
Learning Regions Identification by Unsupervised Methods

VERONIKA HÁJKOVÁ¹, PETR HÁJEK²

¹*Institute of Public Administration and Law*, ²*Institute of System Engineering and Informatics*
Faculty of Economics and Administration, University of Pardubice

Studentská 84, Pardubice

Czech Republic

veronika.hajkova@upce.cz, petr.hajek@upce.cz

Abstract

The paper discusses the importance of knowledge in regional development. The basic notions of learning regions are presented. The input variables are proposed for the modelling of NUTS 2 regions in order to identify learning regions. The identification of the learning regions is realized by unsupervised methods. Data are analyzed by the model merging neural networks and cluster analysis algorithm with the aim of data dimension reduction and, moreover, the model makes it possible to visualize regions in a topological map. The results show on the membership of regions to learning regions.

Key words: Learning regions, Identification, Neural networks.

JEL Classification: C45, R11, O32.

1 Introduction

The concept of regional policy, based on endogenous growth theory and linear model of innovation, lies primarily in the growth of public expenditure on R&D and investment in education [1]. By contrast, institutional approaches are primarily concerned with the institutional aspects of the process of learning and making innovations. Learning ability and innovations making are considered key factors of the regional development in institutional economics. The basis of these concepts lies in the observation that innovations do not arise in isolation of one company, but the potential of their creation is related to the process of learning determined with the relationship of the company and its environment [2]. The environment is considered as a network of relationships among firms and among firms and institutions, as well as a general framework for company operations, i.e. the institutional structure, social values and culture of political and economic relationship between the state and the region in which the firm is embedded. Thus, internal organization of firms, their rooting in the network of formal and informal relationships among themselves as well as the existence of supporting institutions, and the overall socio-cultural environment of the region are important factors for the innovation potential and the learning capacity of firms. The complex defined this way is known as a regional innovation system (RIS) or learning region [3]. Previous studies have been focused on identifying RIS primarily by economic, R&D and educational indicators. However, other factors such as social and cultural capital and infrastructure were not considered. The values of selected variables represented inputs into the models based on statistical methods in previous studies. These methods, however, are constrained with many requirements which are difficult to meet in praxis.

The aim of this study is a model proposal for the identification of learning regions. The selected

input variables characterizing learning regions represent the inputs of the proposed model. The variables concern following characteristics of regions: economy, R&D, education, social and cultural capital and infrastructure. EU regions at NUTS 2 level will be analyzed using suitable unsupervised methods in order to obtain clusters of similar regions. Clusters will be ranked according to the membership to learning regions.

The work is structured as follows. First learning region is defined. On the basis of its characteristics the input variables are designed for learning regions' identification. In Chapter 3 an overview of previous studies in the identification of RIS and learning regions is provided. Furthermore, the methods used for the modeling are characterized. Finally, the experimental results are presented and analyzed.

2 Learning Regions

The competitiveness of modern organization is based on knowledge. The concept of learning region, in this context, shows the way how to mobilize and then use the potentials of all the regional actors for regional development "bottom-up". The model of learning region assumes that regional actors will organize themselves autonomously, and that they take the integral responsibility for regional development [4]. In the field of regional development, tools and policies are searched to ensure economic growth and development. In this context, concepts are discussed such as regional clusters, regional innovation systems, regional innovation networks and learning regions which are attributes of successful development of a number of economies. These concepts can be represented using hierarchical structure as shown in Table 1.

