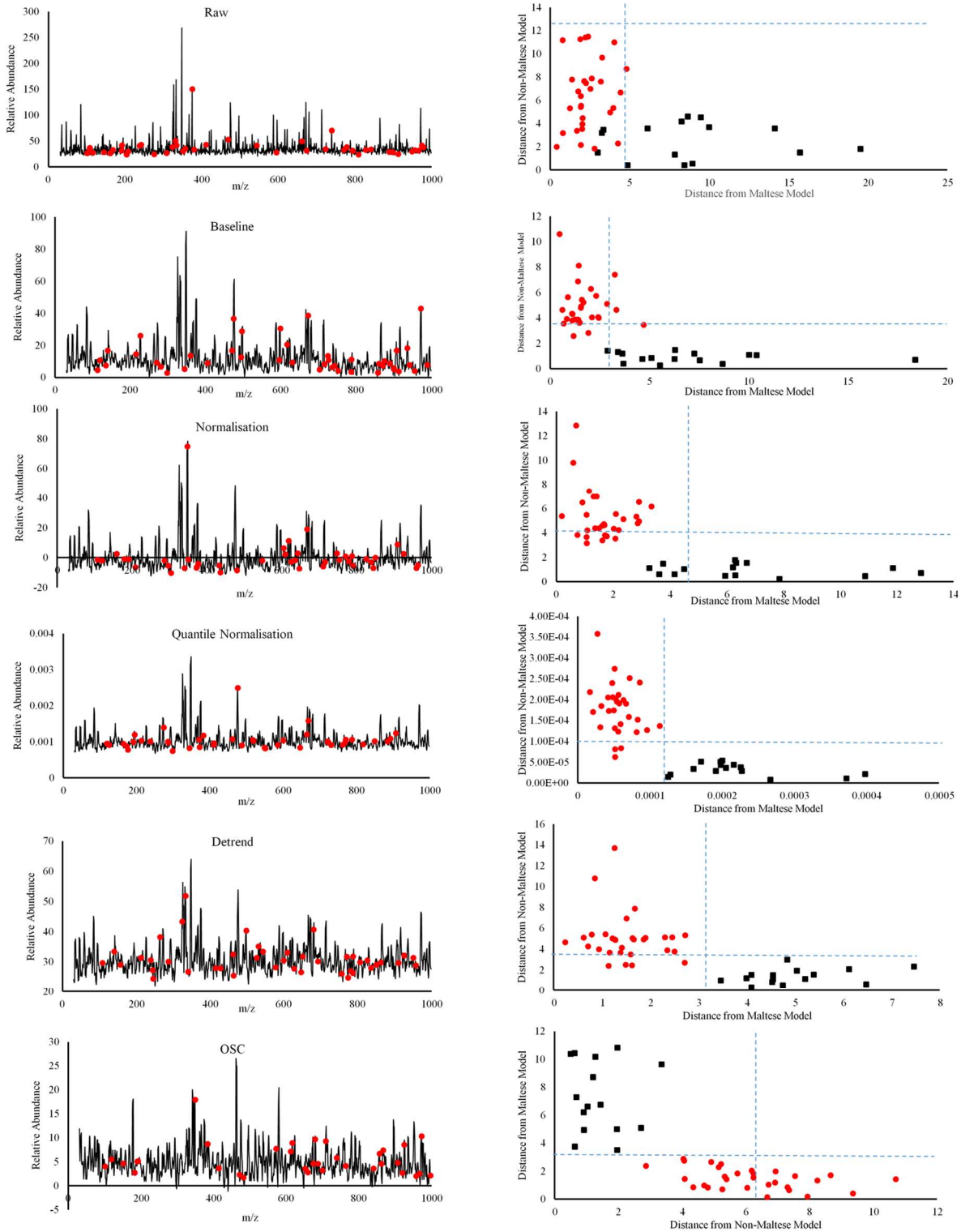


Figure 9.9: (1st column) illustrates the variables which had a VIP score > 0.8 and selected in the SLC-DA (Red dots) for the different spectral pretreatments which were selected for an adjusted PLS-DA and SIMCA. (2nd column) Coomans plot obtained using SIMCA on the selected variables only. The blue dotted lines represent the 25% confidence level.

Figure 9.9 shows that the variables selected during SLC-DA for some of spectral pretreatments were not the major most abundant peaks which would correspond to molecular fragments of the different fatty acids but rather peaks which show a very low intensity but are yet differentiated from the instrumental noise and offer the highest discriminatory power.

Table 9.4: PLS-DA analysis using common variables selected from both SLC-DA and having a VIP>0.8. (Top) the results obtained using LOOCV on the training dataset. (Bottom) the results obtained using LOOCV and 20% of the data as the validation group.

DI-ES+ Mass Spectrometry VIP>0.8 & SLC-DA Spectrum Internal Validation					
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy
Raw	15	69.52	99.92	0.51	100.00
Normalized	13	57.57	99.84	0.35	100.00
Quantile normalized	15	61.45	99.87	0.52	100.00
Baseline	15	82.14	99.89	0.44	100.00
Detrend	12	52.98	99.69	0.48	100.00
Deresolve	15	76.54	99.91	0.27	100.00
SNV	3	18.25	93.45	0.54	100.00
MSC	4	66.98	72.10	0.88	95.65
OSC	15	78.90	99.98	0.15	100.00
Savitzky Golay	15	73.84	99.72	0.21	100.00
1st Derivative	15	67.71	99.80	0.49	100.00
2nd Derivative	15	64.47	99.83	0.44	100.00

DI-ES+ Mass Spectrometery VIP>0.8 & SLC-DA Spectrum External Validation						
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy	% Predictability
Raw	6	44.91	99.06	0.66	100.00	100.00
Normalized	2	14.19	92.95	0.58	95.65	80.00
Quantile normalized	5	28.99	97.85	0.71	100.00	100.00
Baseline	15	84.56	99.96	0.58	100.00	100.00
Detrend	15	66.55	99.97	0.59	100.00	100.00
Deresolve	15	80.23	99.94	0.48	97.83	90.00
SNV	1	10.58	76.14	0.64	100.00	100.00
MSC	1	16.38	33.06	0.99	73.91	70.00
OSC	15	82.72	99.89	0.28	100.00	100.00
Savitzky Golay	15	78.37	99.83	0.45	100.00	100.00
1st Derivative	1	10.12	78.36	0.70	93.48	70.00
2nd Derivative	2	13.61	93.02	0.62	97.83	90.00

Once the variables were selected for each spectral pretreatment a PLS-DA and a SIMCA model was applied in order to determine whether variable selection using the linear method provided ways for a better form of classification. Table 9.4 shows the results obtained from the PLS using the data set composed of variables which had a VIP score > 0.8 and were selected during the SLC-DA analysis. The results show a marked increase in the performance, in terms of the number of latent variables extracted, PRESS, % accuracy and % predictability of the PLS models obtained using the reduced variable date set, suggesting that a variable selection using the two techniques greatly improves the modelling power. Furthermore, a higher number of latent variables were extracted for the majority of the spectral pretreatments resulting in a higher % of variance explained in terms of X and Y. A marked decrease in the PRESS was also noted with the orthogonal signal corrected MS data having a PRESS lower than 0.3, indicating a very strong and robust classification was obtained even when 20% of the data was omitted and leave one out cross-validation was applied. The PLS models obtained for the majority of the spectral pretreatments showed a 100% accuracy during the training phase and 100% predictability during the validation phase.

Table 9.5: Results obtained for SIMCA modelling for the different spectral pretreatments using the only the variables having a VIP > 0.8 and selected in SLC-DA. The values recorded in the table represent the % sensitivity and % specificity of the different models.

Pre-treatment	Maltese EVOOs		Maltese Non-EVOOs	
	% Sensitivity	% Specificity	% Sensitivity	% Specificity
Raw	73.3	100.0	0.0	100.0
Normalized	100.0	100.0	96.8	100.0
Quantile normalized	93.3	100.0	61.3	100.0
Baseline	86.6	100.0	90.3	100.0
Detrend	66.6	100.0	96.8	100.0
Deresolve	60.0	100.0	22.6	100.0
SNV	53.3	100.0	54.8	100.0
MSC	26.7	100.0	22.6	100.0
OSC	100.0	100.0	51.6	100.0
Savitzky Golay	86.7	90.9	6.5	100.0
1st Derivative	93.3	100.0	3.2	100.0
2nd Derivative	73.3	100.0	12.9	100.0

The results obtained showed that all the models obtained using SIMCA on the different spectral pretreatments showed a very good specificity indicating that the two classes can be modelled separately and no sample was misclassified, however for the majority of the models the sensitivity is much lower indicating that whilst the different classes can be discriminated the model failed to completely model the classes separately as some of the samples were still accepted in both the classes. This is mainly attributed to the fact that there are very few signals which can actually discriminate between the different classes compared to a large number of the signal which is present in both of the classes making the models built by SIMCA less sensitive for a particular class. Although the variable selection was carried out in order to maximize the discrimination between the two classes, analyzing each and every variable on its own will increase the sensitivity of the SIMCA, however, it will rout the overall aim of the experiment. From the results obtained it was shown that baseline corrected and normalized MS spectra gave the most promising results in terms of sensitivity and selectivity towards both the Maltese and non-Maltese EVOOs.

9.2.4 Supervised chemometric discriminate analysis techniques – LDA

In order to build the LDA, the selected variables obtained in SLC-DA were arranged in ascending order in terms of their scoring coefficients. A smaller set of variables were selected which consisted of 14 variables which corresponded to 7 of the most positive and 7 most negative standardized scoring coefficients. An LDA was carried out on the training set using only the small set of variables which were selected. The results obtained for the training samples were a visualized on an LDA biplot samples as shown in Figure 9.11 whereby each sample is projected as the scores obtained for the first two discriminate functions. In comparison with SIMCA, LDA avoids the normality problem and confidence interval adjustment making it more reliable methods for classification. The LDA method employs linear decision boundaries, which are defined in order to maximize the ratio of between-class to within-class dispersion (Fisher, 1936). It has been successfully applied to a number of classification problems (Gambarra Neto *et al*., 2009; Gori *et al*., 2012; Riovanto *et al*., 2011; Sinelli *et al*., 2010; Souto *et al*., 2010). When compared with SIMCA and PLS-DA, the LDA method has the disadvantage that the number of training samples must be larger than the number of variables included in the LDA model. In order to fully

satisfy this constraint a smaller number of variables were selected based on the standardized scoring coefficients obtained from the SLC-DA. The standardized scoring coefficients of the variable selected during the SLC-DA were obtained and plotted as shown in Figure 9.11. The importance of these coefficients lies in their use to compute canonical scores in terms of the standardized data often referred to as loadings. They are highly informative when it comes to comparing the relative importance in their discriminatory power of the independent variables. Tentative identification of the most informative peaks was carried out for baseline, detrend, quantile normalized, OSC Savitzky Golay and 1st order derived spectra which gave the most consistent discriminatory models during the LDA.

Table 9.6: Tentative identification of the most informative peaks obtained during the LDA.

Baseline	
926	$[OSL\text{-}2H_2O]^+$ or $[OOO\text{-}2H_2O]^+$
900	$[OLL\text{-}H_2O]^+$
884	$[OOO]^+$
787	$[OOO\text{-}H_2O\text{-}C_{14}]^+$
770	$[OOP\text{-}H_2O\text{-}C_5]$
762	$[OOO\text{-}2H_2O\text{-}C_{12}]$
687	$[PLn]+2K^+$
418	$[PP\text{-}C_{12}]^+$
118	Oleoside-11-methylester fragment
Quantile Normalisation	
787	$[OOO\text{-}H_2O\text{-}C_{14}]^+$
770	$[OOP\text{-}H_2O\text{-}C_5]$
323	Conidendrin fragment
340	Conidendrin fragment
107	Vanillic acid fragment
Detrend	
926	$[OSL\text{-}2H_2O]+$ or $[OOO\text{-}2H_2O]^+$
687	$[PLn]+2K^+$
418	$[PP\text{-}C_{12}]^+$
369	Open ring decarboxilade aldehydic ligstroside oxidized form
237	Open ring decarboxilade aldehydic oleoside oxidized form
127	Oleoside-11-methyl-ester fragment
OSC	
876	$[LLO\text{-}2H_2O]K^+$
852	$[OOO\text{-}H_2O\text{-}CH_2]$
699	$[OL\text{-}CH_2]2K^+$
652	β-methoxy verbascoside
150	Hydroxytyrosol
Savitzky Golay	
631	$[PL\text{-}H_2O]K^+$
550	Oleoside glycoside
454	$[mO\text{-}H_2O]^+$
129	Luteolin fragment
1st Derivative	
773	$[OOO\text{-}H_2O\text{-}C15]K^+$
706	$[OOO\text{-}H_2O\text{-}C_{11}]^+$
643	$[PO\text{-}H_2O]2K^+$
210	Closed ring carboxilade aldehydic form- CH_3OH
209	Oleoside fragment

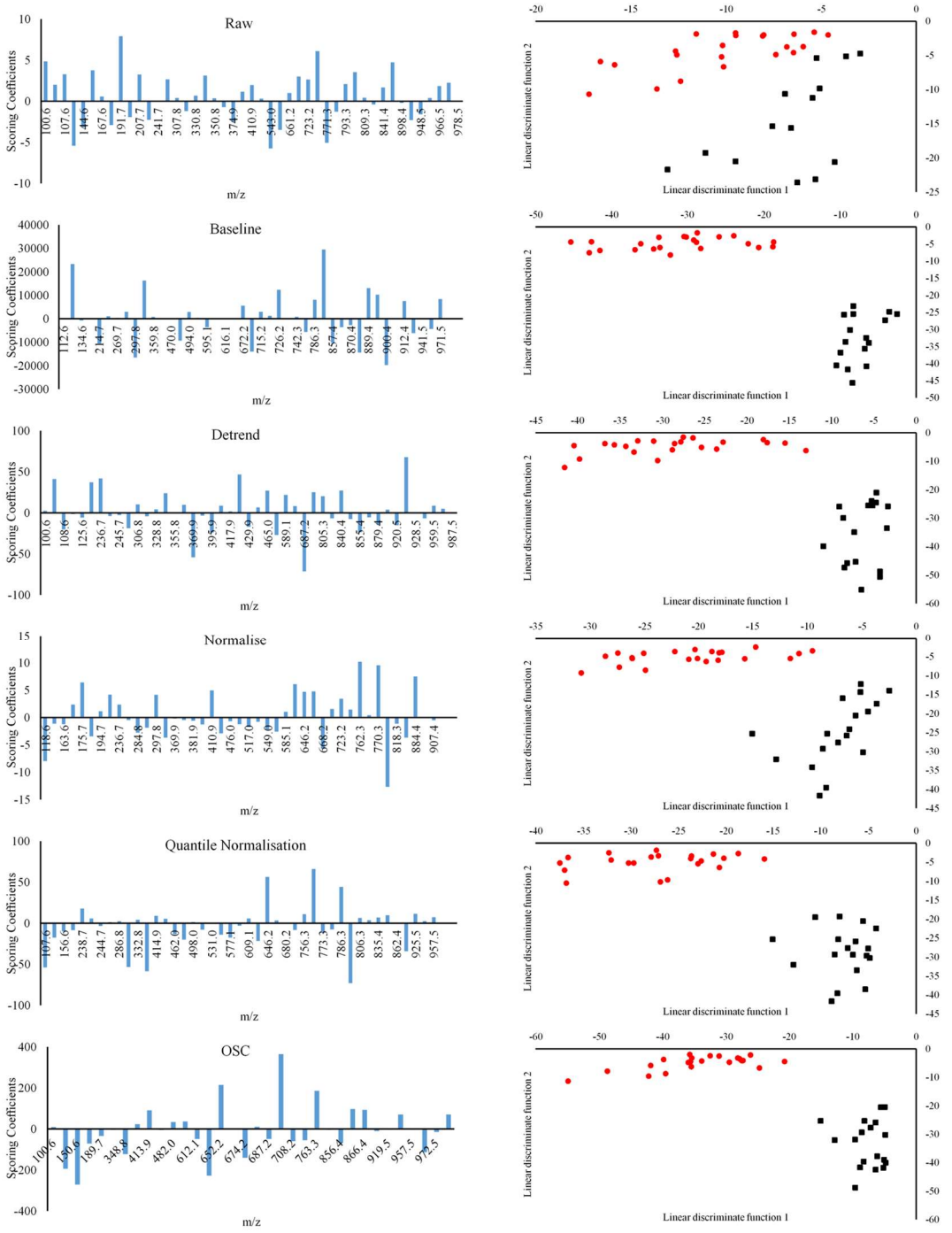


Figure 9.10:(Left) Bar graph showing the standardised scoring coefficients of the variables selected in the SLC-DA for the different spectra pretreatments, which were selected for LDA (Right).

From the Table 9.7, it was shown that during the training phase the LDA models obtained for all the pretreatments ranged from 87-100% in accuracy. The classification model obtained was then repeated on the testing data set, with the exception of raw data and MSC the validation accuracy ranged from 84-100%. From the results obtained it was shown that 1st order and Savitzky Golay derivatisation of the spectra had the highest % accuracy in validation data set (100% for both). Whilst during the PLS analysis and SIMCA both Savitzky Golay and 1st order derivatisation achieved appreciably good prediction models, the models obtained using normalized and OSC data from MS achieved better model performances. These result further confirm that model performance, is highly dependent on the number of the redundant variables and the technique which is used. LDA and PLS models obtained had a higher classification and discriminatory power when compared to SIMCA model. This difference stems from the fact that SIMCA is a soft classification method and thus allows a single object to belong to more than one class, whilst LDA, PLS are a more robust form of classification as the objects are either classified in either one or the other class.

9.2.5 Supervised discriminate chemometric techniques – SVM

The dataset containing only variables which were selected using SLC-DA and having a VIP score > 0.8, were subjected for another classification method, known as support vector machine (SVM). SVM similar to PLS can be used for both classification and regression (Christianini and Shawe-Taylor 2000; Vapnik 1995). During this part of the experiment, SVM was used in the context of classification. Comparison of SVMs with other classification and regression methods found out that they show mostly good performances, although other methods proved to be very competitive (Meyer *et al*. 2003). Table 9.7 shows the results obtained on using SVMs using no Kernel tricks.

Table 9.7: Comparison of the model performance between LDA and SVM

Classification DI-ES+ Mass Spectrometry VIP>0.8 & SLC-DA Dataset				
Pre-treatment	LDA		SVM	
	% Accuracy	% Predictability	% Accuracy	% Predictability
Raw	100.00	76.92	100.00	92.31
Normalized	100.00	92.31	100.00	46.15
Quantile normalized	100.00	92.31	100.00	92.31
Baseline	100.00	84.62	100.00	84.62
Detrend	100.00	92.31	100.00	76.92
Deresolve	97.83	84.62	100.00	92.31
SNV	95.65	84.62	100.00	84.62
MSC	86.96	76.92	100.00	76.92
OSC	100.00	92.31	100.00	92.31
Savitzky Golay	97.83	100.00	100.00	92.31
1st Derivative	97.83	100.00	100.00	76.92
2nd Derivative	100.00	92.31	100.00	92.31

Results obtained from SVM classification are presented in Table 9.7. High rates of accuracy and predictability were obtained for the majority of the spectral pretreatments further validating that SVM classification is highly adaptable to the kind of data used. In the case of SVM, the best classification was obtained using, raw data, quantile normalized data, detrend, OSC, Savitzky Golay, and 2nd order derived spectra with a training accuracy of 100% and a validation predictability of 92%. The trend observed was very similar to both the LDA and PLS models. Similar to LDA and PLS models SVM is also affected by small N large P problem of data sets thus the similarity between the different modelling methods can be attributed to the use of the same starting that set. Furthermore, since no obvious difference was observed when the three different methods were compared, these methods seemed to be equally effective in discriminating the Maltese EVOOs from the non-Maltese EVOOs.

9.2.6 Application of hierarchical cluster analysis

Cluster analysis is a simple yet powerful method to organize samples in the dataset into clusters based on the similarity of their abundance mass profiles. The method follows an agglomerative approach in which the most similar abundance mass profiles are joined together to form a group. These are further joined in a tree structure until all data forms a single group. Application of hierarchical cluster analysis using Ward's method was used on the variables selected using SLC-DA and VIP>0.8 for raw data.

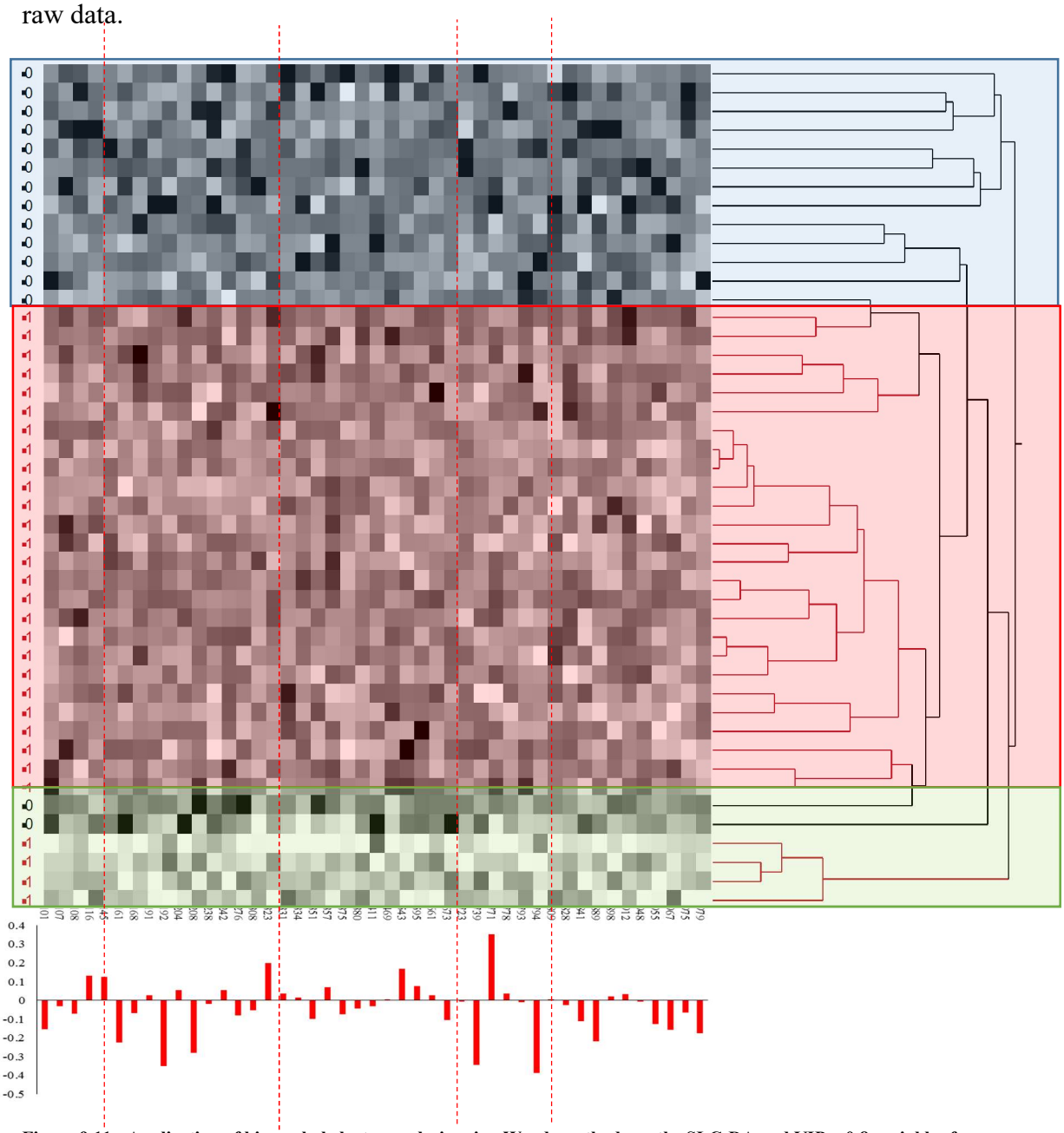


Figure 9.11:. Application of hierarchal cluster analysis using Wards methods on the SLC-DA and VIP > 0.8 variables for the raw DI-MS data. The dendrogram shows three major clusters (Blue) which contained exclusively EVOOs of Maltese origin (0) whilst the Blue and Green cluster contained mainly EVOOs on non-Maltese origin (1). (Top) The β coefficient scores obtained from the PLS-DA analysis of the selected variables.

Cluster analysis using the raw MS data using only the EVOOs derived from monocultivars revealed that very few (2 samples) of the EVOOs of Maltese origin were classified incorrectly. The dendrogram obtained showed the presence of 3 major clusters (marked, Red, Blue, and Green), the blue cluster contained exclusively EVOO's of Maltese origin, whilst the red clusters contained almost exclusively samples of non-Maltese origin, and the green cluster which contained very few samples was composed of both Maltese ('Bajda' and some of the 'Malti') and non-Maltese EVOOs. The standardized β coefficients obtained from PLS-DA analysis are also shown in Figure 9.11. These regression coefficients for the standardized data provide information about the contribution of each predictor, small coefficients make a small contribution to the response prediction. From the Figure 9.11 it was shown that certain regions of the spectra are highly variable between the different classes, such variables would contribute more (terms of β coefficient magnitude) to the prediction formula. The most informative ranges which enable the highest discrimination between EVOOs of Maltese origin from those of non-Maltese origin include 161-323 m/z and the 723-809 m/z which correspond to molecular masses and fragments of minor constituents present EVOOs and diglyceride and triglyceride compounds. This suggests that even though minor unsaponifiable constituents are present in very small amounts, not only the DI-MS was sensitive enough to detect them but also they offered a significant contribution in the discrimination of the Maltese EVOOs.

9.2.7 Whole MS modelling using feed-forward predictive artificial neural networks.

The use of feed-forward predictive neural networks on the MS data as a method for classification was assessed using three different forms of validation, namely 33.3% of data holdback, CV-10 k-fold and excluded row validation. For all the different forms of cross-validation 5 folds were used for 25 iterations, the best models obtained are presented in Table 9.8. The algorithm fitted on the training set was later tested on the validation data and % predictability of the model was obtained. Although PNN was employed as a classification for the geographical origin using FTIR and NMR data, the method employed the use of back propagation artificial neural networks (BP-PNN) rather than feed-forward predictive artificial neural networks (FF-PNN). Backpropagation (BP) works on the principle that after the information has gone through the network in a forward direction and an output has been produced, the error associated with this output is reorganized backwards through the model and weights are adjusted accordingly (Ham and Kostanic, 2001).

Table 9.8 shows % accuracy and % predictability for the different forms of cross-validation. Similar to what was observed in the PLS-DA and SIMCA, spectra treated using MSC had the lowest accuracy and predictability. Conversely the results obtained from the normalized, deresolved baseline corrected and OSC had the highest rates of accuracy and predictability when compared to the other spectral pretreatments over the different forms of cross-validation used. Comparison to the PLS-DA models obtained without any form of variable selection (Table 9.1), FF-PNN had a lower performance especially when it comes to the testing phase, and this observation was similar to the FF-PNN obtained using FTIR data. The lower % precision recorded in the FF-PNN, this observation is coherent to a number of other studies which showed that PLS-DA has a higher sensitivity and performance (Khanmohammadi *et al*., 2011; Efstathios *et al*., 2011; Sampson *et al*., 2011) when compared to PNN. The form of cross-validation had a great effect on the performance of the model obtained especially during the validation step. Holdback and k-fold validation methods are random when compared to the excluded row validation, thus testing models obtained using these forms of cross-validation tend to be overfitted namely due to the small sample size which was used in this part of the study.

Overfitted models tend to have a very good performance during the training phase but they can fail to predict the origin of the sample during the testing phase. The use of excluded row validation, on the other hand, is not random as the experiments defines the testing and training the group thus the model built can be adjusted to cover an equal distribution of samples which belong to different geographical origins in so doing improving not only the models obtained during the training step but also increasing the predictability of the model obtained. in the case of the excluded row validation, it was found that spectra obtained using only raw data had the best model performance reaching a 100% accuracy and predictability none of the other spectral pretreatments reached this level of performance. This is attributed to the higher sensitivity of the FF-PNN, which were able to fully recognize the signal from the instrumental noise, furthermore, although spectral transformations which enable the removal noise including normalization procedures and baseline corrections might have removed minor peaks which are informative than the larger more abundant peaks causing a reduced model performance. On the other hand, derivatisation procedures were found to negatively affect the MS data in the majority of the models obtained not only in FF-PNN, this is mainly attributed to the mixing of noise and signal making it hard for sample discrimination.

Overfitted models tend to have a very good performance during the training phase but they can fail to predict the origin of the sample during the testing phase. The use of excluded row validation, on the other hand, is not random as the experiments defines the testing and training the group thus the model built can be adjusted to cover an equal distribution of samples which belong to different geographical origins in so doing improving not only the models obtained during the training step but also increasing the predictability of the model obtained. in the case of the excluded row validation, it was found that spectra obtained using only raw data had the best model performance reaching a 100% accuracy and predictability none of the other spectral pretreatments reached this level of performance. This is attributed due to the higher sensitivity of the FF-PNN, which were able to fully recognize the signal from the instrumental noise, furthermore, although spectral transformations which enable the removal noise including normalization procedures and baseline corrections might have removed minor peaks which are informative than the larger more abundant peaks causing a reduced model performance. On the other hand, derivatisation procedures

were found to negatively affect the MS data in the majority of the models obtained not only in FF-PNN, this is mainly attributed to the mixing of noise and signal making it hard for sample discrimination.

Table 9.8: % accuracy and % predictability obtained using PNN under different forms of cross-validation.

PNN DI-ES+ Mass Spectrometry Whole Dataset						
CV Type Pre-treatment	Hold back		K-fold		Excluded Row	
	% Accuracy	% Predictability	% Accuracy	% Predictability	% Accuracy	% Predictability
Raw	69.57	66.67	93.48	66.67	100.00	100.00
Normalized	93.48	77.78	97.83	100.00	97.83	100.00
Quantile normalized	82.61	44.44	91.30	55.56	97.83	88.89
Baseline	89.13	77.78	89.13	55.56	91.30	100.00
Detrend	82.61	44.44	93.48	66.67	93.48	66.67
Deresolve	93.48	100.00	97.83	88.89	97.83	88.89
SNV	84.78	66.67	95.65	77.78	95.65	77.78
MSC	91.30	66.67	89.13	55.56	89.13	55.56
OSC	89.13	100.00	95.65	77.78	95.65	77.78
Savitzky Golay	95.65	77.78	95.65	77.78	93.48	66.67
1st Derivative	100.00	100.00	86.96	44.44	89.13	44.44
2nd Derivative	89.13	77.78	93.48	66.67	93.48	77.78

9.3 Application of DI-MS ESI+ for higher order geographical classification

The use of SLC-DA/PLS variable selection was also applied to a more detailed classification of origin, one which is based on the individual country of origin. Preliminary results shown in Figure 9.13 revealed that the variable selection method employed was able to fully discriminate between different samples of based on their producing country. The canonical discriminant analysis showed very high eigenvalues, and by the 3rd canonical function, the model was able to explain 85.85% of the variability with a Wilk's lambda value of 3.39×10^{-7}.

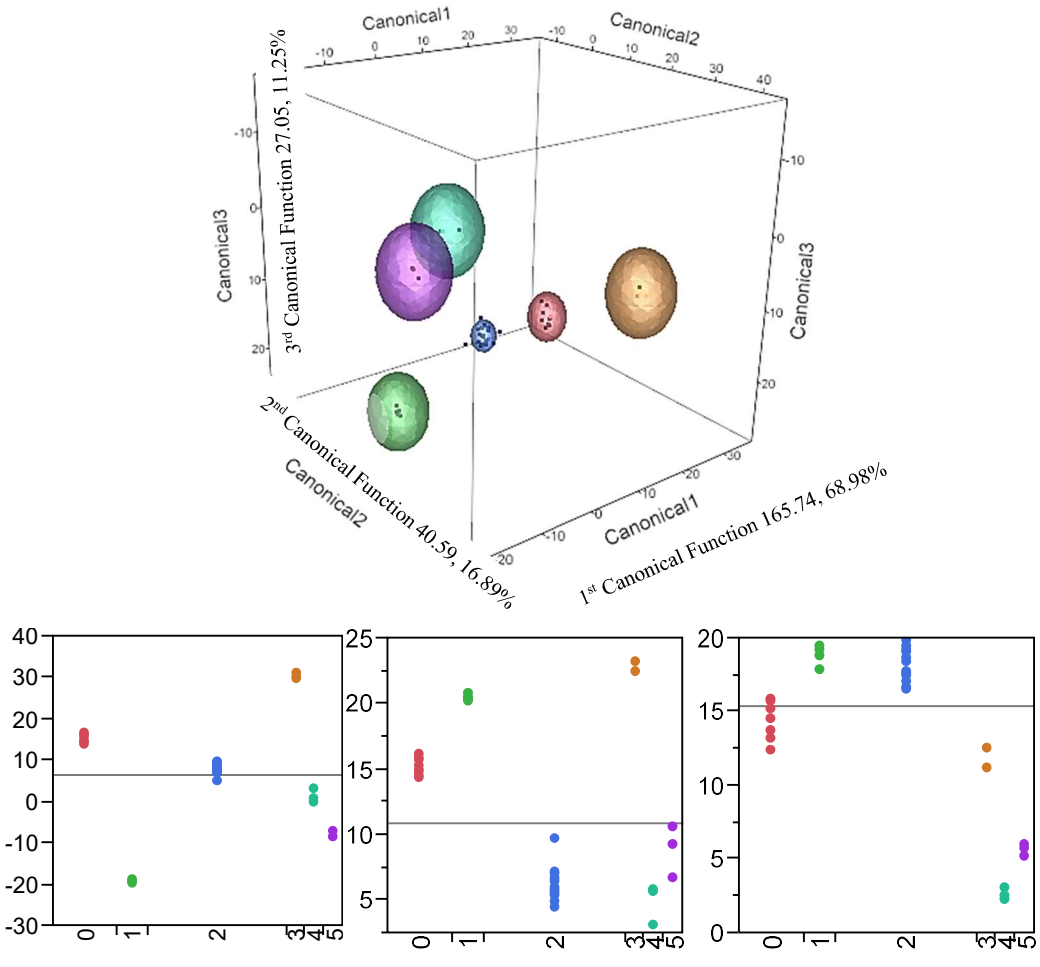


Figure 9.12: (Top) SLC-DA 3D-plot obtained using specific variables. The model obtained was able to classify the olive oils by their country of origin; Maltese indigenous (Red/0); imported cultivars (Green/1); Italian origin (Blue/2); Greek origin (Orange/3); French origin (Turquoise/4) and Spanish origin (Black/5). (Bottom) X-Y fit using the canonical scores obtained for the 1st 2nd and 3rd discriminate function showing that by the 1st function the model was able to almost fully separate olive oils based on their country of origin.

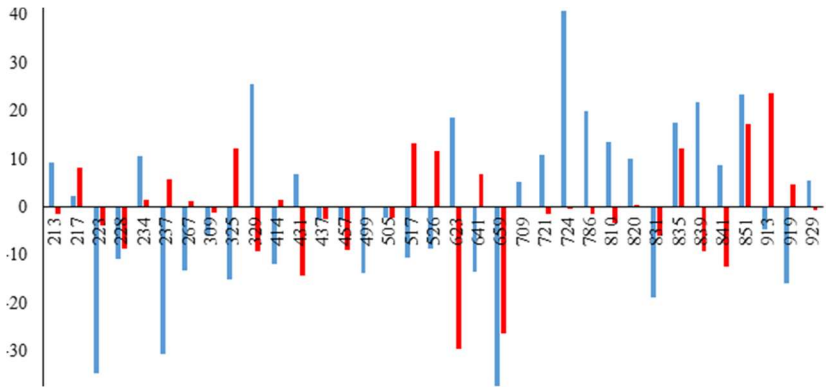
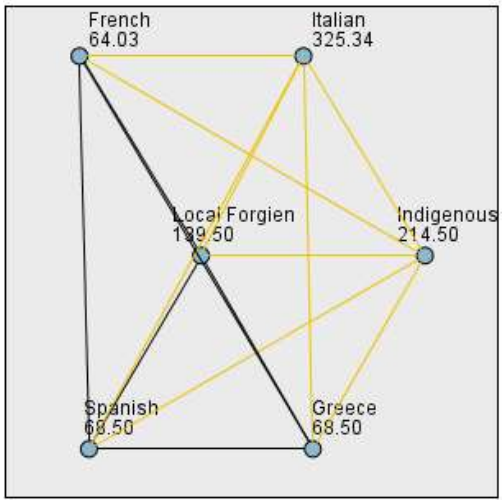


Figure 9.13: Score plot for specific variables (m/z). Blue bars represent the scores obtained in the 1st discriminate function and red bars represent the scores obtained in the 2nd discriminate function.

From the canonical score plot it was observed that the indigenous cultivars were projected very near to cultivars which were derived from Greece and Crete, although the geographical discrimination between the two was possible in the 3rd canonical function, the close proximity between the two countries could provide insights into the history and origin of the indigenous cultivars found in the Maltese islands. Although it is thought that the Phoenicians were the first settlers to introduce olive cultivation within the Maltese islands, very few is known about the presence of olive trees prior that period. Although Malta was never a Greek colony the archaeological records show the introduction of Hellenistic features in architecture and pottery. In the addition, the Greek influence can be further be joined to the name of the island Melite a derivative of the Greek word for honey. In order to provide a definite answer for the geographical origin of the indigenous cultivars samples derived from Middle East are required, nonetheless, it was shown that the indigenous cultivars are not related to the Italian, France and Spanish cultivars, as these were completely discriminated by the 1st canonical function.

Application of One-way ANOVA using Tukey post hoc hypothesis testing on the extracted Mahalanobis distances from the canonical biplot, it was shown that there was a significant difference between the spatial distributions in the geographical origin of the EVOOs. Figure 9.14 shows the p-values obtained for all the pairwise comparisons, the results showed that on the basis of the previously selected m/z values, EVOOs derived from the indigenous cultivars were significantly different from EVOOs derived for other neighbouring countries (France, Italy, Greece and Spain) and from locally grown foreign cultivars.



Comparisons	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adjusted Sig.
Greece-Italian	256.84	24.62	10.43	0.000	0.000
Spanish-Indigenous	146.00	26.34	5.54	0.000	0.000
Greece - Indigenous	146.00	26.33	5.54	0.000	0.000
Spanish - Italian	256.84	24.62	10.43	0.000	0.000
French- Italian	261.30	24.62	10.61	0.000	0.000
Local Foreign - Italian	-185.84	19.82	-9.38	0.000	0.000
Italian - Indigenous	-110.84	15.26	-7.26	0.000	0.000
French- Indigenous	150.47	26.34	5.71	0.000	0.000
Local Foreign - Indigenous	75.00	21.91	3.42	0.001	0.009
French- Local Foreign	75.48	29.22	2.58	0.010	0.147
Spanish - Local Foreign	71.00	29.22	2.43	0.015	0.227
Greece - Local Foreign	71.00	29.22	2.43	0.015	0.227
French- Greece	4.47	32.67	0.14	0.891	1.000
French- Spanish	-4.47	32.67	-0.14	0.891	1.000
Greece -Spanish	0.00	32.67	0.00	1.000	1.000

Figure 9.14: Kruskal-Wallis one-way ANOVA revealed a significant difference in the spatial distribution of observations based on the Mahalanobis distances. Significant difference for pairwise comparisons using post hoc Tukey hypothesis testing are drawn using yellow lines, whilst non-significant differences are shown as black lines.

With regards to the locally grown foreign cultivars, it was shown they were only significantly different from EVOOs derived from Italian countries, whilst no significant difference was found when compared to EVOOs derived from France, Spain and Greece. This observation further highlights the effects of the terroir on the EVOOs. EVOOs derived from locally grown Italian cultivars namely Carolea, Frantoio and Pendolino tend to be significantly different from EVOOs of the same cultivar grown in Italy. Pairwise comparison between EVOOs derived from Spain, Greece, and France showed no significant difference between each pairwise combination, however, this was attributed to the small sample size which was used in order to represent the country of origin. Although a small sample size was used from the aforementioned countries, these were still significantly different from EVOOs derived from indigenous cultivars, indicating that in this case, the differences observed were due to genetic variations within the cultivars. Potential of DI-MS for predicting the genetic variety from which the EVOO was derived was first discussed by Lerma-García *et al*., (2008). In their study genetic varieties of Spanish extra virgin olive oils (Arbequina, Hojiblanca, and Picual) were predicted by direct infusion of the samples in the electrospray ionization source of a mass spectrometer, followed by linear discriminatory analysis. In this study, the same sample preparation method employed by Lerma-García *et al*., (2008) was adopted in order to assess the possibility of predicting and discriminating between EVOOs of different geographical origins. The application of different variable selection techniques revealed that the use of SLC-DA was the most applicable for the extraction of discriminate variables. Coupling of variables extracted by both SLC-DA and PLS showed an improvement on the prediction model generated.

9.3.1 Identification of discriminatory peaks for higher order discrimination

Peak identification was not the primary objective of the study and it was only applied to m/z values which had both a PLS VIP>0.8 and a significant p-value in the SLC-DA. Peaks were identified on the basis of their molecular weight by comparison with the existing literature. Due to MS conditions which were employed, fragmentation of the parent molecular ion was kept to the minimum and fragments smaller than 200 m/z were not considered, however, smaller ions will be considered at a later stage in the study is required. Spectra were recorded under both ESI negative and ESI positive, however, ESI negative mass spectra were dominated by the deprotonated forms of oleic, 281 m/z, and palmitic, 255 m/z which represent the major components of EVOOs. The presence of these dominate peaks reduced the discriminatory power of the analysis. Although Lerma-García *et al*., 2008 employed ESI negative the subsequent data analysis required further processing in order to account for the dominant fatty acid peaks. Similar to the method employed by Alves *et al*., 2010, in this study only positive ion ESI-MS was used in order to yield fingerprints that could distinguish between the geographical origins of EVOOs.

The peaks which were identified can be divided into four major classes, phenolic compounds, tocopherols, fatty acids and their corresponding glycerides. Table 9.9 illustrates the identified peaks and their corresponding molecular ion from. In general, the molecular ion form of free fatty acids and their corresponding tri and diglycerides was observed as their corresponding potassium salt derived from the addition of 40mM KOH to the EVOOs during sample preparation. In the case of phenolic compounds, since most of the compounds present in EVOOs do not readily form potassium salts these were observed as the corresponding parent molecular ion subject to the loss of water. The presence of an aromatic ring and hydroxyl groups which characterize the phenolic structure favours the loss of water from the parent molecular ion in order to yield a positive radical cation. In the case of tocopherols, the molecular ion was observed as the hydride loss molecular fragment. The presence of both an aromatic group and the extended conjugation in the tocopherol molecule highly favours the loss of hydrides from neighbouring saturated carbons in order to allow rearrangement reactions driven by the stability of the extended conjugated radical cation which is formed.

The most informative peaks can be divided into four different mass ranges protonated triacylglycerols potassium salts (TAGs, m/z above 800), diacylglycerols (DAGs, m/z range of 600–650), free fatty acids and simple phenols (m/z range 200-300), and tocopherols and complex phenolic compounds (m/z range 400-600). The most informative peaks extracted from the 1st canonical discriminate function (68.98% of the total variation) corresponded to m/z 223, attributed to the open ring decarboxilade dialdehidic form of elenolic acid, minor free fatty acids including lauric acid m/z 237, stearic acid 329 m/z together with triglyceride compounds namely PPLn Palmitoleic and Linolenic acid (835 m/z), OOP Oleic Palmitoleic triglyceride (851 m/z), and OOO Oleic acid triglyceride (919 m/z). In the case of the 2nd canonical discriminate function oleuropein aglycon, decarboxymethyl dialdehydic form (3,4 DHPEA-DEDA 623 m/z) and the diglyceride of oleic acid (659 m/z) were found to be the most discriminant compounds.

From Table 9.9 a number of minor compounds present within the unsaponifiable fraction of EVOO were selected, these include; tyrosol acetate; acetoxypinoresinol; ligstroside glycoside derivative and oleuropein aglycon decarboxymethyl dialdehydic form (3, 4 DHPEA-DEDA) as potential compounds that enabled the discrimination of EVOOs on the basis of their geographical origin. These results further corroborate the results obtained during the analysis of phenolic extracts by means of HPLC, whereby it was shown that it was shown that both tyrosol acetate and a number of secoiridoid compounds had a high discriminatory potential. The results reported in this study are similar to those reported by Lerma-García *et al.,* 2008, whereby the presence of a dialdehydic form of deacetoxy ligstroside and 10-hydroxyoleuropein aglycone had a significantly higher predictive power in the discrimination of three Spanish cultivars.

Table 9.9: The identification most discriminate m/z values which were identified during variable selection their corresponding parent molecular ion form.

m/z	Name	Molecular ion form
213	Shikimic acid	$[M+K^+]^+$
217	Caprylic acid	$[M+K^+]^+$
223	Open ring decarboxilade dialdehydic form of elenolic acid	$[M+K^+]^+$
228	Eleanolic acid fragment	$[M+K^+]^+$
234	Tyrosol acetate	$[M+K^+]^+$
237	Lauric acid	$[M+K^+]^+$
267	Myristic acid	$[M+K^+]^+$
309	Closed ring carboxilade form elenolic acid	$[M+K^+]^+$
325	Linoleic acid	$[M+K^+]^+$
329	Stearic acid	$[M+K^+]^+$
414	βγ- tocopherol	$[M-H]^+$
431	α-tocopherol	$[M-H]^+$
437	Acetoxypinoresinol	$[M+K^+-H_20]^+$
499	Und1	
505	Ligstroside glycoside derivative	$[M-H_20]^+$
517	Loganic acid glucoside	$[M-H_20]^+$
526	Oleuropein glycoside derivative	$[M-H_20]^+$
623	Oleuropein aglycone decarboxymethyl dialdehydic form (3,4 DHPEA-DEDA)	$[M-H_20]^+$
641	OO oleic acid diglyceride	$[M+K^+-H_20]^+$
659	OO oleic acid diglyceride	$[M+K^+]^+$
709	Phosphatidic acid	$[M+K^+]^+$
721	Steroidal acid ester	$[M+K^+]^+$
724	Octanedioic ester	$[M+K^+-H_20]^+$
810	PPL Palmitoleic and linoleic acid triglyceride	$[M-H_20]^+$
820	PPP Palmitoleic triglyceride	$[M+K^+-H_20]^+$
831	PPLn Palmitoleic and Linolenic acid triglyceride	$[M+K^+]^+$
835	SSM Stearic acid triglyceride	$[M+K^+]^+$
839	PPP Palmitoleic triglyceride	$[M+K^+]^+$
851	OOP Oleic Palmitoleic triglyceride	$[M+K^+]^+$
919	OOO Oleic acid triglyceride	$[M+K^+]^+$
929	SSS Stearic acid triglyceride	$[M+K^+]^+$

Und1: unidentified compound assigned as a possible oleuropein or ligstroside glycoside derivative or

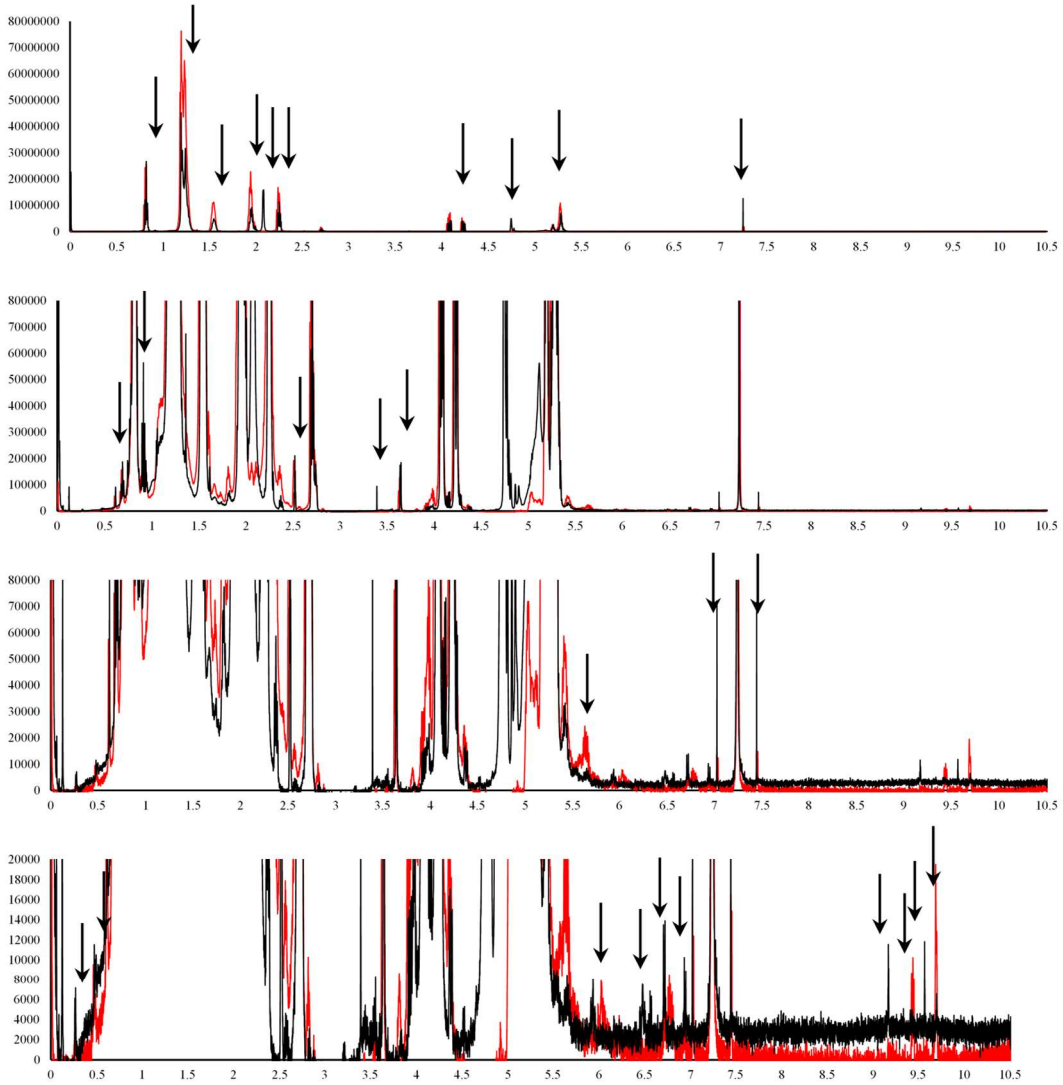
In conclusion, the application of DI-MS fingerprinting in conjunction with multivariate statistical techniques provided a simple, rapid and accurate way to discriminate EVOOs of Maltese origin from non-Maltese EVOOs. From the preliminary assessment using only unsupervised PCA models, significant clustering was observed in the majority of the spectral pretreatments, however, the % variation explained by the first two principal components was very low this was attributed to the high levels of redundant variables. Application of supervised methods of classification namely PLS-DA, FF-PNN, LDA, and SVM were highly effective in classifying local and non-local EVOOs samples. The discrimination of power of the different models obtained was greatly enhanced through the use of a two-stage variable selection procedure. The discrimination of geographical origin using DI-MS under positive ionisation has never been achieved although it was proposed by a number of different authors in the past, including Alves *et al*., 2010; Catharino *et al*., 2005; Lerma-García *et al*., 2008 and 2011; Goodacre *et al*.,2002, however it was never carried out. The combination of variable selection techniques namely PLS and SLC-DA proved to be ideal statistical techniques which enable the extraction of discriminate m/z values. The application of DI-MS ESI (+ve) enabled the extraction of discriminate variables belonging to different chemical classes without the need to preliminary extraction.

10. Application of Proton Nuclear Magnetic Resonance (^{1}H-NMR) spectroscopy for the determination of the geographical origin of olive oil.

During the last decade, nuclear magnetic resonance spectroscopy (NMR) has been shown to be highly effective in the study of properties of oils of vegetable origin (Frankel, 2010; Cañabate-Díaz *et al*., 2007; Murkovic *et al*., 2007; Suárez *et al*., 2008 and Morales *et al*., 2010. The main methods used in NMR include the ^{1}H and ^{13}C NMR spectroscopy as reviewed by Sacchi *et al*., (1997) Vlahov (1999), Guillen *et al*., (2001), Hidalgo *et al*., (2003) and Mannina *et al*.,(2003 & 2010) together with ^{31}P NMR as Dais *et al*., (2010). Apart from target-based analytical approaches, NMR metabolic fingerprinting (Rezzi *et al*., 2005; Alonso-Salces *et al*., 2010 and Longobardi *et al*., 2011) employs the use of whole ^{1}H spectral data to classify a relevant number of samples according to their origin, harvesting, and ageing. In the majority of cases, fingerprinting analysis is used in conjunction with computer-aided sophisticated statistical and mathematical procedures. Although these methods offer a fast, reliable measure for the determination of origin, such methods have been criticized for their complexity by Frankel (2010) due to the necessity of many samples for statistical elaborations and the ensuing difficulty in interpretation of the results.

The combination of ^{1}H NMR fingerprinting with multivariate analysis provides a promising approach to study the profile of olive oils in relation to its geographical origin. The aim of this study was to use two different pulse sequence acquired ^{1}H NMR spectroscopy (^{1}H zg30 and ^{1}H NOESY) in conjunction with chemometrics in order to differentiate the Maltese EVOO's from other EVOO's derived from other countries within the Mediterranean region. In this study, the spectroscopic data were processed both by a discriminant chemometric tools including PLS, SVM, and LDA but also using modelling chemometric tools such as SIMCA and PNN. Moreover, different forms of signal pretreatment were employed in order to enhance the potential of ^{1}H NMR as a tool for authentication purposes. The present work aims at using the NMR profile of olive oils taken as a whole in combination with multivariate statistics to discriminate Maltese EVOOs from EVOOs of various Mediterranean areas. A methodological approach based on the high throughput acquisition of NMR profiles statistically processed in a holistic way in order to provide information regarding the geographical origin was therefore developed.

10.1 ^{1}H NMR peak identification



0.298	-CH_2-(cyclopropanic ring)	4.531	Terpene
0.542	-CH_2-(cyclopropanic ring)	4.779	Terpene
0.687	-CH_3 (C18-steroid group)	5.283	>**C**H**O**COR (glyceryl group)
0.812	-C**H**$_3$ (acyl group)	5.666	Und-Tocopherol
1.197	-(C**H**$_2$)n-(acyl group)	5.915	-C**H**=C**H**-C**H**=C**H**-(cis, trans conjugated dienediene system)
1.548	-OCO-CH_2-C**H**$_2$-(acyl group)	5.955	-C**H**=C**H**-C**H**=C**H**-(cis, trans conjugated dienediene system)
1.955	-C**H**$_2$-CH=CH-(acyl group)	6.563	-C**H**=C**H**-C**H**=C**H**-(cis, trans conjugated dienediene system)
2.085	-C**H**$_2$-CH=CH-(acyl group)	6.724	-Ph-**H** (phenolic ring)
2.265	CH-C**H**$_2$-CH= (acyl group)	6.954	-Ph-**H** (phenolic ring)
2.547	CH-C**H**$_2$-CH= (acyl group)satellite	7.028	Chloroform ^{13}C satellite
2.71	CH-C**H**$_2$-CH= (acyl group)	7.249	Chloroform
2.746	CH-C**H**$_2$-CH= (acyl group)	7.446	Chloroform ^{13}C satellite
3.395	Und-alcohol	9.172	Unk-hydrocarbon
3.694	-C**H**$_2$OCOR (glyceryl group)	9.568	Unk- hydrocarbon
4.101	-C**H**$_2$OCOR (glyceryl group)	9.571	Unk-hydrocarbon
4.227	-C**H**$_2$OCOR (glyceryl group)	9.703	hexanal

Figure 10.1: The major peaks of interest obtained using NMR of EVOOs using zg30 pulse sequence (Black) and NOESY pulse sequence (Red)

The advantage of ^{1}H NMR compared to other spectroscopic techniques is that it provides sharp spectral signals which are a direct reflection of the different chemical environment of the hydrogen isotope. The chemical shifts of the distinct spectral lines are related to the chemical environment of the hydrogen atoms, whilst the intensities of the lines directly correspond to the number of hydrogens (Guillén *et al*., 2001; Mannina and Segre, 2002). Although the quantitative analysis carried out using NMR has been extensively studied, especially regarding the quantification of fatty acids (Sacchi *et al*., 1997; Sacchi *et al*., 1996), the main interest, of this study was to apply chemometric methods using the entire NMR signal. The observed major chemical shifts can be expected to be attributed to triglycerides as these constitute up to 98% of the fraction. These differ in their substitution patterns in terms of length, degree, and kind of unsaturation of the acyl groups resulting in different chemical shifts (Harwood & Aparicio, 2000). The major chemical signals obtained for EVOOs included 4.541 ppm 3.988 ppm 3.636 ppm 2.710 ppm 2.746 ppm 1.244 ppm 1.197 ppm 0.910 ppm and 0.843 ppm which arise from the following functional groups having different number of protons, respectively: methylenic protons in α-glycerol moiety of sn-1,3 diglycerides; methylenic protons in α-glycerol moiety of sn-1,2 diglycerides diallylic protons of linolenic fatty chains; diallylic protons of linoleic fatty chains; methylenic protons of all unsaturated fatty chains; methylenic protons of palmitic and stearic fatty chains; methyl of linolenic fatty chains; methyl of linoleic fatty chains; (Alonso-Salces *et al*., 2010a; Alonso-Salces *et al*., 2010b; Sacchi *et al*., 1996; Mannina and Segre, 2002; Sacco *et al*., 2000).

Whilst the chemical shifts of their ^{1}H signals for the major constituents are well known and very easily identified. The ^{1}H signals of the minor oil components, are only observed by ^{1}H-NMR when their signals do not overlap with those of the main components and their concentrations are high enough to be detected (Guillen & Ruiz, 2001). Minor constituents which are expected to give NMR signals include; mono- and diglycerides, sterols, tocopherols, aliphatic alcohols, hydrocarbons, fatty acids, pigments and phenolic compounds (Harwood & Aparicio, 2000), Figure 10.1 shows the most common ^{1}H-NMR signals of the major and some minor compounds together with their chemical shifts and their assignments to protons of the different functional groups.

Several NMR signals of minor compounds can be observed in ^{1}H-NMR spectra since they did not overlap with those of the triglyceryl protons. The main identified compounds include; cycloartenol at 0.318 ppm and 0.543 ppm, β-sitosterol at 0.669 ppm, stigmasterol at 0.687 ppm, waxes at 0.978 ppm, squalene at 1.662 ppm, sn-1,2 diglyceryl group protons at 3.71 ppm and 5.10 ppm, and three unknown terpenes at 4.571 ppm, 4.648 ppm and 4.699, hexanal at 9.704 ppm and phenolic protons at 5.73, 5.99, 6.55, and 6.75 ppm. These compounds have been as already observed and identified by other authors (Alonso-Salces, Heberger *et al*., 2010; Alonso-Salces, Moreno-Rojas *et al*., 2010; D'Imperio *et al*., 2007; Guillen & Ruiz, 2001; Mannina, Sobolev, & Segre, 2003; Sacchi *et al*., 1996).In the study carried out by Alonso-Salces *et al*., 2010 on the different fractions of unsaponifiable constituents of EVOOs for the determination of geographical origin gave a highly detailed NMR analysis for each unsaponifiable fraction. The fractions studied included the alcohol fraction, sterol fraction, tocopherol fraction, and hydrocarbon fraction each one of them had a distinct chemical signal shown in Table 10.1.

Table 10.1 Chemical shifts observed in $_{1}$H NMR for isolated minor EVOO fractions

Fraction	Chemical Shift (ppm)
Alcohols	0.141, 0.333, 0.38-0.40, 0.55-0.57, 0.558, 0.615, 0.715, 0.747, 0.97, 1.01-1.04, 3.10-3.17, 3.26-3.33, 3.284, 3.293, 3.315, 3.641, 3.723, 4.157, 4.310, 4.65–4.76, 5.256, 7.05, 9.1.
Sterols	0.529, 0.556, 0.683, 0.702, 0.77-0.78, 0.826, 0.834, 0.848, 0.91- 0.94, 0.921, 0.934, 1.009, 1.80-1.88, 2.27-2.31, 2.336, 2.470, 3.49-3.58, 4.771, 4.821, 4.964, 5.30-5.43, 6.92.
Tocopherols	0.156, 0.881, 0.894, 2.131, 2.603, 3.302, 3.330, 3.416, 3.433, 4.099, 4.840, 4.875, 4.836, 4.872, 4.938, 5.53-5.63, 5.988, 6.367
Hydrocarbons	0.884, 1.431, 2.269, 3.287, 3.662, 4.060, 4.539, 4.690, 4.98-5.01, 5.020, 5.703, 5.745, 5.87– 5.98, 6.439, 6.975, 9.365, 9.37–9.41, 9.762.

10.2 Application of chemometrics to NMR Spectra

As previously stated the aim of this study was to build reliable classification models for the traceability of EVOOs from Malta by coupling NMR spectroscopic techniques and chemometrics. To this purpose, NMR spectra of olive oil samples from the Maltese islands and from other Mediterranean countries were collected as described in Section 2.2.1 and analyzed as described in Section 2.2.6.3. In order to obtain spectral fingerprints corresponding to the origin of the sample, discriminant (PLS, LDA) and modelling (SIMCA, PNN) classification approaches were used for two pulse sequences (zg30 and NOESY) and compared.

10.2.1 Unsupervised chemometric techniques – PCA

Different kinds of spectral pretreatments were tested and compared in order to overcome the instrumental limitation and account for scattering and other minor variations which would hinder the performance of the classification models. A total of 12 spectral pretreatment methods were used, in each case, after pretreatment a principal component analysis was carried in order to dimensionally reduce the number of variables into a small set of principal component whilst retaining the information of the larger set. PCA enabled the preliminary identification of which pre-treatment method offered the highest variability and possible clustering. Appendix 16 shows the different forms of spectral pretreatments employed and the corresponding PCA plot for the first two principal components for the ^{1}H NMR acquired using zg30 and NOESY respectively. Results showed a significantly high total explained % variability ranging from 73-83% for the zg30 pulse sequence and slightly lower 69-79% for the NOESY pulse sequence for first two principal components for the pretreated spectra. For the zg30 spectral deresolvation and 2^{nd} order, derivatisation had the highest % variability explained whilst for the NOESY experiment the multiple scatter correction of spectra seemed to improve the total % variability. Whilst the some of the spectral pretreatment methods displayed an improvement in the variability explained from the binned raw data, normalization, detrending, orthogonal signal correction (OSC) and standard normal variate (SNV) showed a lower % variability when compared to the raw data for the zg30 and NOESY experiment. .

Although clustering was observed in the majority of the spectral pretreatments, the obtained clustering did not fully discriminate between the EVOO's of Maltese origin from those obtained for other Mediterranean countries. Only a weak clustering resembling the geographical origin was observed on using PCA. From PC loading (Figure 10.2) it was observed that the chemical shifts observed at 0.8 and 1.2-1.25 ppm for the zg30 and 0.5-1.25 ppm for the NOESY experiment seem to have a larger influence on the first and second principal component separation. These observations suggest that the phytosterol content namely β-sitosterol, campesterol, cyscloartenol together with 1-eicosanol and α-tocopherol which show chemical shifts between 0.5-1.25 ppm have a greater influence on the variation observed along the first two principal components. These signals of minor compounds were found in ^{1}H-NMR spectra recorded because they did not overlap by those of the triglyceryl protons 2-2.4 ppm which contributed far less to the variation observed in the two PC. In the case of zg30 other peaks observed in the 4.7-4.9 ppm range seem to be also influential especially in the 1st PC, these peaks correspond to terpenic compounds present in EVOOs. Alonso-salces *et al*., (2012) identified 3 peaks at 4.571 ppm, 4.648 ppm and 4.699 ppm which were attributed to unknown terpenes during their study on the unsaponifiable fraction of EVOOs.

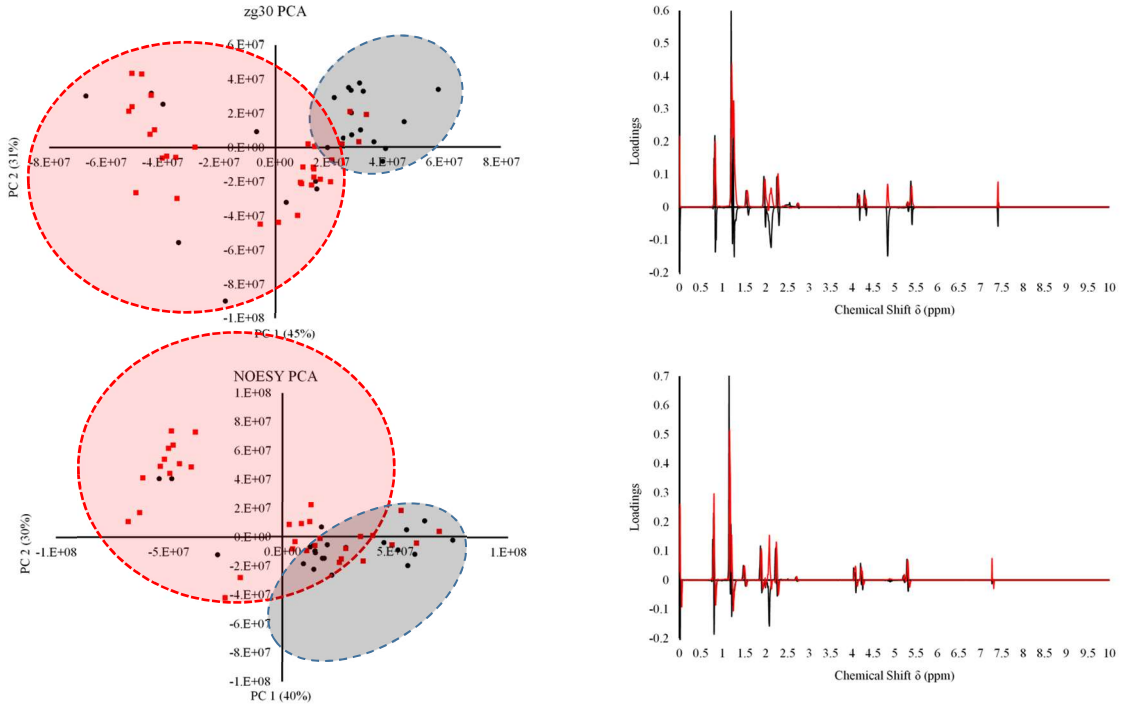


Figure 10.2: PCA biplots (black dots = Maltese red boxes = non-Maltese) and loading plots for PC1 (black line) and PC2 (red line) for the untreated raw data for the zg30 (Top) and NOESY (bottom) NMR spectra.

10.2.2 Supervised chemometric techniques – PLS-DA

In the subsequent step the whole dataset was divided into training and test sets (the former to build the model, the latter to validate it). In order to preserve the diversity in the training and test sets and to account for the fact that different pretreatments had to be tested a unique sample splitting scheme was required. The following method was adopted in order to cover as such variation in the two sets and at the same time being able to compare the outcomes after the different pretreatments. The Maltese and the non-Maltese samples were grouped in an ascending way so that the first 35 samples would represent Maltese EVOO's whilst the rest corresponded to non –Maltese EVOO's. A stratified random sampling method was used in order to exclude 20% of the observation so that they would be retained as the testing set. The remaining 80% of the observation were used to build the training set.

After splitting the data according to the procedure described above, chemometric classification models were built and tested on all the NMR spectral pretreatment using a PLS regression algorithm using JMP 10 and its inbuilt leave one out cross-validation method (LOOCV). Table 10.2 and Table 10.3 show the number

of latent variables extracted, the predicted root mean square error and the % variation explained in terms of X and Y for the different spectral pretreatment methods for the zg30 and NOESY pulse sequence respectively. In the case of the zg30, it was observed that there was a significant drop when the % accuracy (correct classification) between the use of the entire dataset using only LOOCV (Top) and using the 20% cross-validation method. These observations suggest that the model obtained using only LOOCV tended to be overfitted and thus the 20% CV is deemed to be the most reliable method for cross-validation, similar results were also observed in the NOESY experiment.

For the zg30 NMR spectra obtained after detrending, SNV and quantile normalization showed the best model performance with a % accuracy ranging from 93.1-87.9 % and % predictability ranging from 72.2 -81.8%, whilst for the NOESY experiment spectra treated using quantile normalization, orthogonal signal correction, and 1st derivative showed the best performance with an accuracy 77.5-94.8% and predictability of 83.3%. In the case of the zg30 experiment all the spectral pretreatments showed an improvement in the % predictability when compared to the raw data, whilst in the NOESY experiment spectra treated using SNV, MSC and detrending functions showed a lower % predictability and % accuracy when compared to actual non-pretreated raw data. This observation suggests that in the case of NOESY the signal suppression of the major peaks improves the signal to noise ratio, the resulting spectra obtained can be used without the need of extensive pretreatments. Results obtained by Longobardi *et al*., (2012) showed that the presaturation of the dominating lipid signals resulted in increased receiver gain which in turn resulted in a signal-to-noise gain close to 10 compared to the zg30 spectra.

Table 10.2: PLS-DA analysis using the whole NMR data using zg30 pulse sequence. (Top) the results obtained using LOOCV on the training data set and (Bottom) the results obtained using LOOCV and 20% of the data as the validation group

zg30 Whole Spectrum Internal Validation					
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy
Raw data	5	72.00	62.81	0.99	89.66
Normalized	5	70.60	71.45	0.90	94.83
Quantile normalized	3	30.03	77.16	0.91	93.10
Detrend	3	84.52	25.78	0.97	74.14
Deresolve	6	77.32	67.45	0.97	87.93
SNV	1	74.83	4.47	1.10	58.62
MSC	4	84.63	59.48	0.92	89.66
OSC	5	66.18	70.06	0.95	89.66
1st Derivative	1	17.44	31.06	0.96	75.86
Savitzky Golay	8	41.04	96.48	0.97	98.28
2nd Derivative	1	18.94	28.98	0.97	72.41

zg30 Whole Spectrum External Validation						
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy	% Predictability
Raw	4	67.12	61.41	1.05	77.59	27.27
Normalized	9	82.31	94.76	0.85	91.38	63.64
Quantile normalized	4	39.71	91.28	0.89	93.10	72.73
Detrend	8	81.42	85.94	0.93	70.69	36.36
Deresolve	3	86.09	33.62	0.95	87.93	63.64
SNV	8	91.10	86.56	0.92	93.10	81.82
MSC	7	89.93	80.64	0.97	91.38	81.82
OSC	2	51.69	39.14	0.97	72.41	45.45
Savitzky Golay	1	16.92	33.01	1.00	74.14	54.55
1st Derivative	1	19.45	33.61	0.95	68.97	45.45
2nd Derivative	1	17.48	32.08	1.02	77.59	63.64

Table 10.3 PLS-DA analysis using the whole NMR data using NOESY pulse sequence. (Top) the results obtained using LOOCV on the training data set and (Bottom) the results obtained using LOOCV and 20% of the data as the validation group

NOESY Whole Spectrum Internal Validation					
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy
Raw	2	28.06	51.77	0.88	93.45
Normalized	2	31.95	48.59	0.91	84.48
Quantile normalized	6	50.12	93.64	0.89	98.28
Detrend	2	84.71	21.08	0.94	65.52
Deresolve	9	77.50	91.60	0.89	94.83
SNV	1	71.24	10.69	0.98	65.52
MSC	2	92.57	19.65	0.99	72.41
OSC	15	78.18	96.57	0.85	94.83
Savitzky Golay	2	19.36	77.23	0.87	96.55
1st Derivative	2	23.56	72.33	0.81	93.10
2nd Derivative	2	20.16	74.92	0.88	98.28

NOESY Whole Spectrum External Validation						
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy	% Predictability
Raw	7	66.13	92.76	0.92	93.10	75.00
Normalized	1	23.57	27.85	0.95	79.31	75.00
Quantile normalized	1	15.01	38.59	0.93	77.59	83.33
Detrend	2	85.01	26.68	0.94	65.52	58.33
Deresolve	9	79.19	93.04	0.87	89.66	75.00
SNV	1	71.00	11.60	0.98	68.97	66.67
MSC	2	92.86	20.93	1.00	65.52	58.33
OSC	7	65.43	93.02	0.87	94.83	83.33
Savitzky Golay	3	27.58	89.06	0.88	89.66	75.00
1st Derivative	6	49.03	95.22	0.85	94.83	83.33
2nd Derivative	4	37.67	91.67	0.91	89.66	75.00

In order to fully interpret the PLS models obtained, an inspection of the VIP scores was used in order to determine which predictors (variables) are mainly influencing the latent vectors obtained. VIP is an index of how much a single variable contributes to the bilinear model and it is scaled in such a way that indices having VIP larger than 0.8 are considered to be significantly contributing to discrimination. VIP scores > 0.8 for the PLS models built on the differently pretreated NMR data are reported in the first column of Figure 10.5 and Figure 10.6 which for the zg30 and NOESY experiment respectively.

As shown in the Figure 10.3 and in Appendix 16, the VIP>0.8 identified relevant features in the spectra, particularly those centred around the major peaks which were identified in Section 10.2. Furthermore the VIPs also highlight regions in the spectra in which unless magnified no peaks are observed, especially the regions 0-0.9, 2-3.8 ppm, 4-4.4 ppm, 5-6.9 ppm and 7.3-10 ppm, which corresponds to minor unsaponifiable constituents found in EVOOs. Alonso-Salces *et al*., (2015) identified a number of unsaponifiable EVOO constituents which show chemical shifts in the aforementioned regions including: terpenes (4.609 ppm, 4.648 ppm, 4.694 ppm), phenolic compounds (3.487 ppm, 3.855 ppm, 3.950 ppm, 4.125 ppm, 4.55–4.61 ppm, 5.851 ppm, 6.515 ppm, 6.559 ppm, 6.89 ppm, 6.95 ppm, 7.03 ppm, 7.09– 7.21 ppm, 7.165 ppm, 7.185 ppm, 7.32 ppm, 7.53 ppm, 7.69–7.73 ppm, 7.81 ppm, 7.84 ppm), aldehydes (6.02–6.18 ppm, 8.025 ppm, 9.383 ppm, 9.539 ppm, 9.739–9.755 ppm, 9.581 ppm, 9.762 ppm, 9.845 ppm, 9.849 ppm, 9.853 ppm, 9.875 ppm, 9.962 ppm, 9.999 ppm); sterols such as cycloartenol (0.333 ppm, 0.558 ppm, and 0.974 ppm), β-sitosterol (0.826 ppm, 0.834 ppm, 0.848 ppm, 0.921 ppm, and 0.934 ppm), stigmasterol (0.702 ppm).

This step was then followed byd another PLS model this time using only variables which had a VIP score > 0.8. Table 10.4 and 10.5 shows the results obtained on using the adjusted PLS model for the zg30 and NOESY experiments respectively. On comparing the models obtained using variable selection to that previously obtained without any variable selection, an improvement in the overall % accuracy and predictability of the model, was obtained indicating that although binning of data was carried out in order to remove redundant variables the data set obtained still contained an appreciable amount of uninformative variables. These variables were omitted once the corresponding VIP scores from the preliminary PLS model was obtained and the

data set trimmed in such a way that only variables having a VIP score larger than 0.8 were selected. The models obtained using only VIP>0.8 variables showed an increase in both the % X and % Y explained variation together with a lower PRESS indicating an enhanced model performance. In the case of the zg30 experiment it was found that normalized spectra and those pretreated using Savitsky-Golay provided the best results, whilst detrending and SNV which had the lower performance even when compared to the raw data. Furthermore the aforementioned spectral transformation showed a significant decrease in model performance when the VIP > 0.8 data set was used.

In the case of the NOESY experiment the models obtained using VIP > 0.8 showed an increase in the performance when compared to those obtained whole data, however similar to the zg30 experiment SNV and MSC which showed an optimal performance when the whole data set was used, showed a appreciably lower performance when only VIP > 0.8 dataset was used. These observations are indicative that different spectral pretreatments are affected differently to variable selection techniques since each one of them attempts to maximize spectral variations and corrections, thus removal of a small number of predictors can have a devastating effect on the model performance.

Table 10.4 PLS-DA analysis using the VIP>0.8 datasets. (Top) the results obtained using LOOCV on the training dataset. (Bottom) the results obtained using LOOCV and 20% of the data as the validation group

Zg30 VIP>0.8 Spectrum Internal Validation					
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy
Raw	6	78.46	71.20	0.93	93.10
Normalized	5	70.60	71.45	0.90	94.83
Quantile normalized	13	69.92	96.50	0.78	98.28
Detrend	4	92.38	33.51	0.95	72.41
Deresolve	6	77.85	66.83	0.94	87.93
SNV	3	91.69	57.33	0.91	86.21
MSC	4	66.15	60.47	0.91	91.38
OSC	4	57.94	65.13	0.89	89.66
Savitzky Golay	6	52.17	95.16	0.71	98.28
1st Derivative	3	42.95	79.69	0.78	96.55
2nd Derivative	4	45.97	88.43	0.77	98.28

Zg30 VIP>0.8 Spectrum External Validation						
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy	% Predictability
Raw	4	65.34	62.41	0.94	82.76	45.45
Normalized	4	65.80	71.14	0.97	94.83	90.91
Quantile normalized	12	70.18	99.99	0.77	94.83	72.73
Detrend	2	76.70	33.90	0.91	68.97	36.36
Deresolve	4	62.92	59.40	0.98	82.76	45.45
SNV	1	90.28	8.72	1.00	60.34	36.36
MSC	1	33.18	16.33	1.00	67.24	63.64
OSC	10	83.45	93.88	0.91	94.83	72.73
Savitzky Golay	7	55.48	99.61	0.64	98.28	90.91
1st Derivative	8	57.90	99.87	0.66	94.83	72.73
2nd Derivative	12	63.58	100.00	0.65	93.10	63.64

Table 10.5 PLS-DA analysis using the VIP>0.8 datasets. (Top) the results obtained using LOOCV on the training dataset. (Bottom) the results obtained using LOOCV and 20% of the data as the validation group

NOESY VIP>0.8 Spectrum Internal Validation					
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy
Raw	9	71.79	95.73	0.78	98.28
Normalized	2	42.38	50.34	0.87	81.03
Quantile normalized	6	58.62	93.51	0.66	98.28
Detrend	4	88.49	53.30	0.90	86.21
Deresolve	15	88.33	96.42	0.78	94.83
SNV	2	91.52	25.90	0.94	74.14
MSC	1	54.89	12.47	0.96	70.69
OSC	9	72.59	95.37	0.74	100.00
Savitzky Golay	5	50.07	93.77	0.64	98.28
1st Derivative	7	61.79	95.78	0.61	98.28
2nd Derivative	3	40.88	81.89	0.68	98.28

NOESY VIP>0.8 Spectrum External Validation						
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy	% Predictability
Raw	12	80.34	99.96	0.71	93.10	75.00
Normalized	2	91.72	25.85	0.95	94.83	83.33
Quantile normalized	6	60.23	93.17	0.73	96.55	91.67
Detrend	4	93.77	42.07	0.94	74.14	66.67
Deresolve	13	87.78	95.54	0.77	94.83	83.33
SNV	7	73.50	92.57	0.88	70.69	75.00
MSC	1	58.34	12.21	0.96	68.97	75.00
OSC	6	62.91	91.55	0.79	96.55	83.33
Savitzky Golay	4	47.89	91.78	0.63	89.66	75.00
1st Derivative	6	59.23	94.99	0.60	93.10	91.67
2nd Derivative	3	43.05	84.49	0.67	93.10	91.67

10.2.3 Modelling Chemometric techniques – SIMCA

The classification analysis of the EVOO NMR data was then repeated using a modelling approach based on the SIMCA algorithm. The latter is a class modelling algorithm that allows the analyses of one class at a time. For SIMCA analysis two PCA models were built for each spectral pretreatment one for the EVOOs of Maltese origin and the other for the EVOOs of non-Maltese origin. In each case, the optimal number of principal components were chosen in order to obtain optimal model complexity in 10-fold row-wise cross-validation. The use of a two-stage model, one for each category, allowed the comparison between the two categories, thus this permitted the control as to check whether samples are accepted by one, both or none of the modelled classes. The output of SIMCA analysis was assessed by the use of Coomans plot shown in Figure 10.3 and in Appendix 16 for the zg30 and NOESY experiment respectively.

For this experiment given that the NMR data was highly similar and no significant difference in variance was observed (PCA analysis revealed no significant clustering resembling the origin of EVOOs) the significance limit was increased up to 25% rather than the default 5%. These lines (blue) of significance divided the Coomans plot into four different regions: the uppermost left and the lowermost right corresponded to unmistakeable acceptance by a single category model, the lowermost left to acceptance by both classes while the uppermost right to rejection by both category models. From a preliminary survey of the Coomans plot outcome at the 25% level of significance, very few samples showed an unambiguous acceptance by a single category.

The SIMCA analysis classified the majority of the EVOO's in the lowermost left part suggesting that the samples were accepted by both classes and thus failing to discriminate adequately between the two classes. Thus a diagonal (green) line bisecting the plot corresponding to new classification was built in order to represent a new significant boundary so that all the samples lying above are classified as being Maltese, while samples lying below are predicted as from other origins. Table 10.6 and Table 10.7 show the results obtained on using SIMCA modelling chemometric techniques for the classification and discrimination of EVOOs.

Table 10.6: Results obtained for SIMCA modelling for the different spectral pretreatments using the only the variables having a VIP > 0.8 for the zg30 NMR. The values recorded in the table represent the % sensitivity of the different models towards the two classes using the new classification decision boundary.

Application of SIMCA on the VIP>0.8 Data Set for the zg30 NMR			
	% Sensitivity Maltese EVOOs	% Sensitivity Non-Maltese EVOOs	Overall % Accuracy
Raw	76.0	100.0	88.0
Normalized	76.0	88.9	82.4
Quantile normalized	88.0	33.3	60.7
Detrend	100.0	38.9	69.4
Deresolve	88.0	50.0	69.0
SNV	100.0	88.9	94.4
MSC	92.0	38.9	65.4
OSC	92.0	38.9	65.4
Savitzky Golay	100.0	38.9	69.4
1st Der	92.0	33.3	62.7
2nd Der	100.0	44.4	72.2

Table 10.7: Results obtained for SIMCA modelling for the different spectral pretreatments using the only the variables having a VIP > 0.8 for the NOESY. The values recorded in the table represent the % sensitivity of the different models towards the two classes using the new classification decision boundary.

