Techniques for Representation of Regional Clusters in Geographical Information Systems

Adriana REVEIU
Academy of Economic Studies Bucharest
reveiua@ase.ro

This paper provides an overview of visualization techniques adapted for regional clusters presentation in Geographic Information Systems. Clusters are groups of companies and institutions co-located in a specific geographic region and linked by interdependencies in providing a related group of products and services. The regional clusters can be visualized by projecting the data into two-dimensional space or using parallel coordinates. Cluster membership is usually represented by different colours or by dividing clusters into several panels of a grille display. Taking into consideration regional clusters requirements and the multilevel administrative division of the Romania's territory, I used two cartograms: NUTS2- regions and NUTS3- counties, to illustrate the tools for regional clusters representation.
Keywords: *Geographic Information Systems, Regional Clusters, Spatial Statistics, Geographic Data Visualisation*

1 Introduction

In recent years, Geographic Information Systems (GIS) have become an important tool for regional and urban research, and not only, as about 80-90% of data collected and used for regional and environmental information systems are related to geography. [1]

The huge amount of spatial data generated by GIS expansion, the increasing number of geographic informatic applications available, the computerization of a large amount of information sources, and the availability of digital map has increased the opportunity and need for the usage of methods for spatial classification, for both research and applied purposes.

The main problems identified when we try to classify spatial data are:

- The large number of types of areas: administrative, cultural, social, and so on,
- The large number of types of variables available for geographic and economic data representation,
- The non-normal variable distributions, because most geographic data usually have very complex frequency distributions,
- The nonlinear relationships between various variable in analysis,
- The spatial dependency of analysed variable,
- The uncertainty of economic, social and geographic data,
- Systematic non-random variations in spatial representation.

GIS provides: an integrated computing environment for economic, social and environmental data integration, a set of analytical capabilities to operate on topological relationships and on the spatial aspects of the geographical data, a set of interactive and programming tools to manage the non-spatial attributes of economic and social data and tools to combine the non-spatial and spatial attributes of geographical data.

GIS facilitates the integration of disparate data sets, creation of new and derivative data sets, development and analysis of spatially variables.

GIS is connected with the process of obtaining information from spatial data. There are two important issues related with the process of data capture: the information acquisition and the information accuracy.

There are two main techniques to turn data into information: first of all the visualisation of economic and social information on the maps and the statistical and mathematical modelling of spatial and non-spatial attributes of data.

There is a vast array of techniques that have been developed for visualizing spatial data

and it remains a very intense and fruitful area of research.
GIS based visualization of environmental data allows the user to interact with the real world from a distance, through interaction with and immersion into artificial worlds.
The basic elements of a successful visualization and usage of geographical data is the understanding of how human perceive GIS models usage, how people think about space and time and how spatial environments might be better represented using computers and digital data.
There are four principal purposes of the visualization of geographical data:
- **Exploration** - to establish whether and to what the general message of a dataset is sensitive to inclusion or exclusion of particular data elements.
- **Synthesis** - to present the range, the complexity, and detail of data sets in ways that can be quickly assimilated by the users.
- **Presentation** - to communicate the complex message of a representation in an intelligible manner, and to enable the user to understand the expected overall quality of the representation.
- **Analysis** - to provide an environment to support a set of methods and techniques of spatial analysis.

2 The Features of Regional Clusters

In the contemporary knowledge economy, what is crucial is not so much the speed of development and the dissemination of new scientific solutions, as the diversity and spread way in which occurs the innovative process. Companies are present in many contexts and in a range of locations. The consequence is the formation of networks of global links, from ownership to alliances and production partnerships, aimed at organising externality, like the relations with other companies and different socio-economic environments, which can no longer be mastered through the usual form of expansion in size. [2]
The economic activity is concentrated in space, and therefore there is an increased attention over the forces of agglomeration and the role of location in economic development. Porter [3] defines clusters as a group of interconnected companies and associated institutions, close from geographical point of view, working in a particular field and linked by common and complementary elements.
Because of the proximity among them, both in terms of geography and of activities, the clusters constituent enjoy the economic benefits of several types of positive location-specific externalities.
Knox [4] defines a spatial cluster as, a geographically bounded group of occurrences of sufficient size and concentration to be unlikely to have occurred by chance. This is a useful operational definition, but there are very few situations when phenomena are expected to be distributed randomly in space. In most cases an implicit assumption in spatial cluster analysis is that the researcher has accounted for all the factors known, to influence the variable of study.
From a functional perspective, clusters are defined as networks of independent producers of powerful firms, including specialized suppliers, linked to each other in the value-added production channel. [5]
Spatial proximity has grown rapidly in importance, the cluster literature have made the distinction between industrial complexes and industrial clusters on spatial agglomeration of these industrial groups. Spatial proximity of the industrial activities interconnected assumed to influence the performance of these sectors and regional clusters on short and long term [6].
Clusters differ in many dimensions, such as: the type of products and services they produce, the location dynamics they are subject to, their stage of development, and the business environment that surrounds them.
Clusters can be **classified** by the type of product and services they provide. There are clusters in automotive, in financial services, in tourism, in a specific industrial area, and so on. Recent research has pointed out how different locations play different roles. The development of clusters has discouraged many regions with no realistic chance of

