Spatial Continuity of Selected Indicators of Labour Market

Jiří Horák [1], Milan Šimek[2], Monika Šeděnková[1]

[1] Institut geoinformatiky, Hornicko-geologická fakulta, VŠB-TU Ostrava,
17.listopadu 15/2172, 708 33 Ostrava-Poruba, Česká Republika
jiri.horak@vsb.cz, monika.sedenkova@vsb.cz
[2]Katedra národohospodářská, Ekonomická fakulta, VŠB-TU Ostrava, Sokolská třída 33,
702 00, Ostrava, Česká Republika
milan.simek@vsb.cz

Abstract. Social and Economical Geography as well as Sociology apply for measurement of spatial continuity (autocorrelation) usually methods of global and local indicators than geostatistical approaches to a field exploration. Global indicators like Moran I or Geary C indexes are considered to be analogous to basic structural functions of geostatistics, however they utilize a generalised concept of distance and contiguity. Enumerations of local indicators spatial association (LISA) and testing of their significance offer a great contribution to spatial exploratory analysis. Selected methods were applied to 3 time snapshots of the Czech labour market described by the rate of unemployment, the rate of longterm unemployed and rate of unemployed with basic education for the level of municipalities. The evaluation based on global and local indicators are compared with results of usual methods.

Keywords: spatial continuity, geostatistics, global and local indicators of spatial autocorrelation

Abstrakt. Prostorová kontinuita vybraných ukazatelů situace na trhu práce. V oblasti socioekonomické geografie a sociologie se k popisu a zkoumání prostorové kontinuity (autokorelace) používají více metody globálních a lokálních indikátorů než geostatistické postupy zkoumání pole. Globální indikátory jako je Moranovo I nebo Gearyho C kritérium jsou analogické základním strukturálním funkcím geostatistiky, avšak používají zobecněný koncept vzdálenosti a sousedství. Potenciálně velmi přínosné jsou výpočty lokálních indikátorů a testování jejich významnosti, známé pod označením LISA. Vybrané metody byly aplikovány na 3 časových řezech míry nezaměstnanosti, zastoupení dlouhodobě nezaměstnaných a nezaměstnaných se základním vzděláním v obcích ČR. Hodnocení pomocí globálních a lokálních hodnot indikátorů je porovnáno s běžnými metodami.
Klíčová slova: prostorová kontinuita, geostatistika, globální a lokální indikátory prostorové autokorelace

1 Introduction

Characteristics of close objects are more similar than distant objects. It can be seen as of one the basic geographical axioms, expressed by the Tobler's law [12].

The influence of proximity to the value distribution of selected phenomena can be strong or weak and we can recognize slow changes in values or rapid/random changes respectively. Similarity of values is usually denoted as a correlation which can be measured by i.e. Pearson coefficient of correlation, Spearman coefficient of correlation, goodness-of-fit R2 etc. Usually it is applied to evaluate the character and power of the relationship between 2 different phenomena (i.e. temperature and humidity).

If we study a correlation of only 1 phenomenon (i.e. temperature) we speak about a autocorrelation instead of the correlation. The most important form of autocorrelation is *spatial autocorrelation*, where the similarity of values (of 1 phenomenon) separated by some distance is measured. Typically the analysis tries to evaluate how the similarity of values decreases with the increased distance (distance decay effect).

Spatial autocorrelation can be defined as the coincidence between value similarity and locational similarity [2]. Any presence of strong spatial autocorrelation highly influences the application of many statistical, analytical and modelling methods where classic a spatial forms of these methods fail due to the fact that they requires to have samples independent. In such cases we need to apply methods

respecting spatial autocorrelation. I.e. classic regression models are substituted by spatial autoregressive models in various variants like simultaneous autoregressive models, conditional autoregressive models, etc. [3].
This is not the only reason why is the spatial autocorrelation important. The description of spatial autocorrelation, especially unhomogeneity and anisotropy measured by spatial autocorrelation can be deployed to divide the area to homogeneous units or to uncover and understand processes forming the explored situation. The first one represents an effective way how to establish territorial units homogeneous by its internal structure and values.

Finally, the spatial autocorrelation measured in a local scale can be used for determining of hot spots in the area – places where the values here and in surroundings significantly differs from other places. Such hot spots may help to discover places with specific conditions. Measurement of spatial similarity or dissimilarity between values of 1 phenomenon (separated by distance) can be seen as an extension of spatial autocorrelation. We can speak about a *spatial association*. The concept is the same, just for description of spatial similarity not only measures of spatial autocorrelation are deployed – also measures of spatial variability can be applied.

The intention of the paper is to attract attention to the phenomenon of spatial association especially for areal data and to demonstrate an application and a value of contribution for analytical studies. As an example we select the analysis of selected indicators of labour market in the Czech Republic.

2 Global indicators of spatial association

The concept of spatial association and autocorrelation is deeply elaborated and usually applied in geostatistics. Geostatistics is foremost Data Analysis and Spatial Continuity Modelling [7]. Such kind of data analysis is denoted as *structural analysis***.** The aim of structural analysis is to depict the spatial structures in the area both to understanding spatial relationships as well as utilised the final parameters describing the structure in geostatistical methods of interpolation (kriging).

The structural analysis deploys various indicators to measure spatial autocorrelation or variability in the area. Typically variogram (measures of variability depending on distance and direction) and covariance functions (measures of covarince depending on distance and direction) are deployed. Both these functions use only 1 type of distance measure – a metric distance.

Sometimes we need to describe a spatial separation more commonly i.e. separation expressed by the order of neighbours or by the inverse share of common border. This is the case of area data where we do not have any precise measures of distance separated values. Even more, area data is usual form of socioeconomic data where more common measures of distance seem to be more contributing to understand complex phenomena.

The most popular indicators are Moran's I a Geary's Ratio C indexes which can be defined as [3]:

$$I_k = \frac{n * \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^{(k)} * \left(z_i - \bar{z}\right) * \left(z_j - \bar{z}\right)}{\left(\sum_{i=1}^{n} \left(z_i - \bar{z}\right)^2\right) * \left(\sum_{i \neq j} \sum w_{ij}^{(k)}\right)} \quad (1)$$

(2)

$$C_k = \frac{(n-1) * \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^{(k)} * (z_i - z_j)^2}{2 * \left(\sum_{i=1}^{n} (z_i - \bar{z})^2 \right) * \left(\sum_{i \neq j} \sum w_{ij}^{(k)} \right)}$$

where z_i is the value in area unit i (z with the top stripe is mean of z), $w_{ij}^{(k)}$ is the weight between units i and j for the step k.

As you can see, indicators can be calculated for different steps but quite often only the first step is utilised. There are many variants of w_{ij} calculation differs the way how to select neighbours and how to assign weights to them. Some of the variants are in the table 1.

Table 1. Variants of weight matrix calculation

1)	$w_{ij}= 1$	the centroid of area unit *j* is one of the *k* nearest centroid to the unit *i*
	$w_{ij}= 0$	other cases
2)	$w_{ij}= 1$	the centroid of area unit *j* is to the distance δ to the unit *i*
	$w_{ij}= 0$	other cases
3)	$w_{ij}= d^{\gamma}_{ij}$	the distance d_{ij} between the centroid of unit *i* and *j* is lower than distance limit δ from the unit *i* ($\gamma < 0$ denotes steepness of the distance impact)
	$w_{ij}= 0$	other cases
4)	$w_{ij}= 1$	the unit *j* shares the common border with the unit *i*
	$w_{ij}= 0$	other cases
5)	$w_{ij}= l_{ij}/l_i$	l_{ij} is the length of the common border between units *i* and *j; l_i* is the perimeter of the unit i

Some ways of calculation w_{ij} contain the parameters (*k, δ, γ*), which can be optimized are illustrate on fig. 1-2 .