Table 1: Stages of development of regional cooperation

| | |
|-----------------------------|---|
| Regional cluster | The concentration of interconnected companies of the same or related industries in a small geographic area. |
| Regional innovation network | Widely organized cooperation (on the basis of agreements) between firms, stimulated by trust, norms and conventions. |
| Regional innovation system | Cooperation between companies and institutions in the development and dissemination of knowledge in innovation processes. |
| Learning region | Widely organized cooperation of a broader range of civic organizations, companies, institutions and public authorities, which are embedded in social and regional structures. |

Source: [5]

The most relevant characteristics of learning regions can be defined as follows [6]:

- Existence of the higher number of regional actors (municipalities, towns and cities, enterprises, firms, NGOs etc.). Their interactions can facilitate the exchange of information and new ideas.
- Existence of consulting, R&D institutions and transfer centres that cooperate with the other regional actors.
- Regional culture and institutions. This category is the most problematic one since it is hardly possible to stipulate normatively, what should be the character of the culture and the institutions in the region in order to maximize its capacity to learn and to innovate.

The learning regions function as collectors and repositories of knowledge and ideas, and provide the underlying environment or infrastructure which facilitates the flow of knowledge, ideas and learning [7]. Learning regions are increasingly important sources of innovation and economic

growth, and are vehicles for globalization [8]. Key processes of the learning regions can be divided into 3 categories [9]:

- The generation and improvement of the level of know-how at the individual, organisational and regional level.
- The cooperation between regional subjects and diffusion of human capital and know-how in organizations and between organizations.
- The transfer of human capital and the new know-how into practice. In terms of the regional output or economy it means growth of GDP and employment, a higher quality of services and welfare in the region.

Participants are included in interactive learning in learning regions. Actors closely cooperate at the institutional level in the preparation and implementation of regional innovation strategies. Learning regions are constituted as a combination of collective political decisions and local bottom-up activities. The creation of regional development coalitions has strategic importance. Regional development coalitions are long-term models of multilateral cooperation to promote innovation including partners such as local trade unions, economic chambers, venture capital, educational organizations, research institutes and local and regional authorities.

Complex relations among subjects within regions result in the fact that learning regions were largely based on case studies so far [9,10]. Another problem in identifying learning regions is related to different approaches to their definition. One is represented by [11] where only Silicon Valley, Emilia-Romagna and Baden-Württemberg are regarded as actual learning regions. Similarly, in [12] it is indicated that the characteristics of learning region are not met in any region. The alternative approach presented in [13] points to the fact that all regions have some elements of learning regions, they dispose of learning systems respectively. Clearly, it depends on the definition of conditions the region has to meet to be considered as a learning region. Only the synergic effect, resulting from the compliance of all defined requirements, makes growing competitiveness, social inclusion and economic growth possible in these regions.

In [14] four characteristics are defined which are typical for learning regions and, therefore, they can be used to differentiate them from other regions. These characteristics include:

1. Sustainable economic growth coupled with an increase in employment-intensive skills;
2. Social inclusion and social capital formation;
3. Role of different educational strategies for promoting learning regions;
4. An integrated approach to achieve "good governance".

Innovation can be quantified to some extent by expenditure on R&D (both public and private) and the proportion of employees in R&D. Number of computers and patents serve as a measure of technological development. Other factors include the existence of regional innovation policy and instruments for its implementation. Proportion of population with tertiary education is another important variable while technical skills are usually distinguished from the academic. Similarly, it is possible to measure the number of research teams in the region. It is also important to take into account qualitative parameters such as the readiness of people to a change and further education. Social capital can be measured in several ways, e.g. by the crime rate or the participation of people in voluntary associations. Cultural capital is measured by the number of visits to libraries, museums, computer literacy, etc. It is also important to take into account whether the region is attractive measured e.g. by population migration. This way, the quality of

life is taken into account in the region. The relations with other regions can be measured by the capacity of infrastructure. Based on these facts we design the input variables for the identification of learning regions as indicated in Table 2.

