European Research Studies Journal
Volume XXIV, Issue 4B, 2021

pp. 548-569

Typology and Development of Local Administrative Units: Spatial Discriminant Analysis

Submitted 05/10/21, 1st revision 25/10/21, 2nd revision 17/11/21, accepted 30/11/21

Jacek Batóg[1], Barbara Batóg[2]

Abstract:

Purpose: *The aim is the evaluation of the consistency of typology of local administrative units (LAU) with their level of socio-economic development. The authors identified also the impact of the incorporation of spatial relations on the stability of such consistency.*
Design/Methodology/Approach: *The spatial linear discriminant analysis with various weights matrices has been applied. Introducing the spatial relationship between observations in different locations to the classical classification method allows to identify non-measurable spatial factors differentiating different locations as well as the directions of their development.*
Findings: *The results confirmed a high discrepancy between the administrative type and the level of development of Polish communes in 2006-2018. The higher compliance of official typology of examined units with their actual level of development was found when spatial relations were incorporated. The obtained results indicated the marginalisation of medium-sized cities, and underdevelopment of urban-rural communes. The following predictors, new dwellings, employment rate, population density, and registered entities, had the highest power of discrimination in the classification process.*
Practical Implications: *Proposed models could have the practical application for determining the actual type of local administrative unit and designing regional policy. Especially in the case when some intervention is needed in the area of financial external support. Such policy may reduce disparities between the levels of development of the various regions, especially addressing the problems of less developed areas.*
Originality/Value: *This research presents a concept of spatial linear discriminant analysis with a spatial component of regional structure integrated in form of spatial weights matrices. The proposed method was applied to identify the meaning of neighbouring regions characteristics in classification of territorial units, as well as to indicate the group of Polish voivodeships characterized by a high and low degree of correspondence between the actual and expected type of the territorial unit.*

Keywords: *Regional typology, local administrative units, discriminant spatial analysis.*

JEL Codes: *R11, R12, C21, C38.*

Paper Type: *Research study.*

[1]*Institute of Economics and Finance, University of Szczecin https://orcid.org/0000-0003-1413-7692, e-mail: jacek.batog@usz.edu.pl*
[2]*Institute of Economics and Finance, University of Szczecin https://orcid.org/0000-0001-9236-7405, e-mail: barbara.batog@usz.edu.pl*

1. Introduction

The heterogeneity of development among local government units is usually much higher than that occurring at the regional level. The areas most threatened by a permanent social and economic marginalisation are clusters of rural communes and functionally connected small towns, characterised by a peripheral location relative to the largest agglomerations, poor transport accessibility, low human capital and underinvestment in public infrastructure. The areas with a high concentration of development problems are also medium-sized towns losing their socio-economic and administrative functions featuring poor territorial accessibility, population outflow and mismatch between demand and supply on the labour market. A challenge for the regional policy is to reduce the disproportions, in particular in the peripheral areas, located far from large urban centres and at the borders of regions.

The regional policy implemented in Poland since 2010 has reflected the polarisation-diffusion model and focused development activities primarily on the largest agglomerations, assuming that the development impulses arising there will spread to less-developed neighbouring areas. Unfortunately, this type of reasoning has not worked in practice, increasing disproportions in the level of socio-economic development of both urban and rural areas. The currently implemented policy emphasises sustainable development of the whole country, based inter alia on targeted support for specific problem areas, taking into account their specific nature. This requires a correct diagnosis of the potential of particular territories, to determine the areas which should be supported to prevent their further marginalisation (Krajowa …, 2019).

The theory of regional socio-economic development assumes that regional heterogeneity depends on the long-term impact of social, economic, cultural, and political factors (Dyjach, 2013). Such disparities may become a barrier to maintaining the dynamic development of the whole country or the specific region (Adamowicz and Pyra, 2019). Measurement of regional socio-economic performance and identification of regional typologies play a crucial role in research as well as in policy-making when proper typology of regions allows to allocate resources across heterogeneous spaces and define policy priorities (Greco *et al.*, 2018; Latuconsina *et al.*, 2018). According to Bajracharya and Hastings (2018), identification of regional growth patterns and economic development are key attributes of a framework that has generally resulted in acceptable regional outcomes. The observed spatial heterogeneity of socio-economic development not only does not disappear but in many countries tends to worsen despite the use of many instruments supporting the development of peripheral regions (Wójcik and Herbst, 2011; Filipiak and Tarczyńska-Łuniewska, 2020).

At the same time, a regularity is visible in the concentration of regions with the same level of development, which, combined with reduced commuting between regions, causes strong stability of such structures. For instance, Modai-Snir and van Ham

(2018) found that for Tel-Aviv Metropolitan Area the effect of increasing inequality simply replicates the spatial pattern of disparities at the starting point, so high-income neighbourhoods become richer, while low-income neighbourhoods become poorer. It should be noted that living in an urban area in comparison to living in a rural area or small town has more or less advantages depending on the level of development of the country concerned. Current studies indicate that lower risk-of-poverty and social exclusion rates characterise urban areas in less developed European countries (Report …, 2011).

The regional typification process includes the selection of the algorithm used to define the boundaries between types, and the selection of the input data for the algorithm. Such thresholds as well as regional features can be subjective, complex, and may vary by geographical context (Fiaschetti *et al.,* 2021). Current European regional typologies are mostly one-dimensional, based predominantly on administrative data and unable to recognise regional urban-rural gradients (Van Eupen *et al.,* 2012; Holl, 2017; Wandl *et al*., 2014), especially as the meaning of rurality varies depending on the context (Uchida and Nelson, 2009). To similar conclusions in his work came Jacob (2019) who applied an exploratory analysis to the 95 districts of Austria, showing that regions that are either central or peripheral are the exception.

Being based on the local concentration of the population, such typology cannot take into account the presence of economic agglomerations if they occur in neighbouring regions (Brezzi *et al*., 2011). Various authors suggest that a binary urban-rural division is a 'fundamental oversimplification', as their different functions do not produce a fixed barrier delineating the urban and the rural (Partridge *et al*., 2007), and underline that one of the main factors of rural development is the distance to a nearby urban core, the relationship of a rural region to the functional area of the urban core, and the extent to which the rural region offers the potential for high rates of commuting (Gray, 2014).

Laurin *et al.* (2020) examined 98 municipal Canadian counties defining rurality and selecting a set of spatial and industrial variables. They indicated large heterogeneity across rural regions, while the rural concept is not enough to categorize the economic situation of "non-urban" counties. Similar conclusions were drawn by Copus (2015) within the proposition called "new rural economy". A partial solution to the difficulties of regional delimitation may be an inclusion of functional typologies of regions, although Cörvers *et al.* (2009) observed that the numerous functional delimitations are only useful for policy-making if they outperform the administrative delimitations regarding specific relevant indicators. Another proposal could be based on a harmonised methodology to facilitate international regional statistical comparisons and classification along an urban-rural continuum (Applying the Degree …, 2021).

The division of communes into three types – rural, urban-rural and urban – existing in Poland since 1990 was based on urban, administrative, demographic, historical, and geographical criteria. Rural municipalities have no towns on their territory. Urban-rural gminas include both towns within their administrative boundaries and areas outside these towns. However, the borders of urban municipalities coincide with the border of the city-forming the municipality.

The difficulty in planning and implementing regional policies, as well as in distribution and allocation of resources, especially those dedicated to the local level, lies in the discrepancy between the official (administrative) type of local unit and its real character, which depends on the level of development.