Application of SIMCA on the VIP>0.8 Data Set for the NOESY NMR			
	% Sensitivity Maltese EVOOs	% Sensitivity Non-Maltese EVOOs	Overall % Accuracy
Raw	88.0	66.7	77.3
Normalized	80.0	77.8	78.9
Quantile normalized	80.0	80.6	80.3
Detrend	N/A	N/A	N/A
Deresolve	N/A	N/A	N/A
SNV	100.0	88.9	94.4
MSC	84.0	50.0	67.0
OSC	100.0	61.1	80.6
Savitzky Golay	100.0	33.3	66.7
1st Der	96.0	61.1	78.6
2nd Der	100.0	33.3	66.7

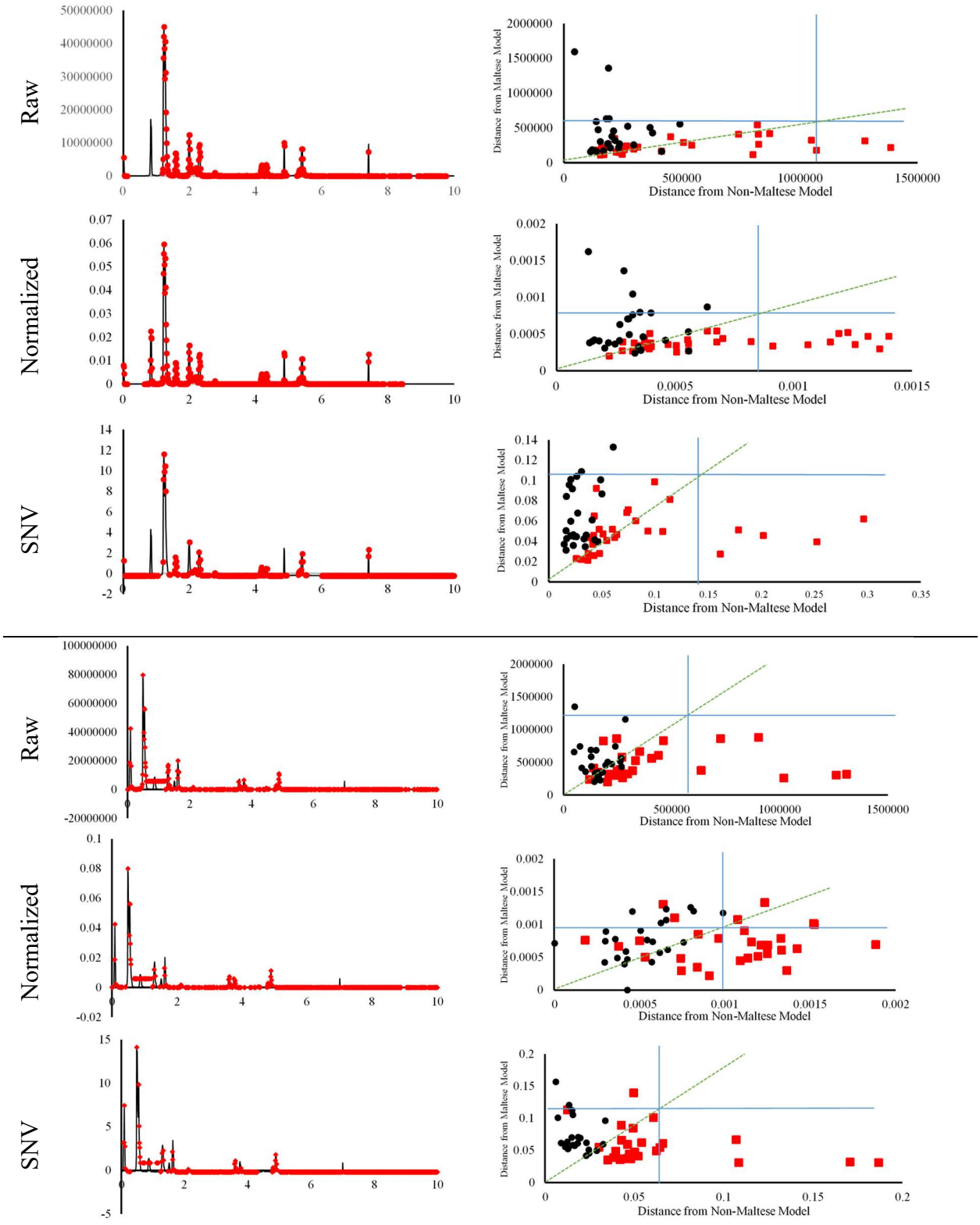


Figure 10.3:(Left) Variables having a VIP score > 0.8 (red dots) for the different spectra pretreatments carried out on zg30 (Top) and NOESY (Bottom) methods. (Right) Coomans plot observed for the corresponding selected variables. The blue dotted lines represent the 25% confidence level whilst the green dotted represent the new classification decision boundary employed

On comparison of the results obtained from using modelling SIMCA to discriminate PLS-DA models obtained, it was found that in general, the SIMCA model had a significantly lower performance when compared to the discriminate models. Although the new classification decision boundary was used, which in its self-provides an overestimation of the model performance, certain spectral pretreatments still failed to fully distinguish between the two classes. It is well known that SIMCA modelling tends to fail when it comes to distinguishing classes in which the data is very similar. Similar results were obtained by Alonso-Salces *et al*., (2012) where it was shown that although good sensitivities were obtained by SIMCA the models obtained had a very low specificity. This can be explained in terms of the fundamental concepts of SIMCA. This technique employs the use of two PCA one for each class, which aims to identify the directions of maximum variability for each class. Since from the preliminary PCA analysis, very few pretreatment models showed signs of clustering based on NMR variability reflecting the geographical origin. Therefore, the hyper-ellipsoids modelled by SIMCA for each class were oriented in similar directions and overlapping, which results in low specificities as explained in Alonso-Salces *et al*., (2012).

NOESY spectra obtained using detrending and deresolve functions were completely overlapping and thus no separation between the different classes was observed. From the results obtained using Standard Normal Variate (SNV), spectral pretreatment method seemed to be the most effective method which enabled an overall accuracy of 94.4% for both the zg30 and NOESY NMR experiments. This can be explained in terms of the actual pretreatment, SNV effectively removes scatter effects by centring and scaling each individual spectrum. The practical difference between SNV and MSC is that each spectrum is standardized using only the data from that spectrum and not the mean spectrum obtained from the entire data set, thus variations belonging to each homologous class are conserved and amplified resulting in a higher variability between the two groups. The increased variability between the two classes enables a more accurate modelling of each class when it comes to SIMCA. The same logic can be used to explain the overall higher % accuracy of the raw data which was observed in the zg30 NMR experiment when compared to the rest of the other spectral pretreatments. Whilst for discriminate analysis obtaining an overall averaged spectrum seemed to be beneficial, the conservation of the sample variation seems to be more beneficial when it comes to modelling with SIMCA, as this variation enables each

class to be modelled separately from the other aided by the increase variation between each sample, which altogether increase the variation of the class.

10.2.4 Application of SLC-DA for variable selection

In order to obtain a more robust method of classification with the use of a smaller number of variables, the VIP data set obtained from the previous analysis was subjected to a stepwise linear canonical discriminate analysis SLC-DA. SLC-DA has performed on both the zg30 and NOESY NMR data from all the pretreatment methods in order to extract only a small amount of highly discriminate variables which would enable an easier and faster discrimination between the origins of EVOOs. This strategy involved a substantial reduction of the dimensionality of the data in such a way that only the variables shown in Figure 10.4 and those in Appendix 16 Section 16.2 were retained. In order to further reduce the number of variables selected from the SLC-DA analysis, a minimum of 14 variables was selected in order to carry out a conventional LDA. During the SLC-DA the variables chosen by applying a forward stepwise variable selection algorithm using JMP 10 using a Wilks' Lambda as a selection criterion and an F-statistic factor to determine the significance of the changes in Lambda when the influence of a new variable is evaluated. The most significant variables were then extracted and their canonical scoring coefficients were plotted as shown in Figure 10.5 and 10.6. The main advantage of using SLC-DA over the convention LDA is the ability to perform a feature selection. Regarding this fact, only those variables which helped to improve classification performance were used whereas variables without discriminant information were discarded.

Figure 10.4 shows that the variables selected during SLC-DA for some of the spectral pretreatments obtained from the zg30 NMR experiment. It was observed that NMR spectra pretreated using; normalization, SNV, MSC and OSC, the majority of the variables selected using the stepwise algorithm seem to be concentrated between 3-7 ppm. This suggests that for these spectral pretreatments peaks corresponding to alcohols, glyceryl groups, terpenes, tocopherols, cis, trans conjugated dienediene system and phenolic compounds. These compounds seem to have a significant effect on the linear discrimination of the two classes. Similarly, NOESY NMR spectra derived from MSC and detrending functions seem also to have a high proportion of variables which are centred in the aforementioned region. On the other hand, spectra obtained from the zg30 experiment derived after detrending and 2^{nd} derivation seem to have a large portion of the selected variables in the 8-10 ppm region which corresponds to hydrocarbons and aldehydes, hexanal in particular. Whilst quantile normalization, deresolvation, 1^{st} and Savitzky Golay derivation of the zg30 experiment, significant variables were picked up along all the chemical shifts.

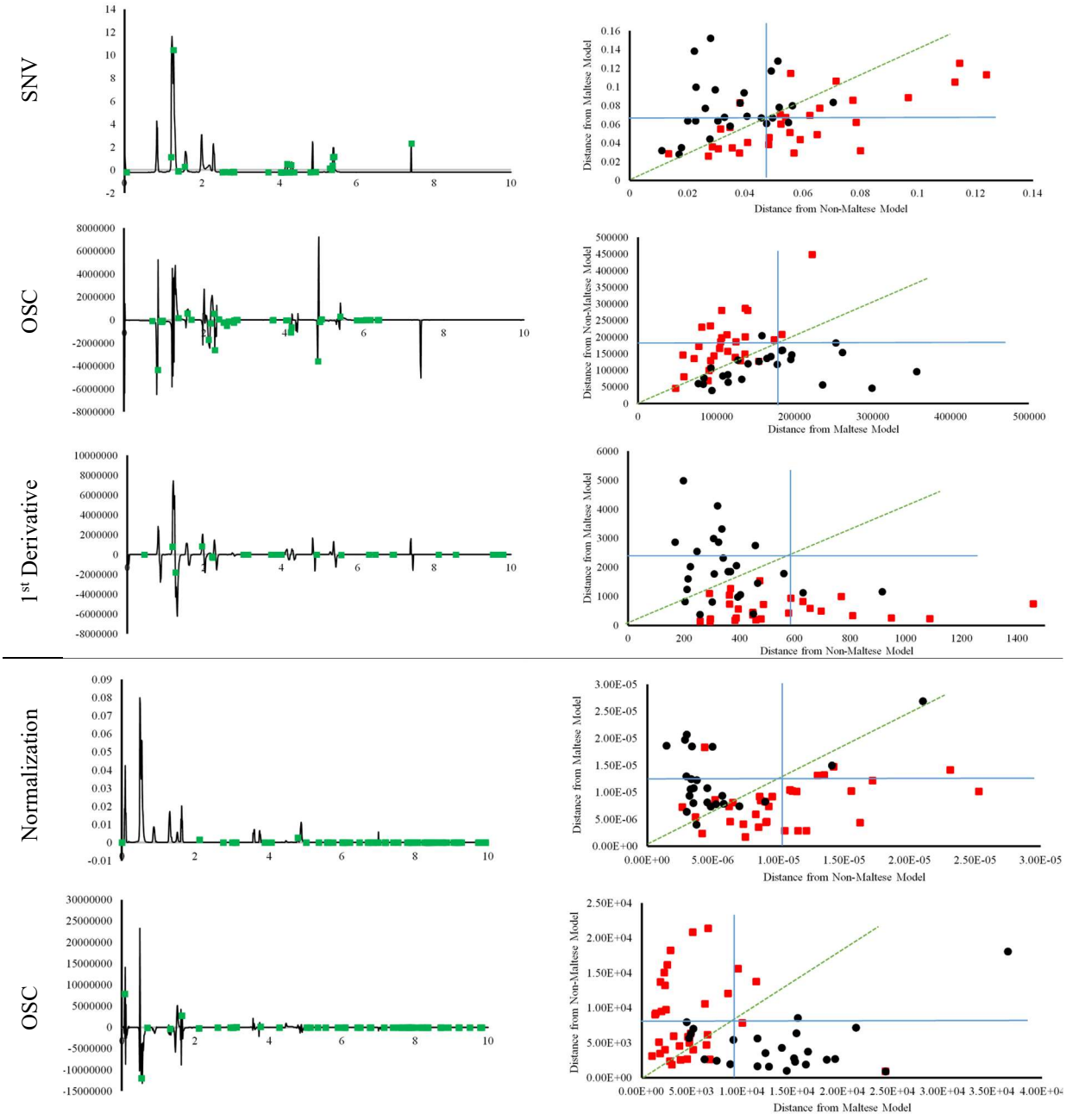


Figure 10.4:(Left) variables which had both a VIP score > 0.8 and selected during the SLC-DA (Green dots) for the different spectra pretreatments carried out on zg30 (Top) and NOESY (Bottom) methods. (Right) Coomans plot observed for the corresponding selected variables. The blue dotted lines represent the 25% confidence level whilst the green dotted represent the new classification decision boundary employed

Table 10.7: PLS-DA analysis using the VIP>0.8 datasets. (Top) the results obtained using LOOCV on the training dataset. (Bottom) the results obtained using LOOCV and 20% of the data as the validation group

zg30 VIP>0.8 & SLC-DA Spectrum Internal Validation					
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy
Raw	6	83.43	80.24	0.67	100.00
Normalized	15	97.25	93.21	0.6	100.00
Quantile normalized	7	59.83	93.27	0.58	100.00
Detrend	15	99.92	90.01	0.89	100.00
Deresolve	9	90.59	85.24	0.65	100.00
SNV	10	92.22	84.39	0.77	98.28
MSC	4	65.17	61.76	0.83	91.38
OSC	9	72.59	95.37	0.74	100.00
Savitzky Golay	15	73.15	95.22	0.41	100.00
1st Derivative	12	73.91	94.27	0.41	100.00
2nd Derivative	13	58.88	95.27	0.45	100.00

zg30 VIP>0.8 & SLC-DA Spectrum External Validation						
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy	% Predictability
Raw	15	97.21	97.66	0.49	91.38	54.55
Normalized	15	97.62	98.09	0.58	98.28	90.91
Quantile normalized	7	60.46	97.79	0.57	98.28	90.91
Detrend	3	72.99	41.38	0.88	74.14	36.36
Deresolve	9	91.57	88.96	0.74	96.55	81.82
SNV	15	97.83	93.19	0.65	68.97	54.55
MSC	1	26.70	25.74	0.98	93.10	63.64
OSC	1	26.31	40.49	0.87	74.14	54.55
Savitzky Golay	15	76.64	99.77	0.26	98.28	90.91
1st Derivative	15	83.52	99.55	0.33	98.28	90.91
2nd Derivative	15	68.04	99.89	0.27	98.28	90.91

Table 10.8 PLS-DA analysis using the VIP>0.8 datasets. (Top) the results obtained using LOOCV on the training dataset. (Bottom) the results obtained using LOOCV and 20% of the data as the validation group

NOESY VIP>0.8 & SLC-DA Spectrum Internal Validation					
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy
Raw	5	60.88	88.13	0.65	100.00
Normalized	9	77.32	92.71	0.78	96.55
Quantile normalized	5	48.30	90.93	0.55	100.00
Detrend	15	99.62	82.82	0.75	98.28
Deresolve	15	88.77	93.52	0.71	100.00
SNV	6	91.76	67.49	0.77	93.10
MSC	10	91.17	78.59	0.88	98.28
OSC	15	88.52	96.10	0.47	96.55
Savitzky Golay	11	56.83	95.43	0.52	100.00
1st Derivative	6	44.73	92.13	0.56	100.00
2nd Derivative	8	55.06	90.72	0.81	100.00

NOESY VIP>0.8 & SLC-DA Spectrum External Validation						
Pre-treatment	Latent Variables	% X	%Y	PRESS	% Accuracy	% Predictability
Raw	10	75.36	94.01	0.68	94.83	75.00
Normalized	2	57.35	50	0.86	84.48	75.00-
Quantile normalized	6	51.65	92.25	0.61	93.10	83.33
Detrend	15	99.78	83.97	0.79	96.55	91.67
Deresolve	13	86.64	93.67	0.72	96.55	83.33
SNV	4	88.19	57.74	0.85	89.66	91.67
MSC	1	39.75	17.25	0.96	70.69	83.33
OSC	6	68.91	88.28	0.69	100.00	100.00
Savitzky Golay	3	31.15	87.05	0.59	100.00	100.00
1st Derivative	4	37.89	88.8	0.64	100.00	100.00
2nd Derivative	2	31.72	65.96	0.86	87.93	66.67

Once the variables were selected for each spectral pretreatment another PLS-DA and SIMCA models were carried out in order to determine whether variable selection using the linear method provided a better form of classification. Table 10.9 and Table 10.10 show the results obtained from the PLS using the data set composed of variables which had a VIP score > 0.8 and were selected during the SLC-DA analysis, for the zg30 and NOESY NMR spectra respectively. The results show that with the exception of spectra treated with MSC for the zg30 NMR data, normalization and 2nd order derivation for the NOESY NMR data all the models obtained from PLS-DA showed a noticeable improvement in the % accuracy and predictability. Notably, the spectra derived after orthogonal signal correction, Savitzky Golay and 1st derivation of the NOESY NMR data reach 100% accuracy in the training model and 100% predictability in the validation dataset. Although improvements were observed in the zg30 spectra none of the spectral pretreatment methods managed to achieve such high levels of performance. These observations suggest that a variable selection using the two techniques greatly improves the modelling power and that NOESY NMR data enables a much more accurate discrimination of the Maltese EVOOs. Similar to what was observed in other spectral techniques which were carried out during this study the PLS-DA models obtained by combining two different types of variable selection procedures ensured a higher number of extracted latent variables resulting in a higher % of variance explained in terms of X and Y and marked decrease in the PRESS.

Application of SIMCA to the data set containing common predictors which had VIP >0.8 and significant effect on the SLC-DA model showed that for the zg30 experiment an improvement in the classification was observed when compared to the VIP>0.8 datasets. On the contrary in the case of NOESY experiment an evident decrease in the classification power was observed when compared to the VIP>0.8, as none of the spectral pretreatments were able to model the two classes separately, thus the Commas plot obtained showed indistinguishable overlapping classes. In the case of NOESY since the NMR data obtained had an increase signal to noise ratio, the removal of a significant number of variables could reduce the ability of the modelling methods to fully discriminate the two classes.

These observations suggest that the application of modelling methods for classification highly depends on a number of variables and can differ substantially from discriminating models as observed in the case of the PLS-DA. In the study carried out by Galtier *et al*., (2011), it was shown that PLS-DA discrimination was better than SIMCA in classification performance for the classification of EVOOs according to their geographical origin. Galtier *et al*., (2011) further state that the main difference between SIMCA and PLS-DA is the criterion which is employed during the model building. In the case of SIMCA PCA submodels are computed with the ultimate the goal of capturing variations within each class, on the other hand, PLS-DA tries to identify the directions in the data space that discriminate classes directly. Due to these fundamental differences SIMCA classification always provides worse results than methods based on PLS analysis. SIMCA performance may be improved when the spectral region is reduced to a small number predictors (Galtier *et al*., 2011), in this experiment it was shown that it also highly dependent on the type of spectral data.

Table 10.9: Results obtained for SIMCA modelling for the different spectral pretreatments using the only the variables having a VIP > 0.8 for the zg30 NMR. The values recorded in the table represent the % sensitivity of the different models towards the two classes using the new classification decision boundary.

Application of SIMCA on the VIP>0.8 Data Set for the zg30 NMR			
	% Sensitivity Maltese EVOOs	%Sensitivity Non-Maltese EVOOs	Overall % Accuracy
Raw	76.0	100.0	88.0
Normalized	76.0	88.9	82.4
Quantile normalized	88.0	33.3	60.7
Detrend	100.0	38.9	69.4
Deresolve	88.0	50.0	69.0
SNV	100.0	88.9	94.4
MSC	92.0	38.9	65.4
OSC	92.0	38.9	65.4
Savitzky Golay	100.0	38.9	69.4
1^{st} Der	92.0	33.3	62.7
2^{nd} Der	100.0	44.4	72.2

Table 10.10: Results obtained for SIMCA modelling for the different spectral pretreatments using the only the variables having a VIP > 0.8 for the NOESY. The values recorded in the table represent the % sensitivity of the different models towards the two classes using the new classification decision boundary.

Application of SIMCA on the VIP>0.8 Data Set for the NOESY NMR			
	% Sensitivity Maltese EVOOs	%Sensitivity Non-Maltese EVOOs	Overall % Accuracy
Raw	88.0	66.7	77.3
Normalized	80.0	77.8	78.9
Quantile normalized	80.0	80.6	80.3
Detrend	N/A	N/A	N/A
Deresolve	N/A	N/A	N/A
SNV	100.0	88.9	94.4
MSC	84.0	50.0	67.0
OSC	100.0	61.1	80.6
Savitzky Golay	100.0	33.3	66.7
1^{st} Der	96.0	61.1	78.6
2^{nd} Der	100.0	33.3	66.7

10.2.5 Supervised chemometric discriminate analysis techniques – LDA

The data sets obtained from using common variables selected using the SLC-DA algorithm and those having a VIP>0.8 were subjected to LDA analysis in order to determine whether the reduction of data will improve the classification models obtained using NMR data.

In comparison with SIMCA, LDA avoids the normality problem and confidence interval adjustment making it a more reliable method for classification. When compared with SIMCA and PLS-DA, the LDA method has the disadvantage that the number of training samples must be larger than the number of variables included in the LDA model. In order to fully satisfy this constraint a smaller number of variables were selected based on the standardized scoring coefficients obtained from the SLC-DA. The standardized scoring coefficients of the variable selected during the SLC-DA were obtained and plotted as shown in Figure 10.5 and Figure 10.6. The importance of these coefficients lies in their use to compute canonical scores in terms of the standardized data often referred to as loadings. They are highly informative when it comes to comparing the relative importance in their discriminatory power of the independent variables.

In order to build the LDA, the selected variables obtained in SLC-DA were arranged in ascending order in terms of their scoring coefficients. A smaller set of variables were selected which consisted of 14 variables which corresponded to 7 of the most positive and 7 most negative standardized scoring coefficients. An LDA was carried out on the training set using only the small set of variables which were selected. The results obtained for the training samples were a visualized on an LDA biplot samples as shown in Figure 10.5 and Figure 10.6 for the zg30 and NOESY experiment respectively. Within each plot, each sample is projected as the scores obtained for the first two discriminate functions.

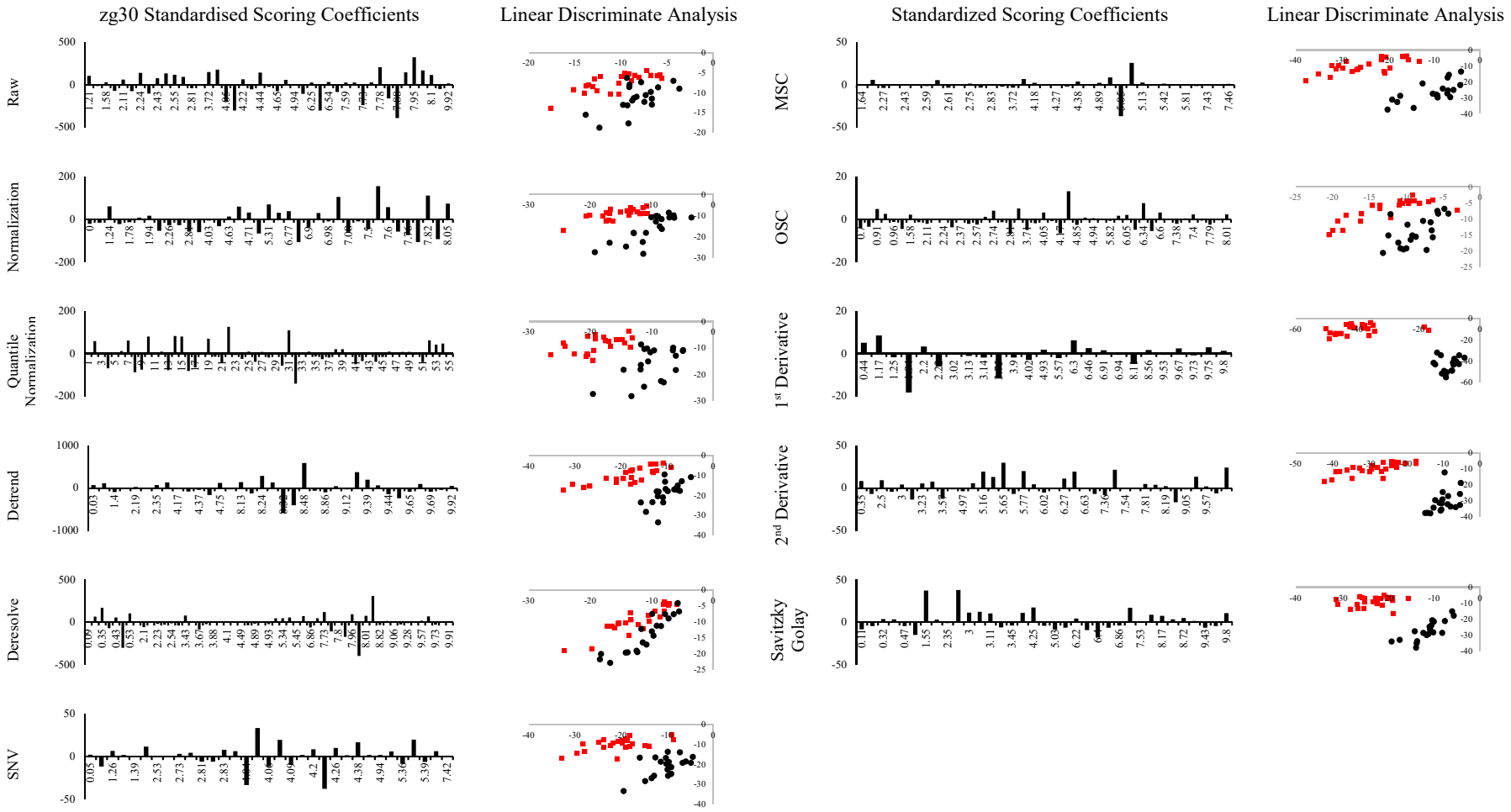


Figure 10.5 (Left) Bar graph showing the standardised scoring coefficients of the variables selected in the SLC-DA for the different zg30 NMR spectral pretreatments, 14 of which were selected for LDA.

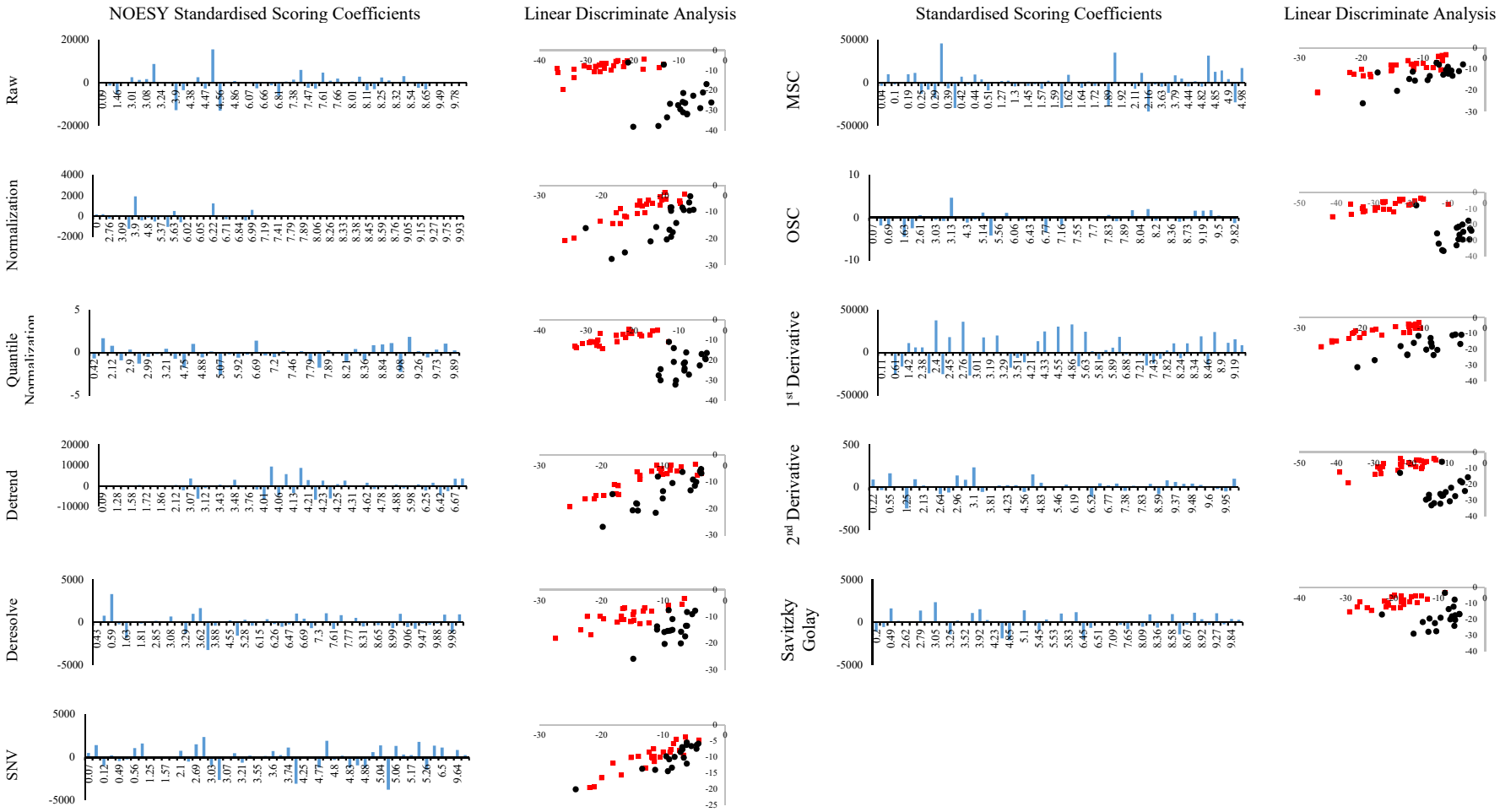


Figure 10.6:(Left) Bar graph showing the standardised scoring coefficients of the variables selected in the SLC-DA for the different NOESY NMR spectral pretreatments, 14 of which were selected for LDA.

From the Table 10.11, it was shown that during the training phase the LDA models obtained for all the pretreatments ranged from 86-100% accuracy for both the zg30 and NOESY NMR data. In the case of the zg30 experiment, it was noted that all the spectral transformations achieved a higher % accuracy during the training phase when compared to the raw data. In the case of the NOESY experiment previously discussed, the actual raw data without the use of spectral transformation yielded a higher discriminatory power. During the validation phase, in the zg30 experiment, only Savitzky Golay derivatization achieved a higher discriminatory power whilst 2nd order derived and quantile normalized spectral had an equivalent discriminatory power to those obtained from using the raw data alone. From the standardized coefficient plot, it can be deduced that compounds having chemical shifts at 1.55, 2.82, 6.2-7.36, and 9.8 ppm have a high discriminatory power and thus offer a more direct form of the discrimination of Maltese EVOOs.

In the case of the NOESY experiment only OSC, 2nd order and Savitzky Golay derived spectra had a higher % accuracy when compared to the raw data, whilst a lower model performance was obtained for the rest of the spectral pretreatments. In the case of OSC, chemical shifts obtained at 1.63, 3.13, 5.50, 6.71, 9.19-9.40 ppm seem to have the highest discriminatory power whilst for the 2nd order derived spectra chemical shifts obtained at 1.25, 2.96-3.10, 4.76, and 9.9 ppm together with chemical shifts observed at 0.49, 2.79, 4.23, 4.85 and 6.45 ppm for the Savitzky Golay derived spectra from being more effective in discerning the Maltese EVOOs from the non-Maltese EVOOs in the NOESY experiment.

Table 10.11: Comparison of the % accuracy and predictability of LDA models for the different spectral pretreatments for zg30 and NOESY NMR.

Linear Discriminate Analysis				
	zg30		NOESY	
	Training	Validation	Training	Validation
Raw	86.05	93.33	100.00	93.33
Normalized	97.67	80.00	95.56	66.67
Quantile normalized	97.67	93.33	97.78	86.67
Detrend	97.67	73.33	88.89	80.00
Deresolve	88.37	73.33	100.00	86.67
SNV	100.00	86.67	86.67	86.67
MSC	97.67	73.33	93.33	86.67
OSC	95.35	80.00	100.00	100.00
Savitzky Golay	97.67	100.00	100.00	100.00
1st Der	97.67	86.67	97.78	66.67
2nd Der	97.67	93.33	100.00	100.00

Nonetheless, it was shown that overall, all the models obtained had a high level of discrimination, in the case of the zg30 experiment 86.6-100% of the excluded samples were successfully classified whilst 66.7-100% accuracy was obtained for the NOESY experiment.

10.2.6 Application of hierarchical cluster analysis

Clustering of samples was carried out using their hierarchical aggregation. The samples are combined according to their distances or similarities to each other. In this experiment, Ward's minimum variance method was used whereby, the distance between two clusters is the equivalent of the ANOVA sum of squares between the two clusters added up over all the variables. Ward's method tends to join clusters with a small number of observations and is strongly biased toward producing clusters with approximately the same number of observations nonetheless it is very sensitive to outliers (Milligan, 1980). Application of hierarchical cluster analysis using Ward's method was used on the variables selected using SLC-DA and VIP>0.8 for Savitzky Golay derivatization was carried out. Savitzky Golay derived spectra were used since they showed the most consist higher rates of % accuracy and % precision through both the zg30 and NOESY experiment as illustrated in Figure 10.7. The cluster analysis carried out on the zg30 Savitzky Golay derived spectra revealed three major clusters as marked in Figure 10.7. The first cluster (Red) was composed mainly of non-Maltese EVOOs however, five Maltese EVOOs were found in this cluster, the second cluster (Green) contained exclusively EVOOs of Maltese origin whilst the last cluster (Blue) contained exclusively non-Maltese EVOOs. Similarly, in the case of NOESY, the first cluster (Green) contained almost exclusively EVOOs of Maltese with the exception of one non-Maltese sample, the second cluster (Blue) and third cluster (Red) contained almost exclusively EVOOs of non-Maltese origin, with the exception of four Maltese EVOOs. These results indicate that although hierarchal cluster analysis might not be an appropriate method for discrimination of EVOOs based on the geographical origin the results obtained offer a confirmation of the results obtained using PLS-DA and LDA.

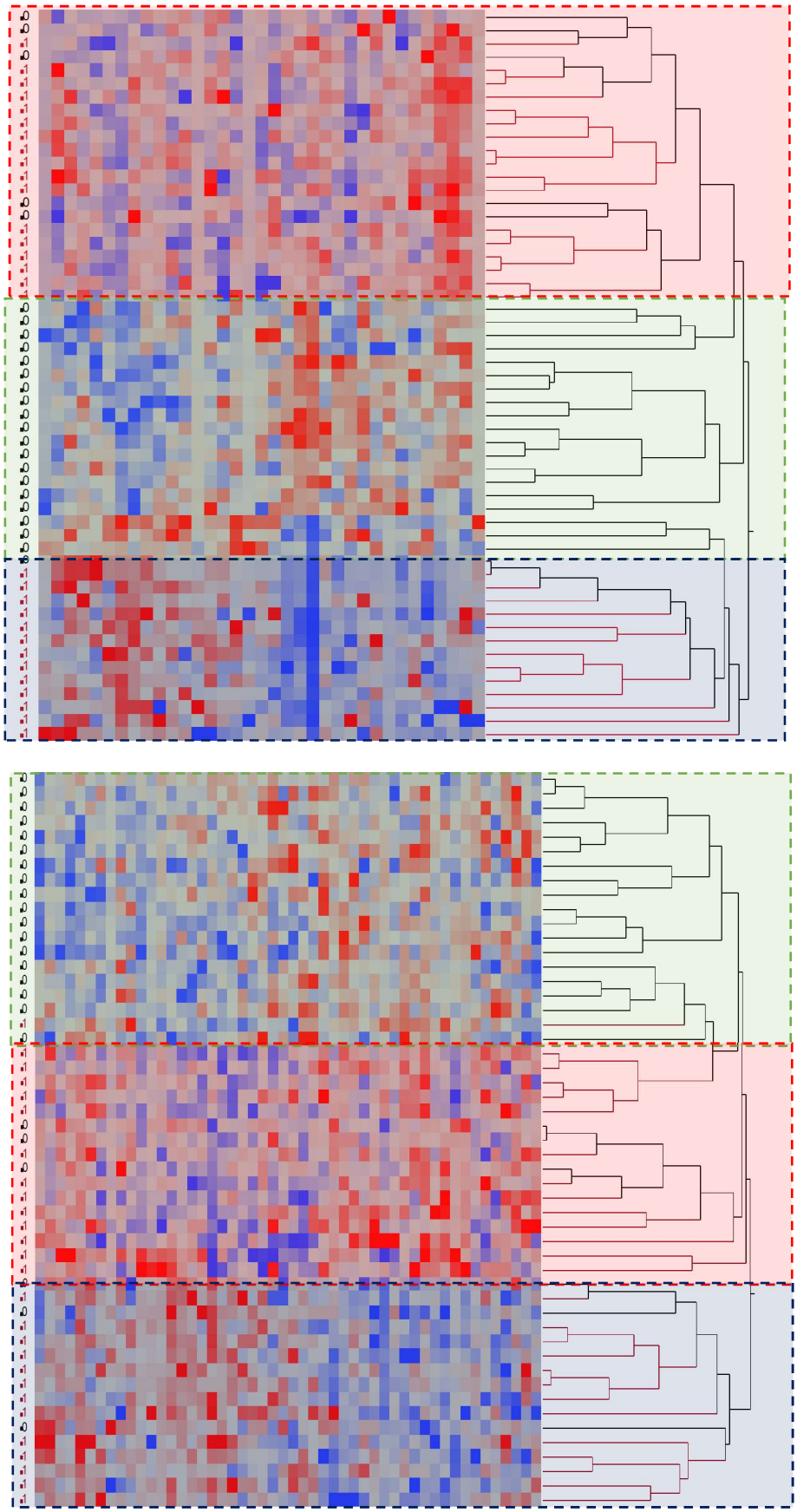


Figure 10.7. Application of hierarchal cluster analysis using Wards methods on the SLC-DA and VIP > 0.8 variables for the Savitzky Golay spectral pretreatment on the zg30 (Top) and NOESY (Bottom) NMR.

10.2.7 Supervised discriminate chemometric techniques – SVM

The dataset containing only variables which were selected using SLC-DA and having a VIP score > 0.8, were subjected to another classification method, known as support vector machine (SVM). SVM similar to PLS can be used for both classification and regression (Christianini and Shawe-Taylor 2000; Vapnik 1995). During this part of the experiment, SVM was used in the context of classification. Table 10.12 and Table 10.13 shows the results obtained on using SVMs using different Kernel tricks including radial, polynomial and sigmoidal on the NMR data obtained from the zg30 and NOESY experiment respectively. The application of kernel functions in SVMs seems to be an advantage, in contrast with other discriminate methods like LDA and PLS-DA where information from all data points is incorporated in order to form the decision boundary, by using the pooled covariance matrix. Unlike PLS-DA and LDA, the SVMs decision boundary is mainly oriented at samples that show no clear classification whether they belong to one or the other class. Although SVM and Kernel functions seem to provide answers for data sets in which the class membership is difficult to obtain, SVMs tend to be more greatly affected by the presence of outliers in the data set causing instabilities in model generated, especially if data outliers are used as support vectors (Steinwart and Christmann 2008).

Table 10.12: Summary SVM models obtained for the zg30 NMR data

Zg30	SVM Type 1									
	Kernel Type Linear									
	Normalized	QNorm	Detrend	Deresolve	SNV	MSC	OSC	Savitzky Golay	1st Derivative	2nd Derivative
Training Accuracy	100.00	83.72	90.70	86.05	74.42	100.00	97.67	88.37	32.56	97.67
Validation	69.77	69.77	67.44	76.74	69.77	51.16	67.44	74.42	62.79	74.42
Predictability	66.67	60.00	80.00	73.33	66.67	80.00	40.00	**93.33***	40.00	73.33
	Kernel Type Polynomial									
Training Accuracy	62.79	55.81	76.74	55.81	90.70	100.00	100.00	74.42	72.09	76.74
Validation	60.47	51.16	67.44	60.47	62.79	58.14	55.81	65.12	65.12	62.79
Predictability	53.33	60.00	80.00	66.67	53.33	33.33	60.00	80.00	66.67	73.33
	Kernel Type Radial Baised									
Training Accuracy	67.44	97.67	74.42	76.74	100.00	100.00	100.00	90.70	97.67	100.00
Validation	55.81	72.09	69.77	65.12	67.44	46.51	76.74	83.72	65.12	72.09
Predictability	66.67	60.00	80.00	80.00	53.33	46.67	46.67	**80.00***	53.33	66.67
	Kernel Type Sigmoid									
Training Accuracy	81.40	88.37	76.74	95.35	76.74	93.02	90.70	72.09	25.58	97.67
Validation	67.44	62.79	67.44	72.09	72.09	58.14	81.40	74.42	69.77	69.77
Predictability	53.33	66.67	80.00	80.00	66.67	93.33	80.00	**93.33***	46.67	73.33

Table 10.13: Summary SVM models obtained for the zg30 NMR data

NOESY	SVM Type 1									
	Kernel Type Linear									
	Normalized	QNorm	Detrend	Deresolve	SNV	MSC	OSC	Savitzky Golay	1st Derivative	2nd Derivative
Training Accuracy	77.78	82.22	93.33	82.22	95.56	97.78	80.00	60.00	93.33	60.00
Validation	77.78	75.56	75.56	68.89	62.22	57.78	77.78	60.00	82.22	62.22
Predictability	66.67	73.33	60.00	73.33	60.00	66.67	40.00	60.00	**80.00***	53.33
	Kernel Type Polynomial									
Training Accuracy	77.78	82.22	73.33	66.67	88.89	95.56	80.00	55.56	82.22	68.89
Validation	77.78	68.89	64.44	60.00	77.78	66.67	55.56	53.33	60.00	68.89
Predictability	**80.00***	60.00	60.00	53.33	73.33	53.33	53.33	60.00	53.33	66.67
	Kernel Type Radial Biased									
Training Accuracy	73.33	73.33	80.00	97.78	84.44	88.89	80.00	97.78	86.67	84.44
Validation	66.67	71.11	68.89	62.22	64.44	66.67	77.78	57.78	75.56	64.44
Predictability	60.00	60.00	**86.67***	60.00	60.00	53.33	80.00	60.00	80.00	60.00
	Kernel Type Sigmoid									
Training Accuracy	88.89	68.89	93.33	91.11	75.56	84.44	82.22	66.67	82.22	62.22
Validation	80.00	71.11	73.33	75.56	68.89	66.67	73.33	66.67	80.00	60.00
Predictability	**86.67***	73.33	80.00	80.00	80.00	73.33	60.00	60.00	73.33	66.67

In the case of SVM, the models were built using different cross-validation forms. In training set which constituted of 80% of the samples was used to build the SVM models. These models were segment validated, whereby the training data set was first partitioned into 10 equally (or nearly equally) sized segments. Subsequently, 10 iterations of training and validation were performed such that within each iteration a different segment of the data was held out for validation while the remaining 9 folds are used for learning. The data matrix was stratified prior split into segments to ensure each segment is a good representative of the whole data set. Once that the model was fitted the % accuracy in the training and validation can be derived. In this experiment, although segment cross-validation is a very good cross-validation method, a second cross-validation method was employed whereby the models fitted on the 80% of the training set were tested on the remaining 20% of the data so that % predictability for each model is obtained.

Table 10.12 shows the SVM results obtained for the zg30 NMR experiment for the four kernel types. In the case of the zg30 experiment, the application of a linear hyperplane, with the exception of 1st order derived spectra the rest of the spectral pretreatments had a training accuracy 74.4-100% which decreased to 51.6-74.4 % in the validation and a predictability in the range of 40-93.3%. It was shown that Savitzky Golay derived spectra had the highest % predictability throughout all the four kernel type functions indicating that the form and shape of the hyperplane do not affect the discriminatory power. Furthermore, it was shown that different pretreatment methods differ in performance depending on the Kernel type function used. In the case of the zg30 experiment, it was found that the detrending and 2nd order derived spectra reached optimal performance using a linear hyperplane, whilst quantile normalized and 1st order derived spectra reach optimal performance using radial baised kernel type. Spectra obtained using deresolve, SNV, MSC, and OSC reached optimal performance under a sigmoidal type kernel-type function. It was noted that polynomial based kernel type was not suitable for none of the zg30 derived spectra.

In the case of the NOESY, it was found that none of the spectral pretreatments had a consistent higher model performance throughout all the different kernel type functions used. For the linear type kernel function, it was found that 1st order derived spectra had the optimal model performance, whilst normalized spectra reached an optimal model performance under the polynomial and sigmoidal type kernel function.

In the case of the radial basis kernel type, spectra pretreated using detrending function reached an optimal performance. On a comparison of the different kernel functions, it was found that the sigmoidal type displayed a generally higher % accuracy in both the training (66.7-93.3%) and validation (66.7-80%) phase together with a higher % accuracy in the predictability (60-86.7%) of the excluded samples.

10.2.8 Application of feed-forward predictive neural networks

The use of feed-forward predictive neural networks on the NMR data as a method for classification was assessed using three different forms of validation, namely 33.3% of data holdback, CV-10 k-fold and excluded row validation. The algorithm fitted on the training set was later tested on the validation data and % predictability of the model was obtained. Table 10.14 shows % accuracy and % predictability for the different forms of cross-validation. Similar, to what was observed in the PLS-DA and SIMCA, NMR raw data derived from the NOESY experiment gave a higher model performance throughout the three different validation methods used. It was shown that whilst the spectral transformation was applied to zg30 NMR data a better model performance was obtained on the use of 33.3% holdback validation with the exception of detrending and OSC. In the case of FF-PNN, matrix scattering-corrected zg30 NMR spectra had the most consistent optimal model performance, followed by the 1st order derived spectra which on the other hand failed in the predictability under the excluded row validation method possibility due to model overfitting.

In the case of NOESY application of FF-PNN to untreated raw data showed a very good model performance as 93.3% of accuracy was observed during the training phase and 91.7% of accuracy was observed during the validation stage. From the results obtained it was shown that only very few spectral pretreatments had a considerable improvement on the FF-PNN models obtained for NOESY data. The only quantile normalized and 1st order derived NOESY NMR spectra showed a consistently higher PNN model performance throughout the three different validation methods used.

Comparison of the zg30 and NOESY PLS-DA models obtained using the whole NMR spectrum, FF-PNN showed a higher model performance. This observation was attributed to the higher flexibility and modelling power of PNNs when compared to PLS-DA. Although PNN is highly flexible and enables the classification of a large number of classes, PNNs often converge on local minima rather than global minima, increasing the risk of overfitting if the training iterations go on too long. In such case the model obtained might start to consider the noise as part of the pattern, this fact could explain the lower % predictability during the validation stage observed when the detrending and 1st order derivatization were applied to zg30 NMR data but not to NOESY data. This observation further confirms that zg30 NMR data to have a lower signal to noise ratio when compared to NOESY.

Table 10.14: Application of FF-PNN on the NMR data using three forms of cross-validation

FF-PNN						
	HoldBack		CV-10		Excluded Row	
	Training	Validation	Training	Validation	Training	Validation
zg30						
Raw	81.03	81.82	96.55	81.82	86.21	90.91
Normalized	94.83	81.82	94.83	100	81.03	63.64
Quantile normalized	98.28	90.91	98.28	90.91	82.76	81.82
Detrend	77.59	54.55	91.38	90.91	75.86	45.45
Deresolve	91.38	90.91	93.1	90.91	79.31	81.82
SNV	96.55	90.91	98.28	100	74.14	72.73
MSC	98.28	100	96.55	90.91	93.1	90.91
OSC	77.59	36.36	89.66	63.64	84.48	81.82
1st Der	96.55	90.91	98.28	100	86.21	45.45
2nd Der	84.48	90.91	98.28	90.91	81.03	63.64
Savitzky Golay	91.38	81.82	98.28	90.91	85	90.91
NOESY						
Raw	93.33	91.67	93.33	91.67	93.33	91.67
Normalized	95	91.67	95	100	93.33	100
Quantile normalized	98.33	100	98.33	100	93.33	100
Detrend	70	83.33	96.67	91.67	95	83.33
Deresolve	96.67	100	96.67	100	91.67	83.33
SNV	93.33	91.67	98.33	100	98.33	100
MSC	93.33	100	93.33	83.33	95	91.67
OSC	81.67	75	91.67	75	90	91.67
1st Der	96.67	100	98.33	100	96.67	91.67
2nd Der	93.33	91.67	98.33	100	93.33	100
Savitzky Golay	88.33	100	98.33	100	90	91.67

In conclusion, it was shown that NMR in conjunction with a number of chemometric methods, provided a cheap, fast and reliable way for the determination of geographical origin of EVOOs, especially when it comes to discrimination of Maltese EVOOs from non-Maltese EVOOs. From the preliminary assessment using only unsupervised PCA models no significant clustering was observed it was attributed to the high levels of similarity between the two classes of EVOOs studied, such method was deemed to be unsatisfactory when it comes to discrimination of geographical origin. Application of supervised methods of classification namely PLS-DA, FF-PNN, LDA and SVM showed to be highly effective in classifying local and non-local EVOOs samples. The use of the variable selection methods significantly increased the effectiveness of PLS-DA models when compared to no variable selection. FF-PNN, SVM and LDA models were also shown to offer similar classification rates to PLS-DA models and thus corroborate the results obtained from the PLS-DA models and put confidence in the use of NMR methods in conjunction with spectral transformation for the classification of Maltese and foreign EVOOs samples. Results showed that different NMR pulse methods can greatly affect the discrimination of EVOOs. NOESY pulse sequence and suppression of strong signal greatly improved the signal to noise ratio and the raw data obtained was more informative when compared to the conventional zg30 pulse sequence. NMR data acquired using zg30 pulse sequence required an extensive spectral elaboration in order to obtain comparable model performance to that of NOESY.

11. Application of spectrofluorometric analysis for the determination of geographical origin of olive oils

Fluorescence spectroscopy is an emergent analytical technique, which presents good sensitivity, with minimal sample preparation. In the case of spectrofluorometric analysis, molecules exhibiting a fluorescence nature are analysed through the simultaneous scanning of the excitation and emission wavelengths resulting in an excitation-emission matrix, known also as a total luminescence spectrum or fluorescence landscape. Rather than using the whole excitation-emission matrix (EEM) adjusting both excitation and emission monochromators to scan in the same instant with a constant wavelength interval between excitation and emission a synchronised excitation-emission spectrum is obtained, this allows narrowing the spectral bands and allows data to handled much easier. Although the SEEFS contain less information than the EEM, they are potentially more informative than single excitation and emission spectra. The selection of an appropriate offset between excitation and emission wavelengths, allows the study of singled out fluorophores, in so doing increasing the sensitivity and selectivity. Coupling this technique to chemometrics enabled the distinction of commercially available samples of virgin olive oils, pure olive oils, and olive pomace oils (Kyriakidis and Skarkalis, 2000; Guimet *et al.,*2004) and determine the overall quality of the olive oil (Guzmána *et al.,*2015)

The aim of this study was to use synchronised excitation-emission spectroscopy associated to chemometrics to differentiate the Maltese EVOO's from other EVOO's derived from other countries within the Mediterranean region. In so developing a quick, easy and cost-saving verification of the origin of EVOOs from the Maltese islands paving the path for the application of protected designation of origin. In this study, the spectroscopic data were collected in the region of 240-700 nm with the wavelength intervals of 10, 30, 60, 80 120 and 185 nm, were processed both by a discriminant chemometric tools including PLS, SVM, and LDA but also using modelling chemometric tools such as SIMCA and PNN. Moreover, different forms of signal pretreatment were employed in order to enhance the potential of FTIR as a tool for authentication purposes.

11.1 Identification of fluorescent peaks

Figure 11.1 shows the full EEM obtained from olive oil dissolved in iso-octane, purified olive oil phenolic extract dissolved in methanol and refined seed oil dissolved in iso-octane, and their corresponding SEEF at different δλ. The total fluorescence spectrum of diluted extra virgin olive oils, measured with the use of right angle geometry, exhibits two intense bands, one with excitation at about 270-330 nm and emission at about 295-360 nm and the second with excitation at about 330-440 nm and emission at about 660-700 nm. In the case of refined seed oil a band located in the intermediate range, with excitation at 280-330 nm and emission at 372-480 nm was observed. The different peaks obtained at the different synchronised spectra were attributed to different fluorophores present within the olive oil.

Similar to the previous studies carried out by Sikorska *et al.,* (2004) and (2005) the results obtained did not show an intense peak at 284 nm as later reported by the same authors in 2008. Analysis of the complete olive oil SEEFS did not reveal the presence of shouldering peaks to the tocopherol peak attributed to the presence of phenolic compounds. However, extraction and analysis of pure phenolic fractions from olive oils displayed a peak maxima centred at around 284 nm which was masked by the prominent tocopherol peak in the complete olive oil, similar results were obtained by Dupuy *et al.,* (2005).

Another important factor which could have masked the fluorescence of the phenolic peak is the solvent effect. Phenolic compounds are poorly soluble in nonpolar solvents (iso-octane), thus the spectrum that was recorded in the iso-octane is different from the spectrum recorded in n-hexane–ethanol mixture as studied by Sikorska *et al.,* (2008). Although phenolic compounds such as caffeic acid and *p*-coumaric acid are insoluble in non-polar solvents, the emission observed in olive oils may originate from oleuropein and secoiridoids derivatives, which have slightly different emission properties and are slightly more soluble in non-polar solvents. Sikorska *et al.,* (2008) showed that the resolution of the phenolic and tocopherol band in the short-wavelength region relying on visual analysis of the three-dimensional EEM is very difficult.

Examination of the total synchronous fluorescence spectra of diluted oils showed dependence of spectral shape and intensity on the wavelength interval (δΔ) used in the measurements, with the presence of particular bands dependent on δΔ. At lower values of δΔ the bandwidths are reduced and the spectrum is simplified as compared to the total fluorescence spectra. The spectra of the virgin olive oil recorded at δ10 nm show two major bands with their maxima at around 310 and 680 nm. These peaks were previously identified by Sikorska *et al.,* (2004, 2005) to be attributed to the presence of tocopherol and the chlorophyll related compounds. These conclusions were further supported by the same authors in 2008 by comparison with respective standards.

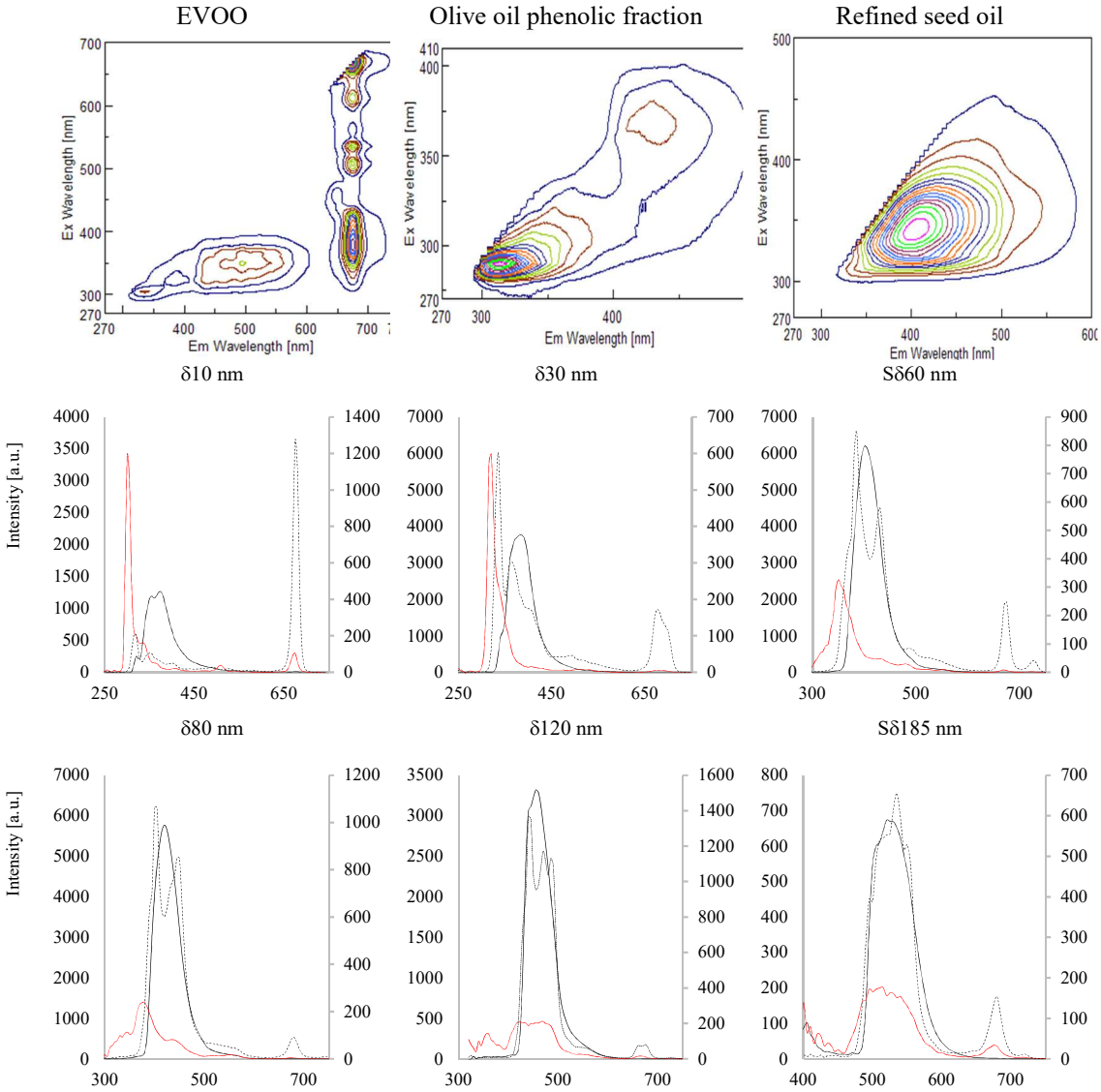


Figure 11.1 : The full EEM obtained for extra virgin olive oil, the phenolic fraction of olive oil and refined seed oils. The corresponding SEEFS measured at δ 10, 30, 60, 80, 120, 185 nm using right angle geometry were obtained. The solid black line corresponds to the refined seed oil, the dotted black line corresponds to complete olive dissolved in iso-octane, whilst the solid red line corresponds to olive phenolic fraction dissolved in methanol.

Appearance of new bands or splitting of existing bands is typically observed with increasing wavelength interval. Emission bands are present in the excitation region below 310 nm, 310-350 nm, 350-380 nm, and above 550 nm in spectra of virgin olive oils (Sikorska *et al*., 2011). Similar spectral characteristics for virgin olive oil were reported by Poulli *et al*., (2006). Through the use of mathematical transformations these changes are further amplified as shown Figure 11.2. From the results obtained it was shown that whilst the δ10 nm showed only emission bands at 310 nm and 670 nm, a bathochromic shift of the 310 nm peak was observed as the δΔ was increased. At δ30 nm two major peaks were observed one at 335 nm and another one at 380 nm as highlighted through the use of quantile normalization and 2nd order derivatisation which corresponds to same peak which was observed for the SEEFS obtained at 10 nm. This peak was attributed to the presence of tocopherol whilst the peak centred 650 nm was attributed to the presence of chlorophyll compounds. Moving to δ60 nm three major peaks were observed, in the region 385-440 nm whilst at δ80nnm only two peaks observed at in this region 405 nm and 440 nm, however as highlighted by the different normalization processes including quantile normalization and SNV two peaks were observed emitting at high wavelengths, 675 nm and 685 nm rather than a single peak as previously observed at δ10 and δ30 nm.

At δ120 nm and δ185 nm the bathochromic shift of lower wavelength is further accentuated as three peaks could be observed at 440, 445 and 470 nm which at δ185 nm these three peaks shift to 510, 540, 555 nm respectively. Furthermore at δ185 nm peaks observed at higher wavelengths split further and become more distinguishable as two maxima were observed 645 nm and 700 nm. The distinct peaks observed at different δΔ cannot be attributed to one single compound. One of the pioneering papers published by Kyriakidis and Skarkalis, (2000) showed that this intense peak was attributed to different forms of tocopherols present within olive oil and their corresponding oxidised derivatives. However, it is still not completely understood what the individual peaks correspond to, since it was later shown that the oxidised derivatives of tocopherols are non-fluorescent compounds (Pollok and Melchert 2004). On examination of synchronous spectra of refining seed (Figure 11.1) oils the peaks are more defined and do not reveal more than one maxima, this suggests that the different emission maxima obtained in the range 350-550 nm can be attributed to both

the tocopherol compounds, but most importantly to phenolic compounds which are present in olive oil and almost completely absent in refined seed oils. The broad peak obtained for the SEEFS of pure olive oil phenolic fraction corroborates these findings.

A number of different studies have been performed in order to fully identify the different peaks observed on the use of SEEFs. The identification of the origin of the particular emission bands rely mainly on comparison to the spectra of chemically pure fluorescent components. The short wavelength band in total fluorescence spectra, which covers the region of 270–330 nm in excitation and 295–360 nm in emission, corresponds to the band at 280–310 nm in the synchronous fluorescence spectra and is mainly assigned to tocopherols and phenols. This assignment has been confirmed by several observations and further confirmed during this study. (Sikorska *et al*., 2004). The presence of a similar peak observed in refined seed oils which do not contain phenolic compounds can be observed at higher δΔ. This observation seems to confirm that tocopherols also contribute to the emission observed in this wavelength range but at a higher wavelength when compared to phenolic compounds. The tocopherols found in the majority of seed and olive oils includes four natural tocopherols (α-, β- , γ-, δ-) and there corresponding unsaturated counterparts tocotrienols (αT_3, βT_3, γT_3, δT_3). Since both tocopherols and tocotrienols share the same basic structure they tend to exhibit very similar UV-absorption spectra and fluorescence properties individual identification is hampered.

The peaks observed at higher wavelengths 525 nm have been attributed to be partly originating from compounds of the vitamin E group, or their derivatives formed upon oxidation (Kyriakidis & Skarkalis, 2000). However the known products of oxidation of -α-, β- , γ-, δ tocopherols, the α-, β- , γ-, δ tocopherolquinones, are all nonfluorescent substances (Pollok & Melchert, 2004) just the identification of these peaks is still not fully understood. Overlapping with this region are phenolic compounds as phenolic acids, phenolic alcohols, hydroxyisochromans, secoiridoids, lignans, and flavonoids which are present in virgin olive oils in very low quantities but given the sensitivity of the instrument their fluorescence contribution could not be excluded (Servili *et al*., 2004). Most of the polyphenols are fluorescent substances, absorbing in the 260-310 nm range and emitting in the near-UV range, with their bands

centred at 310-370 nm (Zandomeneghi and Zandomeneghi, 2005). These phenolic compounds have been detected and identified through the application of HPLC in conjunction with a fluorescent detector by Dupuy *et al*., (2005). Tena *et al*., (2009) recently reported that the fluorescence typical for phenolic components of olive oils showed and emission in the 362-420 nm range.

Analysis of the different spectral pretreatment methods on the different SEEFS it was noted that EVOOs of Maltese origin had more intense peaks observed in the 395-445 nm region at δ60 nm and 645-680 nm as observed at δ120 and δ185 nm. These observations suggest that the EVOOs of derived from the Maltese islands have a higher phenolic/ tocopherol and a higher chlorophyll content. These results corroborate the results obtained using high performance liquid chromatography whereby it was shown that EVOOs of Maltese origin had a significantly higher concentration of secoiridoids and those obtained using UV/Vis absorption analysis for the determination of chlorophyll and carotenoid content.

11.2 Application of chemometrics to SEEFS

As previously stated the aim of this study was to build reliable classification models for the traceability of EVOOs from Malta by coupling synchronised excitation-emission spectroscopy to chemometrics. To this purpose, SEEF spectra of olive oil samples from the Maltese islands and from other Mediterranean countries were collected as described in Section 2.2.1 and analyzed as described in Section 2.2.6.3. In order to obtain spectral fingerprints corresponding to the origin of the sample, discriminant (PLS, LDA) and modelling (SIMCA, PNN) classification approaches were used and compared.

11.2.1 Unsupervised chemometric techniques – PCA

Different kinds of spectral pretreatments were tested and compared in order to overcome the instrumental limitation and account for scattering and other minor variations which would hinder the performance of the classification models. A total of 11 spectral pretreatment methods were used, in each case, after pretreatment a principal component analysis was carried in order to dimensionally reduce the number of variables into a small set of principal component whilst retaining the information of the larger set. PCA enabled the preliminary identification of which pre-treatment method offered the highest variability and possible clustering. Figure 11.2 shows the different forms of spectral pretreatments employed and the corresponding PCA plot for the first two principal components. For the majority of the SEEFs spectral pretreatments, a clustering resembling the geographical origin emerged when the biplot for the first two principal component was analyzed, with δ30 and δ60 nm showing the least clustering. On the other hand synchronized spectra obtained using a 10, 120 and 185 nm wavelength interval showed clustering of data resembling the geographical origin for the majority of the spectral pretreatments. From the % variability explained it was found that for the majority of the SEEFs, the spectra pretreated using the deresolve and the detrending function showed a higher % variability explained for the first two principal complements.

Analysis of the loading plots of the aforementioned spectral pretreatments revealed spectral areas of significant importance which were responsible for the clustering observed in the corresponding biplot. Figure 11.3 shows the variable loading plots for the different SEEFs. In the case of δ10 nm is was shown that peaks obtained at 320 nm and 680 nm corresponding to phenolic and chlorophyll compound fluorescence respectively, had the highest loadings, with the chlorophyll peak being responsible for the majority of the explained variance, whilst the phenolic peak was responsible mainly for the second principal component. For the δ30 nm it was found that the short wavelength emitting compounds 320-360 nm had a higher loading when compared to longer wavelength emitting compounds such chlorophylls are there corresponding derivatives.

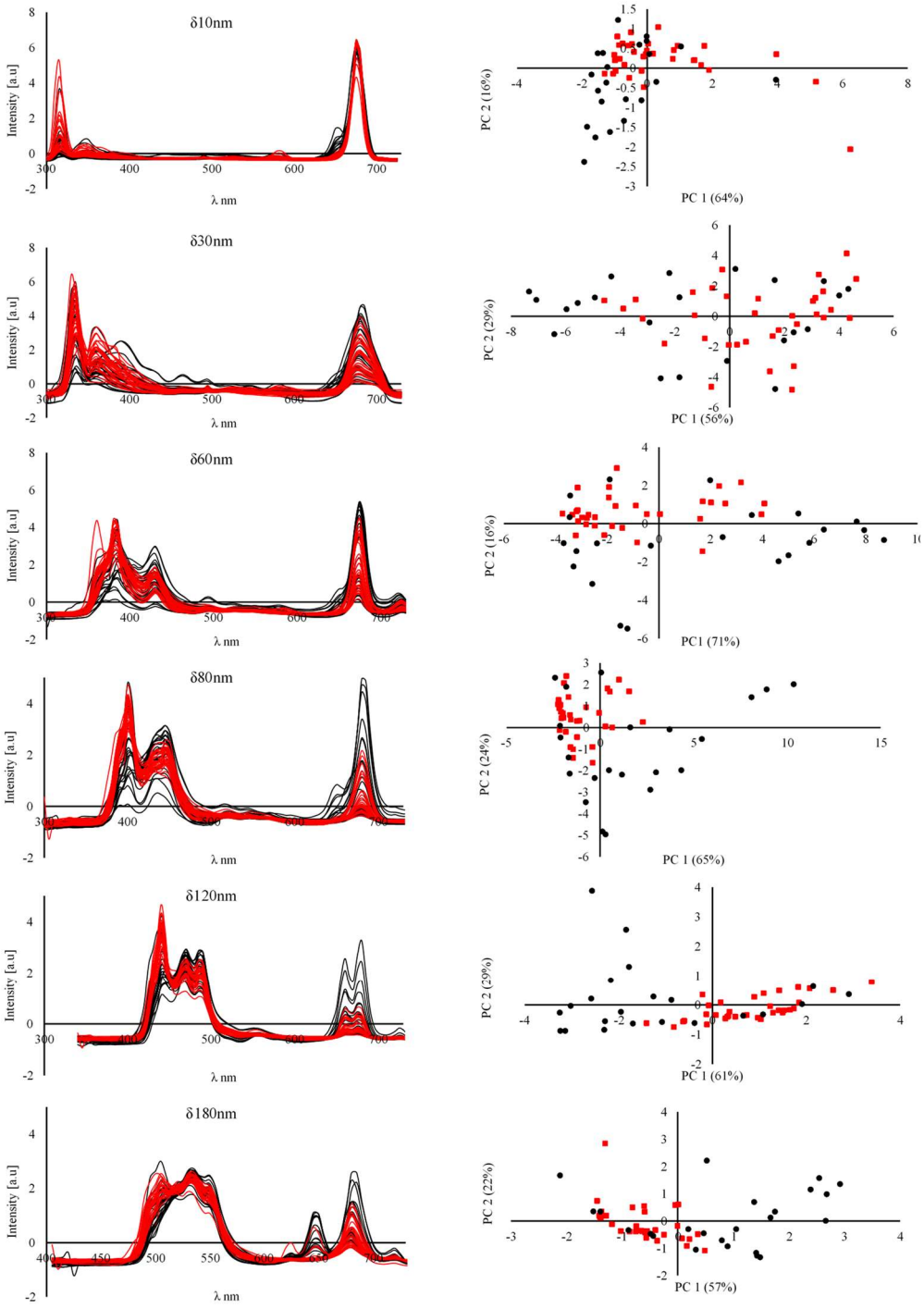


Figure 11.2: Principal component analysis biplot for the different pretreatments obtained for the different SEEF obtained at δ10, 30, 60, 80, 120, 185 nm. EVOOs of Maltese origin are represent as black spectra and black dots whilst non-Maltese EVOOs are represented as red spectra and red squares.

Within the 320-360 nm region compounds namely tocopherols and there corresponding tocotrienol derivatives are expected to emit fluorescent light in the 320-325 nm range which corresponds to the maxima obtained under the 1st principal component, however phenolic compounds namely hydroxybenzoic and hydroxycinnamic acids which emit at slightly higher wavelengths (349-361) explained the remaining variability in the 2nd principal component. Whilst in the case of the δ30 nm the lower wavelength emission compounds had the highest loading in the case of the δ60 nm SEEF, compounds emitting in the 380-420 nm region where responsible for the majority of the explained variation, whilst compounds emitting at lower wavelengths 350-365 nm only contributed to 6% of the explained variation in the 2nd principal component.

The emission of phenolic compounds 349-457 nm was found to be responsible for the majority of the variability explained for the SEEFs at δ80 nm, as two major peaks obtained at 405 nm and another one at 457 nm, possibly corresponding to the emission of p-coumaric and caffeic acid respectively had the largest loadings. Furthermore, the 2nd principal component revealed a shorter wavelength shouldering peak to the 405 which possibly corresponded to gallic acid with emission maxima observed at 382 nm. The importance of the phenolic components present in EVOOs was further emphasised upon examination of the SEEFs obtained at δ120 nm loading plots as compound having a fluorescence in the 435- 500 nm corresponding to nicotinamides (NADPH, NAD) and phenolic compounds namely hydroxycinnamic acids were showed to have the highest loadings in both the 1st and 2nd principal component. The presence of flavin-containing compounds namely FAD, Flavin mononucleotide (riboflavin vitamin B2 derivatives), flavonoids and terpenoids, which emit in the 500-585 nm range showed the highest loadings in the SEEFs obtained at δ185 nm. The emission obtained at 648 nm which corresponds to the emission of chlorophyll b was also evidenced at δ185 nm, however, its loading was lower than those corresponding to flavin, flavonoid, and terpenoid compounds.

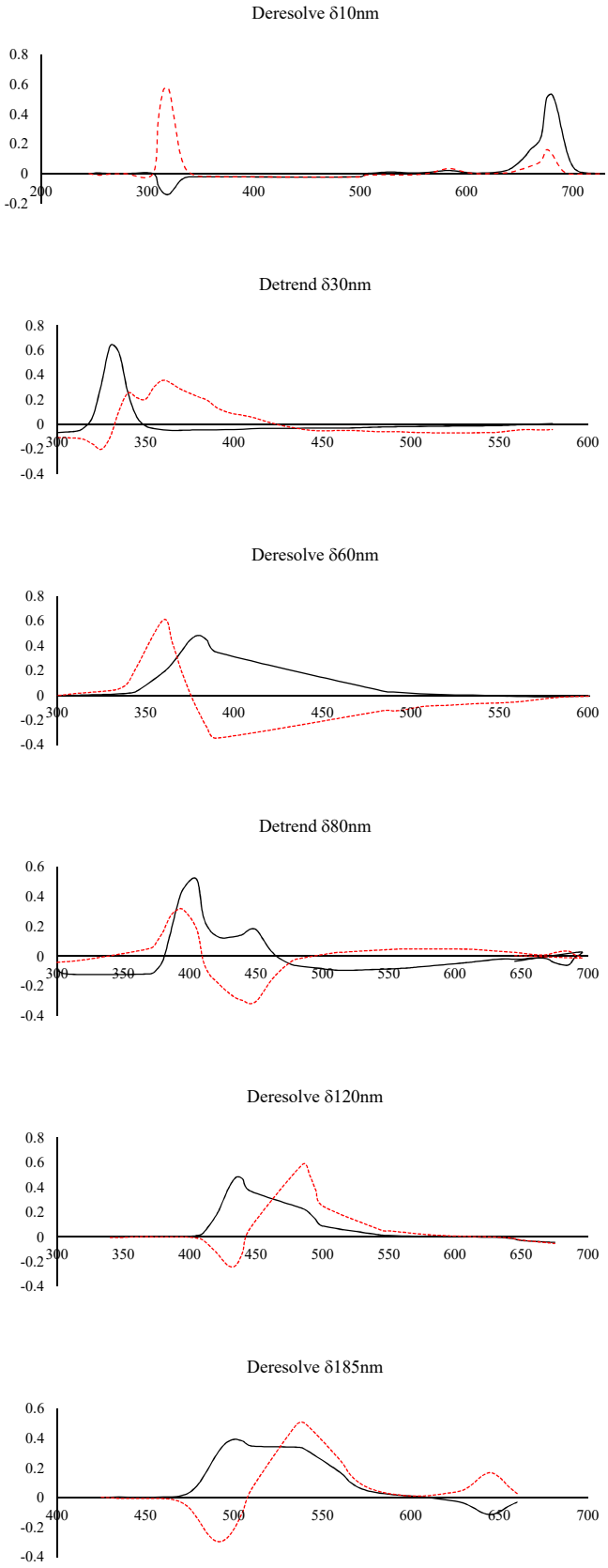


Figure 11.3: Loading plots obtained for the most informative spectral pretreatments obtained for SEEFs at different δnm.

Similar to what was carried to in the other techniques, the next step was to divide the whole dataset into training and test sets (the former to build the model, the latter to validate it). The Maltese and the non-Maltese samples were grouped in an ascending way so that the first 35 samples would represent Maltese EVOO's whilst the rest correspond to non –Maltese EVOO's. A stratified random sampling method was used in order to exclude 20% of the observation so that they would be retained as the testing set. The remaining 80% of the observation were used as to build the training set.

11.2.2 Supervised chemometric techniques – PLS-DA

After splitting the data according to the procedure described above, chemometric classification models were built and tested on all the SEEFs spectral pretreatment using a PLS regression algorithm using JMP 10 and its inbuilt leave one out cross-validation method (LOOCV). Partial least squares regression combines features from and generalizes principal component analysis (PCA) and multiple linear regression. Its goal is to predict a set of dependent variables (geographical origin) from a set of independent variables or predictors (λnm). This prediction is achieved by extracting from the predictors a set of orthogonal factors called latent variables which have the best predictive power. Table 11.1 and Table 11.2 show the number of latent variables extracted, the predicted root mean square error and the % variation explained in terms of X and Y for the different spectral pretreatment methods for each SEEF. From the results obtained with was shown that SEEFs obtained using δ30 and δ60 nm had the lowest performance under both cross-validation methods for the majority of the spectral pretreatments on the other hand it was found that SEEF obtained at δ80 nm had the highest performance. Furthermore it was observed that the PLS-DA models obtained differed not only form SEEF λnm but each SEEF δnm had different model performance depending on the spectral pretreatment used. In the case of δ10 nm it was observed that the model reach optimal performance (100% accuracy and predictability) using the 2nd order derived spectra. At δ80 nm it was found that *Savitzky* Golay, 1st and 2nd order derivatization were equally effective in predicting the geographical origin of EVOOs, as the models obtained using these functions obtained 100% accuracy and predictability.

Table 11.1: PLS-DA model performance using internal validation for the whole SEEF spectrum recorded at different δ nm under different pretreatments.

			Sync 10 nm					**Sync 80 nm**		
	LV	*%X*	*%Y*	*PRESS*	*%Ac*	*LV*	*%X*	*%Y*	*PRESS*	*%Ac*
Raw	5	77.02	77.87	0.70	98.21	7	91.89	78.39	0.73	100.00
Normalized	5	76.50	81.52	0.61	100.00	4	82.63	66.40	0.76	92.86
Quantile normalized	4	63.11	85.66	0.66	100.00	2	68.69	47.48	0.87	76.79
Baseline	13	98.87	92.56	0.67	100.00	10	98.62	82.45	0.64	100.00
Detrend	4	73.57	71.82	0.78	96.43	9	96.09	80.35	0.73	100.00
Deresolve	12	97.35	91.84	0.65	100.00	11	99.50	83.60	0.65	100.00
SNV	9	92.65	90.34	0.65	100.00	6	94.43	70.48	0.74	92.86
MSC	9	93.32	89.91	0.61	100.00	7	98.95	69.03	0.76	92.86
OSC	5	66.18	70.06	0.95	100.00	11	95.69	47.48	0.77	98.21
SG	3	42.36	80.93	0.64	98.21	3	64.69	71.32	0.72	96.43
1st	3	51.67	72.48	0.68	98.21	3	64.69	71.32	0.72	96.43
2nd	4	46.66	85.96	0.62	98.21	2	42.39	64.41	0.79	96.43
			Sync 30 nm					**Sync 120 nm**		
Raw	1	30.79	20.94	0.96	75.00	4	85.13	63.16	0.77	94.64
Normalized	1	40.42	20.25	0.97	73.21	4	80.52	61.66	0.79	94.64
Quantile normalized	2	58.93	41.08	0.95	69.64	4	73.04	69.16	0.76	94.64
Baseline	13	99.22	84.20	0.90	94.64	5	92.89	68.08	0.79	94.64
Detrend	4	77.69	55.66	0.96	91.07	7	93.83	75.58	0.77	96.43
Deresolve	15	99.35	88.25	0.77	96.43	4	45.8	56.13	0.79	89.29
SNV	6	92.68	66.07	0.85	94.64	6	96.52	69.19	0.81	94.64
MSC	8	97.68	72.91	0.81	92.86	8	99.09	75.33	0.78	96.43
OSC	10	95.57	81.93	0.75	96.43	5	86.86	72.11	0.74	98.21
SG	3	54.52	61.70	0.83	92.86	3	65.01	60.95	0.81	91.07
1st	2	30.38	62.91	0.83	98.21	3	65.01	60.95	0.81	92.86
2nd	4	46.66	85.96	0.62	98.21	3	53.7	65.81	0.82	98.21
			Sync 60 nm					**Sync 185 nm**		
Raw	4	81.53	64.15	0.75	96.43	2	66.27	43.49	0.91	87.50
Normalized	4	79.65	69.58	0.72	94.64	1	45.68	26.30	0.94	78.57
Quantile normalized	5	77.44	77.14	0.84	98.21	3	54.65	65.44	0.83	96.43
Baseline	13	99.32	89.73	0.82	100.00	6	93.27	64.14	0.84	96.43
Detrend	4	82.37	62.59	0.76	94.64	6	88.99	61.15	0.77	92.86
Deresolve	14	99.43	90.71	0.69	100.00	2	81.30	34.96	0.89	87.50
SNV	3	76.65	60.82	0.77	92.86	7	94.02	70.77	0.86	96.43
MSC	9	98.38	80.84	0.80	96.43	8	95.82	74.51	0.80	98.21
OSC	8	87.17	83.52	0.72	98.21	4	78.21	63.49	0.85	94.64
SG	9	82.77	90.44	0.75	100.00	14	91.53	89.72	0.88	94.64
1st	3	69.04	62.31	0.78	92.86	3	62.98	56.99	0.82	91.07
2nd	1	18.12	27.99	0.99	100.00	1	34.09	29.25	0.97	82.14

Table 11.2: PLS-DA model performance using internal validation for the whole SEEF spectrum recorded at different δ nm under different pretreatments.

External Validation Whole Spectrum

	Sync 10 nm						**Sync 80 nm**					
	LV	*%X*	*%Y*	*PRESS*	*%A*	*%P*	*LV*	*%X*	*%Y*	*PRESS*	*%A*	*%P*
Raw	3	64.29	74.02	0.77	**98.21**	**100.00**	4	84.44	58.58	0.87	98.21	100.00
Normalized	5	75.96	84.55	0.68	96.43	81.82	4	82.82	62.78	0.82	92.86	100.00
Quantile normalized	4	63.79	88.34	0.66	98.21	90.91	1	43.91	24.03	0.98	89.29	100.00
Baseline	4	86.13	68.35	0.78	91.07	90.91	2	41.37	41.73	0.83	89.29	90.91
Detrend	3	65.35	68.98	0.80	94.64	90.91	15	99.22	89.37	0.80	94.64	100.00
Deresolve	3	64.99	58.53	0.77	85.71	72.73	5	90.72	64.36	0.82	98.21	90.91
SNV	5	88.09	76.52	0.77	94.64	72.73	5	93.65	60.96	0.82	91.07	100.00
MSC	2	53.27	48.48	0.82	85.71	81.82	7	99.07	66.07	0.90	92.86	90.91
OSC	3	57.81	75.63	0.71	98.21	90.91	3	66.00	55.40	0.88	92.86	100.00
SG	3	43.99	81.32	0.71	**98.21**	**100.00**	3	60.89	72.08	0.83	**100.00**	**100.00**
1st	3	52.97	71.89	0.78	96.43	90.91	3	60.89	72.08	0.83	**100.00**	**100.00**
2nd	4	47.97	85.69	0.68	**100.00**	**100.00**	3	52.27	74.31	0.87	**100.00**	**100.00**
	Sync 30 nm						**Sync 120 nm**					
	LV	%X	%Y	PRESS	%A	%P	LV	%X	%Y	PRESS	%A	%P
Raw	1	36.28	20.37	0.98	75.00	63.64	5	87.79	69.98	0.87	98.21	100.00
Normalized	3	74.48	51.72	0.86	87.50	90.91	5	83.76	70.24	0.76	94.64	100.00
Quantile normalized	1	48.63	18.86	0.98	82.14	90.91	3	66.78	63.77	0.78	89.29	81.82
Baseline	1	42.65	16.43	1.02	75.00	63.64	6	94.68	75.53	0.79	**100.00**	**100.00**
Detrend	2	60.19	39.81	0.97	75.00	63.64	5	98.16	60.49	0.82	98.21	100.00
Deresolve	1	37.34	19.75	0.98	87.50	90.91	6	93.27	69.88	0.80	96.43	100.00
SNV	2	66.93	27.09	0.98	80.36	81.82	5	93.10	68.30	0.74	94.64	100.00
MSC	3	88.12	30.95	0.96	76.79	81.82	8	99.13	76.68	0.74	96.43	100.00
OSC	1	38.09	20.94	0.98	73.21	63.64	4	78.30	71.08	0.78	**100.00**	**100.00**
SG	3	49.80	68.16	0.84	**96.43**	**90.91**	3	64.11	67.31	0.85	96.43	90.91
1st	3	49.80	68.16	0.84	**96.43**	**90.91**	3	64.11	67.31	0.85	96.43	90.91
2nd	1	20.44	39.81	0.96	78.57	72.73	3	58.39	69.36	0.90	96.43	100.00
	Sync 60 nm						**Sync 185 nm**					
	LV	%X	%Y	PRESS	%A	%P	LV	%X	%Y	PRESS	%A	%P
Raw	4	83.26	65.69	0.79	**96.43**	**100.00**	1	55.19	19.24	0.97	69.64	63.63
Normalized	4	83.25	65.69	0.79	92.86	81.82	1	49.92	21.61	0.97	73.21	72.73
Quantile normalized	5	80.31	72.42	0.78	96.43	81.82	1	37.90	27.40	0.95	78.57	81.82
Baseline	3	83.60	46.05	0.95	83.93	72.73	1	44.98	19.59	0.94	76.79	81.82
Detrend	7	95.85	79.68	0.77	94.64	90.91	2	80.42	31.25	0.96	87.50	81.82
Deresolve	4	84.24	64.25	0.77	98.21	90.91	6	88.74	58.71	0.87	**87.50**	**100.00**
SNV	3	77.51	63.95	0.76	91.07	72.73	1	55.93	21.52	0.94	73.21	72.73
MSC	14	99.62	92.03	0.78	94.64	90.91	1	58.48	25.66	0.91	83.93	81.82
OSC	4	75.04	73.35	0.87	92.86	81.82	1	50.56	26.94	0.94	80.36	90.91
SG	1	30.43	35.56	1.02	82.14	90.91	1	40.00	26.77	0.96	73.21	63.64
1st	3	70.29	63.59	0.79	91.07	81.82	2	53.19	45.68	0.93	**91.07**	**100.00**
2nd	9	80.93	93.33	0.82	92.86	72.73	1	37.41	25.13	1.01	76.79	63.64

For the δ120 nm it was observed that spectra pretreated using baseline and orthogonal signal correction functions at 100% accuracy and predictability were obtained on the other hand SEEFs obtained at δ185 nm reached optimal performance after deresolve and 1st order derivatization.