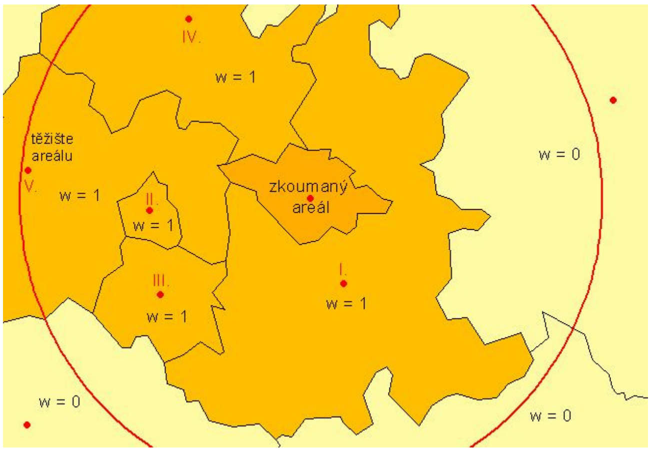

Fig. 1 Neighbors selected as 5 closest polygons, weights are 1 [4]

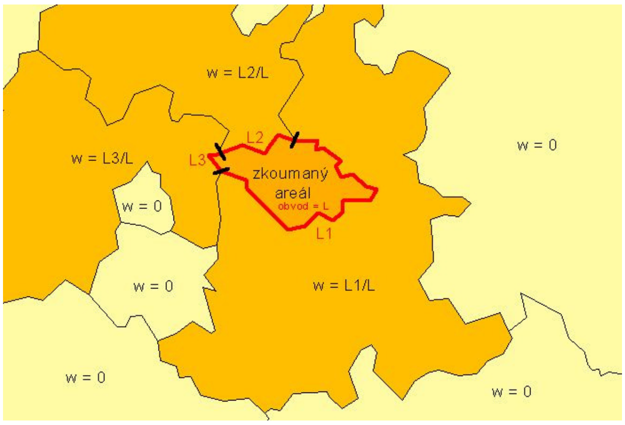

Fig. 2 Neighbors are in touch, weights are assigned according the length of the common border [4]

Moran's I index represents a direct measure of spatial autocorrelation (likewise to covariance function). Geary's Ratio C index is a measure of variability (it is calculated similarly to variogram).
In no spatial autocorrelation exists (it means a random pattern), the expected value of Moran's I is (-1)/(n-1) [9] and the expected value for C is 1 (not depending on the size of file). The clustered pattern carries the C value between 0 and 1, the value of I is higher than for the random pattern. The dispersed pattern is indicated by the C value between 1 and 2, the value of I is lower than for the random pattern.

It is necessary to stressed out that any of the criterions do not provide values exactly in the mentioned interval. It causes real problems with distance matrix formulation. [3] propose a corrective element I' by which I is divided and by that the supposed value intervals can be obtained (similarly for C).

Some authors [11] warn against using the statistical characteristics of spatial autocorrelation indicators for adjusted data that are based on the age population standardization. The population can be usually spatially autocorrelated and that leads to project of this dependence to adjusted data of studied phenomenon. Frequently that phenomenon can indicate spatial autocorrelation even though in reality can be invariable in the whole area. As an appropriate solution is recommended testing of probability based on a simulation.

Instead of any correction factor currently the technology of randomization is applied. The values of attributes (the known values) are assigned to different spatial units randomly. Thus the random process of attribute value simulation is substituted by the random process of selection the place. For the random generated situation assessments of variance are computed and together with estimated mean value we can use for testing of importance of recognised deviation from the random situation. [9].

Moran's I and Geary's Ratio C do not distinguish if the positive autocorrelation takes place between high or low values vice versa. That is why it is possible to deploy an indicator G(d), as an analogy of Ripley's K function [9].

3 Local indicators of spatial association

Moran's I and Geary's Ratio C measure spatial autocorrelation in the whole area. So it is used for global determination of autocorrelation or variability measurement. In some case we need differentiate local autocorrelation situation and find out whether the area is homogeneous (still the same autocorrelation value in the area) or vice versa heterogeneous (the autocorrelation value in individual places significantly differs).

Local Indicators of Spatial Association (LISA) represents the local version of Moran's I and Geary's Ratio C or indicator G ([1] in [8]).

The local Moran's I for unit i is defined as [9]:

$$I_i = r_i * \sum_j w_{ij} * r_j \qquad (3)$$

where r_i and r_j are standardized values of z.

Similarly like a global Moran´s I, high value for local Moran´s I show a clustering of similar values (high or low). A standalone computation of the indicator is not too useful. High values can arise randomly. It is necessary to determine estimated mean value, variance and compute probability of such result (etc. to test if the deviation can be caused randomly).

The expected mean value [1] is $E[I_i] = (-\ w_i)/(n\text{-}1)$. It is necessary to calculate expected mean and variance for each area to evaluate the probability of the given result.
The local version of Geary's Ratio C is less applied due to difficulties [9] by its distribution properties.

$$C_i = c_i * \sum_j w_{ij} * (r_i - r_j)^2 \qquad (4)$$

Exploration analyses try to discover places with an unusual value of spatial autocorrelation. Moran´s diagram draws value *Wz* to *z* and also the result of linear regression that can identify the abnormal values (outliers).

LISA helps to distinguish high autocorrelation between high and low values. It can identify the area with high values (monitoring area is surrounded by high values) which is denoted as a HH type of the area. LL type means a cluster of high correlation where a low value is surrounded by low values. HL resp. LH cluster represents an area, where significantly high autocorrelation is negative and high value is surrounded by low values and vice versa.

The following computations of global and local indicators of spatial association were undertaken in the Geoda environment [13]. The model of rooks' case of spatial weighting was applied according results of studies of German unemployment for different administrative units [10] and practical test showing more compact and stable results for the pattern of Czech municipalities.

The Geoda programme has been used i.e. for spatial analysis of the Czech Party System, where LISA cluster map for district levels brought new view to the institutionalisation and spatial regimes of Czech party system [8].

4 Case study: a Labour market in the Czech Republic

Before starting any study we have to check the basic consistency prerequisites:

- no changes in area unit structure (number, shape etc.) – the number of municipalities in CZ to 31.12.2006 is 6248, for the later used snap shots is 6249. The data of labour market was provided for 6249 municipalities. Thus we have used a fixed structure with 6249 area units for all time snapshots.
- some of area units are not continuous which complicates a computation of weighs for LISA. It negatively influences especially absolute values, where the distribution of values to all individual polygons is necessary. We have used ratio values where the problem can be neglected.

The labour market is influences by many factors, among them selected indicators describing unemployment registered by labour offices are very significant. We have selected 3 main indicators:

the rate of unemployment, the rate of long-term unemployed and rate of unemployed with basic education. To study a local situation the data for the municipal level is deployed.

We have selected 3 time snapshots which help to evaluate not only the state to the selected date but also the development. This way it is possible to assess the stability of the situation and an extent of changes.

