IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT

Building an ANFIS-based Decision Support System for Regional Growth: The Case of European Regions

Journal:	*Transactions on Engineering Management*
Manuscript ID	TEM-18-0059.R2
Manuscript Type:	Research Article
Keywords:	Decision support systems, ANFIS, Forecasting, Regional Growth, Smart Specialisation Strategies
Subject Category:	06 Technology, Innovation Management and Entrepreneurship

Building an ANFIS-based Decision Support System for Regional Growth: The Case of European Regions

Abstract

This paper proposes an ANFIS-based Decision Support System that can provide European policy makers with systematic guidance in allocating and prioritizing scant public resources, whilst accomplishing the best growth performance at regional level. We do so by taking the stance of the Smart Specialisation Strategies which aim at consolidating the regional strengths and make effective and efficient use of public investment in R&D. By applying the ANFIS method we were able to understand how – and to what extent – the competitiveness drivers promoted technological development and how the latter contributes to the economic growth of European regions. We used socio-economic, spatial, and patent-based data to train, test and validate the models. What emerges is that an increase of R&D investments enhances the regional employment rate and the number of patents per capita; in turn, by taking into account the several combinations of specialisation and diversification indicators, this leads to an increase of the regional GDP.

Keywords: Decision Support System, Forecasting, ANFIS, Smart Specialisation Strategies, Regional Growth

1. Introduction

The Smart Specialisation program, fostered by the European Commission (EC), has become one of the main political initiatives to promote growth and economic development at a regional level. The rationale behind the Smart Specialisation program (RIS3 strategy)[1] (European Commission, 2011) is that by concentrating and linking resources to a limited number of priority economic actions, regions can increasingly become competitive and innovative; regions can take advantage of scale, scope, and knowledge spillovers, combining them with their regional strengths and offering a much greater chance of success. Then, if on one side few domains should have the highest concentration of R&D resources (e.g., Bartelsman et al., 1994; Krugman, 1991), on the other side part of the extant literature highlights how important is diversification in order to promote innovation (e.g., McCann and Ortega-Argilés, 2015; Frenken et al., 2007). In this light, whether undertaking a specialisation or diversification strategy becomes an important decision to make.

The aim of this work is exploring to what extent *technological specialisation*[2] and *technological diversification* pay off in terms of wealth creation at regional level. Specifically, we rely on the European Cluster Observatory initiative (established by the European Commission in 2007) which disentangles three types of indicators measuring respectively the competitiveness drivers (consisting of the indicators of public and private R&D investment), intermediate performance (representing the technological development of a region and including both the patent-based

---

[1] Smart Specialization (RIS3 strategy) is defined as "an industrial and innovation framework for regional economies that aims to illustrate how public policies, framework conditions, but especially R&D and innovation investment policies can influence economic, scientific and technological specialization of a region and consequently its productivity, competitiveness and economic growth path. It is a logical continuation in the process of deepening, diversifying and specializing of more general innovation strategies, taking into account regional specificities and inter-regional aspects, and thus a possible way to help advanced economies, as well as emerging economies, to restart economic growth by leveraging innovation led/knowledge-based investments in regions" (OECD, 2013, p.17).

[2] We focus on technology rather than product specialization/diversification (as in Sugheir et al., 2012) as data on technology are more disaggregated (which may be considered as a combination of technologies).

indicators and the socio-economic indicators), and outcomes (growth of a region). More than 60 indicators have been created merging a wide range of sources (e.g., Eurostat, the European Social Survey, national sources, Erawatch, ESPON, ISLA-Bocconi, OECD). We aim at finding the relationship between public and private R&D investments and indicators of technological development (both socio-economic and patent-based), in order to support European policy makers in better allocating scant resources and prioritize investments for the regional economic growth. The spatial dimension of the European regions is also considered. Then, a first research question can be stated:

RQ1: what is the best combination possible of competitiveness drivers and intermediate performance indicators in order to get regional economic growth?