In order to fully interpret the PLS models obtained, an inspection of the VIP scores was used in order to determine which predictors (variables) are mainly influencing the latent vectors obtained. VIP is an index of how much a single variable contributes to the bilinear model and it is scaled in such a way that indices having VIP larger than 0.8 are considered to be significantly contributing to discrimination. VIP scores > 0.8 for the PLS models built on the different pretreated SEEFs are reported in Figure 11.4.

As shown in the figure, for the majority of the SEEFs, the VIP > 0.8 identified relevant features in the spectra, particularly, peaks which corresponded to previously identified compounds. In the case of SEEFs obtained at δ10 nm using baseline, detrending, deresolve, MSC and SNV pretreatments selected variables were mainly concentrated in the 300-400 nm and 600-700 nm region corresponding to phenolic/tocopherol and pigment fluorescence. On the other hand Savikty Golay, 1st and 2nd order derived spectra picked up variables throughout the entire spectrum even between 400-600 nm range, in which no fluorescent compounds were observed using δ10 nm, whilst spectra pretreated using OSC showed that variables were mainly concentrated in the 300-400nnm range. Variables selected in the SEEF obtained at δ30 nm showed that for the majority of the spectral pretreatments were selected throughout the entire spectrum rather than being focused on particular fluorescent peaks. However, spectra pretreated using Savitzky Golay, 1st and 2nd order derivatisation, almost completely exclude variables in the 300-400 nm range. In the case of SEEFs obtained at δ60 nm and δ80nnm only detrended and baselined corrected spectra showed more specific variable selection focused around 350-450 nm and 650-700 nm which correspond to major fluorescent peaks, the other spectral pretreatments had a rather non-specific variable selection.

Analysis of variables selected for the SEEFs obtained at δ120 nm and δ185 nm revealed that even though the major fluorescent peak obtained was centred around 400-500 nm, in the majority of the spectral pretreatments, variable were also selected between the 600-700 nm which correspond to shouldering peaks of chlorophyll pigments and their corresponding derivatives.

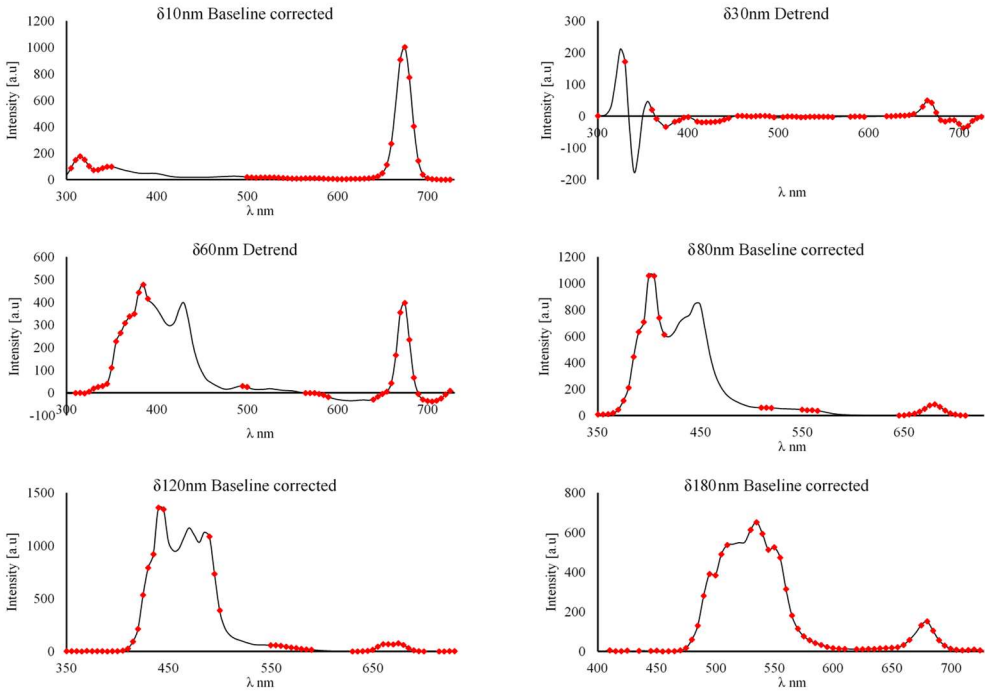


Figure 11.4: Selected variables having a VIP score > 0.8 for the different spectra pretreatments obtained for the different SEEF obtained at δ10, 30, 60, 80, 120, 185 nm

The next step was to build another PLS model this time using only variables which had a VIP score > 0.8. Table 11.3 and Table 11.4 show the results obtained on using the adjusted PLS model for SEEFs. In general comparison of the model performance obtained for the majority of SEEF spectral pretreatments no marked improvements throughout the entire models was observed, furthermore, in certain instances the model performance was in fact lower when compared to the previous models obtained without any variable selection. In the case of SEEF obtained δ10 nm and δ185 nm spectra pretreated using deresolve, detrend, SNV, MSC and OSC functions obtained a higher model performance when compared to those obtained without variable selection however the rest of the spectral pretreatments had comparable or even lower performance. Similar results were obtained for SEEFs obtained at δ30 nm as only spectra obtained after Savitzky Golay, 1st and 2nd order derivatisation showed a marked improvement on using variable selection process. For the SEEFs obtained at δ80 nm and δ120 nm only a marginal improvement was observed for all the spectral pretreatments, however it was only restricted to the training data set (% accuracy) as in the majority of cases the % predictability did not improve, furthermore under certain spectral pretreatments, it actual decreased, upon variable selection. The combined effect of preprocessing and variable selection methods on chemometric model performance may be detrimental to the overall information contained in the data. Appropriate preprocessing selection is therefore a major issue in chemometrics. It is well documented that preprocessing and variable selection approaches are seriously lacking and likely lead to a suboptimal selection of a preprocessing strategy (Engel *et al*., 2013).

Table 11.3: PLS-DA internal validation model performance using the selected VIP>0.8 variables for the SEEFs recorded at different δ nm under different pretreatments

Internal Cross-validation VIP > 0.8

	Sync 10 nm						**Sync 80 nm**			
	LV	*%X*	*%Y*	*PRESS*	*%Ac*	*LV*	*%X*	*%Y*	*PRESS*	*%Ac*
Normalized	5	80.12	79.82	0.62	96.43	6	93.25	80.13	0.65	94.64
Quantile normalized	4	70.87	85.11	0.60	98.21	2	55.27	51.36	0.88	89.28
Baseline	12	98.26	91.40	0.59	100.00	8	97.74	83.43	0.57	100
Detrend	2	50.59	60.77	0.76	94.64	10	98.83	88.55	0.59	89.28
Deresolve	12	98.75	87.90	0.67	100.00	4	84.05	50.91	0.9	100
SNV	10	92.65	90.72	0.58	100.00	7	96.15	80.57	0.67	94.66
MSC	10	96.22	88.48	0.56	98.21	7	98.28	75.4	0.72	94.66
OSC	10	95.23	89.41	0.65	100.00	15	97.42	91.14	0.69	98.29
SG	3	55.12	81.34	0.61	98.21	3	71.17	80.09	0.58	94.66
1st	2	48.40	69.18	0.66	98.21	7	71.17	80.09	0.58	94.66
2nd	4	46.64	85.97	0.62	98.21	9	85.18	84.27	0.84	98.29
	Sync 30 nm						**Sync 120 nm**			
Normalized	15	99.94	82.78	0.69	98.21	4	84.59	54.5	0.85	89.29
Quantile normalized	3	73.55	55.05	0.86	89.29	4	75.16	66.68	0.81	94.64
Baseline	9	98.36	69.18	0.78	92.86	7	95.81	83.02	0.59	100.00
Detrend	13	99.4	87.46	0.6	89.29	15	99.94	91.2	0.61	100.00
Deresolve	11	99.85	78.06	0.65	94.64	6	94.04	78.15	0.62	100.00
SNV	8	96.48	75.34	0.76	96.43	10	98.71	88.88	0.6	100.00
MSC	5	95.74	61.52	0.78	91.07	12	99.4	89.73	0.63	100.00
OSC	10	93.53	84.23	0.73	96.43	4	84.25	60.08	0.82	96.43
SG	15	98.37	83.81	0.6	94.64	10	92.6	92.04	0.62	98.21
1st	7	82.76	77.94	0.61	96.43	10	92.6	92.03	0.62	98.21
2nd	10	86.65	81.85	0.57	98.21	3	64.79	75.91	0.65	96.43
	Sync 60 nm						**Sync 185 nm**			
Normalized	5	89.38	70.96	0.7	94.64	15	99.69	89.5	0.58	100.00
Quantile normalized	5	79.61	73.09	0.82	92.86	3	66.99	63.16	0.78	96.43
Baseline	1	33.06	27.51	1.01	78.57	9	98.02	83.02	0.68	100.00
Detrend	7	94.45	77.36	0.74	100.00	10	99.1	84.62	0.65	89.29
Deresolve	9	98.59	76.55	0.75	100.00	4	85.96	55.18	0.81	100.00
SNV	11	98.99	85.23	0.75	98.21	7	95.22	78.88	0.8	94.64
MSC	9	97.91	82.09	0.8	96.43	6	93.11	77.91	0.71	96.43
OSC	8	91.55	75.53	0.79	94.64	4	85.07	61.31	0.86	91.07
SG	14	95.18	95.97	0.74	98.21	1	38.38	32.27	0.92	82.14
1st	7	76.05	61.68	0.75	91.07	3	69.78	66.4	0.7	89.29
2nd	10	82.52	93.76	0.85	100.00	5	83.89	77.2	0.71	96.43

Table 11.4: PLS-DA external validation model performance using the selected VIP>0.8 variables for the SEEFs recorded at different δ nm under different pretreatments

	External Validation VIP > 0.8											
	Sync 10 nm						**Sync 80 nm**					
	LV	*%X*	*%Y*	*PRESS*	*%A*	*%P*	*LV*	*%X*	*%Y*	*PRESS*	*%A*	*%P*
Normalized	4	68.63	80.39	0.61	96.43	90.91	6	93.13	78.36	0.72	94.64	100.00
Quantile normalized	4	70.88	85.9	0.63	98.21	90.91	2	68.11	44.51	0.9	89.29	100.00
Baseline	9	96.54	84.3	0.75	96.43	81.82	2	63.61	39.41	0.89	80.36	72.73
Detrend	10	98.48	86.28	0.71	96.43	90.91	13	99.74	93.27	0.87	89.29	90.91
Deresolve	3	60.4	70.2	0.74	**100.00**	**100.00**	4	86.04	50.8	0.89	98.21	90.91
SNV	6	87.82	81.87	0.77	**100.00**	**100.00**	5	93.96	70.6	0.74	91.07	100.00
MSC	7	93.09	81.8	0.71	**100.00**	**100.00**	12	99.75	87.69	0.8	**100.00**	**100.00**
OSC	3	67.87	73.88	0.67	98.21	90.91	2	54.98	45.18	0.9	89.29	100.00
SG	2	48.41	74.68	0.63	**98.21**	**100.00**	3	69.48	80.26	0.6	**98.21**	**100.00**
1st	2	49.1	67.37	0.73	96.43	90.91	3	69.48	80.26	0.6	**98.21**	**100.00**
2nd	3	40.95	80.54	0.71	**98.21**	**100.00**	5	71.84	76.25	0.92	94.64	81.82
	Sync 30 nm						**Sync 120 nm**					
Normalized	7	98.11	69.42	0.84	92.86	72.73	4	86.61	61.23	0.75	96.43	100.00
Quantile normalized	2	69.84	42.1	0.88	76.79	72.73	3	69.28	60.81	0.8	91.07	90.91
Baseline	1	43.27	16.64	1.02	75.00	72.73	6	94.89	80.38	0.69	98.21	100.00
Detrend	2	49.42	42.86	0.96	75.00	72.73	9	99.44	82.02	0.72	92.86	81.82
Deresolve	1	34.79	23.16	0.95	89.29	90.91	5	91.61	72.22	0.79	**100.00**	**100.00**
SNV	2	63.78	27.08	0.97	78.57	81.82	5	90.41	76.87	0.64	94.64	100.00
MSC	1	42.35	17.58	0.96	71.43	81.82	10	99.03	87.23	0.71	**100.00**	**100.00**
OSC	1	36.13	22.12	0.98	75.00	72.73	3	75.44	64.08	0.74	**96.43**	**100.00**
SG	4	68.66	74.81	0.81	**98.21**	**100.00**	12	95.39	95.95	0.63	96.43	81.82
1st	3	64.87	69.57	0.71	**94.64**	**100.00**	12	95.39	95.95	0.63	96.43	81.82
2nd	2	46.22	69.46	0.68	**98.21**	**100.00**	3	65.51	77.2	0.73	98.21	100.00
	Sync 60 nm						**Sync 185 nm**					
Normalized	3	80.77	66.12	0.76	91.07	72.73	14	99.36	90.62	0.68	96.43	81.82
Quantile normalized	4	77.78	70.23	0.84	91.07	72.73	3	68.06	61.31	0.87	**96.43**	**100.00**
Baseline	1	34.80	28.76	1.02	80.36	81.82	7	95.7	78.67	0.78	96.43	90.91
Detrend	7	94.39	79.02	0.75	94.64	90.91	10	99.47	84.83	0.72	**92.86**	**100.00**
Deresolve	4	85.88	60.54	0.75	**98.21**	**90.91**	4	86.67	52.03	0.87	98.21	90.91
SNV	3	79.47	64.91	0.74	92.86	81.82	1	47.9	32.79	0.89	83.93	81.82
MSC	13	99.43	90.58	0.84	96.43	90.91	7	96.44	77.12	0.72	**94.64**	**100.00**
OSC	4	77.89	65.04	0.76	91.07	72.73	12	98.73	83.91	0.88	**94.64**	**100.00**
SG	9	86.64	92.46	0.83	96.43	90.91	1	42.11	26.9	0.96	78.57	81.82
1st	3	77.63	62.56	0.76	89.29	90.91	2	60.92	55.99	0.82	87.50	100.00
2nd	9	82.06	93.59	0.83	**98.21**	**90.91**	5	82.49	79.58	0.75	**98.21**	**100.00**

This is mainly attributed to the fact that application of PLS models on a smaller number of variables increases the chances of model overfitting, resulting in models which are able to correctly predict the geographical origin of the training data set but attain a lower model performance when it comes to predicting the geographical origin of the excluded samples. Evidence of this, comes from analysing the models obtained using LOOCV, a less rigorous cross-validation method. Comparison of the PLS models obtained from LOOCV showed a higher performance when compared to those obtained using external cross-validation method, indicating that the models obtained after variables selection a more susceptible to overfitting. However, as it will be demonstrated further on in this study the application of a combined variable selection processes can greatly reduce the chances of overfitting.

11.2.3 Modelling Chemometric techniques – SIMCA

The classification analysis of EVOOs SEEF data was then repeated using a modelling approach based on the SIMCA algorithm. The latter is a class modelling algorithm that allows the analyses one class at a time. For SIMCA analysis two PCA were built for each spectral pretreatment one for the EVOOs of Maltese origin and the other one for the EVOOs of non-Maltese origin. In each case, the optimal number of principal components were chosen in order to obtain optimal model complexity in 10-fold row-wise cross-validation. The use of a two-stage model, one for each category, allowed the comparison between the two categories, thus this allowed us to check whether samples are accepted by one, both or none of the modelled classes. The output of SIMCA analysis was assessed by the use of Coomans plot shown in Figure 11.5. A Coomans plots take the form of a graph where the two axes represent the distance of the samples to each of two class models. The horizontal and vertical lines corresponding to the threshold distances also known as the significance limit can be adjusted up depending on the sensitivity required. One of the major disadvantages of using SIMCA is that one has to set a confidence level, α. If the data are normally distributed, α % (e.g. 5%) of objects belonging to the class will be considered as not belonging to it. For this experiment given that the SEEF data was highly similar the significance limit was increased up to 25% rather than the default 5%

Table 11.5: SIMCA performance given in terms of % sensitivity and % specificity towards the different classes using SEEFs recorded at different δ nm under different pretreatments

	% Sensitivity Maltese	% Sensitivity Non-Maltese	% Specificity Maltese	% Specificity Non-Maltese	% Sensitivity Maltese	% Sensitivity Non-Maltese	% Specificity Maltese	% Specificity Non-Maltese
	Sync 10 nm				Sync 30 nm			
Normalised	87.0	15.2	100.0	100.0	65.2	3.0	97.0	100.0
Quantile normalized	73.9	12.1	100.0	100.0	60.9	9.1	100.0	100.0
Baseline	43.5	27.3	100.0	100.0	21.7	6.1	100.0	100.0
Detrend	78.3	36.4	100.0	100.0	47.8	12.1	100.0	100.0
Deresolve	91.3	24.2	97.0	95.7	56.5	15.2	97.0	95.7
SNV	65.2	9.1	100.0	100.0	60.9	12.1	93.9	100.0
MSC	65.2	12.1	100.0	100.0	47.8	0.0	93.9	100.0
OSC	73.9	33.3	100.0	100.0	52.2	15.2	100.0	100.0
Savitzky Golay	73.9	54.5	100.0	100.0	60.9	6.1	100.0	100.0
1st Der	73.9	15.2	100.0	100.0	43.5	24.2	100.0	100.0
2nd Der	78.3	3.0	100.0	100.0	60.9	36.4	100.0	100.0
	Sync 60 nm				Sync 80 nm			
Normalised	78.3	21.2	97.0	100.0	69.6	15.2	97.0	100.0
Quantile normalized	87.0	6.1	100.0	100.0	69.6	3.0	100.0	100.0
Baseline	30.4	15.2	100.0	100.0	13.0	9.1	100.0	95.7
Detrend	60.9	12.1	100.0	100.0	43.5	6.1	100.0	100.0
Deresolve	60.9	15.2	84.8	95.7	39.1	21.2	97.0	95.7
SNV	78.3	9.1	97.0	100.0	69.6	9.1	100.0	100.0
MSC	87.0	6.1	100.0	100.0	17.4	0.0	100.0	100.0
OSC	65.2	18.2	100.0	100.0	56.5	27.3	100.0	100.0
Savitzky Golay	60.9	6.1	100.0	100.0	52.2	3.0	90.9	100.0
1st Der	73.9	36.4	100.0	100.0	56.5	3.0	90.9	100.0
2nd Der	78.3	6.1	100.0	100.0	82.6	3.0	100.0	100.0
	Sync 120 nm				Sync 185 nm			
Normalised	82.6	6.1	84.8	100.0	69.6	9.1	100.0	100.0
Quantile normalized	69.6	3.0	100.0	100.0	78.3	3.0	100.0	100.0
Baseline	30.4	18.2	93.9	100.0	34.8	6.1	100.0	100.0
Detrend	43.5	27.3	87.9	100.0	73.9	3.0	100.0	100.0
Deresolve	30.4	9.1	84.8	95.7	56.5	18.2	87.9	100.0
SNV	87.0	0.0	97.0	100.0	78.3	9.1	100.0	100.0
MSC	78.3	3.0	97.0	100.0	73.9	9.1	100.0	100.0
OSC	69.6	6.1	100.0	100.0	43.5	6.1	100.0	100.0
Savitzky Golay	52.2	12.1	100.0	100.0	82.6	6.1	100.0	100.0
1st Der	60.9	12.1	97.0	100.0	82.6	18.2	100.0	100.0
2nd Der	69.6	21.2	93.9	95.7	78.3	0.0	97.0	100.0

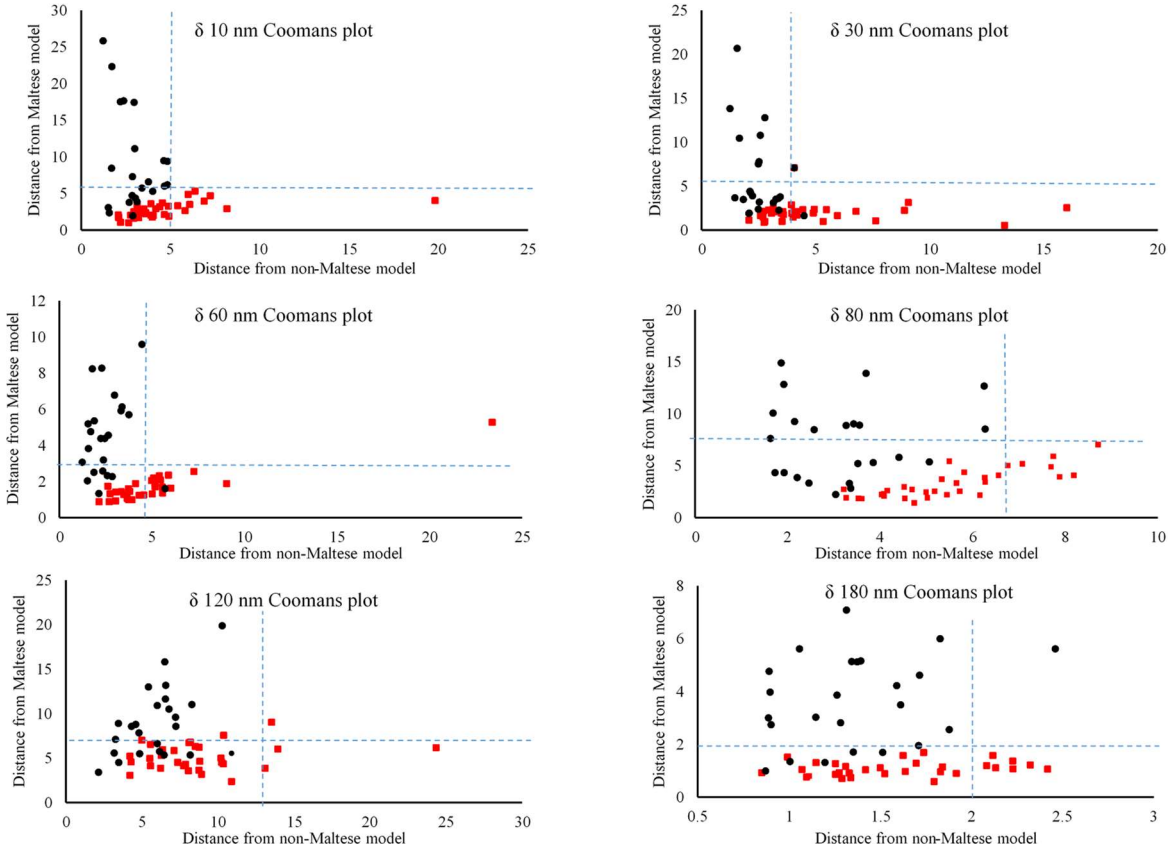


Figure 11.5: Coomans plot obtained using SIMCA for base line corrected spectra obtained for the different SEEF obtained at δ10, 30, 60, 80, 120, 185 nm.The blue dotted lines represent the 25% confidence

Although in previous chemometric analysis using different spectroscopic methods, a new line bisecting the plot corresponding to discriminant classification boundary was built in order to represent a new significant boundary in order to overcome the ubiquity of the non-Maltese EVOOs this could not be applied to SEEF data thus the results presented in Table 11.5 are given in terms of the specificity and sensitivity of the SIMCA models obtained towards both the Maltese and non-Maltese EVOOs. Sensitivity is the percentage of samples from the modelled class that are accepted by the class model, while specificity is the percentage of samples from other classes which are rejected by the class model. It was observed that SIMCA models obtained had an overall very high sensitivity and specificity towards the Maltese EVOOs, however independent of the spectral transformations and the SEEF δnm used, the models obtained had a very low sensitivity for the non-Maltese EVOOs. This was expected since within the non-Maltese EVOO group, EVOOs from different originating countries were used each one of them having a different spectrofluorometric fingerprint, increasing the overall variability of the subset, making it more difficult to be modelled as a single class. However the results obtained further confirm that EVOOs obtained from the Maltese islands tend to be unique in their spectrofluorometric profile independent of the originating cultivar, thus enabling the EVOOs to be fully modelled into one single class which almost completely discriminated from other non-Maltese EVOOs.

The results obtained further illustrate that the performance of the model towards the sensitivity of the Maltese EVOOs is highly dependent on both the spectral transformations used and the SEEF δnm employed. SEEFs obtained using δ10 nm, δ120 nm and δ185 nm had a higher sensitivity towards the Maltese EVOOs when compared to SEEF spectra obtained at δ30 nm, δ60 nm and δ80 nm. These results corroborate the results obtained using PLS-DA. In order to obtain a more robust method of classification with the use of a smaller number of variables, the VIP data set obtained from the previous analysis was subjected to a stepwise linear canonical discriminate analysis SLC-DA.

SLC-DA was performed on the SEEF data from all the pretreatment methods in order to extract only a small amount of highly discriminate variables which would enable an easier and faster discrimination between the origins of EVOOs. This strategy involved a substantial reduction of the dimensionality of the data in such a way that

only the variables shown in Figure 11.8ab were retained. In order to further reduce the number of variables selected from the SLC-DA analysis, a minimum of 18 variables was selected in order to carry out a conventional LDA. During the SLC-DA the variables chosen by applying a forward stepwise variable selection algorithm using JMP 10 using a Wilks' Lambda as a selection criterion and an F-statistic factor to determine the significance of the changes in Lambda when the influence of a new variable is evaluated. The most significant variables were then extracted and their canonical scoring coefficients were plotted as shown in Appendix 17 Section 17.3. The main advantage of using SLC-DA over the convention LDA is the ability to perform a feature selection. Variables which helped to improve classification performance were used whereas variables without discriminant information were discarded. The SIMCA and PLS analysis were carried out on the reduced data, as simplified by Figure 11.8.

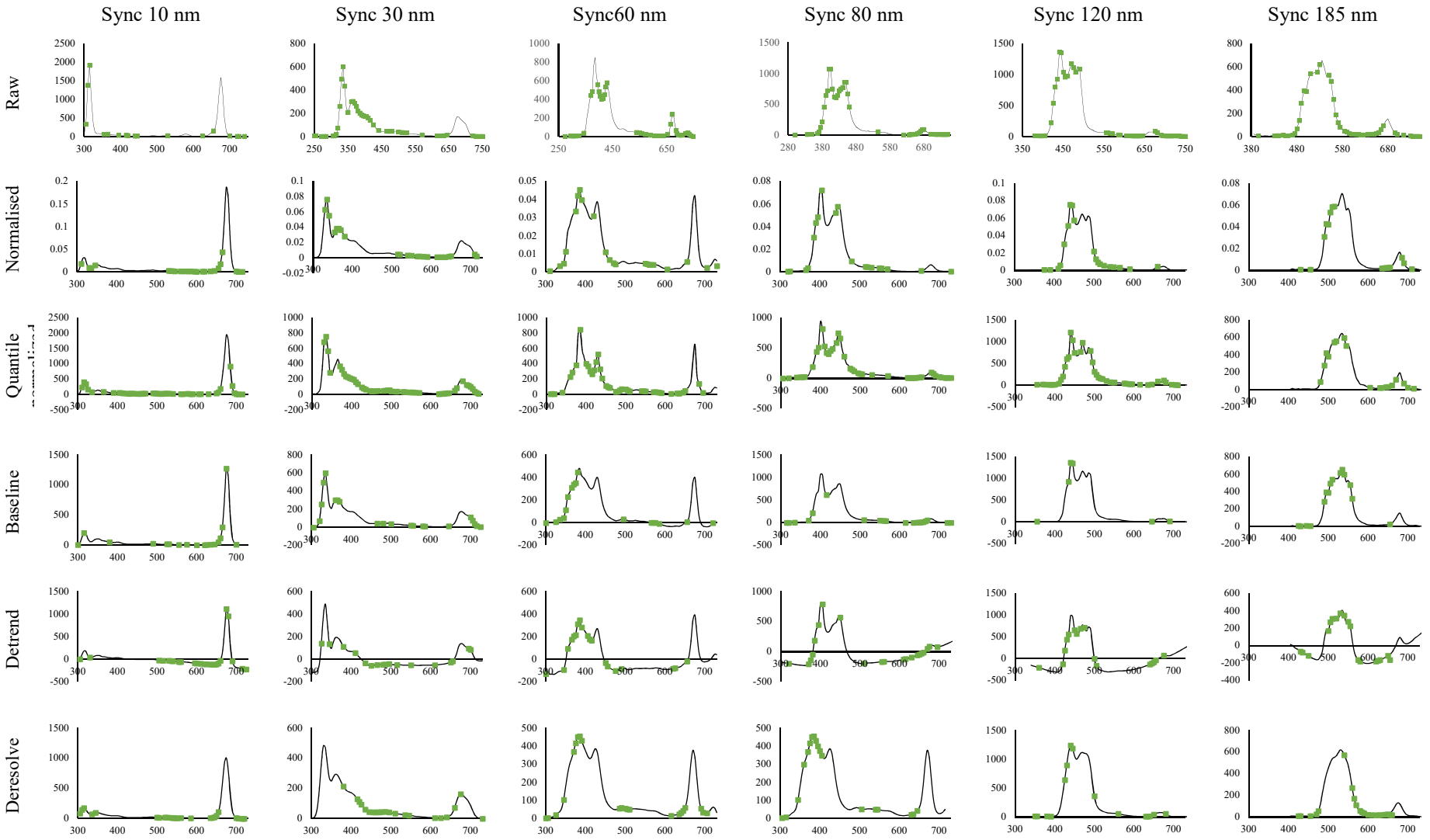


Figure 11.6: Selected variables having a VIP score > 0.8 & selected in the SLC-DA for the different spectra pretreatments obtained for the different SEEF obtained at δ10, 30, 60, 80, 120, 185nm

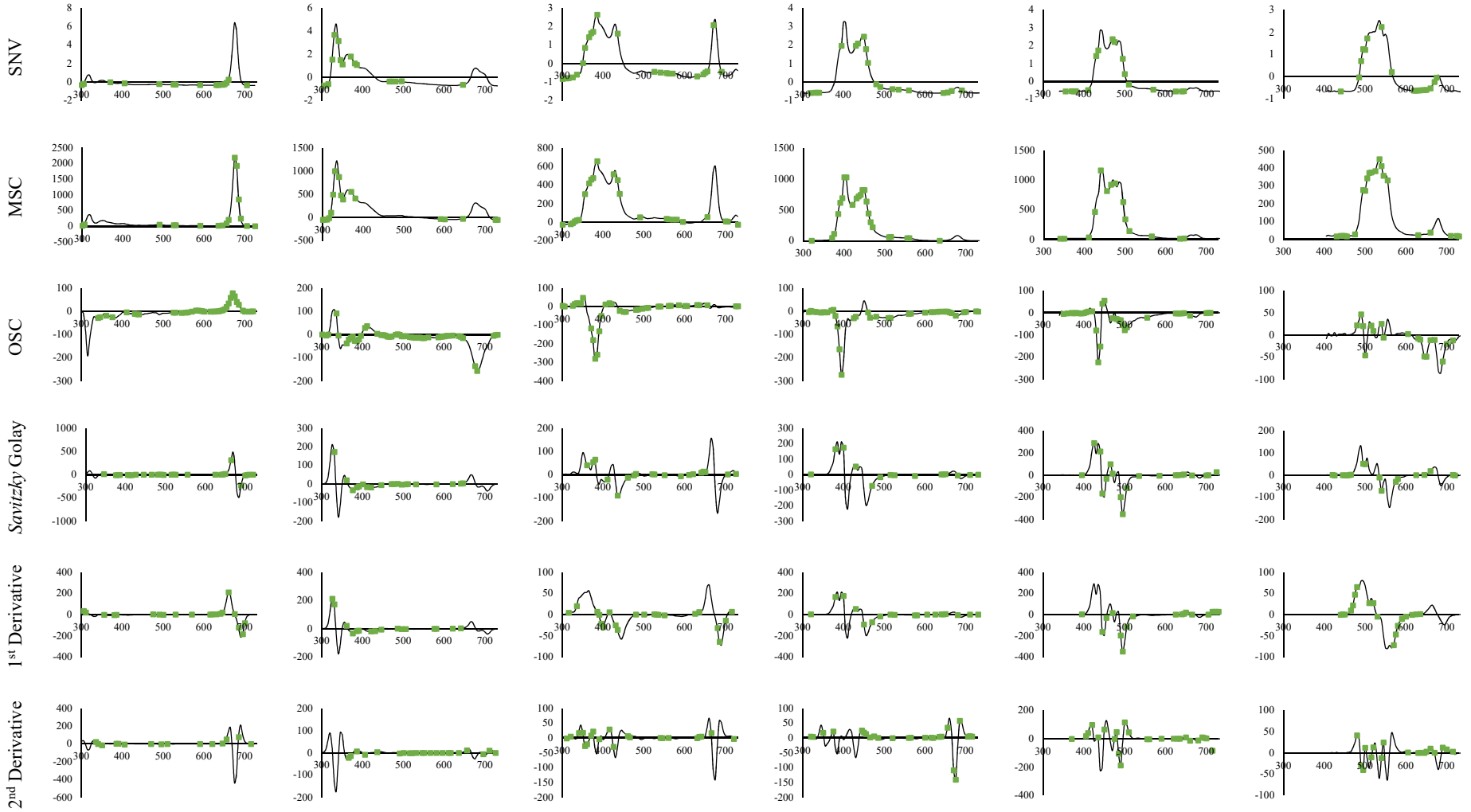


Figure 11.7: Selected variables having a VIP score > 0.8 & selected in the SLC-DA for the different spectra pretreatments obtained for the different SEEF obtained at δ10, 30, 60, 80, 120, 185nm

Figure 11.6 and Figure 11.7 show the variables selected during SLC-DA for all spectral pretreatments. In the case of SEEF obtained at δ10 nm it was found that the for the majority of the spectral pretreatments the stepwise-linear algorithm employed identified variables of significant importance namely in the 300-400 nm and 600-750 nm, whilst only derived spectra after Savitzky Golay, 1st and 2nd order derived δ10 nm spectra, selected variables in the 400-600 nm range. The remaining SEEFs it was found that the 600-750 nm region received less importance as less discriminate variables were selected however, similar to what was observed in SEEF δ10 nm, whereby the derived spectra after Savitzky Golay, 1st and 2nd order derivatisation, discriminate variables were selected throughout the entire spectrum.

Once the variables were selected for each spectral pretreatment a PLS-DA model was applied in order to determine whether variable selection using the linear method provided ways for a better form of classification. Table 11.6 and Table 11.7 show the results obtained from the PLS using the data set composed of variables which had a VIP score > 0.8 and were selected during the SLC-DA analysis. The results show a marked increase in the performance throughout all the spectral pretreatments for all the SEEFs. The increased in performance is given in terms of higher the number of latent variables extracted, lower PRESS, a higher % of variance explained in terms of X and Y together with an increase in both % accuracy and % predictability of the PLS models obtained. These results suggesting that a variable selection using the combination of two techniques greatly improves the modelling power via the removal of collinear and redundant variables. These results further confirm that the SEEF obtained at δ10 nm have a higher discriminatory power when compared to the other SEEFs, as SEEF spectra obtained at 10δnm after deresolve, SNV, MSC, 1st and 2nd order derivatisation enabled the correct classification of the samples used in both the training and in the validation test set.

Table 11.6: PLS-DA performance using internal cross-validation for the selected variables VIP>0.8 & SLC-DA for the SEEFs recorded at different δ nm under different pretreatments

Internal Cross-validation VIP > 0.8 & SLC-DA

		sync 10 nm						**sync 80 nm**		
	LV	*%X*	*%Y*	*PRESS*	*%Ac*	*LV*	*%X*	*%Y*	*PRESS*	*%Ac*
Normalized	11	97.14	90.12	0.56	100.00	3	70.74	54.92	0.77	92.86
Quantile normalized	4	67.83	96.3	0.6	98.21	2	68.7	47.48	0.87	85.71
Baseline	13	99.65	89.39	0.56	100.00	11	99.41	72.89	0.76	96.43
Detrend	9	98.89	75.55	0.68	98.21	10	99.69	73.03	0.79	94.64
Deresolve	14	99.87	90.91	0.46	98.21	15	100	73.35	0.82	94.64
SNV	10	97.69	86.97	0.52	100.00	5	96.59	59.3	0.81	91.07
MSC	5	84.38	78.98	0.59	98.21	3	94.64	41.05	0.86	85.71
OSC	14	98.33	91.4	0.57	100.00	11	95.86	79.68	0.78	94.64
SG	12	91.03	93.85	0.47	100.00	6	83.81	73.18	0.67	96.43
1st	15	99.61	89.48	0.54	100.00	6	82.39	74.35	0.69	96.43
2nd	15	95.24	93.93	0.38	100.00	15	96.69	84.56	0.78	92.86
		sync 30 nm						**sync 120 nm**		
Normalized	15	99.94	82.78	0.69	100.00	15	99.95	81.79	0.67	94.64
Quantile normalized	2	70.75	40.43	0.86	82.14	4	78.43	64.85	0.79	89.29
Baseline	9	98.36	69.18	0.78	92.86	3	87.05	56.59	0.76	92.86
Detrend	13	99.40	87.46	0.60	100.00	12	99.95	69.25	0.89	73.21
Deresolve	11	99.85	78.06	0.65	96.43	1	67.92	23.72	0.92	100.00
SNV	8	96.48	75.34	0.76	96.43	8	98.99	66.24	0.82	92.86
MSC	5	95.74	61.52	0.78	87.50	1	59.82	34.58	0.85	85.71
OSC	13	99.08	84.82	0.82	100.00	2	69.49	46.25	0.84	92.86
SG	15	98.37	83.81	0.60	98.21	15	99.62	77.24	0.82	92.86
1st	7	82.76	77.94	0.61	98.21	8	96.85	65.44	0.81	98.21
2nd	10	86.65	81.85	0.57	96.43	14	98.72	80.49	0.77	94.64
		sync 60 nm						**sync 185 nm**		
Normalized	15	99.9	84.04	0.67	98.21	10	99.69	68.31	0.77	100.00
Quantile normalized	4	74.94	70.39	0.72	94.64	4	78.43	64.85	0.79	91.07
Baseline	2	55.11	40.82	0.88	83.93	1	57.54	12.01	0.98	71.43
Detrend	5	92.36	61.59	0.84	96.43	11	99.89	71.61	0.85	82.14
Deresolve	14	100	77.96	0.71	98.21	10	99.81	67.79	0.8	98.21
SNV	3	77.45	58.52	0.78	91.07	1	46.67	35.79	0.86	85.71
MSC	10	99.17	81.45	0.65	96.43	1	43.61	33.16	0.87	83.93
OSC	11	95.52	82.41	0.74	100.00	2	69.49	46.25	0.84	91.07
SG	8	84.19	77.33	0.77	96.43	3	55.99	62.76	0.78	91.07
1st	6	89.16	69.39	0.71	92.86	2	58.33	49.19	0.81	87.50
2nd	15	95.94	83.8	0.86	98.21	15	99.86	73.47	0.78	94.64

Table 11.7: PLS-DA performance using external cross-validation for the selected variables VIP>0.8 & SLC-DA for the SEEFs recorded at different δ nm under different pretreatments

	External Validation VIP > 0.8 & SLC-DA											
	Sync 10 nm						**Sync 80 nm**					
	LV	*%X*	*%Y*	*PRESS*	*%A*	*%P*	*LV*	*%X*	*%Y*	*PRESS*	*%A*	*%P*
Normalized	3	62.15	77.19	0.61	98.21	100.00	3	73.43	55.39	0.79	91.07	100.00
Quantile normalized	5	70.87	90.27	0.64	98.21	90.91	2	67.56	16.84	0.92	89.29	100.00
Baseline	14	99.93	92.06	0.61	98.21	90.91	10	99.33	71.61	0.85	98.21	100.00
Detrend	15	99.94	92.8	0.6	92.86	90.91	6	97.62	59.16	0.88	89.29	90.91
Deresolve	3	63.78	65.91	0.77	**100.00**	**100.00**	5	93.82	58.02	0.82	98.21	100.00
SNV	9	97.53	85.71	0.67	**100.00**	**100.00**	5	96.67	60.18	0.81	92.86	100.00
MSC	14	99.79	90.6	0.57	**100.00**	**100.00**	3	94.29	40.92	0.89	89.29	90.91
OSC	6	72.77	83.89	0.67	96.43	81.82	3	61.96	61.18	0.84	98.21	90.91
SG	2	50.66	75.11	0.61	96.43	100.00	6	83.97	72.82	0.7	98.21	100.00
1st	15	99.69	88.53	0.72	**100.00**	**100.00**	15	99.66	84.17	0.66	92.86	81.82
2nd	15	96.22	95.21	0.4	**100.00**	**100.00**	13	95.48	86.83	0.69	94.64	100.00
	Sync 30 nm						**Sync 120 nm**					
Normalized	8	98.97	73.10	0.77	96.43	90.91	13	99.87	81.79	0.71	94.64	81.82
Quantile normalized	2	69.72	43.64	0.83	78.57	81.82	3	71.15	61.71	0.78	89.29	81.82
Baseline	7	95.77	63.13	1.03	96.43	100.00	6	99.98	66.97	0.67	98.21	100.00
Detrend	14	99.49	88.75	0.60	85.71	100.00	11	99.94	76.47	0.8	75.00	63.64
Deresolve	3	93.17	44.60	0.85	94.64	90.91	1	70.73	27.5	0.91	98.21	100.00
SNV	9	97.11	46.68	0.91	98.21	100.00	6	97.41	68.41	0.72	92.86	100.00
MSC	6	96.13	64.62	0.86	96.43	100.00	11	99.74	70.87	0.77	96.43	100.00
OSC	1	35.00	18.82	1.00	76.79	72.73	2	67.1	56.47	0.75	92.86	100.00
SG	3	64.56	68.84	0.70	98.21	100.00	2	68.71	54.83	0.76	94.64	90.91
1st	3	61.69	68.63	0.72	96.43	90.91	8	96.78	73.46	0.72	98.21	90.91
2nd	7	79.46	81.67	0.65	96.43	90.91	2	66.31	57.28	0.81	91.07	90.91
	Sync 60 nm						**Sync 185 nm**					
Normalized	14	99.87	88.03	0.7	96.43	90.91	10	99.8	72.79	0.71	98.21	90.91
Quantile normalized	5	81.02	77.61	0.75	96.43	90.91	3	71.15	61.71	0.78	87.50	72.73
Baseline	1	30.89	33.23	0.87	83.93	90.91	10	99.96	64.39	0.89	91.07	90.91
Detrend	12	99.69	81.06	0.75	98.21	100.00	12	99.94	79.27	0.76	87.50	90.91
Deresolve	13	99.99	77.54	0.81	98.21	90.91	8	99.27	60.63	0.91	98.21	100.00
SNV	3	78.05	62.78	0.78	94.64	90.91	11	99.95	67.69	0.8	91.07	90.91
MSC	10	99.3	82.39	0.67	96.43	90.91	7	99.29	67.49	0.73	96.43	100.00
OSC	3	61.67	63.38	0.72	91.07	72.73	2	67.1	56.47	0.75	91.07	90.91
SG	5	74.55	74.91	0.72	96.43	90.91	15	100	81.74	0.8	94.64	100.00
1st	6	90.17	70.46	0.7	94.64	90.91	2	58.95	44.78	0.88	87.50	90.91
2nd	9	84.15	84.27	1.06	98.21	90.91	7	93.86	66.12	0.81	96.43	100.00

SIMCA analysis was carried out again on the different SEEFs spectral pretreatments however the second stage variable selection process negatively affect the overall of the models obtained. This was attributed to the loss of variables which although they might not be contributing for the discrimination of the two EVOO classes, they are responsible for the inherited variation within the subgroup. The loss of this variation within each subgroup made modelling the individual groups far more difficult, as consequences the majority of the models obtained suffered for a lower sensitivity for the respective classes as the majority of the samples could not be modelled separately but rather classified as belonging to both Maltese and non-Maltese EVOOs. The only model which showed an increase in performance, in terms of the sensitivity towards the individual classes were those obtained using SEEF δ10 nm after Savitzky Golay derivatisation. For this particular spectrum, the variables selected by the combination of the two techniques managed to achieve a sensitivity of 91% and 97% for the Maltese and non-Maltese EVOOs respectively, whilst retaining a 100% specificity for both classes. Figure 11.9 shows the Coomans plot obtained for the SEEF obtained δ10 nm after Savitzky Golay derivatisation and two-step variable selection.

Comparison with the previously obtained spectrum, it was evident that very few samples (three in total two Maltese and one non-Maltese) were found to be classified as belonging to both classes. Figure 11.8 also illustrates the modelling power for each class (red bars for the non-Maltese EVOOs and black bars for the Maltese EVOO group). Since as previously demonstrated the SIMCA models are sensitive to the quality of the data used to generate the principal component models. Analysis of both the modelling power and the discriminatory power of the variables used provides a diagnostic assessment of the quality of the data. The modelling power provides insights about how well a variable helps the principal components to model variation. From the Figure 11.9 it was shown that variables selected in the range 300-450 nm and 510-650 nm had a higher modelling power for the Maltese EVOOs, whilst those obtained in the range 280-350 and 465-490 nm had a higher modelling power for modelling the non-Maltese EVOO group. The discriminatory power, which is represented as the green line above the bar graph, describes how well the variables helps the principal components to classify the samples in the correct data set. It was shown that variables selected in the range 345-380 nm and 525-555 nm had a very high discriminatory power for SEEF spectrum obtained at δ10 nm after Savitzky Golay

derivatisation. These ranges correspond to the tocopherol fluorescent peaks and shouldering peaks of chlorophyll pigments.

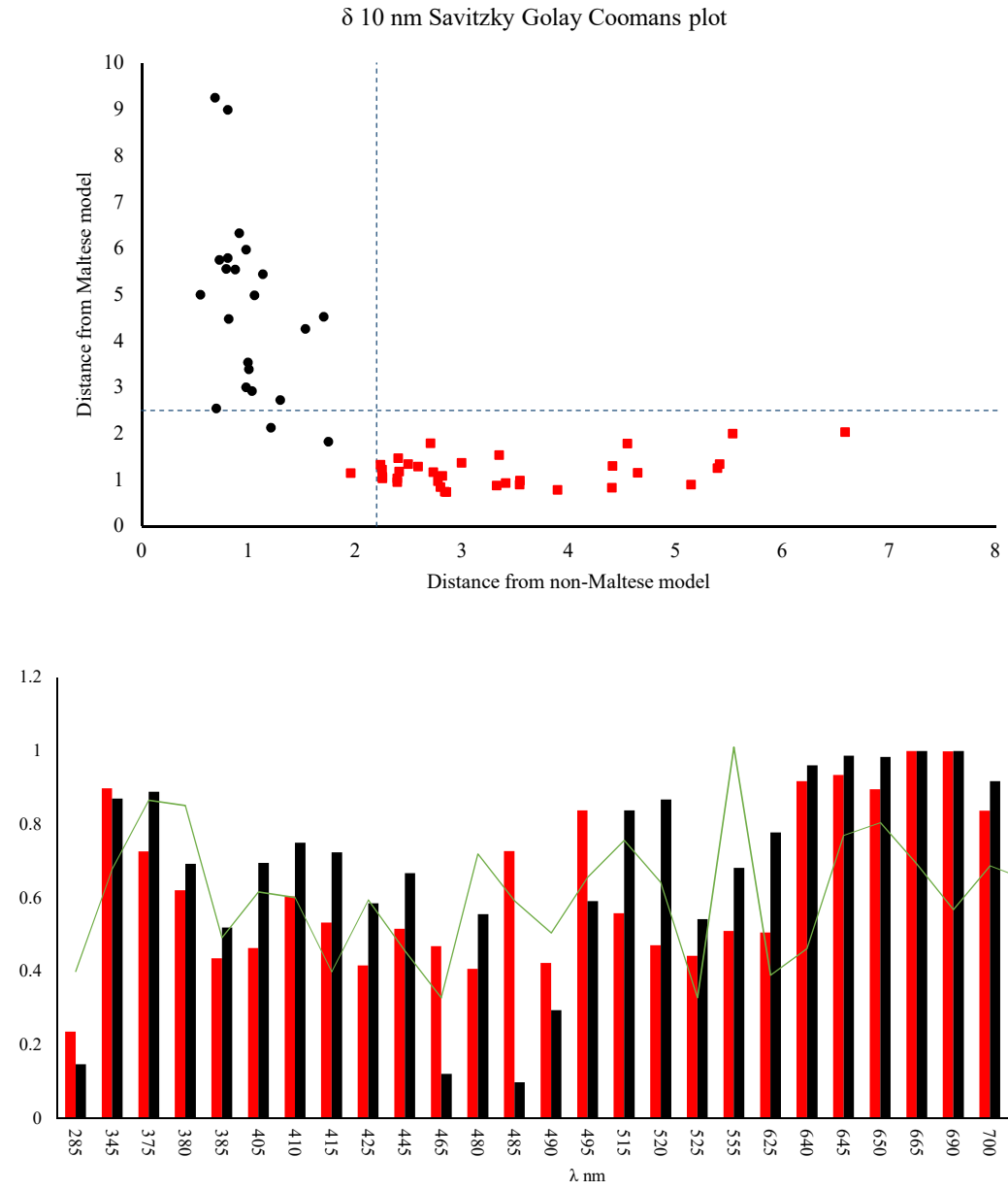


Figure 11. 8: (Top) Coomans plot obtained for the Savitzky Golay derived SEEF obtained at 10 nm. (Bottom) Modelling power of variables with respect to class, (Black= Maltese, Red = non-Maltese origin), the green line represents the discriminatory power for the different variables.

11.2.4 Supervised chemometric discriminate analysis techniques – LDA

In order to build the LDA, the selected variables obtained in SLC-DA were arranged in ascending order in terms of their scoring coefficients. A smaller set of variables were selected which consisted of 14 variables which corresponded to 7 of the most positive and 7 most negative standardized scoring coefficients. An LDA was carried out on the training set using only the small set of variables which were selected. The results obtained for the training samples were a visualized on an LDA biplot samples as shown in Figure 11.9 whereby each sample is projected as the scores obtained for the first two discriminate functions. When compared with SIMCA and PLS-DA, the LDA method has the disadvantage that the number of training samples must be larger than the number of variables included in the LDA model. In order to fully satisfy this constraint a smaller number of variables were selected based on the standardized scoring coefficients obtained from the SLC-DA. The standardized scoring coefficients of the variable selected during the SLC-DA were obtained and plotted as shown in Figure 11.9.

The importance of these coefficients lies in their use to compute canonical scores in terms of the standardized data often referred to as loadings. They are highly informative when it comes to comparing the relative importance in their discriminatory power of the independent variables. Analysis of the standardised scoring coefficients, it was shown that at a SEEF obtained at δ10 nm the majority of the discriminate variables were found in the 650-750 nm range which is concordant to the results obtained using PLS and SIMCA. In the case of SEEF spectra obtained at δ30 nm, the majority of the spectral pretreatments had the most discriminate variables located in the 300-450 nm range whilst only OSC, Savitzky Golay, 1st and 2nd order derived spectra picked up significant variables in the 650-750 nm range. For the reaming SEEF spectra the majority of the spectral pretreatments revealed highly discriminate variables across the entire spectrum, with an increasing number of highly discriminate variables located in 400-550 nm range.

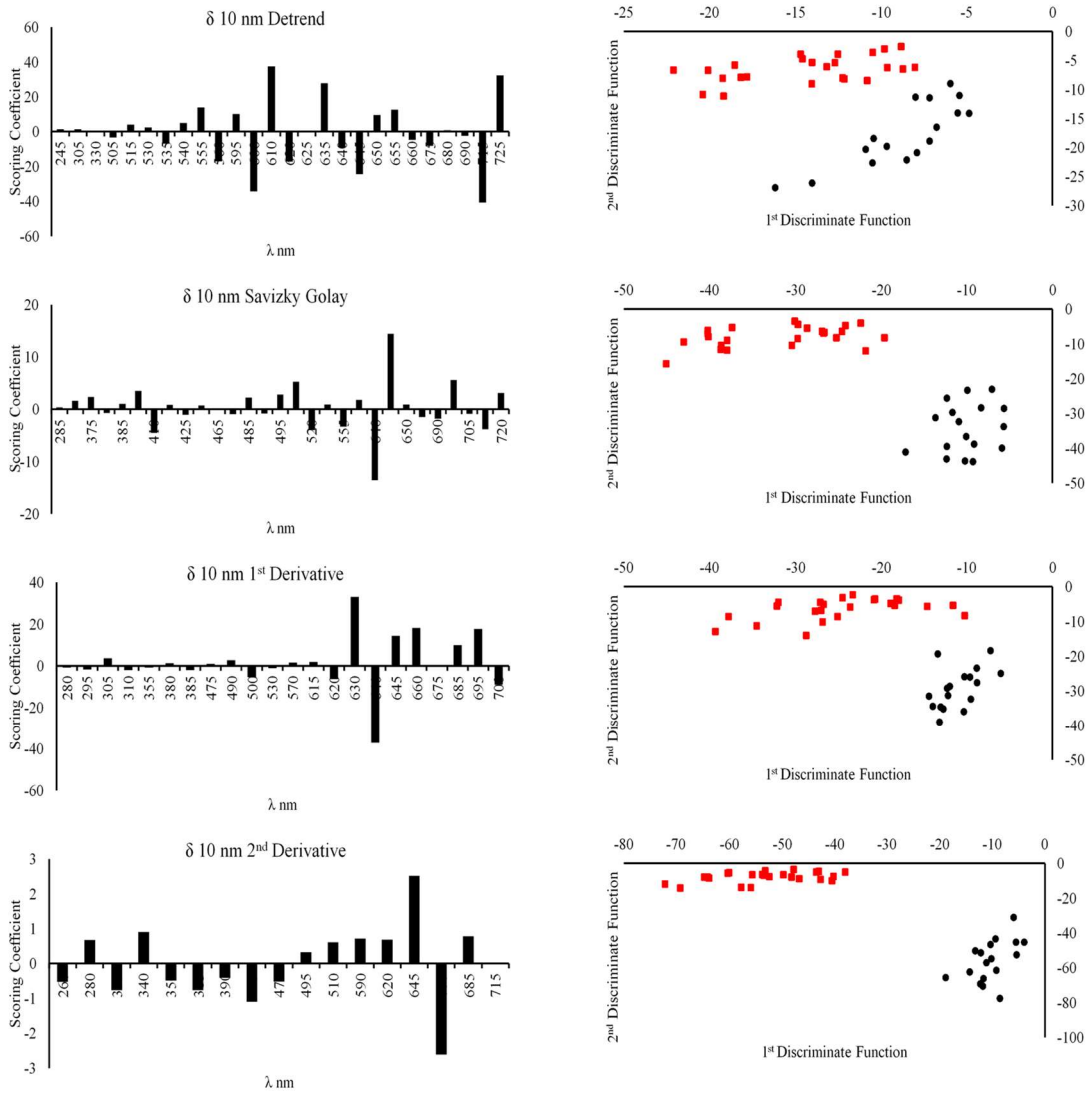


Figure 11.9: LDA carried out on a selected number of variables for the different spectra pretreatments obtained for the different SEEF obtained at δ10.

The classification model obtained was then repeated on the testing data set, with the exception of the detrended spectrum obtained at SEEF δ185, the validation accuracy ranged from 66-100%. From the results obtained in Table 11.8 it was further confirmed that the SEEF spectra obtained at δ10 nm had the best discriminatory power. SEEF spectra at δ10 nm after pretreatment with SNV, MSC, Savitzky Golay, 1st and 2nd order derivation obtained the best model performance (100% accuracy and predictability) when compared to all the SEEF spectra and there corresponding pretreatments. Comparison of the LDA model performance with those obtained using PLS it was found that a similar trend in the overall performance models was observed, however for the majority of the spectral pretreatments a slightly lower performance was observed for the LDA models. Similar to what was observed in other spectroscopic techniques the models obtained using LDA had a higher classification and discriminatory power when compared to SIMCA model. This difference stems from the fact that in that SIMCA is a soft classification method and thus allows a single object to belong to more than one class, whilst LDA, PLS are a more robust form of classification as the objects are either classified in either one or the other class.

Table 11.8: LDA performance given as % accuracy for the training dataset and % predictability for the testing dataset.

LDA	δ10 nm		δ30 nm		δ60 nm		δ80 nm		δ120 nm		δ185 nm	
	% A	%P	% A	%P	% A	%P	% A	%P	% A	%P	% A	%P
Normalized	100.0	93.3	95.0	73.3	97.5	73.3	95.0	66.7	92.5	73.3	95.0	80.0
Quantile normalized	97.5	66.7	77.5	53.3	90.0	66.7	85.0	80.0	87.5	80.0	90.0	80.0
Baseline	100.0	93.3	90.0	80.0	97.5	86.7	92.5	80.0	92.5	86.7	92.5	73.3
Detrend	97.5	100.0	97.5	80.0	90.0	86.7	95.0	66.7	92.5	66.7	97.5	13.3
Deresolve	100.0	86.7	97.5	80.0	92.5	66.7	97.5	86.7	92.5	86.7	95.0	73.3
SNV	**100.0**	**100.0**	100.0	93.3	97.5	46.7	92.5	73.3	82.5	66.7	85.0	73.3
MSC	**100.0**	**100.0**	95.0	80.0	95.0	86.7	87.5	86.7	90.0	73.3	95.0	80.0
OSC	95.0	73.3	87.5	73.3	92.5	73.3	95.0	60.0	85.0	73.3	87.5	73.3
SG	**100.0**	**100.0**	97.5	93.3	95.0	80.0	95.0	66.7	92.5	86.7	85.0	80.0
1st	**100.0**	**100.0**	97.5	86.7	95.0	60.0	95.0	86.7	92.5	93.3	97.5	66.7
2nd	**100.0**	**100.0**	97.5	86.7	97.5	86.7	97.5	80.0	92.5	86.7	97.5	86.7

11.2.5 Supervised discriminate chemometric techniques – SVM

The dataset containing only variables which were selected using SLC-DA and having a VIP score > 0.8, were subjected to another classification method, known as support vector machine (SVM). During this part of the experiment, SVM was used in the context of classification. Classification methods based on SVM employ the use of linear boundaries which are produced between discrete groups in a transformed space. The space in which the groups are projected is in terms of the x-variables which is usually of much higher dimension than the original space generated by the x-variables. The increase in the dimensional space allows the groups to become more linearly separable. The boundary (hyperplane) that separates the two classes projected into this higher dimensional space is known as the maximum margin classifier. Comparison of SVMs with other classification and regression methods found out that they show mostly good performances, although other methods proved to be very competitive (Meyer *et al*. 2003). Table 11.9 shows the results obtained on using SVMs using no Kernel tricks.

Table 11.9: SVM performance given as % accuracy for the training dataset and % predictability for the testing dataset.

SVM	δ10 nm		δ30 nm		δ60 nm		δ80 nm		δ120 nm		δ185 nm	
	% A	%P	% A	%P	% A	%P	% A	%P	% A	%P	% A	%P
Normalized	77.5	73.3	77.5	73.3	85.0	80.0	77.5	80.0	85.0	80.0	80.0	80.0
Quantile normalized	100.0	80.0	100.0	73.3	100.0	66.7	100.0	66.7	100.0	80.0	100.0	80.0
Baseline	**100.0**	**100.0**	97.5	86.7	100.0	80.0	97.5	60.0	95.0	86.7	87.5	60.0
Detrend	100.0	80.0	100.0	86.7	**100.0**	**86.7**	100.0	66.7	92.5	86.7	100.0	73.3
Deresolve	**100.0**	**93.3**	97.5	80.0	100.0	66.7	100.0	73.3	75.0	60.0	100.0	80.0
SNV	95.0	80.0	95.0	86.7	92.5	73.3	85.0	86.7	90.0	86.7	92.5	73.3
MSC	100.0	86.7	100.0	86.7	100.0	80.0	100.0	53.3	90.0	73.3	100.0	60.0
OSC	100.0	80.0	100.0	66.7	100.0	80.0	100.0	46.7	100.0	66.7	100.0	66.7
SG	**100.0**	**93.3**	97.5	73.3	100.0	80.0	97.5	60.0	92.5	86.7	100.0	73.3
1st	100.0	80.0	97.5	80.0	97.5	73.3	97.5	80.0	92.5	93.3	92.5	73.3
2nd	**100.0**	**100.0**	97.5	86.7	100.0	73.3	100.0	80.0	95.0	86.7	100.0	66.7

Comparison of the SVM models obtained to those obtained using PLS, the trends observed were conserved. It was further confirmed that SEEF spectra obtained at δ10 nm had a higher performance when compared to the rest of SEEF spectra, with baseline corrected spectra, detrended, Savitzky Golay and 2nd order derived spectra reaching a %100 accuracy during the testing phase and a 93-100% in testing phase. For the majority of the SEEF spectral pretreatments a lower predictability was observed when compared to PLS and LDA models. In general SVM models obtained had a slightly higher % accuracy during the training phase when compared to other discriminate methods, especially on comparing the SEEF spectra obtained at δ80 nm δ120 nm and δ185 nm. However, when it comes to the validation data set the models obtained have a lower predictability (60-80%). This is mainly attributed to the technique its self, it is well known that SVMs do not perform well on highly imbalanced data sets. These are training data sets in which the number of samples that fall in one of the classes far outnumber those that are a member of the other class. Although, this was tried to be avoid the slightly higher number of non-Maltese EVOO samples might have affected the performance of the SVM models obtained.

It is proposed that SVMs should be carried out on the entire data set rather than a small number of selected variables, this might improve the performance of the models, since SVMs are not affected by the dimensionality of the data since unlike other discriminatory techniques data having lot of features is overcome by the multidimensionality of the SVM models.

11.2.6 Whole SEEF modelling using feed-forward predictive artificial neural networks.

The use of feed-forward predictive neural networks on the MS data as a method for classification was assessed using three different forms of validation, namely 33.3% of data holdback, CV-10 k-fold and excluded row validation. PNN have been employed successfully in the determination of rapeseed oils in EVOOs using SEEF data by Scott *et al*., 2003, however to date the use of PNN for the discrimination of EVOOs has not been employed .The main advantage of a neural network model is that it can efficiently model different response surfaces due to its nonlinearity, allowing a better fit to the data given enough hidden nodes and layer, providing an accurate prediction for kind of data. Unlike other modelling and discriminate methods (SIMCA, LDA, PLS) the main disadvantage of a neural network model is that the results are not easily interpretable, due to presence of intermediate hidden layers which direct path from the X variables to the Y variables, as in the case of regular regression but cannot be fully interpreted. The models obtained for all the different SEEF spectra and their corresponding transformations are presented in Table 11.10 The algorithm fitted on the training set was later tested on the validation data and % predictability of the model was obtained.

From the results obtained it was shown that the performance of the PNN model is dependent not only on the SEEF and the data preprocessing used but also on the cross-validation method employed. In general excluded row validation and 5 fold cross-validation (CV-10) yielded models with a higher performance. This is concordant with other spectroscopic data including FTIR and NMR which was studied and presented later on. This constant trend, was attributed to the fact that whilst CV-10 and excluded row validation rely on, the experimenter to choose the training and validation data set, the 33% holdback is purely random, thus increases the chance that the model is built using uneven or skewed number of samples belonging to different classes. This results in model which is inheritably biased towards one particular class and fails to identify samples of the other class when it comes to the validation data set, causing a lower % predictability.

PNN models further corroborate the trends obtained using PLS on the entire spectrum without any form of feature selection. It was shown that SEEF obtained at

δ10 nm δ80 nm δ120 and δ185 nm had a higher PNN performance throughout all the cross-validation methods used, reaching 100% accuracy and 100% predictability for the majority of spectral pretreatments. Nonetheless an overall higher performance was observed throughout all the SEEF spectra studied for the different pretreatments, independent of the cross-validation method used. This observation was attributed to the higher flexibility and modelling power of PNNs when compared to PLS-DA. Similar to what was observed in the PLS models SEEF spectra pretreated using detrending function and Savitzky Golay, 1st and 2nd order derived spectra had the highest PNN model performance throughout all the cross-validation methods used.

Table 11.10: Results summarizing the FF-ANN model performance with no variable selection using three different cross-validation methods for the different SEEFs pretreatments obtained δ10, 30, 60, 80, 120, 185 nm

δ10 nm PNN

	33% Holdback		10-CV		Excluded Row	
	% A	%P	% A	%P	% A	%P
Normalized	92.73	90.91	94.55	90.91	92.73	90.91
Quantile normalized	94.55	81.82	98.18	90.91	94.55	90.91
Baseline	**100**	**100**	**100**	**100**	**100**	**100**
Detrend	**100**	**100**	**100**	**100**	**100**	**100**
Deresolve	94.55	100	**100**	**100**	96.36	100
SNV	98.18	100	**100**	**100**	**100**	**100**
MSC	**100**	**100**	98.18	90.91	**100**	**100**
OSC	92.73	81.82	98.18	90.91	98.18	100
SG	**100**	**100**	**100**	**100**	**100**	**100**
1st	**100**	**100**	**100**	**100**	**100**	**100**
2nd	**100**	**100**	**100**	**100**	**100**	**100**

δ30 nm PNN

	33% Holdback		10-CV		Excluded Row	
	% A	%P	% A	%P	% A	%P
Normalized	94.55	90.91	96.36	90.91	94.55	90.91
Quantile normalized	87.27	72.73	92.73	81.82	96.36	90.91
Baseline	96.36	100	90.91	90.91	98.18	100
Detrend	94.55	100	94.55	90.91	94.55	90.91
Deresolve	**100**	**100**	**100**	**100**	96.36	100
SNV	92.73	81.82	96.36	90.91	94.55	90.91
MSC	**100**	**100**	98.18	100	92.73	90.91
OSC	89.09	81.82	94.55	81.82	96.36	90.91
SG	94.55	90.91	94.55	90.91	96.36	90.91
1st	96.36	100	96.36	100	92.73	90.91
2nd	94.55	90.91	96.36	100	94.55	90.91

δ60 nm PNN

	33% Holdback		10-CV		Excluded Row	
	% A	%P	% A	%P	% A	%P
Normalized	**100**	**100**	98.18	100	92.73	100
Quantile normalized	94.55	81.82	96.36	90.91	98.18	90.91
Baseline	98.18	90.91	98.18	100	98.18	100
Detrend	**100**	**100**	**100**	**100**	**100**	**100**
Deresolve	**100**	**100**	**100**	**100**	96.36	100
SNV	89.09	72.73	92.73	100	**100**	**100**
MSC	98.18	100	**100**	**100**	87.27	90.91
OSC	96.36	90.91	98.18	90.91	**100**	**100**
SG	**100**	**100**	**100**	**100**	98.18	100
1st	**100**	**100**	**100**	**100**	**100**	**100**
2nd	96.36	100	**100**	**100**	98.18	100

δ80 nm PNN

	33% Holdback		10-CV		Excluded Row	
	% A	%P	% A	%P	% A	%P
Normalized	94.55	100	**100**	**100**	**100**	**100**
Quantile normalized	90.91	90.91	98.18	100	**100**	**100**
Baseline	96.36	100	**100**	**100**	**100**	**100**
Detrend	**100**	**100**	**100**	**100**	**100**	**100**
Deresolve	**100**	**100**	**100**	**100**	**100**	**100**
SNV	98.18	100	**100**	**100**	**100**	**100**
MSC	92.73	100	**100**	**100**	**100**	**100**
OSC	98.18	100	**100**	**100**	**100**	**100**
SG	98.18	100	**100**	**100**	**100**	**100**
1st	**100**	**100**	**100**	**100**	**100**	**100**
2nd	**100**	**100**	**100**	**100**	**100**	**100**

δ120 nm PNN

	33% Holdback		10-CV		Excluded Row	
	% A	%P	% A	%P	% A	%P
Normalized	96.36	100	**100**	**100**	**100**	**100**
Quantile normalized	92.73	90.91	94.55	72.73	96.36	90.91
Baseline	**100**	**100**	**100**	**100**	**100**	**100**
Detrend	98.18	100	**100**	**100**	**100**	**100**
Deresolve	94.55	90.91	**100**	**100**	96.36	90.91
SNV	96.36	90.91	**100**	**100**	**100**	**100**
MSC	96.36	90.91	98.18	100	**100**	**100**
OSC	94.55	90.91	94.55	90.91	96.36	90.91
SG	87.27	81.82	**100**	**100**	96.36	90.91
1st	94.55	100	**100**	**100**	98.18	100
2nd	90.91	90.91	**100**	**100**	96.36	100

δ185 nm PNN

	33% Holdback		10-CV		Excluded Row	
	% A	%P	% A	%P	% A	%P
Normalized	87.27	100	**100**	**100**	**100**	**100**
Quantile normalized	89.09	90.91	96.36	100	98.18	100
Baseline	85.45	100	**100**	**100**	**100**	**100**
Detrend	89.09	90.91	**100**	**100**	96.36	81.82
Deresolve	**100**	**100**	**100**	**100**	**100**	**100**
SNV	94.55	100	**100**	**100**	**100**	**100**
MSC	85.45	100	**100**	**100**	98.18	100
OSC	92.73	90.91	96.36	90.91	**100**	**100**
SG	90.91	100	**100**	**100**	**100**	**100**
1st	89.09	100	**100**	**100**	**100**	**100**
2nd	**100**	**100**	**100**	**100**	**100**	**100**

In conclusion, it was shown SEEF spectra in conjunction with a number of chemometric methods, provided a cheap, fast and reliable way for the determination of geographical origin of EVOOs, especially when it comes to discrimination of Maltese EVOOs from non-Maltese EVOOs. From the preliminary assessment using only unsupervised PCA models, only very few spectral pretreatments for different SEEF spectra managed to identify significant clustering, such method was deemed to be unsatisfactory when it comes to discrimination of geographical origin. Application of supervised methods of classification namely PLS-DA, FF-PNN, LDA and SVM showed to be highly effective in classifying local and non-local EVOOs samples.

The use of the variable selection methods significantly increased the effectiveness of PLS-DA models when compared to no variable selection however this was not the case for SIMCA as lower model performance was recorded for all the spectral pretreatments for all SEEF spectra. FF-PNN, SVM and LDA models were also shown to offer similar classification rates to PLS-DA models and thus corroborate the results obtained from the PLS-DA models, assuring that the use of SEEF methods in conjunction with spectral transformation enables discrimination of Maltese and foreign EVOOs samples. Results showed that different SEEF spectra can greatly affect the discrimination of EVOOs. It was shown that independent of the chemometric technique used SEEF spectra obtained at δ10 nm however a higher model performance. It was shown that the most discriminate variables were those attributed to different concentration of phenolic, tocopherol and chlorophyll compounds. These observations further corroborate the results obtained from using target-specific analysis, whereby it was shown that Maltese EVOO had a significantly higher amount of different non-reducing phenolic compounds and a higher concentration of chlorophyll compounds. The quantification of the individual tocopherol compounds is being proposed as a future study, as a target-specific method for further confirmation of the results obtained using SEEF spectroscopy.

12 Multi-way Data analysis of fluorescence spectroscopic data

Due to the sheer diversity of the fluorophores present in a single olive oil sample it is not adequate to simply measure an emission spectrum at a single emission wavelength to gather sufficient. This problem can be eased through the use of synchronous spectroscopy or multidimensional measurements, assessing the emission spectra at different excitation wavelengths. Synchronous fluorescence spectroscopy, along with 3D-FS, are nowadays accepted to be more suitable for the analysis of complex multi-component samples, than conventional fluorescence spectroscopy. SEEF arises from the simultaneous scanning of excitation and emission wavelengths (λ_{ex}, λ_{em}) with a constant wavelength interval ($\Delta\lambda$) between them rather than scanning the whole EEM as in the case of 3D-FS. The latter method involves scanning a 3-dimensional landscape consisting of the excitation-emission matrix (EEM) and total luminescence (Lenhardt *et al*., 2015). This technique (3D-FS) provides a more detailed and comprehensive description of the numerous fluorophores present (Sergeil *et al*., 2014). Complementary to the study carried out using SEEF, 3D-FS was implemented in this study along with the appropriate multivariate statistical techniques for classification of local and foreign EVOO samples.

In standard multivariate analysis the data is organised in a two-way structure, a matrix of observed variables for each sample, such as in analysis of FT-IR, NMR, DI-MS and SEEF data where the absorbance is determined at set wavelength intervals. This data can be directly analysed through a numerous of bilinear multivariate techniques such as PCA, PLSR, LDA, SVM and FF-NNA. Nonetheless there are cases where the data is more appropriately represented using an additional dimensions. The theory behind the use of 3-way methods like PARAFAC and N-PLS is present in Appendix 18.

12.1.1 PARAFAC modelling

From the results obtained it was determined that a four-component PARAFAC model was the optimum model. This was determined on the basis of explained variance, residual variance, core consistency and split similarity. Similar results were obtained by Guimet *et al*., (2004) whereby four-component model (explained variance 98.7%) was found to be optimal in distinguishing between commercial samples of virgin and pure olive oils. Furthermore, studies conducted by the same authors in 2005 showed that the application PARAFAC as a complementary technique for olive oil characterization indicated that the optimal number of factors was three (98.65% of explained variance). The difference in the optimal number of PARAFAC components obtained by the same authors can be explained in terms of the initial data inputted whilst the study focused on the discrimination of pure olive oils the second study carried out in 2005 focused mainly on relating the quality parameters of different EVOO grades to the EEM.

Table 12.1 showed that on moving further than five components there was a decreased improvement with regards to explained variance and residual variance. Furthermore there was a sharp decrease in core consistency when more than five components were used. A core consistency of zero or lower is a strong indicator that PARAFAC models with more than five components were not stable and thus inappropriate for modelling the EVOO fluorescence. A low core consistency in fact suggests that the components are not just modelling trilinear trends in the data but also other random variation within the data array. Values significantly lower than 100 frequently suggest that the model is either inconsistent or badly constrained and might improve if the model is slightly altered or constrained (Bro, 1998). No improvement was observed for the five-component model even after several replicates. The low value for split similarity for the five-component was used to discount the model as inappropriate and thus the four-component model was chosen as the optimum due to the relatively high core consistency and split similarity. The splits for mode two and mode three (Figure 12.1, 12.2) were quite similar to the full model however variation in the shape of the loadings is expected as fluorophores are very sensitive to slight changes in the pH, moisture content and viscosity of the samples (Dufour and Riaublanc, 1997) and thus the shape of fluorescence spectra will vary.

Table 12.1: Results from non-negative constrained PARAFAC model. The scaled residual variance, explained variance, Core consistency and Split similarities are shown from models having between 1 and 9 components. The 4 component model was deemed to have the optimum number of components. *For presentation reasons the residual variance is scaled to the maximum residual variance.

	Core Consistency %	Explained Variation %	Residual Variation	Split-Half Similarity %
PC1	100.00	62.86	0.048	98.81
PC2	98.80	86.42	0.132	97.60
PC3	90.12	93.12	0.260	96.20
PC4	**91.97**	**95.21**	**0.373**	**88.54**
PC5	83.93	96.47	0.507	59.18
PC6	-169.51	97.19	0.636	44.40
PC7	-72.12	97.58	0.739	0.00
PC8	-699.24	97.83	0.826	0.00
PC9	-11,859.88	98.21	1.000	0.00

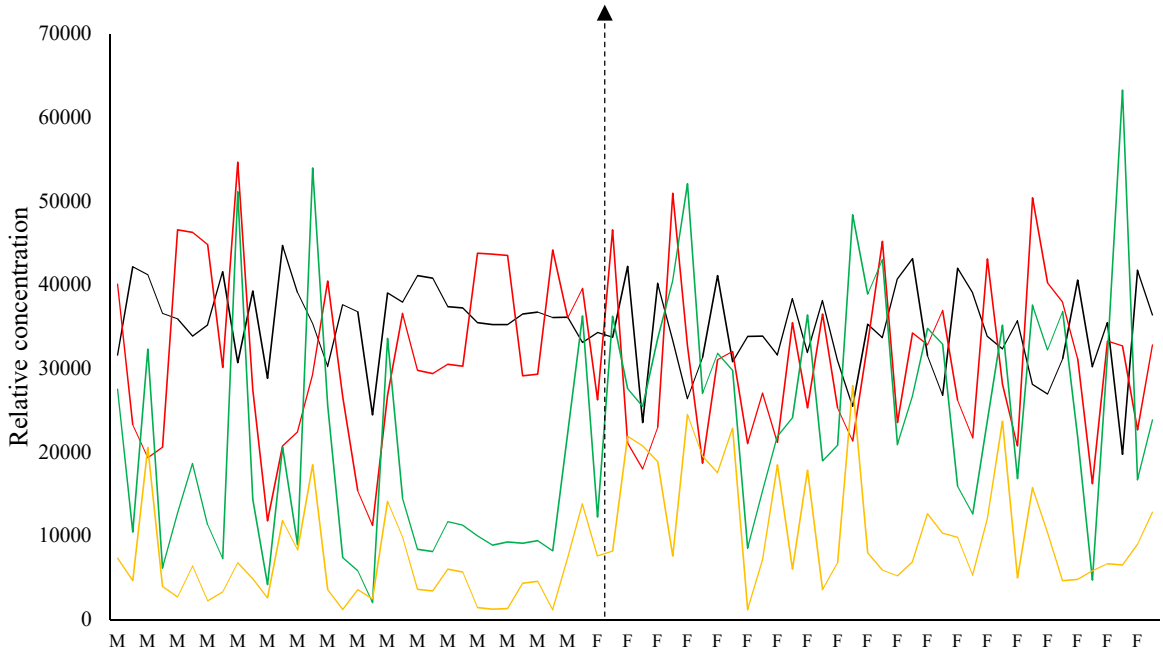


Figure 12.1: Mode 1 loadings (relative concentrations y-axis) from non-negative constrained 4 component PARAFAC model, and how they vary for the different samples (x-axis). Samples of Maltese origin are denoted by the letter M whilst foreign samples are denoted by the letter F.

For each component, the PARAFAC modelling was repeated for at least ten times in order to determine whether the results obtained were consistent. Repeating the modelling process ensures that the results obtained were not as a result of the model being stuck at a local minimum. In order to further prevent the model getting stuck in a local minimum, the SVD algorithm was used for initialisation starting the model from a random starting point each time. Since PARAFAC is an iterative solution which attempts to minimise the sum of squares error it is always advisable to select a random starting point in order to check for true convergence (Bro, 1998).

Unlike what was carried out on the use of simpler spectra no form of data pre-processing apart from trimming was applied before PARAFAC or N-PLS modelling. Multi-way data is often difficult to transform and furthermore any transformations which change the shape of the spectrum considerably will reduce the interpretability of the results. Furthermore, any changes within the profile of the spectra will most probably interfere with the non-negativity constraints imposed on the mode. Whilst in the case of bilinear methods, it is frequently suggested that data is scaled and centred prior to modelling, in the multiway case, centring and scaling are associated with a number of difficulties. In this part of the study Multi-way scaling was not applied since the data was all from the same source (fluorescence) and does not require scaling. In the case of multi-way centring this was investigated on mode 1 only however, it was observed that centring causes the PARAFAC models to perform poorly. In essence, centring causes the data to have negative values in order to provide a mean of 0. This in turn will interferes with the non-negativity constraint of the model rendering it unsolvable. The solution to this is to un-constrain mode 1 in order to allow for modelling of negative values however this returned models which were not chemically interpretable (Bro, 1997). Although centring and scaling were suggested by Guimet *et al*., (2004) and (2005) these were not carried out for the aforementioned reasons. Guimet suggested that if centring is to be applied on a three-way array it needs to be applied across all the modes on the unfolded matrix. Similarly scaling, must be done on the rows of a matrix and not across the columns of the unfolded matrix, as is the case with centring. This involves matrix transformation in order to transpose the columns (ex. emission into rows and then scale) however as suggest by Bro, (1997) not all combinations between centring and scaling are possible when working with three-way data. The use of non-negativity constraints in all modes overcomes the

requirement for centring as shown by split half analysis which gave comparable spectra in the excitation and emission loading modes for different data splits.

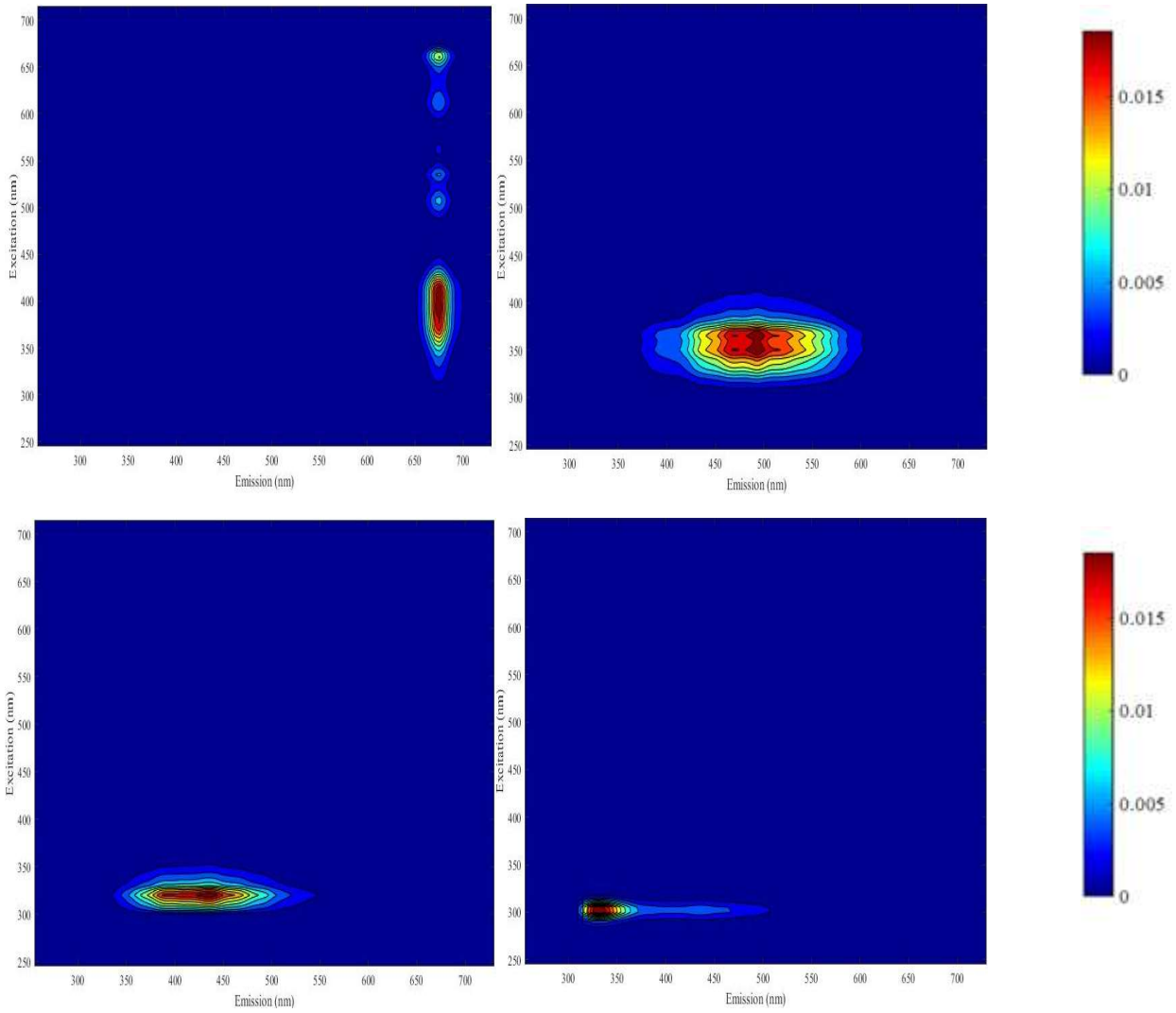


Figure 12.2: Mode 2 (Excitation) and mode 3 (Emission) PARAFAC components from 4 component mode represented as an EEM spectrum

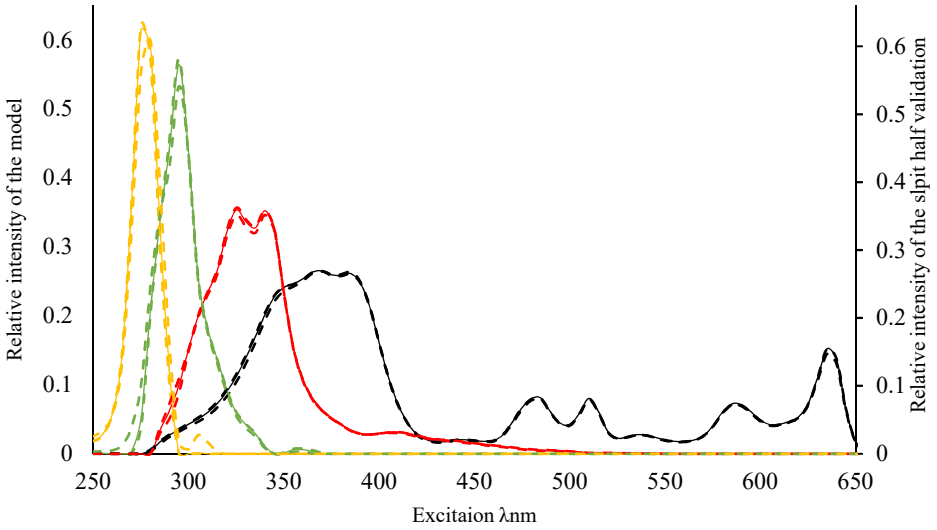


Figure 12.3: Mode 2 loadings (excitation) from non-negative constrained 4 component PARAFAC model, dotted lines represent components from split-half validation models. *1st component (Black), 2nd Component (Red), 3rd Component (Green) and 4th Component (Yellow).*

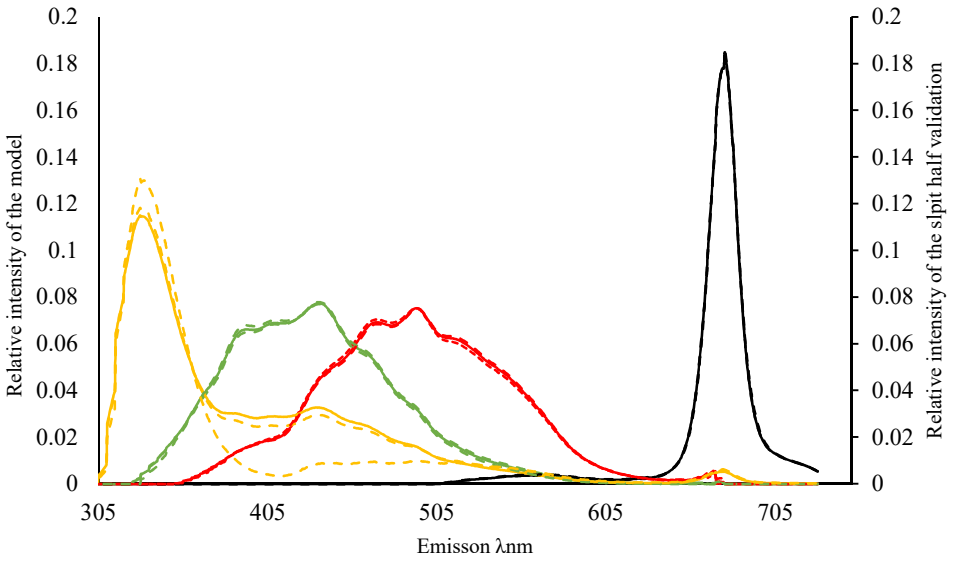


Figure 12.4 Mode 3 loadings (emission) from non-negative constrained 4 component PARAFAC model, dotted lines represent components from split-half validation models. *1st component (Black), 2nd Component (Red), 3rd Component (Green) and 4th Component (Yellow).*

Figure 12.2 shows the four extracted PARAFAC Components on an EEM plot while Figure 12.3 and Figure 12.4 show the four components in the excitation and emission mode respectively. The component maxima are shown in Table 12.2. Assuming that a four-component model is appropriate in describing the data set, it is reasonable to assume that these four components have maxima appertaining to a particular chromophores found in EVOO. Table 12.2 shows the identified chromophores through the observed maxima. The first component was attributed to chlorophylls having an emission band with a maximum at 675 nm (λem), which is associated with the presence of chlorophyll pigments in the samples (Dupuy *et al*., 2005). The emission profile of second factor showed a band with a maximum at 525 nm with an excitation at 325 and 340 nm these were, assigned to oxidation compounds (Dupuy *et al*.,2005). This band (λem=450–650 nm) slightly overlaps with the 3rd component however it's completely absent in the emission profile of the 4th component and 1st component. These results are consistent with the PARAFAC results obtained by Tena *et al*., (2012) and earlier by Guimet *et al*., (2004) and (2005). In the results obtained by Tena *et al*., (2012) the remaining 3rd component showed a characteristic band with a maximum at 350 nm (λem) and 285 nm (λex), which was collectively associated with the presence of tocopherols and phenols as previously identified by Sikorska *et al*., (2005) and Zandomeneghi *et al*., (2005). However in this experiment it was shown that it is possible to distinguish between the 3rd and 4th component.