A traditional approach to analyse the situation of selected indicator is to create choropleth map followed by a visual interpretation. According [14] there are two choroplethic techniques – simple choroplethic mapping and classless choroplethic mapping. The classless choroplethic maps represent the attribute distribution without any generalization but it may create a quite complex pattern and confuse map readers. The result of simple choroplethic mapping depends on the number of classes and class limit determination. We can find many methods for determining class limits in the cartographic literature usually based on the study of the data distribution using graphic (frequency curve, cumulative frequency curve, clinographic curve etc.) or mathematic approaches [15,16]. The methods are separated into three major groups [14]:

1. constant series or equal steps
2. systematically unequal stepped class limits
3. irregular stepped class limits

We have applied three of these methods to demonstrate frequent outputs of choropleth mapping. The first group is presented by equal interval (equal steps based on the range of attribute values [14]) and standard deviation interval (parameters of a normal distribution [14]). The Box-Jenkins method (denoted in following figures as natural breaks) belongs to the third group.
The choropleth maps have been prepared in ArcGIS 9.2.

The application of global and local indicators of spatial association can offer another view to the data and mapped phenomenon. The spatial autocorrelation can help to better understand the phenomenon and its behaviour, to distinguish how much is the phenomenon determined by the local situation or influenced by neighbours or some common factors behind.

The map of regions (NUTS3) (fig.3) will help to better understand locational terms used for description of the results documented in following figures.

Fig. 3 Regions of the Czech Republic

5 Results of local analysis

5.1 Rate of unemployment

Three methods of class interval determination for choropleth map is presented in fig.3-5. We can compare results with a cluster LISA map (fig.6). Equal intervals did not contribute too much to pattern description and the determination of borders for individual regions with similar values is difficult. It is possible to find main territories with higher level of unemployment – large part of the Usti region, peripheral parts of Moravian-Silesian and Olomouc regions, SW part of the South Moravian region and other smaller areas. Better visual interpretation can be achieved using natural breaks and standard deviation classifications (the effect of the last one is enhanced with a bicolour scheme). These methods provide us more plastic view of the situation where the distribution of values are better portrayed. Nevertheless these methods are not efficient to delimit individual regions with high values and establish their borders.
To the opposite LISA cluster map provides clear distinguishing of individual regions where the spatial autocorrelation is high. These regions are quite compact. Some border parts of the CZ are not classified due to the insufficient neighbour's identification which decreases the quality of classification in these areas.

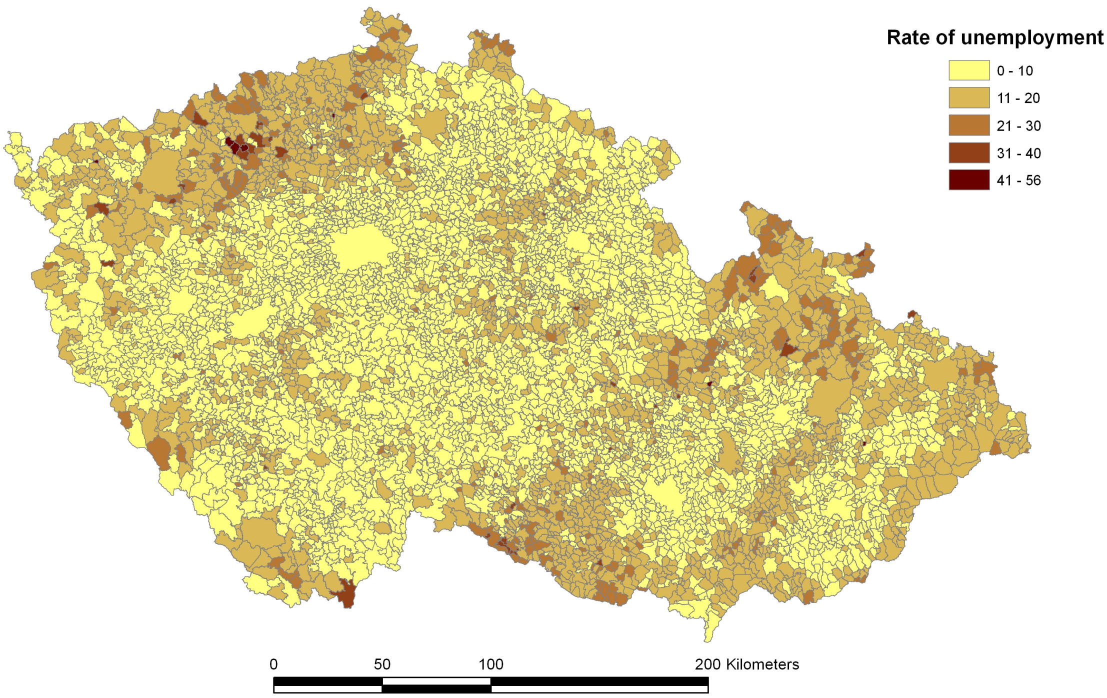

Fig. 3 Rate of unemployment (municipalities, 31.12.2006, equal interval classification)

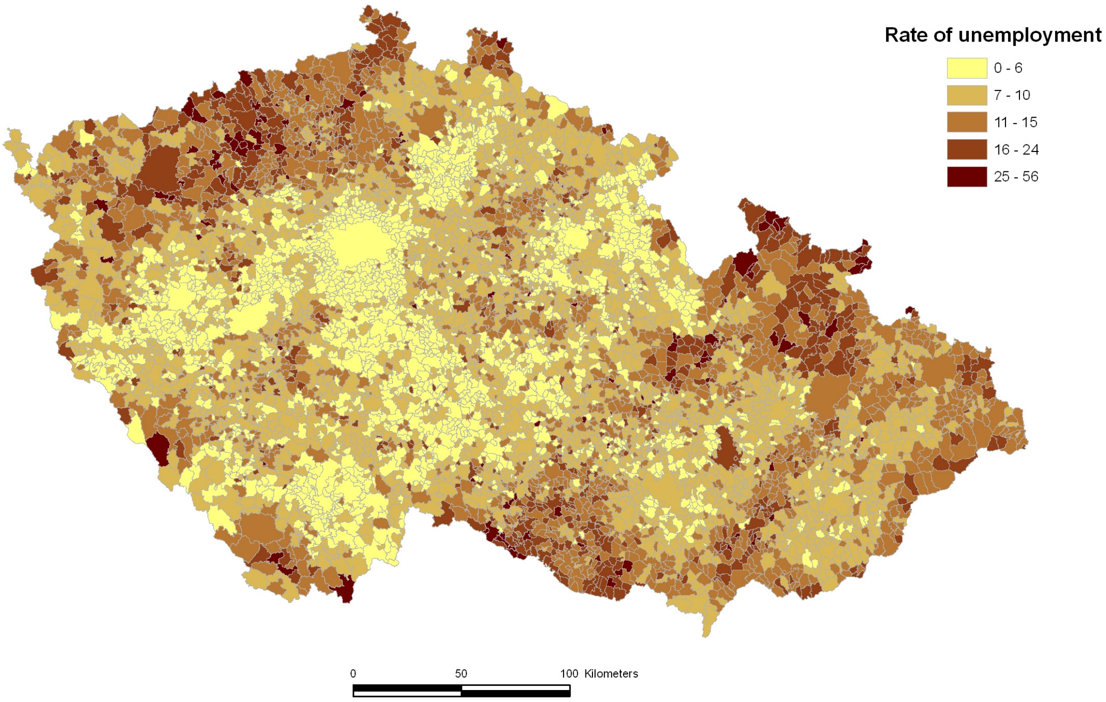

Fig. 4 Rate of unemployment (municipalities, 31.12.2006, natural breaks interval classification)

