Building W Matrices Using Selected Geostatistics Tools: Empirical Examination and Application

Elżbieta Antczak,
Department of Spatial Econometrics,
University of Lodz, Rewolucji 1905 r. Street nr 37, 90-214 Łódź, Poland,
mail: wiszniewska@uni.lodz.pl

Abstract

The construction of a spatial-weight matrix (**W**) is an important problem of spatial econometrics. This matrix considers and expresses the potential for interactions between pairs of observations in various locations. The matrix could be set a priori (exogenously) by the researcher. On the other hand, **W** could be estimated from data (endogenously with ‘dedicated’ weights) and this paper investigates how to determine the values of weights in **W** according to that criteria. To achieve the goal, geostatistical tools (standard deviation ellipsis, semivariograms and surface trend models) were used. Then, in the econometric part of the analysis the effect of applying different variants of matrices was examined. The study was carried out on a sample of about 300 Polish towns and in the time span 2005-2015, as well as for averaged data for the whole period. Variables were related to the quantity of produced waste and economic development. Both ESDA and estimations of spatial panel and SUR models were performed by including particular **W** matrices in the study (exogenous matrix, distance and directional matrices constructed based on data). Received results indicated that 1) geostatistics tools can be effectively used to build **W**, 2) outcomes of applying different matrices did not exclude but rather supplemented one another, 3) the most precise picture of spatial dependences was received by including distance and directional matrices 4) the values of the assessed parameter at the regressors did not significantly change, there was, however, a change in the strength of spatial dependency.

Key words: **W** matrixes, endogenity of **W**, geostatistics, directional matrix, (semi)variograms, trend surface analysis;

1. Introduction

The construction of a spatial-weight matrix (**W**) is an important problem of spatial econometrics (Kooijman, 1976). This matrix considers and expresses the potential for interactions between pairs of observations in various locations (Anselin, 1988). The **W** matrix could be set a priori (**W** specified exogenously) by the researcher, which is not always satisfactory (Angulo et al. 2017). On the other hand, many scientists claim that the **W** matrix could be estimated from data (Harris, 2011).

Kooijman (1976) was one of the first to explicitly tackle the question of estimating the **W** matrix. He suggested that weights be built by maximising the value of Moran's *I*. His procedure aroused many doubts but provoked looking for innovative ways of solving the problem. For example, Lee (1982) showed that the **W** *k*-nearest neighbour problem and other seemingly unrelated problems can be solved efficiently with the Voronoi diagram. In addition, Griffith (1996) proposed finding **W** absorbing spatial effects from data. Fernández et al. (2009) suggested a specification of **W** based on the measure of entropy. Later, Mur and Paelinck (2010) focused on the maximisation of the complete correlation coefficient. Additionally, Getis and Aldstadt (2004) used the local statistical model and the amoeba algorithm. While Stewart and Zhukov (2010) indicted neighbours from the visualisation of spatial effects, Hondroyiannis et al. (2012) and Keleijan and Piras (2014) assumed that elements of **W** are an unknown function of two sets of exogenous variables. In addition, Benjanuvatra and Burridge (2015) as well as Qu and Lee (2015) presented the QML (Quasi-Maximum Likelihood estimator) to estimate weights in **W** directly from data.

Through the empirical application of geostatistical tools (from the standard deviation ellipsis (SDE) and variograms to surface trend models) that were used to construct a range of spatial-weight matrices, an attempt was made to answer the following research questions:

- How do neighbours influence each other: cumulatively, equally, or proportionally to their proximity or via some other measure of decay?
- Should the spatial-weight matrix contain information about the anisotropy of the phenomenon (identical weights without considering the directional character – dispersion or diffusion – of the phenomena in different directions in geographical space and with different intensity)?
- How is the distance of spatial correlations and the degree of the mutual influence of units in space determined?

- Does spatial autocorrelation change solely depending on the distance, or does it also depend on the direction of the courses of the phenomena?
- How are values of weights determined in **W** matrices?
- Do varied values of spatial weights lead to significant differences in the results of the analyses?
- Should weight matrices be different for different years?
- What results of analyses will we receive if we introduce a weight matrix built without considering the nature of the phenomena?

The study was carried out on a sample of about 300 Polish towns for selected years from 2005 through 2015 as well as for the averaged data for the whole period. Variables were related to the quantity of collected municipal mixed waste[1] and economic development of cities.[2] Both ESDA and estimations of spatial panel as well as SUR models were performed by including particular spatial-weight matrices in the study (exogenous matrix, distance, and directional matrices constructed based on data). The research was conducted in geospatial processing programs: ArcMap (by Esri ArcGis), R, SAGA (System for Automated Geoscientific Analyses), and GeoDa (An Introduction to Spatial Data Analysis). The results indicated that selected geostatistics tools can be effectively used to build **W** matrices.

2. Geostatistical Tools in Weight Matrix Construction

Recalling Tobler's first law of geography, the distance and neighbouring relations between different areas can indicate to what degree spatial dependence exists and 'how close places need to be' to be related or spatially autocorrelated. This law makes it clear that spatial relations are not static but evolve over distance.

In the structure of the spatial-weight matrix based on the geographical distance, it can be difficult to determine the maximum distance to which units are interrelated (show similarity in terms of a studied feature resulting from mutual spatial relations). One of the main assumptions of the article is to describe and apply spatial statistics and geostatistics methods (spatial measures of central tendency and semivariograms) based on which the

---

[1] Collected mixed municipal waste (*Waste*) covering waste from households, including bulky waste, similar waste from commerce and trade, office buildings, institutions, and small businesses, yard and garden, street sweepings, contents of litter containers, and market cleansing. Waste from municipal sewage networks and treatment as well as municipal construction and demolition is excluded; http://siteresources.worldbank.org/INTURBANDEVELOPMENT/Resources/336387-334852610766/Chap2.pdf, Accessed: 27.05.2017.

[2] Here, the value of revenue of the city budget in PLN per capita (*R*).

spatial continuity (variation and degree of spatial correlation) of specified phenomena can be effectively characterised depending on the distance.

The variogram ($2\gamma(s_i - s_j)$) is defined as the variance (*Var*) of a random variable whose values are differences in the realisation of the analysed phenomenon in various locations of space *D*:

$$Var[X(s_i) - X(s_j)] = E\{[X(s_i) - X(s_j)]^2\} = 2\gamma(s_i - s_j), \quad (1)$$

where $X(s_i)$ is the realisation of a random variable in location *i* and *j*, for *i* = 1, ..., *N* and *j* = 1, ..., *N*.

In the variogram, each pair of spatial locations is considered twice. For this reason, to describe the differentiation of the variable depending on the distance of measuring points, the semivariogram (semivariance) is analysed. Geostatistics determines it as half of the variogram:

$$\frac{1}{2}Var[X(s_i) - X(s_j)] = \frac{1}{2}E\{[X(s_i) - X(s_j)]^2\} = \gamma(s_i - s_j). \quad (2)$$

One may notice that the values of the semivariogram are only distance functions and are not functions of specific locations. Therefore, the commonly used notation of the semivariogram is one according to the classical formula proposed by Matheron (1965), as follows:

$$\frac{1}{2}Var[X(s_i) - X(s_j + h)] = \frac{1}{2}E\{[X(s_i) - X(s_j + h)]^2\} = \gamma(s_i, s_j + h). \quad (3)$$

Using *h* to mark the vector whose length may depend on both the distance separating two locations (and on the direction of measurement), which is an independent variable of the function of the variogram and semivariogram, one may use the following formula:[3]