The level of development of a local government unit should correspond to its administrative type, as this type depends on its degree of urbanisation. In this case, a two-pronged research problem arises. The first is practical and is related to the assessment of whether an increase in the degree of urbanisation reflected in the type of local government unit is associated with its higher level of development. The second one is methodological and is connected with testing whether taking into account the resources and potential of neighbouring units changes the observed relation between administrative type of local territory and its level of development, with particular consideration given to the impact of how the spatial range of these relations is determined.

The compatibility "type of LAU – LAU level of development" will be assessed by verifying the research hypothesis according to which communes with higher administrative status are not always characterised by the level of socio-economic development warranting their label, so we can observe misclassification errors to some extent.

The current research is an extension of the paper presented in Batóg and Batóg (2020). The main differences between these two studies are the increased size of the research sample, which now includes all Polish communes, the increase in the time-frequency of the sample, which now includes 5 years considered at three-year intervals in the period 2006-2018, and the use of three additional types of spatial weights matrices, which allow an analysis of the robustness of the results obtained to the way spatial relationships are determined within individual voivodships (NUTS, 2).

The main goal of the study is an assessment of the compatibility of the LAU types described by the label given in the administrative delimitation process with their level of socio-economic development in the long term. Additionally, the Authors will identify the impact of the incorporation of spatial relations on the stability of such consistency. The rest of the paper is organized as follows. In the next part, the applied methodology was presented. Data characteristics and received results can be

found in the third section of the study, while the last segment presents concluding remarks and potential directions for future research.

2. Methodology

A variety of research approaches are used in the classification of regional units, for instance, the affinity propagation method proposed by Fiaschetti *et al.* (2021). Data mining methods, such as k-means or Ward's algorithm and other cluster analysis algorithms are relatively often used with regional typologies (Hedlund, 2016; Latuconsina *et al.*, 2018). Many authors apply also Linear Discriminant Analysis (LDA) (Jaba *et al.*, 2006; Batóg and Batóg, 2019; El-Hanjouri and Hamad, 2015). One of the crucial advantages of this method over other algorithms is that we can use it to identify factors with the highest impact on results of classification (discriminant power) among a set of variables X_1, …, X_p. Discriminant function is represented by the linear combinations of the discriminant variables and their coefficients β (Tacq, 2007):

$$Y_j = \beta_0 + \beta_1 x_{1j} + \cdots + \beta_p x_{pj} \tag{1}$$

where:

Y_j – value of a discriminant function for object *j*,
x_{ij} – value of an *i*th discriminant variable for object *j*,
β_i – parameter of discriminant function,
$i = 1, \ldots, p$, p – number of a discriminant variable,
$j = 1, \ldots, n$, n – number of an object.

Estimates of structural parameters β are obtained assuming that their values maximize the ratio of between-group variance B to within-group variance V (McLachlan, 2004):

$$\hat{\beta} = V^{-1}B \tag{2}$$

solving the system of equations:

$$(V^{-1}B - \lambda I)\hat{\beta} = 0 \tag{3}$$

where λ is an eigenvalue, using the characteristic equation:

$$\det(V^{-1}B - \lambda I) \tag{4}$$

Many studies point out (Fiaschetti *et al.*, 2021; Todtling and Trippl, 2005) that regional analysis and classification could be enhanced by integrating a spatial component in which regional typologies are not based just on endogenous characteristics, but also on the characteristics of neighbouring regions.

Determining the spatial relationship between observations in different locations we can identify non-measurable spatial factors differentiating the examined phenomenon between locations (Kopczewska, 2021). Many authors underline also that only the spatial aspects, i.e., regional disparities and spans, indicate socio-economic features that determine the directions of further development and the ability to improve the competitiveness of regions (Malina, 2004). Kopczewska (2008), studying 84 regions from Poland, the Czech Republic, Slovakia, Hungary, Germany, and Austria in 1995-2002, found that growth of GDP in neighbouring regions by 1 percentage point causes an increase in GDP in the studied region by 0.35 percentage point. Many authors underline also that the economic resilience of a specific region is associated with the resilience of nearby regions (Ezcurra and Rios, 2019; Pontarollo and Serpieri, 2018).

It is also worth noting that the choice of the data aggregation level, defining the analysed regional structure, is crucial for the results of the regional analysis. Some authors indicate that modelling performed on more aggregated data may result in absorption of specific effects and averaging of results with smoothing of outliers (Kopczewska, 2008). Other authors emphasize that as the size of the unit under study increases, linkages to places elsewhere become less important compared with the size and scope of activities contained in the area under study (Bosworth and Venhorst, 2018). These observations were the main reason for choosing communes as Local Administrative Units (LAU) under study, which, as Kopczewska (2011) argues, due to their small area and low population, are relatively homogeneous from a socio-economic point of view, and their dense network allows to assess the impact of distance from the regional core and other communes on the level of this development and the processes of clustering and spatial interactions.

The spatial factor can be introduced into LDA in different ways. In contrast to other authors, who for this purpose modify *a priori* or *a posteriori* probabilities of class membership (e.g. Cutillo and Amato, 2008; Steele and Redmond, 2001), the current proposal is based on the application of spatial weights matrix W:

$$Y = X\beta + WX\theta \tag{5}$$

where:

Y – vector of values of discriminant function,
X – matrix of values of discriminant variables, $n \times p$,
W – spatial weights matrix, $n \times n$,
β – vector of non-spatial parameters, $p \times 1$,
θ – vector of spatial parameters, $p \times 1$.

The key issue, when introducing spatial relationships, is to determine the neighbourhood structure that defines the spatial weights matrix. The construction of matrix W results from the assumptions made about the interactions between the studied regions (Kopczewska, 2008). In the literature, there are many proposals on

how matrix W can be built (e.g., Anselin, 1998; Abreu *et al.*, 2004). In this study, we use four kinds of spatial weights matrices.

The first one W_1 called connectivity matrix is based on common borders, when we assume that if two different spatial units have a common border then the elements of spatial weights matrix w_{ij} equal 1 and 0 otherwise:

$$w_{ij} = \begin{cases} 1 & \text{if unit } j \text{ shares a common border with unit } i \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

The second proposition W_2 referred to as distance-based matrix (Fischer and Wang, 2011) assumes that two spatial units i and j are neighbours when the distance d_{ij} between their centroids is less than a given value d:

$$w_{ij} = \begin{cases} 1 & \text{if } d_{ij} < d \ (d > 0) \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

The third matrix W_3 is based on the assumption that we limit the neighbourhood to the k-nearest neighbours:

$$w_{ij} = \begin{cases} 1 & \text{if centroid of } j \text{ is one of the } k \text{ nearest centroids of } i \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

To construct the last matrix W_4 we have to choose *a priori* value d seen in formula (7) in the way that every spatial unit has at least one neighbour:

$$d = \max_{i,j} d_{ij} \tag{9}$$

All spatial weighting matrices were generated using the programme *R* with the assumptions for W_2: $d = 30$ km and W_3: $k = 5$. All other calculations were made using the package STATISTICA 13.1.

