


Creating Spatial Models of Demographic Processes Using Cluster Analysis for Demographic Policy Planning in Bulgaria

Penka KASTREVA*¹, Emilia PATARCHANOVA¹

** Corresponding author*

¹ South-West University "Neofit Rilski", Faculty of Mathematics and Natural Sciences, Blagoevgrad, BULGARIA

✉ penkakastreva@gmail.com  <https://orcid.org/0000-0002-5863-532X>

✉ emilia_patarchanova@swu.bg  <https://orcid.org/0000-0002-1806-557X>

DOI: 10.24193/JSSP.2021.2.05

<https://doi.org/10.24193/JSSP.2021.2.05>

Keywords: *demographic processes, depopulation, spatial models, cluster analysis, Hot Spot Analysis, Spatial Autocorrelation, sustainable municipalities, Bulgaria*

ABSTRACT

Despite the demographic policy conducted by the state, demographic processes in Bulgaria have been negative for more than 30 years, with spatial differences in their manifestation and results. The main goal of our research is to find demographically stable municipalities that can be accepted as a model of demographic policy implementation to achieve positive changes in the population growth. For this purpose we investigated and identified the changes in the main demographic indicators of population for 2011 and 2019, using cluster analysis. We created spatial models of these demographic processes showing that the number of demographically sustainable municipalities is lower than that of the ones in an advanced depopulation process. Several statistical methods (tools) of specialized software - cluster analysis, Hot Spot Analysis, Spatial Autocorrelation were used. Our hypothesis that the demographic stability of a municipality is most strongly influenced by its economy was confirmed. The analysis proved that demographically stable municipalities are represented by the largest cities and economic centres of Bulgaria. A large number of them, located mainly in mountainous and/or rural areas of Bulgaria, are highly depopulated. The significant socioeconomic inequalities in Bulgaria are a major factor that stimulates internal migration to economic centres and deepens the depopulation of vast parts of the country. They are home to older people and, therefore, these municipalities record very low birth rate and high mortality.

1. INTRODUCTION

Bulgaria has been experiencing a demographic crisis since the 1990s. According to data from the National Statistical Institute (NSI, 2019), the population of the country is already below 7 million people (6,951,482 people in 2019). Compared to the beginning of the century (2001), the population of Bulgaria has decreased by 981,502 people. This negative trend is a result of declining birth rate, rising mortality, negative natural growth, active external and

internal migration. The state is trying to conduct a policy by proposing various projects with specific measures to influence the demographic development of Bulgaria. Over the years, different incentives have been introduced to increase birth rate, such as lump sum payments for the birth of a child, paid maternity leave of 410 calendar days for each child, additional leave for raising a child up to 2 years of age, etc. (MLSP, 2006), yet without success. Although these measures are valid for the whole country, there are visible spatial differences - population growth in some territories and

depopulation in others. A brief look at recently published scientific papers shows that scientists are increasingly and actively using GIS and statistical methods, which proves their visible role in modern research. When processing data collected in different scientific fields, most articles quote at least one “p-value” to emphasize the significance of the results they analyse. At the same time, more and more articles present the results of a multifactorial statistical analysis of complex sets of data. This use of statistical methods is also suitable for demographic processes research.

The demographic crisis is not only a national problem. According to the International Institute for Strategic Studies, Europe is also experiencing a long-term demographic decline; its share in the world's population has already halved in the last 60 years (IISS, 2020; European Commission, 2020).

The situation is catastrophic in south-eastern Europe. Population is shrinking and ageing, and fewer children are being born. Since 1990, Albania has lost 12% of its population, Romania 18%, Bulgaria 21% and Moldova 32%, Bosnia and Herzegovina 24%, Serbia 9% and Croatia 15% (Judah, 2021).

These negative trends in population dynamics have been studied by many authors (Willekens, 2014; Shubat et al., 2016; Antipova and Fakeyeva, 2012; Yang et al., 2016; Ocovsky, 1991; Makra et al., 2018) and research institutes, such as the Max Planck Institute for Demographic Research, the Vienna Institute of Demography (VID), the Austrian Academy of Sciences (ÖAW) and the International Institute for Applied Systems Analysis (IIASA). The most serious population decline is recorded in rural and mountainous areas in different countries (Delgado Viñas, 2013; Saito, 2007; Patarchanov, 2010). Persistent depopulation processes are also characteristic of another important group of territories, namely border areas (Grasland, 1990; Hooz, 1992).

In the light of these international trends we decided to investigate the demographic situation in Bulgaria. The main goal is to identify areas with different demographic processes by conducting our study at municipality level which is the lowest spatial scale. Since the statistical methodology, namely the cluster analysis, is considered to be an appropriate tool for this purpose, the cluster analysis with ArcGIS software was used in this research.

The objectives of this study are:

- (1) to identify groups of regions with significant differences in the dynamics of the selected demographic indicators;
- (2) to study and quantify the relationship between the demographic processes and population decline in Bulgaria;
- (3) to create spatial models for identifying clustering trends of the demographic processes that affect the number of population.

2. THEORY AND METHODOLOGY

Clustering is a process of dividing the information into groups, known as clusters, based on several characteristics (variables). The purpose of grouping is that each cluster contains similar objects (observations), and the objects from different clusters differ significantly. The similarity of objects in the cluster is characterized by common properties (attributes). Thus, each object (observation) of the common dataset will belong to a cluster with the closest average value (cluster centres or centroids). The clustering is done by minimizing the sum of squares of the distances between the data and the corresponding centroid of the cluster. Various mathematical procedures are used as a measure of proximity, the most known being the Euclidean distance (Kumar, 2019):

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2}$$

where:

p - number of variables;

d_{ij} - Euclidean distance between i-th j-th spatial object;

$x_{i1}, x_{i2}, \dots, x_{ip}$ and $x_{j1}, x_{j2}, \dots, x_{jp}$ are their variables.

In statistical science, there are various methods for performing cluster analysis, which can be classified as: *Hierarchical cluster analysis* when the number of clusters is not predetermined; *Non - hierarchical cluster analysis* at a predetermined number of clusters (known as the k - means clustering method) and a combination of both methods. In ArcGIS software, the Mapping Clusters tool contains various algorithms for creating clusters: Cluster and Outlier Analysis (Anselin Local Moran's I), Hot Spot Analysis (Getis-Ord G_i^*), Optimized Hot Spot Analysis and Grouping Analysis. The latter creates clusters based on many variables, simultaneously.