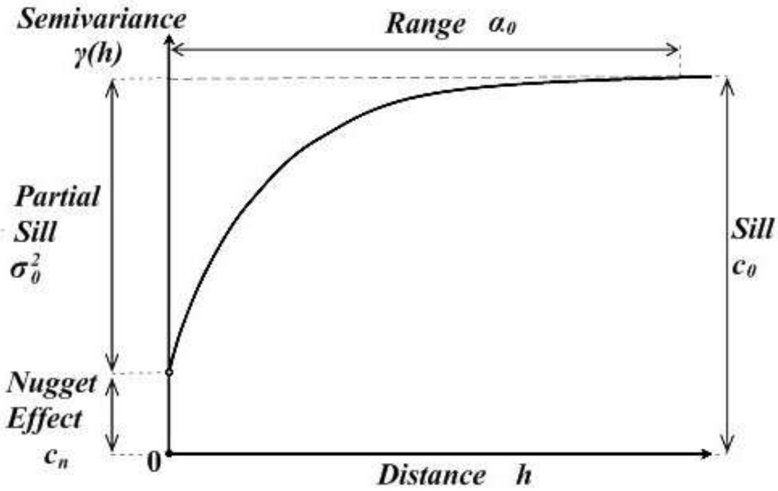
$$\gamma(h) = \frac{1}{2}E\{[X(s+h) - X(s)]^2\}. \quad (4)$$

In the case of an isotropic process of the random field, arguments of the semivariogram function do not depend on direction: $\gamma(h) \equiv \gamma(h)$, where $h = \|h\|$.

According to the graphical representation, when analysing changes in values of the semivariogram with changing *distance* $h = \|h\|$, to describe continuous spatial variability of the process, one may distinguish three specific parameters: *nugget*, *sill*, and *range* (Figure 1).

[3] According to Waller and Gotway (2004).

Figure 1. Theoretical semivariogram of spherical type and its characteristics

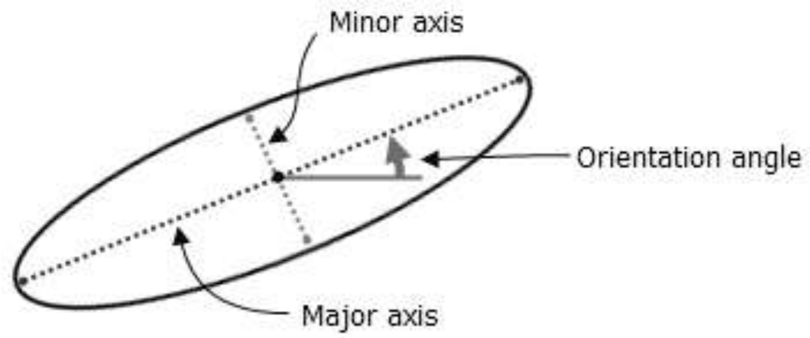


Source: own elaboration based on https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_variogram_a0000000386.htm, accessed on: 24.05.2017.

The *nugget effect* (c_n) is an initial quantity (an absolute term for $h = 0$) of the *semivariogram* function $\gamma(h)$, $\lim_{h \to 0} \gamma(h) = c_0 > 0$. The constant value attained by the semivariogram function, at which further increase of the function is not observed, is called *sill* (c_0), Figure 1. On the other hand, the range (α_0) of distance h from zero to the point where the semivariogram attains approximately 95% of the constant value is called *range*. This expresses the longest distance at which the values of the semivariogram are still correlated. With a further increase in the distance, one can no longer observe autocorrelation, and the semivariogram attains values similar to the constant threshold value and close to the total variance.

With a lack of spatial autocorrelation of the analysed phenomenon, the semivariogram takes the form of a line parallel to the horizontal axis. Step changes show that the variable is not continuous and has highly irregular spatial variability. The high value of the nugget effect especially shows that two observations from very close locations may have significantly different values. If analysing the phenomena is characterised by a specific pattern of spatial changes, then the semivariogram may be a function of variable values (increasing and decreasing; Bao 2000). Directionality shown by the semivariogram may be confirmed by the standard deviational ellipse (SDE; Figure 2).

Figure 2. Components of an ellipse



Source: http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/h-how-zonal-geometry-works.htm, accessed: 13.05.2017.

Standard deviation arises as one of the classical statistical measures for depicting the dispersion of univariate features around the centre. Its evolution in two-dimensional space arrives at the SDE, which was first proposed by Lefever (1926) in 1926. Ever since then, SDE has long served as a versatile geographical information system (GIS) tool for delineating the bivariate distributed features. It is typically employed for sketching the geographical distribution trend of the features concerned by summarising both of their dispersion and orientation. Therefore, it has also been adopted to quantitatively analyse the orientation anisotropy (Wang et al. 2008). The SDE is mainly determined by three measures: average location, dispersion (or concentration), and orientation (Wang 2015).

This paper investigates how to determine the values of weights in **W**. To achieve that, the trend surface analysis (TSA) was used. It describes trends of changes of a phenomenon in a geographical space and, consequently, enables one to come to conclusions concerning agglomeration, dispersion, or spatial global trends and local fluctuations (Chojnicki and Czyż 1975). The TSA is one of the global surface-fitting procedures (i.e., spatial trend identification supporting inference about the nature of the spatial trend estimation of a phenomenon) and one of the oldest mathematical analytical methods used for spatial non-stationarity analysis. The TSA parameters reflect the strength and direction of global and local trends (i.e., systematic heterogeneity of a phenomenon expressing itself with the mean of a spatial process). A random component enables one to grasp the variability of a phenomenon of the mean constant in space, which makes identification of the potential spatial relations possible. According to Chojnicki and Czyż (1975), the TSA method enables one to provide a simplified description of a spatial system of high complexity by means of separating large-scale systematic spatial changes from small-scale local fluctuations. Depending on the spatial structure, variability, and nature of a phenomenon, spatial (surface) trend models take different forms of a function (from linear to a polynomial of any degree).

However, one should note that these models do not provide information on the reasons for formation of a phenomenon in space. Regarding the estimation of regression of a specific dependent variable (z_{ij}), where independent variables in the basic version of the model are orthogonal geographic coordinates ($Xkoor_i$, $Ykoor_j$), the estimation of the model gives an answer to the question regarding whether there are any principles governing the spatial distribution of the analysed phenomenon in the defined region (2), (Chojnicki and Czyż 1975). Depending on the nature of changes of the phenomenon in space, one may discuss a linear spatial trend (5), non-linear spatial trend estimation, for example, expressed by means of a second-degree polynomial, higher *n* degrees, and analogous versions of models with an added set of independent variables. The general form of TSA is as follows:

$$z_{ij} = \alpha_0 + \alpha_1 \cdot Xkoor_i + \alpha_2 \cdot Ykoor_j + \varepsilon_{ij}, \quad (5)$$

where z_{ij} is a dependent variable in a geographical space, $Xkoor_i$, $Ykoor_j$ are independent variables (flat coordinates of geographic location), α_0 denotes the constant, absolute-term, global trend, α_1, ..., α_i are the structural parameters standing near flat coordinates, and ε_{ij} is a random component.

In the surface trend model, the signs of estimates of parameters are subject to estimation. 'Interpretation' of the value makes sense when comparisons are made over time. The sign of the absolute term, the constant α_0, stands for the global trend estimation of the level of the variable (growing or shrinking). The sign of the estimate of the parameter next to variable X_{koori} describes the global trend of the phenomenon from the west to the east of the area, while the sign of the estimate of the parameter next to the variable Y_{koorj} provides information on the global trend from the south to the north. The remainders obtained from the surface trend model give a picture of the local fluctuations and deviations – the so-called spatial local trends – that are specific to a certain unit. The weights in the matrix were given in such a way that the units of a clearly stronger directional spatial trend were attributed with higher values. On the other hand, the outliers, which clearly stand out regarding the level of the analysed phenomenon, were attributed the highest weights in relation to their values.

The following methods were used to set the **W** *a priori* (exogenously) by description:

- the variation of a phenomenon depending on the distance and direction,
- the SDE (allowing one to see if the distribution of features is elongated and hence has a specific orientation),

- the directional variogram (defined above), and
- the surface trend models (a mathematical function or polynomial that describes the variation in data).