3. Data and Results

The selection of data for regional classification is non-trivial because chosen variables define and describe studied regions (Fiaschetti *et al.*, 2021) and a framework we use to assess regional performance is a crucial factor in designing and evaluating regional policy (Greco *et al.*, 2018). Variables used in the analyses of socio-economic development represent diverse concepts and the final set depends also on the availability of statistical data (Nigohosyan and Vutsova, 2018). As it was pointed out before Polish communes are classified as urban (U), urban-rural (UR) and rural (R). The typification of studied 2478 communes localised in 16 voivodships according to their socio-economic development was done in five years

2006, 2009, 2012, 2015 and 2018. The set of diagnostic variables includes 9 variables covering several aspects:

- X_1 - own revenue *per capita*,
- X_2 - national economy entities registered *per* 1000 population,
- X_3 - non-working age population *per* 100 persons of working age,
- X_4 - useful floor area of dwellings completed *per* 1000 population,
- X_5 - investment property expenditure *per capita*,
- X_6 - natural increase *per* 1000 population,
- X_7 - employed persons *per* 1000 population,
- X_8 - consumption of water from water supply systems in households *per capita*,
- X_9 - population density.

The source of all statistical data was the Local Data Bank provided by Statistics Poland. There was assumed that all variables have the same impact on the level of development and no weighing system was introduced. The availability of public statistical data at the level of communes is significantly limited. For this reason, the above set of variables does not include factors describing the state of the environment or expenditures on its protection, as well as variables characterizing the sectoral structure of employment and the intensity of migration.

The latter factor is related to different places of residence and work, and determined by urban-rural interdependences and may have a significant impact on regional resilience across the urban-rural hierarchy (Bosworth and Venhorst, 2018; Silva and Ferreira-Lopes, 2014). The latter, as well as statistically significant spatial patterns of regional resilience, was found by Giannakis and Bruggeman (2020) who used multilevel logistic and multinomial regression models for the European Union NUTS-3 regions.

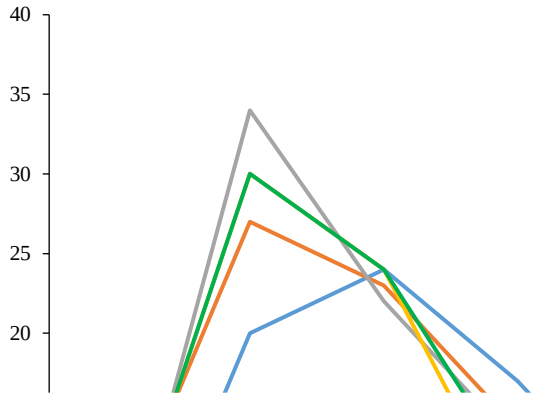
In the process of verifying the research hypothesis according to which we observe a relatively strong incompatibility between the administrative status of communes in Poland nowadays, established more than 30 years ago, and their current level of socio-economic development, 400 discriminant functions were constructed, covering 5 years, 16 voivodships, based on one classical (called W_0) and four different spatial LDA models (called respectively according to the type of spatial weight matrix W_1, W_2, W_3 and W_4). The vast majority of the models obtained were of high quality, as evidenced by the low values of the Wilks' λ statistic (Equation 10, and Figure 1):

$$\Lambda = \prod_{i=k+1}^{q} \frac{1}{1+\lambda_i} \qquad \Lambda \in \langle 0; 1 \rangle \tag{10}$$

where:
q – maximum number of discriminant functions,
λ_i – *i*th eigenvalue.

Due to the large number of calculations performed, the empirical part presents detailed results only for selected voivodships and years. The first concerns detailed results of the discriminant analysis obtained for Opolskie voivodship in 2018. This was followed by a presentation of two voivodships with the highest and lowest classification quality, Lubelskie and Wielkopolskie. The empirical part concludes with a synthesis of all the individual results, which provide a basis for identifying existing regularities, assessing their stability over time, and formulating general conclusions.

Figure 1. *Distributions of Wilks' λ for all analysed variants – all years and voivodships*



Source: *Own elaboration.*

Table 1 presents canonical discriminant functions obtained in classical LDA variant without spatial weights (W_0) for all studied years and basic measures of their quality received for Opolskie voivodship characterised by the highest improvement of classification when the spatial LDA was applied. Only root 1 is presented because the first discriminant function explained in all cases a significant proportion of the variance – almost 90%.

Table 1. *Estimation results for classical LDA (root 1, W_0) – Opolskie voivodship*

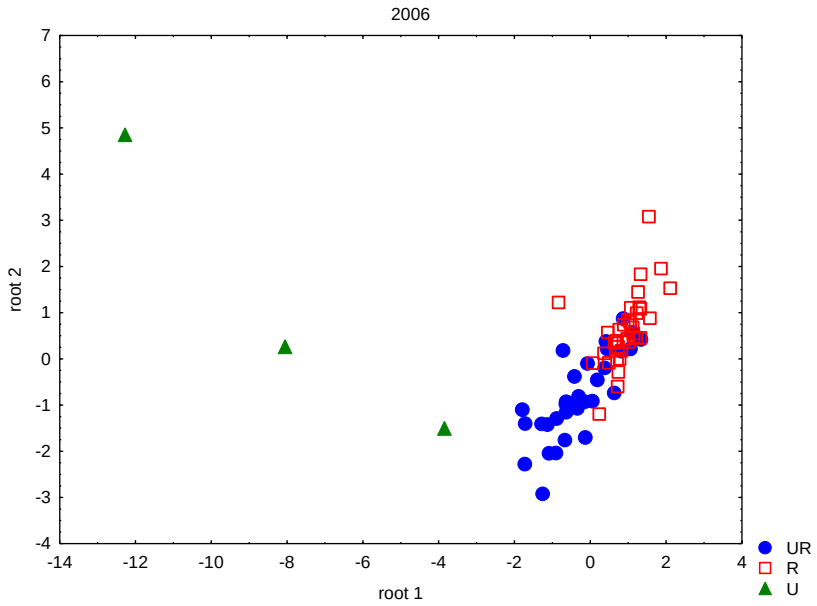
Variable	2006		2009		2012		2015		2018	
	Standardized coefficients	Partial Wilks' λ	Standardized coefficients	Partial Wilks' λ	Standardized coefficients	Partial Wilks' λ	Standardized coefficients	Partial Wilks' λ	Standardized coefficients	Partial Wilks' λ
X_1	-0.185	0.971	0.197	0.984	0.388	0.942	0.286	0.937	-0.333	0.927
X_2	-0.285	0.953	0.389	0.916	0.386	0.926	**0.446**	0.894	**-0.496**	0.891
X_3	-0.254	0.941	0.152	0.983	0.101	0.993	0.084	0.995	-0.067	0.968
X_4	0.172	0.968	**-0.468**	0.818	**-0.554**	0.786	**-0.489**	0.818	**0.584**	0.790
X_5	0.191	0.975	0.177	0.977	-0.229	0.932	0.070	0.915	-0.086	0.984
X_6	-0.048	0.991	-0.086	0.990	0.005	0.998	-0.001	0.996	0.007	0.983
X_7	**-0.402**	0.783	**0.350**	0.880	0.183	0.908	**0.100**	0.844	**-0.430**	0.818
X_8	-0.159	0.977	0.160	0.968	0.075	0.976	-0.154	0.983	0.126	0.976
X_9	**-0.801**	0.483	**0.732**	0.470	**0.754**	0.484	**0.807**	0.465	**-0.764**	0.507
Wilk's λ	0.144		0.140		0.150		0.155		0.165	

χ^2	123.956 (p = 0.000)	125.807 (p = 0.000)	121.290 (p = 0.000)	119.406 (p = 0.000)	115.165 (p = 0.000)
F(18, 120)	10.872 (p = 0.000)	11.147 (p = 0.000)	10.530 (p = 0.000)	10.278 (p = 0.000)	9.726 (p = 0.000)
Eigenvalue	3.376	3.787	3.576	3.320	3.285
Explained variance (%)	0.852	0.885	0.887	0.870	0.889

Source: *Own elaboration.*

The first discriminant function (root 1) discriminates mostly urban communes from urban-rural and rural communes (Figure 2). In all analysed years we can observe a high ability to discriminate objects, measured by the absolute value of structural parameters of discriminant functions, in the case of variables X_4 (useful floor area of dwellings completed *per* 1000 population), X_7 (employed persons *per* 1000 population), and X_9 (population density). In the last two years, the variable X_2 (national economy entities registered *per* 1000 population) joined the set of the most important predictors.