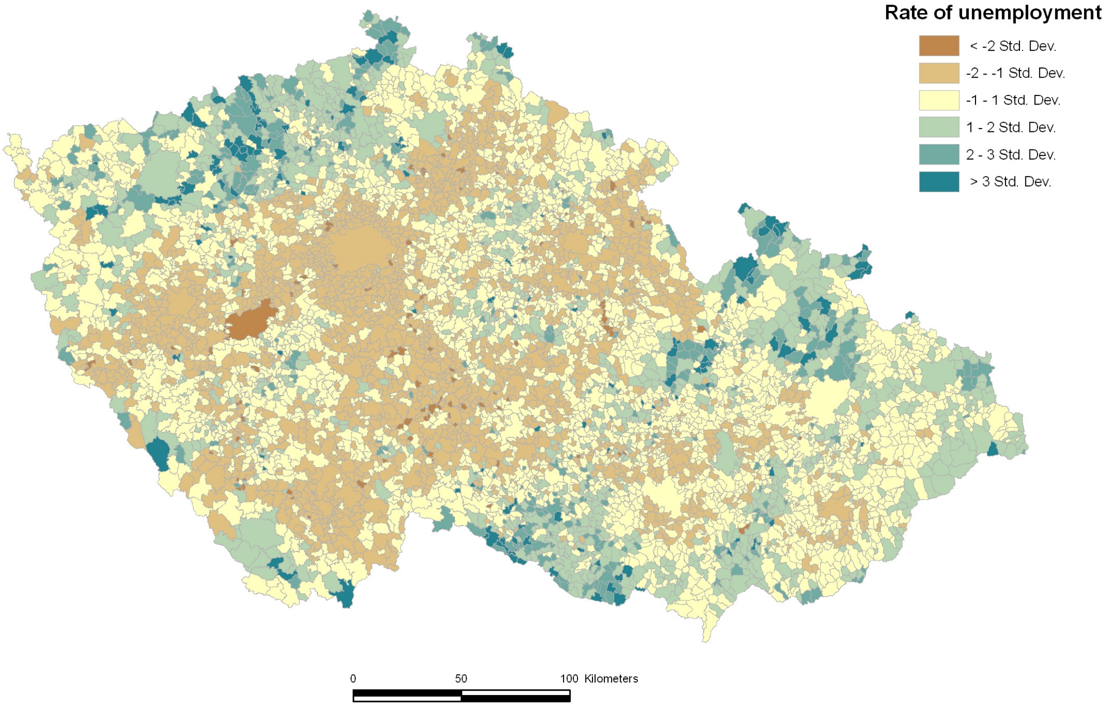

Fig. 5 Rate of unemployment (municipalities, 31.12.2006, standard deviation interval classification)

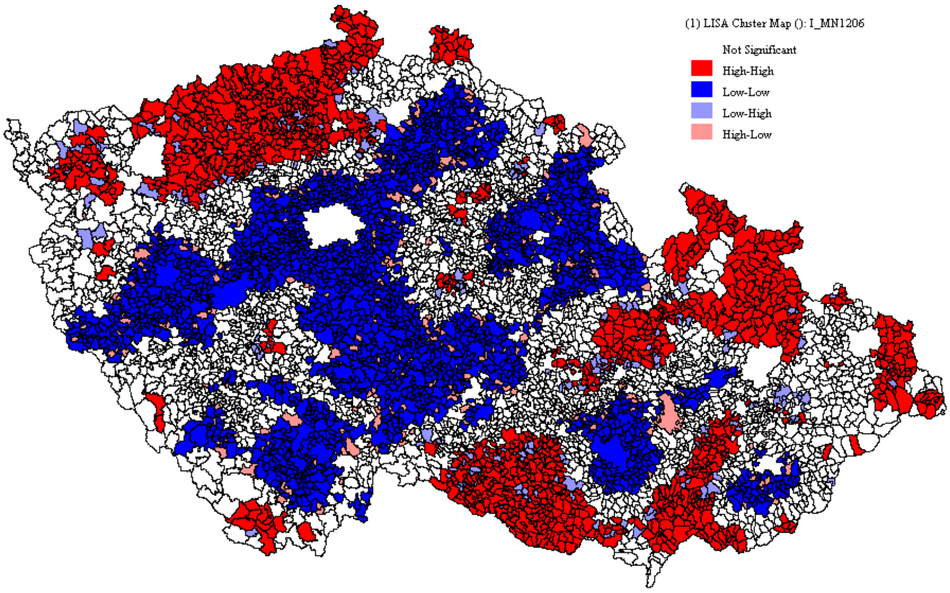

Fig. 6 LISA cluster map of the rate of unemployment (municipalities, 31.12.2006, significant clusters)

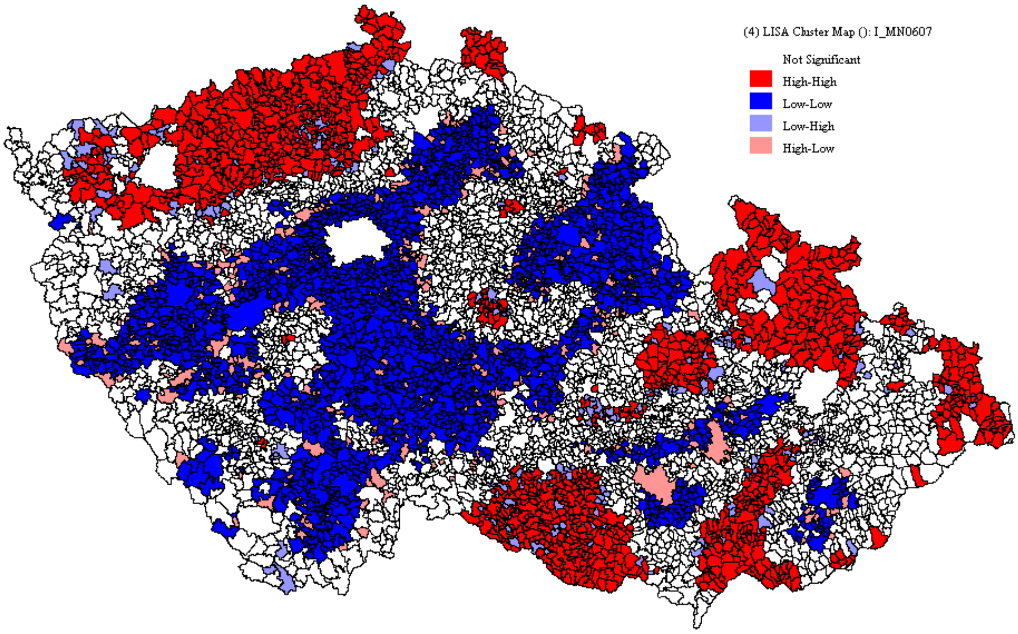

Fig. 7 LISA cluster map of the rate of unemployment (municipalities, 30.6.2007, significant clusters)

The comparison of snaphots of cluster maps shows quite compact and stable clusters where we can identify important high-level clusters for Usti region, outlying municipalities of Moravian-Silesian and Olomouc regions, Znojmo area and Hodonin-Kyjov area of South Moravian region. Important low-level clusters are Central Bohemia extended to Plzen (SW) and Mlada Boleslav (NE) and Vysocina areas, part of Eastern Bohemia region (namely Hradec Karlove region), and area around Ceske Budejovice. Low-level clusters Brno-Olomouc and around Otrokovice are not stable and present significant changes of the shape and extend.

