Associate Professor Titus Felix FURTUNĂ, PhD
The Bucharest University of Economic Studies
Associate Professor Adriana REVEIU, PhD
The Bucharest University of Economic Studies
Professor Marian DÂRDALĂ, PhD
The Bucharest University of Economic Studies
Roman KANALA, PhD
University of Geneva, Switzerland

ANALYSING THE SPATIAL CONCENTRATION OF ECONOMIC ACTIVITIES: A CASE STUDY OF ENERGY INDUSTRY IN ROMANIA

Abstract *The paper proposes a solution, designed to identify and to analyse the spatial concentration of economic activities. Statistical methods of data analysis adapted by us are used to provide relevant results for analysis of economic and spatial data, and to support the research conclusions. The analysis of spatial concentration of economic activities is done at the national and regional levels and is based on two types of data: spatial and economics. A model of analysis, based on Principal Component Analysis (PCA) and cluster analysis methods has been proposed in the paper. A software component has been designed and implemented in order to manage the economic and spatial data, and to develop the data analysis. A case study for energy industry of Romania has been proposed In order to apply the proposed analysis. This kind of analysis is useful for policy makers, entrepreneurs and economic analysts to investigate and evaluate the spatial distribution of economic activities, in order to support their decisions.*

Key words: *PCA, cluster analysis, economic concentrations, GIS, data modelling.*

JEL Classification: **C38, C46, L94, L95**

1. Introduction

Economic activities start, grow and develop in space. Natural resources are spatially distributed in a non-uniform way; typically they are concentrated in some places while they are scarce or absent elsewhere. Economic activity tends to follow the localisation pattern of natural resources. Disparities in the geographical distribution of resources and economic activities generate differences in wages, levels of wealth and well-being, and different degrees of control over local development.

According to McCann (2001), most activities developed in one industry tend to be spatially concentrated. Subsequently, suppliers, support services and commercial activities take place in the immediate vicinity of such localised production activities. At the same time, not all activities are implemented in the same location.

Some activities can be scattered over large areas and goods shipped over large distances.

Spatial behaviour of companies depends both on the attributes of each company itself and upon the regional context. In a region with a prevalent industry sector, any change of its economic performances will have a large aggregate impact on the entire region.

To cope with the two contradictory phenomena, general globalization and administrative management at regional level, detection of economic concentration areas in one industry helps to address cooperation issues between companies. Such cooperation between companies located in the same area triggers synergies in day-to-day operations (such as economy of scale in infrastructure cost sharing and group tariffs for inputs) but also brings new opportunities and win-win situation that both individual companies and the whole region benefit of. In this regard, (Pavlinek, 2005) is distinguishing two attributes of new investments: embeddedness (integration in local and regional economies) and path dependency (greenfield vs brownfield investment).

Identifying concentration areas for economic activities developed in one industry sector is useful in developing this kind of analysis. The approach described in this paper uses the natural widespread of economic activity, function on the localisation of companies acting in a specific industry.

Based on our literature review, the best methods used to identify the spatial concentration of economic activities are: the location quotient method (Porter, 1990), the Ripley's K-method (Porter, 1990), and spatial autocorrelation (Porter, 2003).

This paper proposes an approach based on statistical methods for data analysis to identify spatial economic concentrations, by using economic and geographical data.

The spatial features used to develop the analysis are administrative territorial units at NUTS2 and NUTS3 (Nomenclature of Territorial Units for Statistics) levels.

A software module has been developed and implemented in order to manage the economic and spatial data, to perform the data analysis, and to present the results using a GIS (Geographical Information Systems).

The proposed solution has been tested for a case study applied for energy industry of Romania. In the case study Romania's administrative territorial units are used as spatial data and the following economic indicators: turnover, number of employees and number of companies for each activity as described in NACE rev. 2 (Statistical Classification of Economic Activities).

The structure of this paper is as follows. Section 2 sets out a theoretical framework and describes the methods used to measure the spatial concentration of economic activities. Section 3 presents the research methodology. Sections 4 and 5 introduce the Principal Component Analysis (PCA) and cluster analysis methods used to set up the data analysis. In the sixth section we tackle about the implementation of model proposed for our analysis, and the main research results. Finally, there is a statement resuming the conclusions.

2. Methods to Commeasure the Spatial Concentration of Economic Activities

In order to set up the analysis of spatial economic concentration, all regions of the target geographical space have been characterized by determining concentration, specialization and diversity of economic activities occurring inside each region.

Location quotient method has been used to determine the degree of concentration of one economic activity in a specific region.

In the literature (Andersen et al, 2006), two methods are available for computing location quotient: by using the degree of concentration of population by areas, or in relation with the average level of activities at national level.

In the first option, the location quotient allows measuring the geographic distribution of activities, in relation with the concentration of population as follows:

$$q_{ij} = \frac{p_{ij}}{P_i},$$

where, i is region, j is activity, p_{ij} is the share of region i in the activity j, at national level and P_i is the share of population from region i, in the total population.