achieving a similar level of performance as the top level clusters.
From location point of view, the *local industries* are serving only local markets and are distributed across space approximately according to population size. They might be a kind of a cluster in a more narrow geographical sense, like a part of a city, due to the complementarities in attracting customers, but these effects are not strong enough to influence the development of clusters across regions.
On the other hand, the *natural resource dependent industries* serve global markets and are concentrated across space, in areas in which there are natural resources presented.
Finally, there are many industries that choose their location according to the quality of the cluster-specific business environment. There are the cases of *traded industries* which serve markets in many regions and countries, and concentrate across various geographic locations. The cluster belonging to one industry is strong and its presence is a key part of the attractiveness of a specific location. Understanding the differences between these types of industries is important, because it affects the types of policies that are relevant to upgrade them.
From another point of view, clusters can be classified by the stage of development they have reached. The stage of development depends on two dimensions: on the quality of the external business environment the cluster operates in, and on the progress the cluster has made in mobilizing the potential of its business environment through active cooperation and other internal activities. [7] Researchers have looked at clusters in less developed economies as well as in less developed regions of advanced economies, such as rural regions or inner cities. [8] Most of the theoretical literature suggests that clusters are a factor at every stage of economic development, but that in weaker environments clusters will tend to be weaker and narrower as well. [8] Researchers have focused on the role of cultural factors, institutions, and individual leadership. There is strong view in the literature that cluster dynamics do not occur automatically, but that they depend on and can be reinforced by purposeful action. [8]

3 Techniques for Regional Clusters Data Visualization in GIS

Data visualization can greatly enhance the understanding of multivariate data structures. The regional clusters can be visualized by projecting the data into two-dimensional space or using parallel coordinates. Cluster membership is usually represented by different colours or by dividing clusters into several panels of a grille display. [9] In addition, silhouette plots provide a popular tool for diagnosing the quality of a partition. Some of the popularity of self-organizing feature maps [10] with practitioners in various fields can be explained by the fact that the results can be “easily” visualized.
The majority of modern GIS applications are characterized by sophisticated graphics and this allows GIS to provide effective support for problem of geographical data representation in space.
Cartography is an important source of principles for the design of business graphics. [11] Information visualization techniques have been widely applied in science and geography, but have only been recently integrated into business applications. [12]
The complex nature of spatial data requires GIS to use sophisticated visualization techniques to represent information. It is therefore quite challenging for GIS to provide an interactive interface on the same screen.
Visualization has been recognized in the GIS community as an important aspect of GIS. [13] This may reflect support for the map view of GIS. One limitation of GIS interface designs is that they are seen to provide a means for visualizing results only, rather than providing a comprehensive problem representation for all stages of the problem. [14]
There are a lot of techniques and tools used for data visualisation in GIS. In this paper I present only the tools that could be adapted for representing regional clusters in GIS.
The most important tools and techniques available for regional clusters representation are:

- Cartograms,
- Spatial statistics,
- Spatial query functions,
- Integration of ordinary data stored in databases with spatial data.

Cartograms are unique representations of geographical space. A cartogram is a type of graphic that describes attributes of geographic objects as the object's area. Mapping requirements include the preservation of shape, orientation contiguity, and data that have suitable variation. Successful communication depends on how well the map reader recognizes the shapes of the internal enumeration units, the accuracy of estimating these areas, and effective legend design. Because a cartogram does not depict geographic space, but rather changes the size of objects depending on a certain attribute, a cartogram is not a true map. Cartograms vary on their degree in which geographic space is changed: some appear very similar to a map, however some don't look like a map. [15]

There are three main types of cartograms, each having a very different way of showing attributes of geographic objects: non-contiguous, contiguous and Dorling cartograms.

A **non-contiguous cartogram** is the simplest and easiest type of cartogram. In a non-contiguous cartogram, the geographic objects do not have to maintain connectivity with their adjacent objects. This connectivity is called topology. By freeing the objects from their adjacent objects, they can grow or get smaller in size and still maintain their shape.

In a **contiguous cartogram** is maintained the true topology, the objects remain connected with each other, but this causes great distortion in shape. In this case, the cartographer must make the objects of the appropriate size, to represent the attribute value, but he or she must also maintain the shape of objects as best as possible, so that the cartogram can be easily interpreted.

A **Dorling cartogram** maintains neither shape topology nor object centroids, though it has proven to be a very effective cartogram method. This type of cartogram was named after its inventor, Danny Dorling of the University of Leeds. To create a Dorling cartogram, instead of expanding or shrinking the objects themselves, the cartographer will replace the objects with a uniform shape, usually a circle, of the appropriate size.

Spatial statistics comprises a set of techniques for describing and modelling spatial data. In many ways they extend what the mind and eyes do intuitively, to assess spatial patterns, distributions, trends, processes and relationships. Unlike traditional (non-spatial) statistical techniques, spatial statistical techniques actually use space: area, length, proximity, orientation, or spatial relationships, directly in their mathematics. [16]

A GIS interface can be used to query a **database**, although this requires a more sophisticated interface with the ability to formulate a query using the interactive commands. Database capabilities allow queries to be generated in the GIS, to show only areas selected by attribute value. This type of software also allows spatial database queries, such as selection of a particular region and operations such as buffering or overlay.