Figure 2. *Communes in discriminant space in 2006 for classical LDA (W_0) – Opolskie voivodship*



Source: *Own elaboration.*

Table 2 presents quality measures of the first canonical discriminant function obtained in spatial LDA (when matrix W_1 was used) for Opolskie voivodship in all the surveyed years. All spatial models are characterised by higher quality than the classical LDA functions.

Table 2. *Estimation results for spatial LDA (all years, root 1, W_1) – Opolskie voivodship*

Variable	2006		2009		2012		2015		2018	
	Standardized coefficients	Partial Wilk's λ	Standardized coefficients	Partial Wilk's λ	Standardized coefficients	Partial Wilk's λ	Standardized coefficients	Partial Wilk's λ	Standardized coefficients	Partial Wilk's λ
X_1	-0.081	0.960	-0.311	0.972	**-0.643**	0.874	-0.471	0.920	0.386	0.908
X_2	**-0.560**	0.878	**-0.794**	0.808	**-0.444**	0.874	**-0.770**	0.775	**0.846**	0.789
X_3	-0.418	0.911	-0.181	0.985	-0.332	0.937	-0.163	0.975	-0.003	0.989
X_4	0.411	0.941	0.531	0.902	0.396	0.891	0.569	0.904	**-0.595**	0.876
X_5	0.015	1.000	-0.228	0.969	0.314	0.935	**-0.091**	0.864	0.337	0.937
X_6	-0.092	0.971	0.045	0.988	0.105	0.966	0.026	0.980	-0.118	0.958
X_7	**-0.368**	0.843	-0.196	0.977	0.043	0.962	0.202	0.899	0.213	0.930
X_8	-0.130	0.988	-0.181	0.965	0.120	0.981	0.267	0.964	-0.120	0.989
X_9	**-0.780**	0.580	**-0.719**	0.627	**-0.785**	0.548	**-0.823**	0.620	**0.768**	0.618
WX_1	0.232	0.964	0.145	0.983	-0.168	0.991	0.024	0.975	0.318	0.938
WX_2	0.141	0.970	0.345	0.956	**0.908**	0.838	0.613	0.917	**-0.858**	0.856
WX_3	0.046	0.999	0.107	0.988	-0.251	0.983	-0.021	0.990	0.249	0.928
WX_4	0.026	0.976	0.302	0.940	**-0.579**	0.783	0.085	0.923	0.037	0.984
WX_5	-0.636	0.929	-0.361	0.954	0.268	0.973	-0.024	0.983	-0.115	0.984
WX_6	0.183	0.947	0.087	0.980	0.147	0.990	-0.242	0.969	0.278	0.972
WX_7	-0.352	0.924	-0.402	0.909	-0.401	0.935	-0.118	0.929	-0.059	0.999
WX_8	0.300	0.962	0.137	0.990	0.223	0.930	-0.113	0.960	0.076	0.969
WX_9	0.205	0.982	0.032	0.998	-0.014	0.942	-0.232	0.971	-0.100	0.971
Wilk's λ	0.096		0.094		0.083		0.094		0.109	
χ^2	139.224 (p = 0.000)		140.867 (p = 0.000)		148.037 (p = 0.000)		140.943 (p = 0.000)		131.825 (p = 0.000)	
F(18, 120)	6.295 (p = 0.000)		6.422 (p = 0.000)		6.669 (p = 0.000)		6.428 (p = 0.000)		5.745 (p = 0.000)	
Eigenvalue	4.451		5.038		4.730		4.205		4.290	
Explained variance (%)	0.831		0.868		0.811		0.800		0.854	

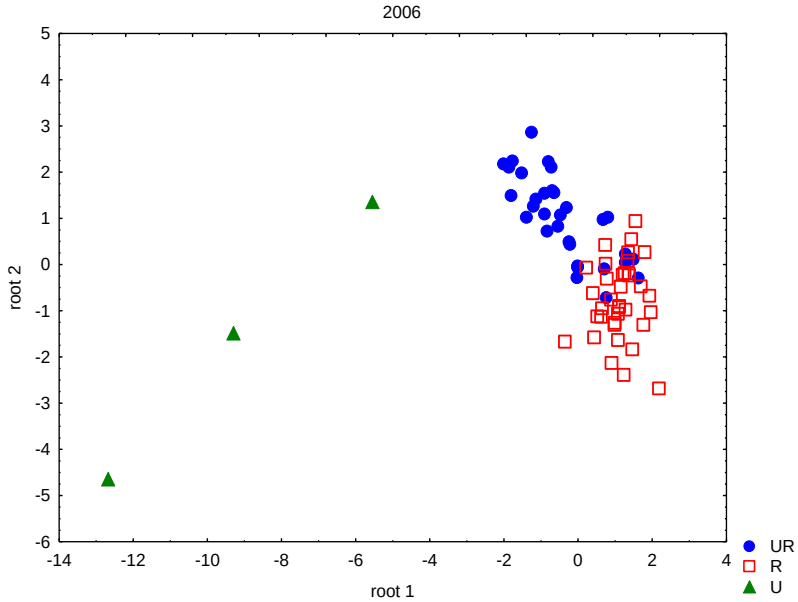
Source: *Own elaboration.*

In comparison to the classical variant of LDA, the importance of the variable X_2 becomes more apparent in the spatial models and the effect of the variables X_4 and X_7 on the classification results disappears. The scatter of points representing classified objects in a spatial variant of LDA is shown in Figure 3. We can observe a smaller common part in which both rural and urban-rural communes are located relating to the classic variant.

The clearest way to evaluate the positive influence of introducing spatial effects into the performed classification is to compare the quality of classification in both considered variants: classical and spatial. The results of such a comparison for all types of spatial weights matrices and all examined years for the Opolskie voivodship are presented in Table 3. In the classic variant, the lowest consistency between the administrative status of communes and their level of development is observed for urban communes (66.7%) and the highest for rural communes – except for 2009. It is worth noting that the incorporation of spatial relationships results in completely

correct recognition of urban communes by the models used – except for the W_3 matrix in 2009.