3. Results

3.1 Spatial Weights Matrices

3.1.1 Geographical Distance Matrices

Geographic distances from the geographic centres of Polish cities, which served to build three (12 total for each year) spatial-weight matrices, were selected based on the occurring statistically significant spatial relationships characterising the discussed phenomenon. To evaluate the scope of the spatial autocorrelation (similarity), a semivariogram was calculated and drawn showing values of semivariance for specific ranges of distances among the compared locations (Table 1 and Figure 3).

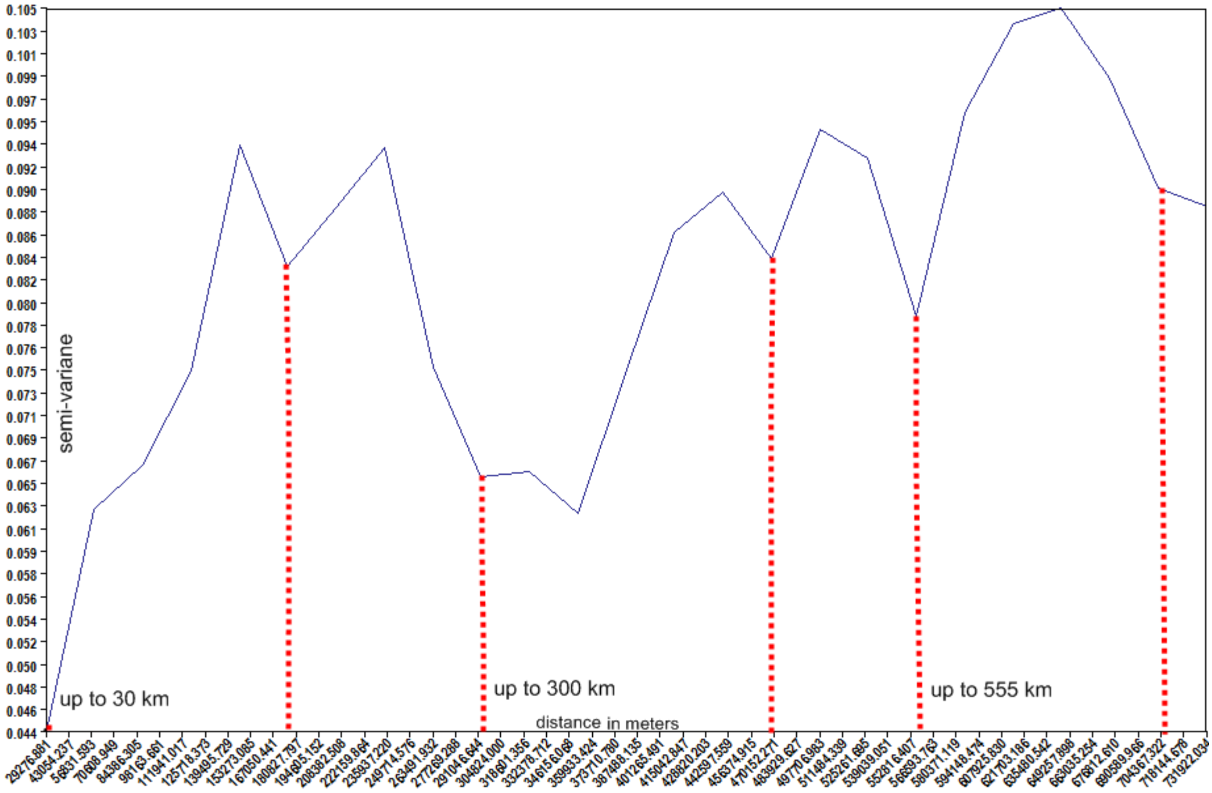
Table 1. Semivariance values of $Waste_{av}$ (average for 2005–2015)[4] and distance range (in metres)

Class	Distance	Semivariance
	$Waste_{av}$	
1	29277	0.044
2	58554	0.063
3	87830	0.067
4	117107	0.075
5	146384	0.093
6	175661	0.083
7	204938	0.088
8	234215	0.093
9	263492	0.075
10	292769	0.066
11	322046	0.066
12	351323	0.063
13	380599	0.075
14	409876	0.086
15	439153	0.089
16	468430	0.084
17	497707	0.095
18	526984	0.092
19	555261	0.079
20	585538	0.096

Note: grey – the most statistically significant values of the spatial autocorrelation
Source: own elaboration in SAGA.

[4] I also built spatial-weight matrices (distance, directional, and random) for the selected years of the time span: 2005, 2010, and 2015 to examine how results differ considering different spatial matrices (available by e-mail: wiszniewska@uni.lodz.pl).

Figure 3. Semi-variogram values of *Waste*$_{av}$ (average for 2005–2015) and distance range (in metres)



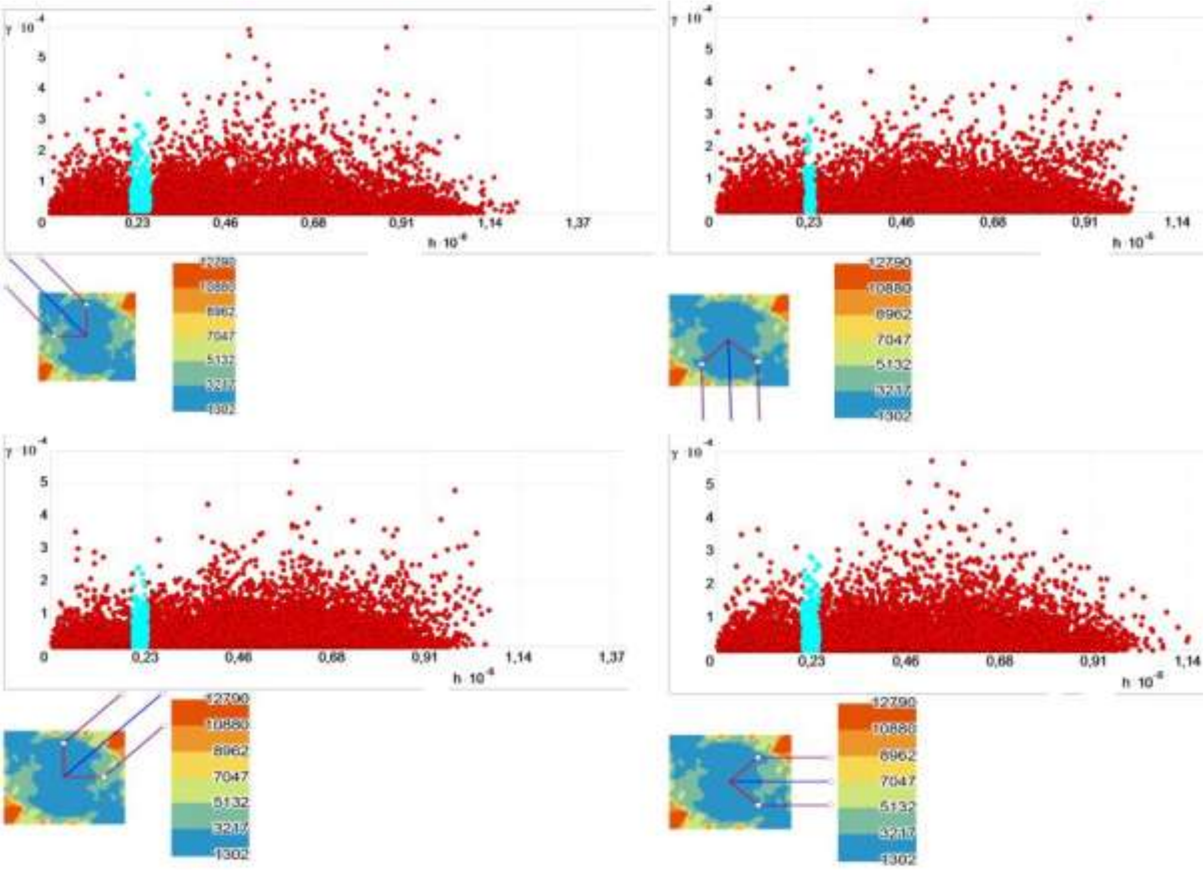
Note: Red lines – the noticeable decrease of semivariance values (range of statistically spatial autocorrelation).
Source: own elaboration in SAGA based on Table 1.