The contribution of the region i at the activity j is computed as follows:

$$p_{ij} = \frac{x_{ij}}{\sum_{k=1}^{n} x_{kj}},$$

where n is the number of the regions and x_{ij} is the level of the activity j in the region i.

The region i is considered specialized in the activity j if the share of this region in the activity j is greater than the share of its population in the total population.

The second way to compute the location quotient is to divide the average level of the activity j from the region i to the national average level of this activity.

The formula of location quotient for this case is:

$$q_{ij} = \frac{\dfrac{x_{ij}}{\sum_{l=1}^{m} x_{il}}}{\dfrac{\sum_{k=1}^{n} x_{kj}}{\sum_{k=1}^{n}\sum_{l=1}^{m} x_{kl}}},$$

where, m is the number of activities.

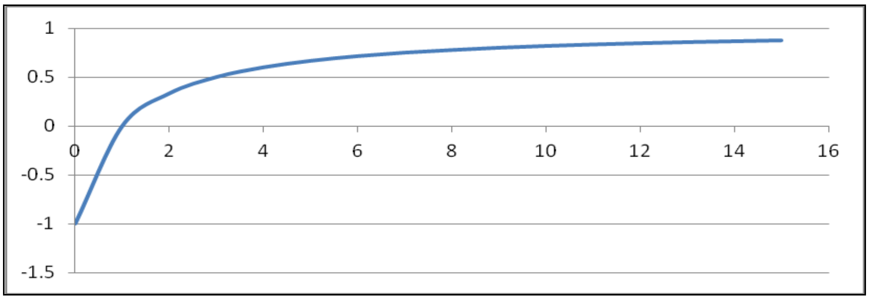
A value of q_{ij} less or equal with 1 suggests that the region *i* is not specialized in the activity *j*. A higher value of q_{ij} shows a region with a higher degree of specialisation.

The coefficient of location may attain high values if all the economic activity is concentrated within a single region. A normalized location coefficient can be calculated by using the formula:

$$qn_{ij} = \frac{q_{ij} - 1}{q_{ij} + 1}.$$

The values of the normalized location quotient vary in the range [-1,1). The relation between *q* and *qn* is illustrated in figure 1. In a region without any activity (*q* = 0) the *qn* value is -1. For the extreme value of *q* (*q*=1), the value of *qn* is 0. We can talk about specialization for values of the normalized location quotient greater than 0. Closer to 1 are the values of the index, the higher is the degree of specialization of the region.

Figure 1. The relationship between the location quotient and the normalized location quotient



The coefficient of specialization is determined by aggregation of location quotients for one activity, from a region. The index is computed at the region level and reflects the degree of specialization of a geographic region. The formula used to calculate the coefficient of specialization is the following: $qsp_i = \frac{1}{m}\sum_{j=1}^{m}\left|qn_{ij}\right|$.

The lower (close to 0) is the value of this index calculates for a region, a more faithful image the region can be considered for the distribution of the economic activities at national level. Greater values for the coefficient of specialization can be obtained from both small and high values of the normalized location quotients. Both the existence of one economic activity over the average level, and the lack of the economic activity define a specific profile for a region.

When the coefficient of specialization is closer to 0, the region does not have a specific profile.

The diversity index can be used to determine the diversity of the economic activities developed at the region level. This index is used to identify the level of homogeneity of one region, from economic point of view. The formula used to compute the diversity index is derived from the Shannon formula, used in the information theory and has the following form:

$$E_i = -\sum_{j=1}^{m} p_{ji} \ln\left(p_{ij}\right), \quad (5)$$

where, p_{ij} is the share of the activity j, developed in region i.
The index can be computed at the national level too.

$$E = -\sum_{j=1}^{m} P_j \ln\left(P_j\right),$$

where, P_j is the weight of the activity, at the national level.
The greater the values of diversity index are, the greater is the diversity of economic activities developed in the specific region.

3. Research Methodology

Spatiality influences the way in which an economic system works. It is a source of economic advantages or disadvantages such as high or low endowments of production factors. It also generates geographical advantages, like the easy, or difficult accessibility of an area, and a high or low endowment of raw materials. Space is also the source of advantages achieving from the cumulative nature of productive processes in space.

In order to analyse the spatial concentration of economic activity from one specific industry we are using two components of analysis: economic and geographic.

The geographic component of our analysis takes into account geographical proximity of territorial administrative units.
Territorial administrative units are represented in GIS using spatial data of polygon type. In this case, geographic parameters used in the analysis are the polygon coordinates of the centroid of the territorial administrative units. The centroid is determined by using the geographical coordinates of a plane.

Both economic and geographical data are used to perform the analysis in this case study: two geographical coordinates and three economic data sets. The number of economic data sets available is usually higher than the number geographic parameters whose number will always remain two. In this case, the geographical parameters will have lesser importance in the clustering process. Regardless of the number of indicators taken into consideration, the share of economic component of our analysis can be adjusted by replacing the indicators, with the first principal component and its weight, accordingly. From informational point of view, first principal component can replace the basic indicators because it represents their informational synthesis.