Figure 3. *Communes in discriminant space in 2006 for spatial LDA (W_1) – Opolskie voivodship*



Source: *Own elaboration.*

Table 3. *The classification quality (percentage of correctly classified communes) for classical and spatial LDA (all years) – Opolskie voivodship*

2006					
Unit	W_0	W_1	W_2	W_3	W_4
U	66.67	100.00	100.00	100.00	100.00
UR	75.00	78.13	81.25	75.00	81.25
R	91.67	100.00	97.22	100.00	97.22
Total	83.10	90.14	90.14	88.73	90.14
2009					
Unit	W_0	W_1	W_2	W_3	W_4
U	66.67	100.00	100.00	66.67	100.00
UR	81.25	87.50	84.38	87.50	90.63
R	80.56	91.67	91.67	94.44	88.89
Total	80.28	90.14	88.73	90.14	90.14
2012					
Unit	W_0	W_1	W_2	W_3	W_4
U	66.67	100.00	100.00	100.00	100.00
UR	78.13	78.13	78.13	75.00	71.88
R	83.33	91.67	86.11	94.44	94.44
Total	80.28	85.92	83.10	85.92	84.51
2015					
Unit	W_0	W_1	W_2	W_3	W_4
U	66.67	100.00	100.00	100.00	100.00

UR	75.00	81.25	84.38	81.25	78.13
R	86.11	88.89	91.67	91.67	86.11
Total	80.28	85.92	88.73	87.32	83.10
			2018		
Unit	W_0	W_1	W_2	W_3	W_4
U	66.67	100.00	100.00	100.00	100.00
UR	75.76	84.85	81.82	75.76	84.85
R	91.43	94.29	94.29	97.14	91.43
Total	83.10	90.14	88.73	87.32	88.73

Source: *Own elaboration.*

Assuming that assigning a given commune to a particular class depends not only on its level of development but also on the socio-economic situation of its neighbours, significantly increases the percentage of correctly classified objects. The highest increase in classification accuracy was observed in 2009 for matrices W_1, W_3 and W_4, and was equal almost to 10 percentage points. In the set of all voivodships, the highest increase in classification accuracy caused by taking into account the influence of the neighbourhood occurred for the Zachodniopomorskie voivodship in the variant of W_1 matrix in 2006 and amounted to 12.28 percentage points.

The classification results for Lubelskie and Wielkopolskie voivodships, which were characterized by the highest and the lowest consistency of administrative labels defining the type of communes with their level of development, are presented in Figures 4 and 5 respectively. The observed difference was equal to almost 15 percentage points and indicated a very clear divergence of the classification results of communes of Wielkopolskie voivodship with their level of socio-economic development. Such a situation can be a serious limitation to the proper allocation of resources under the implemented cohesion policy and regional development in this region.

Figure 4. *Percentage of correctly classified communes in Lubelskie voivodship*

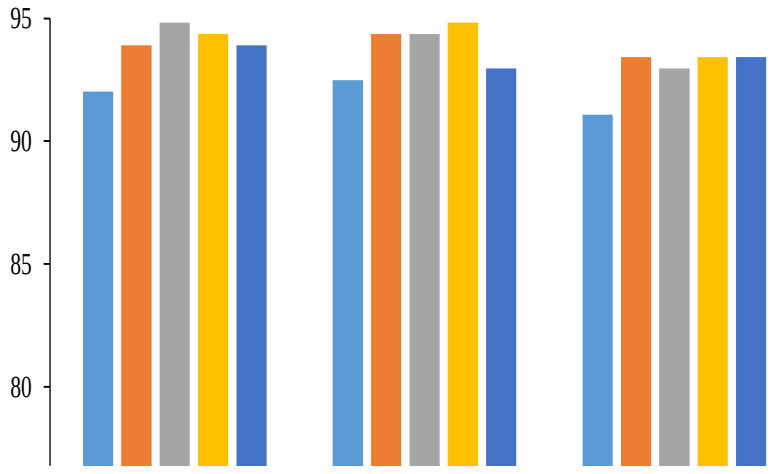
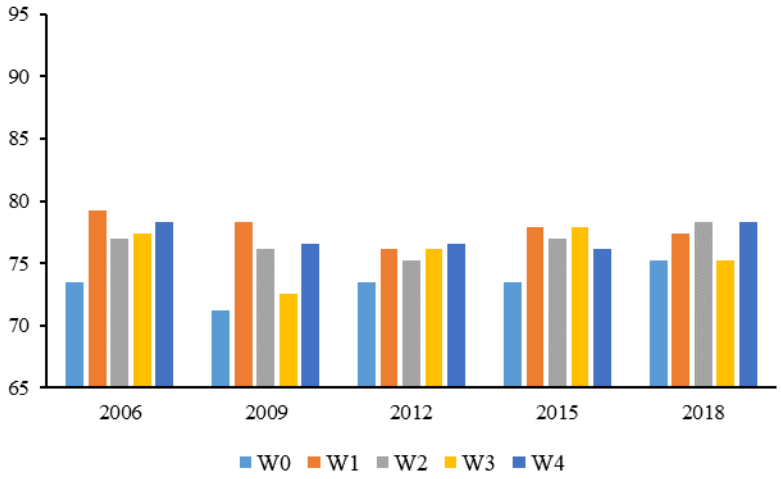


Figure 5. *Percentage of correctly classified communes in Wielkopolskie voivodship*



Source: *Own elaboration.*

Table 4 contains values of measure of classification quality – mean percentage of correctly classified objects – for all studied years and types of spatial relations for the whole country, while Table 5 presents the classification accuracy for all analysed regions and all variants of spatial weights matrices.

Table 4*. Mean percentage of correctly classified communes in analysed years according to spatial matrices*

	W_0	W_1	W_2	W_3	W_4
2006	84.30	86.88	86.84	86.81	86.44
2009	83.56	86.59	86.96	86.18	87.30
2012	82.46	85.13	85.19	85.67	85.22
2015	82.86	85.54	85.74	85.53	85.92
2018	81.87	85.29	85.10	84.16	85.01

Source: *Own elaboration.*

Table 5. *Mean percentage of correctly classified communes in analysed regions – all years and models*

Voivodship	W_0	W_1	W_2	W_3	W_4
Dolnośląskie	79.05	82.25	81.30	80.71	82.01
Kujawsko-pomorskie	83.19	85.14	85.56	85.69	85.56
Lubelskie	90.42	92.86	93.15	92.49	92.58
Lubuskie	82.68	85.12	87.80	87.07	88.05
Łódzkie	89.94	91.53	90.73	90.96	91.30
Małopolskie	84.84	85.27	86.48	84.84	85.38
Mazowieckie	83.44	85.54	86.69	84.84	86.24
Opolskie	81.41	88.45	87.89	87.89	87.32
Podkarpackie	82.00	85.38	83.75	84.75	83.13
Podlaskie	83.05	84.24	86.44	85.08	87.29
Pomorskie	88.62	92.52	92.36	92.85	92.68
Śląskie	79.64	82.75	83.11	82.99	84.19

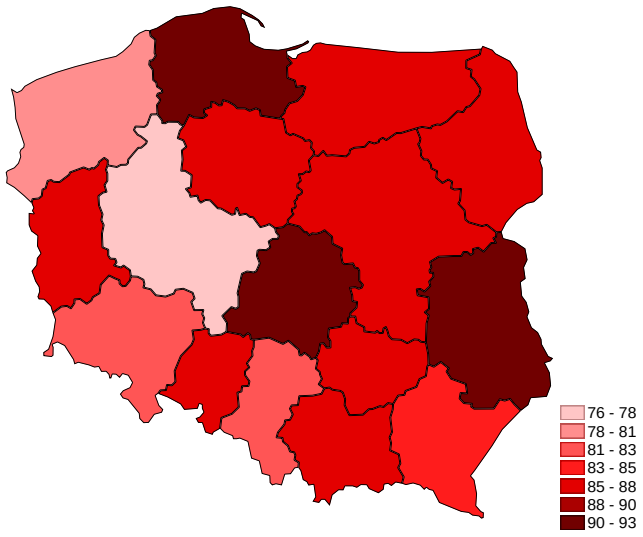
Świętokrzyskie	85.88	85.29	85.49	86.08	85.49
Warmińsko-mazurskie	83.79	87.24	86.21	86.90	86.21
Wielkopolskie	73.36	77.79	76.73	75.84	77.17
Zachodniopomorskie	76.84	82.81	81.75	81.75	81.05

Source: *Own elaboration.*

The analysis of the obtained results leads to several conclusions. The first is the steady decline in the quality of classification of communes over time, which is particularly evident in the classical variant of LDA. The second regularity is that classification quality increases, on average, by about 3 percentage points when spatial weights matrices are introduced into the discriminant functions, especially in variants W_1 and W_4. In the case of individual voivodships, we can also formulate some interesting observations There is one voivodship – Świętokrzyskie, for which classification results are completely insensitive to the use of spatial relationships.