The Analyzing Patterns tools are used before mapping the clusters. The aim is to investigate whether the elements in the dataset or their values are spatially clustered. These tools include statistics: Average Nearest Neighbour, High/Low degree of clustering (Getis-Ord General G), Incremental Spatial Autocorrelation, Multi-Distance Spatial Cluster Analysis (Ripley's K Function), Spatial Autocorrelation (Global Moran's I). Global and local calculations of Moran I index (Global Moran's I, Anselin Local Moran's I and Getis-Ord G_i^*) are able to find out whether the high/low values are closely grouped or more scattered in the spatial model. These methods are used in our research. These algorithms calculate statistics z-scores (standard deviations) and p-values (probabilities), which are measures of statistical significance and show

whether to reject the null hypothesis. Both statistics are associated with the standard normal distribution (ESRI, 2014). The high positive z-scores mean intense grouping of high values (hot spots), and the low negative z-scores indicate more intense grouping of low values (cold spots). In order to correctly interpret the results of the analysis, in the following lines we will mention the most important for these statistics without dwelling on the mathematics in them.

2.1. Global Moran's I

The first step in performing a cluster analysis is to examine the data for the presence of autocorrelation. Moran's I is one way to test for autocorrelation. Moran's Index is a correlation coefficient that has a value from -1 to 1 as: -1 (Negative Spatial Autocorrelation) shows dispersed pattern; 0 is no autocorrelation (perfect randomness i.e. the pattern is neither grouped nor dispersed); +1 (Positive Spatial Autocorrelation) shows clustered pattern (Zhang et al., 2008).

The Spatial Autocorrelation tool (Global Moran's I) calculates five values: Moran's Index, Expected Index, Variance, z-score, and p-value. With their help and the values given in Table 1 for the statistically significant z-score and p-value, we can summarize the following cases when interpreting the results: at very high positive z- score associated with very small p-value, the pattern is clustered; at very low negative z- score associated with very small p-value, the pattern is dispersed; if the z-score is between -1.65 and +1.65, and the p-value is greater than the statistically significant ones (0.10, 0.05, 0.01), then the pattern is random and the null hypothesis cannot be rejected. This means that it is possible that the distribution of characteristic values in the spatial model is the result of random spatial processes. (ESRI, 2014).

Table 1. Critical p-values and z-scores for different confidence levels.

z-score(standard deviations)	p-value (probability)	Confidence Level (%)
< -1.65 or > +1.65	< 0.10	90
< -1.96 or > +1.96	< 0.05	95
< -2.58 or > +2.58	< 0.01	99

Source: ESRI.

2.2. Hot Spot Analysis (Getis-Ord Gi*)

Hot Spot Analysis is performed using Getis-Ord Gi statistics. Applied to an entire spatial dataset, statistically significant local groups of high (or low) values are identified. This statistics results in a z-score and a p-value for each attribute. The reasoning is the same as when checking for the presence of spatial

autocorrelation. A high z-score and small p-value indicate a spatial clustering of high values (hot spots). A low negative z-score and small p-value indicate a spatial clustering of low values (cold spots). Z-score close to zero indicates the absence of spatial clustering.

A new field (Gi_Bin) is automatically created in the attribute table, which reflects the statistical significance with values +/- 3, +/- 2, +/- 1 corresponding to 99%, 95%, 90% confidence levels, and bin = 0 corresponds to statistically insignificant values.

2.3. Cluster and Outlier Analysis (Anselin Local Moran's I statistic)

Anselin Local Moran's statistics identify statistically significant hot spots, cold spots, and spatial outliers. The algorithm calculates a local Moran's Index value, a z-score, p-value and a code representing the cluster type for each statistically significant feature. The following types of statistically significant clusters can be identified, calculated for a 95 percent confidence level: cluster of high values (HH), cluster of low values (LL), outlier in which a high value is surrounded primarily by low values (HL), and an outlier in which a low value is surrounded primarily by high values (LH).

2.4. Grouping Analysis

When using GIS, it is possible to classify objects based on multiple attributes simultaneously (Kastreva and Kastreva, 2019; Kastreva and Kastreva, 2020). This data classification (grouping) process in ArcGIS is performed with the Grouping Analysis tool. Grouping is based on attribute values of the objects and additional space or time constraints. When the "No spatial constraint" parameter is selected for the Spatial Constraints parameter, the analysis is performed with the K-Means algorithm. In this case, the objects are grouped using data only for spatial proximity. In fact, objects may not be in close proximity of one another in time and space in order to be part of the same group. K – Means defines the centre of gravity (centroid), which is the average of the group of objects, after which it is named. The Euclidean distance is used as a measure of proximity between two objects. It is calculated as values between each object of the cluster and its centroid.

The K-Means algorithm in Arc GIS software calculates the parameter R^2 . The greater the value of R^2 for a particular variable, the better the clustering efficiency (ESRI, 2014). The value of R^2 is calculated by the sum of the mean squared errors and these are the differences between the value of each variable and its average in the group (Euclidean distances). In ArcGIS, this difference is designated as SSE (Sum of Squared Error), called global minimum.

Data grouping is represented graphically as a map image, with individual groups displayed in

different colours. Statistical calculations are presented in text and graphical form with a report automatically created by the software, which presents in summary the analysis of the data by groups and by variables.

2.5. Graphical representation of work processes

Figure 1 presents briefly the work processes performed based on the available data.

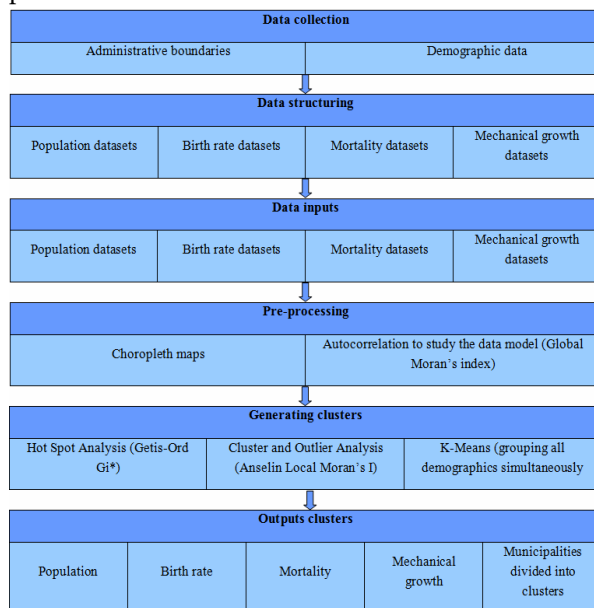


Fig. 1. Simplified scheme of the clustering process.

2.6. Data

Several popular demographic indicators were selected for the realization of cluster analysis: number of population, birth rate, mortality, mechanical growth. Only absolute values of these indicators were used. They are leading indicators for characterizing the demographic situation from the spatial perspective. The data for these indicators is provided by NSI. Two sets of data were used - from the last census in Bulgaria in 2011 and the current (December 31, 2019) data for the population of the country.