The economic component is composed of economic indicators like: productivity, employment, real wages, profits, turnover and gross investment, provided for one or more activities.

A **principal component analysis** (PCA) **method** has been adapted to develop data analysis (Jolliffe, 1986). Principal component analysis method seeks the axis which the cloud of points representing the instances, are closest to each others. The criterion is to have a variance of the projection as big as possible. On the other hand, PCA transforms the initial correlated variables in new uncorrelated variables to each other: the principal components. PCA method is useful to identify and preserves the informational structure of data.

Some steps have been completed to set up the model of analysis, as following:

1. A data based with economic indicators for each territorial administrative unit and for each economic activity, by using the NACE codes, has been create:

$$I_k = \begin{bmatrix} I_{11k} & \dots & I_{1qk} \\ \dots & I_{ijk} & \\ I_{n1k} & .. & I_{nqk} \end{bmatrix}, \text{k} = 1, \text{p},$$

where p is the number of activities, q is number of indicators, n is number of territorial administrative units, I_{ijk} is the value of indicator j for unit i and for activity k.

2. By using the values of economic indicators stored in the data base, the location quotients are computed, for any indicator (j), by using the following formula:

$$q_{ijk} = \frac{I_{ijk} \Big/ \sum_{l=1}^{p} I_{ijl}}{\sum_{u=1}^{n} I_{ujk} \Big/ \sum_{u=1}^{n}\sum_{l=1}^{p} I_{ujl}},$$

3. The normalized location quotients have been calculated by using the formula:

$$x_{ij} = \frac{q_{ij} - 1}{q_{ij} + 1}.$$

4. The C principal components matrix is achieved starting from the X matrix of normalized location quotients (or simply location quotients).

5. A cluster analysis method has been used in order to identify the most appropriate groupings and to design the clusters.

6. A top of the clusters, by using an average location quotient is generating. Only clusters with location quotient greater than 1 have been highlighted.

4. Principal Component Analysis

The principal components (Ruxanda, 2001) are created to satisfy the following conditions:

1. To be a linear combination of the model variables and to have the maximum variance among them;
2. To be uncorrelated among them.

The first condition ensures a maximum representativeness across the initial variables. The last condition ensures components based separation for the significant characteristics of data. Thus, the first principal component takes over the maximum value of the original variables variance. The second component being uncorrelated with the first one, it takes over the maximum value of the remaining variance, and so on. The initial variables are the location quotients, calculated based on the economic indicators. The number of variables is equal with the number of indicators.

The first principal component is calculated by using the formula:

$C_1 = a_{11}X_1 + \ ... + a_{j1}X_j + ... + a_{m1}X_m$,

where m is the number of variables, X_j, $j = 1,m$ are column vectors of the initial variables, a_j, $j = 1,m$ are the linear quotients used to obtain the principal component. The relation: $C_1 = a_{11}X_1 + \ ... + a_{j1}X_j + ... + a_{m1}X_m$ becomes $C_1 = X \cdot a_1$, where $a_1 = \begin{bmatrix} a_{11} \\ ... \\ a_{m1} \end{bmatrix}$, if it is written in a matrix form.

From a geometrical perspective, the problem assumes that the elements space is reorganised by identification of a new orthonormal axes system, according with the principal components. Thus, the C_1 vector contains the elements projection on an axis having as unit vector the a_1 vector ($a_1^t \cdot a_1 = 1$). The component variance is: $\frac{1}{n}\sum_{i=1}^{n} c_{i1}^2$, where n is the number of elements and c_{i1} is the projection of the i element on the axis.

According with the first condition used to determine the principal components; this value must be the maximum. The variation of the first principal component written in a matrix form is:

$$\frac{1}{n}\sum_{i=1}^{n} c_{i1}^2 = \frac{1}{n} C_1^t \cdot C_1 = \frac{1}{n}\left(Xa_1\right)^t Xa_1 = \frac{1}{n} a_1^t X^t X a_1 .$$

The conditions used to identify the first axis and, by default, for the first principal components are:

$$\begin{cases} \underset{a_1}{Max} \frac{1}{n}(a_1)^t X^t X a_1 \\ a_1^t a_1 = 1 \end{cases}$$

The problem is solved by using the multiplier Lagrange method, and results that the a_1 is the eigenvector of the $\frac{1}{n} X^t X$ matrix corresponding to the biggest eigenvalue, $\alpha_1 = \frac{1}{n} a_1^t X^t X a_1$. The others principal components are obtained in the same way by using, in addition, the condition for the orthogonal axis. Thus, for any component, *k*, the conditions are:

$$\begin{cases} \underset{a_k}{Max} \frac{1}{n} a_k^t X^t X a_k \\ a_k^t a_k = 1 \\ a_k^t a_j = 0, j = \overline{1, k-1} \end{cases}$$

The importance of the principal component is given by its variability, $\alpha_k = \frac{1}{n} a_k^t X^t X a_k$. The variants of the principal components are determined as eigenvalues of the correlation matrix. By all means, given up particular principal components, lead to a loss of information. But the amount of lost information is small because the information we gave up contains components with small variability. The loss of information due to decreased dimensionality is:

$$I = 1 - \frac{\sum_{j=1}^{k} \alpha_j}{m},$$

where k is the number of principal components taken into account, and α_j is the j eigenvalue (is the eigenvalue of j order).