Spatial query functions are central tools used in many GIS applications available on the Internet. For many applications, spatial query is the main feature of a GIS application. In other applications, spatial query is the foundation for any advanced geographical analysis. Spatial query may seems to be the routine, but involves complex operations, mainly when the objects of spatial queries are continuously updated, in real time applications.

4 Using Spatial Tools to Represent Romania's Regional Clusters in GIS

In terms of representing regional clusters in GIS, a usual way is to generate cartograms. In this case, a cartogram supposes to define a colours ramp, having a number of colours equal with the number of various kind of clusters defined.

Because of an administrative unit is usually represented in GIS as a polygon, a way to display a cluster on the map is to fill with a proper colour the administrative units that belongs to the cluster.

To define regional clusters related to one domain of activity it is necessary to use a large set of values, of various indicators that can be stored in a database table or as independent data.
Cartogram can be defined taking into consideration multilevel administrative division of the territory. For example, according with the Nomenclature of Territorial Units for Statistics, Romania uses the following levels: NUTS1- macro-regions, NUTS2- regions, NUTS3- counties and LAU2- communes. Analysis can be applied on different levels of administrative organization by aggregating the values of indicators. For example, the value of a indicator for a county can be obtained by aggregating all indicator values, for all communes, belonging to the county.
To exemplify regional cluster representation in GIS, I choose the ESRI (Environmental Systems Research Institute, Inc)'s Arc Gis solution, because it has over a million software users worldwide, and installations at over 5000 universities and is a world leader for the design and development of Geographic Information Systems (GIS) software.[17]
The options available to identify regional clusters in Arc Gis are available in *Spatial Statistics* toolbox and include both statistical functions and general purpose utilities designed for regional clusters.
The options available are grouped in some toolsets, and are presented below.

Measuring geographic distributions

The tools from the *Measuring Geographic Distributions* toolset are descriptive tools, helping to summarize the most important characteristics of a spatial distribution of economic, social and geographic data.
I am exemplifying the usage of these tools using the Romania's administrative maps.
Central feature tool identifies, like in figure 1, the most centrally located feature in a point, line or polygon feature class. This is figured using green colour and is marked on the map with "*central feature*" tag. This is the administrative unit (county) located in the most central position, on the map. Even the Brasov county is not a distinct part of the NUTS2 level, because I used the NUTS3 (counties) level to create the NUTS2 (regions) level, all the counties are available behind the regions.
Directional distribution is the standard deviational ellipse and measures how concentrated features are, around the geographic mean, and whether or not they reveal a directional trend. This is figured on the map using a diagonal pattern and is marked in Figure 2 with "*directional distribution*" tag.