Table 2: The design of input variables for learning regions identification

| Economy | | Education | |
|---------|----------------------------------|------------------------------|--|
| x1 | Regional GDP per capita | x9 | Population with secondary education |
| x2 | Real growth rate of regional GDP | x10 | Population with tertiary education |
| x3 | Employment rate | x11 | Participation in life-long learning |
| x4 | Long-term unemployment share | Social and cultural capital | |
| R&D | | x12 | Population change |
| x5 | Patent applications per capita | x13 | Regular internet users |
| x6 | Public R&D expenditure | x14 | Crude death rate caused by assault |
| x7 | Private R&D expenditure | Relations with other regions | |
| x8 | R&D employment | x15 | Motorways [km] / area [km ²] |

3 Identification of Learning Regions in the EU

In the literature, it is possible to find several studies analyzing RIS in European regions [16]. However, indicators considering cultural and social environment were not included in these studies. Despite of this fact, we will present the overview of the mentioned studies as they preserve most of the information useful for learning regions identification.

There have been two approaches for obtaining a RIS typology [16]. The first one deals with authors who used case studies in order to test previous conceptual works. Cooke [17] combined three types of RIS governance (grassroots, network and interventionist) with other three dimensions of entrepreneurial innovation (localist, interactive and globalised). A typology of 9 groups of RISs has been obtained. Asheim [18] distinguishes between three types of RISs: territorially embedded, regionally networked and regionalised nationals. Tödting and Trippel [19] classify the regions in peripheral, mature industrial and metropolitan regions. The second way to create RIS taxonomies is realized using statistical analysis for a set of regions.

Within the EUROTITE project [20] a set of indicators for learning regions' analysis at NUTS 2 level. This set involves following areas: science (number of publications, public R&D expenditures), technology (patents, private R&D expenditures, share of researchers), education (number and share of students, tertiary students, life-long learning) and performance (GDP, unemployment, long-term unemployment). Moreover, specialization and performance of selected sectors were measured to provide additional information. This report presents the typology of European regions according to their involvement in the knowledge economy. The analysis was, however, not realized for all European regions. For each area (science, technology, etc.) regional profiles were found. Finally, the correlations between these areas were studied and the results show that there can be recognized following regional profiles concerning knowledge economy: Metropolitan regions, North high-tech regions, North scientific regions, British services profile, German high industrial profile, Secondary metropolitan profile, North industrial regions, North Italian and Spanish Industrial Regions and French agro-industrial profile.

Further, Clarysse and Mulder [21] found 6 groups of EU regions considering their GDP, unemployment, R&D expenditures and patents. Similar variables were studied also by [22] with similar results (6 groups – very strong position in knowledge services, ..., staying behind). In [23] 5 types of regions were discovered based on their innovation potential (lack of capacity, average capacity, rich innovation, rich R&D and knowledge centres). In [24], indicators from science and education were used for a hierarchical cluster analysis. The results showed that there are 12

groups of regions according to innovation performance in the EU (NUTS 1 and NUTS 2). A large set of 29 variables (including national environment, regional environment, innovative companies, universities, public administration and demand) was used by [25] resulting in 10 groups of regions.

In [26], the authors studied new EU member states using 25 variables (5 areas – knowledge creation, knowledge absorption, diffusion of knowledge, demand of knowledge and governance). The results of factorial analysis showed 5 specific groups, i.e. capitals, with tertiary growth potential, qualified manufacturing platforms, with industrial challenges, agricultural laggards. New member states were also studied by [16]. Patents, R&D expenditure, employment, education, and economic performance were included for the analysis. The features of the three groups were summarised in the following titles: Regions with a weak economic and technological performance, Restructuring industrial regions with strong weaknesses, Capital-regions specialised in high value-added services. An extension for EU-25 was published in [27]. For the whole EU, 7 types of regions were recognized including Restructuring industrial regions with strong weaknesses, Regions with a weak economic and technological development, Regions with average economic and technological performance, Advanced regions with a certain industrial specialisation, Innovative regions with a high level of economic and technological development, Capital-regions with a certain specialisation in high value-added services, and Innovative capital-regions specialised in high value-added services.

At the level of NUTS 3 regions different indicators are monitored in different EU countries. So, the innovative potential of a region, e.g. [28], and learning regions (see Fig. 1) [29] were analyzed at the national level in the literature.