The 3rd component was identified as belonging to the tocopherols and tocotrienols the wide emission band was attributed to the presence of different isomeric forms of the different classes namely α-β-δ-γ which as previously identified by Eitenmiller *et al*., (2008.) These compounds have an excitation in the range 290-297 nm and emission in the range of 320-324 nm. In the case of the 4th component, this was attributed to the presence of phenolic compounds present EVOOs, Tena *et al*., (2009) showed that phenolic compounds belonging to the secoiridoid class (oleuropein) had an excitation at 270 nm and emission at 310 nm whilst simple phenolic acids (gallic, vanillic, caffeic) and simple phenolic alcohols (tyrosols) shared the same excitation maxima however the emission spanned from 349-457 nm. In view of these results one might think that it is erroneous to assign the 3rd component to tocopherols and the 4th component to phenolic compounds on the basis of what is found in the literature. However, it is first of all, important to note that the values in the

literature were obtained on the analysis of pure compounds alone not in conjugation with each other, thus can synergistic or antagonistic effects on the emission and excitation could be excluded, secondly these were carried out in the presence of polar solvent namely ethanol and methanol. Whilst in this experiment there was no attempt to separate the individual compounds, furthermore the EVOO was dissolved in an aprotic non-polar solvent. The fluorescence of fluorophores is greatly affected by the solvent polarity. Once that the fluorophore has been excited to the first excited singlet state (S(1)), any excess of vibrational energy is lost to surrounding solvent molecules. Solvent molecules assist in stabilizing and further lowering the energy level of the excited state by re-orienting around the excited. This effect is known as solvent relaxation which reduces the energy separation between the ground and excited states, which in turn results in a bathochromic shift of the fluorescence emission. Any increasing the solvent polarity causes a larger reduction in the energy level of the excited state, conversely a decreasing the solvent polarity reduces the solvent effect on the excited state energy level (hypsochromic shift). Apart from solvent polarity itself the polarity of the fluorophore also exacerbates this effect, polar and charged fluorophores exhibit a far stronger effect than non-polar fluorophores. A similar conclusion was drawn by Cheikhousman *et al*. (2005) whereby it was shown that the excitation and emission maxima obtained at λem=380 nm and λex=295 nm agree very well with the respective spectra of α-tocopherols, whilst the hypsochromic shift compared to α- tocopherol observed at λem=300 nm and λex=280 nm, these were attributed to the phenolic compounds.

Table 12.2: Excitation and Emission maxima obtained from the PARAFAC loadings of components 1-4 and the proposed chemical constituents giving rise to the components

	λ_{max} Excitation	λ_{max} Emission	Chemical Class
1st Component	355-400, 480, 510, 596,640	676	Chlorophylls
2nd Component	325,340	434-550	Oxidation Products
3rd Component	295	386-468	Tocopherols
4th Component	280	330	Phenolic Compounds

12.1.1.1 Statistical analysis of the PARAFAC components.

The relative concentration of each component is represented by the scores (mode 1) of the PARAFAC components. These were analysed using both univariate and multivariate statistical analysis. Univariate normality testing for the different components revealed that whilst the first two components, chlorophyll and oxidised by-products concentration in EVOOs had a normal distribution under Shapiro-Wilk's Normality Test with a p-value 0.2673 and 0.5886 respectively. On the other had the 3rd and the 4th component which correspond to the concentration of tocopherols and phenolic compounds respectively, had a non-parametric distribution with p-values of 0.0039 and < 0.0001. Analysis of variance between the Maltese and Foreign EVOOs for the first two components was carried out using ANOVA whilst for the 3rd and 4th component this was carried out using non-parametric Kruskal-Wallis test. From the results obtained it was found that EVOOs of Maltese origin tend to have a marginally significantly higher chlorophyll concentration (p-value 0.090* significant at the 0.1 significance), whilst no significant difference (p-value 0.419) was observed in the concentration of oxidized by-products between the Maltese and foreign EVOOs. The inference of this results suggest that previous results obtained through the use of simple UV-Vis spectroscopy and later by SEEF were once again confirmed using PARAFAC model. Furthermore, these results suggest that all the samples obtained were fresh and that there were no gross outliers in terms of EVOO oxidation.

Non-parametric Kruskal-Wallis test revealed that the Maltese EVOOs had a significantly lower concentration of both the phenolic compounds and tocopherols (p-value < 0.001 for both components). In the case of phenolic content, the results obtained seem to be conflicting with those obtained using HPLC whilst on the other hand they tend to confirm those obtained using microtiter Folin–Ciocalteu assay. This suggests that the major fluorescent phenolic compounds are those which show a more reducing power rather than the more abundant secoiridoids since the latter were found to be significantly higher in Maltese EVOOs. Although, the above hypothesis still needs to be tested in detail through the application of fluorescence-HPLC, whereby we can obtain a fully fluorescent spectrum of each and every eluting species, the results obtained tend to support the aforementioned hypothesis. It was during the application of PARAFAC to EEMs of EVOOs that the first results regarding the composition of

tocopherols in Maltese EVOOs was encountered. In EVOO, more than 90% of total tocopherols are accounted by α-tocopherol, which shows high variation according to soil and climatic conditions and agronomic factors, such as area of origin, cultivar, and fruit ripening stage (Servili *et al*. 2009). Kalogeropoulos and Tsimidou (2014) stated that as a general rule, γ-T content in EVOO is influenced by the genetic factor more than the other tocopherol forms. Furthermore the same authors state that unlike phenolic content the tocopherol content and composition is not very dependent on the degree of olive ripeness and extraction system. Although, a general decrease in tocopherol content was observed during ripening by a number of authors (Solinas 1990; Di Matteo *et al*. 1992; Garcia *et al*. 1996; Gimeno *et al*. 2002) whilst others showed changes in α-tocopherol (Sakouhi *et al*. 2008). It is well known that the EVOOs extracted from olive oils growing at lower altitude and water stressed tend to have a higher total tocopherol content (Osman *et al*. 1994; Mousa *et al*. 1996; Paz Aguilera *et al*. 2005; Tovar *et al*., 2002).

Application of non-parametric correlation analysis on the concentrations of phenolic and tocopherol contents obtained through PARAFAC showed that these two components had a significantly strong positive correlation with each other (ρ=0.772 p-value < 0.001) (Figure 12.5 Top). Furthermore it was also found that the tocopherol content had a significantly slightly positive correlation (ρ=0.335 p-value = 0.004) (Figure 12.5 middle) with the amount of oxidized by-products present in EVOOs. There are a number of factors which affect the oxidative stability of vegetable oils these include the fatty acid composition and antioxidants, mainly tocopherols but also other non-saponifiable constituents. Although tocopherols and tocotrienols, especially α- and γ-tocopherols, act as the major antioxidants in vegetable oils. Their antioxidant behaviour of tocopherols is a complex phenomenon as they tend to show an efficient antioxidants at low concentrations but they gradually lose efficacy as their concentrations increase (Kamal-Eldin 2006). This “prooxidant effect”, of tocopherols could explain the significant positive correlation that was observed between the tocopherol content and amount of oxidised by-products found in EVOO. Kamal-Eldin (2006) also states that the presence of other antioxidants in the oils such as phenolic compounds, may synergize with tocopherols and minimize this loss of antioxidant efficacy. This observation can also explain the significantly slight negative correlation (ρ=-0.306 p-value = 0.009) (Figure 12.5 Bottom) which was observed between the

tocopherol and chlorophyll content. As shown in the EVOO degradation Section 14, during the EVOO degradation process chlorophyll content tend to decrease rapidly during the initial stages of degradation, furthermore it was also shown that during the degradation process there is in fact an increase in phenolic content derived from the hydrolysis complex secoiridoid compounds hence explaining the positive correlation between the tocopherol content and phenolic content.

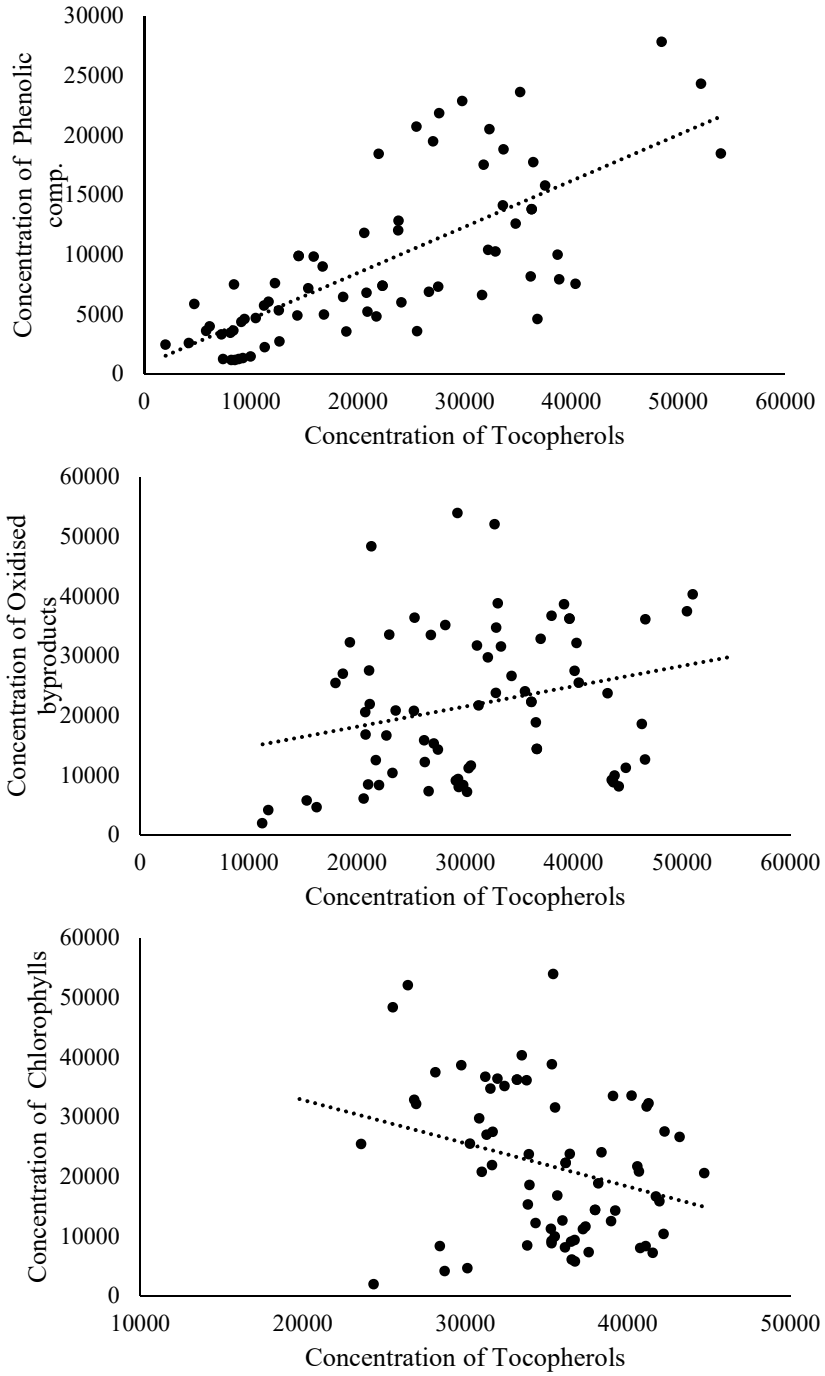


Figure 12.5: Correlation analysis between the individual components extracted

Application of multivariate statistical analysis, was carried out in order to determine whether the components extracted from mode 1 (concentrations) using PARAFAC model would enable discrimination between EVOOs of different geographical origin and whether or not these are suitable predictors for future classifications. A linear discriminate analysis was carried out on the extracted mode concentrations obtained. Discriminant analysis enables the construction of a predictive model for group membership. The prediction model obtained was constructed using a linear combinations of predictor variables. Furthermore, this model also provides insights on those predictor variables which provide the best discrimination between groups. A number of assumptions are made during the construction of discriminate model the main one being that the predictor variables have a multivariate normal distribution, in this case it is know that this is not true since the variables obtained had a significant deviation from normality under a number of multivariate normality tests (Doornik and Hansen omnibus Test, Mardia's, Henze-Zirkler, and Royston p-value < 0.0001). The second most important assumption is that the within-group variance-covariance matrices should be equal across groups. This assumption was tested using Box's M test which although p-value obtained was lower than 0.0001, on further inspection of the log-determinants for the two classes these were found to be equal (Maltese origin = 69.57 Foreign origin = 70.89). Thus equality of group variance was assumed to be true and that the significant value p-value obtained was attributed to the Box's M susceptibility to non-normal data.

Discriminant function analysis is robust even when the homogeneity of variances assumption is not met, provided the data do not contain important outliers. However, a QDA was also done in order to compare the unequal group variance senior. The discriminate function obtained had an eigenvalue of 0.331, 100 % variance explained and a canonical correlation of 0.499. From the results obtained the discriminate function obtained seems to be very robust with a large eigenvalue, high explained variance in the dependent variable. On further inspection of the Wilk's lambda obtained (0.751, p-value < 0.001) a small significant value was obtained indicating that the function obtained has a good discriminatory power and is able to separates cases into groups. The significant p-value obtained for Wilk's lambda indicate that the discriminant function does better than chance at separating the groups. Figure 12.6 shows the standardized discriminant function coefficients (Black Bars)

these indicate the relative importance of the independent variables in predicting the dependent variable. From the results obtained it was shown that the tocopherol content had the largest absolute value, suggesting that this particular class of compounds had the largest discriminating ability. Figure 12.6 also shows the Pearson's correlations of each variable with the discriminant function (Red Bars) obtained from the structure matrix. It was shown that the tocopherol and phenolic content has the largest absolute correlations associated with the discriminant function. These results are in agreement with the results obtained using univariate statistics.

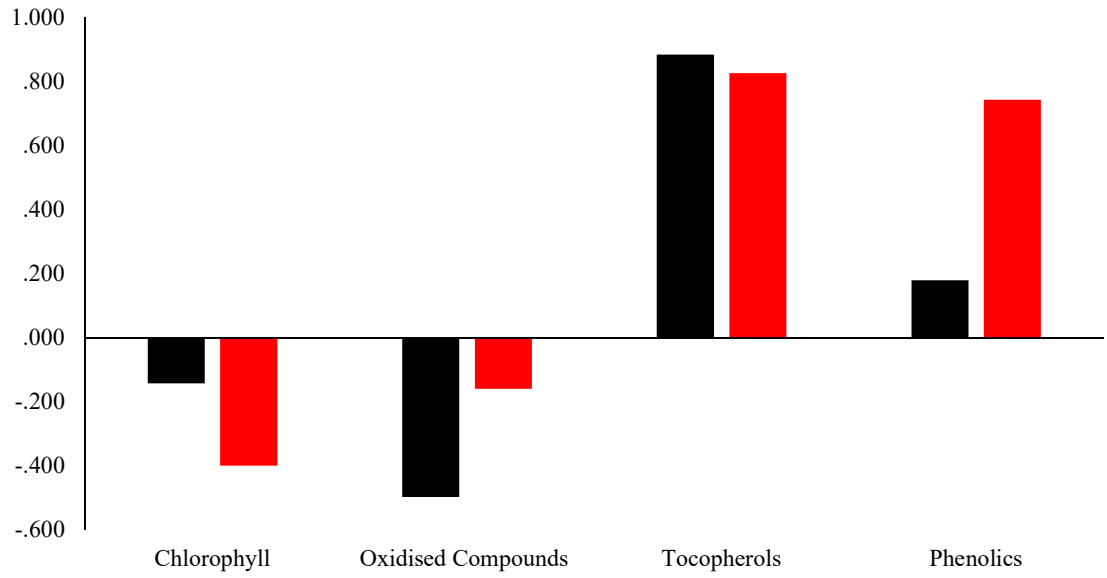


Figure 12.6 : Standardized discriminant function coefficients (Black Bars) and Pearson's correlations of each variable with the discriminant function (Red Bars) obtained from the structure matrix

Analysis of the discriminate model performance, it was shown that the model obtained was able to correctly classified 73.0% of original data and 67.6% of cross-validated grouped cases were correctly classified. Figure 12.7 shows the plotted discriminant scores obtained for each case using the discriminate function obtained. Although there is a good % of data which is correctly classified there is still some overalap between the two classes and when compared to other methods especially FTIR data, the performance obtained tends to be on the low side, thus alternative classification methods need to be tested. In order to further assess the ability of discriminate analysis for EVOO classification quadratic discriminate analysis (QDA) and Mahalanobis distance discriminate analysis (MDA) was carried out. Although the

model performance improved on using the latter type of discriminate analysis (QDA = 77.46% and MDA =80.28% during the cross-validation stage) it was still lower in performance when compared to other methods.

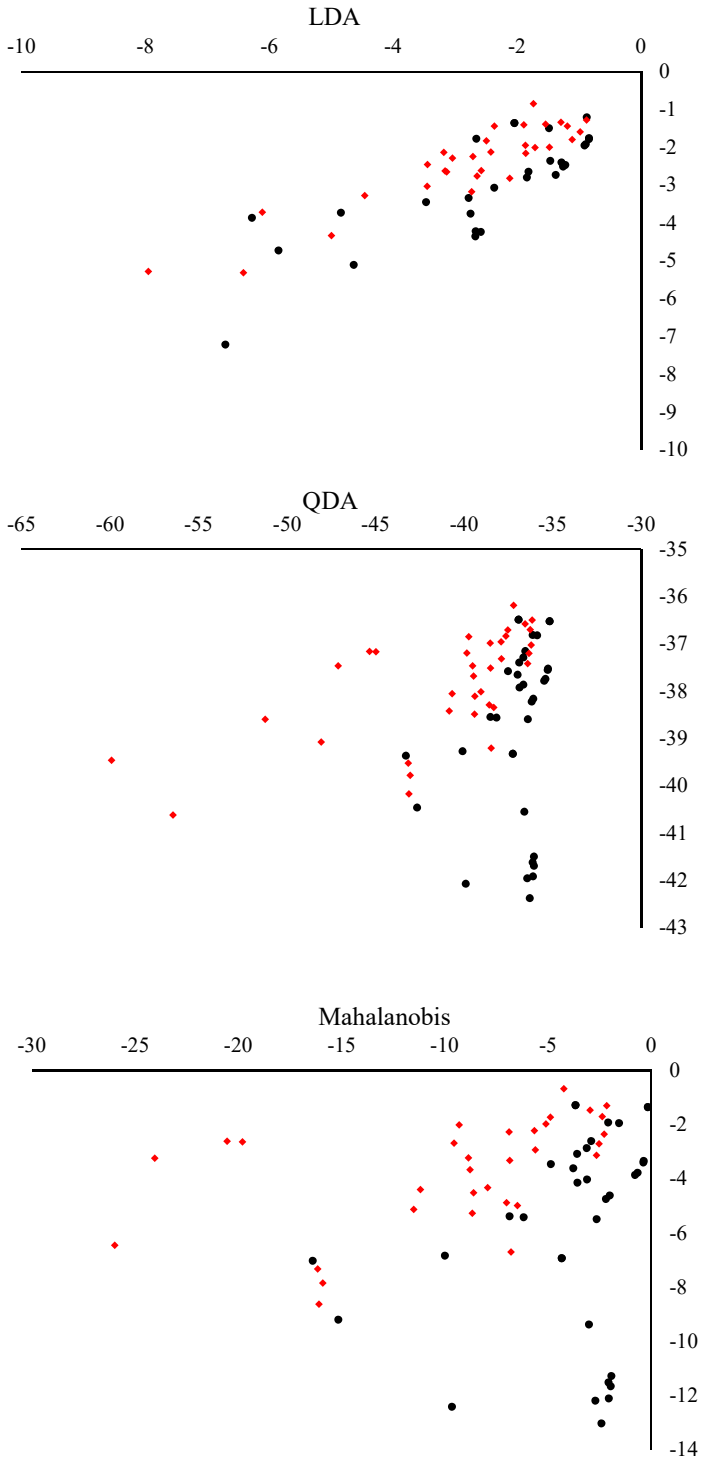


Figure 12.7: Discriminate analysis biplot obtained using different Bayes LDA (Top), QDA (Middle) and non-parametric Fisher (Bottom).

The next multivariate method to which the components in mode 1 were subjected was logistic regression. In this form of regression the parameters of the model are estimated using the maximum likelihood method. That is, the coefficients that make the observed results most likely are selected. Unlike other multivariate methods logistic regression does not make many of the key assumptions of linear regression and general linear models that are based on ordinary least squares algorithms namely regarding linearity, normality, and equal variances. The model starts by determining the probability of the outcome without including any of the parameters also known as the 'null model' (only a constant), thus it is expected that if there is an equal chance that the outcome will be 0 (Maltese) or 1 (Foreign), the overall % of correct classification should be in the 50% range. From the results obtained a 57% of correct classifications was obtained for block 0 during the logistic regression indicating equal chance of the EVOO being Maltese or foreign when no variables are included. The model then compares whether there is a significant difference between the null model and the one in which the variables are included.

From the results obtained tocopherol and phenolic content had a p-value < 0.001, while chlorophyll content had a p-value of 0.049 indicating that the inclusion of these variables cause a significant change in the model predictability when these are included. Only the oxidised by-products (p-value 0.636) seem not to be significantly affecting the performance of the model. The Omnibus test of model coefficient was significant (p-value < 0.001) indicating that there is an effect of the independent variables, taken together, on the dependent variable. On the other hand the Hosmer-Lemeshow Goodness of fit test was not significant (p-value 0.285) this suggests that the model is an adequate fit to the data.

Table 12.3 shows the B regression coefficient obtained for all the predictors form the results obtained it was shown that the model was mainly based on the tocopherol content as it had the largest most significant coefficient. Overall the logistic regression model obtained was able to correctly classify 74.3 % of EVOOs of Maltese origin and 79.5% of foreign EVOOs.

Table 12.3: Logistic regression output on the extracted PARAFAC mode 1 components

							95% C.I.for EXP(B)	
	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Chlorophyll	-6.350	4.978	1.627	1	.202	.002	.000	30.158
Oxidized.Comp	-7.485	2.998	6.234	1	.013	.001	.000	.200
Tocopherols	6.190	2.018	9.407	1	.002	487.859	9.341	25479.675
Phenolics	-.706	1.470	.231	1	.631	.493	.028	8.803
Constant	38.748	25.823	2.252	1	.133	$6.73x10^{16}$		

12.1.2 N-PLS model

In comparison to the PARAFAC models obtained N-PLS-DA analysis, was shown to be quite effective for classification of local and foreign samples as shown in Table 12.4. The advantages of using N-PLS are mainly its robustness (Bro, 1998) and the high classification rates even when no data pre-treatment or variable selection is used. Whilst data pre-treatment might improve the prediction rate by removing unnecessary information and instrumental artefacts, transformations of multi-way arrays are very complex and most transformations which exist for bilinear data or their multi-way equivalents are not readily available for multi-way data.

Table 12.4 illustrates different model parameters on using different number of latent variables, each parameter is represented as an average value and ± 1SD obtained from using 3 different splits. The calibration accuracy and RMSEC indicate model performance during the calibration (training) stage while validation accuracy and RMSECV deal with the model performance of the model with regards to the validation samples. These values along with % explained variance, are indicators of the suitability of the model for prediction. The 12 component model was chosen to be the most suitable model as it had the highest classification rate in the validation stage and a relatively low RMSECV and RMSEC. In fact the 12 component N-PLS model showed better performance when compared to the PLS-DA models on synchronous spectra previously computed using raw data without any form of variable. This is quite significant since the validation procedure for this model was more rigorous than the validation procedure used in the aforementioned PLS-DA models which were only

validated using one split, whilst in the N-PLS models a 3 spilt validation was carried out. Although higher the training accuracy of 3 split N-PLS models was observed on moving from 13 to 19 component N-PLS models these models tend to be unstable (overfitted) as the SD between the different splits tended to increase together with a lower model performance was observed during the validation stage.

Table 12.4: Compiled output obtained from the analysis of the three different splits carried out in DN-PLSR.

No.LV	% Variation X		% Variation Y		Overall % Variation		Training		Validation		RMSEC		RMSEV	
1	62.16	±0.57	57.97	±2.16	69.05	±9.18	70.21	±8.74	67.14	±10.98	0.48	±0.02	0.45	±0.06
2	73.86	±1.13	68.62	±2.07	77.14	±3.78	78.72	±6.49	74.95	±7.34	0.43	±0.02	0.44	±0.04
3	77.93	±1.97	70.60	±1.96	78.10	±3.60	78.01	±6.38	76.21	±6.17	0.41	±0.02	0.39	±0.06
4	85.18	±2.90	78.37	±1.74	80.95	±5.02	83.69	±5.38	78.87	±8.78	0.37	±0.02	0.40	±0.07
5	90.91	±1.48	81.83	±4.44	83.81	±4.36	89.36	±4.43	81.62	±10.82	0.33	±0.04	0.41	±0.08
6	93.25	±1.13	85.38	±3.55	87.62	±4.59	92.91	±4.58	85.45	±9.33	0.30	±0.03	0.40	±0.09
7	94.66	±1.24	87.29	±2.78	87.62	±5.95	93.62	±4.84	85.48	±12.20	0.28	±0.03	0.39	±0.08
8	95.13	±1.20	88.66	±2.14	88.57	±4.29	93.62	±3.94	86.78	±6.85	0.26	±0.02	0.39	±0.06
9	95.40	±1.26	89.67	±1.87	90.00	±3.78	94.33	±3.41	88.37	±6.85	0.25	±0.02	0.40	±0.02
10	95.67	±1.23	90.90	±1.93	91.90	±5.77	96.45	±4.29	90.06	±5.46	0.23	±0.02	0.40	±0.02
11	95.90	±1.27	91.98	±1.17	93.81	±4.36	98.58	±3.78	91.79	±5.49	0.22	±0.02	0.40	±0.02
12	**96.62**	**±1.14**	**93.00**	**±1.15**	**94.76**	**±3.30**	**98.58**	**±2.94**	**93.18**	**±5.19**	**0.20**	**±0.02**	**0.39**	**±0.01**
13	97.12	±0.64	94.06	±0.92	93.81	±2.97	98.58	±3.24	92.25	±5.19	0.19	±0.02	0.40	±0.00
14	97.52	±0.76	94.88	±0.78	92.86	±4.29	98.58	±3.71	91.30	±5.42	0.17	±0.01	0.43	±0.03
15	97.71	±0.64	95.77	±0.45	93.33	±4.36	99.29	±3.87	91.76	±5.66	0.16	±0.01	0.42	±0.05
16	97.86	±0.60	96.32	±0.46	92.86	±2.86	100.00	±4.76	91.11	±7.23	0.15	±0.01	0.42	±0.06
17	98.03	±0.62	96.64	±0.67	93.81	±2.18	100.00	±5.80	92.34	±6.14	0.13	±0.01	0.42	±0.07
18	98.17	±0.68	96.89	±0.71	94.76	±1.65	100.00	±4.39	93.51	±5.16	0.13	±0.02	0.41	±0.06
19	98.40	±0.45	97.47	±0.61	93.33	±3.60	100.00	±3.99	92.00	±7.09	0.11	±0.02	0.42	±0.03
20	98.56	±0.37	97.91	±0.46	61.43	±27.22	63.83	±27.29	66.89	±25.87	0.10	±0.01	0.62	±0.19

The loading components are represented in Figure 12.8, although these did not have distinctive similar shapes to fluorescence spectra as previously found during the PARAFAC analysis. Te likelihood of a distinct chromophore or a set of chromophores which are directly responsible for classification cannot be directly drawn. However, on further inspection it can be observed that in the case of excitation the loading tend to higher in the region of 260 - 320 nm and 370 - 450 nm, while in the case of emission these tend to be centered around 300 – 470 nm and 620 – 680 nm, thus although not distinctively these loading are in fact reflecting the 4 different fluorophores present in EVOO previously identified using PARAFAC. Whilst in the case of PARAFAC each fluorophore was described by a single component in the case of N-PLS each fluorophore is described by 2 or more components which when added together these

reveal revealed the fluorescent profile of the fluorophore. It is the complex interaction of these fluorophores together that is being used for the correct classification of the samples.

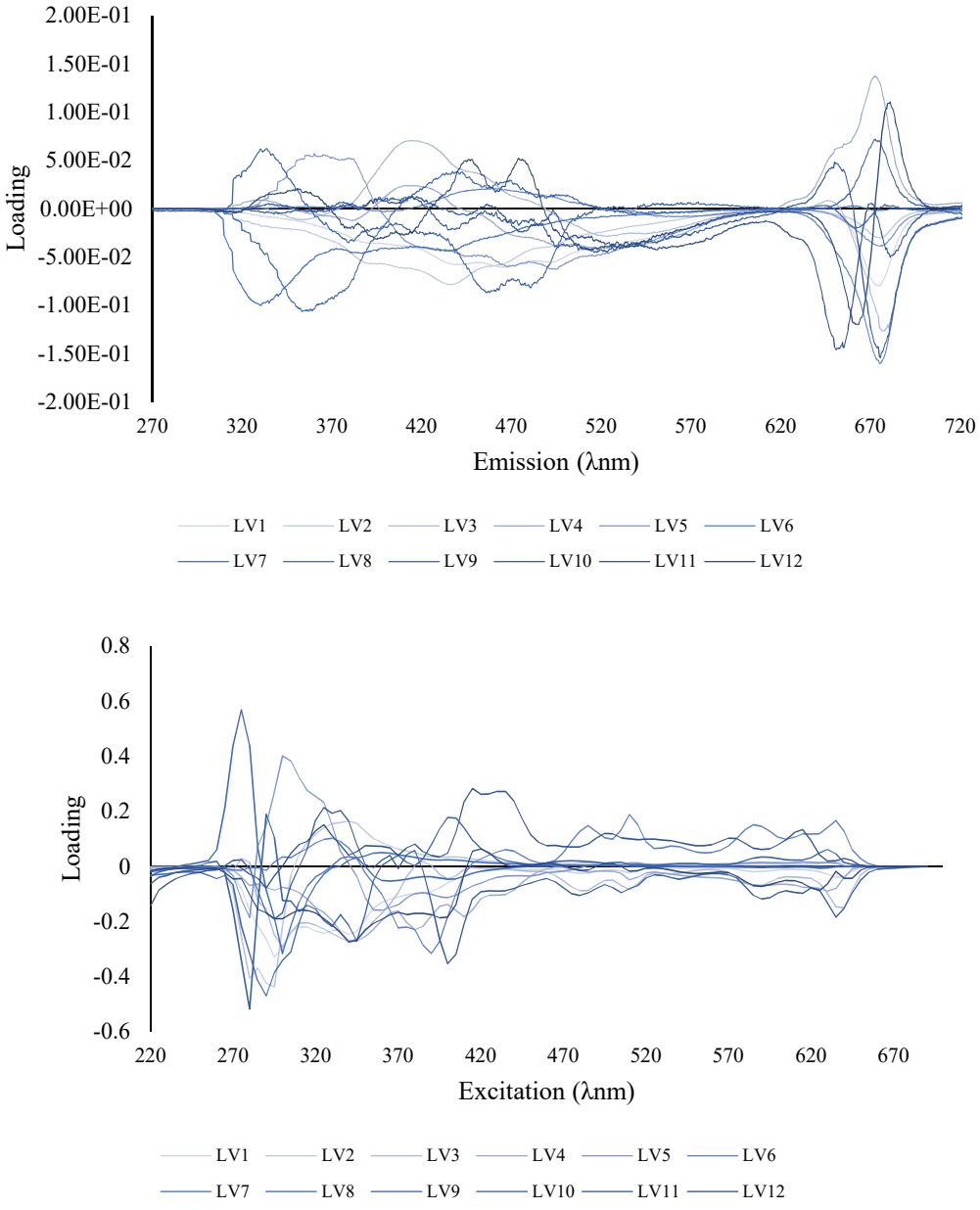


Figure 12.8: Loading obtained for emission (Top) and for excitation (Bottom) using 12 components.

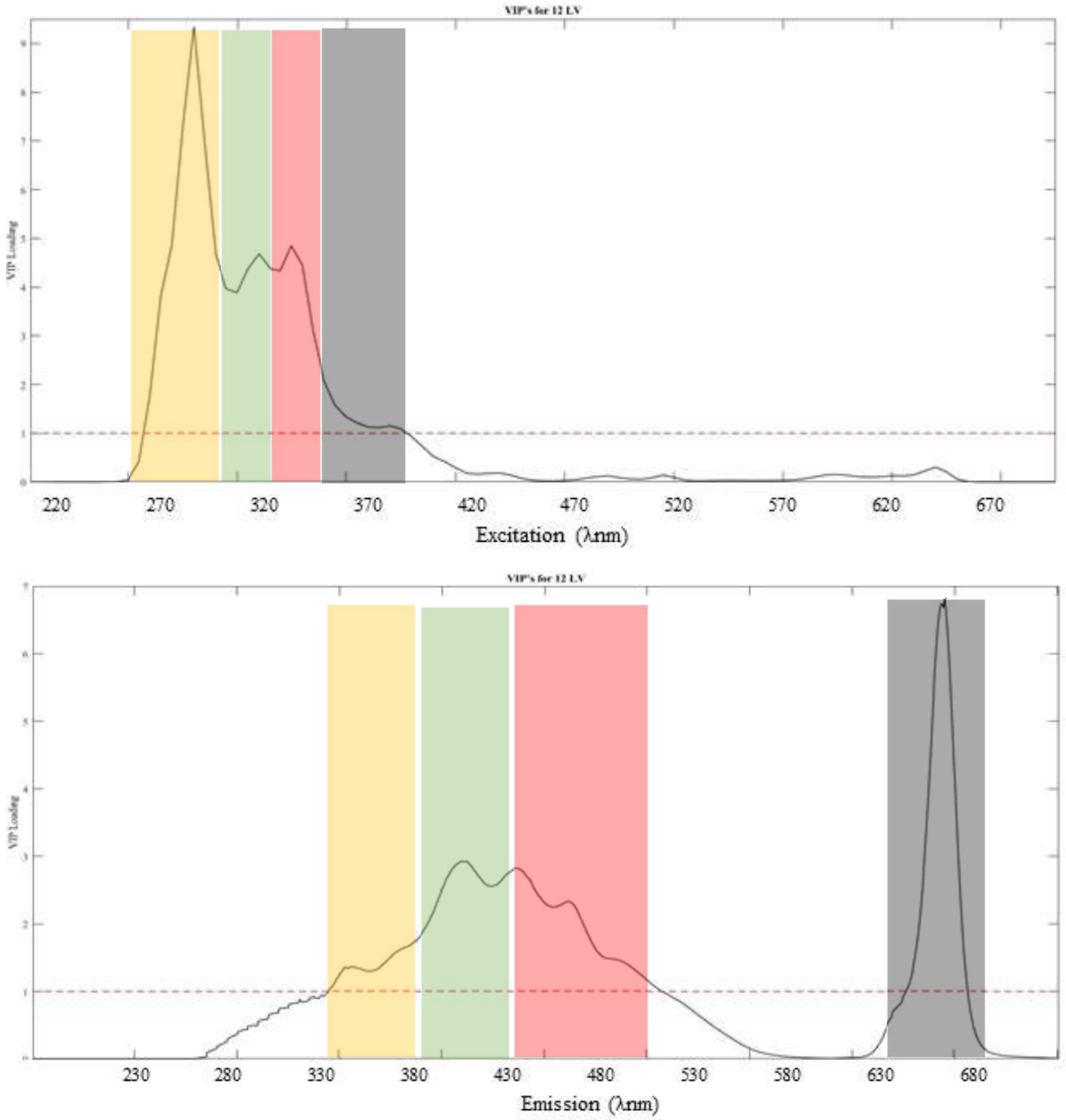


Figure 12.9: VIP loading of DN-PLSR highlighting the 4 main regions of importance as previously identified through PARAFAC.

Figure 12.9 shows a further inspection of the VIP's obtained from N-PLS revealed, that in fact the most discriminate predictors are those associated with the presence of the 4 distinct fluorophores previously determined during the PARAFAC analysis. Thus the 3-way methods, although differing in their discriminatory power between different EVOOs, the underlying concept of a 4 component fluorophore based discrimination tends to be corroborated.

13 Application of fluorescence spectroscopy for the detection of vegetable oil adulterants in Maltese virgin olive oils.

Numerous analytical practices have been established in recent years to safeguard the authenticity of olive oil. These include chromatographic techniques (Bosque-Sendra *et al*., 2012; Baccouri *et al*., 2008) and spectroscopic techniques, such as mass spectrometry (Calvano *et al*., 2012), nuclear magnetic resonance (Fragaki *et al*., 2005), near-infrared spectroscopy (Mignani *et al*., 2011), Raman spectroscopy (Dong *et al*., 2012), chemiluminescence (Papadopoulos *et al*., 2002), UV spectrometry (Jiang, Zheng & Lu, 2013), fluorescence spectroscopy (Sikorska, Khmelinskii & Sikorski, 2012), and synchronous fluorescence (Poulli, Mousdis & Georgiou, 2007). Compared to the other analytical techniques both UV and fluorescence spectroscopic techniques are ideal for the determination of olive oil adulteration, owing to their simplicity, cost-effectiveness, rapidity and non-destructive nature of the analysis. Fluorescence spectroscopy is more sensitive and selective in terms of organic and inorganic compounds than the other spectroscopic methods (Sikorska *et al*., 2004).

The fluorescence emission spectra of olive oils reveals five major bands. The 300–390 nm band provides information about their polyphenolic and tocopherol content (Zandomeneghi *et al*., 2005; Giungato *et al*, 2004). While the low-intensity doublet at 440 and 455 nm corresponds to oxidised fatty acids and phenolic antioxidants in virgin olive oils, which provide greater protection against oxidation of monounsaturated fatty acids (Kyriakidis and Skarkalis 2000). The strong band at 525 nm corresponds to the vitamin E content. The medium intensity band at 681 nm corresponds to the chlorophyll band which most of the time is absent in the rest of the seed oils.

The aim of this part of this study is to focus on the use of fluorescence spectroscopy for the determination of different olive oil adulterants. The aim of the study was to determine the potential of fluorescence spectroscopy and the application of chemometric models including partial least squares and artificial neural networks as a tool for the assessment of olive oil adulterants of Maltese virgin olive oil.

13.1 Excitation and emission fluorescence spectra of adulterants and olive oil.

Three dimensional excitation and emission fluorescence spectra showed variations in both the excitation and emission wavelengths and their corresponding intensity. In general, with the exception of linseed oil, as the concentration of the adulterating seed oil increased there was a shifting of the EEM fluorescence towards 350 nm excitation and 450 nm emission coupled with an increase in the intensity Figure 1 illustrates the changes observed from 100% seed oil (left side) to 25% seed oil. In the majority of the spectra the decrease in the intensity observed within the 350 nm (Ex) and 450 nm (Em) was coupled with an increase in the intensity of bands appearing at an excitation of 330 to 440 nm and an emission of 660 to 700 nm. This peak is attributed to the presence of chlorophyll pigment and their degraded analogue pheophytins, present predominantly in olive oil and to some extent also in cold pressed linseed oil (Gliszczyńska-Świgło *et al*., 2007 and Herchi *et al*., 2012).

Both the seed oil adulterants and olive oil samples studied displayed a strong characteristic band with excitation at 300 to 360 nm and emission at about 350 to 400 nm. This band has been attributed to tocopherols and tocotrienols (Sikorska *et al*., 2004). The maxima of tocopherol emission vary slightly from one oil to another due to differences in the tocopherol composition. The intensity in the absorbance reflects the amount of tocopherols and tocotrienols present in the oil, with the seed adulterating oils showed a much higher concentration of tocopherols, as displayed by the increase in the intensity within this region when compared to seed:olive oil mixtures. Since tocopherols and tocotrienols are a vast class of compounds their spectral characteristics vary. Figures 13.1 and Figure 13.2 show that there are only small changes in excitation and emission maxima, which was attributed to the different extraction procedures to which the seeds were subjected. However, the EEM spectrum of linseed oil differs greatly from the rest of the seed oils; in fact, apart from the emission peak corresponding to the tocopherols and tocotrienols a stronger band was observed at a longer wavelength of 520 nm, which decreased in intensity as the % content of olive oil increased, whilst the peak corresponding to the tocopherols and tocotrienols increased. This is probably due to the presence of fluorophores present in linseed but not in olive oil. These compounds are most probably omega-3,6,9 fatty acids, present in very large amounts in cold pressed linseed oil but that are found at much lower concertation in olive oil (Sauci *et al*., 1994).

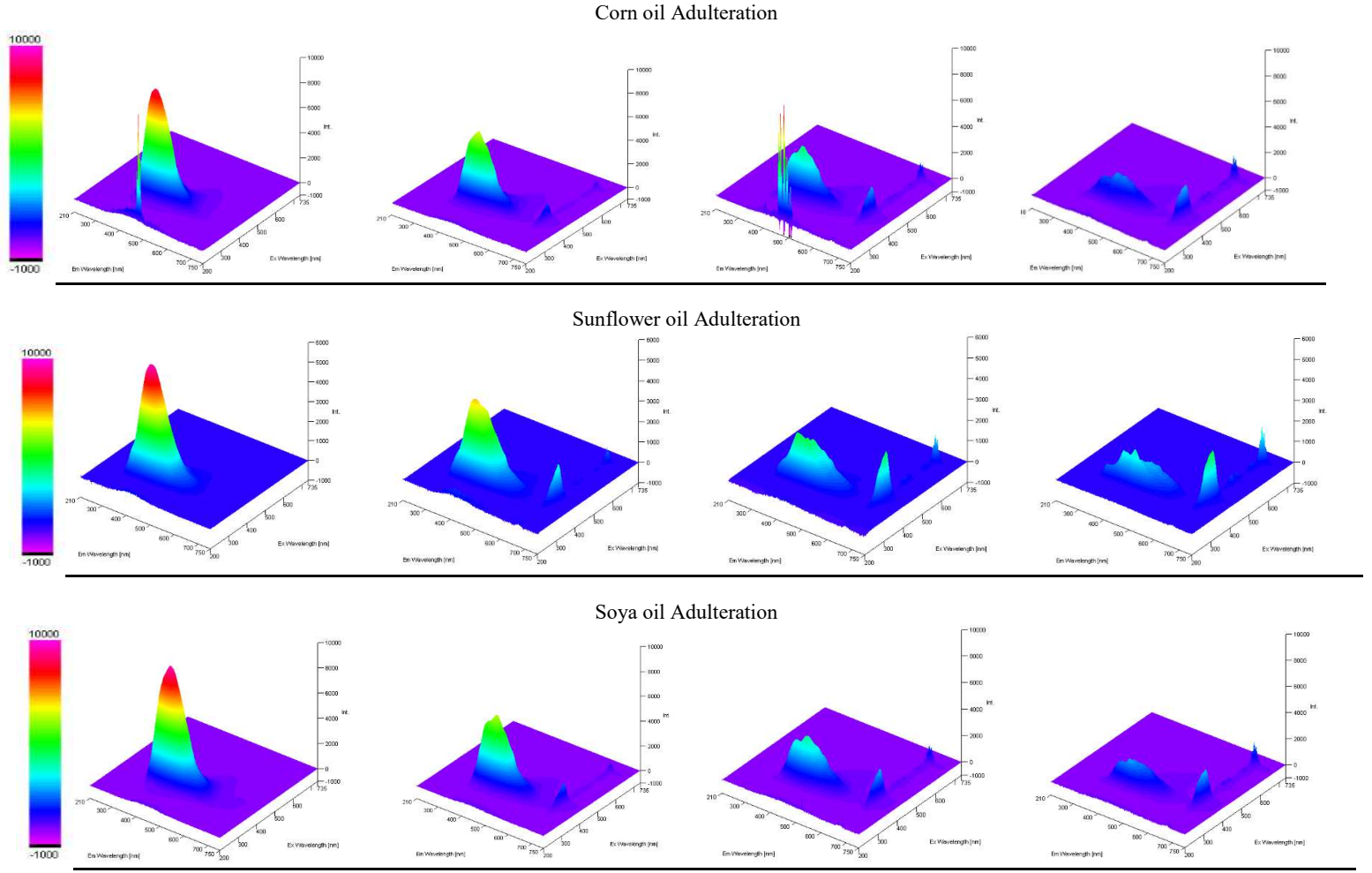


Figure 13.1: 3D EEM's between 210 to 750 nm excitation (axis z) and 210 to 750 nm emission (x axis) against intensity (y axis) for different levels of olive oil adulteration concentrations, 100, 75, 50 and 25% (left to right), for corn (1st row), sunflower (2nd row) and soya bean oil (3rd row).

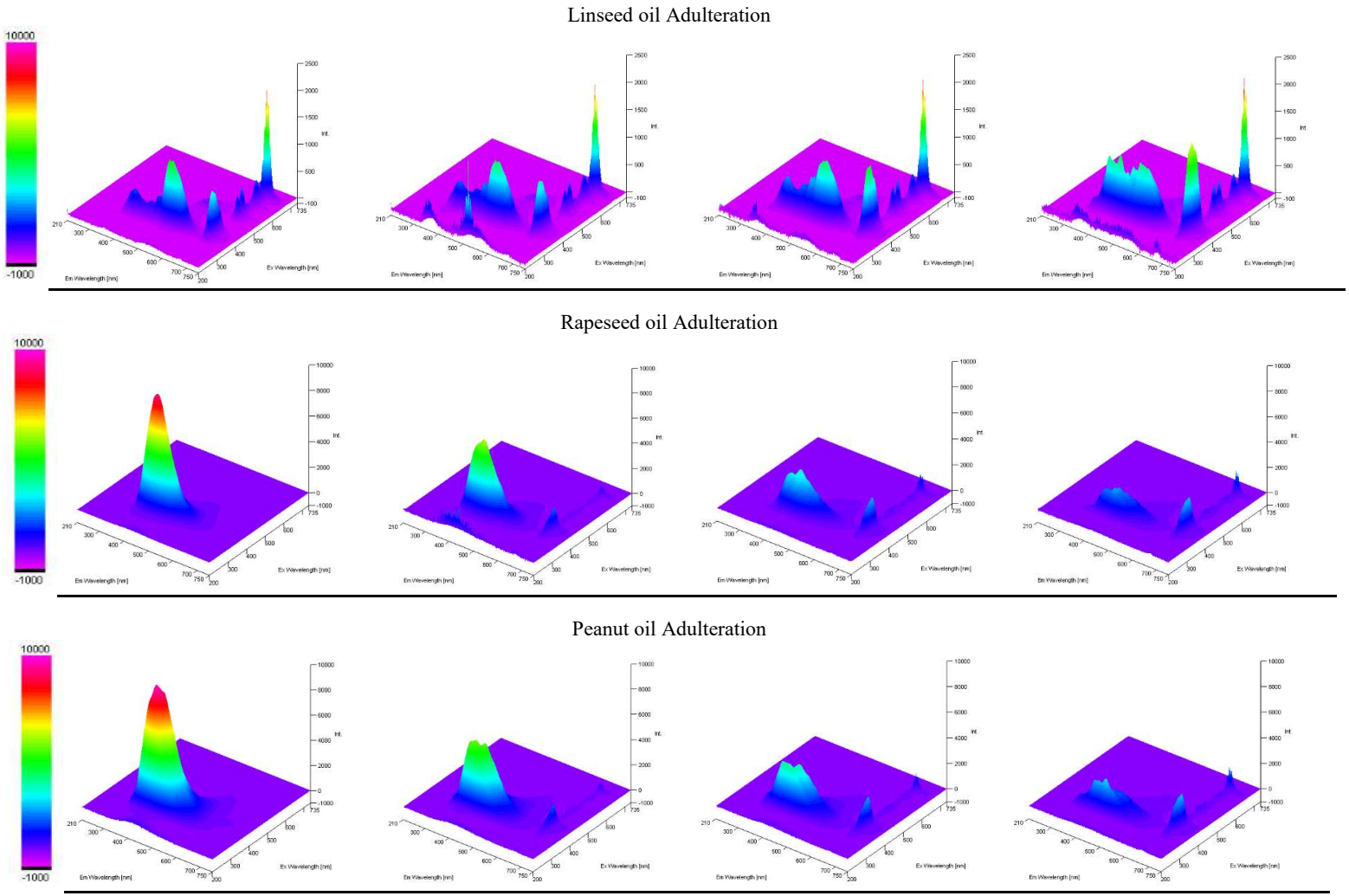
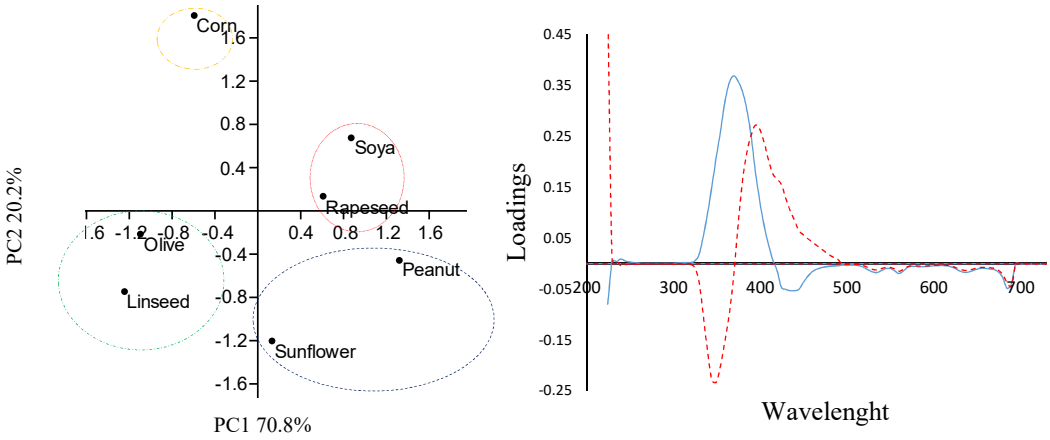


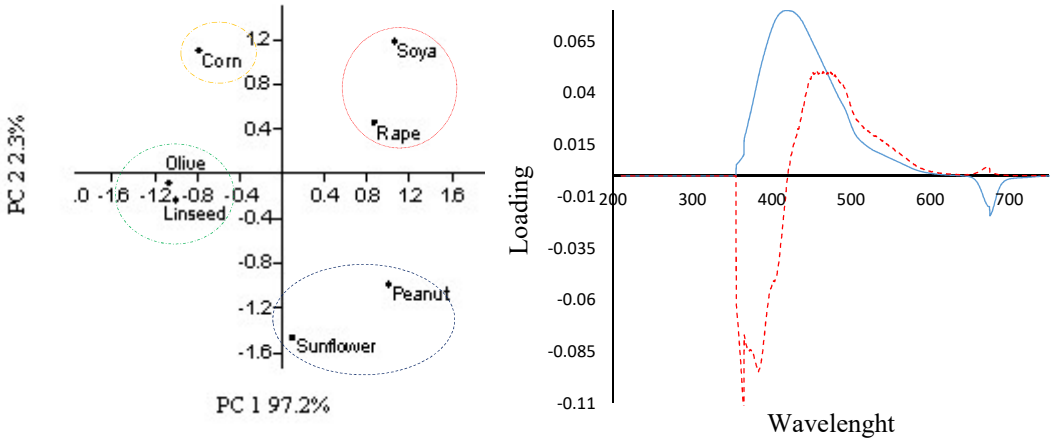
Figure 13.2: 3D EEM's between 210 to 750 nm excitation (axis z) and 210 to 750 nm emission (x axis) against intensity (y axis) for different levels of olive oil adulteration concentrations, 100, 75, 50 and 25% (left to right), for linseed (1st row), rapeseed (2nd row) and peanut oil (3rd row).

13.2 Principal component analysis on the adulterated EVOO

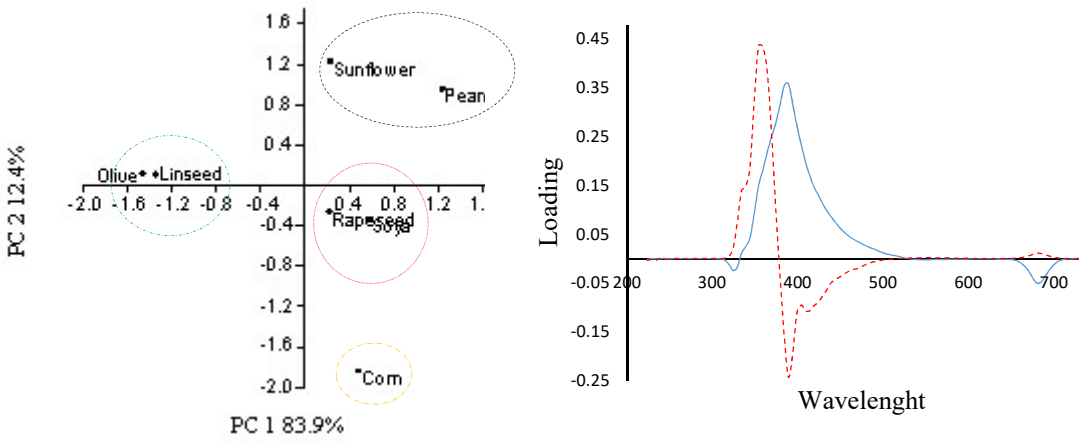
Through PCA, it was found that through the use of only two principal components explained 91.1 % of the variance of the data for the difference between the emission spectra, 99.5% for the excitation spectra and 96.3% for the synchronised spectra at 24 nm. The score plots obtained for each analysis classified the oils in four distinct areas; olive oil and linseed oil separated from the other seed oils (Figure 13.3). This separation is due to the negative scores obtained in PC1, attributed to its high content of compounds emitting at 420 to 485 nm and 600 to 700 nm, together with compounds excited at 680 to 700 nm. Apart from chlorophyll pigments, the presence of a different fatty acid profile will also contribute to variation in the emission and excitation spectra (Maggio *et al*., 2009; Matthäus and Özcan, 2011). Virgin olive oil is rich in oleic acid (55-83%), which is monounsaturated, while corn, soybean and sunflower oils predominantly contain polyunsaturated fatty acids. The clustering of olive oil and linseed oil is attributed to the higher levels of other fluorophores such as tocopherols, β-carotene and phenolic compounds that are refined out of other oils. Under both the excitation and emission PCA, peanut and sunflower oil clustered together whilst soya and rapeseed oil formed a separate cluster, due to similar classes of tocopherols and fatty acids (Kamal-Eldin and Andersson 1997). The pair combination of these two types of oils is due to an emission wavelength maxima centred at 350 nm and excitation wavelength maxima at 420 nm for soya and rapeseed oil, while peanut and sunflower oil showed an emission maxima at a lower wavelength of 330 nm and an excitation centred at 410 nm. This hypsochromic shift is attributed to the different extraction procedures and refining process (Sikorska *et al*., 2004).



(A) Principal Component Analysis Using Emission Spectral Data



(B) Principal Component Analysis Using Excitation Spectral Data



(C) Principal Component Analysis using Synchronised spectral data at 24 nm

Figure 13.3: PCA scores plots (left column) for the discrimination of EVOO and vegetable oil adulterants based on emission spectral data (A), excitation spectral data (B) and synchronised spectral data at 24 nm (C), and their corresponding loading plots (right column). The blue solid lines represent loading scores for PC1 while the red dotted line represents loading scores for PC2 for the different wavelengths (x-axis).

13.3 Synchronous spectra and partial least squares analysis (PLSR)

PLSR was performed on the oil-adulterant mixture samples synchronized spectra. This method models both the dependent (absorbance value at each wavelength) and independent (concentration of olive oil present within the mixture) variables simultaneously to find the latent variables (wavelengths) that will best predict the concentration of olive oil in the mixture. The optimum number of factors calculated using the leave-one-out cross-validation varied depending on the seed oil adulterant. VIP score values for each wavelength are a measure of a variable's importance in modelling both wavelength absorbance and % concentration present in olive oil (Figure 13.4). A value of 0.8 is generally considered to be a small VIP (Eriksson *et al*, 2006) and a red line is drawn on the plot at 0.8. From the variable importance plots one can see that not every wavelength in the spectrum has a VIP > 0.8, however, there was a small region in all spectra which did not vary by changing to a different adulterant. This region was found between 360 to 500 nm and another one at 579 – 664 nm; these regions correspond to the tocopherols/tocotrienols band and the chlorophylls / pheophytins respectively. The predicted values obtained by PLSR (Figure 13.5) were compared to the experimental values. The performance of the prediction model varied depending on the choice of the adulterant, the root means square error ranging from 7.2 for peanut oil adulteration to 0.7 for soya bean oil (Table 2). Although the RMSE was acceptable, the PLSR was repeated using only the wavelengths which had a VIP score > 0.8. The results showed a great improvement on the predicted model and with the exception of peanut oil adulteration, the RMSE was decreased by more than 28% for the remaining oil adulterants.

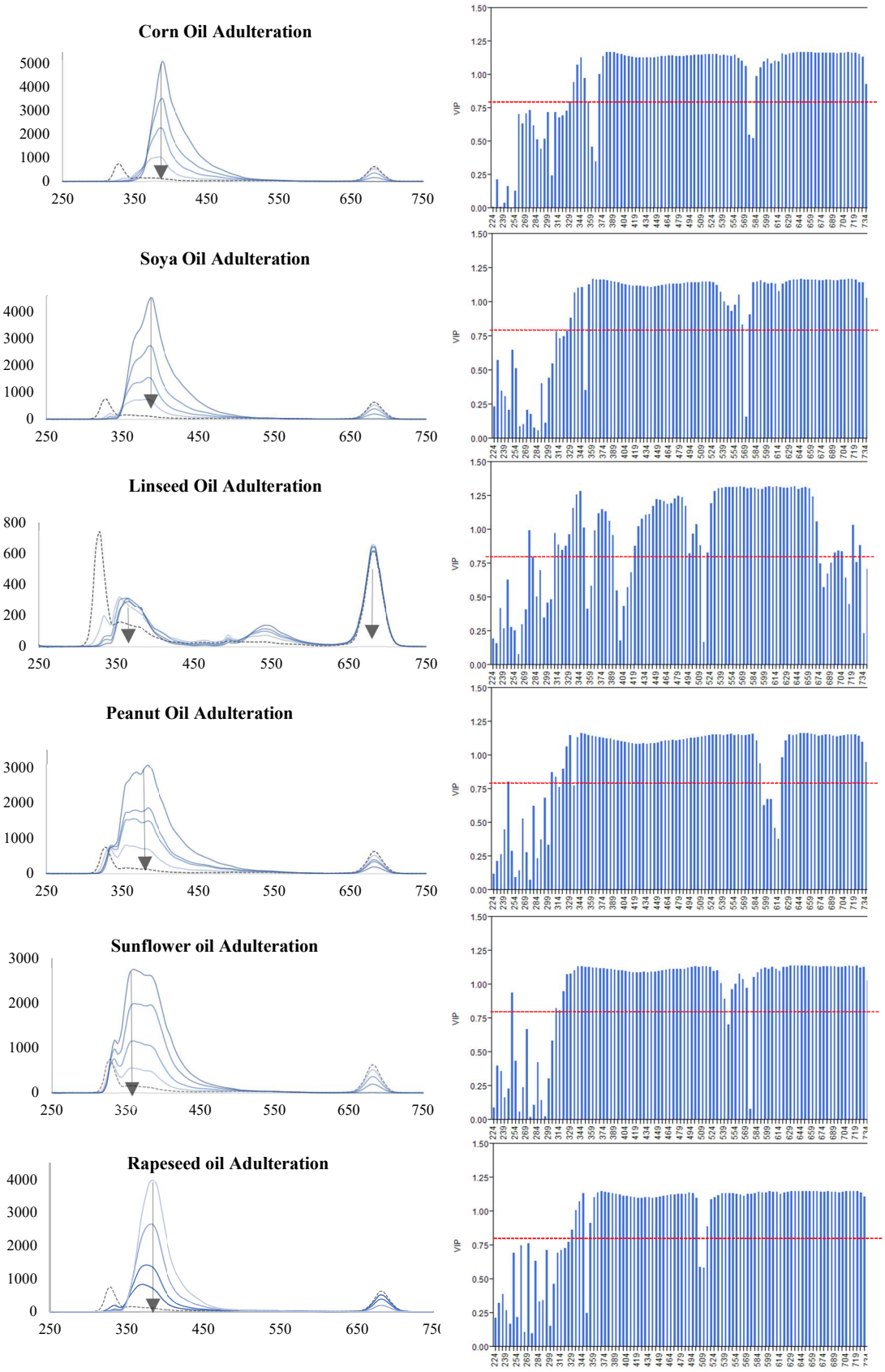


Figure 13.4: Synchronised spectra (left column) obtained at 24 nm with increasing concentration of seed adultertaion (solid blue lines) and olive oil (black dotted line). Arrows indicate the maxima obatined for the different seed oils. The variable importance plots (right column) obtained on using PLSR, where the wavelenghts (x-axis) and their corresponding VIP (y-axis) are plotted. The red dotted line indictes the VIP at 0.8.

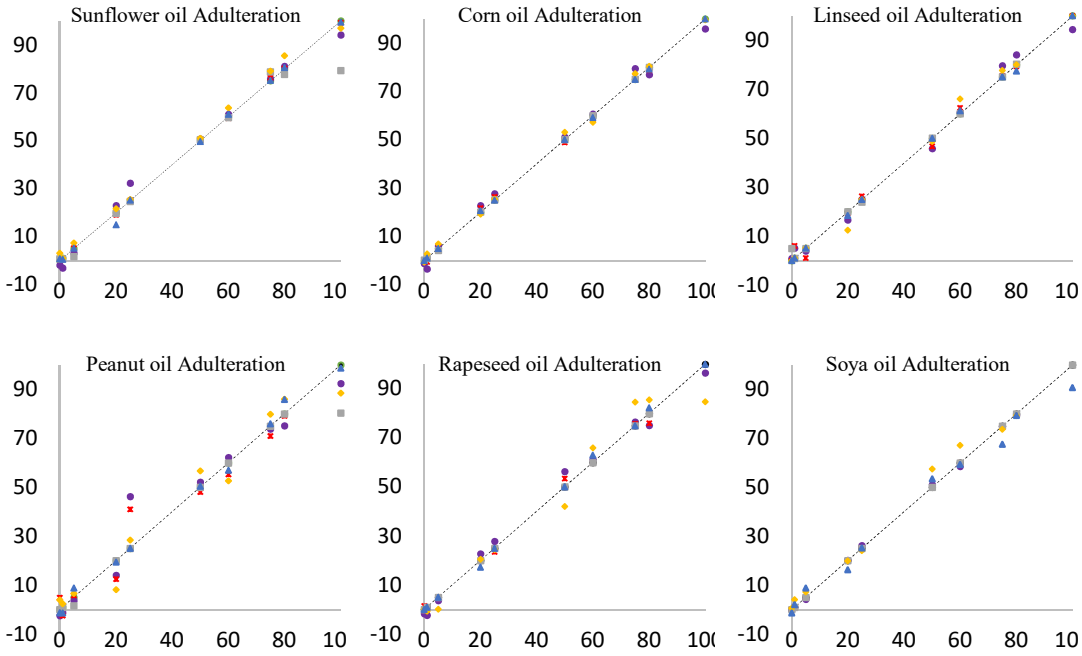


Figure 13.5: Predicted values (y-axis % adulterant) obtained from PLSR (purple circles), adjusted PLSR (red stars), PNN using k-fold (grey boxes), hold back (yellow diamonds) and excluded row cross-validation (blue triangles) against the experimental values (x-axis) for the different adulterating oils.

Table 13.1: Root means square error, root mean standard deviation, $R^{2,}$ and bias error obtained using PLSR and adjusted PLSR whereby the model was re-evaluated using only the wavelengths which had a VIP > 0.8.

	RMSE		RMSD		R2		MBE	
Oil Adulterant	PLSR	Adjusted PLS	PLSR	Adjusted PLS	PLSR	Adjusted PLS	PLSR	Adjusted PLS
Sunflower	3.490	0.646	1.028	0.587	0.995	1.000	-0.400	-0.034
Linseed	3.411	2.453	0.625	2.340	0.995	0.997	0.364	0.417
Corn	2.928	1.257	0.623	1.661	0.996	0.999	-0.404	-0.127
Soya	0.773	0.002	1.509	0.002	1.000	1.000	-0.021	0.000
Peanut	7.665	6.274	2.207	4.732	0.975	0.983	-0.200	-0.327
Rapeseed	3.349	1.850	1.292	0.108	0.995	0.999	-0.329	-0.105

13.4 Predictive artificial neural networks and prediction model analysis

Similar to the PLSR, PNN process large amounts of data in order to obtain a model with lowest error. PNN starts by input signals; in this experiment, the wavelengths were designed to receive various absorbance values from the whole or part of the spectrum. The network processed the data in order to give an output signal which corresponded to the % concentration of olive oil in the adulterated mixture. In this experiment, three different kinds of cross-validation in the neural network. It was found that on using wavelengths which had a VIP > 0.8 as previously determined by the PLSR, the model reached its optimal performance on using the excluded row validation (Table 13.1.). This confirms that the concentrations chosen by the experimenter for cross-validation covered a good concentration range for modelling and testing. For the majority of the seed adulterants, the excluded row cross-validation method showed a lower RMSE, RMSD, MBE and higher R^2 when compared to the other cross-validation methods. On using the whole synchronized spectrum, the model reached optimum prediction on using cross-validated using *k*-fold or CV-10 rather than on using the excluded row holdback (Table 13.2, Table 13.3 and Figure 13.6). This indicates that whilst PLSR improved on using variables which could explain the maximum variation, in the case of PNN the validation method was more dependent on the type of data.

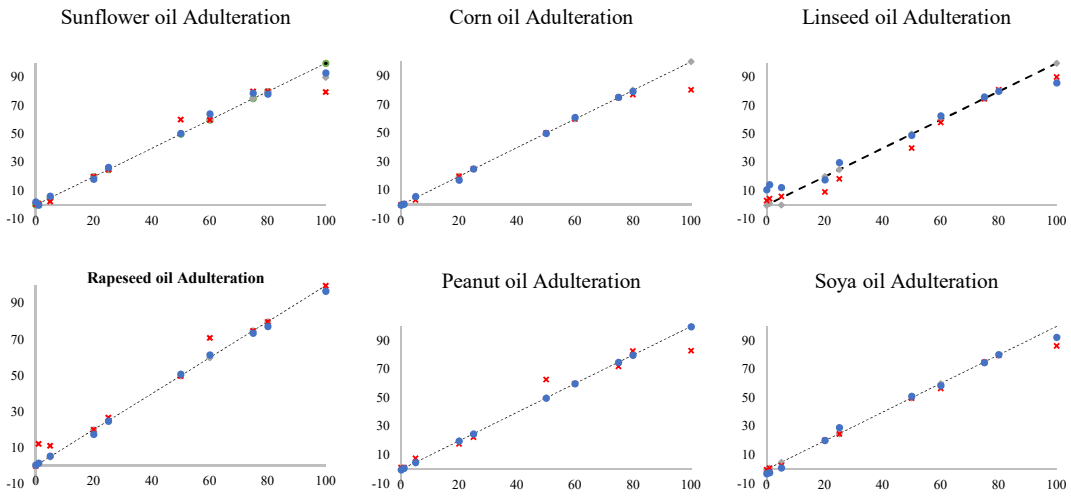


Figure 13.6: Overlaid predicted values (y-axis % adulterant) obtained from using k-fold (grey diamonds), holdback (red stars) and excluded row cross-validation (blue circles) against the experimental values (x-axis) for the different adulterating oils.

Table 13.2: The Root means square error, root mean standard deviation and an R^2 value obtained using the artificial neural network, constructed on the synchronized spectral wavelengths which had a VIP > 0.8. The model was cross-validated using different methods *k*-fold or CV-10 (*k*), Holdback at 0.33% (HB) and excluded row (ER).

	RMSE			RMSD			R^2			MBE		
Oil Adulterant	*k*	HB	ER	*k*	HB	ER	*k*	HB	ER	*k*	HB	ER
Sunflower	6.803	2.940	1.694	1.004	3.727	0.356	0.984	0.998	0.999	-0.111	0.061	-0.075
Linseed	1.609	3.231	1.031	1.231	6.043	0.037	0.999	0.996	1.000	-0.015	-0.027	-0.009
Corn	1.254	1.780	0.420	0.810	2.804	0.003	1.000	0.999	1.000	-0.027	0.217	0.001
Soya	3.018	3.546	1.330	0.704	7.177	0.007	1.000	0.996	0.995	-0.004	0.379	0.146
Peanut	6.314	6.808	2.618	3.014	7.411	0.000	0.987	0.980	0.997	-0.088	0.132	-0.171
Rapeseed	0.001	6.928	1.458	2.967	5.949	0.000	1.000	0.980	0.999	0.000	-0.265	-0.006

Table 13.3: The Root means square error, root mean standard deviation and an R^2 value obtained using the artificial neural network, constructed on the whole synchronized spectrum at 24 nm. The model was cross-validated using different methods *k*-fold or CV-10 (*k*), Holdback at 0.33% (HB) and excluded row (ER).

	RMSE			RMSD			R^2			MBE		
Oil Adulterant	*k*	HB	ER	*k*	HB	ER	*k*	HB	ER	*k*	HB	ER
Sunflower	3.202	7.414	3.110	0.066	0.000	4.230	0.997	0.977	0.996	-0.083	-0.098	-0.080
Linseed	1.655	6.199	7.597	0.017	1.948	2.797	0.999	0.988	0.989	-0.106	0.241	1.468
Corn	0.090	6.351	1.023	0.007	0.005	1.126	1.000	0.989	1.000	-0.026	-0.096	-0.059
Soya	0.001	4.478	3.321	0.000	3.332	1.200	1.000	0.995	0.996	0.000	-0.061	-0.372
Peanut	0.097	11.449	21.201	0.018	5.149	1.195	1.000	0.944	0.789	-0.007	0.087	0.001
Rapeseed	0.862	5.409	1.658	0.000	1.123	1.624	1.000	0.992	0.999	0.036	1.275	0.062

In conclusion, it was showed that spectrofluorimetry can be applied to distinguish among EVOO and vegetable oil blends, coupled with PCA models using the emission, excitation and the synchronized spectra at 24 nm. Fluorescence spectra can be measured quickly and easily, without the need for any pre-treatment of the oil sample, and could possibly be used to determine the level of adulteration of virgin olive oil. The results also show that adjusted PLSR models based on synchronized spectra for detecting the % amount of EVOO in vegetable oil blends had a lower RMSE (0.02–6.27 %) and higher R^2 (0.983–1.000) than those observed on using PLSR on the whole spectrum. This study also showed that PNN provides an alternative chemometric tool for the detection of olive oil adulteration. The performance of the model generated by the PNN is highly dependent both on the type of data input and the mode of cross-validation; for spectral data which had a VIP > 0.8, the excluded row cross-validation was more appropriate while for complete spectral analysis *k*-fold or CV-10 was more appropriate.

14 Thermal Degradation of Maltese extra virgin olive oil, an assessment of stability.

Stability and quality assurance issues of EVOOs are important issues concerning both the producers and sellers. The oxidation of fats and oils is one of the most fundamental reactions in lipid chemistry and itself is the main cause of quality deterioration, determining the shelf life of these products. Unlike, other refined seed oils virgin olive oils, are marketed and ultimately consumed without any chemical treatment such as bleaching. Compared to other seed oils EVOOs tend to have a high oxidative resistance mainly attributed to the high monounsaturated-to-polyunsaturated fatty acid ratio and to the presence of minor polar compounds having powerful antioxidant activity among which polyphenols and tocopherols (Kiritsakis *et al*., 1990). The refining process under which the majority of the marketed seed oils are subjected completely eliminates or drastically reduces the presence of the phenolic compounds, consequently, reducing the oxidative stability of refined oils when compared virgin oils. However, even though virgin olive oil is generally considered as being highly resistant to oxidation, the presence of some minor compounds/ions such as free fatty acids and photosensitizers including pigments and metal ions, tend to act as pro-oxidants. It is the variability of these minor prooxidants that causes the inherited variability in the stability of different virgin olive oils.

In spite of its high stability, virgin olive oil is also susceptible to different forms of oxidative processes. Enzymatic oxidation occurs when the oil is in the drupe and during the extraction process, whereby the presence of enzymes namely lipases, lipoxygenase and polyphenol oxidase catalyse oxidation reaction responsible for the formation of free fatty acids, hydroperoxides, and o-diphenols (further oxidised to o-quinones) respectively. Photo-oxidation occurs when the oil is exposed to light, unlike other vegetable oils EVOOs is consumed containing high amounts of chlorophyll pigments (Psomiadou *et al*., 2001). This is well documented that chlorophyll pigments act as photosensitizers due to the ability to transfer energy from light to triplet oxygen, producing thus singlet oxygen, which then reacts with the unsaturated fatty acids catalysing the hydroperoxide formation. The presence of these pigments makes the EVOO notably vulnerable to light exposure on the shelf reducing the overall quality of the product. The last form of oxidation which EVOOs undergo is known as

autoxidation which mainly occurs during processing and storage when the oil is in contact with oxygen (Frankel, 1985). Autoxidation is the main cause of deterioration of virgin olive oil during its shelf life. It occurs via the interaction of triacylglycerol fatty acids with molecular oxygen in the presence of transition metals or light which yields hydroperoxides by a free radical mechanism. High temperatures accelerate the process of autoxidation. The reaction proceeds through the formation of hydroperoxides, the primary oxidation products, which are unstable and decomposes to produce a range of volatile and non-volatile products. Volatile compounds produced during the lipid autoxidation reaction are mainly aldehydes and ketones, responsible for the rancid defects EVOOs (Angerosa, 2000).

The purpose of this work was to study the changes undergone by three monocultivar EVOOs derived from the Maltese islands namely 'Bidni', 'Malti' and 'Bajda' when subjected to thermal oxidation conditions. Changes were assessed through the study of specific chemical markers of oxidation including free fatty acidity, hydroperoxides, and secondary oxidation compounds. A profiling based approach was used in order to assess the changes in the phenolic composition for each cultivar under study. A fingerprinting based approach in conjunction with chemometric methods was also employed to predict the extent of oxidation through the use spectroscopic data, including spectrofluorimetry, UV-VIS, FTIR, and NMR spectroscopy. Other parameters including the chlorophyll pigment content and colour analysis were also assessed. The combination of target based and fingerprinting based approach was also combined in order to improve the overall predictability of the extent of oxidation.

14.1 Colour changes during thermal oxidation.

Colour is a sensory property with a strong influence on food acceptance as it contributes decisively to the initial perception that one can acquire the condition, ripeness, degree of processing, and other characteristics of foods (Alos, *et al*. 2006). Virgin olive oil is a natural product whose colour depends exclusively on biological compounds such as the chlorophyll and carotenoid pigments, their identification and individual evaluation make it possible to relate oil colour with the content and type of these compounds present. EVOOs is a strong indicator of a quality, changes in the

colour composition can reflect defects that have occurred during blending, storage, crushing, and extraction or the refining process.

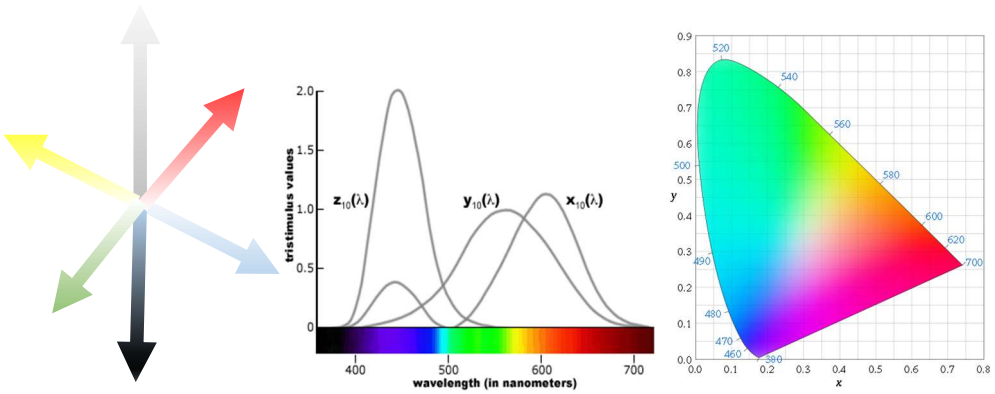


Figure 14.1: Colour systems employed.

The American Oil Chemists' Society (AOCS) proposed four official methods for the determination of colour for fats and oils. These include the Lovibond colour, Wesson colour, spectrophotometer colour and chlorophyll colour. In the present study, CIE L*a*b*, XYZ and xy were used as alternative colour models that might be used in objective oil colour evaluation. L*a*b* is an international standard for colour measurements, adopted by the Commission Internationale d'Eclairage (CIE) in 1976. Where the L* parameter represents the lightness component ranging from 0 to 100, the a* parameter refers to the colour ranges from green to red whilst b* parameter displays the colours from blue to yellow. Both the a* and b* chromatic components range from -120 to 120. The XYZ is an additive colour scheme also known as tristimulus colour space. It works by defining the amounts of three stimuli provided to the eye (l, m, and s). Closely related to the XYZ colour scheme is the xy colour scheme which takes account of the luminance-chrominance state which is mathematically derived from the XYZ colour scheme.

Colour analysis of the EVOOs through the 12-week thermal degradation experiment showed a significant decrease in the L* parameter indicating that in general, thermally oxidised EVOOs tend to lose their deep colour and attain a lighter transparent hue. Conversely, a significant increase in a* chromatic component was observed indicating a transition from the green colour towards a more red colouration, together with a significant decrease in the b* chromatic parameter indicating a

transition from a blue chromatic component towards a yellower colour. Application of the XYZ tristimulus colour space, a significantly moderate negative correlation was observed with respect to the X and Y chromatic component suggesting a decrease in the absorption between 500-700 nm On the other hand, a weak positive increase in absorption in the 400-450 nm range occurs as indicated by the Z chromatic component. In the case of the xy colour scheme, a moderate significant decrease in the two chromatic components was observed, further indicating the loss of the blue-green hues (y) and the red-orange hues (x), resulting in a transparent yellow colouration. From the results obtained it was shown that CIE L*a*b*, chromatic system provides a more robust correlation when compared to XYZ and xy chromatic systems. Morello´ *et al*., (2004) showed that luminosity values (L*) increased in oils after the 12 months storage period, probably as a consequence of the reduction in the pigment content. However, unlike what was observed in the presented study, Morello *et al*., (2004) showed that storage did not appear to have a significant effect on the chromatic ordinate b*, which corresponds to the yellow zone. This suggests that transition of blue to yellow hues is dependent on both the temperature and time rather than time alone. Ayton *et al*, (2012) showed that L* and a* were in general not significantly affected by the temperature at which the oil was stored. Although this observation is not concordant with the results obtained in this study, it important to mention that the maximum temperature to which Ayton *et al*, (2012) subjected the EVOOs was 36°C which was lower than the temperature to which EVOOs were subjected in this study. However, similar to the presented study b* value decreased significantly. This indicates the colour is slightly less yellow than in the initial oil.

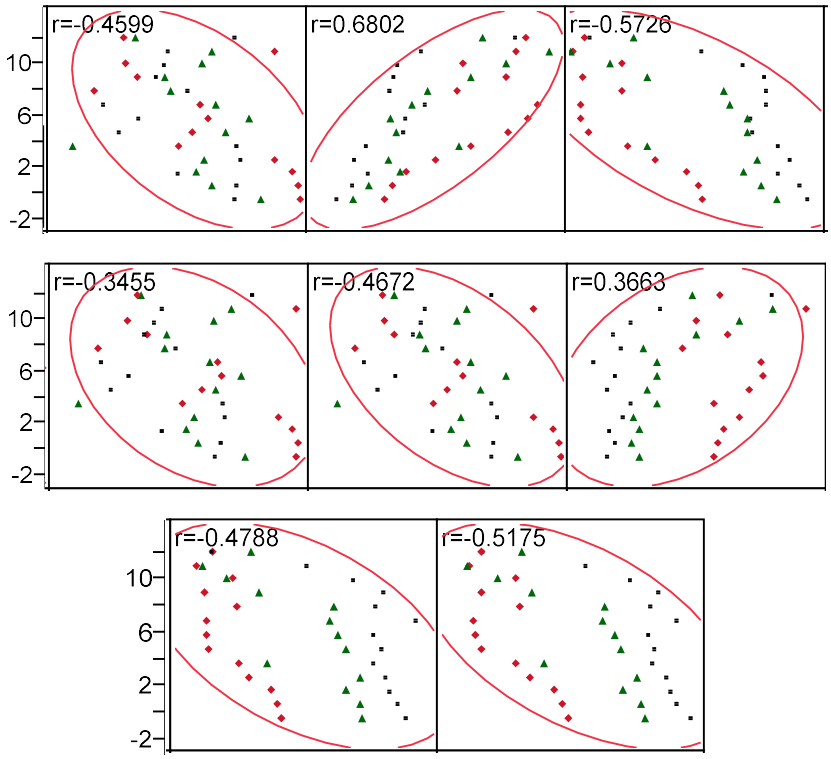


Figure 14.2: Correlation analysis of the different colour system parameters with the extent of thermal degradation. (Top) *L, *a, *b, (Middle) XYZ, (Bottom) xy.