The municipality was chosen as a territorial unit for the realization of our research, because most cluster analyses in Bulgaria were made at a higher hierarchical level – district (Tsvetkov, 2020), which in our opinion has certain limitations. Not a very realistic idea of the state of the demographic processes could be obtained at this level. The reason for this is that differences in population growth exist between urban and rural municipalities although they are part of the same district. Therefore, in cases with a significant disproportion in the demographic processes between an urban centre and a depopulated area, districts are not the best territorial level to define these processes. In contrast, when the cluster analysis is performed at the municipal level, the imbalance in the spatial

manifestation of the state of the demographic processes is obvious.

3. RESULTS AND DISCUSSION

3.1. Preliminary analysis of the demographic processes

To identify the spatial models of the demographic processes, we created choropleth maps, which are quite useful for showing municipalities with high/low values of the demographic indicators and the distribution of variables. In order to better identify clustering trends of demographic processes that affect the number of population, we calculated the difference between the values of demographic indicators for the two years under study. Such maps are designed to illustrate the change in the number of population, birth rate, mortality and mechanical growth. They present the differences in the values of the studied indicators for 2011 and 2019.

Figure 2a shows that in only 28 municipalities there is an increase of the population for the analysed years.

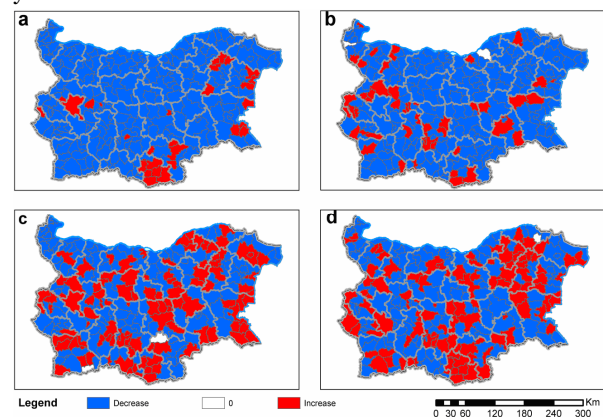


Fig. 2. Choropleth maps for population change for 2011 and 2019: a) map of differences in population number; b) map of differences in birth rate; c) map of differences in death rate; d) map of differences in mechanical growth.

They can be divided in three groups. One of the groups includes municipalities represented by some of the largest cities in the country (Sofia, Plovdiv, and Varna) and medium sized cities such as Shumen and Haskovo. Another group is formed by the Black Sea municipalities (Nessebar, Sozopol, Primorsko, Aksakovo, Devnya, etc.). The third group includes those with Muslim population (municipalities around Kardzhali) and others located in the northeast of the country. We believe that the main reason for population growth in the first and the second groups is due to the higher employment opportunities and incomes.

The spatial differences in the birth rate dynamics in the analysed years 2011 and 2019 are shown in Figure 2b. This negative trend is leading in 212 out of 264 municipalities. More worrying is the fact

that the largest cities in the country, namely Sofia, Plovdiv, Varna, Burgas, Shumen, Stara Zagora, Dobrich, Ruse, Pleven, etc are part of this group. Therefore, the observed population growth is not determined by an increase in the birth rate in the case of the first and second groups. An increase in the birth rate is recorded in only 50 municipalities, whilst it is stable in two. Seven municipalities that record population growth (Fig. 1a) also have an increase in the birth rate. These are mostly small towns or villages (Krumovgrad, Kirkovo, Jebelt, Mineralni Bani, Gorna Malina, Treklyano, etc.), which are not leading economic centres, but have a developed sector of small and medium enterprises, mainly in the services and food industry.

In Figure 1c we can also note the spatial differences in the dynamics of mortality in 2011 and 2019. In the case of 155 municipalities (59%) there is a positive trend in mortality decrease; Hadjidimovo and Dimitrovgrad register no changes in the mortality rate, while in the other 106, mortality increase is observed. Almost all 28 municipalities with an increase in population (Fig. 2a) also show an increase in mortality (Fig. 2c). Exceptions are only the municipalities of Krumovgrad, Aksakovo, Devnya, Beloslav. This analysis shows that mortality dynamics has no direct relation to the changes in population, because this dependence is directly proportional in only four of the 264 municipalities in Bulgaria.

Figure 2d illustrates the spatial differences in the mechanical growth of the population in 2019 compared to 2011. The mechanical growth has negative values in 156 municipalities in the country, i.e. these are territories that are losing their demographic potential. The mechanical growth in 107 municipalities has positive values. According to this indicator, only Alfatar shows no change in value.

When comparing the municipalities with growing population (Fig. 2a) and those with positive mechanical growth (Fig. 2d) two important tendencies are revealed. The leading trend is the increase in the mechanical growth of population in municipalities, although their population is decreasing due to the decreasing birth rate and the increasing mortality rate. Therefore, there is still active internal migration, targeting settlements with more developed economies and services, but not leading to an increase in their population. The second trend covers a small number of municipalities with positive mechanical growth and population growth. On one hand, these are the largest cities and/or economic centres (Stolichna, Plovdiv, Varna, Pirdop), small municipalities in close proximity to them (Bozhurishte, Gorna Malina, Nikolaevo) and those in Black Sea area (Nessebar, Sozopol, Primorsko). The reason for these positive trends is their economic development. On the other hand, these are municipalities from the Kardzhali region and the

Shumen region, traditionally inhabited by Muslims (Varbitsa, Hitrino, Venets, Kaolinovo, N. Kozlevo, Mineralni Bani, etc.).

There is a decrease in the movement of population to the capital city, despite the positive growth registered in 2019. According to NSI (2019), there is a tendency of population decrease in the capital city by almost 50% compared to the previous 10 years. In 2001, in the capital city, the settlers were 13,394 more than the emigrants; in 2011, they decreased to 6,252, while in 2019 they were 3,230. In Plovdiv – the second largest city in the country, the trend is reversed, namely the values of the positive mechanical growth increase due to its accelerated economic development.

No parallel can be drawn between municipalities with increasing mortality and those with negative mechanical growth, which is visible in the cartograms. Territorially, there are very few coincidences, which prove the need for a regional approach to the demographic policy planning, which takes into account the specific features of the demographic situation. This is a reason for continuing with the cluster analysis.

Can the statistical tools for cluster mapping show us more compared to choropleth maps?

3.2. Spatial models of change in population number between 2011 and 2019

The global and local calculations of the Moran index (Global Moran's I, Anselin Local Moran's I and Getis-Ord G_i^*) were used to study the spatial model of change in the number of population.

3.2.1. Global Moran's I

Spatial autocorrelation revealed weak clustering tendencies. The results are given in Figure 3.