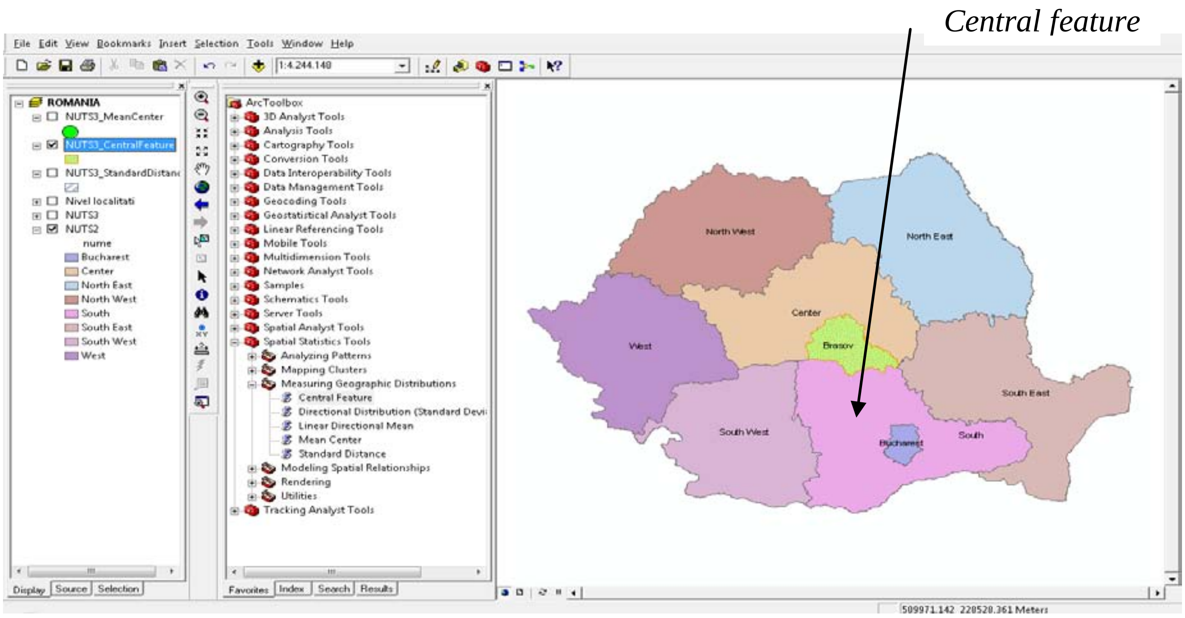


Fig. 1. Using *Central feature* tool

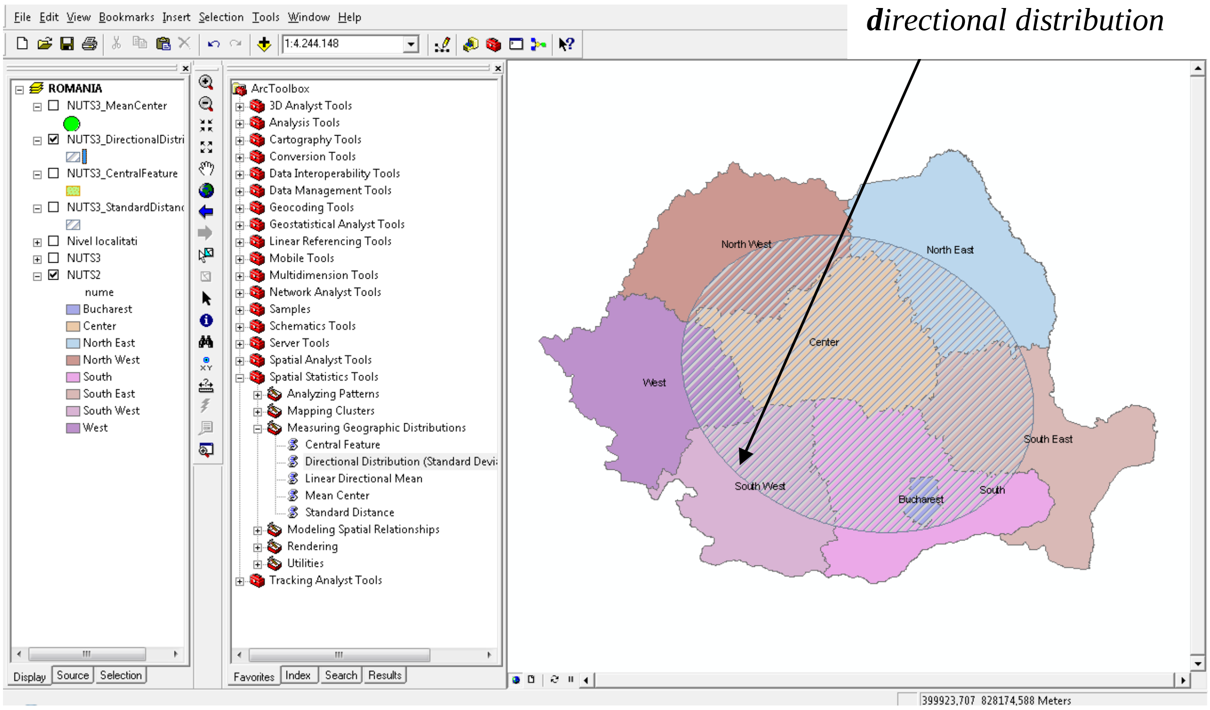


Fig. 2. Using *Directional distribution* tool

Mean center tool identifies the geographic centre for a set of features. This is figured in Figure 3, using green colour and is marked, on the map, with "*mean center*" tag.