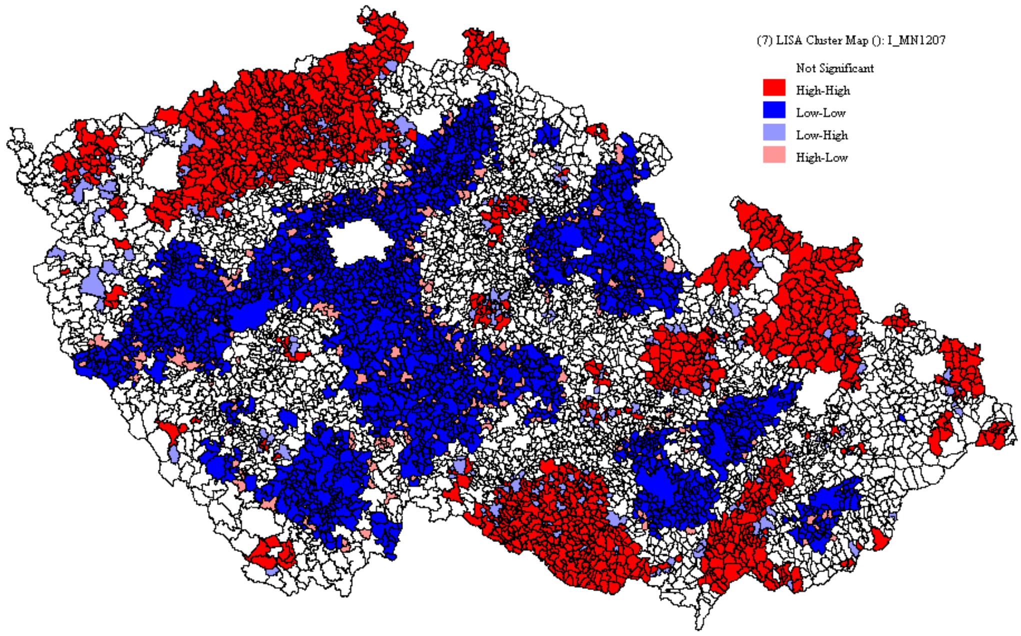

Fig. 8 LISA cluster map of the rate of unemployment (municipalities, 31.12.2007, significant clusters)

5.2 Rate of unemployed with basic education

LISA cluster maps provides clear delimitation of HH and LL regions although the individual clusters are not so compact and stable in time as it was for the rate of unemployment. The high-level clusters are located mainly in peripheral parts of CZ, especially in Usti, Karlovy Vary and Plzen regions. Main area of low-level clustering is along the line Prague-Brno. Transient significant clusters are usually very small and indicate municipalities contrasting to the values around (local extremes).

The development of LISA cluster maps demonstrates the increase of heterogeneity. Such changes are probably induced by a local offer of unqualified vacant places.

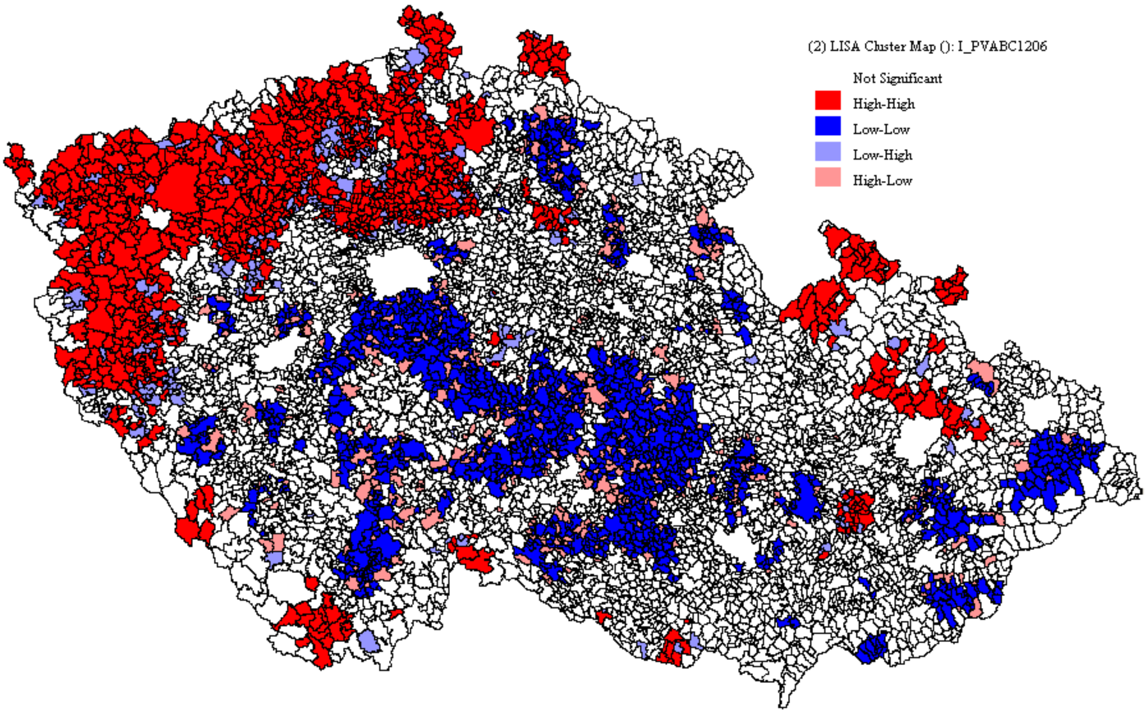

Fig. 12 LISA cluster map of the rate of unemployed with basic education (municipalities, 31.12.2006, significant clusters)