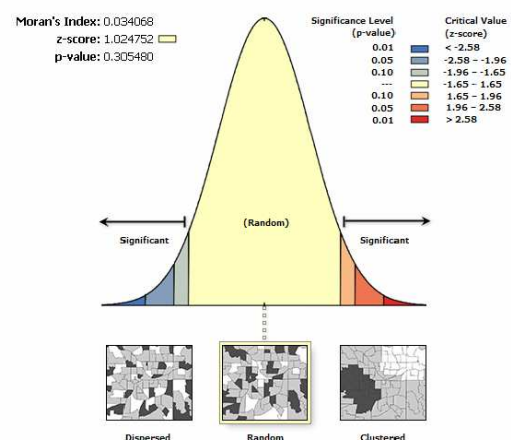


Fig. 3. Spatial Autocorrelation Report.

They were obtained by selecting the following parameters: Conceptualization of Spatial Relationships: Inverse Distance and Distance Threshold: 28403.8914

Meters. The Moran index shows a positive autocorrelation. The z-score (1.024752) falls between -1.65 and +1.65, and the p-value (0.305480) is greater than the critical values shown in Table 1. Therefore, the p-value is not statistically significant. These results show that we cannot reject the null hypothesis and that the pattern does not differ significantly from the random one.

3.2.2. Hot Spot Analysis (Getis-Ord G_i^*)

Random spatial patterns sometimes show some degree of grouping (ESRI, 2014). To verify this and identify the areas with higher rates of increase and decrease in population number, we conducted the Hot Spot Analysis. We used Getis-Ord G_i^* and Anselin Local Moran's I statistics to identify statistically significant hot spots, cold spots, or spatial outliers from the Mapping Clusters tools.

The Hot Spot Analysis performed with the statistical information, which presents the difference in the number of population, shows the following results, reflected in Figure 4a. Some 88 municipalities in the country have negative z-scores (from -3.342444 to -0.004081), but only 16 of them are identified as cold spots (Fig. 4a).

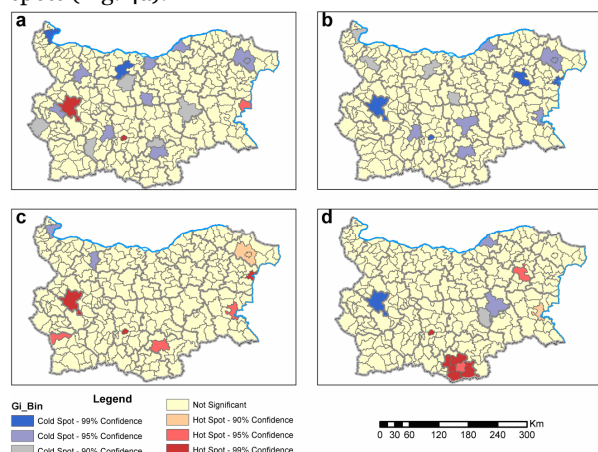


Fig. 4. Hot Spot Analysis of changes in the demographic processes for 2011 and 2019: a) Hot and cold spot distribution of population change; b) Hot and cold spot distribution of differences in birth rate; c) Hot and cold spot distribution of differences in death rate; d) Hot and cold spot distribution of differences in mechanical growth.

The corresponding p-values of the cold spots vary from 0.00083 to 0.092468. These are the municipalities of Pleven, Vidin, Vratsa, Ruse, Pernik, Dobrich, Gabrovo, Svishtov, Haskovo, Pazardzhik, Kyustendil, Sliven, Velingrad, Lovech, Dimitrovgrad. The most significant decline in population is found in Pleven Municipality - by 11,784 people, and the least - in Dimitrovgrad Municipality, by 6,663 people. Cold spots are distributed on three levels of confidence ($G_i^*_{Bin} = 1, 2$ and 3), and in the legend are marked with 90%, 95% and 99% confidence.

There are 176 municipalities with a positive z-score, but unfortunately only three municipalities stand out as hot spots with a significant increase in population, particularly Stoliczna, Plovdiv and Nessebar. Hot spots are distributed on two levels of confidence (95% and 99%). This result is another confirmation of our hypothesis, presented at the beginning of the study. Statistically insignificant values with coefficient $G_i^*_{Bin} = 0$ refer to 245 municipalities (highlighted in beige), although on the choropleth map they have visibly declining population.

The result of Hot Spot Analysis confirms that cold spots with almost random distribution predominate. Weak cluster tendencies are observed in the municipalities of Pleven and Lovech and a second cluster is formed for Dobrich and Dobrich-selska municipalities.

3.2.3. Cluster and Outlier Analysis

Cluster and outlier analysis tools were also used to verify and supplement the Hot Spot Analysis using the Anselin Local Moran's I statistics. This allows detecting areas with abnormalities, if they exist. The picture with the hot/cold spot analysis was supplemented as follows (Fig. 5a).

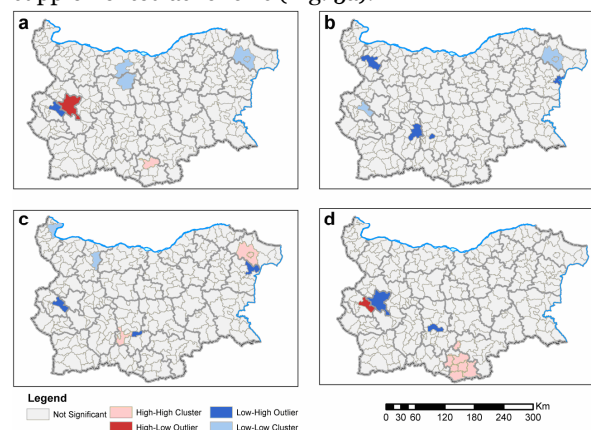


Fig. 5. Cluster and outlier distribution of demographic characteristics for 2011 and 2019: a) Cluster and outlier distribution of the population change for 2011 and 2019; b) Cluster and outlier distribution of differences in birth rate; c) Cluster and outlier distribution of differences in death rate; d) Cluster and outlier distribution of differences in mechanical growth.

Five municipalities with positive values for the coefficient I (LMiIndex) were identified, four of which are clusters with low values (LL) - the municipalities of Lovech, Plevan, Dobrich and Dobrich-selska and Kardzhali with high values (HH). The negative values for LMiIndex are typical for the two municipalities of Stoliczna and Pernik. They fall into the groups of spatial outliers, with a very high value surrounded by low values (HL) - Sofia; and a very low value surrounded by high values (LH) - Pernik. In both cases, the p-value is small enough (from 0 to 0.018193) to

conclude that both clusters and spatial outliers are considered statistically significant. This analysis also shows a weak grouping with mostly cold spots.

Therefore, the demographic indicators and the processes show a weak dependence (correlation) on the changes in population. They confirm the impossibility of the state's current policy on stimulating the birth rate to have positive results. Hence, other aspects of this policy must be sought in order to have a positive outcome.