The data shown in Table 1 and Figure 3 indicate that, for the individual distance ranges, the spatial autocorrelation characterising the municipal waste collection in Polish cities on average reached: 29 km, 180 km, 300 km, 470 km, 555 km, and 705 km (values of the semivariance decreased with distance periodically). This may have been caused by several factors that determine the appeal of a given region and thus the created concept of the transport of waste. To summarise, the results of the analysis indicated that waste collection (management) might be local (regional, social, and urban economic development and waste policy determine the volume of waste streams) and global in nature. Spatial autocorrelation reached up to more than 700 km due to the transboundary shipment of waste. Hence, three matrices were built for selected geographic distances: $\mathbf{W_1}$ – a weight matrix built based on the close distance from the determined geographic centres of specific Polish cities with a circle radius of up to 29 km, $\mathbf{W_2}$ – where the weights were determined depending on the ‘medium’ geographic distance from the geographic centre of a specific city and units contained in a circle with a radius of up to 300 km, and $\mathbf{W_3}$ – based on the circle with a radius of up to 555 km (far distance weight matrix).

3.1.2 Directional Matrix

Examining the semivariogram surfaces (Figure 4), it appears that, on average, there might be directional differences in the semivariogram values of the amount of collected waste in Polish cities from 2005 through 2015. Changing the direction of the links, as shown in Figure 4, some linked locations have values that are quite different, which result in higher semivariogram values. This indicates that cities separated by a selected distance of about 230,000 m in the northwest direction are, on average, more different and higher in terms of waste streams than locations in the south and east. When variation changes more rapidly in one direction than another, it is termed anisotropy.

Figure 4. Semivariogram cloud for *Waste*$_{av}$ with search direction



Source: own elaboration in ArcGIS 9.3.

Considering the assumption of the significant anisotropy in spatial data autocorrelation, the asymmetrical directional matrix was built ($\mathbf{W_4}$). The weight values were initially determined based on the slope (orientation) of two SDEs (Figure 5), and then were made more precise and were confirmed by assessments of the parameters of the spatial trend model (TSA; Formula 7).

The dependent variable in the estimated trend model was waste quantity in Polish cities averaged by years. The form of the model is given by the following formula:

$$Waste_{av.} = \beta_0 + \beta_1 X_{coor} + \beta_2 Y_{coor} + \varepsilon_{it}, \tag{6}$$

where $Waste_{av}$ is the averaged quantities of mixed waste collection in kilograms per capita; X_{coor}, Y_{coor} are the standardised geographic coordinates of the centres of the analysed European cities; β_0, β_1, and β_2 are the structural parameters of the model; and ε_{it} is a random component. Upon the assessment of the parameters, the model takes the following form:

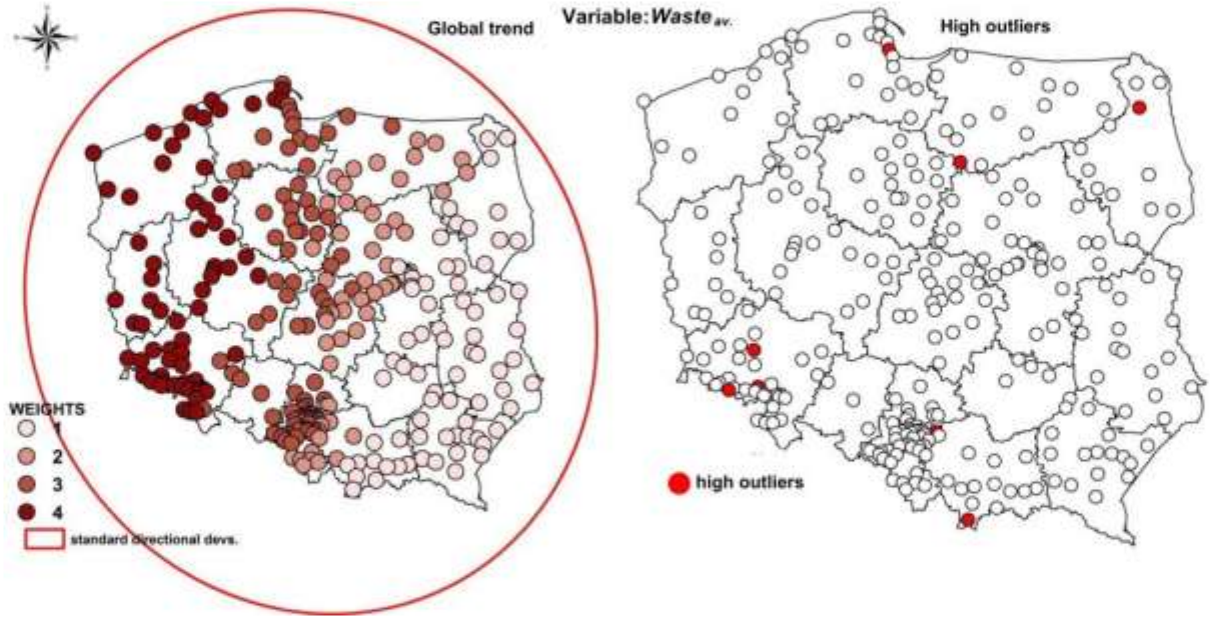
$$\begin{array}{lccc} \hat{Waste}_{av.} = & 273 & -189X_{coor} & -36Y_{coor} \\ t & (3.4) & (-5.5) & (-0.5) \\ \mathrm{S}(b_j) & (79.5) & (34.1) & (81.2) \end{array} \tag{7}$$

Not all estimated parameters of Model 3 were statistically significant at the assumed significance level of $\alpha = 0.05$ (for critical value $t^* = 1.65$). They confirmed the presence of a surface trend in waste collection (global and west-east). Furthermore, the global spatial average trend in the volume of the phenomenon was upward from 2005 to 2015 ($\beta_0 = 273$). Signs of assessed parameters at the *X* coordinate were negative (-189). This indicated a downward spatial trend in the west-east directions (confirmation of the initial assumptions read off the SDE, Figure 5). Therefore, cities in western Poland were characterised by a higher level of the analysed variable, and there was a downward trend among cities in the eastern parts of the country.

Based on the information about spatial trends in the average amount of collected municipal waste in Polish cities from 2005 to 2015, a spatial-weight matrix was built. The matrix considered the occurrence of the spatial trend in such a way that cities in north-western Poland were assigned higher weights, from 3 to 4 reflecting an upward spatial trend (compare Figure 5), whereas central and eastern cities were assigned lower weights, from 1 to 2. Moreover, while analysing the local spatial trends (Formula 7), high outliers were observed where volumes of waste exceeded the average for the studied Polish cities. In matrix $\mathbf{W_4}$, a weight of 5 was assigned to those cities.[5] The cities were Sopot, Karpacz, Augustów, Lidzbark, Legnica, Szczawnica, Karpacz, Sławków, and Zakopane (Figure 5).

[5]The analysis indicated the TSA model defined by Formula 7 as the best under the formal regimes: Jarque-Bera = 5.99 with *p*-value = 0.37, Breusch-Pagan = 3.66 with *p*-value = 0.18 and lower values of Akaike and Schwarz rather than models with high levels of function values.

Figure 5. Spatial trend approximation with the ellipse of two standard deviational distributions and high spatial outliers



Source: own elaboration in ArcGIS 9.3.