How we are going to answer RQ1 depends on the technique we adopt. Although a number of decision support systems (hereafter DSSs) have been proposed for investment decisions (see Table 1), in this paper we rather propose DSSs based on Adaptive Network Fuzzy Inference Systems (ANFIS) (Jang, 1993).

Table 1. Main contributions on DSS for investment decisions

Author(s)	**Title**	**Journal**	**Publication year**	**Method**
Benaroch, M. and Dhar, V.	Controlling the complexity of investment decisions using qualitative reasoning techniques	Decision Support Systems	1995	Expert system based on qualitative reasoning techniques (i.e. qualitative simulation and qualitative synthesis)
Gottschilich, J., and Hinz, O.	A decision support system for stock investment recommendations using collective wisdom	Decision Support Systems	2014	Wisdom of crowd reasoning embedding it into investment decisions and portfolio management
Lourenço, J.C., Morton, A., and Bana E Costa, C.A.	PROBE – A multicriteria decision support system for portfolio robustness evaluation	Decision Support Systems	2012	Multicriteria portfolio decision analysis helpful in situations of limited resources
Ferretti, V., and Montibeller, G.	Key challenges and meta-choices in designing and applying multi-criteria spatial decision support systems	Decision Support Systems	2016	Introducing the spatial dimension for a DSS in order to deal with the spatial distribution of consequences
Fernandez, E., Navarro, J., Duarte, A., and Ibarra, G.	Core: A decision support system for regional competitiveness analysis based on multi-criteria sorting	Decision Support Systems	2013	CORE DSS based on ELECTRE-based preference model used in the framework of the new THESEUS multi-criteria evaluation method for making competitiveness assignments
Saracoglu, B.O.	Selecting industrial investment locations in master plans of countries	European Journal of Industrial Engineering	2013	Decision support procedure based on AHP for the location selection problems in master plans

Yurimoto, S., & Masui, T.	Design of a decision support system for overseas plant location in the EC	International Journal of Production Economics	1995	AHP-based DSS
Burinskiene, M., & Rudzkiene, V.	Comparison of spatial-temporal regional development and sustainable development strategy in Lithuania	International Journal of Strategic Property Management	2004	Decision support method based on multivariate statistical techniques from many probabilistic - statistical methods

Theorized by Jang in 1993, ANFIS is an artificial neural network that is based on Takagi–Sugeno fuzzy inference system. Since it integrates both fuzzy inference systems with fuzzy logic principles and neural networks, it has the potential to capture the benefits of both in a single framework. ANFIS combines two machine learning techniques: Fuzzy Logic and Neural Network; it uses Fuzzy Logic in order to transform given inputs into a desired output through highly interconnected Neural Network processing elements and information connections, which are weighted to map the numerical inputs into an output (Al-Hmouz et al., 2012). According to Chien et al. (2010), ANFIS is a convenient way to simulate the forecasting process and it has rapid learning speed (Chen and Zhang, 2005). ANFIS is composed of rules that can be analyzed and therefore can eventually be changed or adapted to a posteriori knowledge of the problem we try to model. Since during the review of the literature, we found that fuzzy experts-based methods are very widespread (e.g., Achiche et al., 2016; Achiche and Ahmed-Kristensen, 2011; Balazinski et al., 2000), the possibility to insert knowledge before and after modelling could be very important. Finally, the rule-based design characteristic of fuzzy logic makes the interpretability of the model easier compared to simple ANN. Therefore, we decided to choose ANFIS to perform our research and to build the final DSS. As we want to derive heuristics from the learned model, ANFIS is valid choice as it deals well with managing data uncertainties. Also, it is worth noting that we are dealing with a relatively small sample, it is therefore more suitable to have a prior knowledge of how to design the model in terms of rules in each dimension, selection of some features, the shape of membership functions, etc. (Achiche et al., 2013). This will result in more simple and more generalized models. A second research question can be stated:

RQ2: can we design a DSS to help policy makers to better invest public financial resources driving the economic growth of their regions?

