Analysis of Regional Innovation Systems by Neural Networks and Cluster Analysis

Veronika Hájková, and Petr Hájek

Abstract—The paper discusses the importance of regional innovation systems. They are defined as cooperation between companies and institutions in the development and dissemination of knowledge in innovation processes. The input variables are proposed for the analysis of regional innovation systems. They include factors related to economy, R&D and education. 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 regional innovation systems in a topological map. The results show on different typologies of regional innovation systems in the European Union.

Keywords—Neural networks, regional innovation systems, research and development.

I. INTRODUCTION

Innovation becomes a key competitive tool in an era of globalization. There are different approaches to the concept of innovation in economic theory. 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]. 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) [3].

Previous studies have been focused on analyzing RISs primarily by economic, R&D and educational indicators. 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 analysis of the RISs. The selected input variables characterizing the RISs represent the inputs of the proposed model. The variables concern following characteristics of regions: economy, R&D, and education. 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 typologies of the RISs.

The work is structured as follows. First a RIS is defined. On the basis of its characteristics the input variables are designed. Further, an overview of previous studies in the analysis of RISs is provided. Furthermore, the methods used for the modelling are characterized. Finally, the experimental results are presented and analyzed.

II. REGIONAL INNOVATION SYSTEMS IN EUROPEAN UNION

In the field of regional development, tools and policies are investigated to ensure economic growth and development. In this context, the concepts are discussed such as regional clusters, RISs, regional innovation networks, and learning regions which are attributes of successful development of a number of economies.

Since the early 1990s, the concept of the RISs has gained considerable attention from policy makers and academic researchers as a promising analytical framework for advancing our understanding of the innovation process in regional economies [4]. The concept of RISs has no generally accepted definitions, although it is typically understood to be a set of interacting private and public interests, formal institutions, and other organizations that function according to organizational and institutional arrangements and relationships conducive to the generation, use, and dissemination of knowledge [5]. Asheim and Gertler [6] define the RIS as the institutional
Communication and Management in Technological Innovation and Academic Globalization

infrastructure supporting innovation within the production structure of a region.

As the objective of this paper lies in the analysis of the RISs in the EU, it is important to provide a brief description of the specific position of the RISs in the Europe. As stated by [7], the initial position, arising from the European experience, was very different from that in regions in the USA. RISs were thought antidotes in Europe mainly due to following problems. Excellent published science was not exploited commercially. Moreover, there was a market failure in advanced business services that managed knowledge exploitation in e.g. Silicon Valley. Therefore, state intervention substituted market in innovation support. This kind of RIS is referred as an institutional RIS while the RIS oriented on commercial exploitation is called entrepreneurial RIS.

The two key sub-systems in any functioning regional innovation system are, following [8]: the knowledge application and exploitation sub-system; the knowledge generation and diffusion subsystem. The first is principally, but not only, concerned with firms while the second is mainly concerned with public organizations like universities, research institutes, technology transfer agencies, and regional and local governance bodies responsible for innovation support practices and policies. Cooke [7] states that in reality there may be some overlaps since firms conduct knowledge creation activities and universities and public or private research institutes conduct knowledge application activities. The latter is mainly the domain of firms.

In the literature, it is possible to find several studies analyzing RISs in European regions [9]. There have been two approaches for obtaining a RIS typology [9]. The first one deals with authors who used case studies in order to test previous conceptual works [3], [7]. Complex relations among subjects within regions justify this approach. The main objective of these studies is to understand how RISs function, to specify desirable factors and mechanisms for promoting competitiveness and innovation, and to assess the implications for policy [10]. These studies are focused on the impact of different types of RISs in different countries. It is important, therefore, to distinguish between different types of RISs.