It can be inferred from Figure 5 that the directions of the intensity of migration processes in Europe are south-west and north-east. The flattening of the ellipse of two standard deviations (which contains ~95% of observations)[6] indicates a certain spatial trend in waste streams in Polish cities, which is made more precise and confirmed by statistically significant assessments of the surface trend model parameters.

3.1.3 Random Weight Matrix

This matrix (**W$_5$**) was set a priori (specified exogenously). In this part, the nearest neighbours matrix was built, based on the adjacency of the eight nearest cities (the only explanation was that the average number of city neighbours was eight).

3.2 ESDA

This stage of the study seeks an answer to the research question of whether the volume of municipal waste in Polish cities shows statistically significant spatial relationships. Does the strength of these interactions change with distance? Does the application of spatial-weight matrices differentiate analysis results, and are they justified and factually correct?

---

[6] For the formula and detailed descriptions, see Mitchell (2005).

The results of the analysis indicate the presence of significant and varied interregional relationships, which are different for specific years of the study and by the type of **W** matrix (Table 2).

Table 2. Values of global Moran's *I* statistics for waste using W matrices

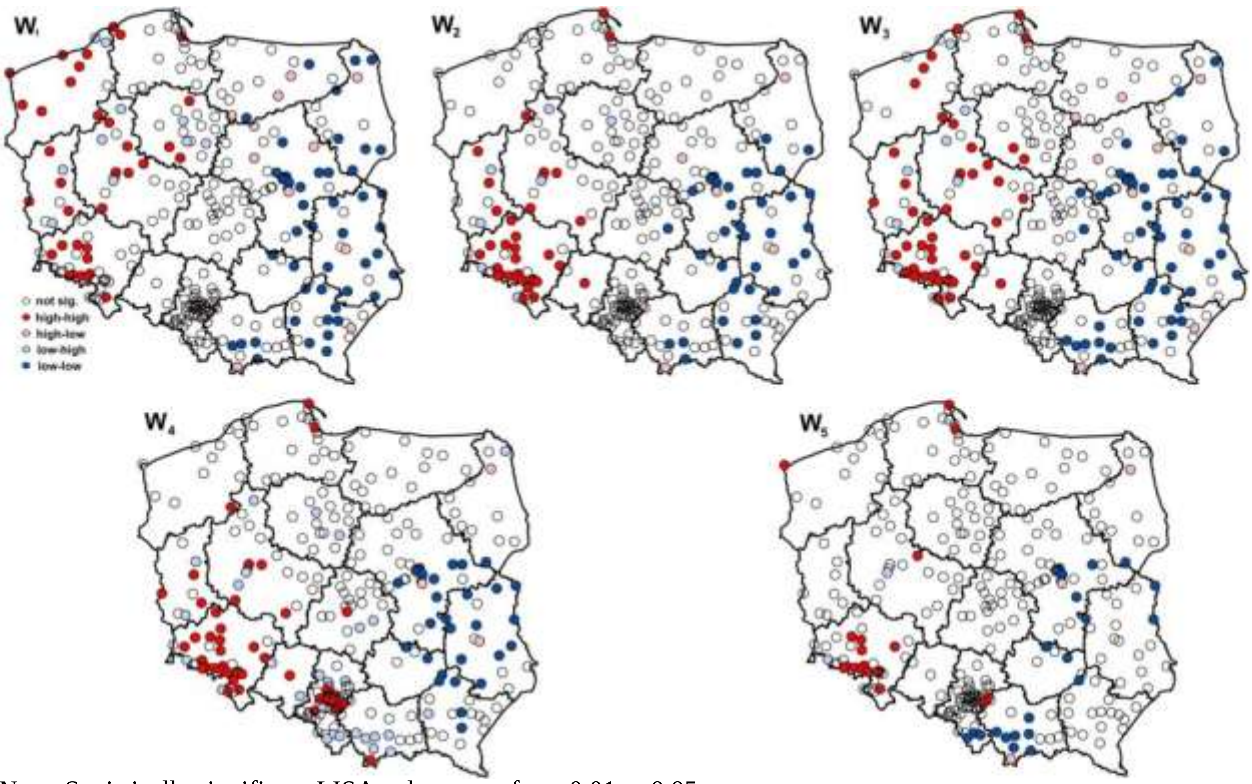
Moran's *I*	**$Waste_{av}$**	**$Waste_{2005}$**	**$Waste_{2010}$**	**$Waste_{2015}$**	**Similarity**
Up to 39 km – close (W_1)	0.20***	0.14***	0.19***	0.11***	**NO** Kruskal-Wallis = 12.18 (difference statistically significant); Dunn's Multiple Comparison Test, comparing all pairs of outcomes indicates significant differences between results obtained using $\mathbf{W_3}$ and $\mathbf{W_4}$)
Up to 300 km – medium (W_2)	0.16***	0.08***	0.15***	0.08***	
Up to 555 km – far (W_3)	0.10***	0.05***	0.08***	0.04***	
Directional (W_4)	0.35***	0.20**	0.36***	0.15***	
Random (W5)	0.26***	0.15***	0.20***	0.15***	

Notes: significance levels: α = 0.10*, 0.05 **, 0.01 ***.
Source: own elaboration in ArcGIS 9.3.

When applying all spatial-weight matrices, positive and statistically significant Moran's *I* statistic values were obtained. This means that, during 2005–2015, the cities displayed a tendency to cluster in space with respect to a similar volume of collected municipal waste. The strongest relationship characterised the amount of mixed collected waste for adjacency as defined by the directional matrix – $\mathbf{W_4}$. In the LISA processes (Figure 6), results from ESDA using this matrix appear more precise and more specific. On average and in the selected years, stronger positive and statistically significant spatial interactions also characterised volumes of waste in cities located close to each other, rather than in cities far away from one another. Moreover, data contained in Table 2 indicate that the strength of statistically significant spatial relationships decreased from 2005 to 2015 (an average flow of 57% in Moran's *I* statistic value in 2015 as compared to 2005). Thus, the production and management of waste in a city still significantly affect the volume of that phenomenon in its adjacent units; however, these processes seem to be more condensed (e.g., due to the new thermal waste treatment plants or a more effective local waste management policy). Overall, the results obtained from global autocorrelation indicate statistically significant differences in strength, specially between results obtained with $\mathbf{W_3}$ and $\mathbf{W_4}$ (Kruskal-Wallis post-test, Dunn's Multiple Comparison Test).

The values of global spatial autocorrelation are conditioned by local spatial regimes. Figure 6 shows the local indices of spatial autocorrelation determined based on matrices $\mathbf{W_1}$, $\mathbf{W_2}$, $\mathbf{W_3}$, $\mathbf{W_4}$, and $\mathbf{W_5}$.

Figure 6. LISA results of *Waste* $_{av.}$ and W matrices



Note: Statistically significant LISA values vary from 0.01 to 0.05.
Source: own elaboration.

According to Figure 6, the general pictures of spatial relations characterising the annual amounts of collected mixed municipal waste are similar regardless of the spatial-weight matrix included in the analysis. Namely, between 2005 and 2015, the cities located in western Poland were characterised by a high level of variable and created classes of similar values. On the other hand, in the units located in the eastern part of the country, in the selected voivodeships in Southern Poland and in Mazowieckie Voivodeship, the annual amounts of collected municipal waste are smaller than in other parts of the country.[7] Such a situation may relate to higher urbanisation rates and levels of economic development of cities in western Poland. On the other hand, the observation of the phenomenon's behaviour near Warsaw, which is the capital of Poland and one of the wealthiest cities in the country (excluded from the research), demonstrates that the amount of annual waste is related to the ‘wealth’ of a city

[7] Map of Polish Voivodeships; see Appendix.

(that is, higher levels of investments, issues related to ecology and environment, and the level of development of a city).