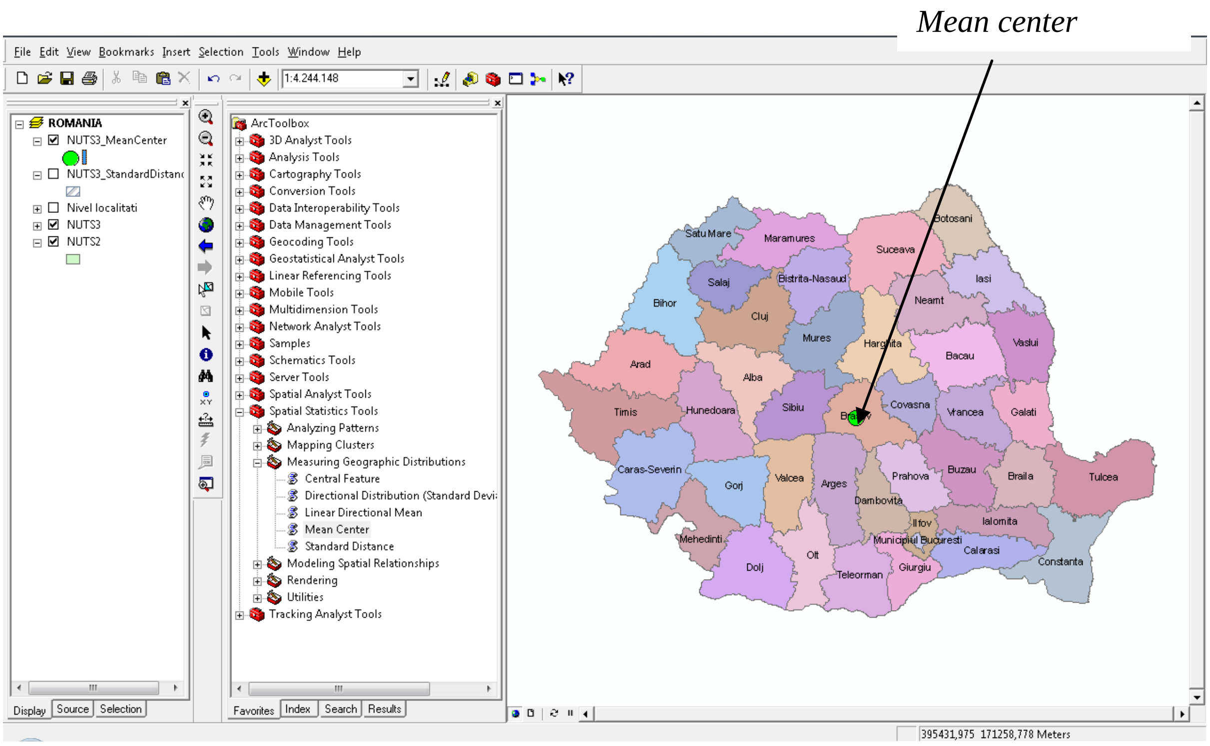


Fig. 3. Using *Mean center* tool

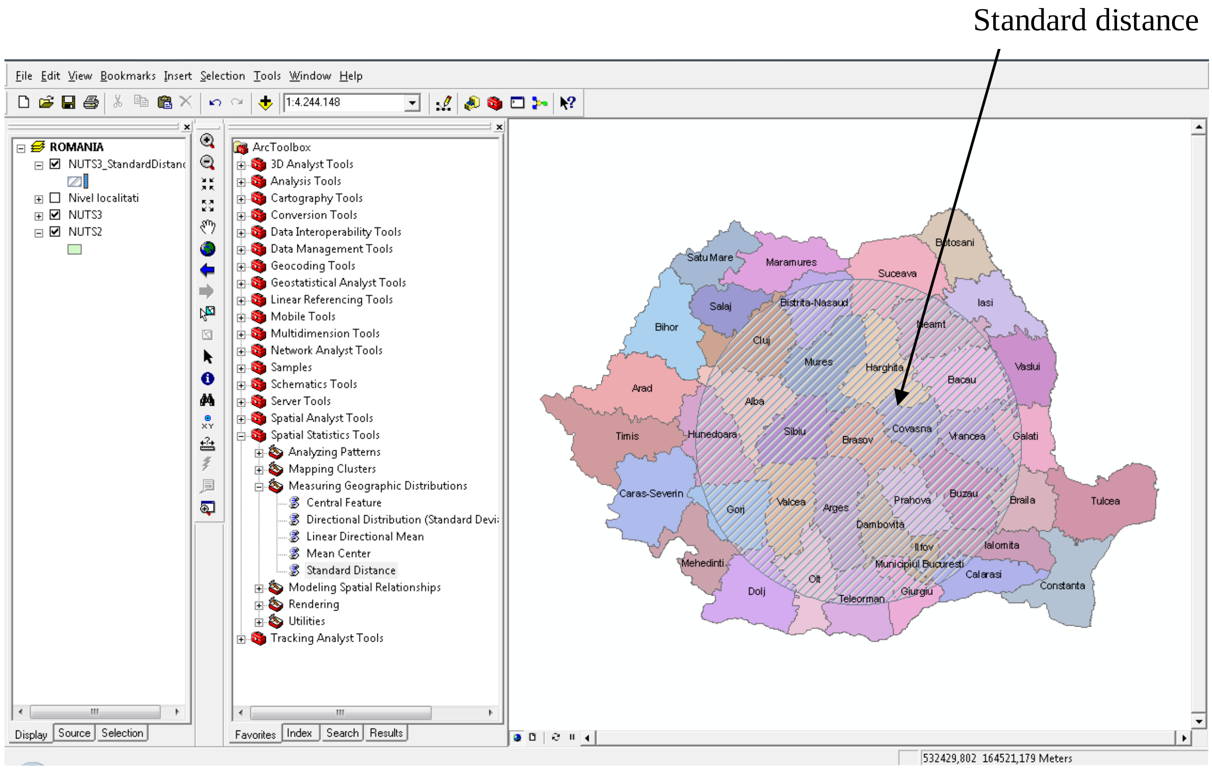


Fig. 4. Using *Standard distance* tool

Standard distance tool measures the degree to which features are concentrated or dispersed around the geographic mean centre. On the map, this is figured using a diagonal pattern and is marked, in figure 4, with "*standard distance*" tag.

Mapping clusters

Tools from the *Mapping Clusters* toolset identify where spatial clustering occurs, and where spatial outliers are located.

In figure 5, the *Anselin's local Moran's I* tool is used for a cluster and outlier analysis, to identify regional clusters, based on the number of employees from Romania's counties. To exemplify the usage of this tool I used the average number of employees from each county of Romania, in 2008. The data source is the Romania Statistical Yearbook, 2009. The goal of this tool is to identify clusters of high or low values as well as spatial outliers, using like a set of features the number of employees and the spatial location of each county.

In the Figure 5 we can see a string of outliers separate clusters of a high number of employees, in the West (Timisoara, Arad) and in the Bucharest region (Bucharest, Ilfov, Prahova) from clusters of low number of employees in the south (Giurgiu, Dambovita, Calarasi).

The *Hot Spot Analysis (Getis-Ord Gi*)* tool is used to identify clusters with high values, named hot spots, and clusters with low values, named cold spots, based on a set of given weighted features.

I applied the Hot Spot Analysis tool to employees' data for Romania's counties, and the results are presented in Figure 6. Running the Hot Spot Analysis tool, the hot spot clusters are in the West of Romania and around the Bucharest and cold spot clusters, areas with high differences between counties, like in the case of Cluj-Salaj area and Harghita and Braila counties, where the difference between number of employees from these counties and the neighbours is high.