Thus, Cooke [11] 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 and Isaksen [12] distinguish between three main groups of RIS in order to capture some conceptual variety and empirical richness in this phenomenon, which resemble the typology of Cooke. The first type is represented by a territorially embedded regional innovation network, where firms base their innovation activity mainly on localised leasing processes stimulated by geographical, social and cultural proximity without many interactions with knowledge organisations. The innovation networks may be further developed into regional networked innovation systems. The firms and organisations are still embedded in a specific region. The networked system is regarded as a regional cluster of firms surrounded by a local ‘supporting’ institutional infrastructure. Regionalised national innovation system stands for the third type of RIS. Industrial branches and the institutional infrastructure are more functionally integrated in national or international innovation systems. The cooperation of the main actors is conducted in order to develop more radical innovations with the use of scientific, formal knowledge. While the networked innovation system represents an endogenous development model, the regionalised national innovation system represents an exogenous development model.

Peripheral, mature industrial and metropolitan regions were classified by [13]. Concretely, RISs analyzed in [14] are classified as follows. Industrialized regions like Wales, Beden-Württemberg, Nordrhein-Westfällen, Brabant, Catalonia, Tampere/ Pirkanmaa and Slovenia which have an industrial structure with a strong position of low and medium technology, a governance structure dominated by public institutions, and a business structure characterized by the important role of multinational companies which are integrated into regional production networks. This type of industrial core regions differs from the industrial or service-oriented business districts dominated by small and medium-sized enterprises (SME) in Denmark or Tuscany. Other types of economic regions are the destroyed industrial regions (e.g. in Eastern Germany). Finally, there are the metropolitan design, research, communication and culture-based service regions described as global or regional cities (e.g. financial district – London, or global media cities – Paris, Munich).

The second way to create RIS taxonomies is realized using statistical analysis for a set of regions. Within the EURODITE project [15] a set of indicators for learning regions’ analysis at NUTS 2 level. This set involves following areas: science, technology, education, and performance. Moreover, specialization and performance of selected sectors were measured to provide additional information. The analysis was, however, not realized for all European regions. Regional profiles were found for each area (science, technology, etc.). 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, etc.

Clarysse and Mulder [16] found 6 groups of EU regions considering their GDP, unemployment, R&D expenditures and patents. Similar variables were studied also by [17] with similar results (6 groups – very strong position in knowledge services, ..., staying behind). In [18] 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 [19], 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 [20] resulting in 10 groups of regions.

In [21], 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 laggers. New member states were also studied by [9]. 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 specialized in high value-added services. An extension for EU-25 was published in [22]. 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; Innovative capital-regions specialised in high value-added services, etc.

Recently, we conducted our research with the objective to identify learning regions [23] in the EU. However, we have been unable to involve all the characteristics of learning regions through the selected variables. It will be necessary to include appropriate social and cultural capital proxies in these input variables in order to achieve the desired outcomes.

III. INPUT VARIABLES DESIGN FOR RISS ANALYSIS

We design the input variables for the analysis of RISs as indicated in Table 1. The selected input variables are related to the ability of a region to generate and absorb knowledge, and its capacity to transform R&D into innovation and economic growth.

The first four indicators were selected to reflect the socio-economic characteristics of a region. They include indicators such as per capita GDP, which can be considered as proxies of the stock of knowledge of a country [24], [25] and the degree of sophistication of its demand [21]. As Stern, Porter, and Furman [26] mention, GDP per capita measure the overall state of a country’s technological development. The employment rate is proxy of the "social filters" of a region, of the regional ability to transform R&D into innovation and economic growth [24], [27].

In addition to these economic indicators, we present also the indicators linked to R&D. Indicators on expenditure on R&D and patents, as in most other studies, are included as proxies for knowledge creation. We distinguish between public and private R&D, as they may carry out different types of research. The results obtained by [24] indicate that R&D investment, as a whole, and higher education R&D investment in peripheral regions of the EU, in particular, are positively associated with innovation. The existence and strength of this association are, however, contingent upon region-specific socio-economic characteristics, which affect the capacity of each region to transform R&D investment into innovation and, eventually, innovation into economic growth [24]. 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 patents serve as a measure of technological development. Moreover, the number of patents represents a proxy of the innovation capacity in a given region and evaluates the productivity of investments in R&D. The value of R&D intensity shows the relative effort of a region to create, disseminate, and exploit knowledge, and it is thus meant to be the main input in the knowledge production function [24].