Table 14.1: (Top) R^2 values (Bottom) gradient obtained from plotting the decrease in color against the extent of thermal degradation for the three different cultivars.

	R^2 Value							
	L	a	b	X	Y	Z	x	y
'Bidni'	-0.33	0.81	-0.77	-0.19	-0.34	0.58	-0.69	-0.74
'Malti'	-0.29	0.85	-0.80	-0.13	-0.30	0.78	-0.80	-0.82
'Bajda'	-0.70	0.76	-0.82	-0.64	-0.70	0.17	-0.80	-0.78
	Gradient							
'Bidni'	-0.27	1.17	-0.30	-0.10	-0.16	0.14	-118.04	-90.37
'Malti'	-0.23	0.95	-0.25	-0.07	-0.14	0.19	-129.99	-91.68
'Bajda'	-0.37	0.92	-0.37	-0.21	-0.20	0.06	-248.03	-146.43

Table 14.1 shows the individual cultivar colour correlation analysis and gradient (as a function of rate). It was observed that the extent of colour changes differ for each cultivar. In general, it was observed that 'Bajda' cultivar exhibited a stronger negative correlation with regards to the L* b*, X and Y chromatic components when compared to EVOOs derived from the 'Bidni' and 'Malti' cultivar. This observation indicates that the EVOOs derived from this leucocarpic cultivar tends to display a linear decrease in the luminosity and yellow parameter coupled with a decrease in absorption 500-700 nm as shown by the X and Y parameters. These results suggest that EVOOs derived from Leucocarpa cultivars behave differently than those derived from other Olea cultivars. It is well known that EVOOs derived from Leucocarpa cultivars tend to have a very low (even absent) concentration of chlorophyll pigments. The reduced concentration of chlorophyll pigments enabled almost a linear decrease in the aforementioned chromatic parameters. The reduced concentration of these compounds makes them more liable to thermal degradation. Furthermore, the rate through which changes in the aforementioned chromatic parameters occurs is faster in the EVOOs derived from Leucocarpa cultivars. Conversely, EVOOs derived from Leucocarpa cultivars had a lower correlation and slower rate in the Z parameter indicating that compounds absorbing the 400-450 nm tend to remain stable during the thermal oxidation process.

From the results obtained it was shown that oils stored at higher temperatures may display a slight change in colour, with the overall loss of luminosity and blueness coupled with an increase in the red hues. Although these changes are difficult to detect with the naked eye subjectively the application of CIE L*a*b*, the chromatic system provides a more direct and less subjective measure of colour. It was observed that the extent of colour change is dependent on the cultivar type, cultivars having a reduced amount of chlorophyll pigments tend to undergo a much more linear and rapid loss of colour. Due to the complex nature of chlorophyll and carotenoid compounds and the complex interaction that occur during the conditions under which they are stored, is further investigated using the entire UV-VIS spectrum.

14.2 Pigment changes during thermal oxidation.

The green-yellowish colour of VOO attributed to different pigments present namely chlorophylls, pheophytins and carotenoids (Cichelli and Pertesana, 2004; Del

Giovine and Fabietti, 2005; Boskou *et al*., 2006). The oxidative stability of EVOO is affected by the presence of these compounds and their derivatives. In the presence of light, chlorophylls and their derivatives are the most active promoters of photosensitized oxidation in virgin olive oil contributing greatly to their susceptibility to oxidation (Interesse *et al*., 1971). Since the experiment was carried out completely in the dark any changes in the pigment composition and the overall oxidation of the oil was solely attributed to thermal decomposition. It is well documented that thermal processing induces structural and chemical changes chlorophyll that often results in a colour change (Canjura *et al*., 1991 and Heaton *et al*., 1996) as shown in the previously. The loss of the green colour is mainly attributed to the conversion of chlorophylls to pheophytins. Under acidic conditions, magnesium in the chlorophyll ring is replaced by two hydrogen ions and green chlorophylls are converted to the olive-brown pheophytins (Mangos and Berger, 1997; Van Boekel 2000). It is possible that the formation of free fatty acids during the thermal degradation process lowers the pH enabling the formation of pheophytins. The second stage of degradation involves the formation of pheophytins derivatives known as pyropheophytins formed by the loss of the carbomethoxy group from pheophytins as a result of further heating (Schwartz and Von Elbe, 1983; Mangos and Berger, 1997).

Figure 14.2 shows the changes occurred during the thermal oxidation experiment to the chlorophyll pigments. Similar to other studies, it was found that chlorophyll is heat-labile and severely affected by the extent of heat treatment. Ayadi and Grati-Kamun (2009), Malheiro *et al*. (2009) and Jaber *et al*. (2012) reported more than 90% chlorophyll loss in olive oil samples. From the results obtained it was shown that for the three cultivars, chlorophyll degradation showed a three-stage decay process, a linear decrease over the first week followed by a stable phase were no significant changes in chlorophyll content was observed between six to ten weeks followed by a rapid decay. These results can be related to the study carried out by Gutiérrez *et al*.,(1991) whereby it was found out that the addition of chlorophyll to a virgin olive oil containing natural pheophytin A did not show any differences in oxidation levels or in Rancimat stabilities after a different storage period. These results suggest that once the compounds are present in sufficient amounts, the increase in their concentration is of minor importance, thus the levels of chlorophylls within the oil tend to stabilize as the degradation process continues. Furthermore, these photosensitizers

may also show slight antioxidant effects on the oils in the dark probably by donating hydrogen to break the free-radical chain reactions (Endo *et al*., 1984, Gutiérrez *et al*.,1991) preventing further loss of pigments.

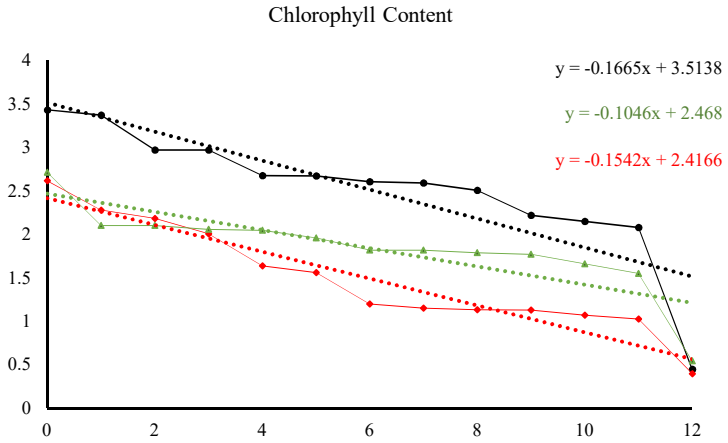


Figure 14.3: Changes in the chlorophyll content of the different cultivars (Black= 'Bidni', Green='Malti', and Red='Bajda')

From the results obtained it was shown that chlorophyll degradation is also cultivar dependent. From Figure 14.3 it was observed that whilst the initial amount of chlorophylls varied between the different cultivars, however, all the cultivars showed that by the end of the 12 weeks thermal degradation period an equivalent amount of chlorophyll pigment was present. Indicating that bleaching of chlorophyll pigments will occur independent of the cultivar. Furthermore, it was observed that the rate of chlorophyll degradation is also cultivar dependent, EVOOs derived from 'Bidni' cultivar had the highest rate of degradation followed by the 'Bajda' cultivar, whilst the EVOOs derived from the 'Malti' cultivar was more resistant to chlorophyll degradation. These results suggest that apart from the initial amount of chlorophyll there are other factors which are affecting the susceptibility, further investigation regarding the presence of antioxidant phenolic compounds will be discussed later on whilst the correlating the observed susceptibility to chlorophyll degradation to the presence of tocopherols is proposed as a further study.

The major carotenoids present in EVOOs are lutein and β-carotene, minor carotenoids including β-cryptoxanthin and epoxidized xanthophylls such as neoxanthin, violaxanthin, antheraxanthin and their furanoid isomers have also be determined by Gandul-Rojas and Minguez-Mosquera, (1996). Along with chlorophyll

compounds, carotenoids are responsible for the colour of the virgin olive oil. Changes within these compounds tend to occur during inadequate storage conditions with high temperatures and/or light. These compounds tend to undergo isomerization or discolouration that alter both the quality of the product and their functional properties. Aparicio-Ruiz, Minguez-Mosquera, and Gandul-Rojas (2011) show that thermal degradation of carotenoids in EVOO flows a first-order kinetic mechanism reactions of isomerization and subsequent degradation to colourless products.

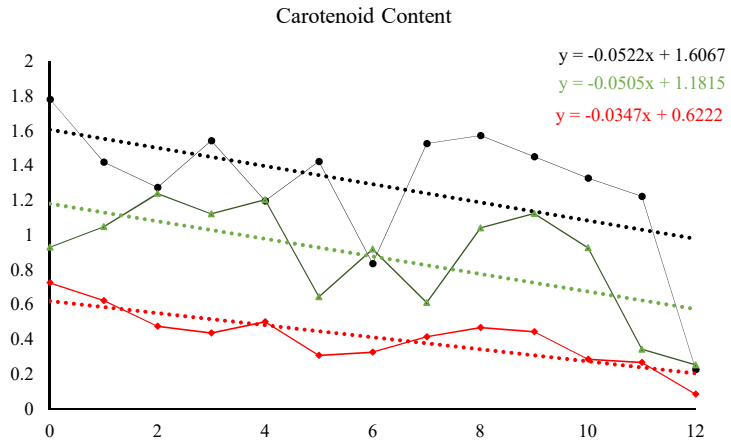


Figure 14.4: Changes in the carotenoid content of the different cultivars (Black= 'Bidni', Green='Malti', and Red='Bajda')

Figure 14.4 shows the concentration of carotenoids present in EVOOs during the 12-week degradation process. Unlike what was observed for the chlorophyll content a far less linear degradation was observed for the 'Bidni' and 'Malti' cultivar, this might be due to inferences due to secondary oxidation products that are formed during the oxidation results giving erroneous results. Nonetheless, an overall decay in the concentration of carotenoid compounds was observed. Comparison of the simplified rates of decay (gradient) of carotenoids to those of chlorophyll. Carotenoid compounds were found degrade at a much slower rate than chlorophyll pigments. It is well known that carotenoids are relatively stable during thermal processing when compared with chlorophylls (Kim *et al*., 2003). However, as the thermal insults are sustained a decrease in the carotenoid content was observed, this was attributed to oxidation and isomerization (Henry et al., 1998; Speek et al., 1988). The oxidation of β-carotene is accompanied by a loss in colour (Goldman *et al*., 1983; Henry *et al*., 1998; Anguelova and Warthesen 2000). Furthermore, the formation of oxidized lipids

may cause oxidation of carotenoids by their reaction with free radicals (Goldman *et al*., 1983).

Similar to the chlorophyll content, the degradation of carotenoid compounds seems also to be cultivar dependent. As expected, EVOOs derived from the 'Bidni' and the 'Malti' cultivars had the highest initial concentration of carotenoid compounds when compared to EVOOs derived from the 'Bajda' cultivar. However, it was shown that the rate at which carotenoid compounds decay is different. EVOOs containing a high amount of carotenoid compounds tend a faster rate of carotenoid decay whilst EVOOs containing a smaller amount of carotenoid compounds tend to lose their carotenoids at a slower rate. These observations further indicate that the rate of carotenoid decay follows a first-order decay kinetics, whereby for a doubling of the initial concentration of the reactant a doubling in the rate of reaction is observed. Further evidence, can be found by comparing the initial concentration of the carotenoids from the different EVOOs cultivars and the gradient of the reaction.

14.3 Colour differences and the effect of pigment degradation on colour.

The oxidation of olive oils leads to the inevitable degradation of natural pigments and thus discolouration (Morello *et al*., 2004). Oxidation products absorbing in the visible spectral range may also contribute to colour changes. Therefore, the olive oil colour may become a critical tool in quality evaluation (Caponio *et al*., 2005; Ceballos *et al*., 2003). The colour difference (ΔE) was calculated according to Sikorski *et al*., (2007) using the formula

$$\Delta E = [(L_0-L)2 +(a_0*-a*)2 +(b_0*-b*)2]1/2$$

L_0 , a_0 *, and b_0 * are the colour parameters of the fresh oil samples. Figure 12.5 shows a graphical representation of the colour changes occurring for each cultivar during the thermal degradation period. It was shown that EVOOs derived from the 'Malti' cultivar were lower than those exhibited by the EVOOs derived from the 'Bidni' and the 'Bajda'. Nonetheless, over 11 weeks the colour changes exhibited by EVOOs derived from the 'Bidni' cultivar were lower than those observed for the 'Malti' cultivar. In the case of EVOOs derived from the 'Bajda' an exponential change in colour was observed after eight weeks of degradation, whilst in the case of 'Bidni', this exponential change in colour was observed only after eleven weeks of degradation.

These results further corroborate the results which were previously obtained and further indicate that other constituents might be affecting the colour change and possibly the overall oxidation of the oil.

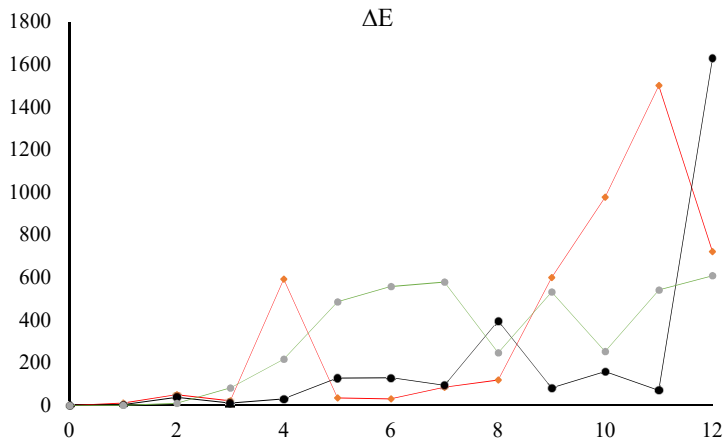


Figure 14.5: Changes in the colour difference of the different cultivars (Black= 'Bidni', Green='Malti' and Red='Bajda')

The relation between colour parameters and chromophore concentrations were assessed via the application of regression analysis. Figure 14.6 shows the correlation between colour parameters and concentrations of individual pigments in extra-virgin oil samples. The correlation coefficients for all of the tested pairs of parameters are also shown in Figure 14.6. Higher correlation coefficients were obtained for the correlations between the colour parameters and the chlorophyll concentration (1st Row), these results are comparable to those obtained by Sikorska *et al*., (2007). The best correlations were found for the a* and b* values with an R^2 value of 0.835, 0.840 (p-value < 0.05 for both). Conversely, a very low correlation was observed between the pigment concentration and luminosity (L*) of the EVOOs. These results, contradict the results obtained by the Sikorska *et al*., (2007), whereby it was shown that the pigment concentration is also highly correlated to the luminosity of the EVOOs. However, the study carried out by Sikorska *et al*., (2007) was a storage study without the application of heat treatment.

Slightly lower regression coefficients were obtained for the carotenoid concentration when compared to the a* and b* colour parameters. This was attributed to their larger stability and significantly lower degradation rates. The best correlation for carotenoids was obtained with the b* value, similar results were also observed by Sikorska *et al*., (2007). Although, lower regression coefficients were observed

between the carotenoid concentration and the L*, a* and b* colour system, conversely, higher regression coefficients were observed between the carotenoid content and the Z colour parameter of the XYZ colour system. The results obtained were as expected since the Z parameter is responsible for the absorption of compounds in the 400-450 nm range which coincides to the absorption of carotenoid compounds.

In the case of the xy colour system, it was observed that the both the chlorophyll and the carotenoid concentration were equally positively correlated to the xy parameters. These results, indicate that as the concentration of these compounds decreases as a result of thermal degradation, a resultant decrease in both the x and y parameter occurs. Since the majority of the fresh EVOOs, a yellow/ green colour a simultaneous decrease in both x and y colour parameters indicate that the colour becomes more centralised with respect to the CIE xy chromatic diagram thus becoming whiter (transparent) in colour.

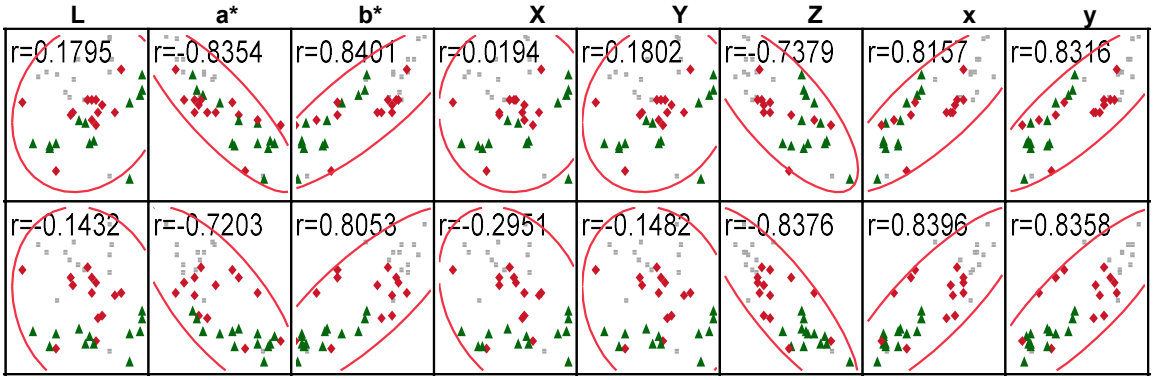


Figure 14.6: Correlation analysis between colour parameters and chromophore concentrations (1st row = chlorophyll content; 2nd row = carotenoid content) for the different cultivars (Black= 'Bidni', Green='Malti' and Red='Bajda') during thermal degradation

14.4 Conjugated Dienes and Trienes during thermal oxidation.

Fatty acids absorb light at particular wavelengths in the UV region and this may be used to determine olive oil quality. Refining causes a change in the configuration of fatty acids and the formation of conjugated dienes and trienes (Angerosa *et al*., 2006). Increased values of K_{232} and K_{268} in olive oil usually indicate the presence of refined oils. Furthermore, autoxidation reactions are also associated with conjugation, due to the formation of either carbon-carbon bonds or carbon-oxygen bonds which cause an increase of absorption in the region between 225 and 325 nm (Boskou, 1996).

Before the thermal stress experiment, the initial K_{232} and K_{270} values were determined and were all below the limits established by the EC Regulation for EVOO category (European Union Commission, 2007), which corresponded to 2.50 and 0.22, respectively. Only the EVOO-derived from the Leucocarpa variety had a higher K_{270} value than the one established by the EC. Similar to other published studies, the oxidative status of the samples were evaluated by monitoring the trend of conjugated dienes (K_{232}) and trienes (K_{270}) during the thermal stress (Hrncirik and Fritsche, 2005; Lerma-García, Simó-Alfonso, *et al*., 2009; Mancebo-Campos, Fregapane, and Desamparados Salvador, 2008). The similar to other studies it was shown that the three samples underwent a significant increase in both the K_{230} and K_{270} value with heating (Allouche et al., 2007; Bendini et al., 2009). However, unlike for the majority of the studies, in this experiment thermal stress was done in order to replicate real situations of mishandling, rather than focusing on the thermal stability of EVOO at cooking and frying temperatures. Figure 14.7 shows changes in the K_{230}, K_{270} and ΔK for the EVOOs derived from three different cultivars. It was shown that EVOOs derived from the 'Bajda' cultivar reached the legal limit (blue dotted line) for both the K_{230} and K_{270} faster than the other cultivars, indicating a lower thermal stability. During heating, a common trend for Δk among samples was not evidenced. K_{232} is related to the formation of hydroperoxides, carboxylic acids, conjugated dienes and conjugated trienes. These compounds are formed during the process of lipid oxidation (Allouche *et al*., 2007). Although in this experiment the EVOOs were not exposed to oxygen it was expected that K_{232} legal limits will be reached at a slower rate than those quoted by other experiments. Whilst K_{232} is a measure of the primary oxidation products in olive oil and thus the results obtained are highly correlated to the accumulation of peroxides in the oils. As primary oxidation products are produced they broke down into secondary oxidation products such as aldehydes and ketones which were measured at K_{270}. These results are very similar to those found by other authors (Krichener *et al*., 2010). The initial lag phase (the slower rate of increase) in the peroxide value, K_{232} and K_{268} values in the oils during thermal stress can be attributed to the antioxidant properties of the polyphenols present.

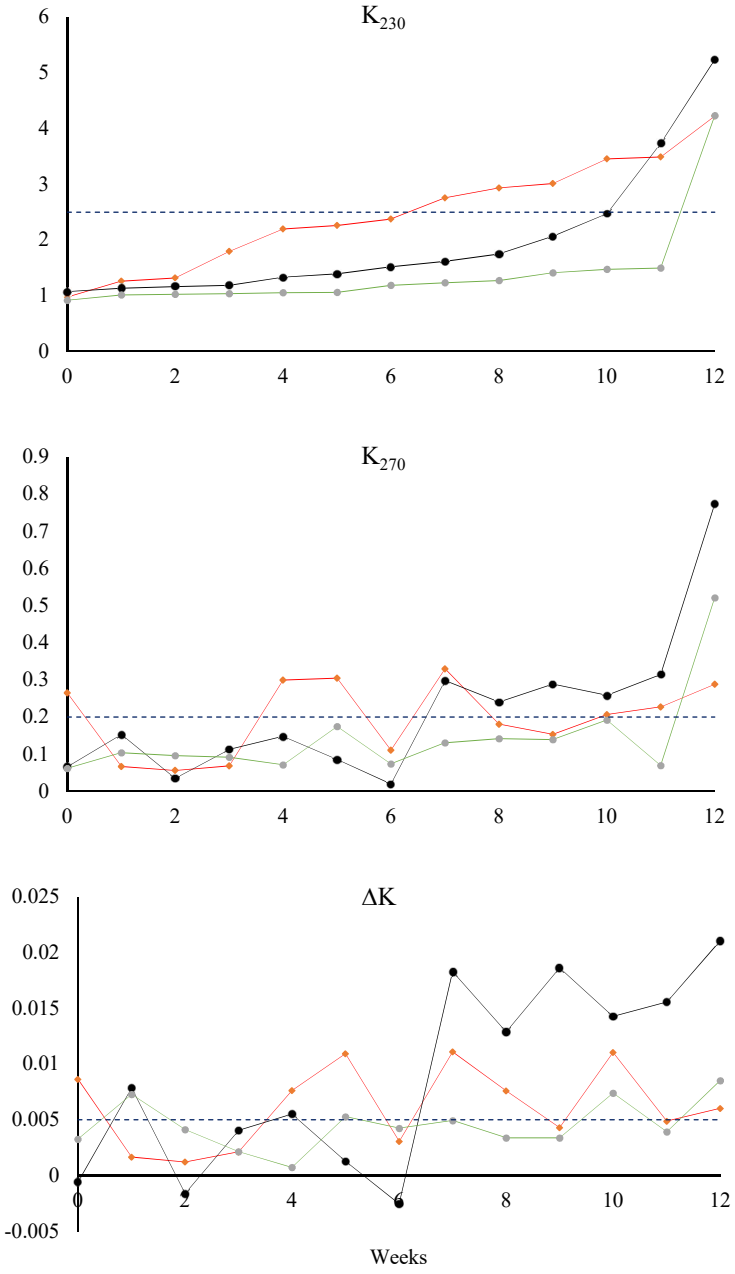


Figure 14.7: Changes in the diene and triene content (Top = K_{230}; Middle = K_{270}, Bottom = ΔK) of the different cultivars (Black= 'Bidni', Green='Malti' and Red='Bajda') during thermal degradation. The dotted line represents the legal threshold for the EVOO to be still considered as extra virgin.

14.4.1 Effects of Chlorophyll pigments on the concentration of conjugated dienes and trienes

Correlation analysis between the K values and the pigment content of the oil revealed an overall negative correlation. This indicates that these pigments displayed an overall antioxidant activity. Table 14.2 shows that stronger negative correlation values were obtained when the individual cultivars were compared, indicating that the antioxidant activity displayed is dependent on the initial concentration of these

pigments. It is well documented that β-Carotene shows antioxidant activity, protecting lipids from free radical auto-oxidation by reacting with peroxyl radicals, ultimately inhibiting propagation reactions and promoting termination of the oxidative chain reactions (Britton *et al*., 2007). Similarly, the effects of chlorophyll and pheophytin on the autoxidation of oils in the dark were investigated by Endo *et al*., (1984). It was shown that both chlorophyll and pheophytin show antioxidant activity retarding the oxidative deterioration of triglycerides in rapeseed and soybean oils at 30 °C. Francisca and Isabel, (1992) showed that in the absence of light, chlorophylls act as antioxidants via the donation of hydrogen radicals.

Table 14.2: Correlation analysis between the conjugated diene and triene content against chlorophyll and carotenoid concentration for EVOOs derived from different cultivars.

		K_{270}	K_{230}	dk
Overall	Ch	-0.69	-0.61	-0.45
	Cr	-0.42	-0.49	-0.13
'Bidni'	Ch	-0.98	-0.87	-0.83
	Cr	-0.74	-0.62	-0.44
'Bajda'	Ch	-0.88	-0.83	-0.39
	Cr	-0.73	-0.63	-0.65
'Malti'	Ch	-0.91	-0.43	-0.92
	Cr	-0.76	-0.45	-0.73

However, although in this experiment it was shown that these pigments display antioxidant activity, in the presence of light these compounds tend to display a prooxidant effects. Prooxidant activity can be defined as a component, such as a metal ion, able to lower the activation energy for the initiation of lipid oxidation (Labuza, 1971). In fact, the prooxidant activity of plant pigments, especially chlorophyll, pheophytin, and phaeophorbide, within oil is well documented in the literature (Shahidi & Wanasundara, 1992). This is attributed to the fact that chlorophylls and their degradation products pheophytins and pheophorbide, act as sensitizers to produce oxygen radicals in the presence of light and atmospheric oxygen. The presence of these compounds increases the rate of oil oxidation (Whang and Peng, 1988). In the presence of high concentrations of oxygen, carotenoids act as prooxidants via the generation of carotenoid radicals which react with atmospheric oxygen to generate carotenoid-peroxyl radicals. The latter compounds act as pro-oxidants by promoting oxidation of unsaturated lipids. Furthermore, Steensen and Min (2000) found that during auto-oxidation of soybean oil in the dark, thermal β-carotene degradation products act as

pro-oxidants, this could explain the sudden increase in the rate of degradation which was observed between 10-12 weeks.

14.5 Application of Chemometrics methods for the evaluation of the extent of thermal degradation

A number of spectroscopic techniques were evaluated to establish their potential in the determination of thermal degradation. These spectroscopic methods were coupled with chemometric methods mainly partial least squares regression (PLSR) and principal component regression (PCR). The aims of this research were to compare the performance of the different methods for the evaluation of the degree of thermal degradation and to develop PLSR and PCR prediction models based on spectrometry for the determination of the extent of thermal degradation. In order to obtain more information from these spectroscopic methods, the extracted spectral data were subjected to mathematical elaboration. In particular, normalization, baseline correction, Multiplicative Scatter Correction (MSC), Standard Normal Variate (SNV) Orthogonal Signal Correction (OSC) and Derivative elaborations namely Savitzky–Golay 1st degree derivatization, first and second derivatization were applied (Iñón *et al*., 2003; Maggio *et al*., 2009).

Both PLSR and PCR are methods which are used to model a response variable in conjunction with a large number of highly correlated or collinear predictors. PCR and PLSR construct components by means of linear combinations of the original predictor variables, the difference between them is that they construct components in different ways. In the case of PCR, the components are generated with the main aim of explaining the maximum observed variability in the predictors without considering the response variable at all. Whilst in the case of PLSR, the response variable is taken into account, and therefore often leads to models that are able to fit the response variable with fewer components. In this study, spectral data obtained from the different EVOO samples obtained at different time intervals obtained from the thermal degradation experiment were used to form the explanatory matrix (X) whilst the extent of degradation given in terms of weeks for which the samples were subjected to heat was used as the dependent matrix (Y). The development of PLSR and PCR prediction models involves two basic steps: training and test phases. In the training phase, 55 % of the data were selected to generate the model. In this phase, leave one out cross-

validation and CV-10 cross-validation was used respectively to choose the optimum number of PLSR and PCR components with the smallest prediction error, which avoids overfitting of the model. After training models, an independent dataset (45 % of the total data) was utilized to test the prediction performance. The residual of the difference between experimental value and prediction value was plotted and validation of the models was performed by comparing differences in root mean square error (RMSE), mean biased error (MBE), mean absolute error (MAE) and coefficient of determination (R^2), which were calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_e - x_p)^2}{N}}$$

$$MBE = \frac{\sum_{i=1}^{n} x_e - x_p}{N}$$

$$MAE = \frac{\sum_{i=1}^{n}|x_e - x_p|}{N}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(x_e - x_p)^2}{\sum_{i=1}^{N}(x_e - x_m)^2}$$

N is the total number of data, x_P represents the predicted value from the model, whereas x_E is the experimental value, x_M is the mean of experimental value. The PLSR was performed using the software JMP Pro (SAS Institute Inc., Cary, NC, USA) whilst PCR was performed using The Unscrambler 10 ® (CamoCamo Process As., Oslo, Norway). MAE and RMSE provide information about the magnitude of the average error but provide no information on the relative size of the average difference between the predicted and experimental values. On the other hand, MBE describes the direction of the error bias. Its value, however, is related to the magnitude of values under investigation. A negative MBE occurs when predictions are smaller in value than observations.

14.5.1 Chemometric analysis of UV-Vis changes during thermal degradation

Figure 14.8 shows the absorption spectra of an extra-virgin olive oil in n-hexane in the visible range obtained after various spectral pretreatment methods as a function of the thermal degradation period. The dotted line represents the R^2 derived from performing a matrix correlation analysis on the entire spectrum in correlation to the weeks of degradation. Typical UV-Vis absorption spectrum of EVOOs displays

two major maxima, the absorption in the 450-520 nm range is attributed to carotenoid pigments. The carotenoid band overlaps with that of the chlorophylls at 380-450 nm, while the characteristic band at 650-700 nm belongs exclusively to chlorophylls and pheophytins. As previously observed from the analysis of the individual pigment concentration, heating had a detrimental effect on a number of pigments present in EVOO. As expected the intensity of the chlorophyll-band decreased to less than half of its initial value after 5 weeks with this band eventually disappearing completely after the 12 weeks of thermal degradation. In contrast, carotenoids were significantly more stable, although their absorption was partially reduced during the thermal process. From the correlation analysis, it was evident that these aforementioned peaks were most negatively correlated with the time of thermal degradation. From the correlation analysis, it was shown that for the majority of the spectral pretreatments the carotenoid and chlorophyll peaks observed at 380-500 nm had the most negative correlation values when compared to 650-700 nm peak. These results provide the first indication that although stable the carotenoid peaks seem to provide a better correlation to the extent of degradation than chlorophyll pigments.

Application of PLSR models for the determination of the extent of thermal degradation showed that whilst during the training phase spectra derived using detrending and MSC function had the best performance with low RMSE, MBE, MAE and high R^2 value, these models failed during the validation step as there was an overall increase in the RMSE and bias. Spectra obtained using second-order derivatization seem to provide the most reliable prediction models as these had the lowest RMSE and bias associated errors during the external validation step. Analysis of the individual cultivars performance under PLSR analysis, it was shown that throughout all the spectral transformations the PLSR models obtained were able to model the thermal degradation of EVOOs derived from the ‘Malti’ cultivar with a higher degree of precision.

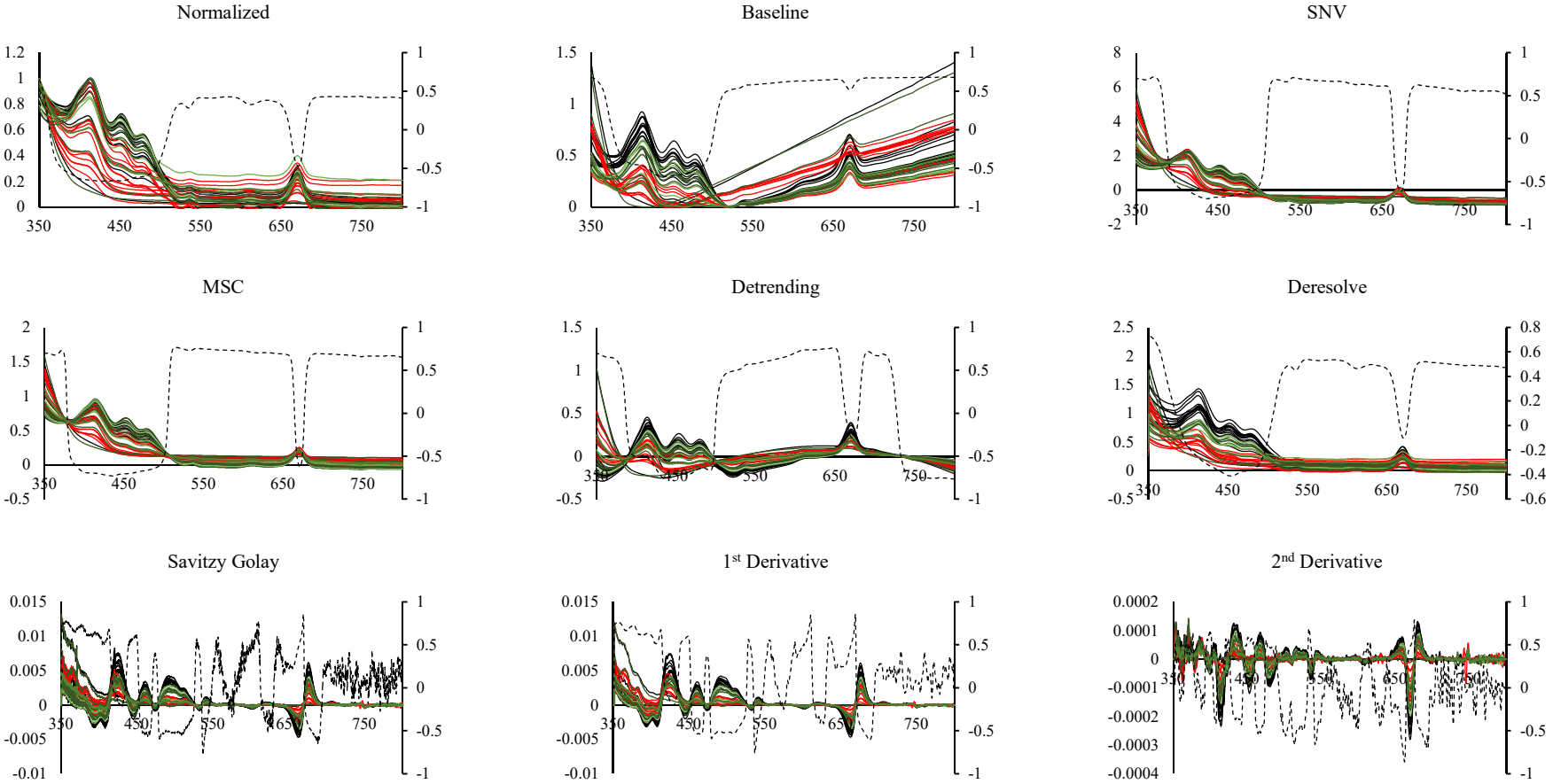


Figure 14.8: Application of different spectral pretreatments on UV-Vis spectra for EVOOs derived from different cultivars (Black= 'Bidni', Green='Malti', and Red='Bajda'). The dotted line represents R^2 value obtained from correlating the different λnm absorbances with the extent of thermal degradation.

In the case of PCR, the models obtained had a lower performance during the training phase when compared to those obtained using PLSR, as a higher RMSE and MAE were observed during PCR training validation models. However, during the testing phase, the PCR models obtained reached a comparable performance to those observed for the PLSR models. From the PCR models obtained it was found that spectra pretreated using the Savitzky Golay and second order derivatization had the optimal performance during the training phase, however, these spectra failed during the testing phase as a higher RMSE and MBE were observed. In the case of PCR models the optimal performance was obtained using the 1st order derived spectra as the errors obtained during the training phase were comparable to those obtained during the testing phase indicating that the model obtained during the training phase is also able to predict the extent of degradation for EVOOs not included in the training phase.

Table 14.3: Performance of the PCR (Top) and PLSR (Bottom) models obtained using UV-Vis data after different spectral pretreatments.

PCR UV-Vis Training											
	Raw	Base	Norm	DS	DT	SNV	MSC	OSC	SG	1D	2D
RMSE	2.096	2.407	2.096	2.145	1.301	1.471	1.105	2.139	1.009	1.206	1.005
MBE	-0.144	-0.207	-0.181	-0.151	-0.089	-0.107	-0.065	-0.151	-0.041	0.021	-0.043
MAE	0.355	0.431	0.366	0.387	0.211	0.315	0.229	0.377	0.169	0.252	0.175
R^2	0.681	0.580	0.681	0.666	0.877	0.843	0.911	0.668	0.926	0.894	0.927
PCR UV-Vis Predictability											
RMSE	2.280	2.235	2.415	2.476	2.820	4.136	1.349	2.370	3.434	0.983	3.546
MBE	0.104	-0.010	0.034	0.147	-0.224	0.166	-0.140	0.127	-0.173	-0.103	-0.180
MAE	0.241	0.305	0.299	0.237	0.296	0.424	0.219	0.249	0.233	0.157	0.237
R^2	0.651	0.653	0.554	0.620	0.769	0.061	0.860	0.636	0.816	0.933	0.822
PLSR UV-Vis Training											
	Raw	Base	Norm	DS	DT	SNV	MSC	OSC	SG	1D	2D
RMSE	2.096	0.426	0.359	0.501	0.137	0.487	0.275	0.794	0.713	0.863	0.078
MBE	-0.144	0.016	0.002	0.003	0.004	0.015	0.024	0.027	0.000	-0.023	0.005
MAE	0.355	0.072	0.055	0.073	0.033	0.114	0.059	0.170	0.147	0.162	0.020
R^2	0.681	0.987	0.991	0.982	0.999	0.983	0.995	0.954	0.963	0.946	1.000
PLSR UV-Vis Predictability											
RMSE	2.280	2.541	2.313	3.323	3.678	2.708	2.451	2.136	1.655	1.539	1.139
MBE	0.104	-0.114	-0.123	-0.118	-0.116	-0.135	-0.117	-0.064	-0.071	-0.064	-0.048
MAE	0.241	0.206	0.187	0.271	0.314	0.185	0.194	0.201	0.153	0.145	0.168
R^2	0.651	0.758	0.826	0.684	0.648	0.798	0.747	0.743	0.886	0.917	0.896

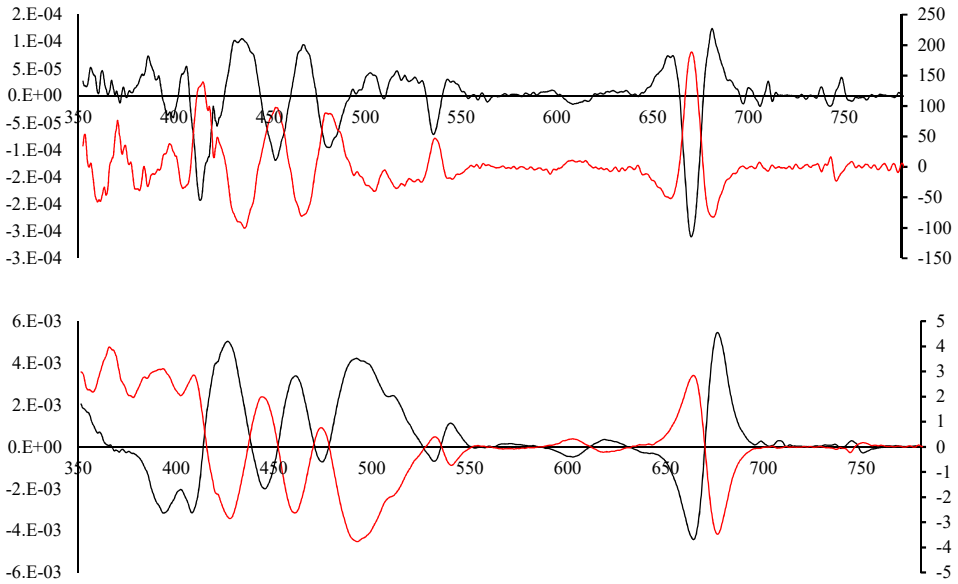


Figure 14.9: (Top) 1st regression coefficient (Red line) obtained from the PCR for the 1st order derived spectra (black line). (Bottom) 1st β-coefficients (Red line) obtained after PLSR for the 2nd order derived spectra (black line).

Detailed analysis of the models which had the highest performance was carried out. In the case of PCR, the models obtained using the 1st order derived spectra were analyzed whilst in the case of the PLSR models obtained using 2nd order derived spectra were analyzed. Figure 14.9 shows the 1st regression coefficient obtained from the PCR for the 1st order derived spectra which explained 76% of the total variation. Analysis of the β- regression coefficients obtained, it was found that the degradation model obtained is highly dependent on the variation observed in the 650-700 nm reaching a maximum at 670 nm, together with the 400-500 nm reaching a maximum at 416 nm, 450nnm and 480 nm. Similar observations were obtained for the analysis of the first β-coefficients obtained after PLSR for the 2nd order derived spectra. The peaks observed were identified as pertaining to β- carotene which shows two major maxima observed at 450 and 480 nm, the maxima observed at 450 nm could also be attributed to the presence of chlorophyll b which absorbs in the same spectral region of β-carotene. The peak observed at 416 nm was attributed to chlorophyll a which under typical UV-Vis conditions it is observed shouldering that of 450 nm as it would be masked by the presence of β-carotene, however, the application of 1st order derivative this peak becomes completely resolved. The peaks observed in the 650-700 nm region namely at the 670 nm was attributed to the presence of degraded chlorophyll

pigments namely pheophytins derived from their chlorophyll pigment counterparts after the loss of the central magnesium ion.

14.5.2 Chemometric analysis of FTIR changes during thermal degradation

In the oxidation experiment EVOOs derived from different cultivars was also assessed using ATR-FTIR spectra, their response to thermal stress was evaluated. Figure 14.10 shows the changes which were observed in the FTIR spectra using different spectral pretreatment methods. It was expected that under thermo-oxidative conditions changes within the infrared spectral absorption bands would occur as a direct result of lipid degradation. However the changes within the FTIR spectra were very subtle, furthermore, no shifts in the exact position of the bands were induced by heat treatment. Furthermore, the preliminary chemometric analysis carried out on the entire dataset proved but be unsuccessful in determining the extent of thermal degradation for the different EVOOs using FTIR spectra, as the performance of both the PLSR and PCR models obtained was very low.

This suggests that a universal model, which could predict the extent of thermal oxidation in EVOOs, could not be applied in the case of FTIR spectra. This was attributed to the very sensitivity of the FTIR spectra since each cultivar has a different chemical composition it cannot be assumed that the linear thermal degradation would be constant throughout, thus the application of a universal spectroscopic method coupled with advanced multivariate statistical analysis would fail. The application of chemometric models representing the thermal decay for each singular cultivar would be more appropriate and thus reduce the chance of model over-fitting. Table 14.4 shows the results obtained when each individual cultivar was modelled using both PLSR and PCR using FTIR data obtained after different spectral transformations. In general, it was observed that as the harsher pretreatment functions the lower was the performance of the models obtained. In fact, for both the PLSR and PCR models obtained spectra directly obtained using from the instrument after a basic ATR correction yielded the highest model performance for all the three cultivars studied, whilst spectra derived after the 2^{nd} order derivatization yielded models with the lowest performance. This indicates that in the FTIR spectra the unprocessed raw data could yield more information regarding the thermal oxidative state of EVOOs.

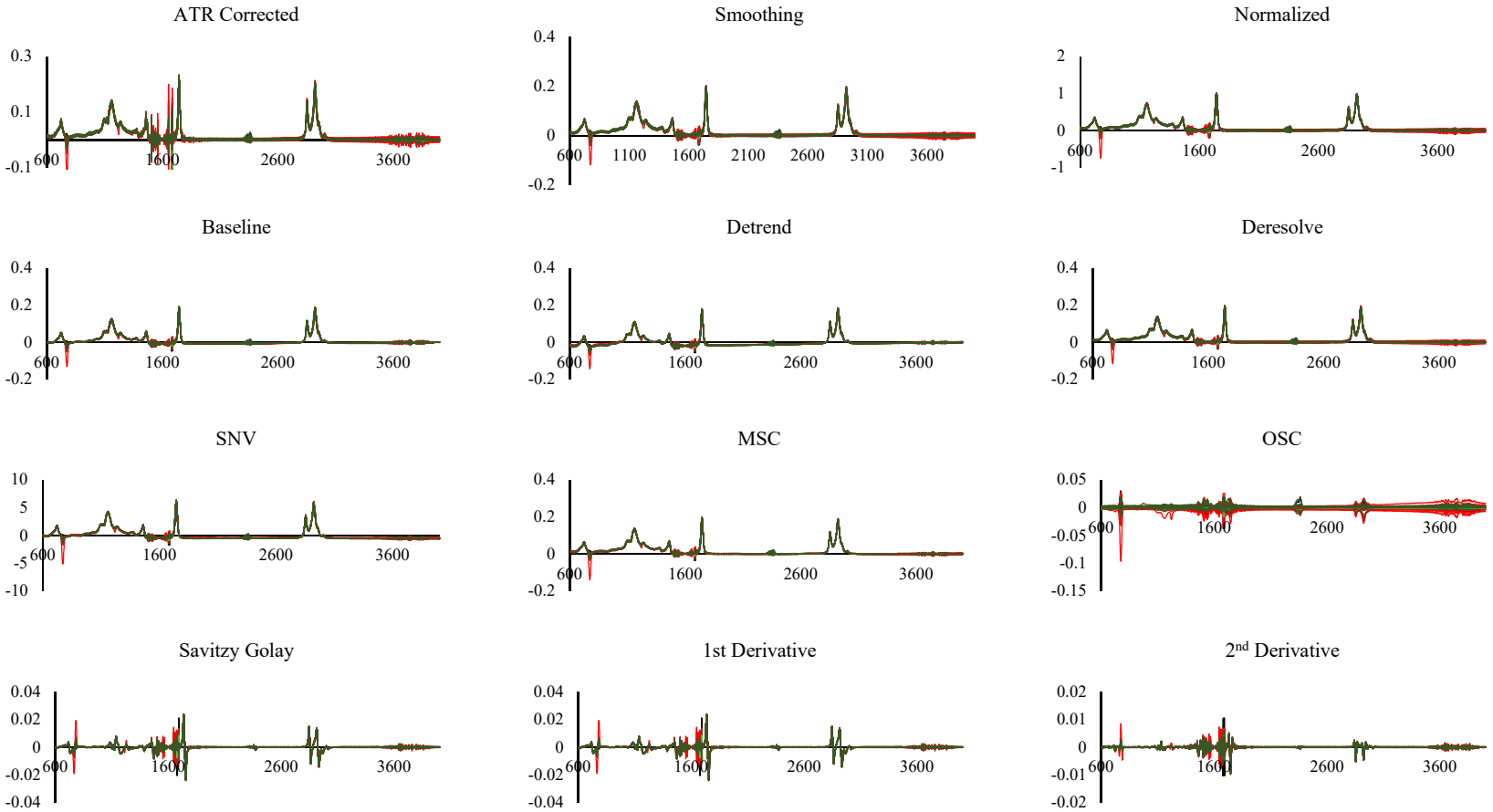


Figure 14.10: Application of different spectral pretreatments on FTIR spectra for EVOOs derived from different cultivars (Black= 'Bidni', Green='Malti', and Red='Bajda').

Table 14.4: Performance of the PCR models obtained during training (Top) and testing (Bottom) using FTIR data after different spectral pretreatments for the different cultivars.

	Raw	Smoothing	Base	Norm	DS	DT	SNV	MSC	OSC	SG	1D	2D
PCR FTIR Training												
						RMSE						
'Bidni'	1.328	1.021	3.027	1.117	2.968	0.993	2.887	2.882	0.332	3.083	3.083	3.006
'Bajda'	1.100	0.820	3.402	0.814	4.019	0.934	2.921	2.965	1.245	3.455	3.455	3.085
'Malti'	1.731	1.569	3.775	1.554	3.743	1.554	3.514	3.512	1.265	3.626	3.626	3.245
						MBE						
'Bidni'	-0.214	-0.164	-0.136	-0.210	-0.138	-0.160	-0.149	-0.149	-0.088	-0.095	-0.095	-0.171
'Bajda'	-0.038	-0.046	-0.183	-0.007	-0.289	-0.082	-0.200	-0.201	-0.030	-0.422	-0.422	-0.380
'Malti'	-0.069	-0.093	-0.607	-0.039	-0.625	-0.097	-0.584	-0.585	-0.061	-0.533	-0.533	-0.468
						MAE						
'Bidni'	0.282	0.213	0.477	0.264	0.495	0.209	0.470	0.469	0.117	0.527	0.527	0.497
'Bajda'	0.195	0.151	0.623	0.238	0.859	0.160	0.538	0.546	0.223	0.835	0.835	0.690
'Malti'	0.206	0.221	0.889	0.190	0.909	0.225	0.873	0.874	0.171	0.821	0.821	0.723
						R^2						
'Bidni'	0.936	0.966	0.342	0.959	0.366	0.968	0.398	0.400	0.994	0.339	0.339	0.353
'Bajda'	0.913	0.951	0.219	0.952	0.154	0.939	0.546	0.539	0.890	0.147	0.147	0.319
'Malti'	0.786	0.831	0.020	0.839	0.029	0.835	0.126	0.128	0.889	0.089	0.089	0.250
PCR FTIR Predictability												
						MBE						
'Bidni'	0.546	0.770	4.572	0.789	4.841	0.781	4.890	4.887	21.480	5.125	5.125	5.175
'Bajda'	1.636	2.755	2.778	2.420	3.650	2.819	1.162	1.170	5.052	2.280	2.280	3.508
'Malti'	0.643	0.486	2.213	1.291	3.141	4.351	1.763	1.793	11.997	1.536	4.087	1.339
						MBE						
'Bidni'	-0.066	-0.087	-0.048	0.030	-0.069	-0.089	-0.061	-0.060	-3.714	-0.070	-0.070	-0.059
'Bajda'	-0.054	0.032	-0.259	-0.014	-0.402	0.040	-0.204	-0.206	0.702	-0.186	-0.186	-0.333
'Malti'	0.092	-0.010	0.061	-0.097	-0.127	0.188	0.028	0.031	1.831	-0.008	-0.140	-0.041
						MAE						
'Bidni'	0.078	0.096	0.670	0.104	0.713	0.094	0.702	0.701	3.714	0.748	0.748	0.750
'Bajda'	0.244	0.317	0.529	0.305	0.698	0.319	0.217	0.219	0.702	0.434	0.434	0.660
'Malti'	0.092	0.041	0.332	0.180	0.403	0.201	0.308	0.311	2.142	0.258	0.394	0.199
						R^2						
'Bidni'	0.991	0.990	0.186	0.986	0.384	0.991	0.436	0.436	0.989	0.389	0.389	0.408
'Bajda'	0.967	0.542	0.629	0.586	0.529	0.508	0.972	0.972	0.973	0.589	0.589	0.547
'Malti'	0.998	0.984	0.773	0.932	0.740	0.027	0.797	0.796	0.744	0.816	0.739	0.976

Table 14.5: Performance of the PLSR models obtained during training (Top) and testing (Bottom) using FTIR data after different spectral pretreatments for the different cultivars.

	Raw	Smoothing	Base	Norm	DS	DT	SNV	MSC	OSC	SG	1D	2D
PLSR FTIR Training												
						RMSE						
'Bidni'	1.320	1.050	3.037	1.144	2.993	1.023	2.407	2.913	0.389	3.023	3.023	3.154
'Bajda'	0.869	0.869	4.044	0.889	4.199	0.936	3.071	3.121	1.305	3.423	3.423	3.158
'Malti'	1.546	1.452	3.713	1.416	3.719	1.431	3.601	3.599	1.262	3.588	3.588	3.212
						MBE						
'Bidni'	-0.201	-0.166	-0.137	-0.211	-0.145	-0.164	-0.086	-0.158	-0.086	-0.126	-0.126	-0.241
'Bajda'	-0.052	-0.052	-0.396	0.008	-0.398	-0.059	-0.313	-0.319	-0.102	-0.382	-0.382	-0.377
'Malti'	-0.120	-0.079	-0.649	-0.062	-0.663	-0.083	-0.575	-0.576	-0.077	-0.541	-0.541	-0.436
						MAE						
'Bidni'	0.270	0.215	0.488	0.263	0.505	0.211	0.334	0.482	0.132	0.471	0.471	0.675
'Bajda'	0.170	0.170	0.956	0.256	0.993	0.162	0.698	0.713	0.262	0.793	0.793	0.702
'Malti'	0.274	0.213	0.927	0.213	0.946	0.216	0.859	0.861	0.185	0.827	0.827	0.692
						R^2						
'Bidni'	0.935	0.963	0.338	0.956	0.355	0.965	0.580	0.388	0.990	0.345	0.345	0.296
'Bajda'	0.945	0.945	0.093	0.944	0.060	0.937	0.416	0.401	0.877	0.202	0.202	0.297
'Malti'	0.829	0.850	0.064	0.861	0.071	0.854	0.120	0.121	0.887	0.120	0.120	0.265
PLSR FTIR Predictability												
						RMSE						
'Bidni'	0.456	0.806	4.602	0.591	4.848	0.827	4.905	4.901	21.392	5.204	5.204	4.914
'Bajda'	1.706	2.539	2.044	2.571	2.087	2.792	1.349	1.372	11.007	2.782	2.782	3.452
'Malti'	0.555	0.529	1.969	1.207	2.978	4.355	1.457	1.494	10.596	1.502	4.100	1.301
						MBE						
'Bidni'	-0.056	-0.088	-0.048	0.084	-0.068	-0.092	-0.059	-0.058	-3.698	-0.056	-0.056	-0.093
'Bajda'	-0.077	-0.080	-0.117	-0.009	-0.126	0.037	-0.140	-0.140	1.710	-0.255	-0.255	-0.320
'Malti'	0.071	-0.010	0.085	-0.083	-0.103	0.188	0.043	0.046	1.572	0.004	-0.152	-0.045
						MAE						
'Bidni'	0.068	0.100	0.679	0.104	0.715	0.100	0.703	0.702	3.698	0.756	0.756	0.726
'Bajda'	0.122	0.369	0.387	0.319	0.396	0.318	0.241	0.245	1.710	0.537	0.537	0.651
'Malti'	0.072	0.049	0.253	0.164	0.334	0.207	0.242	0.246	1.883	0.250	0.402	0.196
						R^2						
'Bidni'	0.994	0.989	0.199	0.996	0.387	0.990	0.449	0.449	0.987	0.424	0.424	0.388
'Bajda'	0.981	0.617	0.643	0.508	0.634	0.513	0.938	0.938	0.894	0.570	0.570	0.545
'Malti'	0.999	0.979	0.857	0.943	0.795	0.024	0.880	0.879	0.756	0.829	0.727	0.982

Analysis of the prediction models for individual cultivars, showed that in general the models obtained for the ‘Bajda’ cultivar had a lower performance when compared to those obtained for the ‘Bidni’ and ‘Malti’ cultivar for the majority of the spectral transformations. This could be indicative that the presence of chlorophyll pigments could be affecting the thermal oxidative process, further emphasizing their antioxidant effect in this experiment. The next step in determining the effects of the thermal degradation process was to evaluate the changes occurring within the individual peaks. The application of β-regression coefficient was used in order to provide a more direct assessment of the variables which were offering the most important impact on the response variable (thermal degradation).

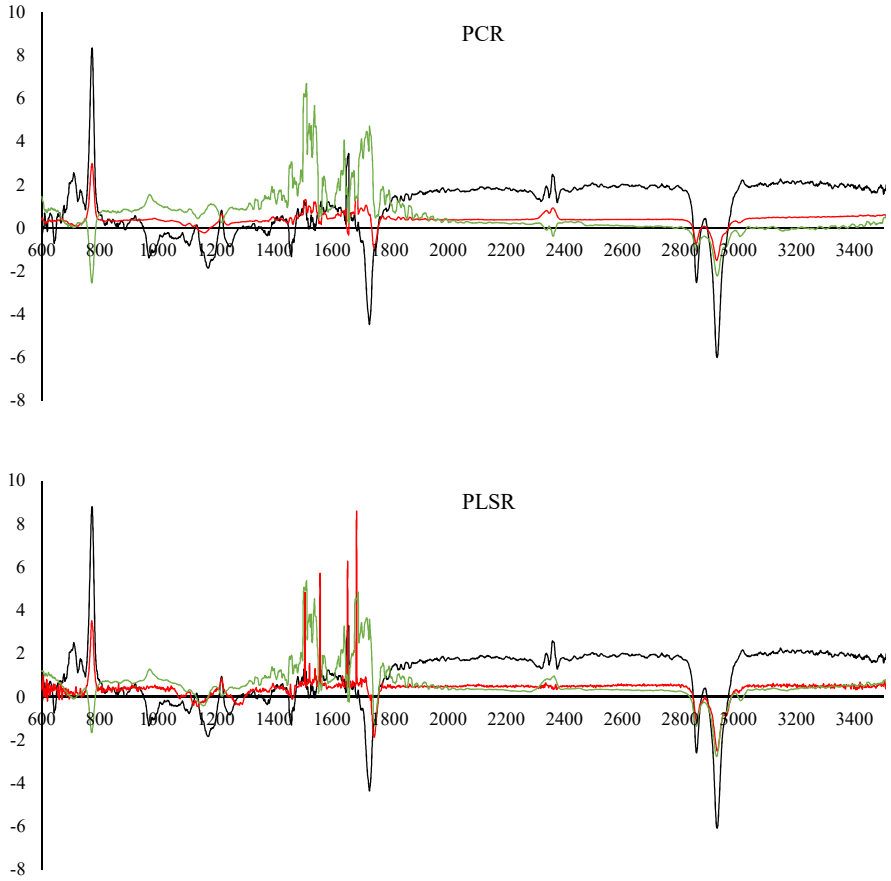


Figure 14.11: β-regression coefficients obtained from PCR (Top) and PLSR (Bottom) models using FTIR data for the different cultivars (Black = ‘Bidni’, Green = ‘Malti’, Red = ‘Bajda’).

From that analysis of the β- regression coefficients it was evident that the models obtained for the 'Bidni' cultivar had a lower input of signal noise when compared to those obtained from the 'Malti' and the 'Bajda' cultivar, thus explaining the lower model performance for the latter cultivars. Nonetheless for the three cultivars the β-regression coefficients revealed that thermo-oxidative stress caused substantial changes throughout the entire infrared spectra. The most obvious changes, which corresponded to variables with higher β-coefficient magnitude were divided into three major regions, 2700- 3006 cm^{-1}, assigned to asymmetric stretching vibration of C–H of aliphatic CH_3 and CH_2 together with C–H stretching vibration of cis double bonds, the C=O region observed at 1720–1750 cm^{-1}, and the fingerprint region (1500−900 cm^{-1}). Within the fingerprint particular peaks of interest were identified namely the 967 cm^{-1} which corresponded to nonconjugated trans double bonds, 987cm^{-1} of the conjugated trans double bonds, together with the 722 cm^{-1} corresponding to cis double bonds of disubstituted olefins.

The sign of the β- regression coefficient for certain spectral regions, indicates changes in the intensity of bands recorded at specific frequencies as a consequence of thermo-oxidative degradation. Wavelengths having a negative β- coefficient implies that every unit increase in X (absorbance), an expected unit decrease in Y (the extent of degradation given in terms of weeks). The negative regression coefficients observed for the peaks observed in the 2700- 3006 cm^{-1} suggest that as the thermal degradation process continues the functional groups responsible for the absorbance in this region diminish. These changes were attributed to changes in the degree of unsaturation of oil samples in response to thermal stress. Changes within the different regions of the FTIR can be explained by cis-trans isomerization and conjugation of double bonds of PUFAs, the formation of secondary oxidation products and changes in the ratio of CH_2 and CH_3 terminal groups (Mallegol *et al*., 1999 and Gonzaga *et al*., 2007).

The loss of the cis double bonds, was mainly attributed to the isomerization to trans groups and/or their breakdown to produce secondary oxidation products, is specific to lipids undergoing oxidation processes as a result of heat stress (Van de Voort *et al*.,1994). During heating formation of primary oxidation products at a rate lower than that of their decomposition could explain the reduction in the absorbance of the cis double bond functional groups. The degradation of hydroperoxides of PUFAs, as primary oxidation products, results in a complex mixture of secondary

products (aldehydes, ketones, acids and esters) and consequently, leads to a decrease in the degree of unsaturation due to the disappearance of double bonds in the cis conformation (Moreno *et al*., 1990). Moharam *et al*., 2010 also, suggested that the decrease in the degree of unsaturation can be associated with a decrease in the level of free radicals that contain cis double bonds formed during the heating process. These radicals decay as a direct result of scavenging by antioxidant compounds such as tocopherols and phenolic compounds, found in oil samples (Moharam *et al*., 2010). Van de Voort *et al*. (1994), showed that the reduction in the unsaturated fatty acids content (18:2 and 18:3 fatty acids) as a result of oxidation.

It was expected that the progressive decrease in the absorbance at 3006 cm^{-1} would be coupled with a decrease in the absorbance at 780 cm^{-1} which belongs to the out of plane cis double bonds bending mode of unsaturated fatty acids. However analysis of the β- regression coefficient showed that for the 'Bidni' and the 'Bajda' cultivar a positive correlation exists between the extent of thermal degradation and the specific absorbance at this peak whilst in the case of the 'Malti' cultivar a negative correlation with the extent of degradation exists. This suggests that whilst the cis double bond degradation could be visualized by the analysis of the 3006cm^{-1} peak, this could not be fully extended to the 780cm^{-1} peak. The presence of minor constituents presents within EVOOs derived from different cultivars might be affecting the susceptibility of the EVOOs to thermal degradation. The difference in the 780cm^{-1} peak correlation provides a further corroborating to the fact that a universal model explaining the thermal degradation for all the different cultivars could not be built using FTIR data.

Unlike what was observed by Alexa *et al*., 2007 during the course of thermal degradation no increases in the saturation of the substrate because the cis-olefinic double bonds of the different acyl groups disappeared during the thermo-oxidation process as the peak at 2854 cm^{-1} corresponding to the symmetric stretching vibration of C–H of aliphatic CH_2 group still retained at negative β-regression

The cultivar dependency on the thermal degradation is further highlighted in the analysis of the fingerprint region 600- 1200cm^{-1}. The band observed at 987 cm^{-1} was associated with bending vibrations of C−H trans, trans and cis, trans conjugated diene groups of hydroperoxides, as previously reported by Guillen *et al*. (2005) and

Poiana *et al*., (2015). Although studies have shown that this band appears as a result of thermal stress, no observable contribution to the regression models obtained using both PLSR and PCR were attributed to this peaks. Guillen *et al*. (2005) showed that hydroperoxide groups are associated with conjugated double bonds. However, similar to our study, Mallegol *et al*., (1999) showed that the intensity of the band attributed to conjugated trans isomers is more important in the study of oil samples oxidation compared to the band assigned to nonconjugated trans isomers groups of hydroperoxides. Although the results obtained might not be consistent with the ones observed in the literature, these can be explained in terms of the methodology. In the literature, in the majority of the thermal degradation studies the oils are exposed to oxygen which facilitates the formation of oxygen-containing degradation products namely, hydrogen peroxides, and secondary oxidation products such as aldehydes or ketones. In this experiment the EVOOs were thermally degraded without the presence of oxygen as the samples were sealed prior to the experiment without a headspace, thus reducing their aerial exposure, hindering the formation of oxygen-containing compounds. Further analysis of peaks corresponding to secondary oxidation products observed at 967 cm^{-1} assigned to aldehydes or ketones that contain isolated trans double bonds (Guillen at al., 2005 and Van de Voort *et al*., 1994).

It became more noticeable, that the reduction of oxygen during the thermal experiment prevented or at least mitigated the formation of these compounds as proven by the lack of any significant contribution to the regression models obtained for all the cultivars under investigation. Other peaks which did not show any contribution to the regression models include the epoxy peak observed at 885cm^{-1}, where epoxides are formed from the thermal decomposition of methyl oleate hydroperoxides (Mallegol *et al*., 1999).

14.5.3 Chemometric analysis of fluorescent changes during thermal degradation

There are several studies that employ the use of fluorescence spectroscopy for monitoring the degradation of extra virgin olive oil under thermoxidizing conditions (Guzmán *et al*., 2015; Cheikhousman *et al*., 2005; Poulli, *et al*., 2009; Tena, Aparicio, and García-González, 2012; Tena, García-González, and Aparicio, 2009). There have also been studies which focus on the characterization and association of fluorescence spectra with quality parameters and typical signals of chemical oxidation including peroxide value, K_{232} and K_{270} values (Guzmán *et al*., 2015, Guimet, Boqué, and Ferré, 2006; Guimet, *et al*., 2005; Sikorska *et al*., 2005; Zandomeneghi, 2006). All the aforementioned studies have revealed that the application of fluorescence spectroscopy provides insights about the quality of EVOOs and its degradation during heating, enabling a more direct evaluation of the oxidative status of oils.

The aim of this part of the study, therefore, was to determine whether fluorescence spectroscopy coupled with three different multivariate regression models provides enough information for the development of chemometric models which could predict the extent of thermal degradation. Similar to the determination of geographical origin synchronous emission spectra were obtained at 6 different wavelength intervals in order to obtain models which represent the major parts of the total luminous spectrum. Each of the 6 different spectra was treated using 12 different spectral transformations and coupled with three multivariate regression models namely PCR, PLSR, and linear SVRM. The performance of the models was assessed using RMSE, MAE, MBE and R^2 values for both the calibration and prediction data sets.

Figure 14.12 shows the SEEF obtained at different wave interval. The fluorescence emission spectra of olive oil show three major peaks corresponding to polyphenol and tocopherol which emit in the 300–390 nm range (Zandomeneghi, Carbonaro, & Caffarata, 2005; Giungato *et al*., 2004). Two smooth peaks at 445 and 475 nm which were identified by Kyriakidis and Skarkalis (2000) as related to the oxidation products of fatty acids whilst the peak observed at 525 nm was derived from vitamin E. However, the study carried out by Kyriakidis and Skarkalis (2000) revealed that the addition of vitamin E that to virgin olive oil increases the fluorescence intensity not only at 525 nm but also at 445 and 475 nm. They suggested that this was due to oxidized vitamin E that fluoresces at about this region. Other compounds which

emit in the 445-475 nm include monounsaturated fatty acids and phenolic antioxidants. The peak at 681 nm is related exclusively to chlorophylls and their corresponding degradation products namely pheophytins.

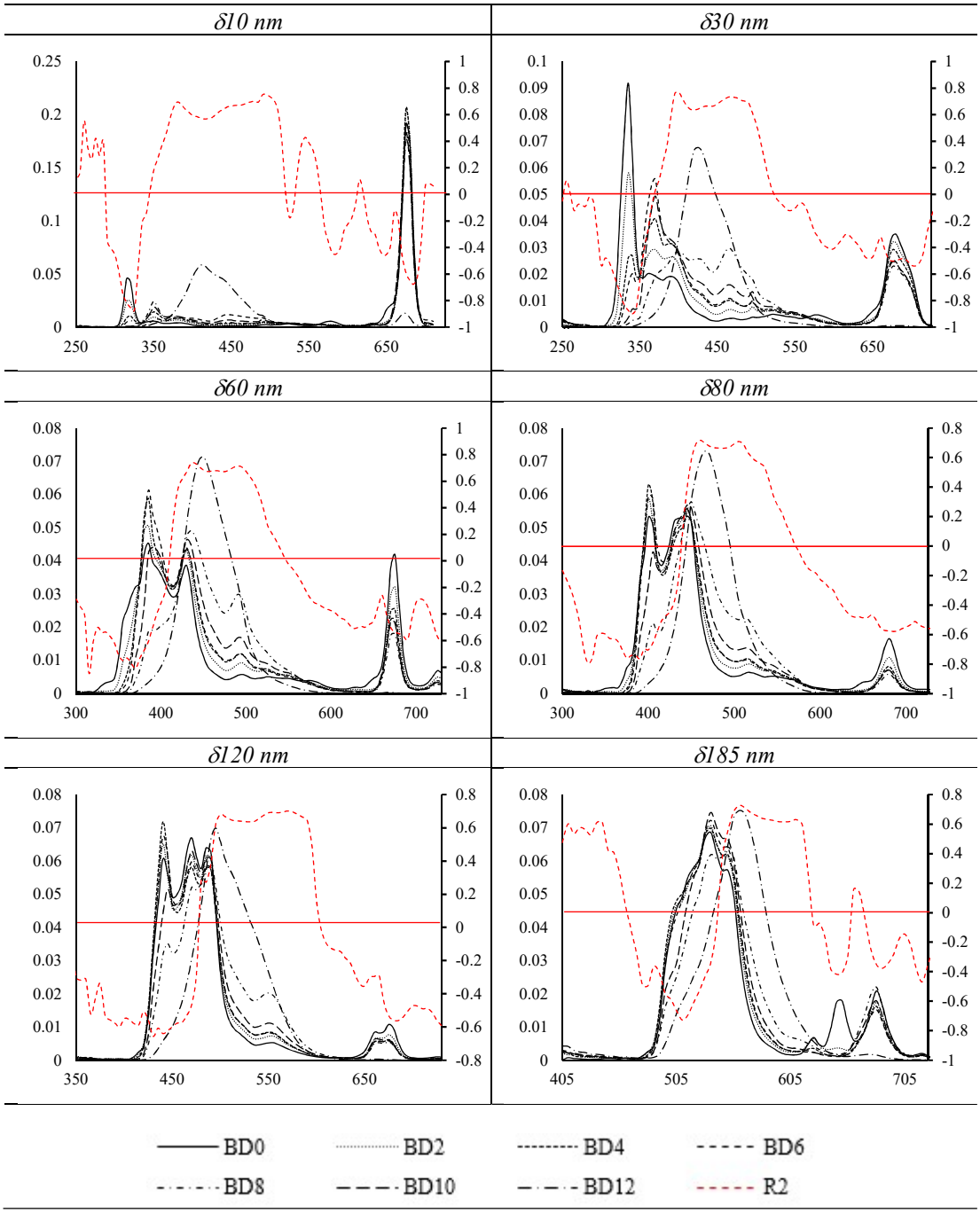


Figure 14.12: Changes in SEEFs recorded at δ10, 30, 60, 80, 120, 185 nm during thermal degradation. The red line represents the R^2 value obtained through a correlation analysis carried out on the intensity of emission and the extent of degradation.

Figure 14.12 also highlights the major changes in the SEEF of EVOOs during the 12-week degradation study. Similar to the studies found in literature the first most

obvious change in the SEEF spectra is a bathochromic effect which takes place as the oil undergoes oxidation. New peaks start to appear at higher wavelengths especially within the 400-600 nm region, these were attributed to a number of different chemical constituents which are developed during the course of degradation. The evolution of these chemical constituents by time can be further assessed using the coefficient of linearity extracted from a matrix correlation which was carried out comparing the absorbance observed at a different wavelength to the extent of thermal degradation. The 400-600 nm region throughout all the SEEF spectra showed a strong significant positive correlation with the extent of thermal degradation. The presence of this peak is namely attributed to the evolution of polar compounds namely phenolic compounds namely tyrosol and phenolic acid derived from the degradation of conjugated secosteroid compounds, which will be presented in the next chapter together with the presence of free fatty acids and their corresponding oxidised products. Poulli, Mousdis, and Georgiou, (2005) showed that the amount of free fatty acids derived from the corresponding hydrolysis of triglycerides during the process of thermal degradation is reflected in the fluorescent component. The free acidity which increases during the degradation increases the amount of oleic, linoleic and palmitic acids. Whilst butyric acid (palmitic acid analogue) and linoleic acid show fluorescence bands at 273 and 325 nm, oleic acid shows a band of fluorescence at 405 nm, thus Poulli, Mousdis, and Georgiou, (2005) concluded that the fluorescence intensity observed at 429–545 nm was due to an increase oleic acid content.

Other studies carried out by Tena *et al*., (2009) showed that the increase in the 400-600 nm fluorescent intensity could not be solely attributed to the presence of free fatty acids formed during the degradation. Tena *et al*., (2009) showed that five phenols namely tyrosol, o-coumaric, vanillic, syringic, and gallic acids spectra recorded in methanol had emission maxima between 349 and 426 nm. Phenolic profiling carried out by Tena *et al*., (2009) on the thermal degraded EVOO samples revealed that after 36 h of thermoxidation the percentage of polar compounds was higher than 25% thus confirming that the peak observed at 400-600 nm could not be solely attributed to the formation of free acids. Furthermore in the study carried out by Tena *et al*., (2009) showed that although the second maxima observed at 540 nm which was recorded also in this study becomes more pronounced on moving to higher wavelength intervals,

was not due to the presence of other compounds but it corresponded to the overtone of 270 nm and, therefore, was neglected.

The fluorescence bands observed in the region 300–370 nm were found to decrease during oxidation this was marked by the strong negative correlation. This peak was attributed to the presence of tocopherols. Non-degraded EVOOs contain different tocopherols (α, β, and γ), of which α-tocopherol is found in the highest concentration (Uceda *et al*., 2001), while tocotrienols are absent (Morales and Tsimidou 2000). Through the use of matrix correlation Tena *et al.,* (2009) confirmed that concentration of α-tocopherol and the fluorescent spectrum showed good correlations in the excitation range of 300-380 nm with the emission range 353-355 nm being the most significant. The band between 630 and 750 nm is associated with chlorophylls and pheophytins according to some authors (Kyriakidis and Skarkalis 2000; Sikorska *et al*., 2005), also decreases with the thermo-oxidation time (Figure 14.12), as pointed out the negative coefficient of linearity these corroborate the previous observation carried out using UV-Vis spectroscopy and are concordant with the study carried out by Poulli, Mousdis, and Georgiou, (2005) .

For the six SEEF spectra, three linear regression models were built for 10 different spectral transformations for a total of 180 regression models. The performance of each model was assessed using RMSE, MAE, MBE and R^2 values, which are summarised in Figure 14.13. In general, it was found that the models obtained using support vector machine regression (SVMR) had a very high performance during the training phase for all the spectral transformation for all the SEEF spectra studied. The SVMR models obtained during the training phase had a very low RMSE, MAE and an R^2 value very close to unity. In comparison, the training models obtained using PLSR and PCR had a lower performance when compared to SVMR models, with PLSR showing a general slightly higher performance than PCR. It was evident that SEEF obtained at δ30 nm after matrix scatters correction yielded very poor results during the training phase, this observation was extended to models obtained using PCR and PLSR, however, this was not evident in the linear models obtained using SVRM. This can be due to, model overfitting in the case of SVRM or else that the SVRM might be actually improving the modelling power of all SEEF spectra.

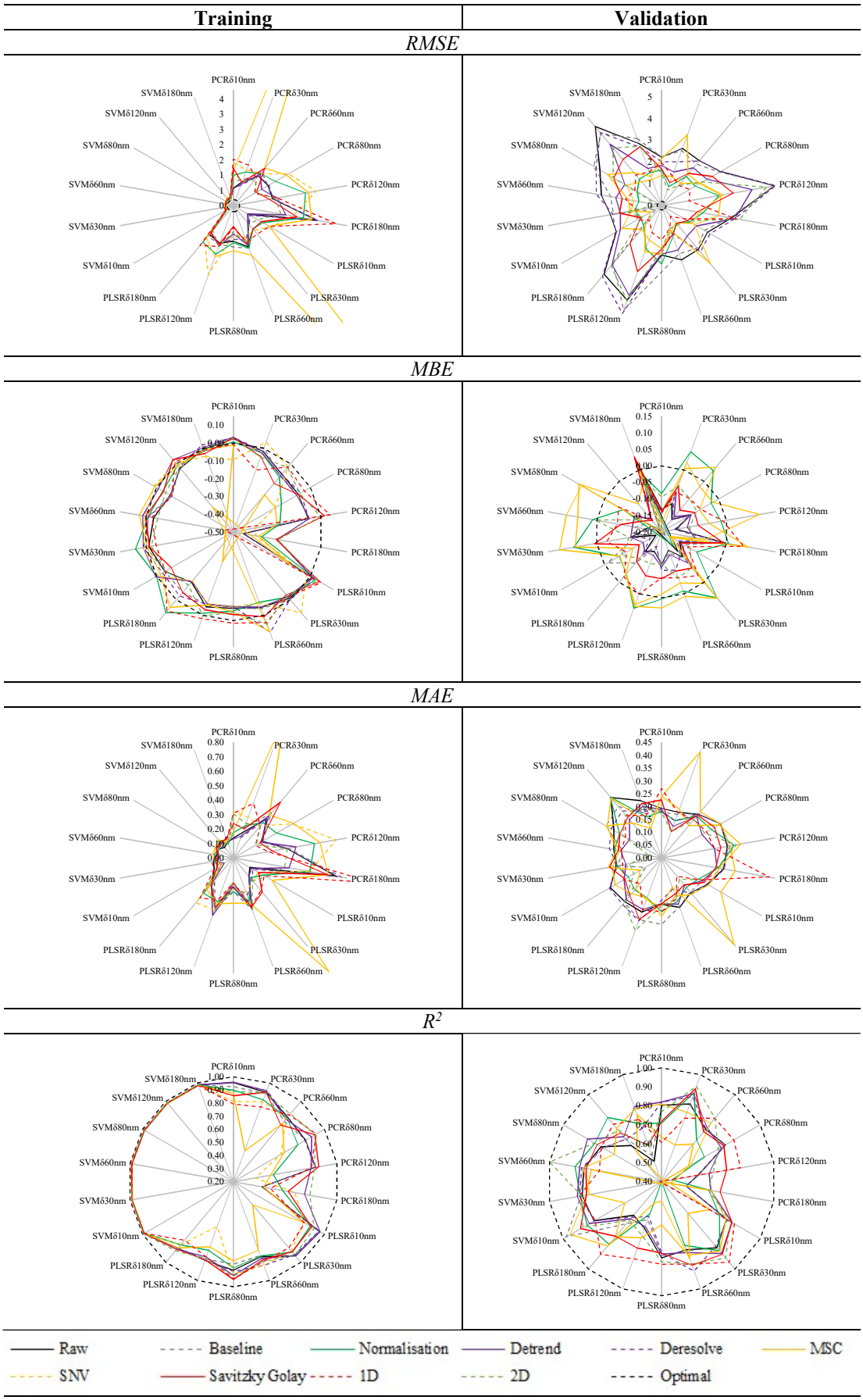


Figure 14.13: Radar plot summarising the performance of the different regression models (PLSR, PCR and SVMR) obtained on the different SEEFs subjected to different spectral pretreatments.

On analysis of the model predictability during the validation stage it was shown the for the majority of the spectral transformation linear regression models obtained using PCR, PLSR and SVRM had no major differences in the predictability of the testing data set. Detailed analysis of the predictability of the models obtained it was found that in general, the model performance is dependent on both the transformation employed and to the wave interval at which SEEF is obtained. Similar results were obtained during the application of SEEF for the determination of geographical origin, this was attributed to the fact that SEEF represents just a very small portion of the total luminous spectrum and thus different SEEF might be reflecting different fluorescent compounds found in EVOOs. In this case, it was found that models obtained using SEEF spectra obtained at δ120 nm, had a lower model performance for the majority of the spectral transformations. This observation was coherent for all the three regression models obtained. On the other hand, SEEF spectra obtained at δ30 nm had the best modelling power as very low RMSE and MAE values were obtained for the majority of the spectral transformations over the three different regression models. This suggests that SEEF obtained at δ30 nm reflect a complete overview of the changes that occur to fluorescent compounds during thermal degradation.

In general, it was observed that 1st order derivatized SEEF spectra had a higher PCR and PLSR model performance at SEEF obtained at intervals δ 60, 80, and 120 nm, whilst in the case of SEEF obtained at δ10 and 180 nm this was restricted to only PLSR models. In the case of PCR models obtained for SEEF δ10 30 and 180 nm, it was found that models reached optimal performance using standard normal variate correction, 2nd order and Savitsky-Golay derivatization respectively. In the case of SVM models, there was no clear spectral transformation that gave a consist higher model performance for the different δ SEEF spectra. These observations suggest that the model performance was dependent on both the spectral transformation and correction to which the SEEF spectrum was subjected and the wave interval at which the SEEF spectrum was obtained. Figure 14.14 shows the β-regression coefficients obtained for the best spectral transformations obtained for PCR and PLSR regression models.

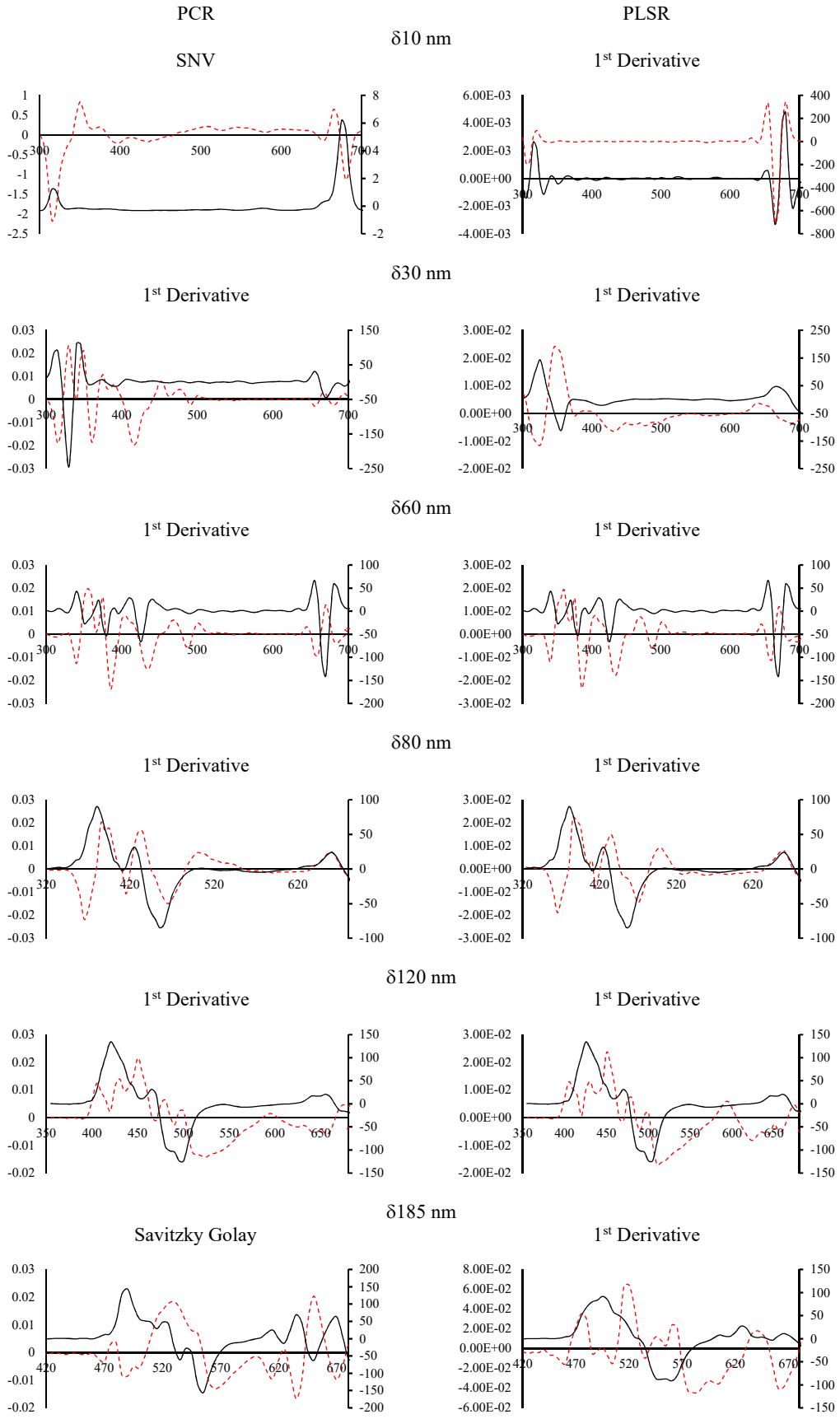


Figure 14.14: β-regression coefficients for the best performing PCR (Left column) and PLSR (Right column) obtained for the different SEEFs at different δ nm.