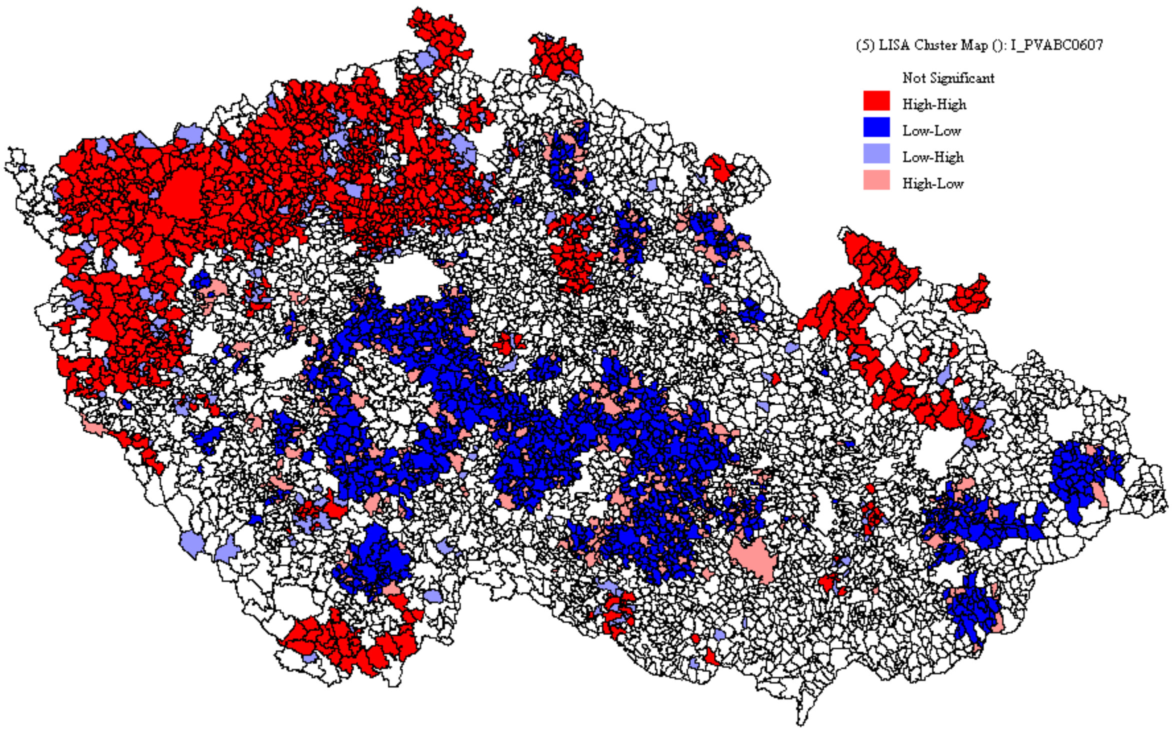

Fig. 13 LISA cluster map of the rate of unemployed with basic education (municipalities, 30.6.2007, significant clusters)

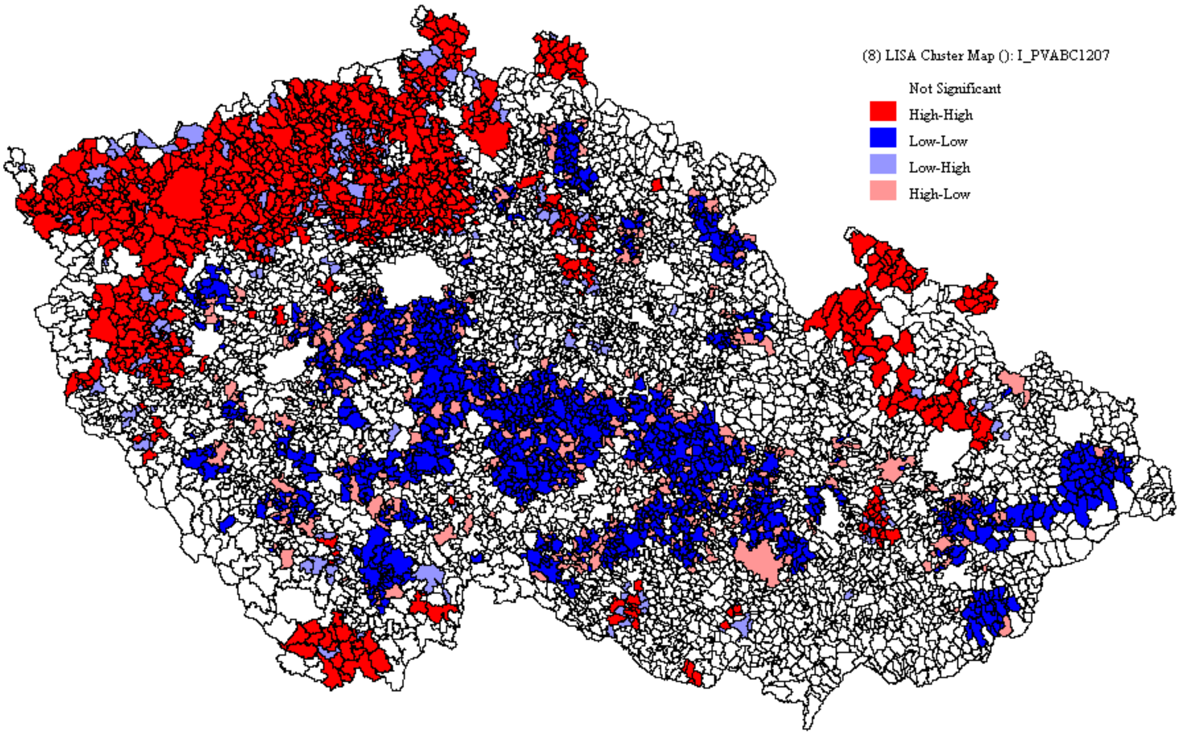

Fig. 14 LISA cluster map of the rate of unemployed with basic education (municipalities, 31.12.2007, significant clusters)

5.3 Rate of long-term unemployed

LISA cluster maps did not show such large homogeneous regions like for the previous indictors. The clusters are small and fragmented.

The temporal development tends to increase homogeneity of both main types of clusters. This feature is probably induced by a decrease of the rate of unemployment and simultaneously an increase of long-term unemployment in the registration. Also the influence of territorial differentiation from the point of view of vacant places availability may take a part.

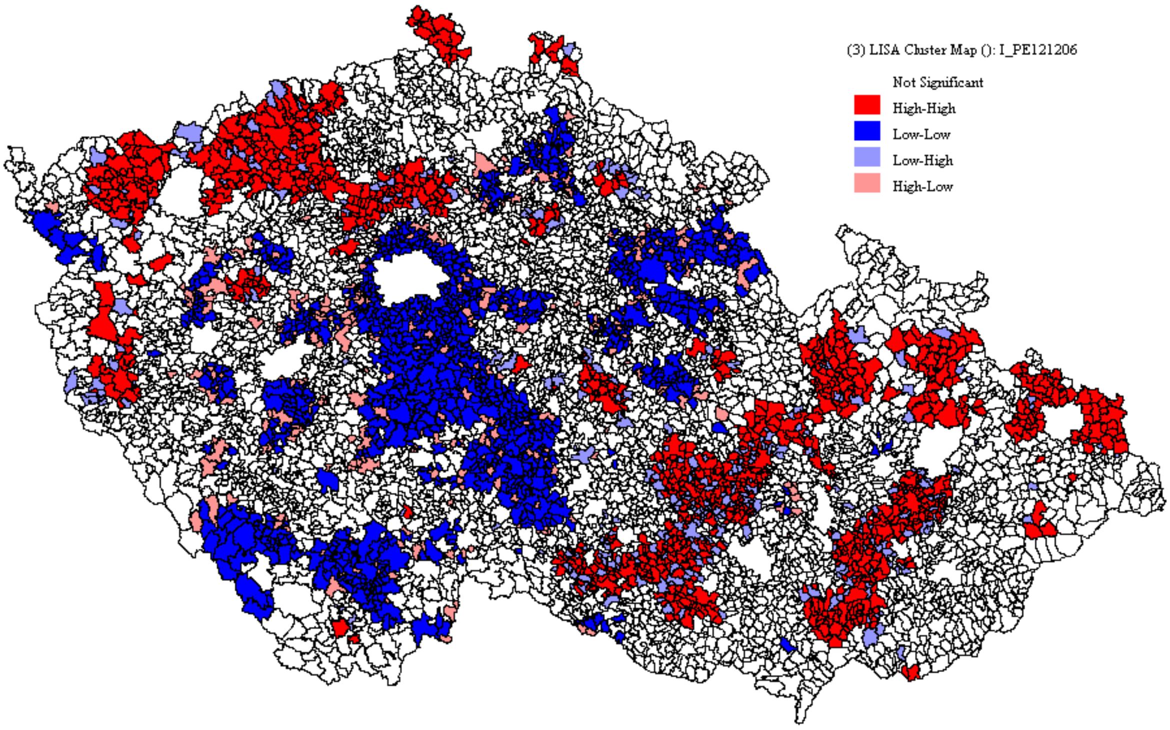

Fig. 18 LISA cluster map of the rate of long-term unemployed (municipalities, 31.12.2006, significant clusters)