The group of regions characterized by the high quality of classification, both in the classical variant and spatial variants of LDA, apart from the above-mentioned Lubelskie voivodship, also includes Łódzkie (89.94 %) and Pomorskie (88.62 %). On the other hand, in the group of regions with the lowest classification accuracy, apart from Wielkopolskie voivodship indicated earlier, there are also Zachodniopomorskie (76.84 %), Dolnośląskie (79.05 %) and Śląskie (79.64 %) (see Figure 6).

Figure 6. *Mean percentage of correctly classified communes in analysed voivodships – all years and all spatial matrices*



Source: *Own elaboration.*

The overall quality of the classification can be analysed in more detail. The direction and level of divergence between the administrative type of communes and their level

of development can also be considered. Table 6 presents shift codes for all years and communes with the following meanings:

- -2 – indicates an urban commune which was classified in the group of rural communes,
- -1 – indicates an urban commune which was classified in the group of urban-rural communes or an urban-rural commune which was classified in the group of rural communes,
- 0 – indicates properly classified commune,
- 1 – indicates an urban-rural commune which was classified in the group of urban communes or a rural commune which was classified in the group of urban-rural communes,
- 2 – indicates a rural commune that was classified in the group of urban communes.

Table 6. *Distribution of shift codes according to spatial matrices – all years and communes*

Shift code	W_0	W_1	W_2	W_3	W_4
-2	92	80	77	68	73
-1	1473	1200	1197	1234	1188
0	10299	10621	10629	10572	10628
1	508	477	476	502	487
2	18	12	11	14	14

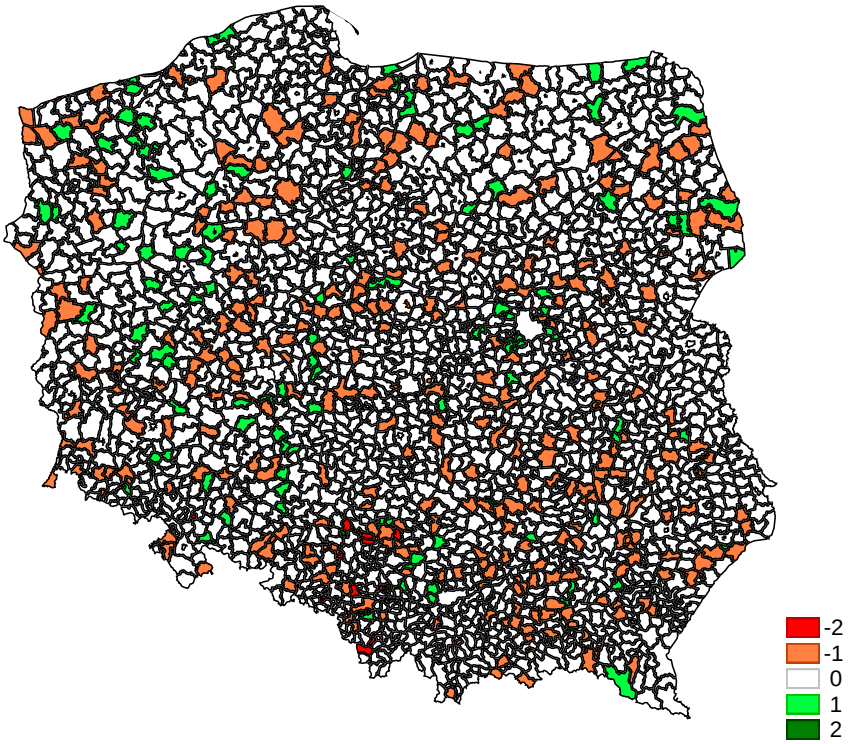
Source: *Own elaboration.*

If we assume that positive values of the shift codes presented in Table 6 indicate a higher level of development in comparison with the administrative label, and negative values indicate a lower one, we may observe a significant predominance of communes whose class, indicated based on the current level of development, concerning their administrative type, was lower. It is worth noting here that the largest fraction among the wrongly classified communes were urban-rural ones, whose level of socio-economic development was on the level of rural communes.

The results obtained, indicating the marginalisation of medium-sized cities, are consistent with the conclusions formulated by Śleszyński *et al*. (2019) who indicate that the main reasons for this phenomenon are monofunctionality, the decline in population (especially educated population of working age), the mismatch between labour demand and supply, declining incomes, and insufficient transport accessibility.

The geographical distribution of correctly and incorrectly classified communes in 2018 for the classical and spatial LDA for the W_4 matrix is shown in Figures 7 and 8 respectively.

Figure 7. *Shift codes for all communes in 2018 for classical LDA*



Source: *Own elaboration.*

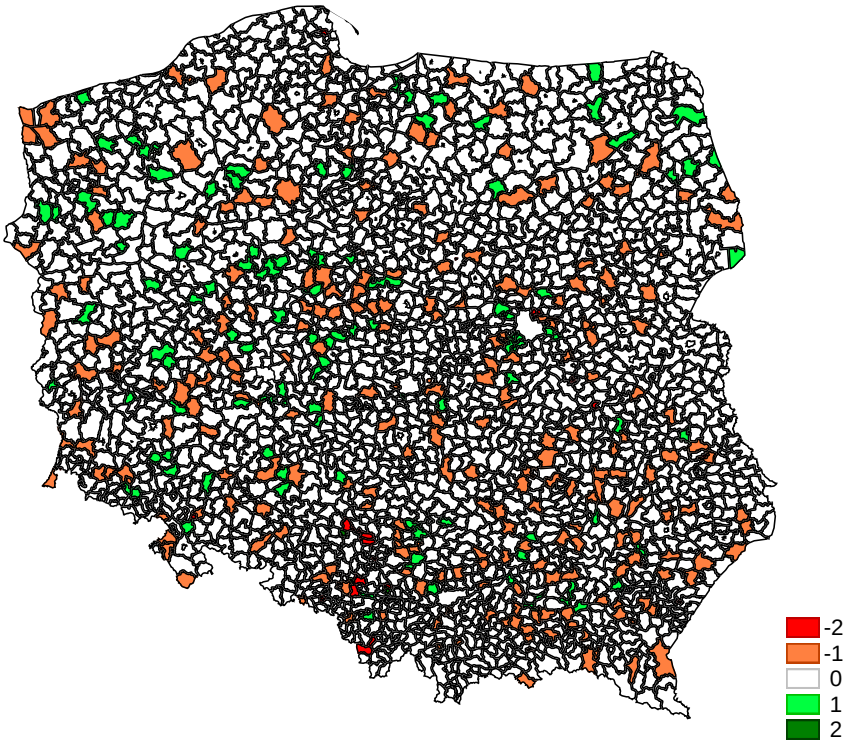
It could be observed that the greatest number of discrepancies of a negative direction, marked in red, occurs among communes in the Śląskie voivodship, located in the south of Poland.