The loss of information is irrelevant in this case because the indicators used to compute the level of concentration of one economic activity are correlated each with the others.

Industries can be aggregated in function of desired clusters. The relative importance of the economic component (compared to the geographic component) is determined by the number of economic indicators included in the analysis. If the number of indicators is big, the share of economic component is high compared to the geographic component. On the other hand, removing some economic indicators from the analysis leads to a loss of relevant information.

In these conditions, the solution proposed in the paper allows the user to choose the proportion between the economic and geographical components of the analysis. For an equal share of the two components, the first principal component method has to be used, with a weight equal to 0.5 for economic indicators, and weights equal to 0.25 for both centroid parameters.

5. Cluster Analysis

The typical goals of a cluster analysis include detecting representative cases from a large dataset to support data reduction, identifying natural clusters in a dataset to give insight into what cases are grouped together. Essentially, cluster analysis identifies clusters that exist in a given dataset (Han et al, 2006).

One advantage of cluster analysis is the multidimensional aspect. The input may contain several indicators for many linked activities.

A cluster is a collection of cases that are more similar to others, than cases from other clusters.

A set of clusters is considered to be of a high quality if the similarity between clusters is low and the similarity of cases within a cluster is high.

Most of the scientific papers published in cluster analysis research area, outline the possibility of using three types of algorithms: hierarchical, non–hierarchical and mixed algorithms.

In this paper, we proposed an improved version of the Bisecting K-Means algorithm. Bisecting K-Means is a mixture between hierarchical and K-means algorithms (Savaresi, Boley, 2001). This algorithm creates a hierarchy, in a top-down manner, as the Greedy technique. In each step, the cluster with the biggest intra-class variance is divided into two clusters, by using the traditional KMeans algorithm. The implementation used in this study is presented in the table 1.

Consider a set of n instances, measured on each of m attributes or variables. The $n \times m$ matrix of values will be denoted by X.

A cluster is stored in a list of integers, indexes of instances, with values from 1 to n. A partition is stored as a list of clusters (L).

The algorithm either provides a predetermined number of clusters (k), or uses a criterion that limits the number of clusters to a value that is considered optimum. The usage of this criterion distinguishes the implementation proposed in this paper by other implementation forms, available for Bisecting KMeans algorithm.

The choice is done by the *fullSplit* parameter that is of Boolean type. At each division step, the variance within the class decreases and the variance among classes are increased with the same value. The optimum criterion proposes the limitation to the partition having the maximum variance within class per cluster.

For any partition, Li:

$T = W_i + B_i$, i = 1,n-1,

where T is the total variance that remains constant, for all partitions, W_i is the within class variance and Bi is the variance among classes.
A partition is more coherent if the within class variance is bigger. Generally, the inter-class variance grows less in each division, but the coherent doesn't uniformly increase.

In the proposed solution we are looking for to identify the partition having the maxim of coherence. This corresponds to the maximum variance within a class divided by the number of the clusters from partition, $\overline{B}_k$:

$$\overline{B}_k = \underset{i}{Max}\frac{B_i}{i+1}\text{, for } i = 1,n\text{-}1.$$

Table 1. Bisecting K-Means

```
Procedure BisectingKMeans(X,C,n,m,k,fullSplit;L)
call Add(L,{1,2,...,n})
// In each step, one cluster is divided into two clusters
pastBetweenVar = 0
for i=1,k-1
    call SelectCluster(L,X,n,m;SplitCluster)
    call KMeans(SplitCluster,X,C,n,m
        ;LeftCluster,RightCluster)
    call DeltaW(SplitCluster,LeftCluster,RightCluster,
        X,n,m;d)
    currentBetweenVar = pastBetweenVar + d
    if currentAvgBVar/(i+1)≥pastAvgBVar/i||fullSplit=true
        call Add(L,LeftCluster)
        call Add(L,RightCluster)
    else
        break;
    endif
endfor
return
end
```

The meaning of the functions is the following:
SelectCluster – the cluster with the maximum variance is selected, from current partition;
KMeans – applies the classical KMeans algorithm to divide a cluster in two clusters;
DeltaW – computes the variance loss within a class at the current division;
Add – adds a cluster at partition.
First division of instances using the classic *KMeans* algorithms is done by applying the first principal component as follows: the instances with values below the average value are included in the first cluster, while the instances with values above the

average value will create the second cluster. In this way, the number of steps for splitting the cases into clusters decreases in a significant way.