3.3. Spatial patterns of the demographic processes and their impact on population decline in Bulgaria

3.3.1. Hot Spot Analysis

Hot Spot Analysis, including spatial autocorrelation, was also performed on the models of demographic processes that have the greatest impact on population decline. The levels of birth and mortality rates and the mechanical growth of population were chosen as such. To identify the clustering trends of the demographic processes, we studied their change for the reference years 2011 and 2019.

Table 2 presents the results of autocorrelation for each of the demographic processes. As can be seen from the values of z-score and p-value, the first two models of birth rate and mortality, similar to that of change in the number of population, are random. The data presenting the changes in the migration processes of population convincingly show a cluster model with less than 1% probability that this model is a result of a chance. Hence, these analyses confirmed once again the previous conclusion, namely: the demographic processes of births and deaths have almost no effect on changes in the population number at the municipal level. However, the results of the analysis of migration processes are not the same in terms of population growth - the cluster model shows a very high degree of influence.

Table 3. Statistics of autocorrelation of the demographic processes.

	Moran's Index	z-score	p-value
Birth rate	-0.001572	0.057245	0.954350
Mortality	0.051873	1.443915	0.148763
Mechanical growth	0.139138	3.767506	0.000165

A Hot Spot Analysis of the changes in the demographic processes was performed to identify the distribution of hot and cold spots. The graphical representation of the models is given in Figures 4b, 4c and 4d. Figure 4b presents only the cold spots on the

levels of birth rate for 2011 and 2019. The lack of hot spots means that there are no municipalities in Bulgaria in which a statistically significant increase in birth rate could be observed and identified by the software. This is a very worrying fact in terms of the possibilities for overcoming the demographic crisis. This also means that birth control measures have not achieved the expected outcome and the policy is not effective. Some 64 municipalities have a negative z-score, of which 15 are defined as cold spots, divided into three categories of statistical significance (Gi_Bin) -3, -2, -1. Statistically significant decrease in birth rate is reported in the municipalities of Stoliczna, Varna, Plovdiv and Shoumen followed by Stara Zagora, Dobrich, Dobrich-Selska, Ruse, Pazardzhik, Burgas, Haskovo, Pleven, Vidin, Gabrovo and Montana. It should be noted that the territory is not only represented by the largest cities, but also by cities that are administrative district centres. This is an indication of change in the pattern of reproductive behaviour of population. The population of reproductive age, which is settled in large urban centres, focuses mainly on professional development and ensuring financial stability, rather than on child raising.

Figure 4c reflects the distribution of hot and cold spots related with the difference in the levels of mortality for the studied years. Some 155 municipalities have a negative z-score. Only in the case of Vidin and Byala Slatina the significance is defined by reduction in mortality. There are 109 municipalities with positive z-score. Eight of them were identified as hotspots that show a significant increase in mortality. These are the municipalities of Sofia, Plovdiv and Varna, followed by Blagoevgrad, Haskovo, Burgas, Dobrich and Dobrich-selska. We assume that this is due to the high levels of stress and an ecologically deteriorating urban environment. They are divided into three categories of statistical significance (Gi_Bin) 3, 2, 1.

Figure 4d presents a cluster model of the mechanical growth of population. Compared to 2011, in 2019 a positive mechanical growth is reported in nine municipalities and a negative one in four. A cluster with a cold spot is identified for two municipalities - Sliven and Nova Zagora, while a hot spot cluster is identified for six others - located around the town of Kardzhali. For the remaining 251 municipalities (with positive and negative z-score) the number of settled or displaced people is statistically insignificant.

3.3.2. Cluster and Outlier Analysis

To identify clusters and spatial outliers, similar to the population change studies, a cluster and outlier analysis was performed for the other models - birth rate, death rate and migration. In the birth rate model (Fig. 5b), three municipalities with positive values for the coefficient I (LMiIndex) were identified, which form

clusters with low LL values (Dobrich, Dobrich-Selska and Pernik). This analysis complements the Hot Spot Analysis with the municipality of Pernik and, together with the neighbouring Stolichna, they form a cluster. The negative values for LMiIndex identify LH type spatial outliers - Varna, Montana, Pazardzhik and Plovdiv. For them we have a significant reduction in the birth rate compared to their neighbouring ones. For these seven municipalities, the p-value varies from 0 to 0.032741, which suggests that the identified clusters and spatial outliers are statistically significant.

Figure 5c presents the model of change in mortality for the studied years. Five municipalities were identified as clusters, with a decrease in mortality in the case of Vidin and Byala Slatina, and an increase in deaths in Rhodopes and Dobrich. Three municipalities (Aksakovo, Sadovo and Pernik) have negative values for LMiIndex, which are spatial outliers of the LH type. The determined clusters and spatial outliers are statistically significant with low p-value ($0.000869 \div 0.039128$).

Figure 5d presents a cluster model of the mechanical population growth for the years under consideration. Seven municipalities with positive values for the coefficient I (LMiIndex) were identified, which form clusters with high values of type HH (the town of Kardzhali and the municipalities of Ardino, Dzhebel, Kirkovo, Krumovgrad, Momchilgrad and Miniralni Bani). Three municipalities have negative values of the coefficient I (LMiIndex). As already mentioned, Stolichna Municipality still retains a positive value for the mechanical growth, but the population is twice less than in 2011. The trend is the same for Maritsa Municipality, the second largest city in the country. These two are defined as spatial outliers of the LH type. They show a decrease in the eviction process, but are surrounded by other locations where there is a reverse trend, which could be explained by more frequent settlement in small towns and villages around the capital (suburbanization and rurbanization). The case with the municipality of Pernik identified by the HL type is similar - less emigrated residents in 2019 are reported.

3.4. A summary of the results

3.4.1. Models of population change

The analyses partially confirm what can be seen on the choropleth map reflecting the change in the number of population. A comparison of the two maps shows that not all municipalities with a declining population in Figure 2a are cold spots (Fig. 4a), and not all municipalities with an increasing population are hot spots. This confirms the theoretical claims that in order for one hotspot to be statistically significant, its corresponding high-value characteristic (attribute) must be bypassed by other high-value one (ESRI, 2014).

Cluster and Outlier Analysis (Fig. 5a) outlined two significant deviations (anomalies) in the above statements concerning the neighbouring municipalities of Stolichna and Pernik. This analysis complements another cluster as a hotspot. In the case of Hot Spot Analysis, the municipality of Kardzhali was allocated to the group *Not significant*, surrounded by others with a slight increase in population. Cluster and Outlier Analysis confirms the clustering for the municipalities of Pleven, Lovech, Dobrich and Dobrich-Selska (cold spots) from the Hot Spot Analysis. The remaining 11 identified by the Hot Spot Analysis tool as cold spots do not form clusters. The two neighbouring municipalities of Dimitrovgrad and Haskovo were rejected as clusters with cold spots due to the high p-values (0.38 and 0.24) compared to the critical values (Table 1).