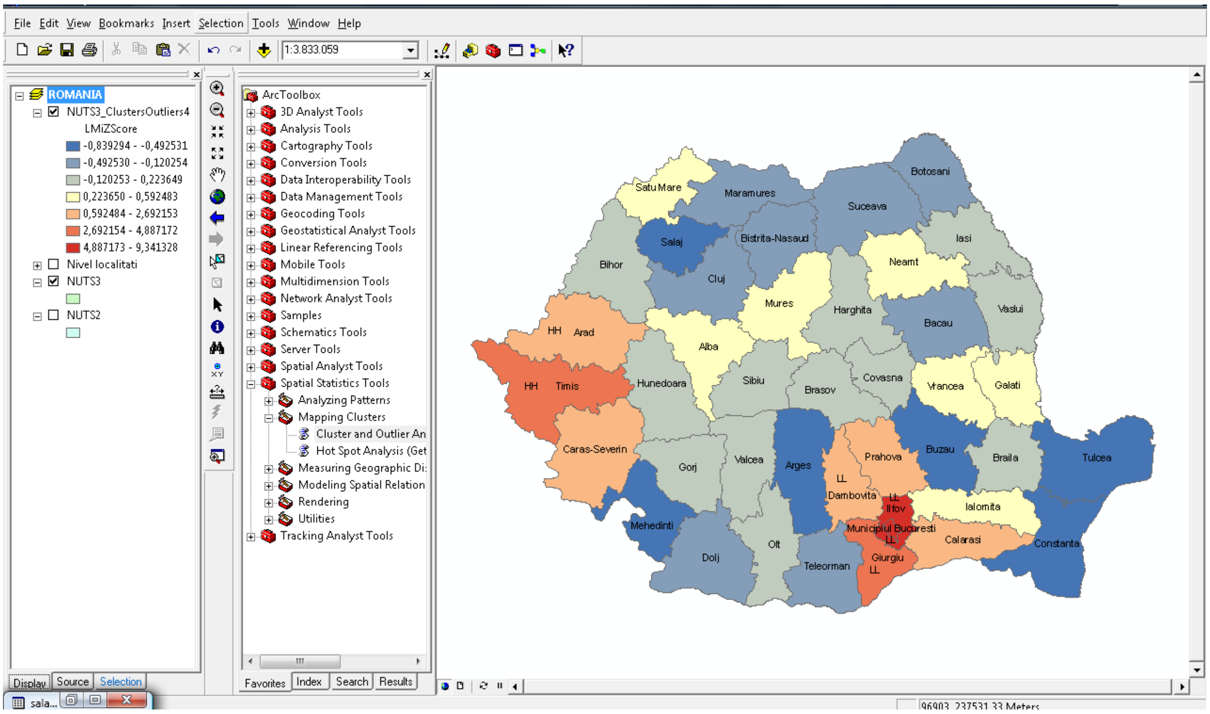


Fig. 5. Using *Anselin's local Moran's I* tool

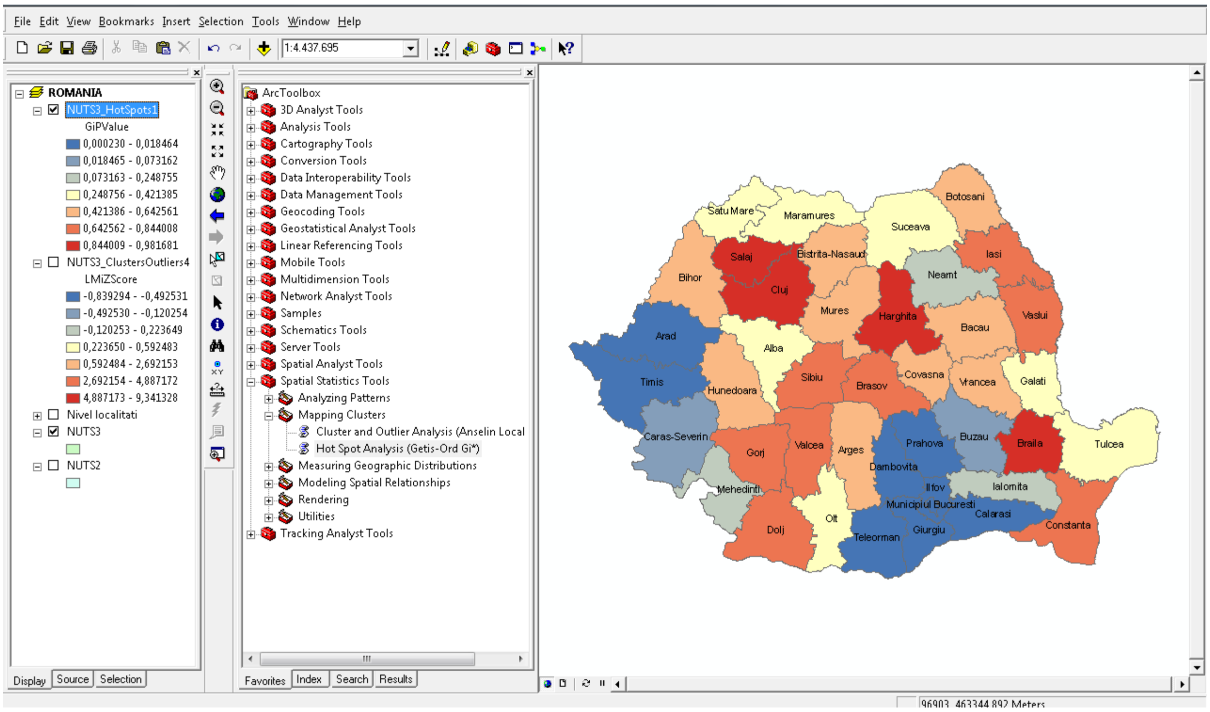


Fig. 6. Using *Hot Spot Analysis* tool

Modelling spatial relationships

The tools from the *Modelling Spatial Relationships* toolset fall into two categories: the first category includes tools designed to help the user to define a conceptual model of spatial relationships and the second class includes ordinary least squares and geographically weighted regression.

The conceptual model is an integral component of spatial modelling and should be selected so that it best represents the structure of spatial dependence among the features being analysed. [18]

The options available for modelling spatial relationships include: inverse distance, fixed distance, polygon contiguity (Rook's and

Queen's case), *k* nearest neighbours, Delaunay triangulation, travel time and travel distance.

The options available in the toolset are:

- *Generate network spatial weights* which build a spatial weights matrix file, specifying spatial relationships among features in a feature class based on a Network dataset.
- *Generate spatial seights matrix* which builds a spatial weights matrix file, specifying spatial relationships among features in a feature class.