But the goal of our research is to perform the cluster analysis on a territorial profile. So, we have to sort the clusters function of specific activity intensity in territorial administrative units. This classification can be done by calculating the location quotient (Ripley, 1981). As it has been presented in the Section 2, the location quotient is calculated as share of one activity developed in a territorial unit, function on its share at the entire level. A value for this quotient less or equal with 1 means that the region is not specialized in that activity field. Values greater than 1 means the region is specialized in that activity more or much more function of the value.

For identifying the clusters, we propose to use an average location coefficient calculated by using the *X* matrix, at the cluster level:

$$\overline{Q_k} = \frac{1}{n_k \cdot m} \sum_{i \in k}^{n_k} \sum_{j=1}^{m} x_{ij} ,$$

where k is the cluster and n_k is the number of the items from cluster.

5. Model Implementation and Main Results

The above model of analysis has been implemented as an *ArcMap* extension associated, with a map document containing Romania's counties (NUTS 3) and regions (NUTS2). Users can select the administrative unit level used to generate clusters, the clustering methods and the name of economic activity for which the analysis has to be performed, using a customized interface.

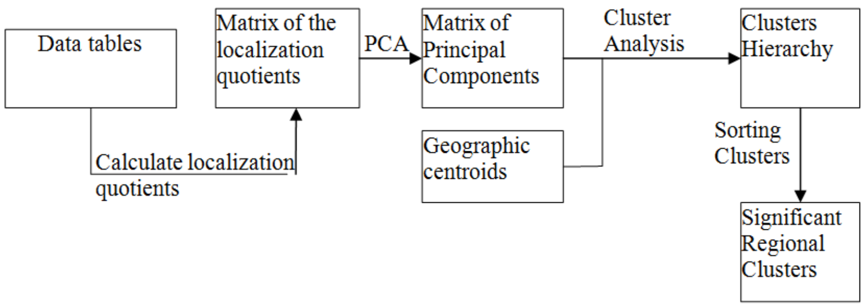
The results are presented in different ways: by using cartograms (Dârdală, Reveiu, 2012), in a graphic form and in a table form.

A database with economic indicators, aggregated based on the NACE code and on the location of companies, has been created by using the indicators collected from Romanian companies that have reported turnover and the number of employees, for year 2011.

Data is summarized by using principal components analysis, followed by the cluster analysis applied to one, two or more main components plus geographic centroid, in order to identify the groups of counties with the same profile.

In the last step we have to order the clusters by using an average location quotient. Only clusters with location quotient greater than 1 will be highlighted. The processing stream is illustrated in the figure 2.

Figure 2. Processing stream of model implementation



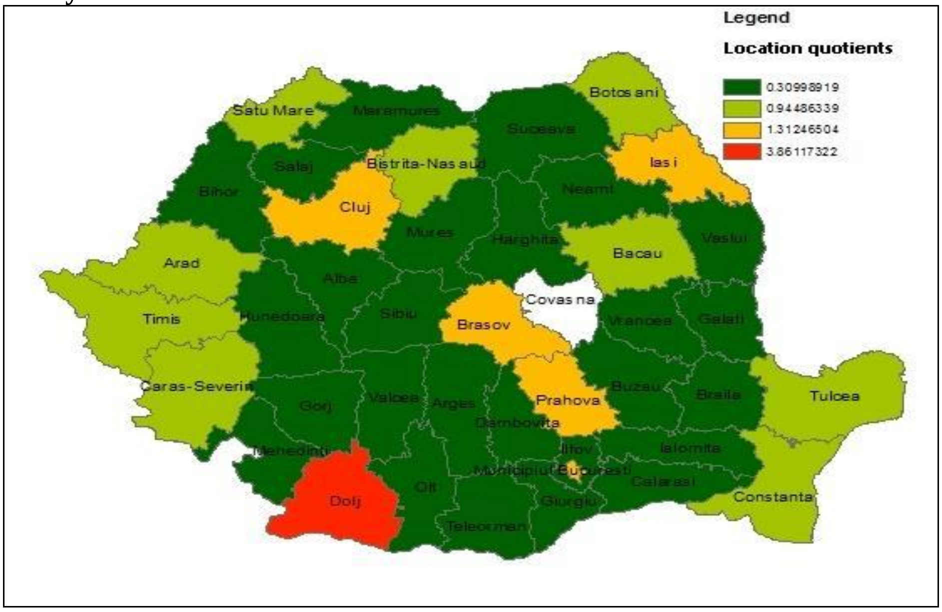
This section presents an analysis of the spatial distribution for the economic activity developed in production and supply of electric and thermal energy, gas, steam and air conditioning (NACE code 35) and its sub-activities.
The main indicators used are: the number of employees in the specified industry, the turnover of companies acting in the industry and the number of companies. For each of these indicators, the location quotients have been calculated. The classification algorithm used is a version of Bisecting KMeans, with improvements presented in the previous section.