By answering these two research questions, we provide with the following contributions. First, we describe a conceptual framework and propose a systematic way to move from the competitiveness drivers to the outcome of regional growth through a comprehensive set of intermediate performance; to the best of our knowledge, this study is one of the few studies building up ANFIS-based DSSs by adopting a holistic view on the regional development and growth variables. Former studies focused on few indicators at a time such as R&D expenditures and government taxation incentives (e.g., Xu and Xu, 2013), type of strategies to be used in order to accomplish technology standards (e.g., Van De Kaa et al., 2014), and with a rather narrow view on specific industries. Second, we take into account the spatial dimension of European regions as this may allow to identify direct and indirect effects as well as spillovers among them. Third, we test the performance of the DSSs and come up with a forecasting tool able to help policy makers to prioritize regional investments more wisely.

2. Literature Review

We searched for articles related to the forecasting techniques in order to understand which one could have been used for the purpose of the present study. Starting from the objectives of the present study, we first defined the characteristics that the method should have in order to perform the research:

- quantitative method;
- method able to extract information from historical data with a good approximation capability;

- method able to perform multivariable – and obviously multivariate – analysis;
- method easy to interpret.

We decided to categorize the articles on the basis of the following recurrent topics:

- Technique: it refers to the general quantitative or qualitative type of forecasting techniques used in the article;
- Field: it refers to the general argument of the article that the statistical techniques are applied to. The fields identified are:
  - RIS3 strategy;
  - Resources allocation;
  - Technology and Innovation development;
- Approach: it is divided into Empirical/Application and Theoretical/Methodological.

Figure 1 summarizes the process:

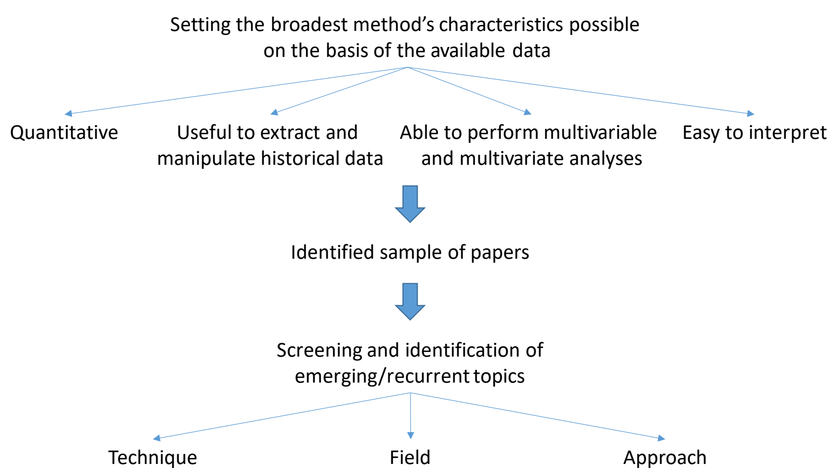

Figure 1. Identification of emerging topics

The next three sections will provide the reader with a comprehensive explanation of the three major emerging topics that stand in the background of our study.