Detailed analysis of the β-regression coefficients obtained for the best performing PCR and PLSR models at different SEEF spectra obtained using different wave intervals are shown in Figure 14.14. As expected it was revealed that the peaks corresponding to fluorescent compounds in EVOOs had the largest β-coefficient magnitude, indicating that the predicted formulas obtained were highly dependent on the presence of these compounds and that the model obtained had minimal influence due to variation arising from instrumental noise and scattering effects. In the case of SEEF spectra in which the optimal models were obtained from one particular spectral pretreatment. The regression coefficients obtained had the same spectral profile indicating that irrelevant of the regression model, the most significant contribution to the regression was due to the same variables. In general as previously expected, negative β- regression coefficients were obtained for the wavelengths which corresponded to the emission of tocopherol and chlorophyll pigments. This indicates that a negative correlation between the emission of these compounds and the extent of thermal degradation. On the other hand, a positive correlation was observed for the peaks obtained between 400-550 nm, further indicating an increase in the fluorescence of these compounds due to degradation.

Analysis of the β- regression coefficients obtained for SEEF obtained at δ10 nm revealed that in fact the peak observed at high wavelengths, which is generally attributed to the chlorophyll pigment emission was actually composed of two separate peaks, one which had a maximum around 650 nm and another one around 680 nm. The presence of these two distinct peaks was further resolved on the application of the 1^{st} derivative function. From the β- regression coefficients obtained, it was found that whilst the peak obtained at 680 nm showed a negative correlation with the extent of thermal degradation the peak observed at 650 nm showed a slightly positive correlation indicating a potential increase during the thermal degradation process. The most plausible explanation for the results obtained was that the peak obtained at higher wavelengths was attributed to the presence of chlorophyll pigment which decreases in concentration during the thermal degradation thus showing a negative correlation. The thermal degradation products obtained from chlorophylls are pheophytins which tend to increase during the thermal degradation progress, thus showing a positive correlation. However, as the thermal degradation process continues pheophytin compounds are a further degraded to non fluorescent pheophorbides and colourless

chlorophyllides (Ward *et al*., 1994). Studies carried out on the degradation of chlorophyll pigments showed that non-degraded chlorophylls tend to have emission at lower wavelengths (≈ 20 nm) compared their corresponding pheophytin derivatives irrelevant of the solvent in which the fluorescence is measured (Ward *et al*., 1994).

14.5.3.1 Application of Synchronous Fluorescence Spectroscopy as predictors for conjugated diene and triene content.

In order to determine the potential use of SEEFs as a measure of conjugated diene and triene content in EVOOs, a PLSR model was built using the previously determined K-value as the Y response matrix whilst the SEEF obtained at δ10 nm was used as the predictor X matrix. Figure 14.5 shows the first factor β-regression coefficients obtained for the different experimentally determined K-values. It was shown that for the different K-values (K_{266}, K_{270}, K_{274}, K_{230} and Δk) regression coefficients were observed across the SEEF spectrum obtained at δ10 nm. As expected, there was a negative correlation with the 330 nm maxima which corresponds to the phenolic and tocopherol content present within the EVOOs, indicating that as the K-values increase during the thermal degradation experiment there is a decrease in both the phenolic and tocopherol content. Similarly, a negative correlation was also observed at the 680 nm maxima which correspond to the chlorophyll to the decrease in chlorophyll pigments during the thermal degradation experiment. In line with was previously observed the β-regression coefficient obtained it was shown the prediction formula obtained for the K-values was also dependent on the 350-500 nm region which showed a positive correlation with the K-values. As expected, this suggests that compounds within this region might be directly related to dienes and trienes which also increase during the thermal degradation process. As regards the peak centred at 666 nm which also showed a positive correlation with the K-values, although this peak was not completely identified, the proposed possibility of being related to pheophytin compounds which increase in tandem with the K-values, is plausible. Analysis of the PLSR model performance obtained it was clear that the potential application of SEEF as predictors of K-values was not equally effective. Table 14.6 summarises the results obtained from the PLSR model, whilst during the training, the model obtained was highly effective in predicting the different K-values, with the exception of the K_{230},

during the validation phase the prediction of K-values was very effective by the PLSR model obtained.

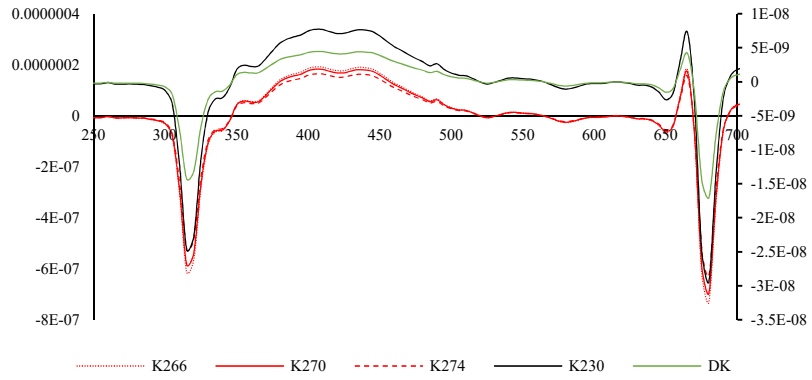


Figure 14.15: 1st-factor β-regression coefficients obtained from the application of PLSR on the use of δ 10 nm as predictors of diene and triene content in thermal degraded EVOO.

Table 14.6: Performance of the PLSR models obtained during training (Top) and testing (Bottom) using SEEFs δ 10 nm data as predictors of diene and triene content.

	Training			
	RMSE	MBE	MAE	R^2
K_{266}	0.001	0.000	0.001	0.857
K_{270}	0.001	0.000	0.001	0.869
K_{274}	0.001	0.000	0.001	0.869
K_{230}	0.000	0.000	0.000	0.229
Δk	0.000	0.000	0.000	0.869
	Testing			
	RMSE	MBE	MAE	R^2
K_{266}	0.005	0.002	0.003	0.724
K_{270}	0.005	0.002	0.003	0.715
K_{274}	0.004	0.002	0.003	0.713
$K2_{30}$	0.000	0.000	0.000	0.762
Δk	0.000	0.000	0.000	0.810

14.5.3.2 Application of Synchronous Fluorescence Spectroscopy as predictors of colour parameters.

The potential application of SEEF spectra obtained at δ10 nm was extended to the prediction of colour parameters previously determined using UV-Vis spectrophotometry. The colour parameters determined using the three previously described methods namely *L *a *b, XYZ, and xy system were used as the response Y matrix whilst the SEEF spectra obtained at δ10 nm was used as the predictor X matrix. Figure 14.16 and Figure 14.17 shows the first β-regression coefficient obtained for the different parameters used in the three colour system. In general it was shown that the regression coefficient is highly dependent on the colour system employed, whilst the parameters present with each colour system tend to follow similar regression coefficients, furthermore whilst the β-regression coefficients of the XYZ and xy system were similar, there was a notable difference in the regression coefficients observed for the *L *a *b colour system.

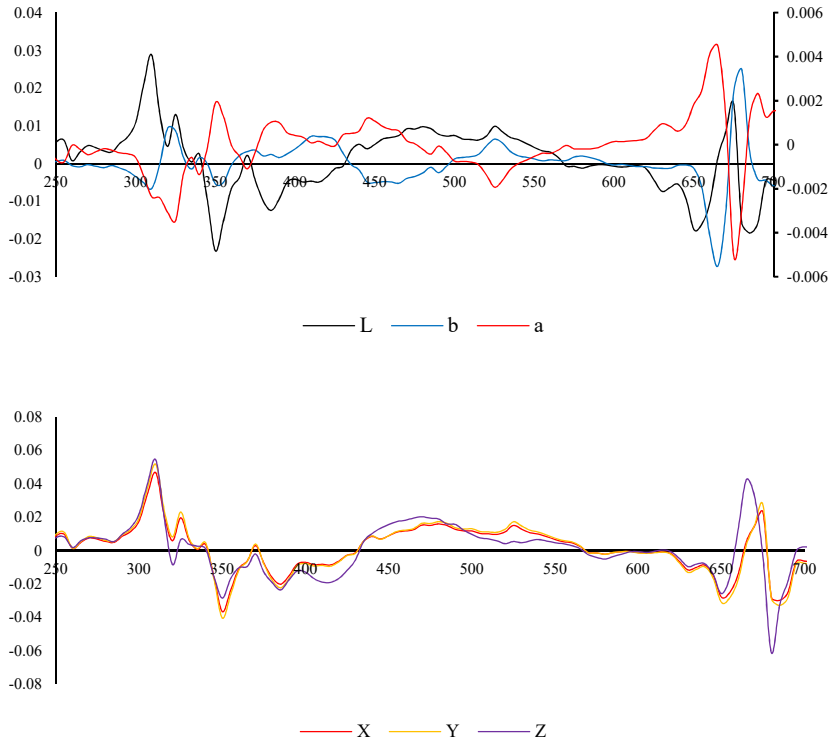


Figure 14.16: 1st factor β-regression coefficients obtained from the application of PLSR on the use of δ 10 nm as predictors of colour parameters (Top) *L *a * b (Bottom) XYZ in thermal degraded EVOO.

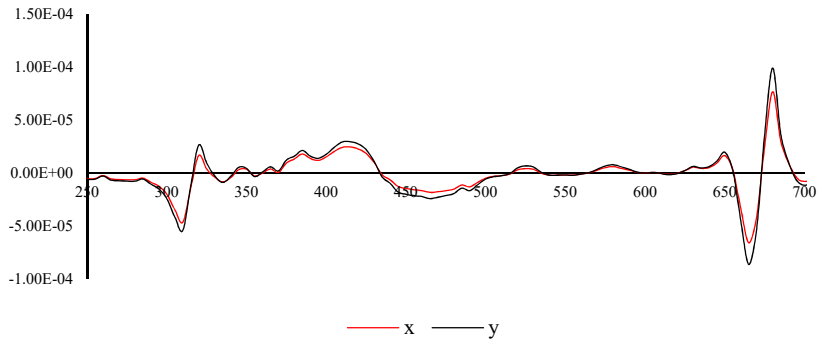


Figure 14.17: 1st-factor β-regression coefficients obtained from the application of PLSR on the use of δ 10 nm as predictors of colour parameters xy in thermal degraded EVOO.

In the case of *L parameter, it was shown that it was positively correlated with the emission observed between 300-350 nm (tocopherol/phenolic content) and 680 nm (chlorophyll content). This suggests that as the luminosity decreases (from dark to light colouration) during thermal oxidation this is coupled with a decrease in the content of this fluorescent compounds. In the case of the *a parameter (+a = red , -a = green) showed minimal contribution to the regression formula, when compared to the *L and *b, however it was shown that this colour parameter was mainly positively correlated with the identified peak observed at 666 nm which suggests that the fluorescence of this compound was responsible for the increase in the red colour parameter. Conversely the a* parameter was strongly negatively correlated with the peak observed at 675 nm which is associated with the chlorophyll pigment responsible for the green colour of EVOO, thus as a* increase (becomes less green and more red), there is decrease in chlorophyll pigments, further corroborating the results observed using UV-Vis spectrophotometry. This also suggests that unidentified pigment emitting at 666 nm is most likely to be a pheophytin derivative as these tend to be less green than their chlorophyll pigment counterpart. Similar to the *a parameter, *b parameter also showed that the 600-700 nm was highly important during the PLSR regression model and were less dependent on the phenolic and tocopherol content. In the case of the *b parameter (+b = yellow, -b = blue) the trend observed for the *a parameter was reversed, whereby the chlorophyll peak was positively correlated to the *b parameter, whilst as the potential pheophytin compounds increase a decrease in the b* parameter, indicating that as the chlorophyll content decreases there is a decrease in blue hues resulting in a more yellowish colour. For the XYZ and xy colour systems whilst the XYZ, parameters showed a positive correlation with the peak observed at

666, a negative correlation with respect to the 680 nm whilst in the case of xy colour system the observation was opposite. This observation suggests that as the amount of chlorophyll decreases this is coupled with a decrease and the amount of pheophytins increases there is a decrease in both red and green hues corroborating the results obtained using *L *a *b colour system.

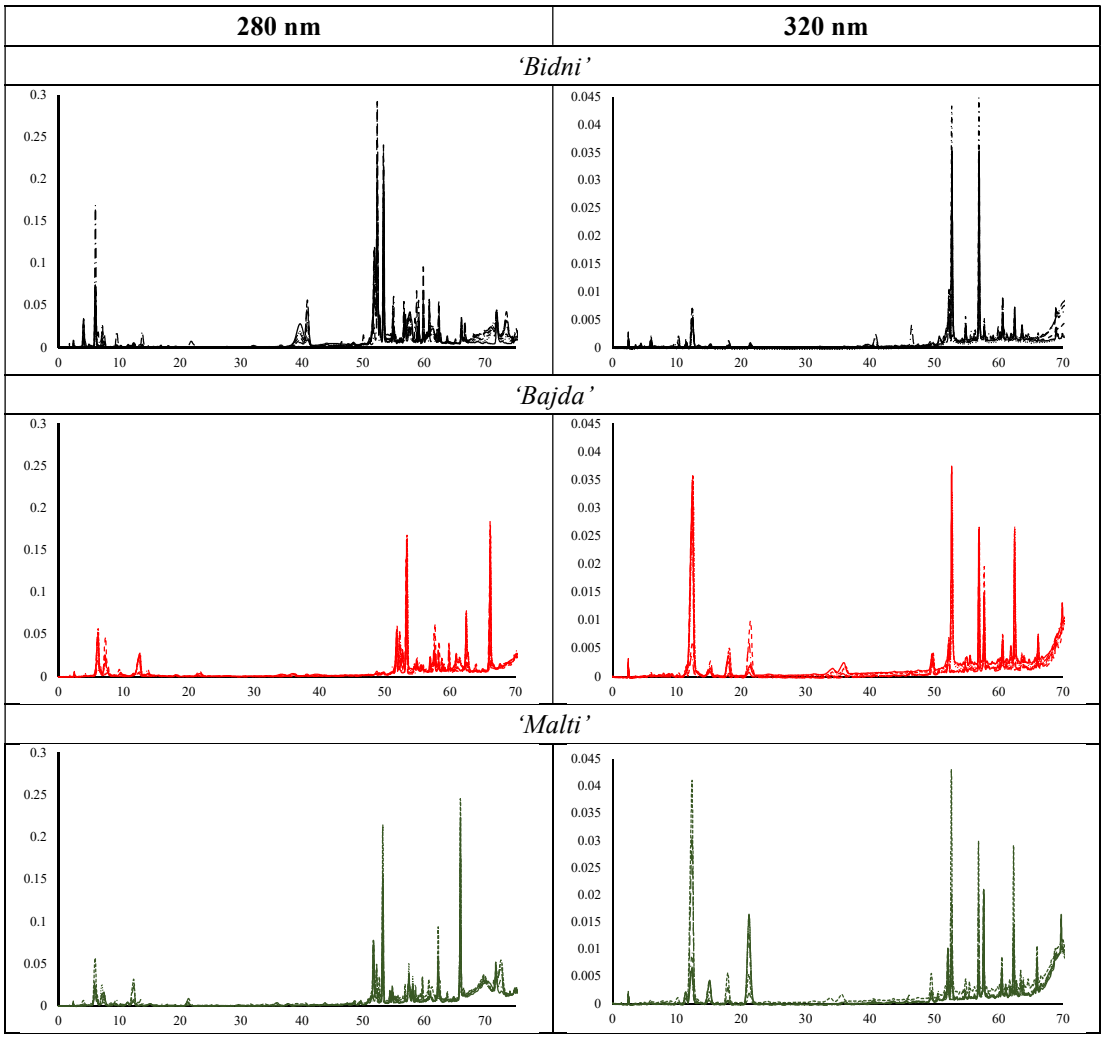
Table 14.7: Performance of the PLSR models obtained during training (Top) and testing (Bottom) using SEEFs δ 10 nm data as predictors of different colour parameters.

	RMSE	MBE	MAE	R^2
		Training		
L	5.195	-0.282	4.581	0.644
a	1.594	2.540	1.203	0.878
b	6.037	36.445	4.580	0.887
X	8.022	64.355	7.071	0.657
Y	9.111	83.012	8.011	0.654
Z	8.651	74.843	6.666	0.916
x	0.013	0.000	0.010	0.900
y	0.017	0.000	0.012	0.904
		Validation		
L	7.256	4.472	5.702	0.368
a	1.745	-0.978	1.154	0.953
b	3.394	2.088	2.626	0.985
X	11.653	6.975	9.068	0.391
Y	12.902	7.969	10.136	0.359
Z	10.667	4.476	8.560	0.909
x	0.008	0.002	0.007	0.973
y	0.011	0.004	0.010	0.973

Analysis of the model performance towards the different parameters in the three colour systems it was shown that SEEF obtained at δ10nnm were able to predict the xy colour parameters it a higher degree of accuracy when compared to other parameters in the other colour systems. Furthermore it was shown that whilst the *L parameter was not accurately predicted the *a and *b parameter of the*L *a *b colour system were far more accurately predicted indicating that the changes from green to red and blue to yellow provide more information than the actual luminosity parameter. In the case of XYZ system the Z parameter was more accurately predicted that the X and Y parameter indicating that compounds absorbing in the 400-500 nm tend to be informative when modelled with respect to the SEEF spectra. These observations suggest that it is possible to determine the colour parameters of EVOOs through the use SEEF, as it is well proven that the main compounds responsible for colour are fluorescent chlorophyll pigments responsible for the blue/green colour and having a first absorption maxima at 450 nm.

14.5.4 Phenolic profile changes during thermal degradation

Changes in the phenolic compounds of virgin olive oils during the accelerated thermal degradation experiment were assessed through the use of reverse phase HPLC. Details of sample preparation, HPLC conditions and elution program employed are given in Section 2.2.3.1.4. Similar to what was observed during the analysis of FTIR degradation, a universal chemometric method which could predict the extent of thermal degradation could not be developed for HPLC profile data. This was mainly attributed to the large differences exhibited within the phenolic profile of the three cultivars under study, thus individual prediction models using PCR, PLSR and SVRM were built for each individual cultivar. Through the application of HPLC and phenolic compound profiles, a large number of studies have reported the discrimination of EVOOs according to their cultivar or geographical origin (Allalout *et al*., 2009). This suggests that the antioxidant content and the ability to withstand oxidation is not constant but it depends on the cultivar, fruit ripening stage, agroclimatic conditions and olive growing techniques (Tovar *et al*., 2002; Uceda and Hermoso, 2001). Therefore, depending on the minor components that are present and their ratio, it could not be expected that a model would be able to monitor and predict the changes in the phenolic profile of all the cultivars. From Figure 14.18 it was clear that the 'Bidni' cultivar exhibited a very different phenolic profile when compared to the 'Malti' and the 'Bajda' cultivar. EVOOs derived from the ''Bidni' cultivar revealed higher concentration of compounds eluting at 39.5 and 40.9min in which the latter was identified through the use of standards, belonging to oleuropein-glycoside, whilst the former was tentatively identified as a dialdehydic form of decarboxymethyl oleuropein aglycone.



3.8	hydroxytyrosol	47.6	dialdehydic form of oleuropein aglycon
6	tyrosol	51.2	dialdehydic form of decarboxymethyl oleuropein aglycone (3,4-DHPEA-EDA)
6.8	protocatechuic acid	51.8	dialdehydic form of decarboxymethyl oleuropein aglycone,
7.2	4-hydrodyphenlyacetic acid	52.1	oxidised form of dialdehydic form of decarboxymethyl ligstroside aglycone
7.6	caffeic acid	52.7	luteion
9.6	vanillic acid	53.3	dialdehydic form of decarboxymethyl ligstroside aglycone (p-HPEA-EDA)
11.5	vanillin	54.6	pinoresinol
12.4*	p-coumaric	55.2	1-acetoxypinoresinol
13.7	hydroxytyrosol acetate	55.5	trans-cinnamic acid
15.23	ferulic acid	56.8	5,7,3'-trihydroxy-4'-methoxyflavone,
18.0*	lutenion-glycoside	57.5	Apigenin
21.7*	tyrosol acetate	58.1	oxidised form of oleuropein aglycon (3,4-DHPEA-EA)
34.0*	Verbascoside	61.3	10-Hydroxy-oleuropein aglycone isomer
36.0*	salicylic acid	61.5	10-Hydroxy-oleuropein aglycone
39.5	olueropein glycoside	62.5*	oleuropein aglycon (3,4-DHPEA-EA)
41.1	Oxidised dialdehydic form of oleuropein aglycon	65.9	oxidised form of aldehydic form of ligstroside aglycone
		66.1	aldehydic form of ligstroside aglycone

Figure 14.18: Changes in the phenolic profile recorded at 280 nm (Left) and 320 nm (Right) during thermal degradation for EVOOs derived from different cultivars. (Bottom) Table of retention times and the compounds which were identified.

The presence of a 3,4-dimethoxycinnamic acid which elutes at 41.1min could not be proven as no absorbance at 320 nm was recorded within the 39-41min region. Furthermore, EVOOs derived from the 'Bidni' cultivar had a notably higher concentration of compounds eluting at 6.1, 53.3, 55.1 and 58.1min which were identified through the use of standards and relative retention times as belonging to tyrosol, dialdehydic form of decarboxymethyl ligstroside aglycone, 1-acetoxypinoresinol and oxidised form of oleuropein aglycone respectively. On the other hand, both the EVOOs derived from the 'Malti' and 'Bajda' cultivar had a notably higher concentration of compounds absorbing at the 320 nm when compared to EVOOs derived from the 'Bidni' cultivar, namely those eluting at 12.3, 15.6, 18.1, 21.7 and 62.5min. These were identified as p-coumaric acid, ferulic acid, luteolin-6-glycoside, tyrosol acetate and oleuropein aglycon. On further analysis of the chromatogram observed at 320 nm a peak eluting at 56.8 min was observed which through the use of standards it was identified as 5,7,3'-trihydroxy-4'-methoxyflavone also knows as diosmetin. The presence of this compound is most of the time neglected in the literature due to its problematic elution with other secoiridoid compounds in EVOOs making its investigation futile. The presence of this O-methylated flavone was first described by Japón-Luján *et al*., (2008), where through the use of LC-MS-MS it was found that this compound had the similar concentration to apigenin which in this experiment was identified at 57.5min.

Application of chemometric methods in order to develop methods which would enable the determination of the extent of thermal degradation of the whole phenolic profile was assessed. PCR, PLSR and linear type Kernel SVRM models were applied on the phenolic data set for each cultivar, the performance of the models obtained is summarised in Table 14.9. In general, it was shown that models obtained using PCR and PLSR had a higher performance than those obtained from SVRM when it comes to the validation. Similar to what was observed in the earlier sections of this study although SVRM models might provide another form of predictive modelling, the performance obtained by these models tends to be very high during the training however, it tends to fail when it comes to the validation data set. This indicates that the models obtained using SVRM tend to be overfitted as higher rates of error are obtained when it comes to predicting the extent of thermal degradation of samples not included in the training set. It was observed that the phenolic profiles obtained at 280

nm tend to be more informative than those obtained at 320 nm. This was mainly attributed to the fact that, for the majority of the models obtained using phenolic profiles obtained at 280 nm a lower RMSE, MAE and MBE compared to their 320 nm counterpart was obtained. Similar results were obtained when the phenolic profiles were employed for the discrimination of geographical origin, thus the results obtained from both the thermal degradation experiment and the determination of geographical origin tend to corroborate each other in that phenolic profiles obtained at 280 nm tend to be more informative. Analysis of the model performances through the different cultivars it was found that thermal degradation of the 'Bidni' cultivar was more effective using both HPLC chromatograms obtained at 280 nm and 320 nm. In the case of EVOOs derived from the 'Bajda' cultivar it was shown that the models obtained at 280 nm tend to have a lower performance when compared to those obtained at 320 nm, conversely in the case of EVOOs derived from the 'Malti' cultivar only the models obtained for HPLC chromatograms obtained at 280 nm were effective in explaining the thermal degradation in terms of changes in the phenolic profile.

Table 14.8: Performance of the PLSR, PCR and SVRM models obtained during training (Left) and testing (Right) using HPLC chromatograms observed at 280 nm as predictors of thermal degradation for EVOOs derived from different cultivars.

280 nm	Training				Validation		
	PCR	PLS	SVM		PCR	PLS	SVM
	'Bidni'						
RMSE	0.920	1.016	0.442		0.654	0.563	0.885
MBE	0.846	1.032	0.196		0.427	0.317	0.784
MAE	0.757	0.789	0.385		0.564	0.491	0.749
R^2	0.977	0.971	0.996		0.986	0.988	0.969
	'Bajda'						
RMSE	1.828	1.846	0.554		1.791	1.691	2.068
MBE	3.341	3.408	0.307		3.207	2.858	4.277
MAE	1.389	1.390	0.532		1.571	1.523	2.043
R^2	0.893	0.891	0.991		0.989	0.988	0.991
	'Malti'						
RMSE	1.244	1.288	0.575		0.524	0.587	0.875
MBE	1.547	1.660	0.331		0.275	0.345	0.765
MAE	0.968	1.040	0.566		0.483	0.483	0.776
R^2	0.953	0.950	0.996		0.999	0.999	0.965

Table 14.9: Performance of the PLSR, PCR and SVRM models obtained during training (Left) and testing (Right) using HPLC chromatograms observed at 320 nm as predictors of thermal degradation for EVOOs derived from different cultivars.

320 nm	Training				Validation		
	PCR	PLS	SVM		PCR	PLS	SVM
	‘Bidni’						
RMSE	1.516	1.491	0.580		1.725	1.400	2.352
MBE	2.297	2.224	0.337		2.974	1.961	5.532
MAE	1.223	1.381	0.576		1.701	1.321	1.944
R^2	0.929	0.932	0.993		0.811	0.858	0.644
	‘Bajda’						
RMSE	1.360	1.407	0.545		1.073	0.854	2.102
MBE	1.850	1.980	0.297		1.152	0.730	4.419
MAE	0.986	1.039	0.523		0.876	0.710	2.065
R^2	0.945	0.943	0.993		0.966	0.983	0.988
	‘Malti’						
RMSE	4.406	4.289	5.160		7.188	6.156	13.379
MBE	19.409	18.397	26.623		51.671	37.892	179.002
MAE	3.853	3.648	4.086		5.379	4.712	8.806
R^2	0.116	-0.118	0.370		-0.783	-0.878	-0.875

From the models obtained, analysis of the β-regression coefficients enabled a more detailed analysis of which compounds had the largest influence on the regression formula. A further assessment of the direction of the β-regression coefficients enables the determination whether which compounds increases or decreases during the thermal degradation process. Figure 14.19 shows the β-regression coefficients obtained for the phenolic profiles observed at 280 and 320 nm, for the EVOOs derived from the three different cultivars. It was shown that for both the chromatograms observed at 280 nm and 320 nm the regression coefficients vary corresponding to specific compounds but also vary from one cultivar to another. This indicates that the degradation of phenolic compounds is primarily affected by the originating cultivar. During the thermal degradation process, the concentration of one particular phenolic compound is in turn affect by the presence or absence of other phenolic compounds. This is indicative of a synergistic and antagonistic relationship between the different phenolic compounds. The presence of some compounds found at higher concentrations can inhibit the antioxidant activity of other compounds, but at lower concentrations, they become less prominent and the antioxidant activity of the other compounds is not hindered. The synergistic effects of phenolic compounds on the observed antioxidant have been well documented however it was never studied in conjugation with its effect on the process

EVOO degradation. It is well known that the combinations of Vitamin C, α-tocopherol and phenolic compounds exhibits a strong synergistic effect due to regeneration of the α-tocopherol (Terao *et al*., 1994; Nardini *et al*., 1997; Liao and Yin, 2000). Interactions and recycling are very common mechanisms in the action of antioxidants which can easily be affected by the thermal degradation process.

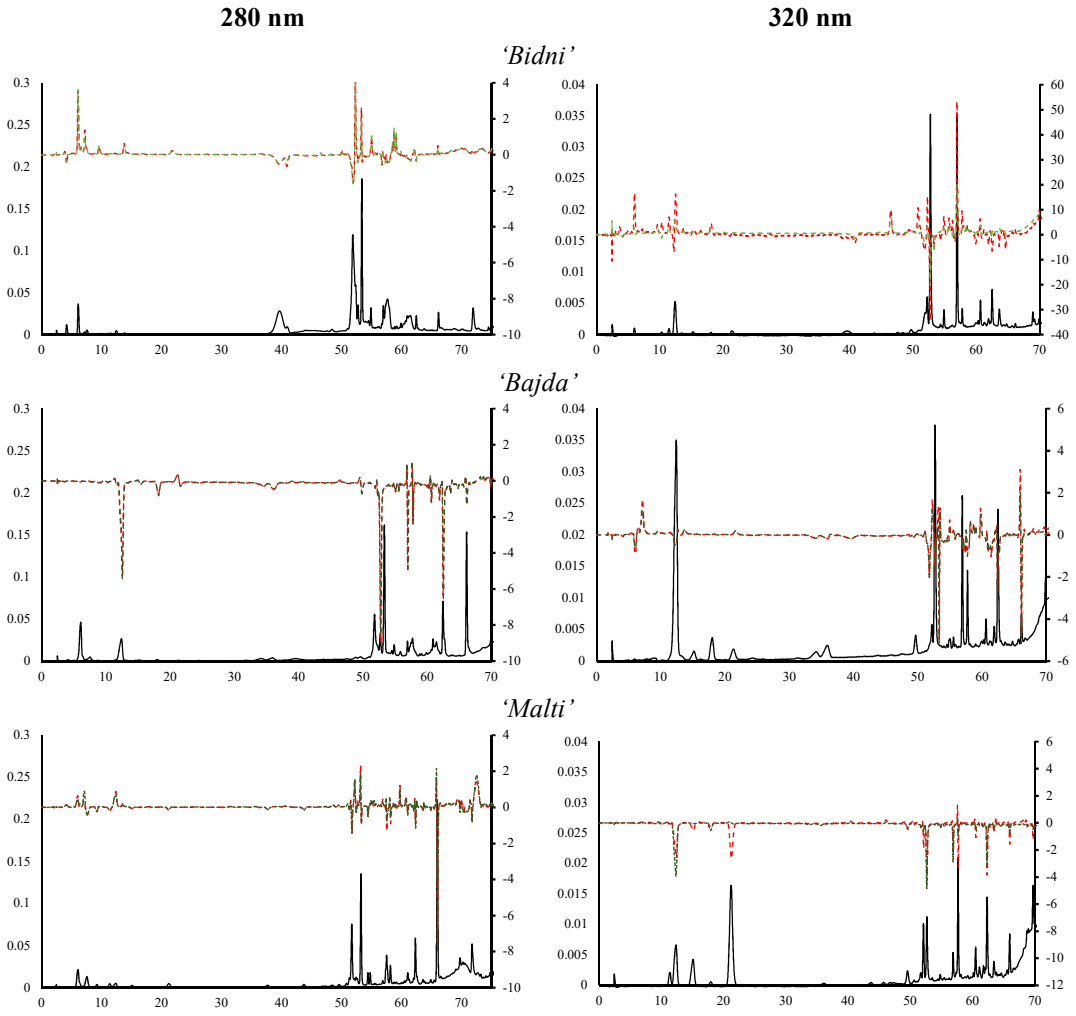


Figure 14.19: β-regression coefficients obtained from the application of PLSR (Red dotted line) PCR (Green dotted line) on the use of phenolic profile chromatogram (Black solid line) recorded at 280 nm (Left) and 320 nm (Right) for EVOOs derived from 'Bidni' (1st row), 'Bajda' (2nd row) and 'Malti' (3rd row).

The first studies concerning the changes in the phenolic profile of EVOOs during storage were carried out by Cinquanta *et al*., (1997). Through the use of HPLC it was shown that in general during storage there is a great increase in the tyrosol and hydroxytyrosol contents which was attributed to the hydrolysis of their complex derivatives in a first stage. The results obtained from the present experiment show that through the use of β-regression coefficients, in the case of the 'Bidni' cultivar there is a positive correlation with both hydroxytyrosol and tyrosol with the extent of thermal degradation, which is linked to a decrease in the dialdehydic form of decarboxymethyl ligstroside aglycone, oleuropein glycoside and dialdehydic form of decarboxymethyl oleuropein glycone. Furthermore in the case of the 'Bidni' cultivar the extent of thermal degradation was accompanied by an increase in the concentration of other phenolic acids namely p-coumaric acid, trans-cinnamic acid together with an increase in oxidized forms of secoiridoid compounds namely oleuropein aglycon and oxidized form of aldehydic form of ligstroside aglycone. Similar results but at different magnitude were observed for the 'Malti' cultivar.

In the studies carried out by Brenes *et al*., (2002) and later by Allouche *et al*., (2007) on the thermal degradation of EVOOs from different cultivars showed that during the first process of degradation there is an actual increase of the 3,4-DHPEA-EDA followed by a later decrease in Picual oils (Brenes *et al*., 2002) this increase in 3,4-DHPEA-EDA was later discovered in Arbequina (Allouche *et al*., 2007). These observations are concordant with the results obtained in this study as an overall positive β-regression coefficient corresponding to 3,4-DHPEA-EDA was observed for both the 'Bidni' and 'Malti' EVOOs. However, this could not be extended to the 'Bajda' cultivar as the concentration of 3,4-DHPEA-EDA was found to decrease with the extent of thermal degradation. As suggested by Allouche *et al*., (2007) and proposed by Brenes *et al*., (2002), the observed increase of 3,4-DHPEA-EDA might be explained by a coelution of 3,4-DHPEA-EDA oxidized. Although the results obtained are concordant to those obtained by Allouche *et al*., (2007) and Brenes *et al*., (2002), Carrasco-Pancorbo (2007) showed that the concentration of 3,4-DHPEA-EDA and 3,4- DHPEA-EA decreased more quickly with the thermal treatment than other phenolic compounds present in olive oil, suggesting a higher antioxidant power. Furthermore, it was shown that hydroxytyrosol acetate (3,4-DHPEA-AC) and p-HPEA-EA were more resistant to heat treatment as their concentration was not affected

to the extent 3,4-DHPEA-EDA and 3,4- DHPEA-EA. In this experiment similar results were obtained, analysis of the β-regression coefficients it was clear that the magnitude of the variables corresponding to hydroxytyrosol acetate p-HPEA-EA was much lower than those corresponding to 3,4-DHPEA-EDA and 3,4- DHPEA-EA. Similar to what was observed by Carrasco-Pancorbo (2007) the amount of (+)-pinoresinol and (+)-1- acetoxypinoresinol were almost unchanged during the thermal degradation process as the peaks corresponding to these had almost zero β-regression coefficient magnitude.

In the case of the 'Bajda' analysis of the β-regression coefficient revealed no apparent increase or decrease in the tyrosol and hydroxytyrosol content. The regression models obtained for the thermal degradation of the 'Bajda' cultivar were more dependent on the decrease of dialdehydic form of decarboxymethyl ligstroside aglycone and oleuropein aglycon (3,4-DHPEA-EA). The thermal degradation of the EVOOs derived from the 'Bajda' cultivar seemed to also be highly dependent on the decrease of different flavonoid compounds present within the oil. Both the profiles observed at 280 nm and 320 nm revealed a strong negative correlation with peaks corresponding to luteion, 5,7,3'-trihydroxy-4'-methoxyflavone, and apigenin.

The strong dependency of the regression models obtained on these flavonoid compounds which absorb at 320 nm could explain the lower rates of errors obtained during modelling of the phenolic profile obtained at 320 nm. This indicates that in the case of the 'Bajda' cultivar the models obtained tend to be more specific towards the concentration of flavonoid compounds. The effects of thermal degradation on the concentration of flavonoid compounds was studied by Allouche *et al*., (2007) in Arbequina oils whereby it was shown that there was a rapid decrease in the concentration of luteolin decreased during heating. Furthermore similar to the results obtained by Allouche *et al*., (2007) it was shown that degradation of luteolin was faster than that of apigenin. This was attributed to the higher structural stability than apigenin. This difference may be attributed to the structure of these compounds, since their antioxidant activity is correlated with the number of phenolic hydroxyls in the molecule (Cao *et al*., 1997). Furthermore Allouche *et al*., (2007) showed the rate of flavonoid degradation is also cultivar dependent as the loss of flavonoid compounds was higher in Arbequina oils than in Picual oils.

14.5.4.1 Predicting K-values through phenolic profiles

The potential application of phenolic profiles as a measure of oxidative stress was extended to the development of regression model for the prediction of K values. PLSR regression was carried using the previously observed K-values as the Y matrix will the corresponding phenolic profile as the X matrix. The same cross-validation methods which were previously applied to the regression models were used for this part of the experiment. Figure 14.20 shows the β-regression coefficients obtained for the different K-values. It was noted that in general in K_{266}, K_{270} and K_{274} had very similar β-regression coefficients, whereby it was identified that there was a strong positive correlation between the K-values and the content of different phenolic compounds. Compounds corresponding to tyrosol, 4-hydrodyphenlyacetic acid, oxidized form of dialdehydic form of decarboxymethyl ligstroside aglycone, p-HPEA-EDA and oxidized form of oleuropein aglycon (3,4-DHPEA-EA) all showed a strong positive correlation with the K_{270} whilst compounds such as dialdehydic form of decarboxymethyl oleuropein aglycone (3,4-DHPEA-EDA), luteolin, apigenin, oleuropein glycoside and aglycon (3,4-DHPEA-EA) showed a strong negative correlation with the increase in the aforementioned K-values. In the case of K_{230} the β-regression coefficient were similar to those observed for K_{270}, however it was found that the concentration of the oxidised form of aldehydic form of ligstroside aglycone had the largest contribution, whilst in the case of ΔK the prediction formula was mainly dependent on the concentration of luteolin and p-HPEA-EDA.

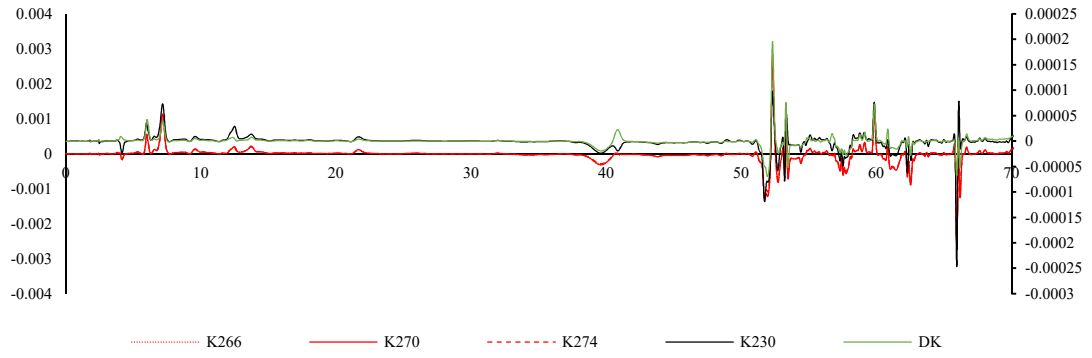


Figure 14.20: 1st factor β-regression coefficients obtained from the application of PLSR on the use of phenolic HPLC chromatograms at 280 nm as predictors of diene and triene content in thermal degraded EVOO.

Analysis of the prediction models obtained it was clear that whilst the models built using the phenolic profiles as a predictors for K-values, were able to predict the K_{266}, K_{270}, K_{274} and ΔK values with a appreciably good accuracy whilst the application of phenolic profile as a predictor of K_{230} was less suitable.

Table 14.10: Performance of the PLSR models obtained during training (Top) and testing (Bottom) using phenolic HPLC chromatograms at 280 nm as predictors of diene and triene content in thermal degraded EVOO.

Training				
	RMSE	MBE	MAE	R^2
K_{266}	0.001	0.000	0.001	0.925
K_{270}	0.001	0.000	0.001	0.939
K_{274}	0.001	0.000	0.001	0.937
K_{230}	0.000	0.000	0.000	0.747
ΔK	0.000	0.000	0.000	1.001
Testing				
	RMSE	MBE	MAE	R^2
K_{266}	0.005	0.002	0.003	0.828
K_{270}	0.004	0.002	0.003	0.824
K_{274}	0.004	0.002	0.003	0.821
K_{230}	0.000	0.000	0.000	0.755
ΔK	0.000	0.000	0.000	0.907

14.5.5 Application of atmospheric-pressure solids-analysis probe (ASAP- MS) for the determination of the extent of olive oil degradation.

The recent commercial development and availability of ambient mass spectrometry techniques has the rapid analysis of chemical residues and contaminants in food (Fussell *et al*., 2010). Furthermore there is an increasing number of ambient-MS methods reported in the literature. Chen *et al*., (2009) classified at least 30 different Ambient-MS techniques into nine groups on the basis of their desorption/volatilization method. All these techniques share the ability to ionize samples under ambient conditions outside of the mass spectrometer inlet enabling a wide range of samples to be analysed directly without any prior derivatization, separation nor extensive preparation.

McEwen *et al*. (2005) developed the atmospheric solids analysis probe (ASAP) technique on the original work by Horning in the 1970s this technique is equivalent to the use of the traditional vacuum solids probe, with the added advantage that the sample can be introduced at atmospheric pressure without the need for a vacuum lock (Fussel *et al*., 2010). In their original paper McEwen *et al*. described a qualitative detection of a range of chemicals using ASAP, including carotenoids in spinach leaves, capsaicin in pepper pods and synthetic compounds used as polymer additives. Furthermore, later on Petucci and Diffendal (2008) reported that ASAP was able to ionize both polar and non–polar small-molecule drug compounds. A number of different ASAP applications that have been reported throughout the fore coming years these included the analysis of inhibitors of the biosynthesis of ergosterol in fungi (McEwen *et al*., 2007); determination of the metabolic profile from urine and bile fluids of different animals (Twohig *et al*., 2010) , analysis of steroids (Rayand Major 2010), determination of tobacco metabolites in urine and saliva (Carrizoa *et al*., 2016) and determination of skin irritants in henna (Chen *et al*., 2016).

In the ASAP technique, samples are introduced directly into the mass spectrometer using a sealed melting-point glass capillary (probe) approximately 10 cm in length. A sample in liquid form (e.g., solvent extract) can be loaded onto the tip of the probe (using a microsyringe or dipping the probe into the liquid), or the glass probe can simply be wiped across the surface of a solid sample. The probe is inserted into the mass spectrometer and the sample is volatilized by heated nitrogen gas as illustrated in Figure 14.21. Since the technique is based on atmospheric pressure

chemical ionisation (APCI), simple mass spectra that typically show to the molecular ion $[M+H]^+$ are produced in positive mode (Cody *et al*., 2005 and Chernetsova *et al*., 2011). The proton source can be water, but methanol or other solvents can be used (Carrizo *et al*., 2015). The production of protonated molecules relies heavily on the molecular interaction/charge-transfer process occurring on the sample surface (Chen *et al*., 2016).

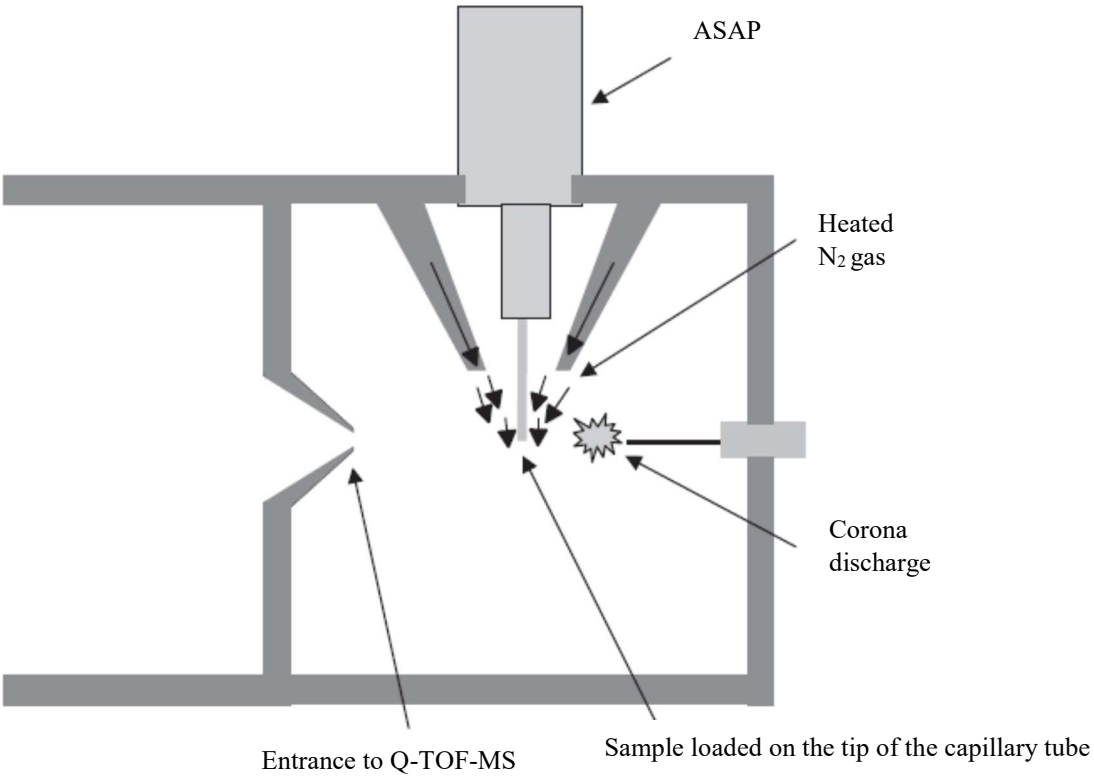


Figure 14.21: Schematic diagram of ASAP MS adopted from Twohig *et al*., (2010)

There are number of different parameters which need to be optimized for the determination of the best total number of ions and the best analyte response include, these include the corona current, sample cone voltage and gas desolvation temperature. Other important parameters which need to be taken care of include the cleanliness of the probe and its position relative to the entrance to the MS inlet. Prior each use the probe was cleaned and conditioned to ensure that all previous contaminates have been removed this was done by heating the probe at 500ºC for at least 30 s, and/ rinsed with an organic solvent.

Coupling an ASAP to a high resolution -Q-TOF enables the accurate determination of mass fragments facilitating the identification of the molecular structure of the compounds. This is an important advantage, especially when it comes to the analysis of complex matrices without any prior treatment. The identification of unknown compounds can be reached with the help of specific software (MassLynx and ChemSpider) and specific chemical databases. Whilst the application of ASAP creates an opportunity for the direct and faster analysis of samples, there are some issues which need to be taken in consideration. The first one being that although it is relatively low-cost accessory to a MS instrument currently requires manual operation, which impacts on sample throughput. Furthermore the analysis of a very small amount of analyte give to concerns relating to interpretation and robustness of data and compliance with EU legislation especially when it comes to the analysis of residues/contaminants in food.

The aim of this study is to enhance the power of ASAP method through coupling these methods to chemometeric methods. The objective is that, through the use of complex mathematical algorithms specific m/z values can be extracted and tentatively identified for a more specific study that would enable the determination of the extent of thermal degradation in EVOOs.

14.5.5.1 Application of Chemometric Analysis.

The total ion count was integrated and the resulting mass spectrum was obtained from each samples. For each sample a resultant mass spectrum was obtained through averaging the three replicate spectra. The spectra were subsequently reduced to 0.5 m/z intervals, baseline corrected and normalized prior any chemometric analysis. Two forms of chemometric methods were used in this study PLS and SVM regression analysis as described in Section 12.6. Two models were tested, one which contained all the samples derived from three different cultivars and another model was built focusing on only one cultivar at a time. Through the application of these two forms of models, an overall picture of the changes in the MS spectrum during thermal degradation could be obtained but also this enabled the identification of difference between the individual EVOOs. This was done through the inspection of the standardized β-regression coefficients extracted from the PLS regression. The magnitude of these coefficients indicates the importance of a particular m/z with

respect to the extent of degradation whilst the sign (positive or negative) give an indication of whether it increases or decrease during the degradation process. in order to improve the sensitivity of the model and increase its prediction power, certain m/z values which corresponded to the major fractions found in EVOOs were omitted in order to reveal other minor compounds which would enable a more accurate prediction of the extent of degradation.

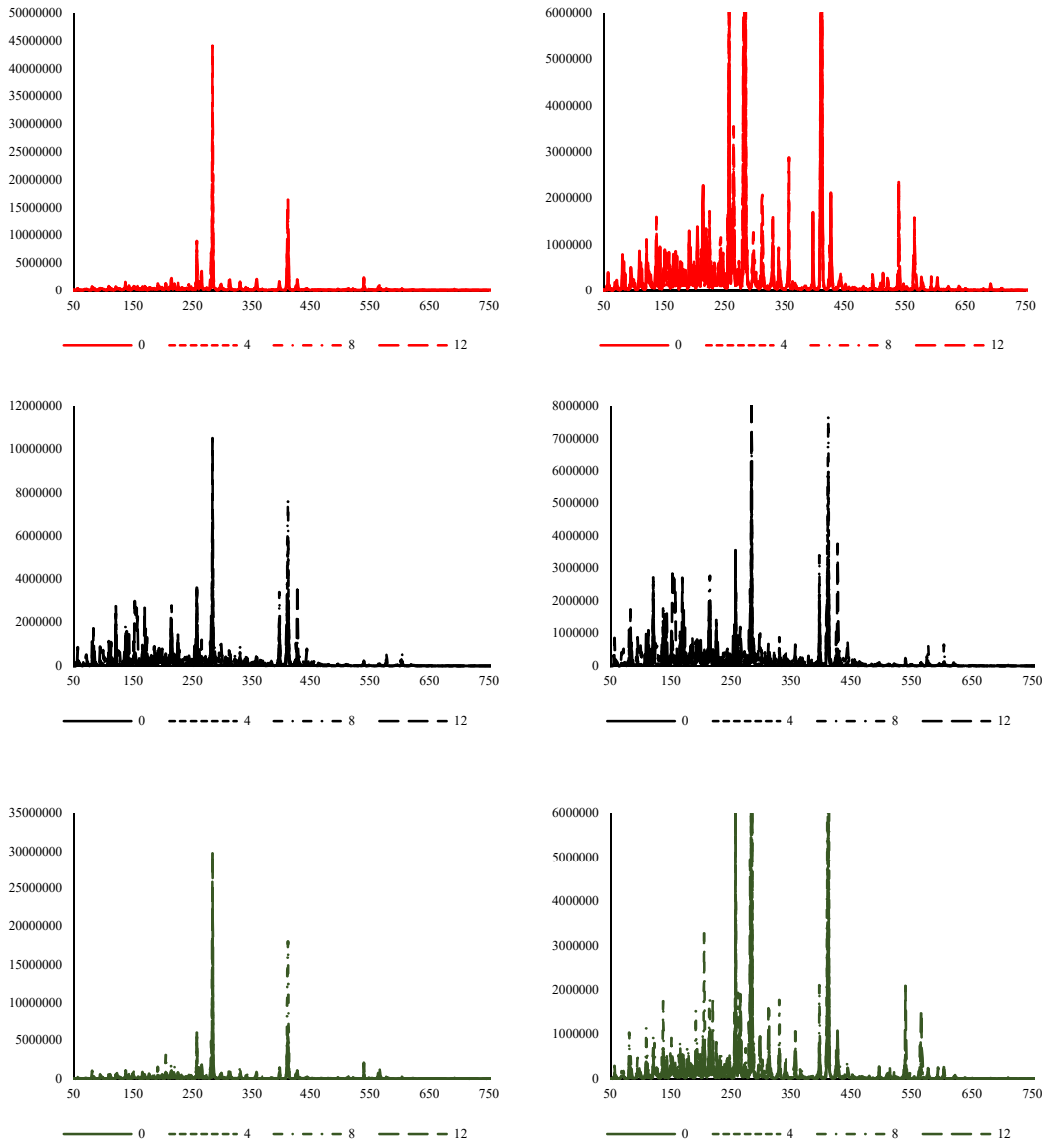


Figure 14.22: Changes in the MS chromatogram of the different cultivars during the 12-week degradation study. Top (Red) 'Bajda', Middle (Black) 'Bidni' and Bottom (Green) 'Malti' EVOO.

As shown in Figure 14.22 whilst the major peak identified was that corresponding to oleic acid, its intensity differed between the different cultivars and during the thermal degradation, suggesting that initial concentration of oleic acid is in fact different between the different cultivars but most importantly its concentration differed according to the individual cultivars during the thermal degradation process. In fact analysis of the PLSR analysis carried out using a universal model showed a marked drop in performance when all the cultivars were included in the model. This further proves that EVOOs derived from different cultivars degrades at a different rate and that different chemical markers are needed to model different EVOOs.

Table 14.11 shows the results obtained on using PLSR modelling on the ASAP-MS spectra obtained for the EVOOs derived from the individual cultivars and the universal model obtained through the combination of all the cultivars. From the results obtained it was shown that the best PLSR models were those applied to applied on the thermal degradation spectra for EVOOs derived from the 'Bajda' cultivar. The differences in model performance further highlights the fact different EVOOs degrade at different rates and the model performance is dependent on different chemical signatures found in EVOOs.

Table 14.11: Performance of the PLSR models carried out on the individual cultivar and on all the cultivars (universal) before and after the removal of the major peaks.

No variable Selection			
	RMSEC	RMSECV	PRESS
'Bajda'	1.762	2.239	2.573
'Bidni'	2.010	2.720	3.047
'Malti'	2.376	2.959	2.327
Universal	3.662	3.662	2.584
Removal of major peaks			
'Bajda'	0.886	1.812	1.888
'Bidni'	1.624	2.052	1.700
'Malti'	2.321	2.755	1.925
Universal	3.128	3.374	2.732

In order to further determine which the chemical compounds are mainly responsible for the EVOO degradation, an inspection of the standardized β-regression coefficients extracted from the PLSR model was performed. Figure 14.23 shows the standardized β-regression coefficients obtained for the individual cultivar EVOOs. In general throughout all the cultivars the peaks corresponded to the free oleic fatty acid tend to dominate the regression coefficient, positive magnitude obtained suggests that there is an increase in the amount of this particular fatty acid during the thermal degradation process. Other compounds of significant importance in terms of their β-regression magnitude are those having a m/z of 412, 427,398 and 265. These were tentatively identified as belonging to stigmasterol, oxidized form of oleoside 11-methyl ester, stigmasterol (- CH_3) and oleic acid (-H_20) respectively. In the case of 'Bidni' and 'Malti' EVOOs peak identified as the molecular ion corresponding to stigmasterol had a negative regression coefficient indicating that for these two cultivars there is a decrease in the amount of this compound, whilst the opposite was found in the case of EVOOs derived from 'Bajda'. Comparison of the regression magnitudes corresponding to the oleoside 11-methyl ester (427 m/z), it was found that this compound had a significant importance when it came to the modelling the thermal degradation of the 'Bidni' cultivar whilst in the case of the 'Malti' cultivar m/z observed at 203 and 310 were highly important in modelling the degradation of this particular EVOO. These were tentatively identified as belonging to an aldehydic form of ligstroside aglycone fragment and to palmitoleic acid monoacylglycerol-like fragment equivalent to an $[RCO+74]^+$.

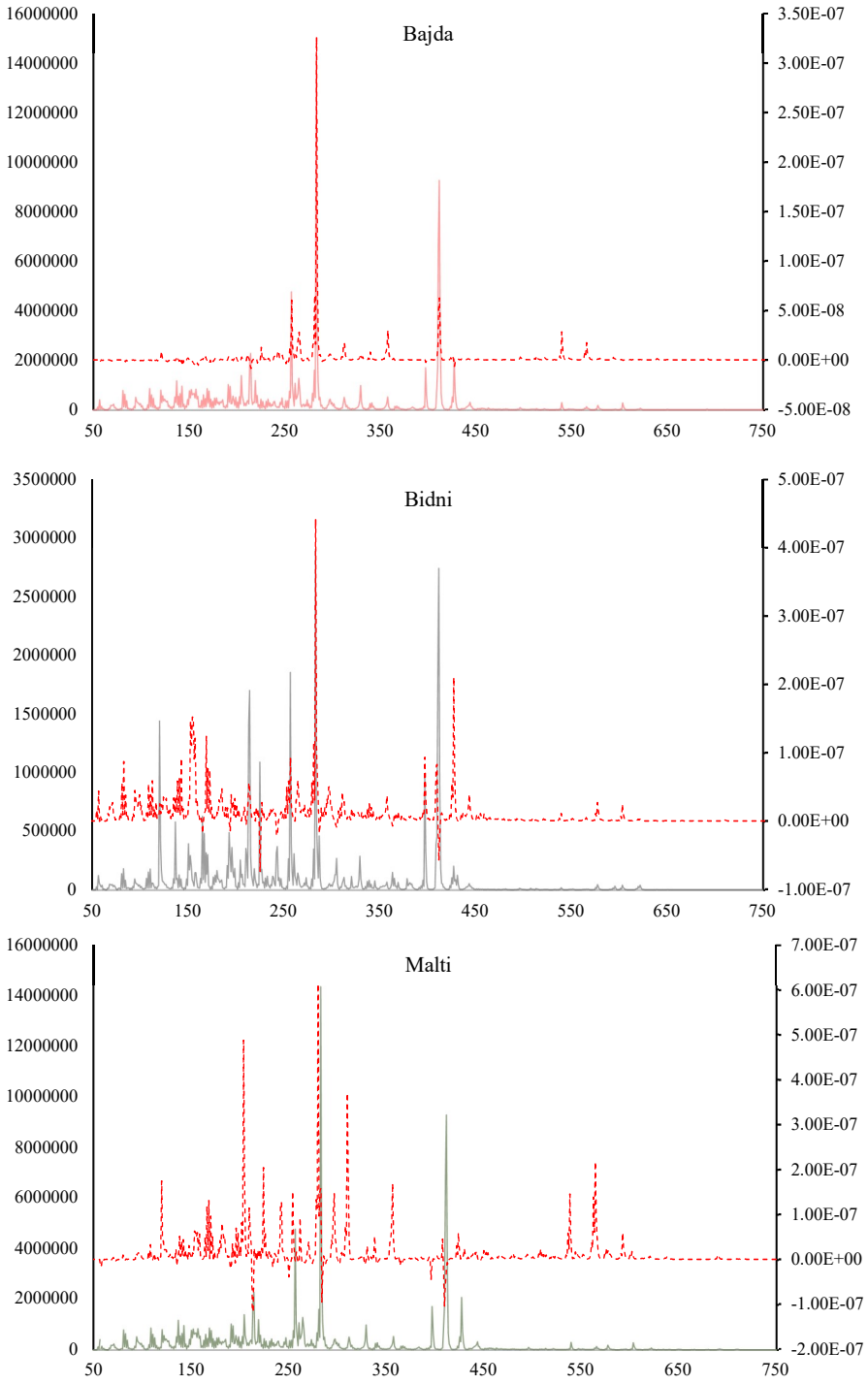


Figure 14.23: The standardised β-regression coefficients (Red dotted line) obtained from the PLSR model for the three individual cultivars, Top (Red) 'Bajda', Middle (Black) 'Bidni' and Bottom (Green) 'Malti'.

Table 14.12 lists the major m/z peaks identified through the analysis of the β-regression coefficients and their tentative identification. The arrows in the last column indicate the direction of the observed regression coefficient. In general it was observed that with the exception of palmitoleic linoleic acid, there was increase in the majority of the rest of the identified diglycerides, mainly in those containing an oleic and linoleic fatty acid in their structure, the polyunsaturated nature of these fatty acid ensures the stabilization of the radical cation formed under the MS conditions.

Although not identified during this study, the increase in the diglycerides was mainly attributed to the degradation of triglycerides. Furthermore, apart from an increase in the diglycerides content linoleic, oleic and palmitic acid monoglycerides were also found to have a positive β-regression coefficient indicative of an increase in content during the thermal degradation study. In the case of free fatty acids only three fatty acids and one methyl ester were identified, namely, palmitoleic, stearic, oleic acid (major peak) and its methyl ester. The aforementioned were found to also increase during the degradation process which is concordant with the increase in free fatty acid content as previously determined during this study.

The extent of thermal degradation was also found to affect the minor compounds present in EVOO in particular the sterols and the phenolic compounds. In the case of sterols were also shown to show some changes during the thermal degradation process, in the case of the 'Bajda' cultivar the peak obtained at 412 m/z which was tentatively identified to belonging to the molecular ion of stigmasterol had a positive β-regression coefficient indicating an increase during degradation. Gutierrez *et al*., (2000) showed that storage time and temperature influenced the percentage composition of the sterol fraction with a sharper increase in stigmasterol in the oils during storage. This was mainly attributed to the hydrolysis of sitosterol during the storage. From the analysis of the coefficients obtained it was shown that in fact the peak obtained at 414 m/z corresponding to β- Sitosterol had a negative coefficients in both the 'Malti' and 'Bidni' EVOO whilst the stigmasterol content had a positive coefficient, indicating that these two compounds could in fact be correlated to each other. In the case of the 'Bajda' cultivar the β- Sitosterol content did not seem to have any importance when it comes to the modelling its thermal degradation. In the study made by Gutierrez *et al*., (2000) also showed that concentration of sitosterol behaved

differently in the two varieties studied during storage. In the case of campesterol tentatively identified at 397 m/z, no general conclusion could be drawn on analyzing the regression coefficient of the cultivar studied as a positive coefficient was obtained for the 'Bidni' cultivar, a negative coefficient was observed for the 'Malti' cultivar whilst in the case of the 'Bajda' this did not seemed to be effective in modelling the regression.

In the case of the phenolic compounds identified, it was found that the major secoiridoid compounds namely ligstroside and oleuropein had a negative regression coefficient (345 m/z: oleuropein derivative, 363 m/z: dialdehydic form of ligstroside; 305 m/z: dialdehydic form of deacetoxy ligstroside). Indicating a decrease during the thermal degradation process this observation was consistent with the results obtained using RP-HPLC presented in the previous section. This observation is also concordant with the literature, Krichene *et al*., (2015) suggested that the ratio of simple to secoiridoid phenolics could be used as indices of the oxidative and hydrolytic degradation of VOO phenolics.

In the case of flavonoids a negative regression coefficient was obtained for both lutein and apigenin, these results further corroborate the results obtained using RP-HPLC. Conversely fragments corresponding to tyrosol, hydroxytyrosol and their acetate derivatives were found to have a positive regression coefficient, which further corroborate the results obtained using RP-HPLC. Further inspection of regression coefficients also revealed that increase in oxidized forms of secoiridoid compounds namely oleuropein aglycon (335 m/z: oxidized product of dialdehydic form of decardoxymethylated oleuropein aglycone) and oxidized form of aldehydic form of ligstroside aglycone (321 m/z oxidized product of dialdehydic form of ligstroside aglycone fragment). However similar to what was encountered during HPLC the full identification of these compounds could not be definitively done due to the postulated existence of free fatty acids masking these compounds having the same m/z. A further MS/MS study is required in order to fully determine the existence of these oxidized complex phenolic compounds in EVOOs, however this was beyond the scope of the study.

Table 14.12: Tentative identification of the major peaks obtained through the analysis of the β-regression coefficient the arrows indicate the direction of the coefficient.

m/z	Name		m/z	Name	
120	tyrosol and Hydroxytyrosol fragment	↑	351	linoleic acid monoglyceride	↑
155	Hydroxytyrosol-hydroxyphenylacetic acid,p-Anisic acid, vanillin	↑	357	Oleic acid monoglyceride	↑
165	Dialdehydic Form of Oleuropein Aglycone Decardoxymethylated fragment	↑	363	Dialdehydic form of ligstroside	↓
167	Vanillic acid	↓	397	Campesterol	↑
183	Oxidized product Dialdehydic Form of Ligstroside Aglycone fragment	↑	412	Stigmasterol-	↑
194	Ferulic acid	↓	414	β-Sitosterol	↓
196	Hydroxytyrosol acetate	↑	427	oxidized form of oleoside 11-methyl ester	↑
210	Palmitoleic acid	↑	540	Palmitic mytrsic acid $[M-H_20]^+$	↑
215	elenolic derivative fragment formed by an aldehyde and the COOH group loss.	↓	574	Palmitoleic oleic acid	↑
225	cyclic structure of elenolic aldehyde	↓	577	Palmitic oleic	↑
243	Elenolic acid	↓	595	Palmitoleic linoleic acid	↓
257	elenolic acid methylester Protonated palmitic acid derived from the thermal degradation of the esters and polyesters.	↑	601	Linoleic oleic acid	↑
270	Apigenin	↓	603	Oleic oleic acid	↑
280	oleic acid	↑			
284	stearic acid	↑			
287	luteolin	↓			
297	Oleic acid methyl ester	↑			
305	Dialdehydic form of deacetoxy ligstroside	↓			
321	Oleic acid $[M- 2H_20]^+$ Oxidized product of Dialdehydic Form of Ligstroside Aglycone fragment	↑			
329	Palmitic acid monoglyceride	↑			
335	linoleic acid monoglyceride $[M- H_20]^+$ Oxidized product of Dialdehydic Form of Decardoxymethylated Oleuropein Aglycone	↑			
345	oleuropein derivative $[M -H - CH_3CH_2OH]^+$	↓			

Following the analysis of the major peaks responsible for modelling the regression, the PLSR was repeated again without the previously identified peaks. The aim was to extract as much information from the MS chromatogram in order to try and improve the regression performance. Inspection of the RMSE obtained during the calibration and validation but also in the prediction phase quoted in Table 14.11 shows that in fact the removal of these peaks improved the performance of the PLSR model. This was attributed to the smaller number of variables which were modelled which increased the overall sensitivity of the model. A further inspection of the β-regression coefficients obtained after the removal of the major peaks indicated that minor compounds which were tentatively identified in Table 14.13 might be as effective if not more in modelling the degradation of EVOOs.

As expected the compounds identified through the β-regression coefficients belonged mainly to the minor fractions in EVOOs namely phenolic and sterol fractions, although diglycerides and free fatty acids of minor form could not excluded. Similar to what observed before, the majority of the molecular ions and fragments belonging to the aldehydic, oxidized and methylated forms of ligstroside and oleuropein had a positive regression coefficients indicating an increase during the thermal degradation process for the most of the cultivars studied.

In the case of the 'Bajda' cultivar the regression model obtained was highly dependent on 4 major m/z which were tentatively identified as belonging to elenolic acid glycoside fragment 223 m/z, desoxy elenolic acid 236 m/z, oxidized form of dialdehydic form of ligstroside aglycon 319 m/z, and a nüzhenide fragment (M-172)m/z.

In the case of the 'Bidni' cultivar the regression model was highly dependent on the increase in stigmasterol and campesterol fragments identified at 369 and 383 m/z and in the decrease at 177 m/z which could be related to either a γ-tocopherol fragment (403 m/z) or else a β-carotene fragment (537 m/z) most probably attributed to the loss of the chain of C_{16} chain. In the case of 'Bidni' the regression was also significantly dependent on the increase of Linoleic- Myristic diglyceride and dimethyl oleuropein observed at 545 m/z and 567 m/z respectively.

For the ‘Malti’ cultivar the regression model obtained was mainly attributed to an increase in compounds and fragments having a m/z between 200-250, tentatively attributed to ligstroside aglycone and elenolic acid fragments. Other compounds having a m/z at 495 and 523 belonging to myristic acid diglyceride and ligstroside respectively have also been found to have a large positive β-regression coefficient magnitude.

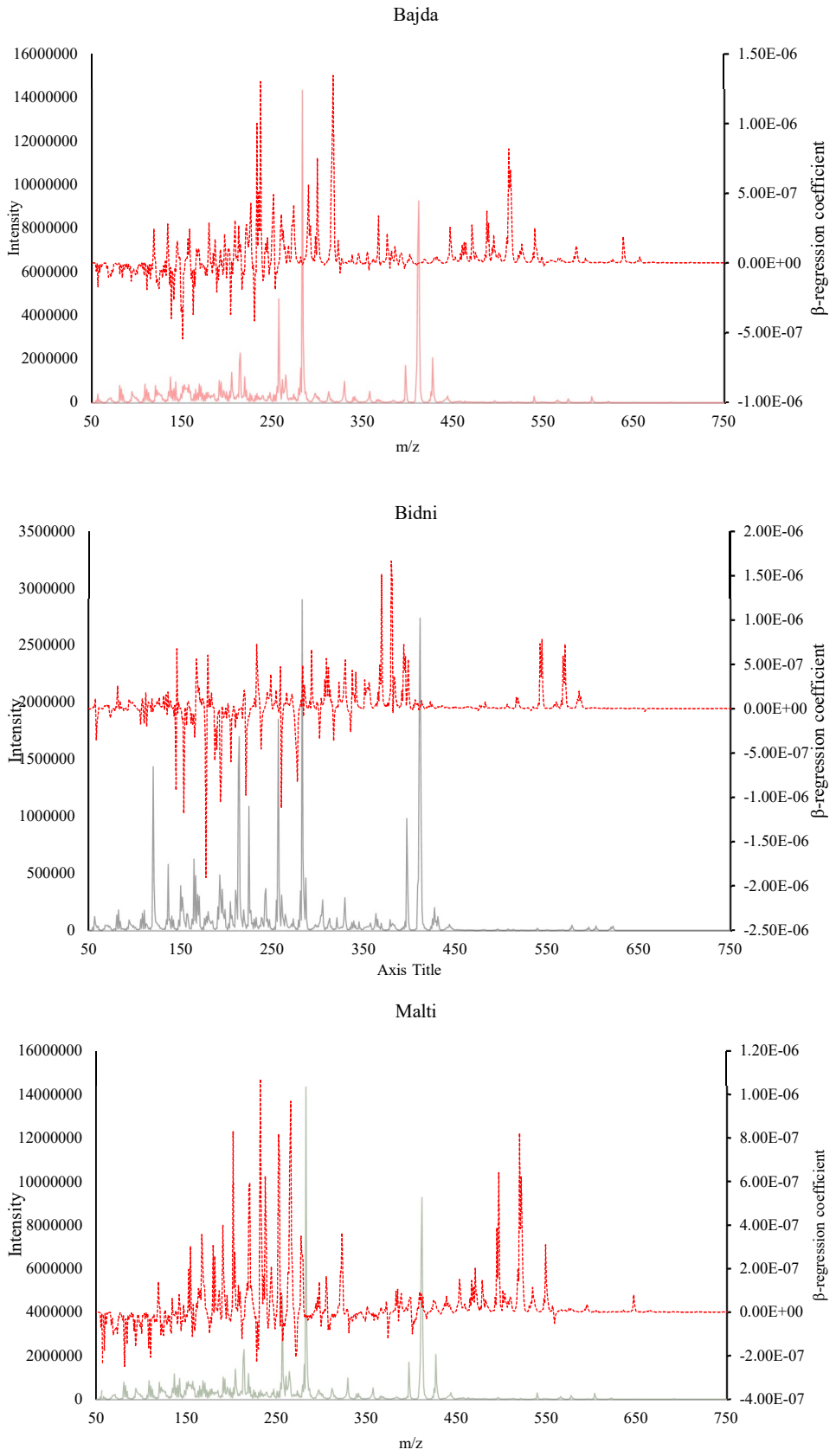


Figure 14.24: The standardised β-regression coefficients (Red dotted line) obtained from the adjusted PLSR model after the removal of the most abundant peaks for the three individual cultivars, Top (Red) 'Bajda', Middle (Black) 'Bidni' and Bottom (Green) 'Malti'.

Table 14.13: Tentative identification of the minor peaks obtained through the analysis of the β-regression coefficient the arrows indicate the direction of the coefficient.

153	dihydroxybenzoic acid or hydroxytyrosol Caffeic acid [- $2H_20$]$^+$	↑
193	Ferulic Acid- or Elenolic Acid [M-H_2O, CH_4O] $^+$	↓
203	Aldehydic Form of Ligstroside Aglycone fragment	↓
223	Elenolic Acid Glycoside- hexose [M- 179 -CO_2] $^+$	↑
236	desoxy elenolic acid	↑
259	Aldehydic Form of Oleuropein Aglycone	↓
265	oleic acid [M- H_2O] $^+$	↑
273	ligstroside glycoside [M- 180] $^+$	↓
277	linoleic acid	↓
291	ligstroside glycoside [M -162] $^+$	↑
300	ligstroside glycoside	↑
319	dialdehydic form of oleuropein aglycon oxidized form of dialdehydic form of ligstroside aglycon	↑
369	Stigmasterol fragment [M-$CH(CH_3)2$] $^+$	↑
377	oleuropein aglycon	↓
383	Campesterol [MH-H_2O] $^+$	↑
392	methyl oleuropein aglycon	↑
397	sitosterol [M-H_2O] $^+$	↓
524	ligstroside	↑
535	Myristic-Heptadecanoic methyl ligstroside isomer [M-OH] $^+$	↑
545	Linoleic- Myristic	↑
549	Myristic Oleic or Palmitoleic-Palmitic	↑
567	dimethyl oleuropein	↑
587	Heptadecanoic- Linoleic	↑
639	dimmer form dialdehydic form of oleuropein aglycon oxidized form of dialdehydic form of ligstroside aglycon dimmer	↑

Table 14.14: Performance of SVMR

	'Bajda'	'Bidni'	'Malti'	All
RMSEC	0.539	0.478	0.535	0.558
RMSEV	1.559	1.712	1.241	2.473
RMSEP	0.301	0.930	0.842	0.801

Following a PLS regression a support vector machine regression analysis SVMR was carried out first on using the combined data set containing all the variables for all the three cultivars and then repeated again for each individual cultivar. The SVM is a powerful method which can handle nonlinearities, and very good results have been reported in the literature. However, it is not so transparent as PCA and PLS and the choice of values for input parameters must be decided from cross-validation to assure a robust model. For this experiment an epsilon (0.1) type SVM algorithm was used together with a linear Kernel type function. A grid search was carried out in order to determine the best number of support vectors extracted between a C values of log10 -2 to +2 at 5 level intervals. The higher is the C parameter the more robust is the regression as errors become more important. From the results obtained it was found that SVMR analysis had a higher model performance when compared to PLSR analysis, as lower RMSE values were obtained for all the cultivars this was attributed to its ability to handle non-linear variables. Nonetheless the same trend was observed in SVRM that is modelling the individual cultivars will results in a model with a higher performance when compared to analysing all the cultivars in the same model.

In conclusion ASAP-MS process enables rapid analysis of EVOO samples although the data generated is not quantitative unlike those obtained using chromatographic methods, it enabled the analysis of a large number of samples with minimal preparation. The application of chemometric methods revealed a number of potential compounds which could be analysed in order to provide insights about the extent of EVOO degradation. Furthermore ASAP-MS further showed that the degradation of EVOOs is highly dependent on the originating cultivar although general method can be built to determine the extent of degradation the performance of this model in predicting the extent of degradation would be lower.

Conclusions

In general, it was observed that monocultivar OOs derived from the Maltese islands had sufficient quality parameters to be classified as extra virgin olive oil as defined by the IOC.

The application of simple sequence repeats screening on genomic DNA extracted from leaves of different olive cultivars showed that the DCA-3 and GAPU 101 primers were enough to distinguish between the different olive cultivars grown in Malta. Genetic analysis showed that the 'Bidni' cultivar consists of single highly homozygous population, similar to the 'Bajda' cultivar showed allelic overlap with the foreign locally grown cultivars. In comparison the 'Malti' cultivar/s consisted of a highly heterogeneous population which is distinct from both the 'Bidni' and foreign cultivar, suggesting that the existing population consists of a number of ancient cultivars which are grouped under one single cultivar.

Elemental analysis using XRF on EVOOs and seed oils revealed the presence of two major concentration related clusters one containing elements of pedological origin whilst the other consisted of heavy metals. Seed oils were found to contain a higher concentration of titanium when compared to EVOOs, whilst EVOOs derived from the Maltese islands had a significantly higher concentration of barium and phosphorus. Application of SLC-DA was able to distinguish between seed oils from EVOOs and distinguish between foreign and locally produced olive oils.

Spectrophotometric analysis of different pigments found in EVOOs showed that in general, the application of pigment concentration for the geographical discrimination of EVOOs is not advisable due to the unstable nature of these compounds. Nonetheless it was found that foreign cultivars grown in the Maltese islands had a significantly higher chlorophyll /carotenoid ratio when compared to EVOOs of the same cultivar derived from other Mediterranean countries. This unexpected results suggests that there are other factors rather than the abiotic stress conditions which are affecting the concentration of these pigments and require further study.

Extraction and quantification of different phenolic classes present in EVOOs showed that the indigenous cultivars had a significantly lower TPC and TFC when compared

to both foreign locally grown cultivars and EVOOs derived from other cultivars not grown in Malta. No significant difference was found in the TdPC between the different EVOOs.

The application of microtiter methods in order to assess the antioxidant activity and radical scavenging activity of phenolic extracts from EVOOs, revealed a similar pattern to the one observed in the analysis of the phenolic content. In fact a significantly strong positive correlation was found between the different TPC and TFC to the observed antioxidant and radical scavenging activity exhibited.

The intrinsic radical scavenging activity of EVOOs was determined via the use of newly developed method. Results obtained further confirmed that phenolic compounds present in EVOOs are the major antioxidant compounds. The development of another method for the determination of the intrinsic radical scavenging activity towards ABTS radical cations is currently under investigation.

Application of PCA using the results obtained from both the antioxidant assays and phenolic content showed that geographical discrimination between the different EVOOs is possible. This was confirmed using canonical discriminant analysis whereby it was shown that with the first canonical function local EVOOs were completely separated from the foreign EVOOs whilst the second canonical function was able to discriminate between EVOOs derived from the indigenous cultivars from EVOOs derived from locally grown foreign cultivars.

Phenolic profiling carried out using HPLC showed that the EVOOs derived from locally grown cultivars had a significantly high concentration of *p*-coumaric acid, tyrosol acetate 3, 4 DHPEA-EDA, *p*-HPEA-EDA and two unidentified phenolic compounds, with *p*-HPEA-EDA and tyrosol acetate being present in significantly higher concentrations in the indigenous cultivars.

Correlation analysis between the individual phenolic compounds present in EVOOs showed that the concentration of gallic acid is positively correlated to that of protocatechuic acid, whilst the latter was significantly correlated to the concentration of 3-Hydroxy-4-methoxycinnamic acid, these observations were explained in terms of the biosynthetic pathways that link the individual phenolic compounds together. Significant positive correlations between different secoiridoid compounds was also

observed in the study, further highlighting the degradation and synthetic pathways which link this class of highly abundant phenolic compounds.

Application of chemometric analysis on chromatographic data revealed no substantial evidence which justifies the use of data bucketing, as the models obtained from PLS analysis showed lower misclassification rates when the whole chromatographic data is taken in consideration. Application of variable selection techniques showed a significant improvement on the models obtained. Cluster analysis showed that chromatographic data obtained at 280 nm was able to fully distinguish between the different geographical origins of EVOOs.

The application of spectrofluorimetry was also extended for the development of analytical models which enable the identification and quantification of possible adulterants present in EVOOs. Adjusted PLS models based on synchronised spectra for detecting the % amount of EVOO in vegetable oil blends had a lower RMSE and higher R^2 than those observed on using PLS on the whole spectrum. This study also showed that PNN provide an alternative chemometric tool for the detection of olive oil adulteration.

The application of FT-MIR-ATR spectra in conjunction with a number of chemometeric methods, was found to provide a cheap, fast and reliable way for the discrimination of Maltese EVOOs from non-Maltese EVOOs. Due to the high level of similarity and collinearity in the data the application of unsupervised PCA models was deemed to be unsatisfactory when it comes to discrimination of geographical origin. Application of supervised methods of classification namely PLS-DA, FF-PNN, LDA and SVM showed to be highly effective in classifying and discriminating local and non-local EVOOs samples. The use of the variable selection methods significantly increased the effectiveness of PLS-DA models when compared to no variable selection. FF-PNN, SVM and LDA models were also shown to offer similar classification rates to PLS-DA models, giving further confidence in the application of FT-MIR.

The application of DI-MS fingerprinting in conjunction with multivariate statistical techniques was successfully employed for both the discrimination of EVOOs of Maltese origin from non-Maltese EVOOs and the discrimination of EVOOs depending on the country of origin. The application of DI-MS under positive ionisation

has never been achieved although it was proposed by a number of different authors in the past, including Alves *et al*., 2010; Catharino *et al*., 2005; Lerma-García *et al*., 2008 and 2011; Goodacre *et al*., 2002, however it was never carried out. From the preliminary assessment using only unsupervised PCA models, significant clustering was observed it was observed in the majority of the spectral pretreatments, however, the % variation explained by the first two principal components was very low this was attributed to the high levels of redundant variables. Inspection of the scores showed that although there was a preliminary clustering based on the geographical origin the loading plots relived that the spectral characteristics were only evident on using higher principal components. The application of SIMCA for the discrimination between Maltese and non-Maltese EVOOs showed that the models obtained under normalisation were sensitive and specific for both the classes. Application of supervised methods of classification namely PLS-DA, FF-PNN, LDA, and SVM showed to be highly effective in classifying local and non-local EVOOs samples. The discrimination of power of the different models obtained was greatly enhanced through the use of a two-stage variable selection procedure.

The application of NMR in conjunction with a number of chemometeric methods showed that similar to other method the application of unsupervised PCA models no significant clustering attributed to the high levels of similarity between the two classes of EVOOs studied, such method was deemed to be unsatisfactory when it comes to discrimination of geographical origin. Application of supervised methods of classification namely PLS-DA, FF-PNN, LDA and SVM showed to be highly effective in classifying local and non-local EVOOs samples. The use of the variable selection methods significantly increased the effectiveness of PLS-DA models when compared to no variable selection. FF-PNN, SVM and LDA models were also shown to offer similar classification rates to PLS-DA models and thus corroborate the results obtained from the PLS-DA models and put confidence in the use of NMR methods in conjunction with spectral transformation for the classification of Maltese and foreign EVOOs samples. Results showed that different NMR pulse methods can greatly affect the discrimination of EVOOs. NOESY pulse sequence and suppression of strong signal greatly improved the signal to noise ratio and the raw data obtained was more informative when compared to the conventional zg30 pulse sequence. NMR data

acquired using zg30 pulse sequence required an extensive spectral elaboration in order to obtain comparable model performance to that of NOESY.

The application of SEEF spectra in conjunction with a number of chemometeric methods, enabled the discrimination of Maltese EVOOs from non-Maltese EVOOs. From the preliminary assessment using only unsupervised PCA models, only very few spectral pretreatments for different SEEF spectra managed to identify significant clustering, whilst supervised methods showed to be highly effective in classifying local and non-local EVOOs samples. It was shown that whilst variable selection methods significantly increased the effectiveness of PLS-DA models it was detrimental for SIMCA as lower model performance was recorded for all the spectral pretreatments for all SEEF spectra. FF-PNN, SVM and LDA models were also shown to offer similar classification rates to PLS-DA models and thus corroborate the results obtained from the PLS-DA models, assuring that the use of SEEF methods in conjunction with spectral transformation enables discrimination of Maltese and foreign EVOOs samples. Results showed that different SEEF spectra can greatly affect the discrimination of EVOOs. It was shown that independent of the chemometric technique used SEEF spectra obtained at δ10 nm had higher model performance. It was shown that the most discriminate variables were those attributed to different concentration of phenolic, tocopherol and chlorophyll compounds. These observations further corroborate the results obtained from using target-specific analysis, whereby it was shown that Maltese EVOO had a significantly higher amount of different non-reducing phenolic compounds and a higher concentration of chlorophyll compounds. The quantification of the individual tocopherol compounds is being proposed as a future study, as a target-specific method for further confirmation of the results obtained using SEEF spectroscopy.

The application of 3D-fluorescence spectroscopy in conjunction with the 3-way methods has proven to be a useful tool for analyzing and interpreting this kind of complex data. Certain identification of the four detected fluorescent components was fully achieved offering a cheap, fast and reliable way for the discrimination of Maltese EVOOs from non-Maltese EVOOs. Although differing in their discriminatory power between different EVOOs, the underlying concept of a 4 component fluorophore based discrimination tends to be corroborated. The application of LDA on the 4 component PARAFAC model in mode 1 was able correctly classified 73.0% of original data and

80.28% of cross-validated grouped whilst the application of 12 latent variable DN-PLSR model was able to correctly classify 98.58% of the original data and 93.18% the cross-validated group, suggest that the latter method offered discriminatory potential for the determination of the authenticity of Maltese EVOOs.