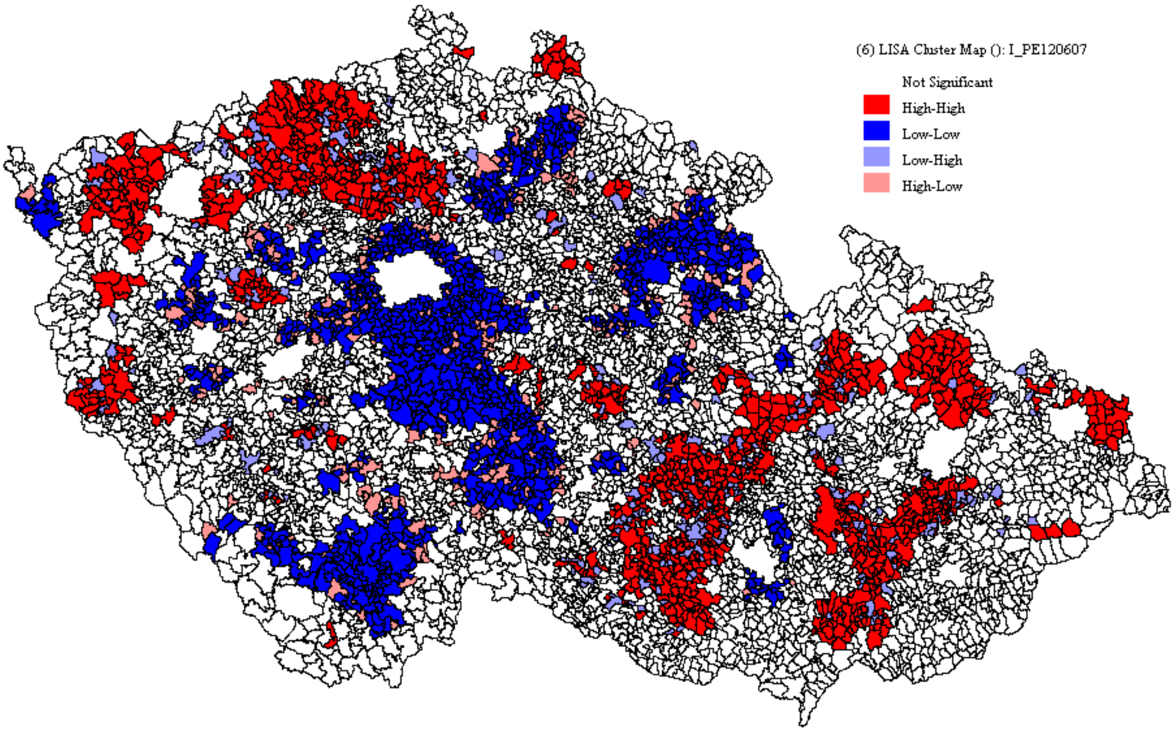

Fig. 19 LISA cluster map of the rate of long-term unemployed (municipalities, 30.6.2007, significant clusters)

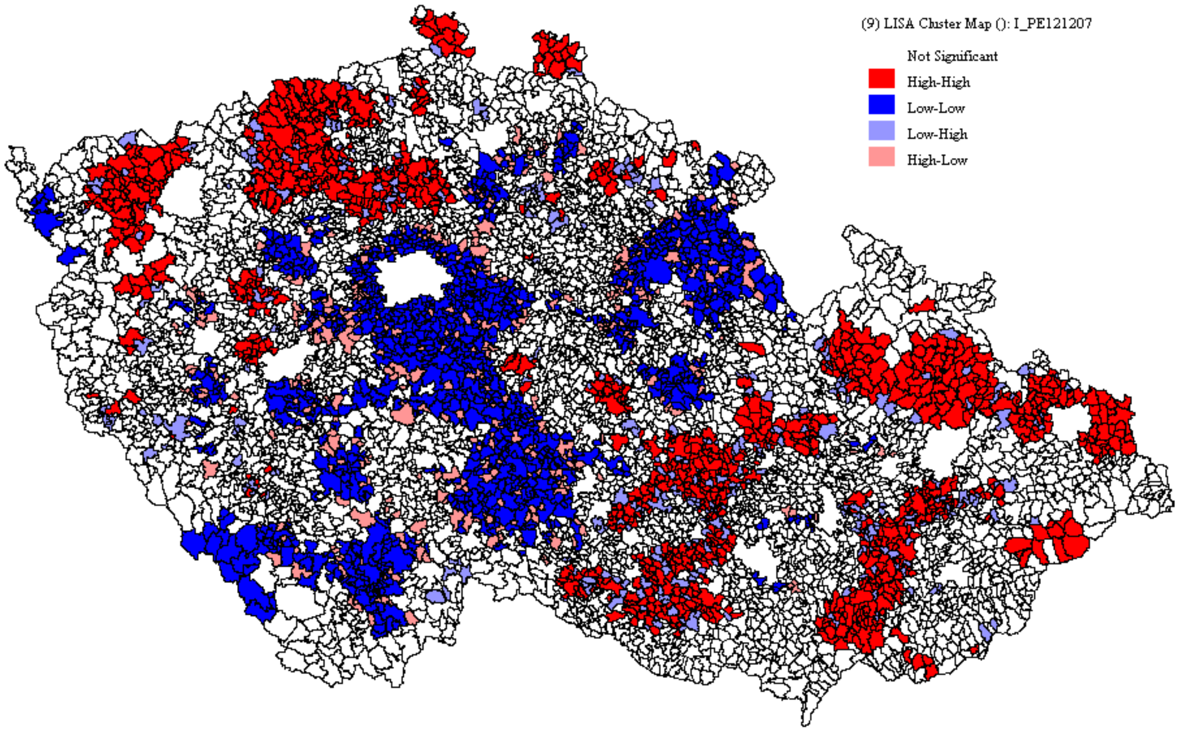

Fig. 20 LISA cluster map of the rate of long-term unemployed (municipalities, 31.12.2007, significant clusters)

5.4 Results of global analysis

The Moran's I index has been computed for all variables in the first time snapshot. Instead of presentation of just one value for the index it is more beneficial to draw a Moran scatter plot. It is the special case of a scatter plot that shows the spatial lag of the variable on the vertical axis and the original variable on the horizontal axis [9]. The variables are standardized so that the units in the graph are expressed in standard deviations. The mean values (zero after standardization) divide the graph into four quadrants which correspond to the basic four types of spatial autocorrelation: high-high (upper right), low-low (lower left), for positive spatial autocorrelation; high-low (lower right) and low-high (upper left), for negative spatial autocorrelation. The slope of the regression line is Moran's I, which is also listed at the top of the graph.