2.1 Smart Specialisation Strategies (RIS 3)

Research concerning Smart Specialisation Strategies (RIS3) is relatively recent. Authors mainly used empirical approaches and a combination of qualitative and quantitative techniques to perform their research. RIS3 envisages the concentration of resources on specific – limited – set of priorities; this approach, which is thought to be as objective and quantifiable as possible, emerged as a consequence of public resources misuses and lack of transparency in public funding decisions (Paliokaitė et al., 2015). Piirainen et al. (2017) argue that in the context of smart specialisation, regional foresight tends to be linked to the development of the policy-making agenda only at an higher level; indeed, while foresight is based on more qualitative information and data (e.g., Nayak, 2010; Cuhls, 2003) making it more prone to opinions when it comes to questioning how the future looks like in the specific area under investigation, forecasting techniques (by their very quantitative nature) can better inform policy makers in setting the R&D agenda leveraging on the most promising (emerging) technological opportunities (e.g., McCann and Ortega-Argilés, 2015; Paliokaitė et al., 2015; Harper and Georghiou,2005; Piirainen et al., 2017). Foresight and forecasting are not mutually exclusive; rather, results from forecasting techniques can inform a better formulation of foresight processes (and vice versa). Paliokaitė et al. (2015), with the purpose of identifying RIS3 priorities, have proposed a combination of analytical and participatory methods, combining both qualitative techniques (qualitative analysis of trend and challenges, surveys, Delphi Method, scenarios and roadmaps) and quantitative ones (statistical analysis, bibliometrics, multiple criteria analysis). Building on their first paper, Paliokaitė et al. (2016) performed another research by focusing on the implementation of roadmaps in Lithuania, built through the opinions of Experts panels, for the implementation of selected RIS3 priorities. Reichardt et al. (2016) have made an empirical research based on patent data focusing on the offshore wind service sector in four regions around the North Sea; they contributed to the smart specialisation literature to the extent that “the empirically corroborated typology of diversification patterns that can be used as an analytical framework for both analysis and anticipation." Finally, Fabbri (2016) proposed an innovative methodological scheme for strategic planning decisions concerning the RIS3 strategies; the scheme concerned is based on foresight, road mapping and large participation processes of experts and stakeholders. However, in the context of our study the focus is on forecasting techniques since there is a lack of attention from extant literature on how such techniques can inform policy making within the RIS3 framework.

2.2 Resource Allocation

Resource allocation concerns the allocation of mainly financial resources in order to improve innovation and development. Most of the articles in this field followed an empirical approach mainly using quantitative techniques. In order to take the proper decisions about resource allocation, R&D decision makers tend to rely on empirical data, such as papers and patent data (Kim, 2010). The relative merits and weaknesses of patent statistics have been widely discussed in prior studies, while showing patent statistics provide still with useful indicators of technological activities at both the firm and the country levels (Wang et al., 2015). Altuntas et al. (2015) used patent-based criteria advancing a new method for technology forecasting in order to efficiently prioritize investments. They did so by starting from the number of patents per year, going through the use of S-shaped curves, calculating technology diffusion speed, patent power and expansion potential of technologies, ending up with investment evaluations. Doha and Kimb (2014), alongside other indicators (e.g., trademark registrations, utility models), measured SMEs' innovation through the number of patents. In order to test the general hypothesis of a positive relationship between government financial support and the innovation of regional SMEs at the firm level, indicators such as firm size, private enterprises, private research organizations and others have been employed in the multiple regression equation. Vargas and Angel (2007) treated the theme of SMEs innovation and development (particularly referring to Mexico) trying to understand through multivariable analysis how the development and the proper allocation of internal resources (financial and not) could be at the basis of competitive advantage and improve business performance. Wang et al. (2015) investigated the link between technological size (number of patents, population, GDP) and technological diversification at the level of province, in order to better take decisions about future investments. Yurynets (2016), by using a technique belonging to the Artificial Intelligence sphere (ANN – Adaptive Neural Network), wanted to investigate the factors that have the greatest influence on the GDP of Ukraine in order to forecast its economic, scientific, technical and innovative development and make better decisions about investments on key areas of interest. Finally, Jun et al. (2017) built up a forecasting model in order to identify the differences between demanding companies and beneficiary companies of R&D supports in Korea and evaluate whether the implementation of the policy was efficient; multivariate analyses, in particular discriminant analysis, were used to construct the model.