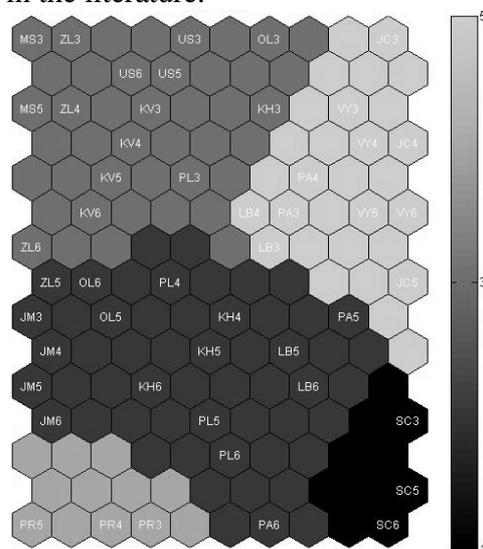


Fig. 1: Learning regions at NUTS 3 level

Legend: PR is Praha, SC is Středočeský, JC is Jihočeský, PL is Plzeňský, KV is Karlovarský, US is Ústecký, LB is Liberecký, KH is Královéhradecký, PA is Pardubický, VY is Vysočina, JM is Jihomoravský, OL is Olomoucký, ZL is Zlínský, MS is Moravskoslezský, and 3,4,5,6 are years 2003,2004,2005,2006.

4 Modelling Methods

In prior studies, factor analysis or cluster analysis were applied for identification of regional typologies (concerning knowledge society). However, these traditional statistical methods are capable to find only linear relations among variables, and together with the influence of multicollinearity or outlying objects the results are not fully reliable. Then when using statistical methods, it is recommendable to apply factor analysis (FA) or principal component analysis (PCA) first, and only then to apply cluster analysis. This way the results are not affected by colinearity. There is, however, still a loss of information (variance) when using FA or PCA. However, the resulting variables are usually easy to interpret based on factor or component loadings. As an example, both the FA and the PCA were applied on our dataset with the aim of finding 2 factors (components) for the visualization in 2D space. Only 50% of the data variance was explained when using these statistical methods. The consequent use of cluster analysis would lead to biased results.

Therefore, we propose to use a combination of neural networks and cluster analysis making it possible to use all variables as they are and, at the same time, to find reliable clusters not affected by outlying objects in 2D space. Economic data are usually in non-linear relations. Thus, it is suitable to realize such a model making it possible to involve these relations and, at the same time, enabling easy interpretation of gained results. This is possible to realize when using unsupervised neural networks.

Kohonen's self-organizing maps (SOMs) [15] are such models of neural networks which utilize competitive learning strategy. The output neurons of the SOM compete for the activity. The SOM is based on unsupervised learning. It is a two-layer neural network with a full connection between layers. The input layer is represented by n neurons serving for the distribution of input values x_i , $i=1,2, \dots, n$. The neurons in the second (competitive) layer are so-called representatives and they are organized into a topological structure (mostly a 2D grid). It determines which neurons neighbors with each other in the SOM. Objects are surrounded by similar objects in the grid but such objects are not always next to each other. Sometimes a cluster of similar objects is divided. Then it is necessary to modify learning parameters of the SOM to achieve good results. Using the SOM certain shortcomings of cluster analysis (e.g. impact of outlying objects) are eliminated. The results of the SOM can be easily visualized. Using the SOM one can discover the structure in the data. It is necessary to apply a clustering algorithm on the adapted SOM in order to find clusters. Then the process of data clustering is realized in two levels. The n objects are reduced to m representatives using the SOM in the first level, while the m representatives are clustered into c clusters in the second level. This way reduction in the computational cost of the process is accomplished. For these reasons, it is suitable to use a combination of the SOM and k-means learning algorithm for the identification of learning regions.