The most precise, detailed, and substantial picture of the occurring spatial relations was obtained by including the matrix of close distance (of a radius up to 30 km), long distance (up to 555 km), and the trend. The analysis of relations with the use of properly designed distance matrices revealed the occurrence of many specific units, the so-called outliers, in intercity relations. That cannot be said when one estimates the results obtained by including the randomly defined spatial-weight matrix. The picture of local relations obtained based on the randomly generated matrix is poor. The scale of the detected statistically significant spatial relations with the use of matrix $\mathbf{W_5}$ is significantly different from the scale of the revealed urban settlements playing a part in spatial processes with the use of other matrices, defined a priori. Several such units are on average twice higher than in the case of the analysis with the use of $\mathbf{W_5}$.

3.3 Spatial Modelling

The aim of the modelling below was not to specify the determinants of the explanatory variable but to try to assess the effect of applying different variants of spatial-weight matrices ($\mathbf{W_1}$-$\mathbf{W_5}$) on the results of an econometric analysis.

Economic theories (e.g., environmental Kuznets curve [EKC][8]) say that the higher the economic development and rate of urbanisation, the greater amount of waste that is produced and collected. Income level and urbanisation are highly correlated, and as disposable incomes and living standards increase, consumption of goods and services correspondingly increase, as does the amount of waste generated. Urban residents produce about twice as much waste as their rural counterparts (World Bank, 2016). On the other hand, the revenues of cities create effective waste management practices (e.g., source reduction, collection, recycling, composting, incineration, and dumping). To describe municipal waste, generation revenues per capita were introduced. Moreover, the ESDA proved that spatial interactions exist and affect the amount of collected waste in Polish cities. To determine whether varied values of

[8] The EKC is a curve (basic version is a second-degree polynomial – the inverted “U”) expressing a change in volume of environmental pollution depending on an increase in economic development. The idea of the classic EKC consists of seeking an inflection point or points for cubic functions (extrema of a function). The inflection point of the basic EKC version is at such a level of economic development, past which a potential drop in environmental pollution begins, for details see Antczak E. (2014) *Economic Development and Transfrontier Shipments of Waste in Poland - Spatio-Temporal Analysis*, ‘Comparative Economic Research. Central and Eastern Europe’, Vol. 17, Iss. 4 (Dec 2014), pp. 5-21.

spatial weights lead to significant differences in the analytical results and to determine the results if we introduce a weight matrix built without considering the nature of phenomena, five spatial weights matrices were introduced to panel and SUR models to determine whether weight matrices should be different for different years. Both a fixed-effects panel model (8) and a simultaneous equations model (12) with spatial autocorrelation of the error component were used, as follows:

$$\begin{aligned}\mathbf{y_t} &= \mathbf{X_t}\boldsymbol{\beta} + \boldsymbol{\alpha} + \boldsymbol{\varepsilon}_\mathbf{t} \\ \boldsymbol{\varepsilon}_\mathbf{t} &= \lambda \mathbf{W}_m \boldsymbol{\varepsilon}_\mathbf{t} + \mathbf{e_t}\end{aligned}, \tag{8}$$

where $\mathbf{y_t}$ is the vector of dependent variables, consisting of one observation on the dependent variable for every unit in the sample $\mathbf{y_t}$ = $(y_{1t}, ..., y_{nt})'$; here, the dependent variable is the amount of mixed municipal collected waste in kilograms per capita in cities from 2005 to 2015 (*Waste*). In addition, $\boldsymbol{\varepsilon}_\mathbf{t}$ is the vector of disturbance terms, and $\mathbf{X_t}$ is the matrix of independent variables, where $\mathbf{X_t}$ = $(X'_{1i}, ..., X'_{nt})'$; here, the revenues (R) of cities in PLN per capita. The $\boldsymbol{\alpha}$ is the vector of dummy variables introduced for each spatial unit as a measure of the variable intercept, $\boldsymbol{\alpha}$ = $(\alpha_1, ..., \alpha_n)'$. The $\boldsymbol{\beta}$ is the vector of parameters, and λ is the spatial autocorrelation coefficient. Moreover, $\mathbf{W}_m$ is the matrix of spatial weights $N \times N$, where m = (1, ..., 5). In addition, $\mathbf{e_t}$ is the vector of independent error terms, obeying the normal distribution; $E(\mathbf{e_t}) = 0, E(\mathbf{e_t}\mathbf{e}'_t) = \boldsymbol{\sigma}^2\mathbf{I}$ and n denotes the number of spatial units (number of cities where n = 1, ..., 279), as follows:

$$\begin{aligned}\mathbf{y}_t &= \mathbf{X}_t\boldsymbol{\gamma}_t + \boldsymbol{\xi}_t \\ \boldsymbol{\xi}_t &= \lambda_t \mathbf{W}_m \boldsymbol{\xi}_t + \boldsymbol{\mu}_t\,, \\ \mathrm{E}[\boldsymbol{\xi}_t, \boldsymbol{\xi}_s^T] &= \boldsymbol{\sigma}_{ts}, \mathbf{I}\end{aligned} \tag{9}$$

where γ_t denotes the vector of parameters, $\boldsymbol{\xi}_t$ is the vector of disturbance terms, and $\boldsymbol{\mu}_t$ is the vector of independent error terms. The remaining variables are as defined in Formula 8. The parameters of models were estimated using maximum likelihood (Tables 3 and 4).

Table 3. Results of the spatial panel estimation

Model	Spatial-Weight Matrix	α_1	λ
$Waste_{it} = \alpha_i + \alpha_1 R_{it} + \varepsilon_{it}$, $\varepsilon_{i,t} = \lambda \mathbf{W}_m \varepsilon_{i,t} + e_{i,t}$	$\mathbf{W_1}$	-0.18***	0.06
	$\mathbf{W_2}$	-0.16***	0.18***

	$\mathbf{W}_3$	-0.15***	0.28**
	$\mathbf{W}_4$	-0.1~~6~~8***	0.05**
	$\mathbf{W}_5$	-0.16***	0.18***

The values reported in parentheses are *p*-values. (*), (**), and (***) denote significance at 10%, 5%, and 1%, respectively. Balanced Panel: $n = 279$, $T = 11$, $N = 3069$. All variables were expressed in logarithms. *Waste* is the dependent variable, the amount of mixed municipal collected waste in kilograms per capita in cities from 2005 through 2015, and *R* denotes the independent variable as revenues of cities in PLN per capita.
Source: Own elaboration in RCran.