Thermal degradation of EVOOs was studied through the use of an oven experiment. Colour analysis of the EVOOs through the 12-week thermal degradation experiment showed a significant decrease in the L* and b*parameter and a significant increase in a* chromatic component was observed indicating a transition from the green colour towards a more red colouration, together with a transition from a blue chromatic component towards a yellower colour. For the three cultivars, chlorophyll degradation showed a three-stage decay process, a linear decrease over the first week followed by a stable phase were no significant changes in chlorophyll content was observed between six to ten weeks followed by a rapid decay. Unlike what was observed for the chlorophyll content a far less linear degradation was observed in the carotenoid concentration for the 'Bidni' and 'Malti' cultivar, this was attributed to the presence of secondary oxidation products formed during the oxidation. It was shown that EVOOs derived from the 'Bajda' cultivar reached the legal limit for both the K_{230} and K_{270} faster than the other cultivars, indicating a very lower thermal stability suggesting that the higher concentration of phenolic compounds, carotenoids and chlorophyll pigments acted as an antioxidant in this experiment. Application of PCR and PLSR methods on UV-Vis spectra obtained using second-order derivatisation seem to provide the most reliable prediction models with 'Malti' cultivar modelled with a higher degree of precision.

Application of chemometeric models on the thermal degradation data obtained through the use of FTIR revealed substantial changes throughout the entire infrared spectra, especially into three major regions, 2700- 3006 cm^{-1}, 1720–1750 cm^{-1}, and the fingerprint region (1500−900 cm^{-1}). Analysis of SEEF spectra revealed a bathochromic effect e as the oil undergoes oxidation. New peaks appeared at higher wavelengths especially within the 400-600 nm region, attributed to the different chemical constituents which are developed during the course of degradation. Apart from free fatty acids and other oxidation products this peak was also attributed to the presence of polar compounds namely phenolic compounds namely tyrosol and

phenolic acid derived from the degradation of conjugated secoiridoid compounds. The latter observation was confirmed through the analysis of the individual phenolic profiles of the cultivars subjected to the thermal degradation study. Similar to what was observed during the determination of geographical origin the phenolic profiles obtained at 280 nm tend to be more informative than those obtained at 320 nm in monitoring the degradation of the 'Bidni' and the 'Malti' cultivar whilst for the 'Bajda' cultivar the chromatogram obtained at 320 nm were the most informative. During the degradation process was characterised by an increase in 3,4-DHPEA-EDA, which was attributed to the coelution of 3,4-DHPEA-EDA oxidized, an increase in the tyrosol and hydroxytyrosol attributed to the hydrolysis of their complex derivatives and overall decrease in flavonoid compounds. Application of ASAP-MS showed that the degradation was characterised by an increase in oleic acid and other free fatty acids but also other diglycerides, together with an increase of hydrogenated form of fatty acids and sterols. ASAP-MS also revealed a decrease in flavonoid compounds and in Dialdehydic form of deacetoxy ligstroside whilst an increase in hydroxytyrosol, p-hydroxyphenylacetic acid, p-anisic acid, vanillin and oxidized product of dialdehydic form of ligstroside aglycone which were in line with the result obtained during the phenolic profiling.

In this study, it was shown that it is possible to establish the uniqueness of Maltese EVOOs, and thus their authenticity for the application of PDO certification, through the use of both genetic and chemical methods. The application of genetic analysis proved to be quite difficult, time consuming and expensive. Thus, rather than providing a cost effective measure to define the authenticity of Maltese EVOOs, its value lies in providing a more specific way to identify unique cultivars grown in the Maltese islands, shed light on mislabelled cultivars, and providing ways and means to identify cultivars which are claimed to be Maltese but in reality they are not. Furthermore, it provided insights about the genetic relationship between different cultivars, obtaining historical insights on the past olive cultivation in the Maltese islands. On the other hand, the application of chemical based methods was very effective in defining the authenticity of the Maltese EVOOs, irrespective of the genetic makeup of the producing cultivars. Indeed it was possible to discriminate between, and thus authenticate, Maltese EVOOs based on the pedoclimatic conditions, which change the chemical composition of the EVOOs, providing a chemical fingerprint that

cannot be replicated elsewhere. Such a chemical fingerprint was identified through the use of a number of different methods targeted to identify markers within both the major saponifiable EVOO fraction as well as within the minor unsaponifiable fraction. These methods ranged from the use of target based methods, such as the quantification of specific phenolic classes using both microtiter based methods and liquid chromatography methods, to the use of spectroscopic data analysis including FTIR, NMR SEEF and 3D-FS. Coupling such methods to advanced multivariate statistical methods not only enabled the discrimination and thus the definition uniqueness of the Maltese EVOOs, but also enabled a direct pinpointing at which variables (chemical species) are the most informative. Although not all the methods required the same extent of manipulation, and the results regarding the sensitivity of the chemical methods to discriminate Maltese EVOOs were different, all the methods employed within this study provided a cost effective way to define the uniqueness of the Maltese EVOO. Furthermore, in some experiments such uniqueness was defined using a number of statistical methods ranging from fully unsupervised methods to supervised methods all of which further assured the results obtained. In addition, different methods used to measure the same chemical makers such as HPLC, 3D-FS and SEEF were concordant with each other in providing a more elegant way to asses such a fingerprint, as the same results could be replicated irrespective of the method employed.

In the light of the definition of such uniqueness counter measures which would predict typical fraudulent behaviour or else mismanagement of EVOOs were assured. This was indeed established through the use of a forced seed oil adulteration study simulating blending to Maltese EVOO with cheaper refined seeds, and forced thermal degradation experiment simulating mishandling of EVOOs. These experiments provided a way to identify and predict both the extent of seed oil adulteration but also the extent of thermal degradation of EVOO with the ultimate aim of discouraging any future fraudulent behaviour. The effects of these experiments on the previously identified chemical markers which defined the Maltese EVOO should be assessed in future work.

Further Work

The application of UPLC MS/MS will be applied in the future through the previously developed HPLC method which will be geometrically transferred for a UPLC. The use of selection ion reactions will be used in order to develop a mass spec method which enable monitoring the content of selective phenolic compounds present in EVOOs. This method aims to identify the peaks which were not previously identified from HPLC due to unavailability of standards and identify possible novel compounds and intermediates. The potential application of this study has been investigated however it was not published as it requires further work, nonetheless a number of compounds some of which are novel have been identified through the use of MS/MS can be found in Appendix 18 Section 18.5.

Forced inter/intra-regional adulteration of EVOOs: mixtures of EVOOs derived from the same cultivar grown in Malta and other Mediterranean countries, furthermore, mixture of EVOOs derived from indigenous cultivars and common non-indigenous once will be prepared and analysed using NMR, spectrofluorimetry, FTIR, and DI-MS with overall aim to develop models and methods capable of determining the concentration of the different EVOOs present within the mixture. This proposed study has been conducted preliminary in an undergraduate thesis published in the chemistry department in 2017 by Ilenia Gatt, however the application of DI-MS or ASAP-MS still needs to be carried out.

Application of SSR markers in the traceability of EVOOs: the first part involves the extraction of DNA from the oil samples using the method developed by Busconi *et al*., (2003) and specialised DNA kits. The extracted DNA will be subjected to qualitative analysis using gel electrophoresis in order to assess DNA degradation. The DNA obtained will be then subjected to the previously identified SSR amplification. The conservation of alleles from the trees to the oil will be assessed as a potential biomolecular marker for olive oil traceability.

DNA sequencing: Full genomic next generation HiSeq 2500 v4 DNA sequencing using shotgun libraries will carried out in the future, in order to fully characterize the indigenous cultivars genome and compare the genome sequences to other databases in order to determine the possible origin of these cultivars.

References

1 Abdel-Hameed, E.S. (2009). Total phenolic contents and free radical scavenging activity of certain Egyptian Ficus species leaf samples. *Food Chemistry*, 114: 1271-1277

2 Adam-Blondon, A.F., M. Sevignac, D. Dron, and H. BPNNerot. (1994). A genetic map of common bean to localize specific resistance genes against anthracnose. *Genome* 37:915-924.

3 Agarwal, M., Shrivastava, N., Padh, H. (2008) Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports*, 27: 617–631.

4 Aghabarati, A., Hosseini, S.M., Esmaeili, A., Maralian, H. (2008). Growth and mineral accumulation in *Olea europaea* L. trees irrigated with municipal effluent. *Research Journal of Environmental Sciences* 2, (4), 281.

5 Aguado, D., Montoya, T., Borras, L., Seco, A., Ferrer, J. (2008). Using SOM and PCA for analysing and interpreting data from a P-removal SBR. *Engineering Applications of Artificial Intelligence*. 21:919–930

6 Aguilera, C.M., Mesa, M.D., Ramirez-Tortosa, M.C., Nestares, M.T., Ros, E. Gil, A. (2004). Sunflower oil does not protect against LDL oxidation as virgin olive oil does in patients with peripheral vascular disease. *Clinical Nutrition*. 23(4), 673-681.

7 Aiamla-or, S., Kaewsuksaeng, S., Shigyo, M., Yamauchi, N., 2010. Impact of UV-B irradiation on chlorophyll degradation and chlorophyll-degrading enzyme activities in stored broccoli (*Brassica oleracea* L. Italica Group) florets. *Food Chemistry* 120, 645–651

8 Alagna, F., Kallenbach, M., Pompa, A., Marchis, F. De, Rao, R., Baldwin, I. T., Baldoni, L. (2016). Olive fruits infested with olive fly larvae respond with an ethylene burst and the emission of specific volatiles. *Food Chemistry* 58(4).

9 Alba, V., Sabetta, W., Blanco, A., Pasqualone, A., Montemurro, C. (2009). Microsatellite markers to identify specific alleles in DNA extracted from monovarietal virgin olive oils. *European Food Research and Technology* 229, 375–382.Albertini, E., Torricelli, R., Bitocchi, E., Raggi, L., Marconi, G., Pollastri, L., Di Minco, G., Battistini, A., Papa, R., Veronesi, F., (2011). Structure

of genetic diversity in *Olea europaea* L. cultivars from central Italy. *Molecular Breeding* 27, 533–547.

10 Alexa, E., Dragomirescu, A., Pop, G., Jianu, C., Dragos D. (2009) The use of FT-IR spectroscopy in the identification of vegetable oils adulteration," Journal of Food Agriculture and Environment, vol. 7, no. 2, pp. 20–24, 2009.

11 Alexieva, V., Sergiev, I., Mapelli, S., Karanov, E. (2001). The effect of drought and ultraviolet radiation on growth and stress markers in pea and wheat. *Plant Cell Environment* 24: 1337–134

12 Alipieva, K., Korkina, L., Orhan, I. E., and Georgiev, M. I. (2014). Verbascoside - A review of its occurrence, biosynthesis and pharmacological significance. *Biotechnology Advances*, 32(6), 1065–1076.

13 Alirezaei, M., Dezfoulian, O., Neamati, S., Rashidipour, M., Tanideh, N., Kheradmand, A. (2012). Oleuropein prevents ethanol induced gastric ulcers via elevation of antioxidant enzyme activities in rats. *Journal of Physiology and Biochemistry*. 68(4), 583–92.

14 Alkan, D., Tokatli, F., and Ozen, B. (2012). Phenolic characterization and geographical classification of commercial extra virgin olive oils produced in Turkey. *Journal of the American Oil Chemists' Society.* 89(2), 261–268.

15 Allalout, A., Krichène, D., Methenni, K., Taamalli, A., Oueslati, I., Daoud, D., & Zarrouk, M. (2009). Characterization of virgin olive oil from Super Intensive Spanish and Greek varieties grown in northern Tunisia. Scientia Horticulturae, 120(1), 77–83.

16 Allouche, Y., Jimenez, A., Gaforio, J.J., Uceda, M., Beltran, G. 2007. How Heating Affects Extra Virgin Olive Oil Quality Indexes and Chemical Composition. Journal of Agricultural and Food Chemistry, 55:9646–9654.

17 Almonor, G. O., Fenner, G. P., Wilson, R. F. (1998). Temperature effects on tocopherol composition in soybeans with genetically improved oil quality. *Journal of the American Oil Chemists' Society* 75, 591−596

18 Alonso-Salces, R. M., Héberger, K., Holland, M. V., Moreno-Rojas, J. M., Mariani, C., Bellan, G., Guillou, C. (2010). Multivariate analysis of NMR fingerprint of the unsaponifiable fraction of virgin olive oils for authentication purposes. *Food Chemistry*, *118*(4), 956–965. https://doi.org/10.1016/j.foodchem.2008.09.061

19 Alonso-Salces, R. M., Moreno-Rojas, J. M., Holland, M. V., Reniero, F., Guillou, C., and Heberger, K. (2010). Virgin Olive Oil Authentication by Multivariate Analyses of ^{1}H NMR Fingerprints and ^{13}C and ^{2}H Data. *Journal of Agricultural and Food Chemistry*. 58(9), 5586-5596.

20 Alós, E., Cercós, M., Rodrigo, M.J., Zacarías, L., Talón, M. (2006). Regulation of Color Break in Citrus Fruits . Changes in Pigment Profiling and Gene Expression Induced by Gibberellins and Nitrate, Two Ripening Retardants, *Journal of Agricultural Food Chemistry* 4888–4895.

21 Al-Ruqaie, I., Al-Khalifah, N. S., and Shanavaskhan, a. E. (2015). Morphological cladistic analysis of eight popular Olive (*Olea europaea* L.) cultivars grown in Saudi Arabia using Numerical Taxonomic System for personal computer to detect phyletic relationship and their proximate fruit composition. *Saudi Journal of Biological Sciences.* 23(1), 115–121.

22 Alves, J.D., Neto, W.B., Mitsutake, H., Alves, P.S.P., Augusti, R. (2010). Extra virgin (EV) and ordinary (ON) olive oils: distinction and detection of adulteration (EV with ON) as determined by direct infusion electrospray ionization mass spectrometry and chemometric approaches. *Rapid Communication Mass Spectrometery*. 24 1875–1880.

23 Amane, M., Lumaret, R., Hany, V., Ouazzani, N., Debain, C., Vivier, G., Deguilloux, M.F. (1999). Chloroplast-DNA variation in cultivated and wild olive (*Olea europaea* L.). *Theoretical and Applied Genetics.* 99(1-2) 133-139.

24 Amane, M., Ouazzani, N., Lumaret, R., Debain, C. (2000). Chloroplast DNA variation in the wild and cultivated olives (*Olea europaea* L.) of Morocco. *Euphytica* 116, 59-64

25 Amiot, M.J., Fleurietm, A., Macheix, J.J. (1989) Accumulation of oleuropein derivatives during maturation. *Phytochemistry* 28:67–9.

26 Andersen, C. M., Bro, R. (2003). Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. *Journal of Chemometrics* 17, (4), 200-215.

27 Andrewes P., J. L. H. C. Busch, T. De Joode, A. Groenewegen, H. Alexandre, (2003) Sensory Proper- ties of Virgin Olive Oil Polyphenols : Identification of Deacetoxy-ligstroside Aglycon as a Key Contributor to Pungency, *Journal of Agricultural and Food Chemistry*, 51, 1415-1420,.

28 Andrewes, P., Busch, J. L. H. C., de Joode, T. (2003). Sensory properties of virgin olive oil polyphenols: identification of deacetoxy-ligstroside aglycon as a key contributor to pungency, *Journal of Agricultural and Food Chemistry* 51, 1415–20.

29 Angerosa, F., and L. Di Giacinto. (1995). Quality characteristics of virgin olive oil in relation to crushing method. *Note IIRivista Italiana delle Sostanze Grasse*72:1–4

30 Angerosa, F., d'Alessandro, N., Basti, C., Vito, R. (1998). Biogeneration of volatile compounds in virgin olive oil: their evolution in relation to malaxation time. *Journal of Agricultural and Food Chemistry* 46, 2940–2944.

31 Angerosa, F., Mostallino, R., Basti, C., Vito, R. (2000). Virgin olive oil odour notes: their relationships with volatile compounds from the lipoxygenase pathway and secoiridoid compounds. *Food Chemistry* 68 (3), 283–287.

32 Angiolillo, A., Mencuccini, M. Baldoni, L. (1999) Olive genetic diversity assessed using amplified polymorphic fragment length polymorphisms. *Theoretical and Applied Genetics*, 98, 411–421

33 Anguelova, T., Warthesen, J. (2000). Degradation of lycopene, β-carotene, and α–carotene during lipid peroxidation *Journal of Food Science* 65(1):71–5.

34 Aparicio, R. and Luna, G. (2002). Characterisation of monovarietal virgin olive oils. *European Journal of Lipid Science and Technology* 104, 614–627.

35 Aparicio-Ruiz, R., and Gandul-Rojas, B. (2014). Decoloration kinetics of chlorophylls and carotenoids in virgin olive oil by autoxidation. *Food Research International* 65, 199–206.

36 Aparicio-Ruiz, R., Mínguez-Mosquera, M. I., & Gandul-Rojas, B. (2011). Thermal degradation kinetics of lutein, β-carotene and β-cryptoxanthin in virgin olive oils. *Journal of Food Composition and Analysis*, 24(6), 811–820.

37 Aquino, F. J. T., Augusti, R., Alves, J. D. O., Diniz, M. E. R., Morais, S. a L., Alves, B. H. P. Sabino, A. a. (2014). Direct infusion electrospray ionization mass spectrometry applied to the detection of forgeries: Roasted coffees adulterated with their husks. *Microchemical Journal* 117, 127–132.

38 Aranda F., Gómez-Alonso S., Rivera Del Álamo R.M., Salvador M.D. and Fregapane G. (2004). Triglyceride, total and 2-position fatty acid composition of Cornicabra virgin olive oil: Comparison with other Spanish cultivars. *Food Chemistry* 86, 485-492,

39 Araujo, A., Rocha, L.L, Tomazela, D, M., Sawaya, A.C.H.F., Catharino, R.R, Eberlin, M.N. (2005). Electrospray ionization mass spectrometry fingerprinting of beer. *Analyst* 130: 884.

40 Arif, I. A., Baker, M. A., Khan, H. A., Ashamed, A., Al Farhan, A. H., Al Hamadan, A. A., Al Sadoon, M., Bahkali, A. H., Shobrak, M. (2010). A simple Method for DNA Extraction from Mature Date Palm Leaves: Impact of Sand Grinding and Composition of Lysis Buffer. *International Journal of Molecular Sciences* 11, 3149–57.

41 Arnan, X. López, B. C. Martínez-Vilalta, J. Estorach, M. Poyatos, R. (2012). The age of monumental olive trees (*Olea europaea*) in northeastern Spain. *Dendrochronologia* 30:1 11-14

42 Arnow, L.E. (1937). Colorimetric detennination of the components of 3,4-dihydroxyphenylalanine- tyrosine mixtures. *The Journal of Biological Chemistry* 18:531.

43 Artajo, L. S., Romero, M. P., Morello, J. R. (2006). Enrichment of refined olive oil with phenolic compounds: evaluation of their antioxidant activity and their effect on the bitter index, *Journal of Agricultural and Food Chemistry* 54, 6079–88.

44 Ayadi, M.A., Grati-Kamun, N.H. (2009). Physico- Chemical Change and Heat Stability of Extra Virgin Olive Oils Flavored by Selected Tunisian Aromatic Plants. *Food Chemistry and Toxicology*, (47): 2613– 2619.

45 Ayed R, Grati-Kamoun N, Moreau F, Rebaï A. (2009).Comparative study of microsatel- lite profiles of DNA from oil and leaves of two Tunisian olive cultivars. *European Food Research and Technology* 229 757–762.

46 Ayton, J., Mailer, R. J., & Graham, K. (2012). The effect of storage conditions on extra virgin olive oil quality. Australian Goverment.

47 Baccouri, O., Cerretani, L., Bendini, A., Caboni, M.F., Zarrouk, M., Pirrone, L., Ben ,Miled D,D. (2007). Preliminary chemical characterization of Tunisian monovarietal virgin olive oils and comparison with Sicilian ones. *European Journal of Lipid Science and Technology* 109: 1208-1217.

48 Baccouri, O., Guerfel, M., Baccouri, B., Cerretani, L., Bendini, A., Lercker, G., Zarrouk, M., Miled, D.D.B. (2008). Chemical composition and oxidative stability of Tunisian monovarietal virgin olive oils with regard to fruit ripening. *Food Chemistry*. 109 (4), 743-754.

49 Baldioli , M. , Servili , M. , Perretti , G. , Montedoro , G.F. (1996) . Antioxidant activity of tocopherols and phenolic compounds of virgin olive oil . *Journal of the American Oil Chemists' Society* 73 , 1589 – 1593

50 Baldoni, L., Cultrera, N.G., Mariotti, R., Ricciolini, C. (2009). A consensus list of microsatellite markers for olive genotyping. *Molecular Breeding* 24: 213-231.

51 Baldoni, L., Tosti, N., Ricciolini, C., Belaj, A. (2006). Genetic structure of wild and cultivated olives in the Central Mediterranean Basin. *PNNuals of Botonay*. 98: 935-942.

52 Ballus, C. A., Meinhart, A. D., de Souza Campos, F. A., Godoy, H. T. (2015). Total Phenolics of Virgin Olive Oils Highly Correlate with the Hydrogen Atom Transfer Mechanism of Antioxidant Capacity. *Journal of the American Oil Chemists' Society* 92, 843–85.

53 Banilas, G., Hatzopoulos, P. (2013). Genetics and Molecular Biology of Olives. In: Handbook of Olive Oil, Analysis and Properties, R. Aparicio and J. Harwood (Eds.) Springer Science Business Media New York

54 Banilas, G., Minas, J., Gregoriou, C., Demoliou, C. (2003). Genetic diversity among accessions of an ancient olive variety of Cyprus. *Genome* 46: 370-376.

55 Barranco, D., Cimato, A., Fiorino, P., Rallo, L., Touzani, A., Castan˜ eda, C., Serafı´ n, F., Trujillo, I. (2000). World Catalogue of Olive Varieties. International Olive Oil Council, Madrid, Spain, 360 pp.

56 Barranco, D., Rallo, L. (1985). Las variedades de olivo cultivades en Espana. *Olivae* 9, 16–22.

57 Barranco, D., Rallo, L. (2000). Olive cultivars in Spain. *Hortechnology* 10 (1), 107–110.

58 Bartolini, G., Prevost, G., Messeri, C., Carignani, G., Menini, U.G. (1998). Olive Germplasm. Cultivars and World-Wide Collections. FAO, Rome

59 Bastida , S. , Sánchez-Muniz , F.J. (2001) . Thermal oxidation of olive oil, sunflower oil and a mix of both oils during forty discontinuous domestic fryings of different foods. *Food Science and Technology International* 7 , 15 – 21 .

60 Bates, M.H. (1988). Land farming of reserve pit fluids and sludges: Fates of selected contaminants. *Water Res* 22:793-797.

61 Bautista, R., Crespillo, R., Canovas, F.M. (2003). Identification of olive-tree cultivars with SCAR markers. *Euphytica*, Vol.129, pp. 33-41.

62 Beauchamp, G. K., Keast, R. S. J., Morel, D. (2005). Ibuprofen- like activity in extra-virgin olive oil, *Nature*, 437, 45–6

63 Beckman, C,H. (2000) Phenolic-storing cells: keys to programmed cell death and periderm formation in wilt disease resistance and in general defence responses in plants. *Physiological and Molecular Plant Pathology* 57:101–110

64 Belaj, A., Caballero, J.M., Barranco, D. (2003). Genetic characterization and identification of new accessions from Syria in an olive germplasm bank by means of RAPD markers. *Euphytica* 134:261-268.

65 Belaj, A., Muñoz-Diez, C., Baldoni, L., Satovic, Z. (2010). Genetic diversity and relationships of wild and cultivated olives at regional level in Spain. *Scientia Horticulturae* 124: 323-330.

66 Belaj, A., Ojeda, M.A., Muñoz, C., Rodríguez, E., Díaz, A., Rallo, P., De la Rosa, R., Barranco, D., Rallo, L., Trujillo, I. (2006). The use of molecular markers to characterize olive (*Olea europaea* L.) germplasm and its wild relatives. OliveBioTeq Conference. pp.391-396.

67 Belaj, A., Satovic, Z., Rallo, L. Trujillo I. (2004). Optimal use of RAPD markers for varietal identification in olive (*Olea europaea* L.) germplasm collections. *Journal of the American Society for Horticultural Science* 129(2) 266-270.

68 Belaj, A., Trujillo, L., Rallo, L. (2002). RAPD's analysis supports the autochthone origin of olive cultivars. *Acta Horticulturae* 586 83-86.

69 Beltran G., C. Del Rio, S., Sanchez. (2004). Seasonal Changes in Olive Oil Fruit Characteristics and Oil Accumulation During Ripening Process, *Journal of the Science of Food and Agriculture* 84, 1783-1790,.

70 Beltran, G., Aguilera, M. P., Del Rio, C., Sanchez, S., and Martinez, L. (2005). Influence of fruit ripening process on the natural antioxidant content of Hojiblanca virgin olive oils. *Food Chemistry*,89, 207–215.

71 Bendini, A., Cerretani, L., Di Virgilio, F., Belloni, P., Bonoli-Carbognin, M., and Lercker, G. (2007). Preliminary evaluation of the application of the FTIR spectroscopy to control the geographic origin and quality of virgin olive oils. *Journal of Food Quality* 30, 424–437.

72 Bendini, A., Cerretani, L., Salvador, M. D., Fregapane, G., & Lercker, G. (2010). Stability of the Sensory Quality of Virgin Olive Oil during Storage - An Overview. Italian Food & Beverage Technology, LX, 5–18.

73 Benedet, J. A. and Shibamoto T. (2008). Role of Transition Metals, Fe(II), Cr(II), Pb(II), and Cd(II) in Lipid Peroxidation. *Food Chemistry* 107,165-168.

74 Benincasa, C., Lewis, J., Perri, E., Sindona, G., and Tagarelli, A. (2007). Determination of trace element in Italian virgin olive oils and their characterization according to geographical origin by statistical analysis. *Analytica Chimica Acta*, 585, 366–370.

75 Benito, M., Lasa, J. M., Gracia, P., Oria, R., Sanchez-Gimeno, A. C (2010) Olive oil quality along ripening in super- intensive Arbequina orchards. *European Food Research and Technology*

76 Benjelloun B, Talou T, Delmas M, Gaset A. (1991). Oxidation of rapeseed oil: effect of metal traces. *Journal of the American Oil Chemists' Society* 68:210-211.

77 Berenguer, M.J., Vossen, P.M., Grattan, S.R., Connel, J.H., Polito., V.S. (2006).Tree irrigation levels for optimum chemical and sensory properties of olive oil. *HortScience* 41:427–432.

78 Besnard, G., Baradat, P., Bervillé, A. (2001). Genetic relationships in the olive (*Olea europaea* L.) reflect multilocal selection of cultivars *Theory of Applied Genetics* 102:251–255

79 Besnard, G., Baradat, P., Breton, C., Khadari, B. and Berville´,A (2001a) Olive domestication from structure of wild and cultivated populations using nuclear RAPDs and mitochondrial RFLPs. *Genetics, Selection, Evolution,* 33 251–S268.

80 Besnard, G., Baradat, P., Breton, C., Khadari, B., Bervillé, A. (2001). Olive domestication from structure of oleasters and cultivars using RAPDs and mitochondrial RFLP. *Genetics Selection Evolution*, 33, 251–S268.

81 Besnard, G., Baradat, P., Chevalier, D., Tagmount, A., Berville´, A. (2001b) Genetic differentiation in the olive complex (*Olea europaea*) revealed by RAPDs and RFLPs in the rRNA genes. *Genetic Resources and Crop Evolution,* 48, 165–182.

82 Besnard, G., Bervillé, A. (2000). Multiple origins of the Mediterranean olive (*Olea europaea* L. ssp. *europaea*) based upon mitochondrial DNA polymorphisms. Compte Rendu Académie des Sciences, *Sciences de la Vie*, 323 173–181.

83 Besnard, G., Khadari, B., Baradat, P. Bervillé, A. (2002). Combination of chloroplast and mitochondrial DNA polymorphisms to study cytoplasmic genetic

differentiation in the olive complex (*Olea europaea* L.). *Theoretical and Applied Genetics* 105, 139-144.

84 Besnard, G., Khadari, B., Navascués, M., Fernández-Mazuecos, M., El Bakkali, a, Arrigo, N, Savolainen, V. (2013). The complex history of the olive tree: from Late Quaternary diversification of Mediterranean lineages to primary domestication in the northern Levant. Proceedings. *Biological Sciences The Royal Society*, 280(1756), 2012-2833.

85 Besnard, G., Khadari, B., Villemur, P., Bervillé, A. (2000). A cytoplasmic male sterility in olive cultivars (*Olea europaea* L.): phenotypic, genetic and molecular approaches. *Theoretical and Applied Genetics*.100, 1018–1024.

86 Bianco, A., Mazzei, R. A., Melchioni, C., Romero, G., Scarpati, M. L., Soriero, A., Uccella, N. (1998). Microcomponents of olive oils III. Glucosides of 2(3,4-dihydroxy-phenyl)ethanol. *Food Chemistry* 63, 461-464.

87 Bisignano , G. , Tomaino , A. , Lo Cascio , R. , Crisafi , G. , Uccella , N. , Saija , A. , (1999). On the *in vitro* antimicrobial activity of oleuropein and hydroxy-tyrosol. *Journal of Pharmacology and Pharmacotherapeutics.* 51, 971 – 974.

88 Blekas , G. , Tsimidou , M. , Boskou , D. (1995) . Contribution of α -tocopherol to olive oil stability . *Food Chemistry* 52 , 289 – 294

89 Bodek, I., Lyman, W.J., Reehl, W.F. (1988). Environmental inorganic chemistry: Properties, processes, and estimation methods. New York, NY: Pergamon Press, 7.3.1-7.3-4,

90 Boerjan, W., Ralph, J., Baucher, M. (2003). Lignin biosynthesis. *PNNual Review of Plant Biology* 54:519–546

91 Bohnert, H.J., Nelson, D.E., Jensen, R.G. (1995). Adaptations to environmental stress. *Plant Cell* 7: 1099-1111.

92 Borg, J (1922). Cultivation And Diseases Of Fruit Trees In The Maltese Islands, Malta Government Printing Office pp. 98–113.

93 Bors, W., Heller, W., Michel, C. Saran, M. (1990). Flavonoids as Antioxidants: Determination of radical-scavenging efficiencies. *Methods Enzymology*. 186, 343-355.

94 Bortoleto, G. G., Pataca, L. C. M., and Bueno, M. I. M. S. (2005). A new application of X-ray scattering using principal component analysis – classification of vegetable oils. Analytica Chimica Acta, 539(1-2), 283–287.

95 Boskou, D. (1996). Olive oil composition, in: Olive oil chemistry and technology, Boskou, D. (Ed), pp. 52–83. AOC Press, Champaign, Illinois, USA

96 Boskou, D. (2002). Olive oil. In FD Gunstone, Vegetable oils in food technology: composition, properties, and uses (pp. 244–277). Oxford: Blackwell.

97 Boskou, D., Blekas, G., Tsimidou, M. (2006). Olive Oil Composition. In: Olive Oil: Chemistry and Technology (Editor: Boskou, D.). AOCS Press, IL. pp. 41–72.

98 Bosque-Sendra, J.M., Cuadros-Rodrí guez, L., Ruiz-Samblás, C., De La Mata, A.P. (2012). Combining chromatography and chemometrics for the characterization and authentication of fats and oils from triacylglycerol compositional data: A review. Analytica Chimica Acta. 724, 1-11.

99 Bracci, T., Sebastiani, L., Busconi, M., Fogher, C. (2009). SSR markers reveal the uniqueness of olive cultivars from the Italian region of Liguria. Scientia Horticulturae 122: 209-215.

100 Brenes M, García A, Dobarganes MC, Velasco J, Romero C (2002a) Influ- ence of thermal treatments simulating cooking processes on the polyphenol content in virgin olive oil. Journal of Agricultural and Food Chemistry 50, 5962-5967

101 Brenes M., Garcia A., Garcia P., Garrido A. (2000) Rapid and complete extraction of phenols from olive oil and determination by means of a coulometric electrode array system. Journal of Agricultural and Food Chemistry 48:5178-5183

102 Brenes, M., A. Garcıa, P. Garcıa, A. Garrido. (2001). Acid hydrolysis of secoiridoid aglycons during storage of virgin olive oil. Journal of Agricultural and Food Chemistry 49:5609–5614.

103 Breton, C., Claux, D., Metton, I., Skorski, G., Berville, A. (2004). Comparative study of methods for DNA preparation from olive oil samples to identify cultivar SSR alleles in commercial oil samples: possible forensic applications. Journal of Agricultural and Food Chemistry 52(3) 531-537.

104 Breton, C., Tersac, M., Berville´, A. (2006) Genetic diversity and gene flow between the wild olive (oleaster, Olea euro- paea L.) and the olive: several Plio-Pleistocene refuge zones in the Mediterranean basin suggested by simple sequence repeats analysis. Journal of Biogeography 33, 1916–1928

105 Breviglieri N, Battaglia E (1954) Ricerche cariologiche in Olea europaea L. Caryologia 6:271–283

106 Britton G (1995) UV/visible spectroscopy. In Britton G, Liaaen-Jensen S, Pfander H (eds), Carotenoids: spec- troscopy, vol 1B. Birkhäuser Verlag, Basel, pp 13-63

107 Britton, G., Liaaen-Jensen and Pfander, H. (eds) (2007) Carotenoids – Handbook. Compiled First Edition, Birkhauser Basel book, by A.Z. Mercadante e E.S. Egeland

108 Britz, S. J., Kremer, D. F. (2002). Warm temperatures or drought during seed maturation increase free a-tocopherol in seeds of soybean (Glycine max [L.] Merr.). Journal of Agricultural and Food Chemistry 50, 6058−6063.

109 Bro R. (1998) Multi-way analysis in the food industry, theory algorithms and applications, University of Amsterdam, (PhD Thesis).

110 Bro R. (1999). Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis, Chemometrics and Intelligent Laboratory Systems 46 133–147

111 Bro, R. (1997) PARAFAC. Tutorial and applications. Chemometrics and Intelligent Laboratory Systems 38, (2), 149-171.

112 Bronzini de Caraffa, V., J. Maury, C. Gambotti, C. Breton, A. Bervillé, and J. GiPNNettini. (2002). Mitochondrial DNA varia- tion and RAPD mark oleasters, olive and feral olive from western and eastern Mediterranean. Theoretical and Applied Genetics 104: 1209–1216.

113 Brooks, C.J.W. Steroids: sterols and bile acids. In: COFFEY, S. Rodd's chemistry of carbon compounds: a modern comprehensive treatise. Steroids. 2.ed. Elservier, 1970. v.2, part.4, 500p

114 Brufau, G., Canela, M.A., Rafecas, M. (2008) Phytosterols: physiologic and metabolic aspects related to cholesterol- lowering properties. Nutrition Research, 28(4).217-225.

115 Brune, M., Hallberg, L., Skanberg, A.B. (1991). Determination of iron-binding phe- nolic groups in foods. Journal of Food Science 56:128-131, 167.

116 Bucci, R., Magri, A. D., Magri, A. L., Marini, D., Marini, F. (2002). Chemical authentification of extra virgin olive oil varieties by supervised chemometric procedures. Journal of Agricultural and Food Chemistry 50, 413-418.

117 Buiarelli, F., Di Berardino, S., Coccioli, F., Jasionowska, R., Russo, M. V. (2004) Determination of phenolic acids in olive oil by capillary electrophoresis. Analytica Chimica Acta 94, 699– 705.

118 C. J. Mecklin and D. J. Mundfrom. A monte carlo comparison of the type I and type II error rates of tests of multivariate normality. Journal of Statistical Computation and Simulation, 75(2):93–107, 2005.

119 Cabrera-Vique, C., Bouzas, P. R., Oliveras-López, M. J. (2012). Determination of trace elements in extra virgin olive oils: A pilot study on the geographical characterisation. Food Chemistry 134(1), 434–439.

120 Calvano, C.D., De Ceglie, C., D'Accolti, L., Zambonin, C.G. (2012). MALDI-TOF mass spectrometry detection of extra-virgin olive oil adulteration with hazelnut oil by analysis of phospholipids using an ionic liquid as matrix and extraction solvent. Food Chemistry. 134(2), 1192-1198.

121 Campbell , S. , Stone , W. , Whaley , S. , Krishnan , K. (2003). Development of gamma (γ)-tocopherol as a colorectal cancer chemopreventive agent. Critical Reviews in Oncology/Hematology 47 , 249 – 259 .

122 Canjura F,L. and Schwartz S.J. (1991). Separation of chlorophyll compounds and their polar derivatives by high-performance liquid chromatography. Journal of the Science of Food and Agriculture , 39: 1102-1105.

123 Cantini, C., Cimato, A., Sani, G., 1999. Morphological evaluation of olive germplasm present in Tuscany region. Euphytica 109, 173– 181

124 Caponio, F., Alloggio, V., Gomes, T. (1999). Phenolic compounds of virgin olive oil: influence of paste preparation techniques. Food Chemistry 64, 203–209.

125 Caponio, F., Bilancia, M.T., Pasqualone, A., Sikorska, E. & Gomes, T. (2005). Influence of the exposure to light on extra virgin olive oil quality during storage. European Food Research and Technology, 221, 92–98.

126 CapPNNesi C, Palchetti I, Mascini M, Parenti A. 2000. Electrochemical sensor and biosensor for polyphenols detection in olive oils. Food Chemistry 71, 553-562.

127 Cardoso, S.M., Guyot, S., Marnet, N., Lopes-Da-Silva, J.A., Silva, M.S.A, Renard C.M.G.C, Coimbra, M.A. (2006). Identification of oleuropein oligomers in olive pulp and pomace. Journal of Agricultural and Food Chemistry 86: 1495-1502.

128 Carrasco-Pancorbo A, Cerretani L, Bendini A, Segura-Carretero A, Lercker G, Fernández-Gutiérrez A (2007) Evaluation of the influence of thermal oxidation on the phenolic composition and on the antioxidant activity of extra-virgin olive oils. Journal of Agricultural and Food Chemistry 55, 4771-4780

129 Carrasco-Pancorbo A., Cerretani L., Bendini A., A., Segura-Carretero A., Del Carlo M., Gallina-Toschi T., Lercker G., Compagnone D.,Fernández-Gutiérrez A. (2005). Evaluation of the antioxidant ca- pacity of individual phenolic compounds in virgin olive oil. Journal of Agricultural and Food Chemistry53: 8918.

130 Carrasco-Pancorbo, A., Cerretani, L., Bendini, A., Segura-Carretero, A., Gallina-Toschi, T., Fernandez-Gutierrez, A. (2005). Analytical determination of polyphenols in olive oils. Journal of separation science 28:837-858.

131 Carrasco-Pancorbo, A., Go´mez-Caravaca, A. M.,Cerretani, L., Bendini, A. (2006). A simple and rapid electro- phoretic method to characterize simple phenols, lignans, complex phenols, phenolic acids, and flavonoids in extra virgen olive oil. Journal of separation science 29, 2221–2233.

132 Carriero, F., Fontanazza, G., Cellini, F. Giorio, G. (2002). Identification of simple sequence repeats (SSRs) in olive (Olea europaea L.). Theoretical and Applied Genetics 104, 301-307

133 Carrizo, D., Nerin, I., Domeno, C., Alfaro, B & Nerin, C., (2016). Direct screening of tobacco indicators in urine and saliva by atmospheric pressure solid analysis probe coupled to quadruple-time of flight mass spectrometry (ASAP-MSQ-TOF). Journal of Pharmaceutical and Biomedical Analysis, 124, 149 – 156

134 Cartoni, G. P., Coccioli, F., Jasionowska, R., Ramirez, D. (2000). HPLC analysis of the benzoic and cinnamic acids in edible vegetable oils. Italian Journal of Food Science 12, 163–173.

135 Catharine, R. R., Cunha, I. B. S., Fogaça, A. O., Facco, E. M. P., Godoy, H. T., Daudt, C. E., Sawaya, A. C. H. F. (2006). Characterization of must and wine of six varieties of grapes by direct infusion electrospray ionization mass spectrometry. Journal of Mass Spectrometry 41(2), 185–190.

136 Catharino, R. R., Haddad, R., Cabrini, L. G., Cunha, I. B. S., Sawaya, A. C. H. F., and Eberlin, M. N. (2005). Characterization of vegetable oils by electrospray ionization mass spectrometry fingerprinting: Classification, quality, adulteration, and aging. Analytical Chemistry 77, 7429–7433.

137 Cavalli, J.F. , Fernandez , X. , Lizzani-Cuvelier , L. , Loiseau , A.M. (2003). Comparative study of different extraction techniques for the analysis of virgin olive oil aroma . Journal of Agricultural and Food Chemistry 51 , 7709 – 7716

138 Ceballos, C., Moyano, M.J., Vicario, I.M., Alba, J., Heredia, F.J. (2003) Chromatic evolution of virgin olive oils submitted to an accelerated oxidation. Journal of the American Oil Chemists' Society 80:257–262

139 Çelikkol Akçay, U., Özkan, G., Şan, B., Dolgun, O., Dağdelen, A., and Bozdoğan Konuşkan, D. (2014). Genetic stability in a predominating turkish olive cultivar, Gemlik, assessed by RAPD, microsatellite, and AFLP marker systems. Turkish Journal of Botany 38(3), 430–438.

140 Cerretani, L., Bendini, A., Del Caro, A., Piga, A., Vacca, V., Caboni, M.F., Toschi, T.G. (2006). Preliminary characterisation of virgin olive oils obtained from different cultivars in Sardinia. European Food Research Technology 222:354– 361.

141 Cerretani, L., Motilva, M. J., Romero, M. P., Bendini, A., and Lercker, G. (2008). Pigment profile and chromatic parameters of monovarietal virgin olive oils from different Italian cultivars. European Journal of Food Research and Technology 226, 1251–1258.

142 Cevoli, C., Cerretani, L., Gori, A., Caboni, M. F., Gallina Toschi, T., & Fabbri, A. (2011). Classification of Pecorino cheeses using electronic nose combined with artificial neural network and comparison with GC–MS analysis of volatile compounds. Food Chemistry, 129(3), 1315–1319.

143 Cheikhousman, R., Zude, M., Bouveresse, D. J., Rutledge, D. N., & Birlouez-Aragon, I. (2005). Fluorescence spectroscopy for monitoring deterioration of extra virgin olive oil during heating. Analytical and Bioanalytical Chemistry, 382(6), 1438–1443.

144 Chen, B. H., Liu, M. H. (1998). Relationship between chlorophyll a and β-carotene in a lipid-containing model system during illumination. Food Chemistry 63, 207-213.

145 Chen, H., Gamez, G., & Zenobi, R. (2009). What Can We Learn from Ambient Ionization Techniques. Journal of the American Society for Mass Spectrometry, 20(11), 1947–1963.

146 Chen, W., Nkosi, T. A. N., Combrinck, S., Viljoen, A. M., & Cartwright-Jones, C. (2016). Rapid analysis of the skin irritant p-phenylenediamine (PPD) in henna products using atmospheric solids analysis probe mass spectrometry. Journal of Pharmaceutical and Biomedical Analysis, 128, 119–125.

147 Chernetsova, E.S., Ovcharov, M.V., Khomyakov, Y.Y, Bochkov, P.O., Varlamov, A.V. (2011). The use of DART mass spectrometry for express confirmation of empirical formulas of heterocyclic compounds Russian. Chemistry Bulletin., 59 2014-2015

148 Cho Ruk, K., Kurukote, J., Supprung, P. and Vetayasuporn, S. (2006). Perennial plants in the phytoremediation of lead contaminated soils. Biotechnology. 5(1): 1–4

149 Choe, E., & Min, D. B. (2007). Chemistry of Deep-Fat Frying Oils. Journal of Food Science, 72(5), R77–R86.

150 Christensen, J., Miquel Becker, E., & Frederiksen, C.S. (2005). Fluorescence spectroscopy and PARAFAC in the analysis of yogurt. Chemometrics and Intelligent Laboratory Systems, 75, 201–208.

151 Christianini, N. and Shawe-Taylor, J. 2000. An Introduction to Support Vektor Machines and Other Kernel-Based Learning Methods. Cambridge University Press, Cambridge, NY,.

152 Ciasca, A. (1999) 'Phoenicia' , in S. Moscati (ed.), The Phoenicians (New York: Rizzoli): 168-84

153 Cichelli, A., & Pertesana, G. P. (2004). High-performance liquid chromatographic analysis of chlorophylls, pheophytins and carotenoids in virgin olive oils: chemometric approach to variety classification. Journal of Chromatography A, 1046 (1) 141-146.

154 Cinquanta, L., Esti, M., and La Notte, E. (1997). Evolution of phenolic compounds in virgin olive oil during storage. Journal of the American Oil Chemists' Society 74:1259–1264.

155 Cipriani, G., Marrazzo, M.T., Marconi, R., Cimato, A,. (2002). Microsatellite markers isolated in olive (Olea europaea L.) are suitable for individual fingerprinting and reveal polymorphism within ancient cultivars. Theory of Applied Genetics 104: 223-228.

156 Coco, F.L., Ceccon, L., Ciraolo, L., Novelli, V. (2003). Determination of cadmium (II) and zinc (II) in olive oilsby derivative potentiometric stripping analysis. Food Control 14, 55–59.

157 Cody, R.B., Laramee, J.A. , Durst, H..D. (2005) Versatile new ion source for the analysis of materials in open air under ambient conditions Analytical Chemistry., 77 ,2297-2302

158 Cordeiro, A.I., Sanchez-Sevilla, J.F., Alvarez-Tinaut, M.C.,Gomez-Jimenez, M.C. (2008). Genetic diversity assessment in Portugal accessions of Olea europaea by RAPD markers. Biologia Plantarum 52(4) 642-647.

159 Cunha, S.C., Amaral, J.S., Fernandes, J.O. , Oliveira , M.B.P.P. (2006). Quantification of tocopherols and tocotrienols in Portuguese olive oils using HPLC with three different detection systems . Journal of Agricultural and Food Chemistry 54 , 3351 – 3356 .

160 Curran-Everett, D., & Benos, D. J. (2007). Last Word on Perspectives "Guidelines for reporting statistics in journals published by the American Physiological Society: the sequel." Advances in Physiology Education, 31(4), 306–307. https://doi.org/10.1152/advan.00089.2007

161 Cuvelier, M.E., Richard, H., Berset, C. (1996). Antioxidative activity and phenolic composition of pilot-plant and commercial extracts of sage and rosemary. Journal of the American Oil Chemists' Society 73, 645–652.

162 D'Imperio, M., Dugo, G., Alfa, M., Mannina, L., and Segre, A. L. (2007). Statistical analysis on Sicilian olive oils. Food Chemistry, 102(3), 956–965.

163 Dabbou, S., Dabbou, S., Selvaggini, R., Urbani, S., Taticchi, A., Servili, M., and Hammami, M. (2011). Comparison of the chemical composition and the organoleptic profile of virgin olive oil from two wild and two cultivated Tunisian Olea europaea. Chemistry and Biodiversity 8(1), 189–202.

164 Dag, A., Kerem, Z., Yogev, N., Zipori, I., Lavee, S., Ben-David, E. (2011). Influence of time of harvest and maturity index on olive oil yield and quality. Scientia Horticulturae 358-366.

165 Dagdelen, A., Tümen, G., Özcan, M. M., Dündar, E. (2013). Phenolics profiles of olive fruits (Olea europaea L.) and oils from Ayvalik, Domat and Gemlik varieties at different ripening stages. Food Chemistry 136(1), 41–45.

166 Damak, N., Bouaziz, M., Ayadi, M., Sayadi, S., Damak, M. (2008). Effect of the maturation process on the phenolic fractions, fatty acids, and antioxidant activity of the Chétoui olive fruit cultivar. Journal of Agricultural and Food Chemistry 56: 1560-1566.

167 Dantas, T.N.C., Neto, A.A.D., Moura, M.C.P.A., Neto, E.L.B., Forte, K.R., Leite, R.H.L. (2003). Heavy metals extraction by microemulsions. *Water Reaserach* 37, 2709-2717

168 De Belie, N., Sivertsvik, M., De Baerdemaeker. J. (2003). Differences in chewing sounds of dry-crisp snacks by multivariate data analysis. *Journal of Sound and Vibration*. 266(3):625–643.

169 De La Rosa, R., James, C.M., Tobutt, K.R. (2002). Isolation and characterization of polymorphic microsatellites in olive (Olea europaea L.) and their transferability to other genera in the Oleaceae. *Molecular Ecology Notes* 2: 265-267.

170 De Luca, M., Terouzi, W., Ioele, G., Kzaiber, F., Oussama, A., Oliverio, F., Ragno, G. (2011). Derivative FTIR spectroscopy for cluster analysis and classification of morocco olive oils. *Food Chemistry*, 124(3),

171 De Luca, Michele Terouzi, Wafa Ioele, Giuseppina Kzaiber, Fouzia Oussama, Abdelkhalek Oliverio, Filomena Tauler, Romà Ragno. (2011). Derivative FTIR spectroscopy for cluster analysis and classification of morocco olive oils *Food Chemistry* 124 3 1113-1118

172 De Stefano, G., Piacquadio, P., Servili, M., Di Giovacchino, L., Sciancalepore, V. 1999. Effect of extraction systems on the phenolic composition of virgin olive oils. *European Journal of Lipid Science and Technology* 101:328–332

173 De Wald, M.G., Moore, G.A., Sherman, W.B. (1988). Identification of pineapple cultivars by isozyme genotypes. *Journal of the American Society for Horticultural Science* 113:935-938.

174 Del Giovine, L., Fabietti, F. (2005). Copper Chlorophyll in Olive Oils: Identification and Determination by LIF Capillary Electrophoresis. *Journal of Food Control,* 16:267–272

175 Del Rio, C., (1994). Preliminary agronomical characterization of 131cultivars introduced in the olive germplasm bank of Cordoba in in March 1987. *Acta Horticulturae* 356, 110–115

176 Demesure, B., Sodzi, N. Petit, J. R. (1995). A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* 4, 129–131.

177 Di Giovacchino, L., Costantini N., Serraiocco A., Surricchio G., Basti C. (2001). Natural antioxidants and volatilecompoundsof virgin olive oils obtained by two or three-phases centrifugal decanters. *European Journal of Lipid Science and Technology* 103:279–285.

178 Di Matteo, M., S. Spagna Musso, G. Grassog, and G. Bufalo. (1992). Caratterizzazione agronomica e merceologica in relazione al grado di maturazione delle produzioni di alcune cultivar olearie della provincia di Avellino. *Rivista della Societa Italiana di Scienza dell Alimentazione* 21:35–36

179 Diaz, A., Martin, A., Rallo, P., Barranco, D., De La Rosa, R. (2006) Self incompatibility of 'Arbequina' and 'Picual' olive assessed by SSR markers. *Journal of the American Society for Horticultural Science* 131, 250-255.

180 Dieana, S., Gessa, C., Marchetti, M., Usai, M. (1995). Phenolic acid redox properties: pH influence on iron (III) reduction by caffeic acid. *Soil Science Society of America Journal* 59:1301-1307

181 Dietrich, W. F., Weber, J. L., Nickerson, D. A., and Kwok, P-Y (1999). Identification and analysis of DNA polymorphisms. In B. Birren, E. D. Green, P. Hieter, S. Klapholz, Meyers, R. M. H. Riethman, and J. Roskams (Eds.), Genome analysis a laboratory manual New York: Cold Spring Harbor Lab Press.

182 Divya, O., & Mishra, A. K. (2007). Multivariate methods on the excitation emission matrix fluorescence spectroscopic data of diesel-kerosene mixtures: A comparative study. *Analytica Chimica Acta*, 592(1), 82–90.

183 Dolde, D., Vlahakis, C., Hazebroek, J. (1999). Tocopherols in breeding lines and effects of planting location, fatty acid composition, and temperature during development. *Journal of the American Oil Chemists' Society* 76, 349−355

184 Doleschall, F., Kemény, Z., Recseg, K., Kõvári, K. (2002). A new analytical method to monitor lipid peroxidation during bleaching. *European Journal of Lipid Science and Technology* 104

185 Doležel, J., Bartoš, J., Voglmayr, H., Greilhuber, J. (2003). Nuclear DNA content and genome size of trout and human. *Cytometry* 51: 127–128.

186 Dong, W., Zhang, Y., Zhang, B., Wang, X. (2012). Quantitative analysis of adulteration of extra virgin olive oil using Raman spectroscopy improved by Bayesian framework least squares support vector machines. *Analytical Methods*. 4, 2772-2777.

187 Douzane, M., Tamendjari, A., Abdi, A. K., Daas, M. S., Mehdid, F., Bellal, M. M. (2013). Phenolic compounds in mono-cultivar extra virgin olive oils from Algeria *Grasas y Aceites* 64 3 285-294

188 Doveri , S. , O'Sullivan , D.M. , Lee , D. (2006) . Non-concordance between genetic profiles of olive oil and fruit: a cautionary note to the use of DNA markers

for provenance testing . *Journal of Agricultural and Food Chemistry* 54 , 9221 – 9226 .

189 Doyle, J.J., Doyle, J.L., 1990. Isolation of plant DNA from fresh tissue. *Focus* 12, 13–15.

190 Dufour, E., & Riaublanc, A. (1997). Potentiality of spectroscopic methods for the characterisation of dairy products. I. Front face fluorescence study of raw, heated and homogenised milks. Lait, 77, 657–670

191 Dupuy, N., Le Dreau, Y., Ollivier, D., Artaud, J., Pinatel, C., Kister J. (2005). Origin of French Virgin Olive Oil Registered Designation of Origins Predicted by Chemometric Analysis of Synchronous Excitation-Emission Fluorescence Spectra. *Journal of Agricultural and Food Chemistry* 53(24) 9361-9368.

192 Ebrahimzadeh, M.A., Bahramian, F. (2009). Antioxidant activity of Crataegus pentagina subsp. elbursis fruits extracts used in traditional medicine in Iran. *Pakistan Journal of Biological Sciences* 12(5): 413-419.

193 Ebrahimzadeh, M.A., Pourmorad, F., Bekhradnia, A.R. (2008). Iron chelating activity screening, phenol and flavonoid content of some medicinal plants from Iran. *African Journal of Biotechnology* 32: 43-49.

194 Eisner , J. , Iverson , J.L. , Mozingo , A.K. , Firestone , D. (1965) . Gas chromatography of unsaponifiable matter. III. Identification of hydrocarbons, aliphatic alcohols, tocopherols, triterpenoid alcohols and sterols present in olive oils . *European Journal of Lipid Science and Technology*. 48 , 417 – 433.

195 Elliott AC, Woodward WA. Statistical analysis quick reference guidebook with SPSS examples. 1st ed. London: Sage Publications; 2007

196 Ellis, B.E. (1983). Production of hydroxyphenylethanol glycosides in suspension cultures of *Syringa vulgaris. Phytochemistry* 22,1941–3.

197 Endo, Y., Usuki, R., Kaneda, T. (1984). Prooxidant activities of chlorophylls and their decomposition products on the photoxidation of methyl linoleate. *Journal of the American Oil Chemists' Society* 61: 781

198 Ercisli, S., Barut, E., Ipek, A. (2009). Molecular characterization of olive cultivars using amplified fragment length polymorphism markers. *Genetics and Molecular Research* 8: 414–419.

199 Eriksson, L., Johansson, E., Kettaneh-Wold, N., Trygg, J., Wikstrom, C., and Wold, S. (2006), Multi- and Megavariate Data Analysis Basic Principals and Applications (Part I), Chapter 4, Umetrics.

200 Esti, M., Cinquanta, L., La Notte, E., Phenolic compounds in different olive varieties. *Journal of Agricultural and Food Chemistry* 1998, 46, 32– 35.

201 European Commission Regulation (EC) No 1989/2003 of 6 November 2003 amending regulation (EEC) No 2658/1991 on the characteristics of olive oil and olive pomace oil and the relevant method of analysis. *Official Journal of the European Union* L295, 13 November 2003

202 European Commission Regulation (EC) No 2568/1991 of 11 July 1991 on the characteristics of olive oil and olive pomace oil and the relevant method of analysis. *Official Journal of the European Union* L248, 5 September 1991

203 Everitt, B. (1987). Graphical Techniques for Multivariate Data. HeinemPNN Educational Books, London, United Kingdom.

204 F. Husson, J. Josse, S. Le, and J. Mazet. FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R, 2014. URL http://CRAN.R-project.org/package=FactoMineR. R package version 1.26. [p151]

205 Fabbri, A., Hormaza, J., Polito, V. (1995) Random Amplified Polymorphic DNA Analysis of Olive (*Olea europaea* L.) Cultivars. *Journal of the American Society for Horticultural Science* 120, 538–542.

206 Favati, F., Caporale, G., Bertuccioli, M. (1994) Rapid determi- nation of phenol content in extra virgin olive oil. *Grasas Aceites* 45, 68–70.

207 Fernández-Pachón, M. S. (2004). Anioxidant activity of wines and relation with their polihenolic composition. *Analytica Chimica Acta* 513, 113-118.

208 Ferruzzi, M.G., Blakeslee, J. (2007). Digestion, absorption, and cancer preventive activity of dietary chlorophyll derivatives. *Nutrition Research* 27,1–12.

209 Field A. Discovering statistics using SPSS. 3 ed. London: SAGE publications Ltd; 2009. p. 822.

210 Fontanazza, G., Patumi, M., Solinas, M., and Serraiocco, A. (1993). Influence of cultivars of the composition and quality of olive oil. *Acta Horticulturae* 356, 358–361.

211 Foote, C.S., Chang, Y.C., Denny, R.W. (1970). Chemistry of singlet oxygen X. Carotenoid quenching parallels biological protection. *Journal of the American Oil Chemists' Society* 92:5216-5218.

212 Forcadell, M. L, Lopez, M. C, Torre, M. C. (1988).Classification of virgin olive oils from different origin by discriminant analysis *Rivista Italiana Delle Sostanze Grasse* 65,213-214.

213 Forina, M., Oliveri, P., Lanteri, S., Casale, M. (2008). Class-modelling techniques, classic and new, for old and new problems. *Chemometrics and Intelligent Laboratory Systems* 93, 132–148.

214 Fragaki, G., Spyros, A., Siragakis, G., Salivaras, E., Dais, P. (2005). Detection of extra virgin olive oil adulteration with lampante olive oil and refined olive oil using nuclear magnetic resonance spectroscopy and multivariate statistical analysis. *Journal of Agricultural and Food Chemistry*. 53(8), 2810-2816.

215 Franceschi, V.R., Krokene, P., Christiansen, E., Krekling, T (2005) Anatomical and chemical defenses of conifer bark against bark beetles and other pests. *New Phytologist* 167(2):353–375

216 Francisca, G. and Isabel M. (1992). Action of chlorophylls on the stability of virgin olive oil. *Journal of the American Oil Chemists' Society* 69:866–871.

217 Franco, M. N., Galeano-Díaz, T., López, O., Fernández-Bolaños, J. G., Sánchez, J., De Miguel, C. Martín-Vertedor, D. (2014). Phenolic compounds and antioxidant capacity of virgin olive oil. *Food Chemistry* 163, 289–98.

218 Frankel, E. N. (2005). Lipid oxidation, (2nd ed.). UK: Barnes and Associates

219 Frankel, N. (1998). Lipid oxidation. Dundee, Scotland: The Oily Press Ltd.

220 Freeman, B. L., Eggett, D. L., Parker, T. L. (2010). Synergistic and antagonistic interactions of phenolic compounds found in navel oranges. *Journal of Food Science* 75(6), 570–576.

221 Fregapane, G., Lavelli, V., León, S., Kapuralin, J., Salvador, M.D. (2006). Effect of filtration on virgin olive oil stability during storage. *European Journal of Lipid Science and Technology* 108, 134-142

222 Fregapane, G., Salvador, M. D. (2013). Production of superior quality extra virgin olive oil modulating the content and profile of its minor components. *Food Research International* 54(2), 1907–1914.

223 Fuleki, T.j Francis, F.J. (1968) Quantitative methods for anthocyanins. 1. Extraction and deter- mination of total anthocyanin in cranberries. *Journal of Food Science* 33:72

224 Fuller, M.P. Griffiths, P.R. (1978) Diffuse Reflectance Measurements by Infrared Fourier Transform Spectrometry. *Analytical Chemistry* 50, 1906-1910.

225 Funk, C., Brodelius, P. E. (1992). Phenylpropanoid Metabolism in Suspension Cultures of Vanilla planifolia Andr. : IV. Induction of Vanillic Acid Formation. *Plant Physiology* 99, 256–262.

226 Fussell, R. J., Chan, D., & Sharman, M. (2010). An assessment of atmospheric-pressure solids-analysis probes for the detection of chemicals in food. TrAC - Trends in Analytical Chemistry, 29(11), 1326–1335.

227 Fussell, R.J., Chan, D., Sharman, M. (2010). An assessment of atmospheric-pressure solids-analysis probes for the detection of chemicals in food, Trends Analytical Chemistry 29 1326–1335

228 Galtier O., Dupuy N., Le Dréau Y.,Ollivier D., Pinatel C., Kister J. Artaud J. (2007). Geographical origins compositions of virgin olive oils determined by chemometric analysis of NIR spectra. *Analytica Chimica Acta* 136-144

229 Gandul-Rojas, B., Minguez-Mosquera,M. J. (1996). Chlorophyll and carotenoid compo- sition in virgin olive oils fromvarious Spanish olive varieties. *Journal of the Science of Food and Agriculture* 72,31–39.

230 Gandul-Rojas, B., Roca-L.Cepero, M., Mínguez-Mosquera, M. I. (2000). Use of chloro- phyll and carotenoid pigment composition to determine authenticity of virgin olive oil. *Journal of the American Oil Chemists Society* 77,853–858.

231 Gapor, M.T., Hazrina, A.R. (2000) Squalene in Oils and Fats, Palm Oil. 32, 36–40

232 Garcia, J.M., Gutiérrez, F., Castellano, J.M., Perdiguero, S., Morilla, A., Albi, M.A. (1996). Influence of storage temperature on fruit ripening and olive oil quality. Journal of Agricultural and Food Chemistry. 44: 264-267

233 García-González, D.L., Tena, N., Aparicio, R., (2010). Quality characterization of the new virgin olive oil var. 'Sikitita' by phenols and volatile compounds. *Journal of Agricultural and Food Chemistry* 58: 8357-8364.

234 Garratt, D.C., (1964): The quantitative analysis of Drugs. Volume 3. Chapman and Hall ltd, Japan, 456-458.

235 Gasteiger, J., Engel, T. (2003). Chemoinformatics—A Textbook. Wiley-VCH, Weinheim, Germany.

236 Gatt, I., Lia, F., Farrugia, C (2017) Determination of intra and extra cultivar blending of foreign and Maltese extra virgin olive oils using derivative FTIR, spectrofluorescence and NMR. Undergraduate thesis. University of Malta Chemistry Department.

237 Gemas, V.J.V., Almadanim, M.C., Tenreiro, R., Martins, A., Fevereiro, P. (2004). Genetic diversity in the Olive tree (*Olea europaea* L. subsp. *europaea*) cultivated in Portugal revealed by RAPD and ISSR markers. *Genetic Resource and Crop Evolution* 5,501-511.

238 Gemas, V.J.V., Rijo-Johansen, M.J., Tenreiro, R. and Fevereiro, P. (2000). Inter- and intra- varietal analysis of tree *Olea europaea* L. cultivars using RAPD technique. *Journal of Horticultural Science and Biotechnology* 75(3), 312-319.

239 Gilani , A.H. , Khan , A. , Ghayur , M.N. , (2006). Ca^{2+} antagonist and cholinergic activities explain the medicinal use of olive in gut disorders. *Nutrition Research* 26 , 277 – 283 .

240 Gill, S. C., Von Hippel, P. H. (1989). Calculation of Protein Extinction Coefficients from Amino Acid Sequence Data. *Analytical Biochemistry* 182, 319–26.

241 Gimeno, E., Castellote, A. I., Lamuela-Raventó´s, R. M., Torre,M. C., Lopez-Sabater, M. C. (2000). Rapid determination of vitamin E in vegetable oils by reversed-phase high-performance liquid chromatography. *Journal of Chromatography A* 881, 251-254.

242 Gimeno, E., Fito, M., Lamuela-Raventos, R.M., Castellote, A.I., Covas, M., Farré, M. (2002). Effect of ingestion of virgin olive oil on human low- density lipoprotein composition. E*uropean Journal of Clinical Nutrition* : 114-120.

243 Giuffrida, D., Salvo, f., Salvo, A., Pera,L., Dugo, G. (2007) Pigments composition in monovarietal virgin olive oils from various Sicilian olive varieties. *Journal of Food Chemistry* 101: 833-837.

244 Giungato, P., Aveni, M., Rana, R., Notarnicola, L. (2004).Modifications induced by extra virgin olive oil frying processes. *Industrie Alimentari*. 43, 369–375

245 Gliszczyńska-Świgło,A., Sikorska, E.,Khmelinskii, I., Sikorski M. (2007). Tocopherol Content in Edible Plant Oils. *Polish Journal of Food and Nutrition Science* 57,(4),157-161

246 Goldman M, Horev B, Saguy I. (1983). Decolorization of β-carotene in model systems simulating dehydrated foods. Mechanism and kinetic principles. Journal of Food Science 48(3):751–4.

247 Gomes, S., Martins-Lopes, P., Lima-Brito, J., Meirinhos, J., Lopes, J., Martins, A., Guedes- Pinto, H. (2008). Evidence of Clonal Variation in Olive 'Verdeal-

Transmontana' cultivar using RAPD, ISSR and SSR Markers. *Journal of Horticultural Science and Biotechnology* 83(4), 395-400.

248 Gomes, S., Martins-Lopes, P., Lopes, L., Guedes-Pinto, H. (2009). Assessing genetic diversity in *Olea europaea* L. using ISSR and SSR markers. *Plant Molecular Biology Reports* 123, 82-89.

249 Gómez-Caravaca, A.M., Cerretani, L., Bendini, A., Segura- Carretero, A., Fernández-Gutiérrez, A., Del Carlo, M., Compagnone, D., Cichelli, A. (2008). Effects of fly attack (*Bactrocera oleae*) on the phenolic profile and selected chemical parameters of olive oil. *Journal of Agricultural and Food Chemistry* 56, 4577-4583.

250 Gomez-Rico, A., Fregapane, G., Desamparados Salvador, M. (2008). Effect of cultivar and ripening on minor components in Spanish olive fruits and their corresponding virgin olive oils. *Food Research International* 41,433-440.

251 Gómez-Rico, A., Salvador, M. D., La Greca, M., and Fregapane, G. (2006). Phenolic and volatile compounds of extra virgin olive oil (*Olea* europea L. cv. Cornicabra) with regards to fruit ripening and irrigation management. *Journal of Agricultural and Food Chemistry*, 54, 7130–7136.

252 Gómez-Rico, A., Salvador, M. D., Moriana, A., Pérez, D., Olmedilla, N., Ribas, F. (2007). Influence of different irrigation strategies in a traditional Cornicabra cv. olive orchard on virgin olive oil composition and quality. *Food Chemistry*, 100, 568–578.

253 Gong , F. , Liang , Y.-Z. , Xie , P. , Chau , F.-T. (2003). Information theory applied to chromatographic fingerprint of herbal medicine for quality control. *Journal of Chromatography A* 1002 , 25 – 40

254 Gonzaga F.B., Pasquini C. (2006) A new method for determination of the oxidative stability of edible oils at frying temperatures using near infrared emission spectroscopy, Analytica Chimica Acta, 570, 129-135.

255 González-Rodríguez, J., Pérez-Juan, P., Luque de Castro, M. D. (2003). Superheated liquids for extraction of solid residues from winemaking processes. *Analytical and Bioanalytical Chemistry* 7−8, 1190−1195.

256 González-Rodríguez, J., Pérez-Juan, P., Luque de Castro, M. D. (2004). Use of superheated liquids for the extraction of non-volatile compounds from wood: Liquid chromatography studies. *Journal of Chromatography A*.

257 Goodacre , R. , Vaidyanathan , S. , Bianchi , G. , Kell , D.B. (2002). Metabolic profiling using direct infusion electrospray ionisation mass spectrometry for the characterisation of olive oils. *Analyst* 127, 1457 – 1462.

258 Gordon, A. D. (1999) Classification. Chapman and Hall, CRC, Boca Raton, FL.

259 Grasso, S., Siracusa, L., Spatafora, C., Renis, M., Tringali C. (2007). Hydroxytyrosol lipophilic analogues: enzymatic synthesis, radical scavenging activity and DNA oxidative damage protection . *Bioorganic Chemistry* 35 , 137 – 152 .

260 Grati, K.N., Ouazzani, N., Trigui, A. (2002). Characterizing isozymes of some Tunisian olive (*Olea europaea* L.) cultivars. *Acta Horticulturae* 586, 137–140.

261 Grati-Kamoun, N., Mahmoud, F.L., Rebaı, A., Gargouri, A., Panaud, O. (2006) Genetic Diversity of Tunisian Olive Tree (*Olea europaea* L.) Cultivar Assessed by AFLP Markers. *Genetic Resources and Crop Evolution* 53, 265– 275.

262 Grattan, S.R., Berenguer, M.J., Connell, J.H., Polito, V.S., Vossen, P.M. (2006). Olive oil production as influenced by different quantities of applied water. *Agricultural Water Management* 85, 133–140.

263 Graversen, H.B., Becker, E.M., Skibsted, L.H., Andersen, M.L. (2008). Antioxidant synergism between fruit juice and alpha-tocopherol. A comparison between high phenolic black chokeberry (*Aronia melanocarpa*) and high ascorbic blackcurrant (*Ribes nigrum*) *European Food Research and Technology* 226 (4), 737–743.

264 Green, P., Wickens, G. E. (2002). The *Olea europaea* complex. In The Davis and Hedge Festschrift: Plant Taxonomy, Phytogeography and Related Subjects, Tan, K., Ed., Edinburgh University Press, 989, 287–299.

265 Grigg, D. (2001). Olive oil, the Mediterranean and the world. *Geography Journal* 53:163–172.

266 Guerfel, M., Ouni, Y., Taamalli, A., Boujnah, D., Stefanoudaki, E., Zarrouk, M. (2009). Effect of location on virgin olive oils of the two main Tunisian olive cultivars. *European Journal of Lipid Science and Technology* 111, 926–932.

267 Guilbault, G. G. (1999). Practical Fluorescence, Marcel Dekker, ISBN 0824712633, New York

268 Guillen M.D., Cabo N. (2002) Fourier transform infrared spectra data versus peroxide and anisidine values to determine oxidative stability of edible oils, Food Chemistry 77, 503-510.

269 Guimet, F., Ferre ´, J., Boque ´, R. (2004a) Cluster analysis applied to the exploratory analysis of commercial Spanish olive oils by means of excitation-emission fluorescence spectroscopy. *Journal of Agricultural and Food Chemistry* 52, 6673-6679.

270 Guimet, F., Ferré, J., Boqué, R. (2005). Rapid detection of olive pomace oil adulteration in extra virgin olive oils from the protected denomination of origin Siurana using excitation– emission fluorescence spectroscopy and three-way methods for analysis. *Analytica Chimica Acta* 544, 143–152.

271 Guimet, F., Ferre, J., Boque, R., & Rius, F. X. (2004). Application of unfold principal component analysis and parallel factor analysis to the exploratory analysis of olive oils by means of excitation-emission matrix fluorescence spectroscopy. Analytica Chimica Acta, 515(1), 75–85.

272 Guimet, F., Ferre, J., Boque, R., Vidal, M., & Garcia, J. (2005). Excitation-emission fluorescence spectroscopy combined with three-way methods of analysis as a complementary technique for olive oil characterization. Journal of Agricultural and Food Chemistry, 53(24), 9319–9328.

273 Guimet, F., Ferre, J., Boque´, R., Rius, F. X. (2004b). Application of unfold principal component analysis and parallel factor analysis to the exploratory analysis of olive oils by means of excitation- emission matrix fluorescence spectroscopy. *Analytica Chimica Acta* 515,75-85.

274 Gutiérrez, F., Varona, I., & Albi, M. A. (2000). Relation of acidity and sensory quality with sterol content of olive oil from stored fruit. Journal of Agricultural and Food Chemistry, 48(4), 1106–1110.

275 Gutiérrez-Rosales F., Rios J.J., Gomez-Rey M.A.L. (2003): Main polyphenols in the bitter taste of virgin olive oil. Structural confirmation by on-line high-performance liquid chromatography electrospray ionization mass spectrometry. *Journal of Agricultural and Food Chemistry*, 51: 6021–6025.

276 Gutiérrez-Rosales, F., Garrido-Fernández, J., Gallardo-Guerrero, L., Gandul-Rojas, B., & Minguez-Mosquera, M. I. (1992). Action of chlorophylls on the stability of virgin olive oil. Journal of the American Oil Chemists Society, 69(9), 866–871.

277 Guzmán, E., Baeten, V., Pierna, J. A. F., & García-Mesa, J. a. (2015). Evaluation of the overall quality of olive oil using fluorescence spectroscopy. Food Chemistry, 173, 927–34.

278 Guzmán, E., Baeten, V., Pierna, J. A. F., & García-Mesa, J. a. (2015). Evaluation of the overall quality of olive oil using fluorescence spectroscopy. Food Chemistry, 173, 927–34.

279 Haaland, D. M., Thomas, E. V. (1988). Partial least-squares methods for spectral analysis in relation to other quantitative calibration methods and the extraction of quantitative information. *Analytical Chemistry* 60, 1193-1202.

280 Haddada, Faouzia M. Krichène, Dhouha Manai, Hédia Oueslati, Imen Daoud, Douja Zarrouk, Mokhtar. (2008). Analytical evaluation of six monovarietal virgin olive oils from Northern Tunisia *European Journal of Lipid Science and Technology* 110, 10, 905-913

281 Hakim, I. R., Kammoun, N. G., Makhloufi, E., and Rebaï, A. (2010). Discovery and potential of SNP markers in characterization of Tunisian olive germplasm. *Diversity* 2, 17–27.

282 Hamdi, H.K., Castellon, R. (2005). Oleuropein, a non-toxic olive iridoid, is an anti-tumor agent and cytoskeleton disruptor. *Biochemical and Biophysical Research Communications.* 334, 769 – 778.

283 Harold L. (1997). Squalene, olive oil, and cancer risk, A review and hypothesis. Cancer Epidemiology Biomarkers and Prevention, 6, 1101–1103.

284 Harshman, R.A., Lundy, M.E. (1994). Parafac-parallel factor-analysis, Computational Statistics & Data Analysis . 18 (1994) 39–72

285 Harwood J., Aparicio R. (2000). Handbook of Olive Oil, Analysis and Properties. Gaithersburg Maryland,Aspen Publ. Inc.

286 Hashemi, P., Delfan, B., Ghiasvand, a. R., Alborzi, M., Raeisi, F. (2010). A study of the effects of cultivation variety, collection time, and climate on the amount of oleuropein in olive leaves. *Acta Chromatographica* 22, 133–140.

287 Hatzopoulos, P., G. Banilas, K. GiPNNoulia, F. Gazis, N. Nikoloudakis, D. Milioni, and K. Haralampidis. 2002. Breeding, molecular markers and molecular biology of the olive tree. *European Journal of Lipid Science and Technology* 104,574–586.

288 Heaton, J.W and Marangoni, A.G. (1996). Chlorophyll degradation in processed foods and senescent plant tissues. Trends Food Science and Technology, 7: 8-15.

289 Heijnen, C.G., Haenen, G.R., van Acker, F.A., van der Vijgh, W.J. Bast, A. (2001). Flavonoids as peroxynitrite scavengers, the role of the hydroxyl groups. *Toxicology in vitro* 15, 3-6.

290 Hendrikse, P. W., Slikkerveer, F. J., Folkersma, A., Dieffenbacher, A. (1991). Determination of Lead in Oils and Fats by Direct Graphite Furnace Atomic Absorption Spectrometry. *Pure and Applied Chemistry* 63 (8). 1183-1190.

291 Henry, L.K., Catignani, G.L., Schwartz, S.J. (1998). The influence of carotenoids and tocopherols on the stability of safflower seed oil during heat-catalyzed oxidation. Journal of the American Oil Chemists' Society 75(10):1399–402.

292 Herchi, W., Arráez-Román, D., Boukhchina, S., Kalle, H., Segura-Carretero, A., Fernández-Gutierrez, A. (2012). A review of the methods used in the determination of flaxseed components. *African Journal of Biotechnology* 11(4), 724-731.

293 Hernández, P., Martín, A., Dorado, G. (1999). Development of SCARs by direct sequencing of RAPD products, practical tool for the introgression and marker-assisted selection of wheat. *Molecular Breeding* 5, 245-253.

294 Hildebrand, D.H., Terao, J., Kito, M., (1984). Phospholipids plus tocopherols increase soybean oil stability. *Journal of the American Oil Chemists Society* 61, 552–555.

295 HPNNachi, H. (2010). Genetic Relationships between Cultivated and Wild Olive Trees (*Olea europaea* L. Var. *Europaea* and Var. *sylvestris*) Based on Nuclear and Chloroplast SSR Markers. *Natural Resources*, 01(02), 95–103.

296 Hrncirik K, Fritsche S. (2004). Comparability and reliability of different techniques for the determination of phenolic compounds in virgin olive oil. *European Journal of Lipid Science and Technology* 106, 540-549.

297 Hrncirik, K., Fritsche, S. (2005). Relation between the endogenous antioxidant system and the quality of extra virgin olive oil under accelerated storage conditions. *Journal of Agricultural and Food Chemistry*, 53(6), 2103–2110.

298 Huang, D., Boxin, O. U., Prior, R. L. (2005). The chemistry behind antioxidant capacity assays. *Journal of Agricultural and Food Chemistry* 53(6), 1841–1856.

299 Inarejos-García, A., Gómez-Rico, A., Salvador, M. D., Fregapane, G. (2009). Influence of malaxation conditions on virgin olive oil yield, overall quality and composition. *European Food Research and Technology* 228, 671–677.

300 Iñón, F. A., Garrigues, J. M., Garrigues, S., Molina, A., & de la Guardia, M. (2003). Selection of calibration set samples in determination of olive oil acidity by partial least squares-attenuated total reflectance-Fourier transform infrared spectroscopy. Analytica Chimica Acta, 489, 59–75

301 Interesse, F. S. Ruggiero, P. Vitagliano M. (1971) Autoxidation of olive oil: influence of chlorophyll pigments. Industrial Agriculture. 9 318-323

302 Interesse, F. S., Ruggiero, P., Vitagliano, M. (1971). Autoxidation of olive oil: influence of chlorophyll pigments. *Industrial Agriculture*. 9, 318-323.

303 International Olive Council (IOC) (2011). COI/T.15/NCNo 3/Rev. 6. Trade standards applying to olive oils and olive-pomace oils. Madrid, Spain.

304 Intrieri, M.C., Muleo, R., Buiatti, M. (2007). Chloroplast DNA polymorphisms as molecular markers to identify cultivars of, *Olea europaea* L. *The Journal of Horticultural Science and Biotechnology* 82, 109–113.

305 Ismail, M., Mariod, A., Bagalkotkar, G., Sy Ling, H. (2009). Fatty acid composition and antioxidant activity of oils from two cultivars of Cantaloupe extracted by supercritical fluid extraction. *Grasas Y Aceites* 61(1), 37–44.

306 Itoh, T., Yoshida, K., Yatsu, T., Tamura, T., Matsumoto, T. (1981). Triterpene alcohols and sterols of Spanish olive oils. *Journal of the American Oil Chemists' Society* 58, 545–550.

307 Jaber, H., Ayadi, M., Makni, J., Rigane, G., Sayadi, S., Bouaziz, M. (2012). Stabilization of Refined Olive Oil by Enrichment with Chlorophyll Pigments Extracted from Chemlali Olive Leaves. European Journal of Lipid Science and Technology, 114:1274–1283.

308 Jacob, R. A., Klevay, L. M. (1975). Determination of Trace Amounts of Copper and Zinc in Edible Fats and Oils by Acid Extraction and Atomic Absorption Spectrophotometry. *Analytical Chemistry* 47 (4) 741-743.

309 Japón-Luján R, Janeiro, P., De Castro, M.D.L. (2008). Solid–liquid transfer of biophenols from olive leaves for the enrichment of edible oils by a dynamic ultrasound-assisted approach. Journal of Agriculture and Food Chemistry 56:7231–5.

310 Jemai, H., Bouaziz, M., Sayadi, S. (2009). Phenolic composition, sugar contents and antioxidant activity of Tunisian sweet olive cultivar with regard to fruit ripening. *Journal of Agricultural and Food Chemistry* 57, 2961-2968.

311 Jiang, L., Zheng H.,Lu, H. (2013). Application of UV spectrometry and chemometric models for detecting olive oil-vegetable oil blends adulteration. *Journal of Food Science and Technology* 52, 479-485.

312 Jiménez de Blas O del Valle González A. (1996). Determination of sterols by capillary column gas chromatography. Differentiation among different types of

olive oil, virgin, refined, and solvent-extracted. *Journal of the American Oil Chemists' Society* 73, 1685–1689.

313 Jung, M.Y., Yoon, S.H., Min, D.B. (1989). Effects of processing steps on the contents of minor compounds and oxidation of soybean oil. *Journal of the American Oil Chemists Society* 66, 18–120.

314 Kabata-Pendias A, Pendias H. (1984). Trace elements in soils and plants. Boca Raton, FL, CRC Press, Inc.

315 Kaewsuksaeng, S., Yamauchi, N., Funamoto, Y., Aiamlaor, S., Mori, T., Shigyo, M., Kanlayanarat, S. (2010). Partially purification of Mg-dechelatase in relation to chlorophyll degradation in broccoli (*Brassica oleracea* L. Italica Group) florets. *Acta Horticulturae* 875, 509–514.

316 Kalogeropoulos, N., & Tsimidou, M. (2014). Antioxidants in Greek Virgin Olive Oils. Antioxidants, 3(2), 387–413.

317 Kalua, C.M., Allen, M.S., Bedgood, J.D.R., Bishop, A.G., Prenzler, P.D., Robards, K. (2007) Olive oil volatile compounds, flavour development and quality, a critical review. *Food Chemistry* 100,273–286

318 Kamal-Eldin, A. (2006). Effect of fatty acids and tocopherols on the oxidative stability of vegetable oils. European Journal of Lipid Science and Technology, 58, 1051e1061.

319 Kamal-Eldin, A., Andersson, R. (1997). Multivariate Study of the Correlation between Tocopherol Content and Fatty Acid Composition in Vegetable Oils. *Journal of the American Oil Chemists' Society*.74, 375-380.

320 Kambourakis, S., Draths, K.M., Frost, J.W. (2000). Synthesis of gallic acid and pyrogallol from glucose, replacing natural product isolation with microbial catalysis. *Journal of the American Oil Chemists' Society* 122,9042–9043.

321 Kapellakis , I.E. , Tsagarakis , K.P. , Crowther , J.C. (2008). Olive oil his- tory, production and by-product management . *Reviews in Environmental Science and Bio/Technology* 7 , 1 – 26 .

322 Karim, M. M., Karim, S. F. E., Rana, A. A., Masum, S. M., Mondol, A., & Israt, S. S. (2015). ATR-FTIR spectroscopy and chemometric techniques for the identification of edible vegetable oils, *50*(4), 233–240.

323 Karp, A., Seberg, O., Buiatti, M. (1996). Molecular techniques in the assessment of botanical diversity. *PNNals of Botany* 78 143-149.

324 Kato, N., Shiroya, M., Yoshida, S., Hasegawa, M. (1968) Biosynthesis of gallic acid by a ho- mogenate of the leaves of *Pelargonium inquinanl*. *Botanical magazine Tokyo* 81,506.

325 Katsiotis, A., Hagidimitriou, M., Douka, A., Hatzopoulos, P. (1998). Genomic organization, sequence interrelationships, and physical localization using in situ hybridization of two tandemly repeated DNA sequences in the genus. *Olea. Genome* 41,527–534.

326 KaufmPNN, L., Rousseeuw, P. J. (1999) Finding Groups of Data. Wiley, New York.

327 Keceli, T., Gordon, M. H. (2002) Ferric ions reduce the antioxidant activity of the phenolic fraction of virgin olive oil *Journal of Food Science* 67 3 943-947

328 Kettaneh, N., Berglund, A., and Wold, S. (2005). PCA and PLS with very large data sets. *Computational Statistics and Data Analysis* 48(1), 69–85.