5 Experimental Results

The input data for the modelling are represented by values of input variables x_1, x_2, \dots, x_{15} for NUTS 2 regions of the EU25. Data on 265 regions have been obtained from 2003 to 2006. Further, only the results for the years 2003 and 2006 will be presented to compare results over time. Distances among representatives can be visualized using the map in Fig. 2. Similar regions on the map are located close to each other. Moreover, the optimal number of clusters was determined, that is the number at which the highest quality of clustering is achieved (measured by the Dunn index of clustering quality). In terms of clustering quality, the number of clusters of ten is optimal for the year 2003; while it is nine for 2006, see Fig. 2 and Fig. 3. The interpretation of clusters is possible based on the values of input variables for individual representatives on the

map (grid). These are presented in Fig. 5 and Fig. 6 (Appendix), where the dark colours indicate low values and light colours indicate high values of input variables. Representatives on the map of clusters (e.g. Fig. 3) are located on the same positions in Fig. 5. The whole clusters of regions can be characterized based on the values of representatives.

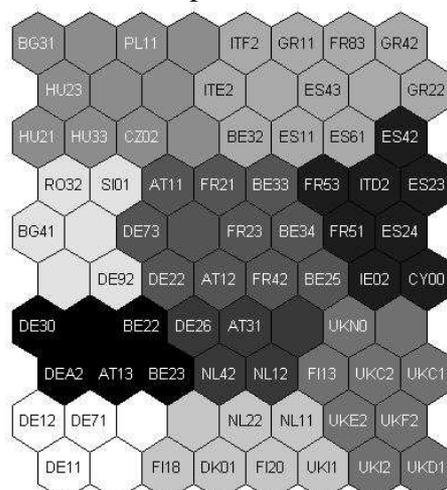


Fig. 3: Clustering of regions using k-means algorithm in 2003

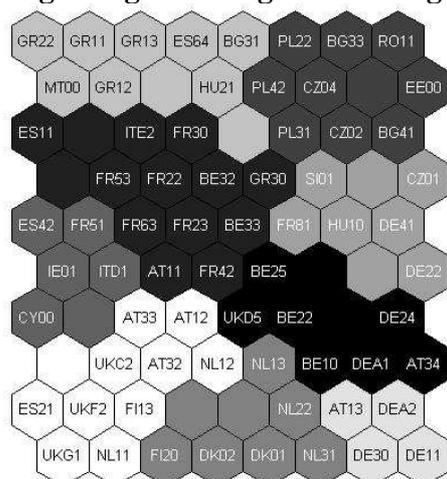


Fig. 4: Clustering of regions using k-means algorithm in 2006

The results are similar in the monitored period. It is evident from Fig. 2 and Fig. 3 that maps are rotated, i.e. regions located on the right of the map in 2003 are located on the left in 2006 and vice versa. This results from the fact that each map was trained individually. However, this rotation does not affect results. Still, similar regions are located close to each other. There were only partial changes in the map structure. The economic performance of regions converged over the period. In the area of R&D, especially advanced regions of Central and Eastern Europe (SI01, CZ01, FR81, ...) converged to the most advanced regions by the year 2006. This is due to high public spending in R&D, accompanied by a rise in private spending and a high number of graduates in tertiary education.

Clusters can be ordered based on the interpretation of the values in Fig. 5 and Fig. 6. They are ordered from those regions which are the least similar to the concept of learning region defined in

chapter 2 to those which are the most similar to this concept. Following description results from the year 2006 (in comparison to 2003).

Cluster 9 (PL22, BG33, RO11, ...) represents the undeveloped regions of the new Member States. The only positive factors were high growth in GDP and a high proportion of graduates in tertiary education in 2006. Since 2003 moderate employment growth occurred in these regions. Regions remain backward in particular in the areas of economy, R&D and social and cultural capital. Cluster 8 (GR22, GR11, GR13, ...) represents the backward regions of Southeastern Europe. A high GDP growth occurred in 2003 in these regions but, at the same time, low employment rate and high long-term unemployment rate are typical for these regions. They are strongly retarded in the area of R&D and education. They also become less attractive to residents.