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>DESIGN OF INPUT VARIABLES FOR RISS ANALYSIS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Economy</strong></td>
<td></td>
</tr>
<tr>
<td>x₁</td>
<td>Regional GDP per capita</td>
</tr>
<tr>
<td>x₂</td>
<td>Real growth rate of regional GDP</td>
</tr>
<tr>
<td>x₃</td>
<td>Employment rate</td>
</tr>
<tr>
<td>x₄</td>
<td>Long-term unemployment share</td>
</tr>
<tr>
<td><strong>R&amp;D</strong></td>
<td></td>
</tr>
<tr>
<td>x₅</td>
<td>Patent applications per capita</td>
</tr>
<tr>
<td>x₆</td>
<td>Public R&amp;D expenditure</td>
</tr>
<tr>
<td>x₇</td>
<td>Private R&amp;D expenditure</td>
</tr>
<tr>
<td>x₈</td>
<td>R&amp;D employment</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>x₀</td>
<td>Population with secondary education</td>
</tr>
<tr>
<td>x₁₀</td>
<td>Population with tertiary education</td>
</tr>
<tr>
<td>x₁₄</td>
<td>Participation in life-long learning</td>
</tr>
</tbody>
</table>

Our study also introduces indicators to proxy the knowledge and technological absorptive capacity of a region. The indicators related to education and human resources in science and technology virtually match those included in the European Innovation Scoreboard 2006, and distinguish, as [18], between general education and the qualification of human resources linked to R&D activities. Proportion of population with tertiary education is another important variable while technical skills are usually distinguished from the academic. It is also important to take into account qualitative parameters such as the readiness of people to a change and further education (participation in life-long learning). This analysis adopts a measure of educational attainment as a proxy of the level of skills. The higher the level of attainment is, the greater are the skills in a society and, therefore, the greater is its capacity to transform R&D into innovation. This study uses the share of the adult population that has attained secondary education as a proxy to denote the skills in a region.

IV. ANALYTICAL METHODS

In previous studies, factor analysis or cluster analysis were applied for the analysis of RIS typologies. 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 (CA). This way the results are not affected by multicolinearity. There is, however, still a loss of information (variance) when using FA or PCA. On the other hand, the resulting variables are usually easy to interpret based on factor or component loadings. When only a low proportion of the data variance is explained then the consequent use of CA would lead to biased results. Therefore, we propose to use a combination of neural networks and CA 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 the two-dimensional 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 the gained results. This is possible to conduct through the use of unsupervised neural networks. In this study we will apply Kohonen’s self-organizing maps (SOMs) [28] for the analysis of RISs. SOMs 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 neurons serving for the distribution of input patterns \( p_i \), \( i=1,2, \ldots, n \). 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 neighbours 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.

The learning algorithm of the SOM works as follows. First, the distances are computed between pattern \( p_i \) and synapse weights \( w_{ij} \) of all neurons in the competitive layer. The winning neuron \( j^* \) (Best Matching Unit, BMU) is chosen, for which the distance \( d_j \) from the given pattern \( p_i \) is minimum. The output of this neuron is active, while the outputs of other neurons are inactive. The aim of the SOM learning is to approximate the probability density of the real input vectors \( p_i \in \mathbb{R}^n \) by the finite number of representatives \( w_{ij} \in \mathbb{R}^n \), where \( j=1, 2, \ldots, m \). When the representatives \( w_{ij} \) are identified, the representative \( w_{ij^*} \) of the BMU is assigned to each vector \( p_i \). After the BMUs are found, the adaptation of synapse weights \( w_{ij} \) follows. The principle of the sequential learning algorithm [28] is the fact, that the representatives \( w_{ij^*} \) of the BMU and its topological neighbours move towards the actual input vector \( p_i \), according to the relation

\[
w_{ij}(t+1) = w_{ij}(t) + \eta(t) \times h(j^*, j) \times (p_i(t) - w_{ij}(t)),
\]

where \( \eta(t) \in (0,1) \) is the learning rate and \( t \) is time. In the learning process of the SOM, the representatives \( w_{ij} \) in the neighbourhood of the BMU are also adapted. 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 CA 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.