- *Geographically weighted regression* is a local form of linear regression used to model spatially varying relationships among a set of data variables.
- *Ordinary least squares regression* performs global linear regression to model the relationships among a set of data variables.

Regression analysis could be used to model, examine and explore the spatial relationships, in order to understand the factors behind the observed spatial patterns or to predict spatial outputs.

In the Figure 7, I exemplify the usage of *Geographically weighted regression* tool to model spatially relationship between two data sets: the number of employees and the average gross nominal monthly earning from each county of Romania.

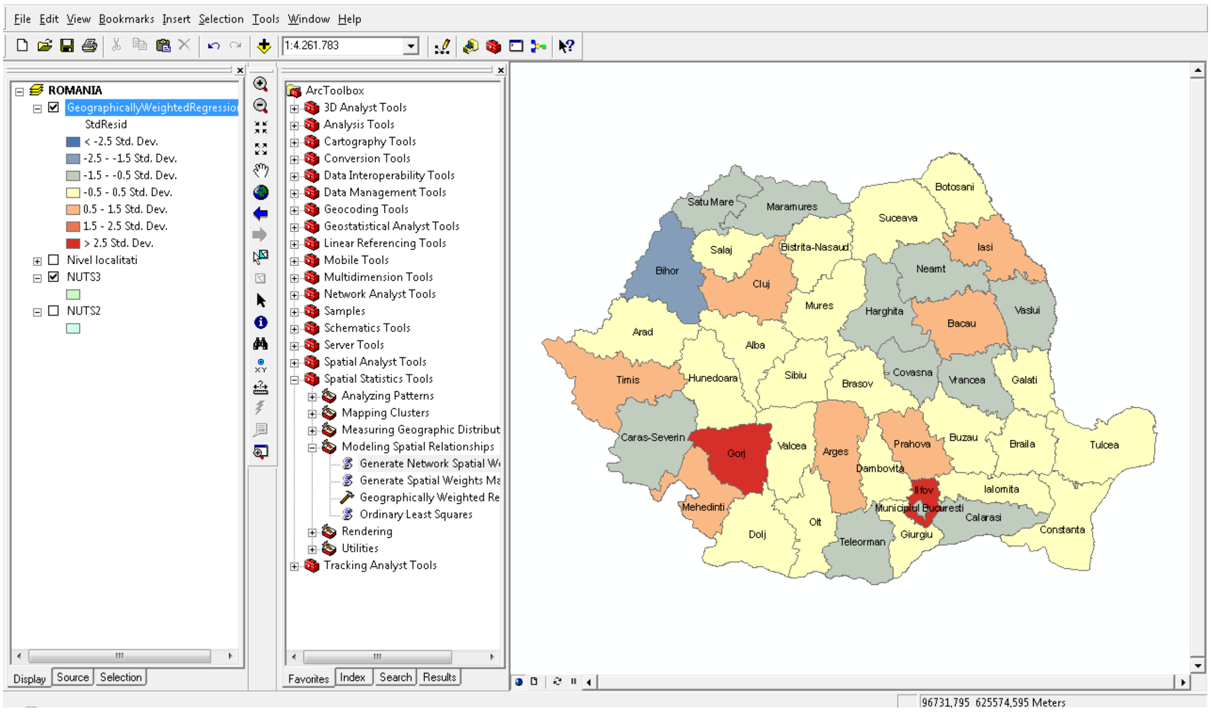


Fig. 7. Using *Geographically weighted regression* tool

OLS (*Ordinary Least Squares*) is a global model which creates a single equation to represent the relationship between the independent variable and each of explanatory variables.

Global models, like OLS, are based on the assumption that relationships are static and consistent across the entire study area. When the relationships perform differently in separate parts of the study area, the global model becomes less effective.