Cluster 7 (ES11, ITE2, FR30, ...) represents the developed regions of Southern Europe. These regions have a high GDP per capita. However, they are average in R&D and education. Cluster 6 (ES42, FR51, IE01, ...) represents the developed regions of France, Spain and Italy especially. By 2006, there was an economic growth in the regions. The regions have moderate R&D and education (they managed to increase the number of patents by 2006). Their competitive advantage lies in the high quality of life. As a result, these regions are very attractive for people. Cluster 5 (SI01, CZ01, FR81, ...) represents the advanced regions of Central and Eastern Europe. Rapid economic growth is typical for these regions, and they are investing in R&D and education. They remain still less attractive for the people. Cluster 4 (AT33, AT12, AT32, ...) represents the regions of Austria and Great Britain especially. They are economically very strong, but only average in R&D with emphasis on secondary and long-life education.

Cluster 3 (BE25, UKD5, BE22, ...) represents the advanced industrial regions of Germany, Belgium and Great Britain. A high number of patents and a growing number of researchers and tertiary education are typical for this cluster. Moreover, they are equipped with developed social and cultural capital and technical infrastructure. Cluster 2 (NL13, NL22, FI20, ...) represents high-tech regions of Netherlands, Finland and Denmark. They are economically prosperous, with a high proportion of investment in R&D and with a high proportion of researchers. Cluster 1 (AT13, DEA2, DE30, ...) represents the regions which are closest to the characteristics of learning regions. They are either the metropolitan regions (Vienna, Berlin, Hamburg, etc.) or high-tech industrial regions (Karlsruhe, Freiburg, Stuttgart, etc.), see Fig. 4. They are economically highly developed with an emphasis on R&D and education. A high public and private investment in R&D are typical for these regions but, at the same time, they also produce a large number of patents. Moreover, they are also equipped with the developed social and cultural capital and technical infrastructure.

6 Conclusion

In our paper we discussed the importance of learning regions and the methods for their identification. Following a design of input variables the modeling was realized using neural networks and cluster analysis algorithm. The results show that the concept of learning region, as defined in the paper, is realized in the advanced regions of Germany, Netherlands, Austria and France.

When compared to prior studies, the similar results as in this work were obtained also by [20] in particular. The same number of 9 clusters of regions was found. In addition, the profiles of regions in [20] are very similar to those in this study despite the fact that the analysis in [20] was realized by factor analysis and cluster analysis. Moreover, slightly different input variables were used in our study. We included additional input variables (social and cultural capital, relations