V. Analysis of Results

The input data for the modelling are represented by values of input variables \( x_1, x_2, \ldots, x_{11} \) 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 maps in Fig. 1 and Fig. 2. Similar regions on the map are located close to each other. Moreover, the optimal number of clusters was determined in order to achieve the highest quality of clustering (measured by the Dunn index of clustering quality). AS a result, the number of clusters \( c \) of nine is optimal for both the years 2003 and 2006.

![Fig. 1 Clustering of regions by SOM and k-means algorithm in 2003](image1)

![Fig. 2 Clustering of regions by SOM and k-means algorithm in 2006](image2)

The results are similar in the monitored period. It is evident from Fig. 1 and Fig. 2 that regions are located in similar positions on both maps.
The interpretation of clusters is possible based on the values of input variables for individual representatives on the map (grid). These values are presented in Fig. 3 and Fig. 4. Clusters can be ordered based on this interpretation. They are labelled by classes of RISs ordered from those regions which are the least similar to the concept of RIS to those which are the most similar to this concept. This can be done as follows.

Class 1 (cluster 7 (in 2003) and 8 (in 2006)) represents the undeveloped regions of the new Member States, such as most Czech and Polish regions, etc. Their positive characteristics involve a high growth in GDP and a high proportion of graduates in tertiary education.

Class 2 (clusters 4 and 2) 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. In 2006 many regions from this class moved to class 1 (regions from Bulgaria, Romania, and Slovakia). On the other hand, the other regions of Southern Europe arrived in this class 2 (Portuguese and Italian regions).

Class 3 (clusters 9 and 1) represents the developed regions of Southern Europe, e.g. Spanish, French and rich Italian regions. These regions have a high GDP per capita. However, they are average in R&D and education. In 2006 Portuguese and Greek regions moved to class 2, while several regions from class 4 approached class 3. Class 4 (clusters 8 and 5) stands for the developed regions of France, Austria and Belgium in 2003. In 2006 the structure of this class changed so that French, Belgian, and less developed German regions are located in this class. They have a high expenditure on public R&D. Though, they are average in all the other variables.

Class 5 (clusters 5 and 7) represents the advanced regions of Central and Eastern Europe, such as CZ01, HU10, SK01, DE41, DEG0, etc. They are investing in R&D and education.

Class 6 (clusters 1 and 6 in 2006) was created from clusters 8 and 1 of the year 2003. This class contains a wide range of Austrian, rich Spanish and French, Dutch, Swedish, and British regions. They are economically very strong, but only average in R&D with emphasis on secondary and long-life education.

Class 7 (clusters 6 and 9) represents high-tech industrial regions of Germany (Oberbayern, Karlsruhe, Freiburg, Stuttgart, etc.). 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.

Class 8 (clusters 2 and 4) involves Metropolitan regions such as Vienna, Brussels, Paris, Stockholm, and other highly developed regions from Belgium, Finland, Sweden or Great Britain are located in this class. They can be distinguished by means of a very high private investment in R&D and proportion of researchers, and by a highly educated population. From the economic point of view, they have the highest GDP per capital, high employment rate, and the lowest long-term unemployment rate.

Class 9 (clusters 3 and 3) the most developed RISs and it is represented by high-tech regions of Great Britain, Netherlands, Finland and Denmark. They are economically prosperous, with a medium proportion of investment in R&D and with a high proportion of researchers. The same as for class 6, the emphasis is put on secondary and long-life education.

Considering the development of the RISs over time, especially advanced regions of Central and Eastern Europe (e.g. SI01, CZ01) 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.