To analyse the economic industry with NACE code 351, namely *Electric power generation, transmission and distribution* industry, we considered two scenarios:

1. The clustering is done by using the location quotients calculated based on the economic indicators, without taking into consideration the geographic component. The results are figured on the map with the territorial distribution of the economic activities, as is shown in figure 3. The algorithm of classification generates a division of Romania's counties in four groups. In the legend of the map, the average values per cluster, for the values normalized location quotients are displayed. The clusters are figured by using a colour ramp, and are ordered according with the average values of location quotients.

In this scenario only the economic indicators are included in the analysis. It doesn't take into consideration the spatial coordinators of the economic activities analysed. This explains a slight spatial dispersion of the elements component of each cluster.

Figure 3. Spatial distribution of the clusters identified in *Electricity, gas, steam and air conditioning supply* industries, by using only the economic component of analysis



The cluster with the highest economic activity has only one unit, namely Dolj County. This result could be explained by the natural resources available in this county: power plants burning coal extracted in this area. The specific profile of each cluster can be determined through detailed analysis done for each sub-activity. In the table 2, normalized location quotients are presented for the following sub-activities: *production* (NACE 3511), *transmission* (NACE 3512), *distribution* (NACE 3513) and *commercialization* (NACE 3514).

In the table 2, the results for the first two clusters have been presented, with the average value of normalized location quotient greater than 1. We see that in the second cluster (with the average value of 0.94486339), the administrative units with activities of transmission, distribution and commercialization of electricity.

Table 2. Structure of clusters with an intense activity in *Electric power generation, transmission and distribution* industry, function on their sub-activities

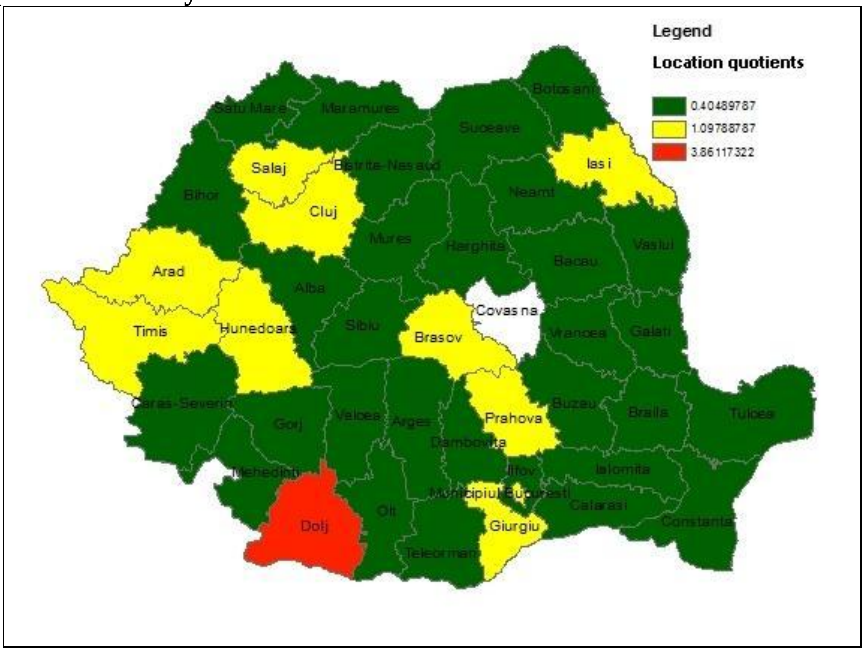
Unit\Caen	NACE=3511 Production of electricity	NACE=3512 Transmission of electricity	NACE=3513 Distribution of electricity	NACE=3514 Trade of electricity
	Cluster 1			
Dolj	7.7	0	4.138	5.385
	Cluster 2			

Brasov	0.561	2.427	2.4	2.074
Iasi	0.263	0	4.019	1.904
Bucharest	1.277	2.434	1.081	7.803
Cluj	0.448	0	2.16	1.917
Prahova	0.757	2.619	2.853	0.441

2. Identifying the spatial clusters, by using the first principal component and the centroid values.

By taking into consideration the centroid values, the resulted clusters are strongly influenced by the geographic proximity among administrative territorial units. The resulted clusters are more consistent, from territorial distribution point of view, and are spread into more compact areas. The weight of principal component can vary function on the importance of economic factor. The territorial distribution map presented in figure 4 uses a share of 60% for the principal component.

Figure 4. **Concentration areas in *Electric power generation, transmission and distribution* industries, determined by using both economic and geographic components of analysis**



The production of gas and distribution of gaseous fuels through pipelines (NACE code 352), is concentrated in few counties. The most stable partition has two clusters, one cluster with only one county (Mures), and the second cluster with the remaining counties. However, for a more detailed highlighting of the distribution of this economic activity in space, we have built a partition with four clusters, as shown in table 3.