-
- [3] DOLOREUX, D.: What we Should Know about Regional Systems of Innovation. In: *Technology in Society*. No. 24 (2002), pp. 243 – 263.
- [4] JEŽEK, J.: Učení se, Sítě a Samoregulace jako Základní Stavební Kameny Učících se Regionů. In: *Národní a Regionálna Ekonomika VI*. pp. 140 – 144.
- [5] SKOKAN, K.: Systémy Inovací v Regionálním Rozvoji. In: *Ekonomická revue*. Vol. 8, No.4 (2005), pp. 12 – 25.
- [6] KERN, J., MALINOVSKÝ, J., SUCHÁČEK, J.: Learning Regions: Towards a Container Conception. In: *Learning Regions in Theory and Practice* (Eds. Kern, J., Malinovský, J., Sucháček, J.). (2007), pp. 11 – 19.
- [7] FLORIDA, R.: Calibrating the Learning Region. In: *Local and Regional Systems of Innovation* (Eds. De la Mothe, J., Paquet, G.). (1998), pp. 19 – 28.
- [8] FLORIDA, R.: Toward the Learning Region. In: *Futures*. Vol. 27, No. 5 (1995), pp. 527 – 536.
- [9] KULHÁNEK, L.: Theory of Learning Regions from an Subjective View. In: *Learning Regions in Theory and Practice* (Eds. Kern, J., Malinovský, J., Sucháček, J.). (2007), pp. 47 – 56.
- [10] NIJKAMP, P., VAN GEENHUIZEN, M.: Lessons from Learning Regions: Policymaking in an Evolutionary Context. In: *Research Memorandum*. (2002).
- [11] COOKE, P., MORGAN, K.: *The associational economy*. Oxford: Oxford University Press, 1998.
- [12] KEATING, J., BADENHORST, A., et al. *Victoria as a learning region. Background Report*, 2002.
- [13] DOLOREUX, D., PARTO, S.: *Regional innovation systems: A critical review*, 2004.
- [14] TOLAND, J., YOONG, P.: Learning Regions in New Zealand: The Role of ICT. In: *International Journal of Education and Development using ICT*. Vol.1, No.4 (2005).
- [15] KOHONEN, T.: *Self-organizing maps*. New York: Springer-Verlag, 2001.
- [16] AGUADO, R., GIBAJA, J.J., et al.: Regional Innovation Systems in EU-10: A Typology and Policy Recommendations. In: *Proc of the 4th Globelics Conference at Mexico City*. (Sep., 2008), pp. 1 – 23.
- [17] COOKE, P.: Introduction: Origins of the Concept. In: *Regional Innovation Systems. The role of governances in a globalized world* (eds. Braczyk, H.J. et al.). (1998), pp. 20 – 25.
- [18] ASHEIM, B.: Sistemas Regionales de Innovación y bases de Conocimiento Diferenciadas: Un Marco Teórico Analítico. In: *Sistemas Regionales de Innovación: Nuevas Formas de Análisis y Medición* (eds. Buesa, M., Heijs, J.). (2007), pp. 65 – 89.
- [19] TÖDTLING, F., TRIPPL, M.: One Size fits All? Towards a Differentiated Regional Innovation Policy Approach. In: *Research Policy*. No. 34 (2002), pp. 1203 – 1219.
- [20] CARRINCAZEAUX, Ch., GASCHET, F., BECUE, M.: Typology of European regions – Final report. FP6 Integrated Project Eurodite: Regional Trajectories To the Knowledge Economy: A Dynamic Model, 2003.

-
- [21] CLARYSSE, B., MUL DUR, U. : Regional Cohesion in Europe? An Analysis of how EU Public RTD Support Influences the Techno-economic Regional Landscape. European Commission, Directorate General for Science, Research and Development. In: *Working Papers*. No. 1 (Jan., 1999).
- [22] BRUIJN, P., LAGENDIJK, A.: Regional Innovation Systems in the Lisbon Strategy. In: *European Planning Studies*. Vol. 13, No. 8 (2005), pp. 1153 – 1172.
- [23] ECOTEC: The Territorial Impact of EU Research and Development Policies. ESPON 2.1.2, 2005.
- [24] HOLLANDERS, H.: 2006 European Regional Innovation Scoreboard (2006 RIS). European Trend Chart on Innovation, 2007
- [25] MARTÍNEZ PELLITERO, M.: Los sistemas regionales de innovación en Europa: tipología y eficiencia. In: *Sistemas Regionales de Innovación: Nuevas Formas de Análisis y Medición* (eds. Buesa, M., Heijs, J.). (2007), pp. 215 – 256.
- [26] MULLER, E., NAUWELAERS, C.: Enlarging the ERA: Identifying Priorities for Regional Policy focusing on Research and Technological Development in the New Members States and Candidate Countries. In: *Final report COP6-CT.2004. 00001*. (2005).
- [27] NAVARRO, M., GIBAJA, J.J. et al.: Patterns of Innovation in the EU-25 regions: A Typology and Policy Recommendations. In: *Working Papers 200804*. (2004).
- [28] VITURKA, M.: Hodnocení Regionální Kvality Podnikatelského Prostředí z Pohledu Inovačního Potenciálu. In: *3rd Central European Conference in Regional Science*. (2005), pp. 458 – 469.
- [29] HÁJKOVÁ, V., HÁJEK, P.: Identifikace Učících se Regionů. In: *Modelování a Rozhodování v Podnikání i ve Veřejné Správě 2009*. (2009), in press.