3.4.2. Models of birth rate

From the birth rate models presented in Fig. 2b, Fig. 4b and Fig. 5b the following concluding remarks can be drawn. The choropleth (Fig 2b) shows a decrease in the birth rate in most municipalities (212), but also an increase in 50 of them. For the Hot Spot Analysis (Fig 4b) this increase is insignificant and only 15 cold spots are identified. They refer to the municipalities with the largest cities. Cluster and Outlier Analysis reveals statistically significant clusters and anomalies. In this case they are seven.

The comparison of the two figures 4b and 5b shows that the statistics of Anselin Local Moran complements the Hot Spot Analysis, with the municipality of Pernik as a cold spot. At the same time, the municipalities of Vidin, Pleven, Ruse, Gabrovo, Shumen, Stara Zagora, Haskovo and Burgas are defined neither as clusters nor as outliers, although they are cold spots in Hot Spot Analysis. For them the Anselin Local Moran algorithm determines high p-value (from 0.203188 to 0.944238), higher than the critical ones, according to Table 1. Thus, the two models together (Fig 4b and Fig 5b) show weak clustering. The birth rate in all municipalities in Bulgaria suggests that they cannot be defined as demographically sustainable, even those with the largest cities. At the same time, these cities are dominated by people of reproductive age who strive to create material stability, rather than raise children.

3.4.3. Mortality models

The mortality models are presented in Fig. 2c, Fig. 4c and Fig. 5c. The cartogram (Fig. 2c) shows a reduction in mortality in 155 municipalities, but the Hot Spot Analysis (Fig 4c) identifies only two of them as cold spots with statistical significance, which were identified as clusters (Fig. 5c) using Cluster and Outlier Analysis.

Eight out of 106 municipalities (Fig. 2c) with an increase in mortality were identified by Hot Spot

Analysis as hot spots (Fig. 4c), and only three of them were identified as clusters by the Anselin Local Moran's I algorithm (Fig. 5c). In this case, Cluster and Outlier Analysis adds another one as a hot spot (Rhodopes). These models also show weak clustering. Although the mortality is declining in more than half of the municipalities in Bulgaria, it is statistically insignificant and cannot influence their demographic sustainability. There are several exceptions located near the largest cities (Pernik Municipality, the municipalities around Plovdiv and around Varna). In recent years, large investments were made in their economy, infrastructure and services, hence, they became attractive to younger people, especially young families, and this changed their demographic profile.

3.4.4. Models of the mechanical growth

The models of mechanical growth in the country have a definite cluster character, as can be seen from Table 4. The domination of the eviction process is typical for most municipalities in the country, but with the Hot Spot Analysis it turns out that only four of them register negative values of the mechanical growth. The reverse process of dominant settlement is present in nine municipalities, six of which are located in the district of Kardzhali (Fig. 5d). The Cluster and Outlier Analysis algorithm defines three municipalities as spatial outliers (the municipalities of Sofia - Stolichna, Pernik and Maritsa) (Fig. 5d).

3.5. Spatial models of the population taking into account the impact of all demographic indicators simultaneously

The K-Means function was used to analyse which groups of municipalities contained similar data simultaneously for the four variables: change in the number of population (Pop_19-11), birth rate (Born_19-11), mortality rate (Died_19-11), and mechanical growth

Table 4. Variable statistics.

Variables	Overall Mean	R ²	Mean					
			1	2	3	4	5	6
Pop_19-11	-1472.51	0.85	-1632.34	1156.00	-867.39	32175.00	5587.50	-7133.28
Born_19-11	-36.08	0.86	-24.29	-6.50	-14.13	-667.00	-490.00	-151.08
Died_19-11	-0.23	0.79	-42.25	22.33	6.49	461.00	500.00	30.00
Growth_19-11	7.80	0.74	-55.39	1057.00	11.08	-3022.00	1131.00	-30.12

The following conclusions originate from the summarized statistics of the whole dataset and their visualization in the Boxplot graphs (Fig. 7):

1). The closest values to the average differences of the whole dataset (for 264 municipalities) are reported for the following indicators and groups: a change in the number of population (Pop_19-11) - first (blue) and third (green) group; birth rate (Born_19-11) -

(Growth_19-11). The cluster analysis was performed with ArcGIS software. Six groups were created after the evaluation of the optimal number of groups. The result of the cluster analysis is a new vector layer in the map, which contains all municipalities, divided into the following groups (Fig. 6): First (blue) group - 80; Second (red) group - 6; Third (green) group - 150; Fourth (orange) group - 1; Fifth (purple) group - 2; Sixth (brown) group - 25.

In the grouping analysis of municipalities, the summarized statistics Mean, Std. Dev., Min, and Max and R² were calculated, both for the whole dataset and for each group individually. Table 4 presents some of them. Theoretically, the higher the value of R² for a particular variable, the better the grouping efficiency (ESRI, 2014). The values of R² for the four variables are closest to each other when grouped into six groups. The most well-known measure for performing the similarity, the Euclidean distance, was chosen to perform the classification and no spatial constraint was considered for the Spatial Constraints parameter.

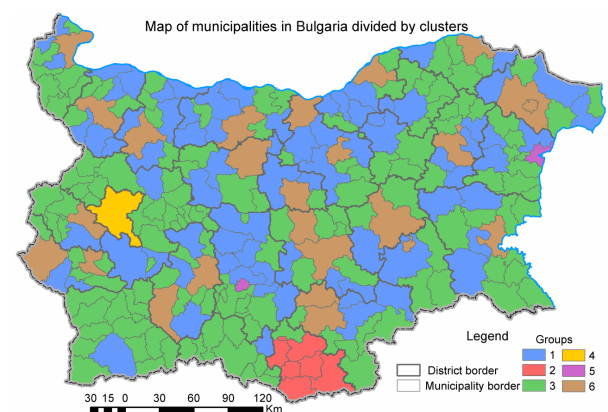


Fig. 6. Grouping analysis of all demographic indicators, simultaneously, for 2011 and 2019.

All variables reflect differences in the values of demographic indicators for the two years, 2019 and 2011.

first (blue), second (red) and third (green) group; mechanical growth (Growth_19-11) - first (blue), third (green) group and sixth (brown) group.

This means that the data for these indicators are similar for the indicated groups. This confirms that the change in population number is more strongly influenced by the changes in the mechanical growth and in the birth rate.

2). Higher than average differences of the whole dataset are reported for the following indicators and groups: mortality (Died_19-11) - second (red), third (green) and sixth (brown) group. In these groups there is an increase in mortality above the national average difference (-0.23) for 2011 and 2019. The municipalities in these groups have higher mortality rate than average because they are dominated by a population of older age groups. These are rural, mountainous and/or border areas.