Table 4. Results of the SUR estimation

Model	Spatial-Weight Matrix	Year	α_{0n}	α_{1n}	λ_{nnn}
$Waste_{2005} = \alpha_{00} + \alpha_{10} R_{2005} + \varepsilon_{30}$, $\varepsilon_{30} = \lambda_{100} \mathbf{W}_m \varepsilon_{30} + e_{70}$	**$\mathbf{W}_1$**	**2005**	1.86***	0.13	-0.75**
	$\mathbf{W}_2$		1.84***	0.14	0.21
	$\mathbf{W}_3$		1.87***	0.13	-0.06
	$\mathbf{W}_4$		1.88***	0.12	0.17**
	$\mathbf{W}_5$		1.83***	0.14	0.16**
$Waste_{2006} = \alpha_{01} + \alpha_{11} R_{2006} + \varepsilon_{35}$, $\varepsilon_{35} = \lambda_{101} \mathbf{W}_m \varepsilon_{35} + e_{71}$	**$\mathbf{W}_1$**	**2006**	1.94***	0.11	-0.55*
	$\mathbf{W}_2$		1.91***	0.12	0.22
	$\mathbf{W}_3$		1.95***	0.11	0.02
	$\mathbf{W}_4$		1.94***	0.10	0.18
	$\mathbf{W}_5$		1.91***	0.12	0.11
$Waste_{2007} = \alpha_{02} + \alpha_{12} R_{2007} + \varepsilon_{40}$, $\varepsilon_{40} = \lambda_{102} \mathbf{W}_m \varepsilon_{40} + e_{72}$	**$\mathbf{W}_1$**	**2007**	2.10***	0.06	-0.62
	$\mathbf{W}_2$		2.08***	0.07	0.17
	$\mathbf{W}_3$		2.11***	0.06	-0.07
	$\mathbf{W}_4$		2.08***	0.07	0.05
	$\mathbf{W}_5$		2.08***	0.07	0.05
$Waste_{2008} = \alpha_{03} + \alpha_{13} R_{2008} + \varepsilon_{45}$, $\varepsilon_{45} = \lambda_{103} \mathbf{W}_m \varepsilon_{45} + e_{73}$	**$\mathbf{W}_1$**	**2008**	1.88***	0.12	-0.88
	$\mathbf{W}_2$		2.06***	0.06	-0.02
	$\mathbf{W}_3$		2.09***	0.05	-0.93*
	$\mathbf{W}_4$		2.06***	0.07	-0.02
	$\mathbf{W}_5$		2.06***	0.07	0.001
$Waste_{2009} = \alpha_{04} + \alpha_{14} R_{2009} + \varepsilon_{50}$, $\varepsilon_{50} = \lambda_{104} \mathbf{W}_m \varepsilon_{50} + e_{74}$	**$\mathbf{W}_1$**	**2009**	2.49***	-0.06	0.37***
	$\mathbf{W}_2$		2.51***	-0.07	0.29**
	$\mathbf{W}_3$		2.49***	-0.06	0.46**
	$\mathbf{W}_4$		2.49***	-0.06	0.13**
	$\mathbf{W}_5$		2.49***	-0.06	0.13**
$Waste_{2010} = \alpha_{05} + \alpha_{15} R_{2010} + \varepsilon_{55}$, $\varepsilon_{55} = \lambda_{105} \mathbf{W}_m \varepsilon_{55} + e_{75}$	**$\mathbf{W}_1$**	**2010**	2.68***	-0.12*	0.16***
	$\mathbf{W}_2$		2.69***	-0.12**	0.29**
	$\mathbf{W}_3$		2.68***	-0.12*	0.40**
	$\mathbf{W}_4$		2.67***	-0.12*	0.16**
	$\mathbf{W}_5$		2.67***	-0.15*	0.17
$Waste_{2011} = \alpha_{06} + \alpha_{16} R_{2011} + \varepsilon_{60}$, $\varepsilon_{60} = \lambda_{106} \mathbf{W}_m \varepsilon_{60} + e_{76}$	**$\mathbf{W}_1$**	**2011**	2.58***	-0.09	0.07
	$\mathbf{W}_2$		2.57***	-0.08	0.24
	$\mathbf{W}_3$		2.58***	-0.09	0.17**
	$\mathbf{W}_4$		2.59***	-0.10*	0.17

	$\mathbf{W}_5$		2.58***	-0.09	0.17**
$Waste_{2012} = \alpha_{07} + \alpha_{17} R_{2012} + \varepsilon_{65}$, $\varepsilon_{65} = \lambda_{107} \mathbf{W}_m \varepsilon_{65} + e_{77}$	$\mathbf{W}_1$	**2012**	2.61***	-0.09*	-0.07
	$\mathbf{W}_2$		2.61***	-0.10*	0.20
	$\mathbf{W}_3$		2.61***	-0.10*	0.27*
	$\mathbf{W}_4$		2.61***	-0.10**	0.11
	$\mathbf{W}_5$		2.61***	-0.10*	0.11
$Waste_{2013} = \alpha_{08} + \alpha_{18} R_{2013} + \varepsilon_{80}$, $\varepsilon_{80} = \lambda_{108} \mathbf{W}_m \varepsilon_{80} + e_{78}$	$\mathbf{W}_1$	**2013**	2.26***	0.002	0.22
	$\mathbf{W}_2$		2.25***	0.003	0.41***
	$\mathbf{W}_3$		2.24***	0.005	0.41**
	$\mathbf{W}_4$		2.20***	0.02	0.19
	$\mathbf{W}_5$		2.20***	0.02	0.31***
$Waste_{2014} = \alpha_{09} + \alpha_{19} R_{2014} + \varepsilon_{85}$, $\varepsilon_{85} = \lambda_{109} \mathbf{W}_m \varepsilon_{85} + e_{79}$	$\mathbf{W}_1$	**2014**	2.49***	-0.06	-0.34
	$\mathbf{W}_2$		2.51***	-0.07	0.40**
	$\mathbf{W}_3$		2.48***	-0.06	0.21
	$\mathbf{W}_4$		2.45***	-0.05	0.24***
	$\mathbf{W}_5$		2.45***	-0.05	0.24***
$Waste_{2015} = \alpha_{010} + \alpha_{1} R_{2015} + \varepsilon_{90}$, $\varepsilon_{90} = \lambda_{110} \mathbf{W}_m \varepsilon_{90} + e_{91}$	$\mathbf{W}_1$	**2015**	2.92***	-0.19*	0.19
	$\mathbf{W}_2$		2.88***	-0.17	0.29
	$\mathbf{W}_3$		2.91***	-0.18*	0.31
	$\mathbf{W}_4$		2.82***	-0.16	0.22**
	$\mathbf{W}_5$		2.82***	-0.16	0.22**
Ws' application: for α_0 and α_1, Kruskal-Wallis: difference not statistically significant; Dunn's multiple test, comparing all pairs of columns: not significant. For $\boldsymbol{\lambda}$, statistically significant differences between results from $\mathbf{W}_1$ and $\mathbf{W}_2$;					

The values reported in parentheses are *p*-values. (*), (**), and (***) denote significance at 10%, 5%, and 1%, respectively; $n = 279$, $T = 11$, $N = 3069$. All variables were expressed in logarithms. *Waste* is the dependent variable, the amount of mixed municipal collected waste in kilograms per capita in cities from 2005 through 2015, and *R* denotes the independent variable as revenues of cities in PLN per capita.
Source: Own elaboration in RCran.

This part of the article offers a spatial econometric analysis. However, the modelling was not aimed at specifying determinants of the annual amounts of collected mixed municipal waste but at evaluating the effect of using different variants of spatial-weight matrices ($\mathbf{W}_1$-$\mathbf{W}_5$) on the results of modelling. Moreover, there was an attempt made to identify the differences that occurred as a result of using the particular weight matrices. The results indicated the following:

- substantive and formal correctness of conclusions; therefore, the designed matrices are correct;
- consistency and stability of the modelling results, that is, statistically significant similarity of the values of parameter estimates;
- differences in the values of estimates of parameters reflecting the existing spatial processes; selection of spatial weights matrix should be dictated by the research

objective, and application of different matrices may differentiate conclusions concerning spatial processes.

Given the statistical significance of the estimates of parameters, spatial panel data models unequivocally indicate that, with a 1% increase in revenue to the city budget, there is an average decrease of approximately 0.2% in the annual amount of collected mixed municipal waste with other fixed factors. Nevertheless, using different spatial-weight matrices has an effect on the value of the estimate of parameter λ. Still, the direction of correlation remains unchanged. Spatial effects were the strongest when the average and long-distance matrices were included. Eventually, spatial relations of random factors in 'neighbouring' cities (defined in weight matrices) have an influence on shaping of the amount of waste in a particular unit, and the selection of the matrix moulds/determines the strength and significance of these interactions.