Appendix

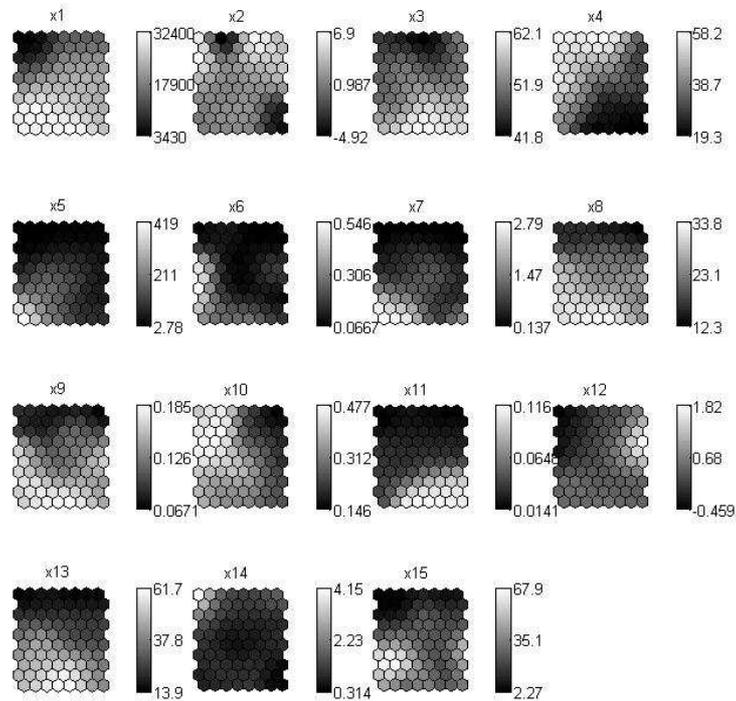


Fig. 5: Values of input variables for the representatives of regions in 2003

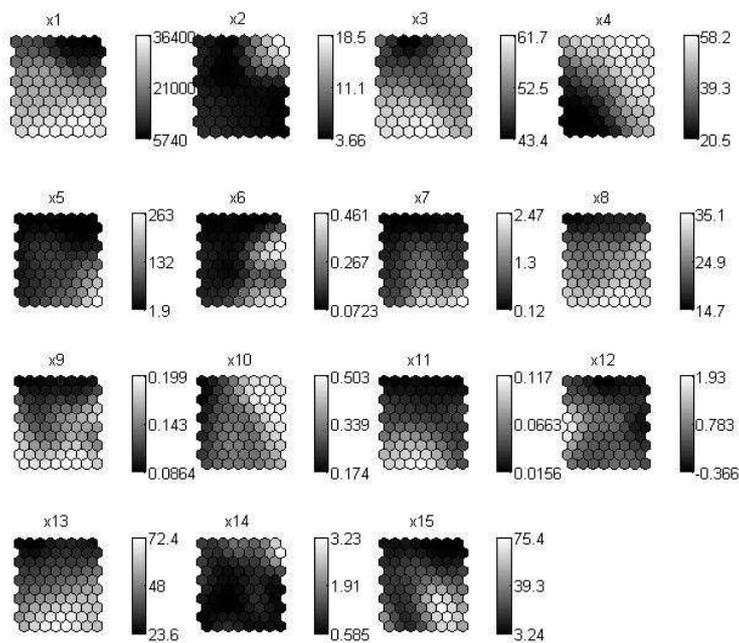


Fig. 6: Values of input variables for the representatives of regions in 2006