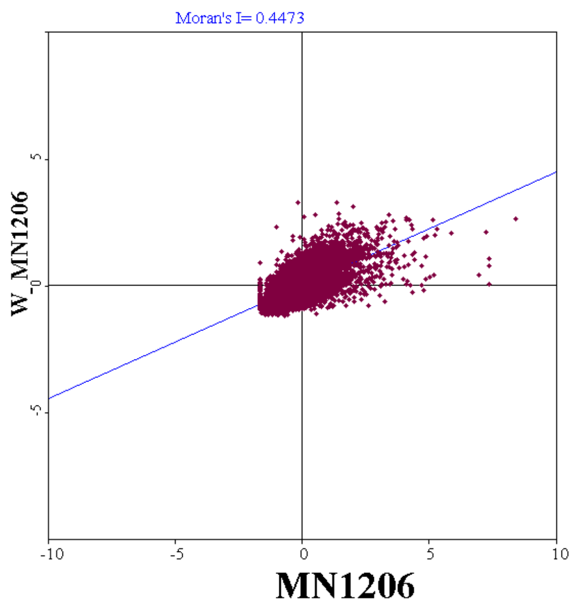

Fig. 21 Moran’s I for the rate of unemployment (municipalities, 31.12.2006)

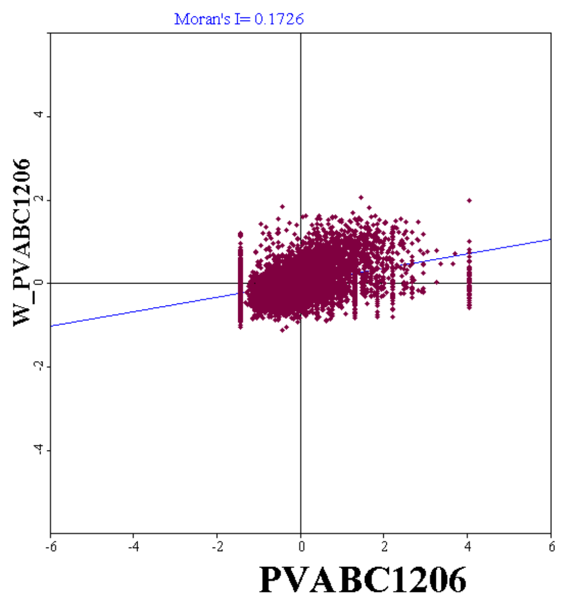

Fig. 22 Moran’s I for the rate of unemployed with basic education (municipalities, 31.12.2006)

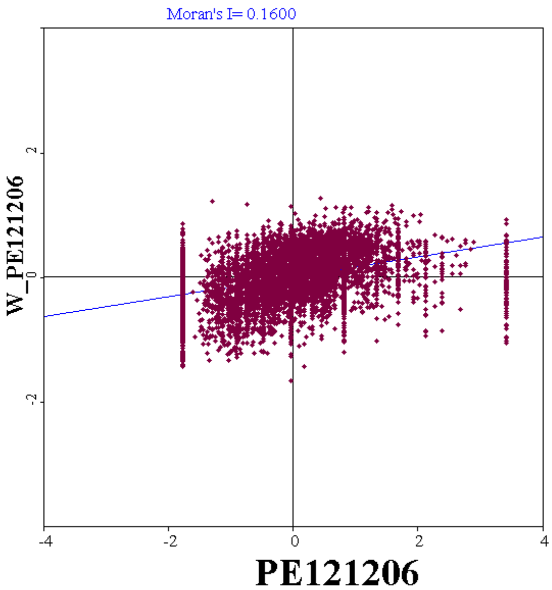

Fig. 23 Moran's I for the rate of long-term unemployed (municipalities, 31.12.2006)

The global Moran's I indicator shows a significant and positive level of spatial autocorrelation for the rate of unemployment (RU) (fig. 21). The values of RU are surrounded by the similar values which invokes the high positive spatial autocorrelation. The asymmetry of the cloud in the graph seems to reflect skew distribution of RU values.

The positive however low spatial autocorrelation can be seen for other indicators. The rate of unemployed with basic education shows Moran's I=0,17 (fig. 22) where the high-high quadrant is more populated. The long-term unemployment (PE12) (fig.23) presents even lower value of Moran's I which is attributed to higher variability of PE12.

In both last plots we can see several vertical lines. Such effect is caused by several frequent values of the variable where each of them is surrounded by a practically full spectrum of values in their neighbourhood.

The randomisations with 999 permutations prove in all cases the significance of results for the level of 0.001.

6 Conclusion

The utilisation of the indicators of spatial association offers new possibilities in the frame of spatial analysis of the labour market as well as for an application of employment policies:

- Better identification of microregions with a distinct, specific regime of labour market
- Monitoring of processes determining creation of new homogeneous (micro-)regions and its stability
- Identification of homogeneous regions across district/region borders which can be useful mainly for preparatory analysis and for the realisation of projects supported from European Social Fund on different levels.
- Local and regional analyses of labour market requested by incoming foreign investors to evaluate the share of selected professions in the concerned area.

The presented results demonstrate some possibilities but also stay new challenges in front of us. The labour market should be more studied using different weight matrices, bivariate global and local indicators and also better tested with randomisation process. Also the possibility of combination spatial autocorrelation approach with *Gini index* computation [6] can bring more information requested for deeper study and understanding of complex processes of labour market.

References

[1] Anselin, L: Local indicators of spatial association – LISA. *Geographical* Analysis. 27. 1995. p. 93-115.

[2] Anselin, L.: Spatial econometrics. In Companion to Econometrics, Baltagi B. (ed.) Basil Blackwell. Oxford, 2001.

[3] Bailey, T.C., Gatrell, A.C. Interactive spatial data analysis. *Essex, Longman Scientific & Technical*, 1995.

[4] Hanuš, F.: Časoprostorové vyhlazení socioekonomických dat. *Disertation thesis, Ostrava 2007. 63 pages*.

[5] Horák, J.: Prostorová analýza dat*. Institute of geoinformatics*, 2006, Ostrava.

[6] Guillain, R., Gallo Le, J.: Measuring agglomeration: An Exploratory Spatial Analysis Approach Applied to the case of Paris and its surroundings. *Proceedings of 1st workshop on Spatial Econometrics,* Rome, 2006.

[7] Journal, A.: Fundamentals of Geostatistics in Five Lessons. *American Geophysical Union*, 1989. 39 p.

[8] Kouba, K.: Prostorová analýza českého stranického systému. Institucionalizace a prostorové režimy. *Czech Sociological Review, 2007, Vol. 43, No. 5*. p. 1017-1037.

[9] Lee, J., Wong, D.: Statistical analysis with ArcView GIS. *USA, John Wiley & Sons Inc.,* 2001, 192 pages, ISBN 0-471-34874-0.

[10] Patuelli, R., Griffith D. A., Tiefelsdorf, M., Nijkamp, P.: The use of spatial filtering techniques: the spatial and space-time structure of German unemployment data. *Proceedings of 1st workshop on Spatial Econometrics*, Rome, 2006.

[11] Smans, M., Esteve, J.: Practical approaches to disease mapping. In *Elliott P., Cuzick J., Engish D., Stern R.(ed.): Geographical & Environmental Epidemiology. New York, Oxford University Press, 1996, s. 141-150, ISBN 0-19-262235-8.*

[12] Tobler, W.: A computer movie simulating urban growth in the Detroit region. *Economic Geography, 1970, Vol. 46, no. 2*, p.234-240.

[13] Anselin, L., Syabri I., Kho, Y.: GeoDa: An introduction to spatial data analysis. *Geographical Analysis. Vol. 38, no. 1, p. 5-22, 2005.*

[14] Robinson A., Morrison J, Muehrcke P., Kimerling A., Guptill S.: Elements of Cartography. 6th edition. Wiley, 1995, 674 pages. ISBN 0-471-55579-7.

[15] Kraak M., Ormeling F.: Cartography. Visualization of spatial data. 3rd edition, Longman, 1998, 222 pages. ISBN 0-582-25953-3.

[16] Voženílek, V.: Cartography for GIS - geovisualization and map communication. *Olomouc, UP, 140 pages, 2005. ISBN 80-244-1047-8*