Local models, like GWR (*Geographically Weighted Regression*), create an equation for every feature in the dataset, adjusting each one using the target feature and its neighbours. Nearby features have a higher weight in the calibration than features that are farther away. So, the relationships we are trying to model are allowed to change over the study area.

5 Conclusions

The huge amount of spatial data generated by GIS expansion, the increasing number of geographic informatic applications available, the computerization of a large amount of information sources, and the availability of digital map has increased the opportunity and need for the usage of methods for spatial classification, for both research and applied purposes.

The paper identified two main techniques to turn data into information: the visualisation of economic and social information on the maps, and the statistical and mathematical modelling of spatial and non-spatial attributes of data.

The main purposes of the visualization of geographical data distinguished in this paper are: exploration, synthesis, presentation and analysis.

The environmental component is very important in the process of defining regional clusters, but it's quantification is difficult to be realised compared with socio-economic data quantification. For these reasons, the usage of specialised tools available in professional Geographical Information Systems it is a good solution, provides the best results and provides a way to integrate the spatial data with socio-economic and statistical data.

Acknowledgments

This work was supported by the project "Performance and excellence in postdoctoral research in Romanian economics science domain", contract POSDRU/89/1.5/S/59184.

References

[1] W. Huxhold, *An introduction to urban geographic information systems*, Oxford University Press, 1991.

[2] R. A. Boschma, R. C. Kloosterman, "Learning from Clusters. A Critical Assessment from an Economic-Geographical Perspective", *GeoJournal Library*, Vol. 80, Springer-Verlag Berlin Heidelberg, 2005.

[3] M. Porter, *Clusters and the New Economics of Competition*, Harvard Business Review, 1998.

[4] E. G. Knox, *Detection of clusters*, in Elliott P (ed) *Methodology of enquiries into disease clustering.*, Small Area Health Statistics Unit, London, pg.17-20.

[5] T. J.A. Roelandt, P. den Hertog, *Boosting innovation; the cluster approach*, 1999, Paris, OECD.

[6] P. Maskell, *Towards a Knowledge-based Theory of the Geographical Cluster*, Industrial and Corporate Change, 2001, pg. 919-941.

[7] M. Enright, M., *Regional clusters and economic development: A research agenda*, Business Networks: Prospects for Regional Development, Walter de Gruyter, Berlin, 1996.

[8] M. Porter, *The Competitive Advantage of the Inner City, In: On Competition, Boston*, Harvard Business School Press, 1998.

[9] R. Becker, W. Cleveland, M.-J Shyu, "The visual design and control of trellis display", *Journal of Computational and Graphical Statistics*, vol. 5, 1996, pg. 123-155.

[10] T. Kohonen., *Self-organization and Associative Memory*, 3rd edn, Springer, New York, 1989.

[11] G. DeSanctis, "Computer graphics as decision aids: Directions for research", *Decision Sciences*, vol. 15, no. 4, 1984, pg. 463-487.

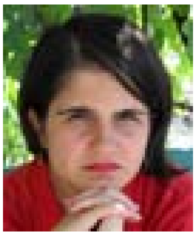
[12] C. Speier, M. G. Morris, The influence of query interface design on decisionmaking Performance, *MIS Quarterly*, Vol. 27, No. 3, pg. 397-423, 2003.

[13] B. P. Buttenfield, W. A. Mackaness, *Geographical Information Systems, Volume 1: Principles*, Harlow, Essex, UK: Longman Scientific & Technica, 1991, pg. 427-443.

[14] A. D. Blaser, M. Sester, M. J. Egenhofer, "Visualization in an early stage of the problem-solving process in GIS", *Computers & Geosciences*, Vol. 26, No. 1, pg. 57-66, 2000.

[15] D. D. Borden, *Cartography: Thematic Map Design*, 4th edition, William C Brown Publishing House, 1996.

[16] L. Scott, A. Getis, *Spatial statistics. Encyclopedia of geographic informations*, Sage, Thousand Oaks.
[17] M. Fischer, A. Getis, *Handbook of Applied Spatial Analysis*, Springer-Verlag Berlin Heidelberg, 2010.
[18] A. Getis, J. Aldstadt, "Constructing the spatial weights matrix using a local statistic", *Geogr Anal*, vol. 36, no. 2, pg. 90-104.
[19] I. Armaş., M. Dârdală, A. Reveiu, T. F. Furtună, *Spatial Modeling of Urban Environmental Vulnerability to Seismic Risk. Case Study: The Historical Center of the Bucharest Municipality – Romania*, 3rd WSEAS International Conference on Urban Rehabilitation and Sustainability (URES'10), University of Algarve, Faro, Portugal, November, 3-5, 2010, in Volume: Advances in Urban Rehabilitation and Sustainability, Published by WSEAS Press, 2010, pg. 49-54.



Adriana REVEIU has graduated the Faculty of Cybernetics, Statistics and Economic Informatics. She holds a PhD diploma in Economic Cybernetics and Statistics. She is associate professor in Economic Informatics field and branches within Department of Economic Informatics at faculty of Cybernetics, Statistics and Economic Informatics from Academy of Economic Studies. She is the author and co-author of 10 books and over 50 articles in journal and proceedings of national and international conferences, symposiums, workshops in the fields of multimedia, GIS, communications, learning systems and data management.