Table 3. Structure of clusters having intense activity, function on their sub-activities for *Manufacture of gas distribution of gaseous fuels through mains* industries

Unit\Caen	NACE=3521 Manufacture of gas	NACE=3522 Distribution of gaseous fuels through mains	NACE=3523 Trade of gas through mains
	Cluster 1		
Mures	0	11.061	35.526
	Cluster 2		
Botosani	11.965	0	0
Giurgiu	10.528	0	0
	Cluster 3		
Tulcea	0	2.316	0
Iasi	0	1.862	0

The first three clusters have the average value of location quotient greater than 1, and the results are presented in figure 5. The analysis of internal structure of this activity points out two regions with strong production activity, in the second cluster, and a significant presence of distribution activity, in the third cluster. Mures County is distinguished in the distribution and trade of gas activities, being in the neighbourhood of the main gas tank operated in Romania.

The use of geographic component in our analysis highlights a more intensively presence of this activity in the East and Southeast regions of Romania, as presented in figure 6. This phenomenon could be explained by the presence of natural gas distribution and commercialization by pipeline activities in these regions, in connection with the gas imports from Russian Federation.

Figure 5. Concentration areas in *Manufacture of gas; distribution of gaseous fuels through mains* industries, computed from economic indicators only

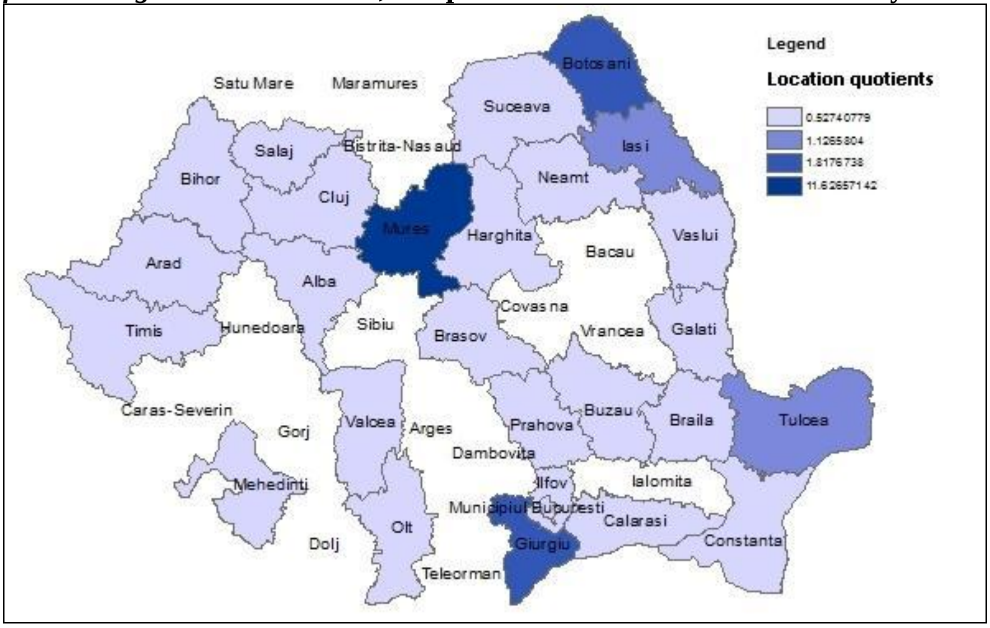
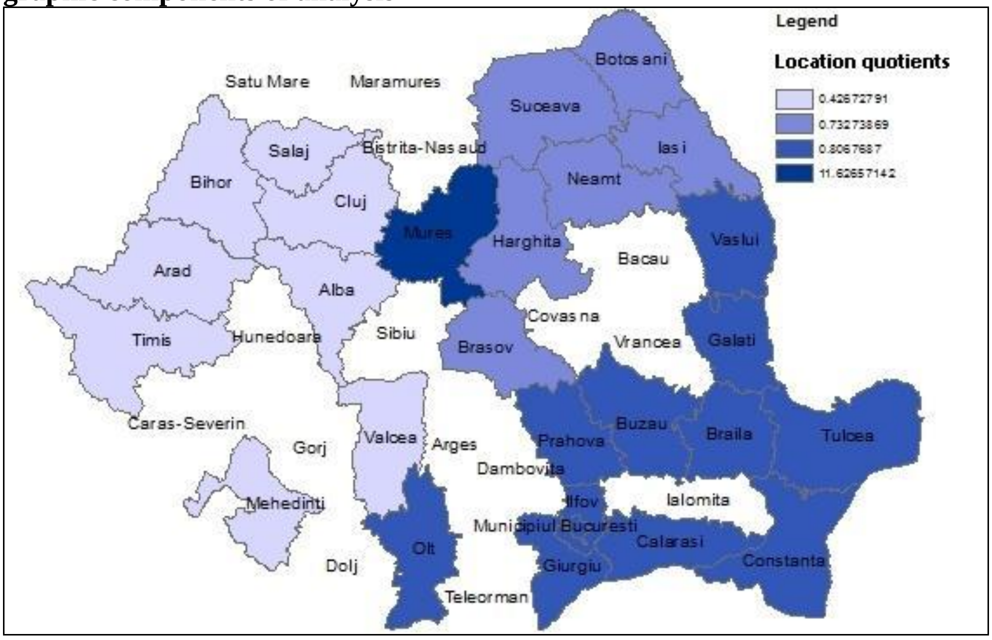