329 Kimura, M. (1968) Evolutionary rate at the molecular level. *Nature* 217, 624–626.

330 Kimura, M. (1983) The Neutral Theory of Molecular Evolution. Cambridge Univ. Press, Cambridge

331 Kiosseoglou, B., Vlachopoulou, I., Boskou, D. (1987). Esterified 4-monomethyl- and 4,4-dimethyl- sterols in some vegetable oils. *Grasas y Aceites* 38, 102-103

332 Kiritsakis, A.K.(1990). Chemistry of olive oil. In: Kiritsakis AK (editor): Olive Oil. Champaign, Illinois: American Oil Chemist's Society; 25–55.

333 Kiritsakis, A.K., Nauos, G.D., Polymenoupoulos, Z., Thomai, T., Sfakiotakis, E.Y., (1998). Effect of fruit storage conditions on olive oil quality. *Journal of the American Oil Chemists' Society* 75, 721–724.

334 Kirwan, J. A., Weber, R. J. M., Broadhurst, D. I., Viant, M. R. (2014). Direct infusion mass spectrometry metabolomics dataset, a benchmark for data processing and quality control. *Scientific Data*.

335 Kombargi, W.S., Michelakis, S.E., Petrakis, C.A. (1998). Effect of olive surface waxes on oviposition by *Bactrocera oleae* (Diptera, *Tephritidae*). *Journal of Economic Entomology* 91, 993–998

336 Koulman, A.L, Tapper, B.A., Fraser, K., Cao, M., Lane, G.A., Rasmussen, S. (2007). High-throughput direct-infusion ion trap mass spectrometry, a new method for metabolomics. *Rapid Commun Mass Spectrometery* 21(3),421-8.

337 Kraus, T.E., Mckersie, B.D.,Fletcher, R.A. (1995). Paclobutrazole induced tolerance of wheat leaves to paraquat may involve antioxidant enzyme activity. *Journal of Plant Physiology* 145, 570–576.

338 Krichene, D., Allalout, A., Mancebo-Campos, V., Salvador, M. D., Zarrouk, M. and Fregapane, G. (2010). Stability of virgin olive oil and behaviour of its natural antioxidants under medium temperature accelerated storage conditions. Food Chemistry 121(1): 171-177.

339 Krichene, D., Salvador, M. D., & Fregapane, G. (2015). Stability of Virgin Olive Oil Phenolic Compounds during Long-Term Storage (18 Months) at Temperatures of 5-50°C. Journal of Agricultural and Food Chemistry, 63(30), 6779–6786.

340 Krichene, D., Taamalli, W., Daoud, D., Salvador, M. D., Fregapane, G., Zarrouk, M. (2007). Phenolic compounds, tocopherols and other minor components in virgin olive oils of some Tunisian varieties. *Journal of Food Biochemistry* 31, 179–194

341 Krinsky, N.I. (1989) Antioxidant functions of carotenoids. *Free Radical Biology and Medicine* 7,617-635

342 Kruzlicova, D., Mocak, J., Balla, B., Petka, J., Farkova, M., Havel, J. (2009). Classification of Slovak white wines using artificial neural networks and discriminant techniques. *Food Chemistry*, 112(4), 1046–1052

343 Kubo, I., Matsumoto, A., Takase, I. (1998). A multichemical defense mechanism of bitter olive *Olea europaea* (*Oleaceae*). Is oleuropein a phytoalexin precursor? *Journal of Chemical Ecology* 11, 251–263

344 Kumar, K., & Kumar Mishra, A. (2015). Parallel factor (PARAFAC) analysis on total synchronous fluorescence spectroscopy (TSFS) data sets in excitation-emission matrix fluorescence (EEMF) layout: Certain practical aspects. Chemometrics and Intelligent Laboratory Systems, 147, 121–130.

345 Kumar, K., Mishra, A.K. (2012).Application of parallel factor analysis to total synchronous fluorescence spectrum of dilute multifluorophoric solutions: addressing the issue of lack of trilinearity in total synchronous fluorescence data set, Analytica Chimica Acta, 755 (2012) 37–45.

346 Kuroda, M., Qzawa, T., Imagawa, H. (1990). Changes in chloroplast peroxidase activities in relation to chlorophyll loss in barley leaf segments. *Physiologia plantarum* 80, 555-560.

347 Kyriakidis, N. B., & Skarkalis, P. (2000). Fluorescence spectra measure- ment of olive oil and other vegetable oils. Journal of AOAC International,83(6), 1435–1439

348 Kyriakidis, N. B., Skarkalis, P. (2000).Fluorescence spectra measurement of olive oil and other vegetable oils. *Journal of the American Oil Chemistry Society* 83, 1435–1439

349 Labuza, T.P. (1971) Kinetics of lipid oxidation in foods. CRC Critical Reviews Food Science Technology., 2, 355–405.

350 Lagas, P., Loch, J.P.G., Bom, C.M. (1984). The behavior of barium in a landfill and the underlying soil. *Water Air and Soil Pollution* 22,121-129.

351 Lakowicz, J. R. (2006). Principles of Fluorescence Spectroscopy, Third edn. Kluwer Academic/Plenum Publishers, ISBN 0387312781, New York

352 Lamzira, Z., Ghabbour, N., Rokni, Y., Thonart, P., Chihib, N. E., Berrabah, M., Asehraou, A. (2014). Characterization of phenolic profile of moroccan picholine olive variety. *Journal of Materials and Environmental Science* 5(2), 490–497.

353 Lanfer-Marquez, U.M., Barros, R.M.C., Sinnecker, P. (2005). Antioxidant activity of chlorophylls and their derivatives. *Food Research International* 38,885–91.

354 Lavee, S. (1986). Olive. In S. P. Monselise (Ed.), Handbook of fruit set and development (pp. 261–274). *Boca Raton*, FL, CRC Press.

355 Lavee, S., Wodner, M. (1991). Factors affecting the nature of oil accumulation in fruit of olive. *The Journal of Horticultural Science and Biotechnology* 66, 583-591

356 Lawson, W.R., Goulter, K.C., Henry, R.J., Kong, G.A., Kochman, J.K. (1998) Marker-assisted selection for two rust resistance genes in sunflower. *Molecular Breeding* 4, 227-234

357 Lazzez, A., Vichi, S., Grati Kammoun, N., Ne, M., Arous, J., Khlif, M., Romerp, A., Cossentini. M. (2011). A four year study to determine the optimal harvesting period for Tunisian Chemlali olives. *European Journal of Lipid Science and Technology* 113(6),796–807

358 Leardi, R. Paganuzzi, V. (1987). Characterization of the origin of extra virgin olive oils by chemometric methods applied to the sterols fraction. *Rivista Italiana delle Sostanze Grasse* 64, 131–136.

359 Lee, Jiyeun Lee, Yoosung Choe, Eunok (2007) Temperature dependence of the autoxidation and antioxidants of soybean , sunflower , and olive oil. *European Food Research and Technology* 226,239–246

360 Leja, M., Mareczek, G., Wyzgolik, G., Klepacz-Baniak, J., Czekońska, K. Antioxidative properties of bee pollen in selected plant species. *Food Chemistry* 2007, 100, 237–240.

361 Lenhardt, L., Bro, R., Zeković, I., Dramićanin, T., Dramićanin, M. D. (2015) Fluorescence Spectroscopy Coupled with PARAFAC and PLS DA for Characterization and Classification of Honey. Food Chemistry, 175, 284–291.

362 Leonardis, D.A., Macciola, V., Felice, D.M. (2000). Copper and iron determination in edible vegetable oils by graphite furnace atomic absorption spectrometry after extraction with diluted nitric acid. *International Journal of Food Science and Technology* 35, 371-375.

363 Lerma-García, J.M., Herrero-Martínez, G. Ramis-Ramos, E. F. Sim´o-Alfonso. (2008). Prediction of the genetic variety of Spanishextra virgin olive oils using fatty acid and phenolic compound profiles established by direct infusion mass spectrometry. *Food Chemistry* 108, 1142.

364 Lia, F and Buhagiar, J (2014). Chemical characterisation of grapevine waste and bioactivity on human cancer cell lines. Master Thesis. University of Malta Biology Department

365 Liang, Y., Kvalheim, O. M. (1996). Robust methods for multivariate analysis – A tutorial analysis. *Chemometrics and Intelligent Laboratory Systems*, 32, 1–10.

366 Liao, K.-L. & Yin M. C (2000).: Individual and combined antiocidant effects of seven phenolic agents in human erythrocyte membrane ghosts and phosphatidylcholine liposome systems: importance of the partition coefficient. Journal of Agricultural and Food Chemistry, 48, 2266 - 2270

367 Lien, E.J., Ren, S., Bui, H.H., Wang R. (1999). Quantitative structure-activity relationship analysis of phenolic antioxidants. *Free Radical Biology and Medicine* 26, 285–294

368 Litt, M., Luty, J.A. (1989). A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. A*merican Journal of Human Genetic* 44 397-401.

369 Llorent-Martinez, E. J., Ortega-Barrales, P., Fernandez-de Cordóva, M. L., Ruiz-Medina, A. (2011). Analysis of the Legistated Metals in Different Categories of Olive and Olive-pomace Oils. *Food Control* 22, 221-225.

370 Lo Scalzo, R., Scarpati, M.L., Verzegnassi, B., Vita, G. (1994). *Olea europaea* chemicals repellent to Bactrocera oleae females. *Journal of Chemical Ecology* 20, 1813–1823

371 Longobardi, F., Ventrella, a., Napoli, C., Humpfer, E., Schütz, B., Schäfer, H., Sacco, a. (2012). Classification of olive oils according to geographical origin by using 1H NMR fingerprinting combined with multivariate analysis. *Food Chemistry*, *130*(1), 177–183. https://doi.org/10.1016/j.foodchem.2011.06.045

372 Lonso, Ä. M., Regapane, E. Z., Giuseppe, F., Alvador, M. D.,Esamparados, S., Ordon, Michael, H.G. (2003)Changes in Phenolic Composition and Antioxidant Activity of Virgin Olive Oil during Frying *Journal of Agricultural and Food Chemistry*2003, 51, 667−672

373 Loureiro, J., Rodriguez, E., Costa, A., Santos, C. (2007). Nuclear DNA content estimations in wild olive (*Olea europaea* L. ssp. *europaea* var. sylvestris Brot.) and Portuguese cultivars of O. *europaea* using flow cytometry. Genet Resour Crop Evol 54,21–25

374 Luchetti, F. (2002). Importance and future of olive oil in the world market – an introduction to olive oil. *European Journal of Lipid Science and Technology* 104,559-563.

375 Lumaret, R., Amane, M., Ouazzani, N., Baldoni, L., Debain, C. (2000). Chloroplast DNA variation in the cultivated and wild olive taxa of the genus *Olea* L. T*heoretical and Applied Genetics* 101(4), 547–553.

376 Lumaret, R., N. Ouazzani. (2001). Ancient wild olives in Mediterranean forests. *Nature* 413, 700.

377 Lumaret, R., Ouazzani, N., Michaud, H., Villemur, P. (1997). Cultivated olive and oleaster, two closely connected partners of the same species (*Olea europaea* L.), evidence from allozyme polymorphism. *Bocconea* 7, 39–42.

378 Luna, G., Morales, M.T., Aparicio, R. (2006). Characterization of 39 varietal virgin olive oils by their volatile composition. *Food Chemistry* 98, 243–252.

379 Mabry, T.J., Markham, K.R., Thomas, M.B. (1970). The systematic identification of fiavonoids. Springer-Verlag, Berlin, 50-56.

380 Maggio, R. M, Kaufman, T.S, Carlo, M. D., Cerretani, L., Bendini, A., Cichelli, A., Compagnone, D. (2009). Monitoring of fatty acid composition in virgin olive oil by Fourier transformed infra- red spectroscopy coupled with partial least square. *Food Chemistry*. 114, 1549–1554

381 Maggio, R. M., Kaufman, T. S., Del Carlo, M., Cerretanil, L., Bendini, A., Cichelli, A., et al. (2009). Monitoring of fatty acid composition in virgin olive oil by Fourier transformed infrared spectroscopy coupled with partial least squares. Food Chemistry, 114, 1549–1554.

382 Mahedero, M. C., Diaz, N. M., Muñoz de la Peña, A., Espinosa Mansilla, A., Gonzalez Gomez, D., Bohoyo Gil, D. (2005). Strategies for solving matrix effects in the analysis of sulfathiazole in honey samples using three-way photochemically induced fluorescence data. Talanta, 65, (3), 806-813.

383 Mailer, R.J., Ayton, J., Graham, K. (2010).The influence of growing region, cultivar and harvest timing on the diversity of Australian olive oil. *Journal of the American Oil Chemistry Society* 87,877–884.

384 Malheiro, R., Oliveira, I., Vilas-Boas, M., Falcão, S., Bento, A., Pereira, J.A. (2009). Effect of Microwave Heating with Different Exposure Times on Physical and Chemical Parameters of Olive Oil. Food and Chemical Toxicology, 47:92–9

385 Mallégol, J., Gardette, J.-L., & Lemaire, J. (2000). Long-term behavior of oil-based varnishes and paints. Fate of hydroperoxides in dring oils. Journal of the American Oil Chemists' Society, 77(3), 249–255.

386 MALSIS (Maltese Soil Information System). 2004. Soil geographic database of the Maltese Islands. National Soil Unit, Ministry for Rural Affairs and the Environment, Malta

387 Mancebo-Campos, V., Fregapane, G., & Salvador, M. D. (2008). Kinetic study for the development of an accelerated oxidative stability test to estimate virgin olive oil potential shelf life. European Journal of Lipid Science and Technology, 110(10), 969–976.

388 Mangos, T.J., Berger, R.G. (1997). Determination of major chlorophyll degradation products. Z. Lebensm.Unters. Forsch. A, 204: 345-350.

389 ManivPNNan, P., Abdul, Jaleel, C, Sankar, B., Kishorekumar, A., Somasundaram, R., Lakshmanan, G.M.A., PPNNeerselvam, R. (2007). Growth,

biochemical modifications and proline metabolism in *Helianthus PNNuus L*. as induced by drought stress. *Colloids and Surfaces Biointerfaces* 59, 141–149.

390 Mannina, L., Marini, F., Gobbino, M., Sobolev, a. P., & Capitani, D. (2010). NMR and chemometrics in tracing European olive oils: The case study of Ligurian samples. *Talanta*, *80*(5), 2141–2148. https://doi.org/10.1016/j.talanta.2009.11.021

391 Manzi , P. , Panfili , G. , Esti , M. , Pizzoferrato , L. (1998) . Natural antioxidant in the unsaponifiable fraction of virgin olive oils from different cultivars. *Journal of the Science of Food and Agriculture* 77 , 115 – 120

392 Marfil, R., Cabrera-Vique, C., Gimenez, R., Bouzas, P. R., Martinez, O., Sánchez, J. A. (2008). Metal content and physicochemical parameters used as quality criteria in virgin argan oil, Influence of the extraction method. *Journal of Agricultural and Food Chemistry*, 56, 7279–7284.

393 Mariani, C., Lanzani, A., Fedeli, E. (1987). La presentia di cere nell'olio di girasole in funzione di alcuni trattamenti technologici. *Rivista Italiana delle Sostanze Grasse* 13-16.

394 Marini, D., Grassi, L., Balestrieri, F., Pascucci, E. (1990). Analisi spettrofotofluorimetrica dell'olio di oliva. Possibilita` di applcazione. *Rivista Italiana delle Sostanze Grasse* 67,95-99.

395 Martens, H. (1979) Factor analysis of chemical mixtures. *Analytica Chimica Acta* 112, 423-442.

396 Martens, H. Nres, T. (1989) Multivariate Calibration, John Wiley, Chichester, Sussex.

397 Martins-Lopes, P., Gomes, S., Lima-Brito, J., Lopes, J. Guedes-Pinto, H. (2009). Assessment of clonal genetic variability in *Olea europaea* L. 'Cobrançosa' by molecular markers. *Scientia Horticulturae* 123, 82-89.

398 Martins-Lopes, P., Lima-Brito, J., Gomes, S., Meirinhos, J., Santos, L., Guedes-Pinto, H. (2007). RAPD and ISSR molecular markers in *Olea europaea* L., Genetic variability and cultivar identification. *Genetic Resource and Crop Evolution* 54,117-128.

399 Massart, D. L., KaufmPNN, L. (1983). The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis. Wiley, New York.

400 Mateos , R. , Dominguez , M.M. , Espartero , J.L. , Cert , A. (2003). Antioxidant effect of phenolic compounds, alpha- tocopherol, and other other minor

components in virgin olive oil . *Journal of Agricultural and Food Chemistry* 51 , 7170 – 7175 .

401 Mateos, R., Espartero, J. L., Trujillo, M., Ríos, J. J., León-Camacho, M., Alcudia, F., Cert, A. (2001). Determination of phenols, flavones, and lignans in virgin olive oils by solid-phase extraction and high-performance liquid chromatography with diode array ultraviolet detection. *Journal of Agricultural and Food Chemistry* 49, 2185-2192

402 Matthäus B., Özcan M., M. (2011). Determination of fatty acid, tocopherol, sterol contents and 1,2- and 1,3-Diacylglycerols in four different virgin olive oil. *Journal of Food Processing and Technology* 2,117.

403 Mazzitelli, O., Calleja, A., Sardella, D., & Farrugia, C. (2015). Analysis of the molecular diversity of *Olea europaea* in the Mediterranean Island of Malta. *Genetic Resources and Crop Evolution* 1021–1027.

404 McDougall, G., Martinussen, I., Stewart, D. (2008). Towards fruitful metabolomics, High throughput analyses of polyphenol composition in berries using direct infusion mass spectrometry. *Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences* 871(2), 362–369.

405 McEwen, C,N., McKay, R.G., Larsen, B.S. (2005). Analysis of solids, liquids, and biological tissues using solids probe introduction at atmospheric pressure on commercial LC/MS instruments. Analytical Chemistry; 77: 7826.

406 McEwen, C., Gutteridge, S. (2007). Analysis of the inhibition of the ergosterol pathway in fungi using the atmospheric solids analysis probe (ASAP) method. Journal of the American Society for Mass Spectrometry; 18: 1274.

407 Mei, L., McClements, D.J., Wu, J., Decker, E.A. (1998). Iron-catalyzed lipid oxidation in emulsions as affected by surfactant, pH and NaCl. *Food Chemistry* 61,307-312.

408 Meira, M., Quintella, C. M., dos Santos Tanajura, A., da Silva, H. G. R., Fernando, J. D. S., da Costa Neto, P. R., Pepe, J. M., Santos, M. A., Nascimento, L. L. (2011). Determination of the Oxidation Stability of Biodiesel and Oils by Spectrofluorimetry and Multivariate Calibration. *Talanta* 85, 430-434.

409 Mekuria, G., Collins, G., Sedgley, M. (1999). Genetic variability between different accessions of some common commercial olive cultivars. *Journal of Horticultural Science and Biotechnology* 74, 309–314.

410 Melchert, H.-U., Pabel, E. (2000). Quantitative determination of α,β,γ-, and δ-tocopherols in human serum by high-performance liquid chromatography and gas chromatography-mass spec- trometry as trimethylsilyl derivatives with a two-step sample preparation. *Journal of Chromatography A* 896, 209-215.

411 Menendez, J.A., Papadimitropoulou, A., Vellon, L., Lupu, R. (2006). A genomic explanation connecting "Mediterranean diet", olive oil and cancer, oleic acid, the main monounsaturated fatty acid of olive oil, induces formation of inhibitory " PEA3 transcription factor-PEA3 DNA binding site " complexes at the Her-2/neu (erbB-2) oncogene promoter in breast, ovarian and stomach cancer cells. *European Journal of Cancer* 42, 2425 – 2432 .

412 Merken, H.M. Beecher, G.R. (2000). Measurements of food flavonoids by high-performance liquid chromatography, a review. *Journal of Agricultural and Food Chemistry*, 48,577–599.

413 Meyer, D., Leisch, F., Hornik, K. (2003). The Support Vector Machine Under Test. Neurocomputing, 55, 169–186.

414 Mignani, A.G., Ciaccheri, L., Ottevaere, H., Thienpont, H., Conte, L., Marega, M. (2011). Visible and near-infrared absorption spectroscopy by an integrating sphere and optical fibers for quantifying and discriminating the adulteration of extra virgin olive oil from Tuscany. *Analytical and Bioanaytical Chemistry*. 399(3), 1315–1324.

415 Miles , E.A. , Zoubouli , P. , Calder , P.C. (2005). Differential anti-inflammatory effects of phenolic compounds from extra virgin olive oil identified in human whole blood cultures. *Nutrition*. 21, 389 – 394.

416 Millar, S., Robert, P., Devaux, M. F., Guy, R. C. E., and Maris, P., (1996). Near-infrared spectroscopic measurements of structural changes in starch-containing extruded products. *Applied Spectroscopy* 50 (9), 1134-1139.

417 Minguez Mosquera , M.I. , Gallardo Guerrero , L. (1995) . Anomalous trans-formation of chloroplastic pigments in Gordal variety olives during processing for tables olives . J. Food Protect. 58 , 1241 – 1248 .

418 Mínguez-Mosquera, M. I., Gandul-Rojas, B., Gallardo-Guerrero, M. L., Roca, M., Jarén- Galán, M. (2007). Chlorophylls. InW. J. Hurst (Ed.), Methods of analysis in functional foods and added nutraceuticals (pp. 337–400) (2nd ed.). Boca Raton, FL, CRC Press LLC

419 Mínguez-Mosquera, M.I, Garrido-Fernandez, J. (1989). Chlorophyll and carotenoid presence in olive fruit. *Journal of Agricultural and Food Chemistry* 37,1-7

420 Minguez-Mosquera, M.I., Rejano, L., Gandul, B., Sanchez, A.H., Garrido, J. (1991). Color-pigment correlation in virgin olive oil. *Journal of the American Oil Chemists Society* 68, 322–337.

421 Minguez-Mosquera,M. I., Jarén-Galán, M. (1999). Quantitative and qualitative changes associated with heat treatments in the carotenoid content of paprika oleoresins. *Journal of Agricultural and Food Chemistry* 47, 4379–4383.

422 Mishima, K., Tanaka, T., Pu, F., Egashira, N., Iwasaki, K., Hidaka, R., Matsunaga, K., Takata, J., Karube, Y, Fujiwara, M. (2003). Vitamin E isoforms α -tocotrienol and γ -tocopherol prevent cerebral infarction in mice. *Neuroscience Letters* 337 , 56 – 60 .

423 Mohan, M., Nair, S., Bhagwat, A., Krishna, T.G., Yano, M., Bhatia, C.R., Sasaki, T. (1997). Genome mapping, molecular markers and markers-assisted selection in crop plants. *Molecular Breeding* 3, 87-103.

424 Moharam M.A., Abbas L.M. (2010) A study on the effect of microwave heating on the properties of edible oils using FTIR spectroscopy, African Journal of Microbiology Research 4, 1921-1927.

425 Moller, J.K.S., Catharino, R.R., Eberlin, M.N. (2005) Electrospray ionization mass spectrometry fingerprinting of whisky, immediate proof of origin and authenticity. *Analyst* 130, 890

426 Montedoro, G.F., Garofolo, L. (1984). Caratteristiche qualitative degli oli vergini di oliva. Influenza di alcune variabili, varietà, ambiente, conservazione, estrazione, condizionamento del prodotto finito. *Rivista Italiana delle Sostanze Grasse* LXI, 3-11.

427 Montedoro, G.F., Servili, M. Baldioli, M. Miniati, E. (1992a). Simple and hydrolizable phenolic compounds in virgin olive oil. 1. Their extraction, and quantitative and semiquantitative evaluation by HPLC. *Journal of Agricultural and Food Chemistry* 40,1571–1576.

428 Montedoro, G.F., Servili, M., Baldioli, M., Miniati, E. (1992b). Simple and hydrolizable phenolic compounds in virgin olive oil. 2. Initial characterization of the hydrolyzable fraction. *Journal of Agricultural and Food Chemistry* 40,1577–1580.

429 Montedoro, G.F., Taticchi, A., Esposto, S., Selvaggini, R., Urbani, S., Servili, M. (2007). Antioxidants in virgin olive oil (Review). *Olea* 26, 5-13.

430 Montedoro, G.F.,Servili. M., Baldioli, M. (1993). Simple and hydrolyzable phenolic compounds in virgin olive oil. 3. Spectroscopic characterizations of the secoridoid derivatives. *Journal of Agricultural and Food Chemistry* 41,2228–2234.

431 Monteleone, E., Caporale, G., Lencioni, L., Favati, F. and Bertuccioli, M. 1995. Optimization of virgin olive oil quality in relation to fruit ripening and storage. In Food Flavours, Generation, Analysis and Process Influence (G. Charalambous, ed.) pp. 397–418, Elsevier Science B.V., Amsterdam, The Netherlands.

432 Montemurro, C., Simeone, R., Pasqualone, A., Ferrara, E., Blanco, A. (2005). Genetic relationships and cultivar identification using AFLP and SSR markers. *Journal of Horticultural Science and Biotechnology* 80, 105–110.

433 Morales, M. T., Rios, J. J., Aparicio, R. (1997). Changes in the volatile composition of virgin olive oil during oxidation, Flavors and off-flavors. *Journal of Agricultural and Food Chemistry* 45, 2666–2673

434 Morales, M. T., Tsimidou, M. (2000). The role of volatile compounds and polyphenols in olive oil sensory quality Handbook of Olive Oil: Analysis and Propertie, p. 393. J. Harwood and R. Aparicio (Eds.), Aspen Publishers: Gaithersburg, Maryland.

435 Morales, M.T., Alonso, M.,V., Rios, J.J., Aparicio, R. (1995). Virgin Olive Oil Aroma, Relationship between Volatile Compounds and Sensory Attributes by Chemiometrics *Journal of Agricultural and Food Chemistry* 43, 2925-2931

436 Morales, M.T., Tsimidou, M. (2000). The role of volatile compounds and polyphenols in olive oil sensory quality. p. 393–458. In, J. Harwood, R. Aparicio (eds.), Handbook of olive oil. Analysis and properties. Aspen Publ., Gaithersburg, MD.

437 Moran, J.F., Klucas, R.V., Grayer, R.J., Abian, J., Becana, M. (1997). Complexes of iron with phenolic compounds from soybean nodules and other legume tissues, Proox- idant and antioxidant properties. *Free Radical Biol and Medicine* 22,861-870.

438 Morello, J. R., Romero, M. P., Ramo, T., Motilva, M. J. (2005). Evaluation of L-phenyllanine ammonia-lyase activity and phenolic profile in olive drupe from fruit setting period to harvesting time. *Plant Science* 168, 65–72.

439 Morelló, J.R., Motilva M.J., Tovar M.J., Romero M.P. (2004). Changes in commercial virgin olive oil (cv Arbequina) during storage, with special emphasis on the phenolic fraction. *Food Chemistry* 85, 357–364.

440 Morello, J.R., Motilva, M.J., Tovar, M.J., Romero, M.P. (2004). Changes in commercial virgin olive oil (cv Arbequina) during storage, with special emphasis on the phenolic fraction. Food Chemistry 85:357–364.

441 Moreno M., Olivares M., López A., Adelantado J., Reig B. (1999) Analytical evaluation of polyunsaturated fatty acids degradation during thermal oxidation of edible oils by Fourier transform infrared spectroscopy, Talanta, 50, 269-275.

442 Mousa, Y. M., Gerasopoulos, D., Metzidakis, I., Kiritsakis, A., Effect of altitude on fruit and oil quality characteristics of ‘‘Mastoides’’ olives. *Journal of the Science of Food and Agriculture* 1996, 71, 345–350

443 Mousa, Y.M., D. Gerasopoulos, I. Metzidakis, A. Kiritsakis. 1996. Effect of altitude on fruit and oil quality characteristics of ‘Mastoides’ olives. *Journal of Agricultural and Food Chemistry* 71,345–350.

444 Mumm R, Hilker, M. (2006). Direct and indirect chemical defence of pine against folivorous insects. *Trends Plant Science* 11(7),351–358.

445 Muñoz de la Peña, A., Mora Diez, N., Mahedero García, M. C., Bohoyo Gil, D., Cañada- Cañada, F. (2007). A chemometric sensor for determining sulphaguanidine residues in honey samples. Talanta 73, (2), 304-313.

446 Muzzalupo, I., Stefanizzi, F., Perri, E. (2009). Evaluation of olives cultivated in southern Italy by simple sequence repeat markers. *HortScience* 44, 582-588.

447 Naqvi, N. I., Chattoo, B. B. Development of a sequence characterized amplified region (SCAR) based indirect selection method for a dominant blast-resistance gene in rice. Genome 1996, 39,26-30.

448 Nardini, M., D´Aquino, M., Tomassi, G., Gentili, V., Di Felice, M. & Scanccini, C. (1995) Inhibition of human low-density lipoprotein oxidation by caffeic acid and other hydroxycinnamic acid derivatives. Free Radical Biology and Medicine 19, 5, 541 – 552.

449 Naz , S., Siddiqi, R., Sheikh, H., Sayeed, S.A. (2005) . Deterioration of olive, corn and soybean oils due to air, light, heat and deep-frying. *Food Research International* 38 , 127 – 134.

450 Ndou, T. & Warner, I. M. (1991). Applications of Multidimensional Absorption and Luminescence Spectroscopies in Analytical Chemistry. Chemical Reviews, 91(4), 493-507, ISSN 1520-6890

451 Negi, M.S., Devic, M., Delseny, M., Lakshmikumaran, M. (2000) Identification of AFLP fragments linked to seed coat colour in Brassica juncea and conversion to a SCAR marker for rapid selection. *Theory of Applied Genetics* 101,146–152

452 Neish, A.C. (1960). Biosynthetic pathways of aromatic compounds. In, Machlis L, BriggsWR (eds) *PNNual review of plant physiology* 11, 55–59

453 Nguyen,D.V. and Rocke,D.M. (2002) Tumor classification by partial least squares using gene expression data. Bioinformatics, 18, 39–50

454 Nicklas, T.A., Hampl, J.S., Taylor, C.A., Thompson, V.J., Heird, W.C. (2004). Monounsaturated fatty acid intake by childrens and adults, temporal trends and demographic differences. *Nutrition Reviews* 62, 132–141.

455 Nikoloudakis, N., Banilas, G., Gazis, F., Hatzopoulos, P., Metzidakis, J. (2003). Discrimination and genetic diversity among cultivated olives of Greece using RAPD markers. *Journal of the American Society for Horticultural Science* 128(5), 741–746.

456 Noreen Z, Ashraf, M. (2009). Assessment of variation in antioxidative defense system in salt-treated pea (*Pisum sativum*) cultivars and its putative use as salinity tolerance markers. *Journal of Plant Physiology* 166, 1764–1774

457 Obied, H. K., Karuso, P., Prenzler, P. D., Robards, K. (2007). Novel secoiridoids with antioxidant activity from Australian olive mill waste. *Journal of Agricultural and Food Chemistry* 55(8), 2848–2853.

458 Ocakoglu, D., Tokatli, F., Ozen, B., Korel, F. (2009). Distribution of simple phenols, phenolic acids and flavonoids in Turkish monovarietal extra virgin olive oils for two harvest years. *Food Chemistry* 113, 401–410.

459 Ockels, F.S., Eyles, A., McPherson, B.A., Wood, D,L., Bonello, P. (2007) Phenolic chemistry of coast live oak response to Phytophthora ramorum infection. *Journal of Chemical Ecology* 33(9),1721–1732.

460 Oliveras-Lopez, M.J., Innocenti, M., Giaccherini, C., Ieri, F., Romani, A., Mulinacci, N. (2007). Study of the phenolic composition of Spanish and Italian

monocultivar extra virgin olive oils, distribution of lignans, secoiridoidic, simple phenols and flavonoids. *Talanta* 73,726–732

461 Ollivier, D., Artaud, J., Pinatel, C., Durbec, J. P., Guere, M. (2003). Triacylglycerol and fatty acid compositions of french virgin olive oils. Characterization by chemometrics. *Journal of Agricultural and Food Chemistry* 51, 5723-5731

462 Omar, S.H. (2010) Oleuropein in olive and its pharmacological effects. *Scientia Pharmaceutica* 78(2),133–154.

463 Osman, M., Metzidakis, I., Gerasopoulas, D., Kiritsakis, A. (1994). Qualitative changes in olive oil of fruits collected from trees grown at two altitudes—Rivista Italiana delle Sostanze Grasse 71, 187-19

464 Ossipov, V., Bonner, C., Ossipova, S., Jensen, R.(2000). Broad-specificity quinate (shikimate) dehydrogenase from *Pinus taeda* needles. *Plant* Physiology Biochemistry 38,923–928.

465 Ouazzani, N., Lumaret, R., Villemur, P. (1994). An evaluation of approach the genetic variability of olive tree using enzymatic markers. Acta Hortica 356, 91–94.

466 Ouazzani, N., Lumaret, R., Villemur, P. (1996). Genetic variation in the olive tree (*Olea europaea* L.) cultivated in Morocco. *Euphytica* 91, 9-20.

467 Ouazzani, N., Lumaret, R., Villemur, P., Di Giusto, F. (1993). Leaf allozyme variation in cultivated and wild olive trees (*Olea europaea*). *Journal Hereditary* 84, 34–42.

468 Ouni, Y., Taamalli, A., Gómez-Caravaca, A. M., Segura-Carretero, A., Fenández- Gutiérrez, A., Zarrouk, M. (2011). Characterisation and quantification of phenolic compounds of extra-virgin olive oils according to their geographical origin by a rapid and resolutive LC-ESI-TOF MS method. *Food Chemistry* 127, 1263–1267.

469 Owen, C.A., Bita, E.C., Banilas, G. (2005). AFLP reveals structural details of genetic diversity within cultivated olive germoplasm from the Eastern Mediterranean. *Theory of Applied Genetics* 110, 1169-1176.

470 Owen, R.W., Mier, W., Giacosa, A., Hull, W.E., Spiegelhalder, B., Bartsch, H. (2000). Olive- oil consumption and health, the possibile role of antioxidants. *Food Chemistry Toxicology* 38,647–659.

471 Pafundo, S., Agrimonti, C., Maestri, E., Marmiroli, N. (2007). Applicability of SCAR markers to food genomics, olive oil traceability. *Journal of Agricultural and Food Chemistry* 55, 6052–6059.

472 Pafundo, S., Agrimonti, C., Marmiroli, N. (2005) Traceability of Plant Contribution in Olive Oil by Amplified Fragment Length Polymorphisms. *Journal of Agricultural and Food Chemistry* 53, 6995-7002

473 Paganuzzi, V. (1997). Current Possible Adulterations of Olive Oil. *Rivista Italiana Delle Sostanze Grasse* 74, 49

474 Pallant J. SPSS survival manual, a step by step guide to data analysis using SPSS for windows. 3 ed. Sydney: McGraw Hill; 2007. pp. 179–200.

475 Palmer, J.D. (1987). Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. *The American Naturalist* 130,S6-S29

476 Papadopoulos, K., Triantis, T., Tzikis, C.H., Nikokavoura, A., Dimotikali, D. (2002). Investigations of extra virgin olive oils with seed oils using weak chemiluminescence. *Analytica Chimica Acta* 464(1), 135-140.

477 Papaseit, Totosaus, G. (1986). The quality of virgin olive oil as a function of color. *Grasas y Aceites* 37, 204-206.

478 Paran, I., Michelmore, R. (1993). Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theory of Applied Genetics* 85,985–93.

479 Parcerisa, J., Casals, I., Boatella, J., Codony, R., Rafecas, M. (2000). Analysis of olive and hazelnut oil mixtures by high-performance liquid chromatography-atmospheric pressure chemical ionisation mass spectrometry of triacylglycerols and gas-liquid chroma- tography of non-saponifiable compounds (tocopherols and sterols). *Journal of Chromatography A* 881, 149-158.

480 Pasqualone, A., Di Rienzo, V., Blanco, A., Summo, C., Caponio, F., & Montemurro, C. (2012). Characterization of virgin olive oil from Leucocarpa cultivar by chemical and DNA analysis. *Food Research International* 47(2), 188–193.

481 Pasqualone, A., Montemurro, C., Caponio, F., Blanco, A. (2004) Identification of virgin olive oil from different cultivars by analysis of DNA microsatellites. *Journal of Agricultural and Food Chemistry* 52, 1068-1071.

482 Pastori, G.M., Trippi, V.S. (1992). Oxidative stress induces high rate of glutathione reductase synthesis in a drought-resistant maize strain. *Plant Cell Physiology* 33, 957–961.

483 Patumi, M., D'Andri A, R., Fontanaza, G., Morelli, G., Gioro, P. Sorrentino, G. (1999). Yield and oil quality of intensively trained trees of three cultivars of olive (*Olea europaea* L.) under different irri- gation regimes. *Journal of Horticultural Science and Biotechnology* 74, 729–737.

484 Paz Aguilera, M., Beltran, G., Ortega, D., Fernandez, A., Jimenez, A., Uceda, M. (2005). Characterization of virgin olive oil of Italian olive cultivars Frantoio and Leccino, grown in Andalusia. *Food Chemistry* 89,387–391.

485 Pedrielli, P., Pedulci, G. F., Skibsted, L. H. (2001). Antioxidant mechanism of flavonoids. Solvent effect on rate constant for chain- breaking reaction of quercetin and epicatechin in autoxidation of methyl linoleate. *Journal of Agricultural and Food Chemistry*, 4, 3034–3040

486 Pellegrini, N., Visioli, F., Buratti, S., Brighenti, F. (2001). Direct analysis of total antioxidant activity of olive oil and studies on the influence of heating. *Journal of Agricultural and Food Chemistry* 49, 2532-2538.

487 Peng, J., Lu, F., Ralph, J. (1999). Isochroman lignin trirners from DFRC-degraded *Pinus taeda. Phytochemistry* 50, 659–666

488 Perez- Camino, M.C., Cert, A., (1999). Quantitative determination of hydroxy penta- cyclic triterpene acids in vegetable oils. *Journal of Agricultural and Food Chemistry* 47, 1558–1562.

489 Pérez, A. G., León, L., Pascual, M., Romero-Segura, C., Sánchez-Ortiz, A., de la Rosa R. (2014). Variability of virgin olive oil phenolic compounds in a segregating progeny from a single cross in *Olea europaea* L. and sensory and nutritional quality implications. *PLoS ONE* 9(3)

490 Perez-Camino, M.C., Moreda, W., Mateos, R., Cert, A. (2003). Simultaneous determi- nation of long-chain aliphatic aldehydes and waxes in olive oils. *Journal of Chromatography A* 983, 283–288.

491 Pérez-Gávez, A., Jarén-Galán, M., Mínguez-Mosquera, M. I. (2000). Effect of high-temperature degradative processes on ketocarotenoids present in paprika oleoresins. *Journal of Agricultural and Food Chemistry* 48, 2966–297

492 Pérez-Jiménez, J., and Saura-Calixto, F. (2006). Effect of solvent and certain food constituents on different antioxidant capacity assays. *Food Research International* 39(7), 791–800.

493 Perri, E., Raffaelli, A., Sindona, G. (1999). Quantitation of oleuropein in virgin olive oil by ionspray mass spectrometry-selected reaction monitoring. *Journal of Agricultural and Food Chemistry* 47 , 4156 – 4160 .

494 Petrakis, P.V., Agiomyrgianaki, A., Christophoridou, S., Spyros, A., Dais, P. (2008) Geographical characterization of Greek virgin olive oils (Cv. Koroneiki) using ^{1}H and ^{31}P NMR fingerprinting with canonical discriminant analysis and classification binary trees. *Journal of Agricultural and Food Chemistry* 56, 3200–3207.

495 Petucci C., Diffendal J. (2008). Mass spectroscopy letter. Journal of Mass Spectrometry. 43 1565.

496 Peyrat-Maillard, M.N., Cuvelier, M.E., Berset, C. (2003). Antioxidant activity of phenolic compounds in 2,2'-azobis(2-amidinopropane) dihydrochloride (AAPH)-induced oxidation, synergistic and antagonistic effects. *Journal of the American Oil Chemists' Society* 80, 1007–12.

497 Pietta , P.G. (2000) . Flavonoids as antioxidants. *Journal of Natural Products* 63 , 1035 – 1042

498 Pinelli, P., Galardi, C., Mulinacci, N., Vincieri, F. F., Cimato, A., Romani A. (2003). Minor polar compound and fatty acid analyses in monocultivar virgin olive oils from Tuscany. *Food Chemistry* 80, 331-336

499 Pinelo, M., Manzocco, L., Nuez, M. J., Nicoli, M. C. (2004). Solvent effect on quercetin antioxidant capacity. *Food Chemistry*, 88, 201–207.

500 Pirisi , F.M., Angioni, A., Cabras, P., Garau, V.L., Sanjust di Teulada, M.T., Kaim dos Santos, M., Bandino, G. (1997). Phenolic compounds in virgin olive oils – I. Low-wavelength quantitative determination of complex phenols by high-performance liquid chromatogramphy under isocratic. *Journal of Chromatography A* 768 , 207 – 213

501 Pizarro, C., Rodríguez-Tecedor, S., Pérez-del-Notario, N., González-Sáiz, J. M. (2011). Recognition of volatile compounds as markers in geographical discrimination of Spanish extra virgin olive oils by chemometric analysis of non-specific chromatography volatile profiles. *Journal of Chromatography. A* 1218(3), 518–23.

502 Poiana, M.-A., Alexa, E., Munteanu, M.-F., Gligor, R., Moigradean, D., & Mateescu, C. (2015). Use of ATR-FTIR spectroscopy to detect the changes in extra virgin olive oil by adulteration with soybean oil and high temperature heat treatment. Open Chemistry, 13(1), 689–698.

503 Pollok, D., Melchert, H. U. (2004). Determination of alpha-Tocopherolquinone in Human Serum Samples by Liquid Chromatography with Fluorescence Detection and OnLine Post-Column Derivatization. *Journal of Chromatography A* 1056 (2), 257-262.

504 Pontikis, C. A., M. Loukas, M., G. Kousounis G. (1980). The use of biochemical markers to distinguish olive cultivars. *Journal of Horticultural Science and Biotechnology* 55,333– 343.

505 Poulli, K. I., Mousdis, G. a., & Georgiou, C. a. (2007). Rapid synchronous fluorescence method for virgin olive oil adulteration assessment. Food Chemistry, 105(1), 369–375.

506 Poulli, K. I., Mousdis, G. A., Georgiou, C. A. (2005). Classification of edible and lampante virgin olive oil based on synchronous fluorescence and total luminescence spectroscopy. *Analytica Chimica Acta* 542, 151–156.

507 Poulli, K.I., Mousdis, G.A., Georgiou, C.A. (2007). Rapid synchronous fluorescence method for virgin olive oil adulteration assessment. *Food Chemistry*. 105, 369-375.

508 Pannelli, G., Alfei, B., D Ambrosio, A., Rosati, S., Famiani, F. (2000). Varietà di Olivo in Umbria. Pliniana, Perugia, Italy. 98.

509 Pridham, J. B. (1960). Oligosaccharides and associated glycosidases in aspen tissues *Biochemistry Journal* (76), 13–20

510 Procter, H.R., Paessler, J. (1955). Leitfaden fiir gerbereichemische untersuchungen. S-78, Berlin (1901). Quoted In, Schmidt, O.T. Natiirilime gerbstoffe.ln, Moderne methoden der pfanzenanalyze. Paech, K., Tracey, M.V. (eds.) Springer-Verlag, Berlin, 517-548

511 Psomiadou, E., Tsimidou, M. (2002). Stability of virgin olive oil. 1. Autooxidation studies. *Journal of Agricultural and Food Chemistry* 50, 716–721

512 Rafalski, J.A., Tingey, S.V. (1993). Genetic diagnostics in plant breeding, RAPD, microsatellites and machines. *Trends in Genetics* 9, 275-279.

513 Rahmani, M., Saari-Csallany, A. (1998). Role of minor constituents in the photoxidation of virgin olive oil. *Journal of the American Oil Chemists Society* 75, 837–843

514 Rallo, P., Dorado, G., Martin, A. (2000). Development of Simple Sequence Repeats (SSRs) in Olive tree (*Olea europaea* L.). *Theoretical and Applied Genetics* 101, 984–989.

515 Ramadan, M.F., Moersel, J.T. (2006) Screening of the antiradical action of vegetable oils. *Journal of Food Composition and Analysis* 19, 838-842

516 Ramos, A.F., Santos, F.L. (2010). Yield and olive oil characteristics of a low–density orchard (cv. Cordovil) subjected to different irrigation regimes. *Agricultural Water Management* 97, 363–373.

517 Ranalli, A., Ferrante, M.L., De Mattia, G., Costantini, N. (1999). Analytical evaluation of virgin olive oil of first and second extraction. *Journal of Agricultural and Food Chemistry* 47, 417–424.

518 Ranalli, A., Modesti, G. (1999).Processingtechnologiesandbiotechnologies affect the composition of green and yellow lipochromes and the chromatic features of virgin olive oil, in PFT Sevilla, in Proceedings of the 1st International Congresson Pigmentsin Food Technology Seville, Spain.

519 Rao, C.V., Newmark, H.L. , Reddy , B.S. (1998) . Chemopreventive effect of squalene on colon cancer . *Carcinogenesis* 19 , 287 – 290 .

520 Ray, A. D., Major, H. (2010). Molecular ions and protonated molecules observed in the atmospheric solids analysis probe analysis of steroids, 174, 169–174

521 Reale, S., Doveri, S., Diaz, A., Angiolillo, A., Lucentini, L., Pilla, F., Martín, A., Donini, P., Lee, D. (2006). SNP-based markers for discriminating olive (*Olea europaea* L.) cultivars. *Genome* 49(9) 1193-1205.

522 Reiners, J., Grosch, W. (1998). Odorants of virgin olive oils with different flavor profiles. *Journal of Agricultural and Food Chemistry* 46(7), 2754–2763

523 Reiter, B., Lechner, M., Lorbeer, E., Aiccholz, R. (1999). Isolation and characterization of wax esters in fennel and caraway seed oils by SPE–GC. *Journal of High Resolution Chromatography* 22 (9) 514–520.

524 Rezzi, S., Axelson, D. E., Reniero, F., Mariani, C., & Guillou, C. (2005). Classification of olive oils using high throughput flow 1 H NMR fingerprinting with principal component analysis , linear discriminant analysis and probabilistic neural networks &, *552*, 13–24. https://doi.org/10.1016/j.aca.2005.07.057

525 Rice-Evans, C.A., Miller, N.J., Paganga, G. (1996) .Structure-antioxidant activity relationships of flavonoids and phenolic acids. *Free Radical Biology and Medicine* 20, 933–956.

526 Rigane, G., Salem, R.B., Sayadi, S., Bouaziz, M. (2011). Phenolic composition, isola- tion, and structure of a new deoxyloganic acid derivative from Dhokar and Gemri-Dhokar olive cultivars. *Journal of Food Science* 76, 965–973

527 Ripa, V., De Rose, F., Caravita, M. L., Parise, M. R. (2008). Qualitative evaluation of olive oils from new olive selections and effects of genotype and environment on oil quality. *Advances in Horticultural Science* 22, 95–103

528 Ripley, B. D. (1996). Pattern Recognition and Neural Networks. Cambridge University Press, Cambridge, NY.

529 Roberts, W. G., Gordon, M. H. (2003). Determination of the total antioxidant activity of fruits and vegetables by a liposome assay. *Journal of Agricultural of Food Chemistry*, 51, 1486–1493.

530 Roca, M., Gandul-Rojas, B., Gallardo-Guerrero, L., Minguez-Mosquera, M. I. (2003). Pigment parameters determining Spanish virgin olive oil authenticity, Stability during storage. *Journal of the American Oil Chemists Society* 80(12), 1237–1240.

531 Roca, M., Mínguez-Mosquera, I. (2001). Change in the natural ratio between chlorophylls and carotenoids in olive fruit during processing for virgin olive oil. *Journal of the American Oil Chemists' Society* 78, 133–138.

532 Rogers, S.O., Bendich, A.J. (1987). Ribosomal RNA genes in plants, variability in copy number and in the intergenic spacer. *Plant Molecular Biology* 9, 509–520.

533 Romani, A., Mulinacci, N., Pinelli, P., Vincieri, F., Cimato, A. (1996). Polyphenolic content in five tuscany cultivars of *Olea europaea* L. *Journal of Agricultural and Food Chemistry* 47, 964–967

534 Romero, M.P., Tovar, M.J., Ramo, T., Motilva, M.J. (2003). Effect of crop season on the composition of virgin olive oil with protected designation of origin "Les Garrigues". *Journal of the American Oil Chemists' Society* 80, 423-430.

535 Rotondi, A., Bendini, A., Cerratani, L., Mari, M., Lecker, G., Toschi, T.G. (2004). Effect of olive ripening degree on the oxidative stability and organoleptic proprieties of Cv Nostrana di Brisighella extra virgin olive oil. *Journal of Agricultural and Food Chemistry* 52,3649–54.

536 Roubos, K., Moustakas, M., Aravanopoulos, F.A. (2010). Molecular identification of Greek olive (*Olea europaea*) cultivars based on microsatellite loci. *Genetics and Molecular Research* 9, 1865-1876.

537 Rovellini, P., Cortesi, N., Fedeli, E. (1997). Analysis of flavonoids from *Olea europaea* by HPLC-UV and HPLC-electrospray-MS. *Rivista Italiana delle Sostanze Grasse* 74, 273-279.

538 Rugini, E., De Pace, C., Gutierrez-Pesce, P., Muleo, R. (2011). Wild Crop Relatives, Genomic and Breeding Resources, Temperate Fruits (Kole C, ed.). Springer-Verlag, Berlin, Heidelberg.

539 Ryan, D., Antolovich, M., Prenzler, P., Robards, K., Lavee, S. (2002). Biotransformations of phenolic compounds in *Olea europaea* L. *Scientia Horticulturae* 92, 147–176.

540 Sabir, S.M., Maqsood, H., Hayat, I., Khan, M.Q., Khaliq, A. Elemental and nutritional analysis of sea Buckthorn (Hippophae rhamnoides ssp. turkestanica) berries of Pakistani origin. J. Med. Food, v.8, n.4, p.518- 522, 2005.

541 Saimaru, H, Orihara, Y. (2010) Biosynthesis of acteoside in cultured cells of *Olea europaea*. *Journal of Natural Medicines* 64, 139–45.

542 Saimaru, H., Orihara, Y., Tansakul, P., Kang, Y.H., Shibuya, M., Ebizuka, Y. (2007). Production of triterpene acids by cell suspension cultures of *Olea europaea*. *Chemical and Pharmaceutical Bulletin* 55, 784–788

543 Sairam, R.K (1994). Effect of moisture stress on physiological activities of two contrasting wheat genotypes. *Indian Journal of Experimental Biology*., 32, 594–597.

544 Saitta, M., Lo Curto, S., Salvo, F., Di Bella, G., Dugo, G. (2002). Gas chromatographic–tandem mass spectrometric identification of phenolic compounds in Sicilian olive oils. *Analytica Chimica Acta* 466(2),335–44.

545 Sakouhi, F., Harrabi, S.,. Absalon, C., Sbei, K., Boukhchina, S., and Kallel, H. (2008). a- Tocopherol and fatty acid contents of some Tunisian table olives (Olea europaea L.): Changesin their composition during ripening and processing. Food Chemistry.108:833–839

546 Salvador, M.D., F. Aranda, S. Go´mez-Alonso, G. Fregapane. (2001). Cornicabra virgin olive oil, A study of five crops seasons. Composition, quality and oxidative stability. *Food Chemistry* 74,267–274.

547 Samaniego Sánchez, C., Troncoso González, A. M., García-Parrilla, M. C., Quesada Granados, J. J., López García de la Serrana, H., López Martínez, M. C. (2007). Different radical scavenging tests in virgin olive oil and their relation to the total phenol content. *Analytica Chimica Acta* 593(1), 103–107.

548 Sánchez de Medina, V., Priego-Capote, F., de Castro, M. D. L. (2015). Characterization of monovarietal virgin olive oils by phenols profiling. *Talanta* 132, 424–32.

549 Sanchez, A.M., Carmona, M., Ordoudi, S. A., Tsimidou, M. Z., Alonso,G. L. (2008). Kinetics of individual crocetin ester degradation in aqueous extracts of saffron (Crocus sativus L.) upon thermal treatment in the dark. *Journal of Agricultural and Food Chemistry*, 56, 1627–1637.

550 Sanchez, F.J., Manzanares, M., de Andres, E.F., Tenorio, J.L., Ayerbe, L. (1998) Turgor maintenance, osmotic adjustment and soluble sugar and praline accumulation in 49 pea cultivars in response to water stress. *Field Crops Research* 59, 225–235.

551 Sarri, V., Baldoni, L., Porceddu, A., Cultrera, N.G.M., Contento, A., Frediani, M., Belaj, A., Trujillo, I., Cionini, P.G. (2006). Microsatellite markers are powerful tools for discriminating among olive cultivars and assigning them to geographically defined populations. *Genome* 49, 1606–1615.

552 Sauci, Fahman, Kraut. (1994). Food Composition and Nutrition Tables, 5th edition.

553 Sawatsky, M. L., Clyde, M., & Meek, F. (2015). Partial Least Squares regression in the social sciences. The Quantitative Method for Psychology, 11(2), 52–62. Retrieved from http://www.tqmp.org/RegularArticles/vol11-2/p052/p052.pdf

554 Sawaya, A.C.H.F., Tomazela, D.M., Cunha, I.B.S., Bankova, V.S., Marcucci, M.C., Custodio, A.R., Eberlin, M.N. (2004) Electrospray ionization mass spectrometry fingerprinting of propolis. *Analyst* 129, 739.

555 Scarpati, M.L., Lo Scalzo, R., Vita, G., Gambacorta, A. (1996). Chemiotropic behaviour of female olive fly (*Bactrocera oleae Gmel*.) on *Olea europaea* L. *Journal of Chemical Ecology* 22, 1027–1036

556 Schafer-Schuchardt (1988) L'Oliva, La Grande Storia Di Un Piccolo Frutto Bari, Favia, Brossura editoriale illustrate

557 Schwartz S.J. and Von Elbe J.H. (1983). Kinetics of chlorophyll degradation to pyropheophytin in vegetables Journal of Food Science 48:1303-1306

558 Schwarz, K., Bertelsen, G., Nissen, L.R., Gardner, P.T., Heinonen, M.I., Hopia, A., Huynh-Ba, T., Lambelet, P., McPhail, D., Skibsted, L.H., Tijburg, L. (2000). Investigation of plant extracts for the protection of processed foods against lipid oxidation. Comparison of antioxidant assays based on radical scavenging, lipid oxidation and analysis of the principal antioxidant compounds. *European Food Research and Technology* 21, 319–328

559 Scott, S. M., James, D., Ali, Z., O'Hare, W. T. & Rowell, F. J. (2003). Total Luminescence Spectroscopy With Pattern Recognition for Classification of Edible Oils. Analyst, Vol.128, No.7, pp. 966-973, ISSN 0003-2654

560 Seddon, J. M., Ajani, U. A., Sperduto, R. D., Hiller, R., Blair, N., Burton, T. C. (1994). Dietary carotenoids, vitamin-A, vitamin-C, and vitamin-E, and advanced age-related macular degeneration. *Journal of the American Medical Association*, 272,1413–1420.

561 Sefc, K.M., Lopes, M.S., Mendonça, D., Rodrigues dos Santos, M., Câmara Machado, M.L., Câmara Machado, A. (2000). Identification of microsatellite loci in olive (*Olea europaea*) and their characterization in Italian and Iberian olive trees. *Molecular Ecology* 9, 1171–1173.

562 Sena, M. M., Poppi, R. J. (2004). N-way PLS applied to simultaneous spectrophotometric determination of acetylsalicylic acid, paracetamol and caffeine. Journal of Pharmaceutical and Biomedical Analysis, 34(1), 27–34.

563 Serani, A., Piacenti, D. (1992). Kinetics of pheophytin R photodecom- position in extra virgin olive oil. *Journal of the American Oil Chemists' Society* 69, 469-470.

564 Sergiel, I., Pohl, P.; Biesaga, M., Mironczyk, A. (2014). Suitability of Three-Dimensional Synchronous Fluorescence Spectroscopy for Fingerprint Analysis of Honey Samples with Reference to Their Phenolic Profiles. Food Chemistry 145, 319–326.

565 Servili, M., Esposto, S., Fabiani, R., Urbani, S., Taticchi, A., Mariucci, F., Selvaggini, R., Montedoro, G. F. (2009). Phenolic compounds in olive oil, antioxidant, health and organoleptic activities according to their chemical structure. *Inflammopharmacology* 17 76–84

566 Servili, M., Montedoro, G. (2002). Contribution of phenolic compounds to virgin olive oil quality. *European Journal of Lipid Science and Technology*, 104(9-10), 602–613.

567 Servili, M., Montedoro, G.F. (2002). Contribution of phenolic compounds to virgin olive oil quality. *European Journal of Lipid Science and Technology* 104,602– 613.

568 Servili, M., R. Selvaggini, S. Esposto, A. Taticchi, G.F. Montedoro, G. Morozzi. (2004). Health and sensory properties of virgin olive oil hydrophilic phenols, Agronomic and technological aspect of production that affect their occurence in the oil. *Journal of Chromatography A* 1054,113–127

569 Servili, M., S. Esposto, A. Taticchi, S. Urbani, I. Di Maio, B. Sordini, R. Selvaggini, G.F. Montedoro, and F. Angerosa. (2009). Volatile compounds of virgin olive oil: Their importance in the sensory quality. p. 45–77. In: L. Bertiand J. Maury (eds.), Advances in olive research

570 Sesli, M., Yeğenoğlu, E. D. (2010). Determination of the genetic relationships between wild olive (*Olea europaea* oleaster) varieties grown in the Aegean region. *Genetics and Molecular Research* 9(2), 884–90.

571 Sesli, M., Yegenoglu, E.D. (2009). Genetic analysis on wild olives by using RAPD markers. *African Journal of Agriculture Research* 8, 707-712.

572 Shaffer, R. E., Rose-Pehrsson, S. L., & McGill R. A. (1998). Multiway anaysis of preconcentrator-sampled surface acoustic wave chemical sensor array data. *Field Analytical Chemistry and Technology*, *2*, 179-192.

573 Shahidi, F., & Wanasundara, P. D. (1992). Phenolic Antioxidants. Critical Reviews in Food Science and Nutrition, 32, 67–103.

574 Shahriari, M., Omrani, A., Falahati-Anbaran, A., Ghareyazei, B., Nankali, A. (2008). Identification of Iranian olive cultivars by using RAPD and microsatellite markers. *Acta Horticulturae* 791, 109-115.

575 Shi, J., Le Maguer, M. (2000). Lycopene in tomatoes, Chemical and physical properties affected by food processing. *Critical Reviews in Food Science and Nutrition* 40,1–42.

576 Shibuya, M., Zhang, H., Endo, A., Shishikura, K., Kushiro, T., Ebizuka, Y. (1999). Two branches of the lupeol synthase gene in the molecular evolution of plant oxidosqualene cyclases. *European Journal of Biochemistry* 266,302–307

577 Sikorska E. (2008a). Metody fluorescencyjne w badniach żywno?ci, AE Poznań. ISBN 978-83- 7417-360-5, Poznań, Poland

578 Sikorska E.; Gliszczyńska-wigo A.; Khmelinskii I. & Sikorski M. (2005b). Synchronous Fluorescence Spectroscopy of Edible Vegetable Oils.

Quantification of Tocopherols. Journal of Agricultural and Food Chemistry, Vol.53, No.18, pp. 6988-6994, ISSN 1520-5118

579 Sikorska E.; Górecki T.; Khmelinskii I. V.; Sikorski M. & Kozioł J. (2005a). Classification of Edible Oils Using Synchronous Scanning Fluorescence Spectroscopy. Food Chemistry, Vol.89, No.2, pp. 217-225, ISSN 0308-8146

580 Sikorska E.; Khmelinskii I. V.; Sikorski M.; Caponio F.; Bilancia M. T.; Pasqualone A. & Gomes T. (2008b). Fluorescence Spectroscopy in Monitoring of Extra Virgin Olive Oil During Storage. International Journal of Food Science and Technology, Vol.43, No.1- 2, pp. 52-61, ISSN 1365-2621

581 Sikorska E.; Romaniuk A.; Khmelinskii I. V.; Herance R.; Bourdelande J. L.; Sikorski M. & Kozioł J. (2004). Characterization of Edible Oils Using Total Luminescence Spectroscopy. Journal of Fluorescence, Vol.14, No.1, pp. 25-35, ISSN 1573-4994

582 Sikorska, E., Go´recki, T., Silorski, M., Khmelinskii, I. V. M., Koziol, J. Classification of edible oils using synchronous scanning fluorescence spectroscopy. *Food Chemistry*2005, 89, 217- 225.

583 Sikorska, E., Khmelinskii, I., Sikorski, M. (2012). Analysis of olive oils by fluorescence spectroscopy, Methods and applications. In Boskou Dimitrios (Ed.), Agricultural and Biological Sciences – Olive Oil – Constituents, Quality, Health Properties and Bioconversions 63-88.

584 Sikorska, E., Romaniuk, A., Khmelinskii, I. V., Herance, R., Bourdelande, J. L., Silorski, M., Koziol, J. (2004).Characterization of edible oils using total luminescence spectroscopy. *Journal Fluorescence* 14,25-35.

585 Sikorska, E., Romaniuk, A., Khmelinskii, IV., Herance, R., Bourdelande, J.L., Sikorski, M., Koziol, J. (2004). Characterization of edible oils using total luminescence spectroscopy. *Journal of Fluorescence* 14, 25-35.

586 Sikwese, F.E. Duodu, K.G. (2007). Antioxidant Effect of A Crude Phenolic Extract from Sorghum Bran in Sunflower Oil in the Presence of Ferric Ions. *Food Chemistry*, 104, 324-331.

587 Siliani, S., Mattei, A., Benevieri Innocenti, L., Zanoni, B. (2006). Bitter taste and phenolic compounds in extra virgin olive oil, An empirical relationship. *Journal of Food Quality* 29, 431–441

588 Sillanpaa, M., Jansson, H. (1992) Status of Cadmium, Lead, Cobalt and Selenium in Soils and Plants of Thirty Countries, FAO, Gene`ve, Switzerland.

589 Singleton, V.L., Orthofer, R., Lamuela-Raventós, R.M., Lester, P. (1999). Analysis of total phenols and other oxidation substrates and antioxidants by means of Folin-Ciocalteu reagent. *Methods in Enzymology* 299,152-178.

590 Siquet, C., Paiva-Martins, F., Lima, J.L., Reis, S., Borges, F. (2006). Antioxidant profile of dihydroxy- and trihydroxyphenolic acids--a structure-activity relationship study. *Free Radical Research* 40(4),433-42.

591 Skevin, D., Rade, D.,Strucelj, D. (2003). The Influence of Variety and Harvest Time on the Bitter- ness and Phenolic Compounds of Olive Oil, *European Journal of Lipid Science and Technology* 105, 536-541.

592 Smirnoff, N. (1995) Antioxidant systems and plant response to the environment. In, Smirnoff V (Ed.), Environment and Plant Metabolism, Flexibility and Acclimation, BIOS Scientific Publishers, Oxford, UK.

593 Smith , T.J. (2000) . Squalene, potential chemopreventive agent. *Expert Opinion on Investigational Drugs* 9, 1841 – 1848.

594 Società Botanica Italiana (1894). Bullettino della Società Botanica Italiana. Firenze, Pellas.

595 Solari, M.E., Vernet, J.L. (1992). Late glacial and Holocene vegetation of the Corbières based on charcoal analysis at the Cova de L'Espérit (Salses, Pyrénées Orientales, France). *Review of Palaeobotany and Palynology* 71,111–120

596 Solinas, M. (1987). HRGC analysis of phenolic components in virgin olive oils in relation to the ripening and the variety of olives. *La Rivista Italiana delle Sostanze Grasse* 64, 255-262

597 Solinas, M. (1990). La qualita dell olio di oliva e i fattori che la influenzano. Atti Convegno Problematiche qualitative dell olio di oliva, Sassari, Italy, November 6, 1990. p. 23–56

598 Solinas, V., Deiana, S., Gessa, C., Pistidda, C., Rausa, R. (1996). Reduction of the Fe(III)- desferrioxamine-B complexes by caffeic acid, A reduction mechanism of bio- chemical significance. *Soil Biology and Biochemistry* 28,649-654.

599 Speek, A.J., Speek-Saichua, S., Schreurs, W.H.P. (1988). Total carotenoid and β-carotene contents of Thai vegetables and the effect of processing. Food Chemistry 27(4):245–57.

600 Spencer, C.M., Cai, Y., Martin, R., Gaffney, S.H., Goulding, P.N., Magnoloto, D., Lilley, T.H., Haslam, E. (1988). Polyphenol complexation. Some thoughts and observations. *Phytochemistry* 27,2397–2409.

601 Srilaong, V., Aiamlaor, S., Soontornwat, A., Shigyo, M., Yamauchi, N. (2011). UV-B irradiation retards chlorophyll degradation in lime (*Citrus latifolia Tan*.) fruit. *Postharvest Biology and Technology* 59, 110–112.

602 Stahl, W., Sies, H. (2003). Antioxidant activity of carotenoids. *Molecular Aspects of Medicine* 24(6),345-51.

603 Stalikas , C.D. (2007) . Extraction, separation, and detection methods for phenolic acids and flavonoids . *Journal of Separation Science* . 30, 3268 – 3295.

604 Steenson, D. F., & Min, D. B. (2000). Effects of β-carotene and lycopene thermal degradation products on the oxidative stability of soybean oil. JAOCS, Journal of the American Oil Chemists' Society, 77(11), 1153–1160.

605 Stepien, P., Klobus, G. (2006). Water relations and photosynthesis in Cucumis sativus L. leaves under salt stre*ss. Biologia Plantarum* 50(4), 610–616.

606 Stewart, C.R. (1981). Proline accumulation, Biochemical aspects. In, Paleg LG, Aspinall D (Eds), Physiology and Biochemistry of drought resistance in plants 243-251

607 Sung, J., Lee, J. (2010) Antioxidant and antiproliferative activities of grape seeds from different cultivars. *Food Science and Biotechnology* 19 (2),321-326.

608 Talhaoui, N., Gómez-Caravaca, A. M., León, L., De la Rosa, R., Segura-Carretero, A., Fernández-Gutiérrez, A. (2014). Determination of phenolic compounds of 'Sikitita' olive leaves by HPLC-DAD-TOF- MS. Comparison with its parents 'Arbequina' and 'Picual' olive leaves. *Journal of Agricultural and Food Chemistry* 58,28−34.

609 Talhaoui, N., Gómez-Caravaca, A. M., Roldán, C., León, L., De la Rosa, R., Fernández-Gutiérrez, A., and Segura-Carretero, A. (2015). Chemometric Analysis for the Evaluation of Phenolic Patterns in Olive Leaves from Six Cultivars at Different Growth Stages. *Journal of Agricultural and Food Chemistry.*

610 Tanksley, S.D., Orton, T.J. (1983). Isozymes in plant genetics and breeding. Elsevier, Amsterdam.

611 Tao, R., Sugiura, A. 1987. Cultivars identification of japanese persimmon by leaf isozymes. *HortScience* 22,932-935.

612 Tautz, D., Trick, M., Dover, G.A. (1986). Cryptic simplicity in DNA is a major source of genetic variation. *Nature* 322, 652-656.

613 Telesiński, A., Nowak, J., Smolik, B., Dubowska, A., Skrzypiec, N. (2008). Effect of soil salinity on activity of antioxidant enzymes and content of ascorbic acid and phenols in bean (*Pheseolus vulgaris* L.) plants. *Journal of Elementology* 13,401–409

614 Tena N., Garcia-Gonzalez, D. L., Aparicio R. (2009). Evaluation of Virgin Olive Oil Thermal Deterioration by Fluorescence Spectroscopy. *Journal of Agricultural and Food Chemistry* 57(22), 10505-10511.

615 Tena, N., Aparicio, R., & García-González, D. L. (2012). Chemical changes of thermoxidized virgin olive oil determined by excitation-emission fluorescence spectroscopy (EEFS). Food Research International, 45(1), 103–108.

616 Terao, J., Piskula, M. & Yao, Q (1994) Protective effect of epicatechin, epicatechin gallate, and quercetin on lipid peroxidation in phospholipid bilayers. Archives of Biochemistry and Biophysics, 308, 1, 278 – 284.

617 Terral, J.F., Arnold-Simard, G. (1996). Beginnings of olive cultivation in eastern Spain in relation to Holocene bioclimatic changes. *Quaternary Science Reviews* 46,176–185.

618 Theriault, A., Chao, J. T., Wang, Q., Gapor, A., Adeli, K. (1999). Tocotrienol, a review of its therapeutic potential. *Clinical Biochemistry* 32, 309-319.

619 Tovar, M. J., Motilva, M. J., & Romero, M. P. (2001). Changes in the phenolic composition of virgin olive oil from young trees (Olea europaea L. cv. Arbequina) grown under linear irrigation strategies. Journal of Agricultural and Food Chemistry, 49(11), 5502–5508.

620 Tovar, M. J., Motilva, M. J., Paz Romero, M. (2001). Changes in the phenolic composition of virgin olive oil from young trees (*Olea europaea* L. cv. Arbequina) grown under linear irrigation strategies, *Journal of Agricultural and Food Chemistry* 49, 5502–8.

621 Tovar, M.J., Motilva, M.J., Romero, M.P. (2001). Changes in the phenolic composition of virgin olive oil from young trees (*Olea europaea* L. cv. Arbequina) grown under linear irrigation strategies. *Journal of Agricultural and Food Chemistry* 49(11),5 502-585

622 Trujillo, I., L. Rallo, E.A. Carbonell, M.J. Asins. (1990). Isoenzymatic variability of olive cultivars according to their origin. *Acta Horticulturae* 286, 137–140.

623 Trump, D.H (1966) Skorba , A Neolithic Temple in Malta Society of Antiquaries of London

624 Tsai, S.Y., Huang, S.J., Mau, J.L. (2006) Antioxidant properties of hot water extracts from Agrocybe cylindracea. *Food Chemistry* 98, 868–875.

625 Tsao, R., Deng, Z. (2004). Separation procedures for naturally occurring antioxidant phytochemicals, *Journal of Chromatography B* 812, 85–99, 2004.

626 Tsimidou, M., Lytridou, M., Boskou, D., Pappa-Lousi, F., Kotsifaki, A., Petrakis C. (1996). On the determination of minor phenolic acids of virgin olive oil by RP-HPLC. *Grasas Aceites* 47,151–157.

627 Tulipani, S., Marzban, G., Herndl, A., Laimer, M., Mezzetti, B., Battino, M. (2011). Influence of environmental and genetic factors on health-related compounds in strawberry, *Food Chemistry* 124, 906–913

628 Tura, D., Gigliotti, C., Pedo, S., Failla, O., Bassi, D., Serraiocco, A. (2007). Influence of cultivar and site of cultivation on levels of lipophilic and hydrophilic antioxidants in virgin olive oils (*Olea* europea L.) and correlations with oxidative stability. *Scientia Horticulturae* 112, 108-119.

629 Twohig, M., Shockcor, J. P., Wilsfon, I. D., Nicholson, J. K., & Plumb, R. S. (2010). Use of an atmospheric solids analysis probe (ASAP) for high throughput screening of biological fluids: Preliminary applications on urine and bile. Journal of Proteome Research, 9(7), 3590–3597.

630 Uceda, M., & Hermoso, M. (2001). La calidad del aceite de oliva. En: “El cultivo del olivo”. Mundi-Prensa, Madrid, 588-614.

631 Uceda, M., Hermoso M. (1998) “La calidad del aceite de oliva,” in El Cultivo del Olivo,D.Barranco, R. Fernandez-Escobar, and L. Rallo, Eds., 547–572, Junta de Andalucıa EdicionesMundi Prensa,Madrid, Spain.

632 Uceda, M., Hermoso, M. (1998). La calidad del aceite de oliva. In: D. Barranco, R. Fernandez- Escobar, L. Rallo, editors. El Cultivo del Olivo.Madrid, Spain: Junta de Andalucıa Ediciones Mundi-Prensa. p 547–572.

633 Uceda, M., Hermoso, M., García-Ortiz, A., Jiménez, A., Beltran, G. (1999). Intraspecific Variation of Oil Contents and the Characteristics of Oils in Olive Cultivars. *Acta Horticulturae* 474(2), 659–652.

634 Umamaheswari, M., Asokkumar, K., Somasundaram, A., Sivashanmugam, T., Subhadradevi, V., Ravi, T.K. (2007). Xanthine oxidase inhibitory activity of some Indian medical plants. *Journal of Ethnopharmacology* 109, 547–551.

635 Uri, N. (1961). Mechanisms of autoxidation. In W. O. Lundberg (Ed.), Autoxidation and antioxidants. New York, John Wiley and Sons.

636 Van Boekel MAJS. (2000). Kinetic modelling in food science: a case study on chlorophyll degradation in olives. Journal of the Science of Food and Agriculture 80: 3-9.

637 Van de Voort F., Ismail A., Sedman J., Emo G. (1994) Monitoring the oxidation of edible oils by FTIR spectroscopy Journal of the American Oil Chemists' Society, 71, 243-253.

638 van der Sluis A, Dekker N, Van Boekel. (2005). Activity and concentration of polyphenolic antioxidants in apple juice, Stability during storage *Journal of Agricultural and Food Chemistry*53, 1073-1080

639 Vapnik, V.N., 1995. The nature of statistical learning theory, Cambridge University Press, New York.

640 Vapnik, V.N., 1998. Statistical Learning Theory, Wiley, New York. Venables,

641 Varzakas, T.H., Zakynthinos, G., Arapoglou, D.(2010). Fruit ripening in relationship to oil quality and some quality characteristics of the Greek olive cultivar koroneiki. Int *Journal of Food Science* 22,401–407

642 Vázquez-Roncero A., Janer del Valle C., aner del Valle M.L. (1973). Determinación de los polifenoles totales del aceite de oliva. *Grasas y Aceites*. 24, 350.

643 Verbruggen, N., Hermans, C. (2008). Proline accumulation in plants, a review. *Amino Acids* 35, 753 759.

644 Vichi, S., Castellote, A.I., Pizzale, L., Conte, L.S., Buxaderas, S., Lo ´pez Tamames, E. (2003). Analysis of virgin olive oil volatile compounds by headspace solid-phase microextraction coupled to gas chromatography with mass spectrometric and flame ionization detection. J*ournal of Chromatography A* 983, 19–33.

645 Vichi, S., Pizzale, L., Conte, L.S., Buxaderas, S., Lopez-Tamames, E. (2005). Simultaneous determination of volatile and semi-volatile aro- matic hydrocarbons in virgin olive oil by headspace solid-phase microextraction coupled to gas chromatography/mass spectrometry. *Journal of Chromatography A* 1090, 146 – 154 .

646 Vietina, M., Agrimonti, C., Marmiroli, M., Bonas, U., Marmiroli, N. (2011). Applicability of SSR markers to the traceability of monovarietal olive oils. *Journal of the Science of Food and Agriculture* 91(8), 1381–91.

647 Villano, D., Fernandez-Pachon, M.S., Troncoso, A.M., Garcia-Parilla., M.C. (2004) “The Antioxidant activity of wines determined by the ABTS method, influence of sample dilution and time”. *Talanta* 64, 501-509.

648 Vinha, A.F., Ferreres, F., Silva, B.M., Valentao, P., Gonçalves, A., Pereria J.A., Oliveira, M.B., Seabra, R.M., Andrade, P.B. (2005). Phenolic profiles of Portuquese olive fruits (*Olea europaea* L.), influences of cultivar and geographical origin. *Food Chemistry* 89,561–568.

649 Visioli, F., Galli, C. (1998). Olive oil phenols and their potential effects on human health. *Journal of Agricultural and Food Chemistry* 46,4292-4296.

650 Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeu, M. (1995). AFLP, a new technique for DNA fingerprinting. *Nucleic Acids Research* 23, 4407–4414

651 Vossen, P. (2007). Olive oil, history, production, and characteristics of the world’s classic oils. *HortScience* 42(5), 1093–1100.

652 W. N. Venables and B. D. Ripley. Modern Applied Statistics with S. Springer, New York, fourth edition, 2002. URL http://www.stats.ox.ac.uk/pub/MASS4. ISBN 0-387-95457-0. [p151]

653 Wall, M.E., Kelley, E.G. (1947) Determination and nature of leaf sterols. *Analytical Chemistry* 19(9),677-683.

654 Wallis, C., Eyles, A., Chorbadjian, R., McSpadden, Gardener, B., Hansen, R., Cipollini, D., Herms, D.A., Bonello, P. (2008). Systemic induction of phloem secondary metabolism and its relationship to resistance to a canker pathogen in Austrian pine. *New Phytologist* 177(3), 767–778

655 Wanasundara, P.K.J.P.D., Wanasundara, U.N., Shahidi, F. (1999). Changes in flax (*Linum usitatissimum L.*) Seed Lipids during Germination. *Journal of the American Oil Chemists Society* 76, 41-48.

656 Wang, T., Hicks, K.B., Moreau, R., 2002. Antioxidant activity of phytosterols, oryzanol and other phytosterols conjugates. *Journal of the American Oil Chemists Society* 79, 1201–1206.

657 Ward, K., Scarth, R., Daun, J. K., Thorsteinson, C. T. (1994). Comparison of High- Performance Liquid Chromatography and Spectrophotometry to Measure Chlorophyll in Canola Seed and Oil. Journal of the American Oil Chemists' Society, Vol.71, No.9, pp. 931-934, ISSN 0003-021X

658 Warner, K., Frankel, E.N. (1987). Effect of β-carotene on light stability of soybean oil. *Journal of the American Oil Chemists Society* 64, 213–218

659 Waterman, E., Lockwood, B. (2007). Active Components and Clinical Applications of Olive Oil. *Alternative medicine review* 12(4), 331–342

660 Weeden, N.F., Lamb., R.C. (1985). Identification of apple cultivars by isozyme phenotypes. *Journal of the American Society for Horticultural Science* 110, 509-515.

661 Welsh, J., McClelland., J. (19900. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* 18, 7213–7218.

662 Werner, R.A., RossmPNN, A., Schwarz, C., Bacher, A., Schmidt, H.L., Eisenreich, W. (2004). Biosynthesis of gallic acid in Rhus typhina, discrimination between alternative pathways from natural oxygen isotope abundance. *Phytochemistry*. 65, 2809–2813.

663 Wieland ,Weil, Über das Krötengift., 46, 3315 (1913).

664 Wiesman, Z., Avidan, N., Lavee, S., Quebedeaux, B. (1998). Molecular characterization of common olive varieties in Israel and the West bank using randomly amplified polymorphic DNA (RAPD) markers. *Journal of the American Society for Horticultural Science* 123, 837–841

665 Wiesman, Z., D. Itzhak, N. Ben Dom. (2004). Optimization of saline water level for sustainable Barnea olive and oil production in desert conditions. *Journal of the American Society for Horticultural Science* 100, 257–266.

666 Williams, G. K., A. R. Kubelik, K. L. Livak, J. A. Rafalaski, and S. V. Tingey. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18, 6531–6535.

667 Wilson, A. C., Carlson, S. S., White, T. J. (1977) Biochemical evolution. *PNNual Reviews in Biochemistry* 46, 573–639.

668 Wittrup, C. (2000). Comparison of chemometric methods for classification of fungal extracts based on rapid fluorescence spectroscopy. Journal of Chemometrics, 14(5–6): 765–776

669 Wittrup, C. 2000. Comparison of chemometric methods for classification of fungal extracts based on rapid fluorescence spectroscopy. *Journal of Chemometrics*, 14(5–6): 765–776.

670 Wold, S. (1994), PLS for Multivariate Linear Modelling , QSAR: Chemometric Methods in Molecular Design. Methods and Principals in Medicinal Chemistry.

671 Wold, S. (1994). PLS for Multivariate Linear Modelling , QSAR, Chemometric Methods in Molecular Design. *Methods and Principals in Medicinal Chemistry*

672 Wu, K.S., Tanksley, S.D. (1993). Abundance, polymorphism and genetic mapping of microsatellites in rice. *Molecular General Genetics* 241, 225-235.

673 Yamamoto, K., Niki, E. (1988). Interaction of α-tocopherol with iron, Antioxidant and prooxidant effects of α-tocopherol in the oxidation of lipids in aqueous disper- sions in the presence of iron. *Biochimica Biophysica Acta* 958,19-23.

674 Yamauchi, R., Tatsumu, Y., Asano, M., Kato, K., Ueno, Y. (1988). Effect of metal salts and fructose on the oxidation of methyl linoleate in emulsions. *Agricultural and Biological Chemistry* 52,849-850.

675 Yokozawa, T., Chen, C.P., Dong, E., Tanaka, T., Nonaka, G.I., Nishioka, I. (1998). Study on the inhibitory effect of tPNNins and flavonoids against the 1,1-diphenyl-2-picrylhydrazyl radical. *Biochemical Pharmacology* 56, 213–222.

676 Yorulmaz, A., Erinc, H., Tekin, A. (2013). Changes in Olive and Olive Oil Characteristics During Maturation *Journal of the American Oil Chemists' Society* 90, 647–658

677 Yousfi, K., Cert, R.M., García, J.M. (2006). Changes in quality and phenolic compounds of virgin olive oils during objectively described fruit maturation. *European Food Research and Technology* 223, 117-124.

678 Youssefi, O., Guido, F., Daoud, D., Mokhtar, Z. (2011). Effect of cultivar on minor components in Tunisia olive fruits cultivated in microclimate. *The Journal of Horticultural Science and Biotechnology* 3 (1), 13–20.

679 Zandomeneghi, M. (2006). Fluorescence spectroscopy excitation-emission methods combined with three-way analysis as a complementary technique for the characterization of olive oil. *Journal of Agricultural and Food Chemistry*. 54 (14) 5214-5215

680 Zandomeneghi, M., Zandomeneghi, G. (2005). Comment on Cluster Analysis Applied to the Exploratory Analysis of Commercial Spanish Olive Oils by Means of ExcitationEmission Fluorescence Spectroscopy. *Journal of Agricultural and Food Chemistry* 53(14), 5829-5830.

681 Zeder, M.A., Emshwiller, E., Smith, B.D. Bradley, D.G. (2006) Documenting domestication, the intersection of genetics and archaeology. *Trends in Genetics* 22, 139–155.

682 Zeiner, M., Juranovic-Cindric, I., Skevin, D. (2010). Characterization of extra virgin olive oils derived from the Croatian cultivar Oblica. *European Journal on Lipid Science and Technology* 112, 1248–1252

683 Zeiner, M., Steffan, I., Cindric, I. J. (2005). Determination of Trace Elements in Olive Oil by ICP-AES and ETA-AAS, A Pilot Study on The Geographical Characterization. *Microchemical Journal* 81, 171-176.

684 Zhao, D., Kim, S. M., Pan, C. H., Chung, D. (2014). Effects of heating, aerial exposure and illumination on stability of fucoxanthin in canola oil. *Food Chemistry* 145, 505–513

685 Zheng, H., Lu, H. F. (2011). Use of kinetic, Weibull and PLSR models to predict the retention of ascorbic acid, total phenols and antioxidant activity during storage of pasteurized pineapple juice. *Food Science Technology* 44, 1273–1281

686 Zobayed S, Afreen F, Kozai T (2005). Temperature stress can alter the photosynthetic efficiency and secondary metabolite concentrations in St. John's Wort. *Plant Physiology Biochemistry* 43, 977–984

687 Zohary, D. (1994). The wild genetic resources of the cultivated olive. *Acta Horticulturae* 356, 62-65.

688 Zunin, P., Boggia , R., Lanteri , S., Leardi, R., De Andreis , R., Evangelisti , F. (2004) . Direct thermal extraction and gas chromatographic-mass spectrometric determination of volatile compounds of extra-virgin olive oils. *Journal of Chromatography A* 1023 , 271 – 276 .

689 Zunin, P., Boggia, R., Salvadeo, P., Evangelisti, F. (2005). Geographical traceability of West Liguria extravirgin olive oils by the analysis of volatile terpenoid hydrocarbons. *Journal of Chromatography A* 1089 , 243- 249

690 Zupan, J. , Gasteiger, J., (1999). Neural Networks in Chemistry and Drug Design , 2nd edn . Wiley-VCH , Weinheim

691 Zupan, J., Novic, M. , Li , X. , Gasteiger, J., (1994) . Classification of multicomponent analytical data of olive oils using different neural networks. *Analytica Chimica Acta* 292 , 219 – 234

692 Zurada, J.M. (1992). Introduction to Articial Neural Systems, West Publishing Company.