3). Lower than average differences of the whole dataset are reported for the following indicators and groups: mortality (Died_19-11) - first (blue) group. Only in this group a decrease in mortality above the national average difference (-0.23) for 2011 and 2019 is observed. This group includes small municipalities, located usually in the interior part of the country, forming the so-called "inner periphery". They are the most affected by depopulation.

4). Significant deviations from the mean differences of the whole dataset are reported for the following indicators and groups: change in the number of population (Pop_19-11) - second (red), fourth (orange), and fifth (purple) group. In these groups there is an increase in population number above the national average difference (-1472.51) for 2011 and 2019. The most significant is the increase recorded in the municipalities of Stolichna, Plovdiv and Varna; mortality (Died_19-11) - fourth (orange) and fifth (purple) group. In the municipalities of Stolichna, Plovdiv and Varna there is a significant increase in mortality above the national average difference (-0.23) for 2011 and 2019; birth rate (Born_19-11) - fourth (orange) and fifth (purple) group. In the municipalities of Stolichna, Plovdiv and Varna there is a significant decrease in the birth rate above the national average difference (-36.08) for 2011 and 2019; mechanical growth (Growth_19-11) – the largest emigration above the national average difference (7.80) for 2011 and 2019 is reported in the fourth (orange) group, respectively Sofia Municipality. The largest settlement is observed in the sixth (brown) group or most municipalities in Kardzhali district, followed by the fifth (purple) group, corresponding to the municipalities of Plovdiv and Varna. The most possible reasons for this outcome are the investments made in the regional economy of the second and third largest cities in the country in the last few years (2014-2019), related to the creation of economic zones and centres for innovation and development. The newly created jobs led to attracting migrants and higher levels of positive mechanical growth.

The grouping algorithm of municipalities takes place simultaneously for the four processes - population, birth rate, mortality and mechanical growth. This analysis confirms the conclusions we brought forward, with Hot Spot Analysis and Cluster

and Outlier Analysis, when we analysed the variables separately.

Cluster and demographic analyses have also been performed by other scientists conducting various population researches (Antipova and Fakeyeva, 2012; Delgado Viñas, 2013; Makra et al., 2018; Ocovsky, 1991; Judah, 2021). The results of our study provide evidence that the demographic processes in Bulgaria are similar to those in the countries of Southeast Europe and Russia (Shubat et al., 2016; Tsvetkov, 2020; Judah, 2021).

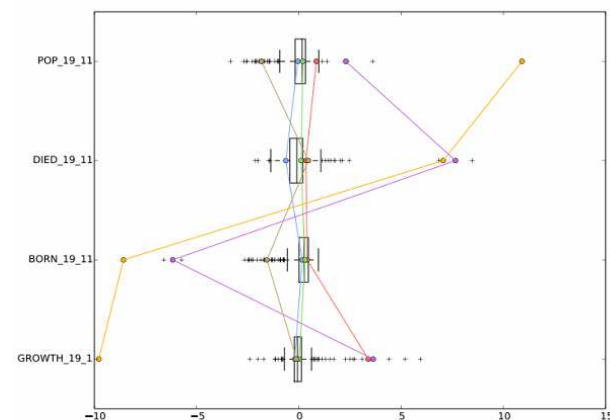


Fig. 7. Parallel Box Plot.

4. CONCLUSION

We conducted a study on demographic processes in Bulgaria and examined them at the lowest administrative territorial level, which facilitated the analysis of the causes of the demographic trends in the country. Our purpose was to establish the existence of demographically sustainable municipalities. To achieve it we used cluster analysis. The performed analysis showed a cluster model of data presenting the mechanical growth for 2011 and 2019. The models for the change in the number of population and the change in the birth and mortality rates revealed random spatial models with a weak tendency for clustering. The analyses of the demographic processes proved that migration has the strongest influence on population dynamics. There is no such dependence on birth rate and mortality. Our hypothesis was proved. The demographic sustainability of municipalities is most strongly influenced by their economy, which is the main reason for migration. This analysis confirms that in the municipalities in which the largest cities in the country are located, as well as some of the ones bordering them, will continue the process of concentration of population. This trend is not typical for the municipalities in which small towns are located and for the rural ones. The economic factors will be leading for this migration process.

Negative demographic processes prevail in the smaller cities and villages. Important differences between municipalities regarding the demographic

processes were revealed. A slight increase in birth rate is typical for certain municipalities and it is due to economic and religious factors. Higher mortality rate is typical for most mountainous and/or border area municipalities. A group with a very advanced depopulation process and low mortality rate was identified, forming the so-called inner periphery.

The cluster analysis confirmed the impossibility to stabilize the demographic processes in the country through measures for birth rate stimulation. In our opinion, the policy aimed at economic development, job creation, and investments in infrastructure and services will be the most effective. The measures must reflect the unique combination of economic, political, socio-cultural, psychological and religious characteristics of the local population (Naidenov, 2017; Petrov, 2020). These dynamically changing data and trends require the implementation of geographic information systems linked to the application of the roadmap and the INSPIRE Directive (Andreeva and Andreev, 2020; Kastreva et al., 2018). A step towards modernizing activities in the society is the introduction of e-government and digitalization of some services (Karadzhova, 2019; Patarchanov, 2012).

Finally, we should mention that our data is limited to four indicators for the characterization of demographic processes, because they are the most representative. From a research point of view, other indicators could always be added, which are likely to give additional emphasis to the created spatial models. Further investigations could be made in this direction.

REFERENCES

- Andreeva P., Andreev A.** (2020), On the problems related to the implementation of geographic information systems in Bulgaria. [online], URL: https://www.aadcf.nvu.bg/scientific_events/papers/df2020.pdf. Accessed on 28.11.2021
- Antipova E., Fakeyeva L.** (2012), Settlement system of Belarus: spatial and temporal trends at the end of 20th and the beginning of the 21st centuries. *Journal of Settlements and Spatial Planning*, 3 (2), 129–139. URL: http://geografie.ubbcluj.ro/ccau/jssp/arhiva_2_2012/09JSSPo22012.pdf
- Delgado Viñas C.** (2013), Population Dynamics of Spanish Mountain Areas: Case Study of Two Regions in the Cantabrian Mountains (Spain). *Journal of Settlements and Spatial Planning*, Special Issue (2), 207-217. URL: https://geografie.ubbcluj.ro/ccau/jssp/arhiva_si2_2013/04JSSPSIo22013.pdf
- ESRI** (2014), An overview of the Mapping Clusters toolset. [online], URL: <https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/an-overview-of-the-mapping-clusters-toolset.htm>. Accessed on 05.06.2021