Figure 6. The concentration areas in *Manufacture of gas, distribution of gaseous fuels through mains* industries, determined by both economic and geographic components of analysis



6. Conclusions

The method of analysis proposed in this paper succeeded to identify clusters of territorial administrative units having similar features from purely economic point of view, or by taking into consideration both the economic indicators and the proximity of the administrative units.

It is based on principal component analysis and cluster analysis methods and represents an improved version of the Bisecting K-Means algorithm. The proposed algorithm, which is a mixture between hierarchical and K-means algorithms, contains a limiting criterion to an optimal number of clusters. Partitions created in this way are more stable and have more significance. This algorithm creates a similar top-down hierarchy as the Greedy technique.

The results of this analysis are useful for policy makers, entrepreneurs and economic analysts to investigate and evaluate the spatial distribution of economic activities, for decision making.

The results of this research are important for the next stages of our research aiming to design a sustainable energy model for Romania.

Acknowledgment

This work was supported by the Swiss Enlargement Contribution in the framework of the Romanian-Swiss Research Programme (Grant IZERZ0_142217).

REFERENCES

[1] **Andersen, T., Bjerre, M., Wise Hansson, E. (2006), T*he Cluster Benchmarking Project***; Nordic Innovation Centre, FORA;

[2] **Anselin, L., Florax, R. J. G. M., Rey, S. J. (2004), *Advances in Spatial Econometrics. Methodology, Tools and Applications* ;** *Springer;*

[3] **Arbia, G. (2001), *The Role of Spatial Effects in the Empirical Analysis of Regional Concentration***; Journal of Geographical Systems, Vol. 3, No.3, pp. 271-281;

[4] **Dârdală, M., Reveiu, A. (2012)**, ***GIS as a Supporting Tool in Analysis the Spatial Distribution of Economic Activities***; 3rd World Conference on Information Technology, 14-17 November, 2012, Barcelona, Spain;

[5] **Delgado, M., Porter, M.E., Stern, S. (2010)**, ***Clusters and Entrepreneurship***; Journal of Economic Geography, Vol. 10, issue 4;

[6] **Diday, E., Pouget, J., Lemaire, J., Testu, F. (1985), *Elements d'analyse de donnee***, Dunod, Paris;

[7] **Fischer, M., Getis, A. (2010)**, ***Handbook of Applied Spatial Analysis***; *Springer-Verlag Berlin Heidelberg*;

[8] **Han, J., Kamber, M., Kauffman, M. (2006),** ***Data Mining: Concepts and Techniques***, 2nd Edition, *Elsevier*;
[9] **Haining, R. (2005**), ***Spatial Data Analysis. Theory and Practice*; ** Cambridge University Press;
[10] **Jolliffe, I. T. (1986),** ***Principal Component Analysis***. *Springer-Verlag;*
[11] **McCann, P. (2001),** ***Urban and Regional Economics***; *Oxford University Press;*
[12] **Murtagh, F., Heck, A. (1987),** ***Multivariate Data analysis***; Dordrecht, Holland;
[13] **Pavlinek, P. (2005),** ***Transformation of the Central and East European Passenger Car Industry: Selective Peripheral Integration through Foreign Direct Investment***, in: Turnock, D. (ed.), Foreign Direct Investment and Regional Development in East Central Europe and the Former Soviet Union, *Ashgate Publishing*;
[14] **Porter, M.E. (1990),** ***The Competitive Advantage of Nations***; Free Press;
[15] **Porter, M.E. (1998),** ***Clusters and the New Economics of Competition***; Harvard Business Review;
[16] **Porter, M.E. (2003),** ***The Economic Performance of Regions***; Harvard Business School;
[17] **Ripley, B. D. (1981),** ***Spatial Statistics***. *Wiley*, New York;
[18] **Ruxanda, Gh. (2001),** ***Analiza Datelor*** ; *ASE Publishing, Bucharest;*
[19] **Savaresi, S. M., Boley, D.L. (2001),** ***On the Performance of Bisecting K-means and PDDP***; Proceedings of the First SIAM International Conference on Data Mining (ICDM-2001);
[20] **Smeureanu, I., Ruxanda, Gh. (2013),** ***Considerations on the Stochastic Approach in E conomics***; Amfiteatru Economic, Vol. XV, No. 34.