Taking into account all estimated models, we can identify variables characterised by the highest discriminant power. Table 7 presents how many times the given variable was considered to have high discriminant power in every variant of the spatial matrix. In the classic LDA, the most influential variables include:

X_9 – population density,
X_7 – employed persons *per* 1000 population,
X_2 – national economy entities registered *per* 1000 population,
X_4 – useful floor area of dwellings completed *per* 1000 population.

The introduction of spatial weights matrix strengthens the importance of variables X_2 and X_1. This means that the determination of the class of a given commune is significantly influenced by the level of entrepreneurship and own revenue *per capita* of neighbouring communes.

***Figure 8**. Shift codes for all communes in 2018 for spatial LDA – matrix W_4*



***Source:** Own elaboration.*

***Table 7**. The most influential variables – all years, communes and models*

Variable	W_0	W_1	W_2	W_3	W_4
X_1	28	58	56	59	60
X_2	55	105	105	114	106
X_3	17	33	36	40	35
X_4	53	57	71	73	64
X_5	8	18	17	25	19
X_6	5	9	11	25	7
X_7	56	67	61	71	66
X_8	5	13	9	20	15
X_9	80	88	93	96	84

***Source:** Own elaboration.*

4. Conclusions

The level of regional development depends on many diverse determinants, including economic, demographic, environmental, and social factors, as well as trends observed in the labour market. Development processes may be supported or hindered by interventions undertaken within the framework of the regional policy

implemented and by decisions concerning the level of financial external support. In this context, it is important to direct properly resources aimed at supporting the development of less developed areas, and it is crucial to identify properly the types of areas to determine the set of potential beneficiaries. The development of smaller urban centres and rural areas with a lower development potential depends, inter alia, on the degree of functional integration with the main urban centres of the region, as well as the possibility of using unique internal resources. Policy objectives should be based on a region's distinctive structures and knowledge base, building capacities for smart specialisation strategies to increase their competitive advantage, to ensure the harmonious development of both urban and non-urban areas, as well strengthening urban-rural linkages. Such policy should contribute to reducing disparities between the levels of development of the various regions, especially addressing the problems of disadvantaged rural areas in improving the resilience of their communities, accessing basic services, enhancing attractiveness for investment and connectivity to large markets (Regulation …, 2021).

However, we should be aware that the evolution of urban and rural areas, does not depend only on its endogenous sources of development, but also on the potential and development of their neighbourhoods level, as well as the intensity of interactions with them. Therefore, in the process of identifying the type of territorial unit, which is compatible not only with its administrative character but above all with its level of socio-economic development, the characteristics of neighbouring units should be taken into account. One of the tools that allow the classification of objects is linear discriminant analysis. The results obtained indicate that if the spatial relationships between the surveyed regions are incorporated when LDA is applied, then we observe higher accuracy of classification because we reduce the value of classification errors caused not only by the incompatibility of the socio-economic development of communes with their administrative status but also by interactions with neighbouring regions.

It seems that proposed models, given their high quality, could have the practical application for determining the kind of local administrative unit and consequently become the crucial factor in both designing and evaluating regional policy. Especially in the case that someone needs to designate an administrative label for the new local unit or the modified one, as it was in the case of Zachodniopomorskie voivodship in 2019 when the commune Ostrowice became bankrupt.

The analyses carried out can form the basis for further research, in which models of discriminant analysis would be used in the classification of European regional and local units. It could be also interesting to determine the impact of outliers and territorial accessibility on the results of such classifications, as well as to provide robustness analysis of applying different sets of predictors and generating sub-samples. It would also be an interesting area of analysis to determine the impact of the degree of data aggregation, i.e. the size of the study areas. This factor, as indicated by LeSage (1999), appears to be important because the administrative

boundaries for collecting data do not accurately reflect the nature of the underlying process of generating the sample data. The removal of boundaries between regional units, whether within a country or internationally, can significantly change the results obtained and allow a better assessment of whether we observe a marginalisation of peripheral areas relative to the core of a region. Such a conclusion would confirm the hypotheses on the existence of regional divergence, the limited spatial impact of the regional core and the declining importance of location in the development process.

References:

Abreu, M., de Groot, H.L.F., Florax, R.J.G.M. 2004. Space and Growth: A Survey of Empirical Evidence and Methods. Tinbergen Institute Discussion Paper, TI 2004-129/3, Amsterdam.

Adamowicz, M., Pyra, M, 2019. Links Between the Level of Local and Regional Development – Problems of Measuring. In: Proceedings of the 2019 International Conference "Economic Science for Rural Development", 51, Jelgava, LLU ESAF, 9-10, 14-22. doi:10.22616/ESRD.2019.0522019.

Anselin, L. 1988. Spatial Econometrics: Methods and Models. Dordrecht, Kluwer Academic.

Applying the Degree of Urbanisation. A Methodological Manual to Define Cities, Towns and Rural Areas for International Comparisons. 2021. European Union/FAO/UN-Habitat/OECD/The World Bank.

Batóg, B., Batóg, J. 2020. Classification of Local Administrative Units in Poland: Spatial Approach. In: Advances in Intelligent Systems and Computing, 1173, W. Zamojski, J. Mazurkiewicz, J. Sugier, T. Walkowiak, J. Kacprzyk (Eds). Theory and Applications of Dependable Computer Systems, Proceedings of the Fifteenth International Conference on Dependability of Computer Systems DepCoS-RELCOMEX, June 29-July 3, Brunów, Poland, 31-40. Springer, Cham, Switzerland. doi:10.1007/978-3-030-48256-5_4.

Batóg, J., Batóg, B. 2019. The Application of Discriminant Analysis to the Identification of Key Factors of the Development of Polish Cities. Folia Oeconomica. Acta Universitatis Lodziensis, 4(343), 181-194. doi:10.18778/0208-6018.343.11.

Bajracharya, B., Hastings, P. 2018. A Regional, Strategic Growth-Management Approach to Urban and Peri-Urban Development in South East Queensland, Australia. Journal of Regional and City Planning, 29(3), 210-233. doi:10.5614/jrcp.2018.29.3.3.

Bosworth, G., Venhorst, V. 2018. Economic linkages between urban and rural regions – what's in it for the rural? Regional Studies, 52(8), 1075-1085. doi:10.1080/00343404.2017.1339868.

Copus, A. 2015. The new rural economy and macro-scale patterns, In Copus, A., P. de Lima, Ed., Territorial cohesion in rural Europe: The relational turn in rural development, 11-34. New York, Routledge.

Cörvers, F., Hensen, M., Bongaerts, D. 2009. Delimitation and Coherence of Functional and Administrative Regions. Regional Studies, 43(1), 19-31.

Cutillo, L., Amato, U. 2008. Localized Empirical Discriminant Analysis. Computational Statistics and Data Analysis, 52(11), 4966-4978.

Dyjach, K. 2013. Teorie rozwoju regionalnego wobec zróżnicowań międzyregionalnych, Theories of the Regional Development in View of Interregional Disparities. Annales Universitatis Mariae Curie-Skłodowska. Sectio H Oeconomica, 47(1), 49-59.

El-Hanjouri, M.M.R., Hamad, S.B. 2015. Using Cluster Analysis and Discriminant Analysis Methods in Classification with Application on Standard of Living Family in Palestinian Areas. International Journal of Statistics and Applications, 5(5), 213-222. doi:10.5923/j.statistics.20150505.05.