- ESRI** (2014), How Grouping Analysis works. [online], URL: <https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/grouping-analysis.htm>. Accessed on 05.06.2021
- ESRI** (2014), Spatial Autocorrelation (Global Moran's I) [online], URL: <https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/spatial-autocorrelation.htm>. Accessed on 05.06.2021
- European Commission** (2020), European Commission Report on the Impact of Demographic Change [online], URL: https://ec.europa.eu/info/sites/default/files/demographic_report_2020_n.pdf. Accessed on 27.11.2021
- Grasland C.** (1990), Potentiel de population, interaction spatiale et frontières: des deux Allemagnes à l'unification. [Population potential, spatial interaction, and borders: from two German states to unification]. *L'Espace Géographique*, 19-20 (3), 243-254. DOI: <https://doi.org/10.3406/spgeo.1990.2992>. (Article in French)
- Hooz I.** (1992), Depopulation of border settlements. *Hungarian statistical review*, 70 (12), 1005-1017. (Article in Hungarian)
- IISS (International Institute for Strategic Studies)** (2020), Europe's uneven demographic decline. *Strategic Comments*, 26 (8), Pages ix-xi, DOI: <https://doi.org/10.1080/13567888.2020.1856490>
- Judah T.** (2021), Southeast Europe's Looming Demographic Crisis. [online], URL: <https://www.helvetas.org/en/eastern-europe/about-us/follow-us/helvetas-mosaic/article/March2021/demographic-decline-southeast-europe>. Accessed on 27.11.2021
- Kastreva K., Chamova T., Tarnev I.** (2018), Български патиентски регистър за spinalna muskulna atrofiya kato chast ot mezhdunarodniya SMA registър - analiz na klinichni dannii. [Bulgarian patient registry for spinal muscular atrophy as part of the international SMA registry - analysis of clinical data]. *Българска неврологиya*, 19 (1), 15-19, ISSN 1311-8641. (Article in Bulgarian)
- Kastreva P., Kastreva K.** (2019), Klasifikatsiya na statisticheskite dannii v GIS sreda. [Classification of statistical data in GIS]. *Geodeziya, kartografiya i zemeustroistvo*, 5 (6), 20-24, ISSN 0324 – 1610, Sofia, Bulgaria. (Article in Bulgarian)
- Kastreva K., Kastreva P.** (2020), Classification of multiple numerical attributes in ArcGIS environment. *Micro Macro & Mezzo Geo Information*, 14, 7-18 UDC: 519.85:004.6]:[528.9:616.83(497.2)
- Karadzhova G.** (2019), Building an e-government in Bulgaria in the context of European development. In: *Multidisciplinary Journal of Science, Education and Art*. 701-711, [online], URL: <http://www.usb->

blagoevgrad.swu.bg/media/2082/godishnik.pdf.
Accessed on 30.11.2021

Makra Z., Bajmocy P., Balogh A. (2018), Villages on the Edge of Extinction - the Hungarian Situation. *Journal of Settlements and Spatial Planning*. DOI: <https://doi.org/10.24193/JSSP.2018.1.04>

Kumar V. (2019), Cluster Analysis: Basic Concepts and Algorithms. In: Tan P., Steinbach M., Karpatne A., Kumar V. (eds) *Introduction to Data Mining*. 2nd edition, Pearson, 487-568, [online], URL: <http://www-users.cs.umn.edu/~kumar/dmbook/ch8.pdf>. Accessed on 29.11.2021

MLSP (Ministry of Labour and Social Policy) (2006), *Natsionalna strategija za demografsko razvitie v republika Bulgaria (2006 - 2020)*. [National strategy for demographic development of the population in the republic Bulgaria (2012-2030)]. [online], URL: <https://www.strategy.bg/StrategicDocuments/View.aspx?lang=bg-BG&Id=444>. Accessed on 20.12.2021

NSI (National Statistical Institute) (2019), *Population and demographic processes in 2019*. [online], URL: <https://www.nsi.bg/en/content/18126/прессъобщение/population-and-demographic-processes-2019>. Accessed on 29.11.2021

Naydenov K. (2017), Influence of the demographic factor on the regional social and economic development. *SGEM* 4 (1), 363-370 DOI: <https://doi.org/10.5593/sgemsocial2017/41/S18.046>

Ocovsky S. (1991) *Aktualna problematika geografie obyvatel'stva a sidel* [Current problems in the geography of population and settlements]. *Geografický časopis*, 43(3): 250-7. (Article in Slovak)

Patarchanov P. (2010), Agricultural mountain areas in Bulgaria – problems, challenges, decisions. *Journal of Settlements and Spatial Planning*. 1 (2), 145-150. [online], URL: https://geografie.ubbcluj.ro/ccau/jssp/arhiva2_2010/07JSSPo22010.pdf. Accessed on 29.11.2021

Patarchanov P. (2012), Role and Place of Alternative Tourism Development in Mountain Areas. *Journal of Settlements and Spatial Planning*. 1, 149-155. [online], URL:

https://geografie.ubbcluj.ro/ccau/jssp/arhiva_si1_2012/17JSSPSIo12012.pdf. Accessed on 24.02.2021

Petrov K. (2020), *Regionalna demografia*. [Regional Demography]. UNSS, Sofia, Bulgaria. ISBN: 9786192322670. (Book in Bulgarian)

Saito Y. (2007), Depopulation and Rural Spatial Management: A Case Study in Shimoina County, *Journal of Asian Architecture and Building Engineering*, 6 (2), 259-266. DOI: <https://doi.org/10.3130/jaabe.6.259>

Shubat O., Bagirova A., Makhabat A., Ivlev A. (2016), The Use of Cluster Analysis for Demographic Policy Development: Evidence from Russia. *ECMS 2016 Proceedings* edited by: Thorsen Claus, Frank Herrmann, Michael Manitz, Oliver Rose, European Council for Modeling and Simulation. DOI: <http://dx.doi.org/10.7148/2016-0159>. ISBN: 978-0-9932440-2-5

Tsvetkov A. (2020), Clustering of districts according to their socio-economic status. [online] URL: <https://www.regionalprofiles.bg/bg/top2015/>. Accessed on 29.11.2021

Willekens F. (2015), *Demographic Transitions in Europe and the World*. In: Matthijs, K., Neels, K., Timmerman, C., Haers J. (eds) *Population Change in Europe, the Middle-East and North Africa*. 1st Edition, London, Routledge, 13-44. DOI: <https://doi.org/10.4324/9781315601496> [online]

Yang R., Xu Q., Long H. (2016), Spatial distribution characteristics and optimized reconstruction analysis of China's rural settlements during the process of rapid urbanization, *Journal of Rural Studies*, 47 (B), 413-424. DOI: <https://doi.org/10.1016/j.jrurstud.2016.05.013>

Zhang C., Luo L., Xu W., Ledwith V. (2008), Use of local Moran's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland. *Science of The Total Environment*, 398 (1-3), 212-221. DOI: <https://doi.org/10.1016/j.scitotenv.2008.03.011>