Considering the employment structure of the RISs, the regions in classes 1 and 2 have a large agricultural sector compared to other regions (see Table II). On the other hand, they have a small financial sector. Services and construction sector are typical for class 3. Class 4 have an average structure of employment. Classes 5, 6, and 7 are oriented on private and public services. Industrial regions are located in class 8. Class 9 corresponds to metropolitan regions, i.e. orientation on services, financial sector, and public sector.
regional innovation systems, regional innovation networks and learning regions which are attributes of successful development of a number of economies [30].

We discussed the current issues related to RISs. Similarly to previous studies we designed input variables for the analysis which was realized by using SOM and K-means algorithm. The representation of RISs in 2-dimensional grid makes it possible to illustrate the closeness of the RISs which corresponds to that one in the input space (i.e. of 11 input variables). Moreover, we can see the moving of the RISs in time. The results show that the concept of RIS is realized in the selected regions of Austria, France, Belgium, Sweden, Finland, Netherlands and Great Britain. In these regions, the most of the population is employed in the sector of services.

We have to mention that there is a limitation in our analysis that results from the use of the NUTS II classification developed by the Eurostat. The use of this classification for regional analysis is not simple and presents an important limit having to do with the choice of a geographical unit of analysis. The regions defined within NUTS II do not necessarily correspond to homogenous and self-contained regions in the broad sense [31].

### TABLE II

<table>
<thead>
<tr>
<th>c</th>
<th>Agric</th>
<th>Indus</th>
<th>Const</th>
<th>Servi</th>
<th>Finan</th>
<th>Publi</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.0</td>
<td>18.0</td>
<td>9.5</td>
<td>25.1</td>
<td>10.6</td>
<td>31.7</td>
</tr>
<tr>
<td>2</td>
<td>11.4</td>
<td>16.3</td>
<td>9.8</td>
<td>26.8</td>
<td>7.5</td>
<td>28.2</td>
</tr>
<tr>
<td>3</td>
<td>2.1</td>
<td>14.0</td>
<td>7.7</td>
<td>25.5</td>
<td>16.9</td>
<td>33.8</td>
</tr>
<tr>
<td>4</td>
<td>2.3</td>
<td>15.8</td>
<td>6.3</td>
<td>23.8</td>
<td>16.7</td>
<td>35.1</td>
</tr>
<tr>
<td>5</td>
<td>2.9</td>
<td>21.3</td>
<td>7.1</td>
<td>24.0</td>
<td>11.5</td>
<td>33.2</td>
</tr>
<tr>
<td>6</td>
<td>3.6</td>
<td>17.3</td>
<td>9.0</td>
<td>26.1</td>
<td>11.4</td>
<td>32.6</td>
</tr>
<tr>
<td>7</td>
<td>2.5</td>
<td>16.1</td>
<td>7.6</td>
<td>25.2</td>
<td>15.0</td>
<td>33.5</td>
</tr>
<tr>
<td>8</td>
<td>14.5</td>
<td>26.4</td>
<td>7.4</td>
<td>23.0</td>
<td>6.2</td>
<td>22.5</td>
</tr>
<tr>
<td>9</td>
<td>2.4</td>
<td>26.8</td>
<td>6.3</td>
<td>23.1</td>
<td>13.6</td>
<td>27.7</td>
</tr>
<tr>
<td>total</td>
<td>6.3</td>
<td>19.2</td>
<td>8.2</td>
<td>24.9</td>
<td>11.1</td>
<td>30.3</td>
</tr>
</tbody>
</table>

Agric - agriculture, hunting, forestry and fishing; 
Indus - industry excluding construction, 
Const - construction; 
Servi - wholesale and retail trade, repair of motor vehicles, motorcycles and personal and household goods; hotels and restaurants; transport, storage and communication; 
Finan - financial intermediation; real estate, renting and business activities; 
Publi - public administration and defence, compulsory social security; education; health and social work; other community, social and personal service activities; private households.

### REFERENCES