Ezcurra, R., Rios, V. 2019. Quality of government and regional resilience in the European Union. Evidence from the Great Recession. Papers in Regional Science, 98(3), 1267-1290. doi:10.1111/pirs.12417.

Fiaschetti, M., Graziano, M., Heumann, W.B. 2021. A data-based approach to identifying regional typologies and exemplars across the urban–rural gradient in Europe using affinity propagation. Regional Studies, Forthcoming special issue: Regional Approaches to data, environment, and society, 1-16. doi:10.1080/00343404.2021.1871598.

Filipiak, B.Z., Tarczyńska-Łuniewska, M. 2020, Socio-Economic Potential of Regions – Theory and Practice. Folia Oeconomica Stetinensia, 20(1), 95-116. doi:10.2478/foli-2020-0006.

Fischer, M.M., Wang, J. 2011, Spatial Data Analysis. Models, Methods and Techniques. Heidelberg, Dordrecht, London, New York, Springer.

Giannakis, E., Bruggeman, A. 2020. Regional disparities in economic resilience in the European Union across the urban–rural divide. Regional Studies, 54(9), 1200-1213. doi:10.1080/00343404.2019.1698720.

Gray, D. 2014. Economic approaches to the rural. In: G. Bosworth, P. Somerville, Eds., Interpreting rurality: multidisciplinary approaches, 32-53. Abingdon, Routledge.

Greco, S., Ishizaka, A., Matarazzo, B., Torrisi, G. 2018. Stochastic multi-attribute acceptability analysis. SMAA: an application to the ranking of Italian regions. Regional Studies, 52(4), 585-600. doi:10.1080/00343404.2017.1347612.

Hedlund, M. 2016. Mapping the socioeconomic landscape of rural Sweden: Towards a typology of rural areas. Regional Studies, 50(3), 460-474. doi:10.1080/00343404.2014.924618.

Holl, A. 2017. Local employment growth patterns and the Great Recession: The case of Spain. Journal of Regional Science, 58(4), 837-863. doi:10.1111/jors.12392.

Jaba, E., Jemna, V.D. Viorica, D., Lacatusu, T. 2006. Discriminant Analysis in the Study of Romanian Regional Economic Development in View of European Integration. Accessed from: SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=931613.

Kopczewska, K. 2008. Renta geograficzna a rozwój społeczno-gospodarczy. Geographical pension and socio-economic development. Warszawa, Poland: CeDeWu.pl Wydawnictwa Fachowe.

Kopczewska, K. 2011. Rola sektora publicznego w przestrzennym rozwoju państwa. The role of the public sector in the spatial development of the country. Warszawa, Poland, CeDeWu.pl Wydawnictwa Fachowe.

Kopczewska, K. 2021. Applied Spatial Statistics and Econometrics Data Analysis in R. London and New York, Routledge.

Krajowa Strategia Rozwoju Regionalnego 2030. Rozwój społecznie wrażliwy i terytorialnie zrównoważony. 2019. Warszawa, Ministerstwo Inwestycji i Rozwoju.

Latuconsina, Z., Rustiadi, E., Sahara, S. 2018. An Analysis of the Level of Development in Malang Regency Based on a Typology of Development Regions. Journal of Regional and City Planning, 29(1), 1-17. doi:10.5614/jrcp.2018.29.1.1.

Laurin, F., Pronovost, S., Carrier, M. 2020. The end of the urban-rural dichotomy? Towards a new regional typology for SME performance. Journal of Rural Studies, 80, 53-75.

LeSage, J.P. 1999. Spatial Econometrics. Toledo, University of Toledo.

Malina, A. 2004. Wielowymiarowa analiza przestrzennego zróżnicowania struktury gospodarki Polski według województw. A Multidimensional Analysis of the Spatial Differentiation of Poland's Economic Structure by Voivodship. Kraków, Poland, Wydawnictwo Akademii Ekonomicznej w Krakowie.

McLachlan, G.J. 2004. Discriminant Analysis and Statistical Pattern Recognition. Hoboken: John Wiley & Sons.

Modai-Snir, T., van Ham, M. 2018. Neighbourhood change and spatial polarization: The roles of increasing inequality and divergent urban development. Cities, 82, 108-118. doi:10.1016/j.cities.2018.05.009.

Nigohosyan, D., Vutsova, A. 2018. The 2014-2020 European Regional Development Fund Indicators: The Incomplete Evolution. Social Indicator Research, 137(2), 559-577. doi:10.1007/s11205-017-1610-8.

Partridge, M., Bollman, R., Olfert, R.M., Alasia, A. 2007. Riding the wave of urban growth in the countryside: Spread, backwash, or stagnation? Land Economics, 83(2), 128-152. doi:10.3368/le.83.2.128.

Pontarollo, N., Serpieri, C. 2020. A composite policy tool to measure territorial resilience capacity. Socio-Economic Planning Sciences, 70. doi:10.1016/j.seps.2018.11.006.

Regulation, EU, 2021/1058 of the European Parliament and of the Council of 24 June 2021 on the European Regional Development Fund and on the Cohesion Fund. 2021. Official Journal of the European Union 231/60.

Report from the Commission to the European Parliament and the Council. Seventh progress report on economic, social and territorial cohesion. 2011. European Commission, SEC(2011, 1372 final, Brussels.

Silva, R., Ferreira-Lopes, A. 2014. A Regional Development Index for Portugal. Social Indicator Research, 118(3), 1055-1085. doi:10.1007/s11205-013-0455-z.

Śleszyński, P., Bański, J., Degórski, M., Komornicki, T. 2019. Aktualizacja delimitacji obszarów problemowych na 2018 rok, 2018 Update of the Delimitation of Problem Areas. Cooperation: M. Mazur, M. Stępniak, Polska Akademia Nauk Instytut Geografii i Przestrzennego Zagospodarowania, Warszawa.

Steele, B.M., Redmond, L.R. 2001. A Method of Exploiting Spatial Information for Improving Classification Rules: Application to the Construction of Polygon-based Land Cover Maps. International Journal of Remote Sensing, 22(16), 3143-3166.

Tacq, J. 2007. Multivariate Analysis in Social Science Research. London, Sage Publications.

Todtling, F., Trippl, M. 2005. One Size Fits All? Towards a Differentiated Regional Innovation Policy Approach. Research Policy, 34(8), 1203-1219.

Uchida, H., Nelson, A. 2009. Agglomeration index: towards a new measure of urban concentration. Background paper for the World Bank's World Development Report 2009. https://openknowledge.worldbank.org/handle/10986/9039.

Van Eupen, M., Metzger, J.M., Pérez-Soba, M., Verburg, N.P., Van Doorn, A., Bunce, R.G.H. 2012. A rural typology for strategic European policies. Land Use Policy, 29(3), 473-482. doi:10.1016/j.landusepol.2011.07.007.

Wandl, A., Nadin, V., Zonneveld, W., Rooij, R. 2014. Beyond urban–rural classifications: Characterising and mapping territories-in-between across Europe. Landscape and Urban Planning, 130, 50-63. doi:10.1016/j.landurbplan.2014.06.010.

Wójcik, P., Herbst, M. 2011. Obszary polaryzacji i dyfuzji rozwoju społeczno-gospodarczego w Polsce. Próba delimitacji, Areas of polarisation and diffusion of socio-economic development in Poland. Delimitation attempt. Report for MRR, unpublished, University of Warsaw, Warszawa, Poland.