On the other hand, the added value regarding the results obtained from the SUR modelling is the fact that it is possible to not only explain the reasons for the shaping of the phenomenon, considering spatial processes, but also indicates a specific point in time where these processes were of the highest significance (Table 4). First, spatial autocorrelation of random components did not have an effect every year on the amount of collected municipal waste. Moreover, spatial effects were stronger in 2005, 2009 through 2010, and 2013 through 2015 with the use of different weight matrices. Furthermore, city revenues did not play a role every year in the decrease in the amount of municipal waste. However, the goodness-of-fit tests confirmed a lack of differences in the values of absolute terms and estimates of parameters next to the variable R. On the other hand, statistically significant differences were observed in the values of estimates of parameters λ especially with the use of matrices $\mathbf{W_1}$ and $\mathbf{W_2}$ (in the spatial panel data model, and spatial autocorrelation did not have a significant influence on shaping the amount of waste with the use of matrix $\mathbf{W_1}$, Table 3).

4. Discussion

Based on the results from both ESDA and econometric modelling, one may assume a main research hypothesis that the methods and tools of geostatistics can be used to specify and develop a spatial-weight matrix. Moreover, using the matrix in modelling does not have an influence on the stability of models over time. The results of the analysis are substantially correct and not mutually exclusive. The results of spatial modelling were consistent and confirmed by proper statistical tests. Nevertheless, the values of estimates of spatial autocorrelation parameters were different. This concerns both signs of estimates and strength of influence as well as statistical significance. However, it was found that it resulted from a type of specific matrix included in the analysis and the nature and dynamics of the phenomenon. Still, there are many unexplained issues and questions left that inspire further research, such as the following:

- How can a directional matrix be built based on the semivariogram?
- Is there a possibility of attributing values to weights in the matrix based on this geostatistical measure?
- Should one include a matrix designed for each period in the model, and how can it be done?
- Is the construction of the matrix for the particular periods of the research justified and essential?
- Is the construction of other or specific matrices essential for external variables compared to internal variables in SLX models?

5. Conclusions

Substantially, results of the analyses of applying different spatial-weight matrices did not exclude but rather supplemented (enriched and extended) one another. The most precise picture of spatial (global and local) dependences was received by including the selected distance and directional matrices in the analysis. Moreover, models with different weight matrices (directional, selected distance, or exogenous matrix) were sensitive to a change in that matrix in such a way that the values of the assessed parameter at the regressors did not significantly change. In a few cases, considering the direction of the influence on the endogenous variable, it determined a slight increase of the values. There was, however, a change in the assessed value of the spatial autoregression or autocorrelation parameter (strength of influence). Nevertheless, modelling results were still substantially accurate, and the application of different (endogenous with ‘dedicated’ weights) matrices was justified. A problem appeared to be the inclusion of matrices whose spatial weights change recurrently (from period to period) in panel econometric modelling. In addition, the creation of such matrices is time-consuming.

References

1. Aldstadt, J., Getis A., 2006. Using AMOEBA to Create a Spatial Weights Matrix and Identify Spatial Clusters. Geogr. Anal. 38, 327-343.
2. Angulo A., Burridge P., Mur J., 2017. Testing for Breaks in the Weighting Matrix. Documentos de Trabajo, Facultad de Ciencias Económicas y Empresariales, Universidad de Zaragoza. http://www.dteconz.unizar.es/DT2017-01.pdf (accessed 17.08.16).
3. Anselin L., 1988. Spatial Econometrics: Methods and Models, Springer, Netherlands.
4. Antczak, E., 2014. Economic Development and Transfrontier Shipments of Waste in Poland - Spatio-Temporal Analysis. Comparative Economic Research. Central and Eastern Europe. 17 (4) 5-21.
5. Bao S.D., 2000. Soil and Agricultural Chemistry Analysis. Chin. Agric. Press, Beijing, 178-200.
6. Benjanuvatra, S., Burridge, P., 2015. QML estimation of the spatial weight matrix in the MR-SAR model, http://rri.wvu.edu/wp-content/uploads/2013/07/Fullpaper_3.A.3.pdf (accessed 17.08.16).
7. Chojnicki, Z., Czyż T., 1975. Zastosowania analizy trendu powierzchniowego w geografii [Trend surface analysis in geography], Przegląd Geograficzny T. XLVII, Z. 2, 235-261.
8. Elhorst, J., 2014, Spatial Econometrics: From Cross-Sectional Data to Spatial Panels, Springer: New York.
9. Fernández, E., Mayor, M., Rodríguez, J., 2009. Estimating spatial autoregressive models by GME-GCE techniques, Int. Reg. Sci. Rev. 32, 148-172.
10. Getis, A, Aldstadt, J., 2004. Constructing the Spatial Weights Matrix Using a Local Statistic Spatial, Geogr. Anal. 36, 90-104.
11. Griffith, D. A., 1996. Some Guidelines for Specifying the Geographic Weights Matrix, in Arlinhaus, S., L., (Eds.), Practical Handbook of Spatial Statistics, Boca Raton: CRC, pp. 65-82.
12. Harris, R., Moffat, J., Kravtsova, V., 2011. In search of 'W', Spat. Econ. Anal. 6, 249–270.
13. Hondroyiannis, G., Kelejian, H., H., Tavlas, G., S., 2009. Spatial Aspects of Contagion Among Emerging Economies, Spat. Econ. Anal. 4, 191-211
14. Kelejian, H., Piras, G., 2014, Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes, Reg. Sci. Urb. Econ. 46,140-149.
15. Kooijman, S., 1976. Some Remarks on the Statistical Analysis of Grids Especially with Respect to Ecology, Ann. Syst. Res. 5, 113-132.
16. Krige, D., G., 1996. A practical analysis of the effects of spatial structure and of data available and accessed, on conditional biases in ordinary kriging, in Baafi, E.,Y., Schofield, N., A., (Eds), Geostatistics Wollongong, Kluwer Press, pp. 799-810.
17. Lee, D., T., 1982. On k-Nearest Neighbor Voronoi Diagrams in the Plane, IEEE Trans.Comp., Vol. C-31, No. 6, 478-487.
18. Lefever, D., W., 1926, Measuring Geographic Concentration by Means of the Standard Deviational Ellipse, Am. J. Sociol. 32(1), 88–94.
19. Matheron, G., 1963. Principles of Geostatistics, Econ. Geol. 58, 1246–1266.
20. Mur, J., Paelinck, J., 2010. Deriving the W-matrix via p-median complete correlation analysis of residuals, Ann. Reg. Sci. 47, 253–267.

21. Qu, X., Lee L., 2015. Estimating a spatial autoregressive model with an endogenous spatial weight matrix. J.Econom. 184(2), 209-232.
22. Robinson, M., Dietrich, S., 2016. An Introduction to Spatial Autocorrelation and Kriging. https://sites.ualberta.ca/~lkgray/uploads/7/3/6/2/7362679/slides-autocorrelation_kriging.pdf (accessed 17.06.24).
23. Stewart, M., Zhukov, Y., 2010. Choosing Your Neighbors: The Sensitivity of Geographical Di_usion in International Relations. http://isites.harvard.edu/fs/docs/icb.topic668784.files/Stewart_Zhukov_ISA.pdf (accessed 17.06.24).
24. Tobler, W., 1970. A computer movie simulating urban growth in the Detroit region. Econ. Geog. 46 234-240.
25. Waller, L., A., Gotway C., A., 2004. Applied Spatial Statistics for Public Health Data. Wiley, New Jersey.
26. Wang, B., Shi, B., Inyang, H., I., (2008). GIS-based quantitative analysis of orientation anisotropy of contaminant barrier particles using standard deviational ellipse, Soil Sedim. Contam. 17(4), 437–447.
27. Wang, B., Shi, W., Miao, Z., (2015). Confidence Analysis of Standard Deviational Ellipse and Its Extension into Higher Dimensional Euclidean Space, PLoS One. 10(3): e0118537.
28. Zawadzki, J., 2002, Badania korelacji przestrzennych zawartości wybranych pierwiastków śladowych w glebach Warszawy i okolic, Ochrona Środowiska. 24 (4), 17-26.

Appendix. Map of Polish Voivodeships