2.3 Technology and Innovation development

Contributions belonging to this field are homogeneously distributed through the years and research has been conducted following both an empirical/application and a theoretical/methodological approach using, without a relevant majority, qualitative and quantitative techniques. Hence, due to the large quantity of methods employed for various purposes, even not properly inherent to our study, we have just made some observations about the statistical techniques used in this field. One of the most recurrent quantitative forecasting techniques is the S-shaped curve which entails estimating future performance by finding the fitting growth curve to a set of technological performance data (Chen et al., 2011). Robertson et al. (2007) shed new light on innovative product diffusion by modeling S-shaped curves; Kim (2010), in order to make a forecast-based evaluation of technology in South Korea, used both expert interviews and growth curve fitting; Intepe and Koc (2012) used patent data in order to investigate how S-Curve could help to disentangle both the technological and evolutionary trajectory of the 3D TV technology. Among the qualitative techniques, instead, Delphi method and Scenarios are the most recurrent methods: Varho et al. (2016), by focusing on the renewal energy sector in Finland, implemented the Delphi method and Hierarchical Cluster analysis to collect and analyze data while representing future scenarios; Weber and Rinchel (2016) used a scenario-based approach to sectoral innovation foresight; Okuwada (2013) used scenarios to identifying expected areas of future innovation. Visualization technique, such as roadmap (Sarkkinen and Kässi, 2014), multi-path mapping (Robinson and Propp, 2008), generative topology maps (Song, 2014) are also used as a practical tool for effective anticipation and management of standardization. Multiple regression analysis is also used in this field: Bartels et al. (2012) outlined innovation policies by identifying to what extent knowledge institutions, governments, and business corporations shape National Innovation Systems; Sarkkinen and Tuomo (2014) investigated the relations of factors and systems influencing innovation. Finally, several modelling techniques have been applied for technological and innovation forecasting: System dynamics for exploring systems structure in order to increase the understanding of industry's system behavior (Hsiao and Liu, 2012); exploratory modelling and analysis, in order to understand the impact of information dynamic on innovation diffusion (Yücel and Van Daalen, 2011).

Overall, the former three sections introduce the reader with the major emerging topics characterizing the theoretical background of our study. These three sections are instrumental to create the link with the two research questions (RQ1 and RQ2) stated upfront in the Introduction section. Concerning RQ1 (*what is the best combination possible of competitiveness drivers and intermediate performance indicators in order to get regional economic growth?*), the three sections highlight the need to link the regional economic growth with some antecedents; precisely, while RIS3 deal with the general rationale behind European actions for development and growth, Resources Allocation identifies the need to put emphasis on finding smart ways to allocate scant resources for specific actions able to stimulate and sustain the regional economic growth. The third section on Technology and Innovation development emphasizes the need to pay attention on the antecedents of the regional economic growth that can be inflected in terms of competitiveness drivers and intermediate performance. The competitiveness drivers are a collection of specialisation indicators, firms behavior indicators, and business environment indicators; whereas intermediate performance mainly deal with indicators measuring the employment rate, number of enterprises, enterprises growth, number of patents, scientific publications, and labour productivity. Once selected the indicators, we can be able to answer RQ2 (*can we design a DSS to help policy makers to better invest public financial resources driving the economic growth of their regions?*) by building up an ANFIS-based DSS and testing it in order to check whether – and to what extent – the resulting forecasting can support policy makers in improving their investments decisions.

3. Method and Data

3.1 Research Setting

In order to better understand the logic used in the present project, it is important to underline that both diversification strategies (Frenken et al., 2007) – Related Variety (RV) and Unrelated Variety (UV) – which describe respectively the extent to which a region is diversified in different 'main classes' of the Fraunhofer Technology Classification and the extent to which a region is diversified in similar Fraunhofer domains, and the specialisation strategies – Herfindahl as an indicator measuring the level of concentration of applications and Number of Specialized Fields – are considered as alternative to foster innovation at regional level[3]. The present study relies upon data collected from Eurostat and PATSTAT[4], socio-economic data, patent-based indicators and spatial

[3] Detailed definitions of each Fraunhofer domain are available at the following link: http://www.wipo.int/export/sites/www/ipstats/en/statistics/patents/pdf/wipo_ipc_technology.pdf
[4] https://www.epo.org/searching-for-patents/business/patstat.html#tab-1

data for 134 EU-27 regions disaggregated at NUTS 2 level[5], for the time window 2002-2011. Previous research shows that spatial dependence among proximate European regions is an important indicator for assessing growth opportunities (Le Gallo et al., 2011; Van Oort et al., 2014; Boschma et al., 2014; Cortinovis and Van Oort, 2015; Content and Frenken, 2016) and exploiting knowledge spillovers (Hidalgo et al., 2007; Bahar et al., 2014).

3.2 Selecting the forecasting technique

On the basis of what already exists in literature (e.g., Eerola and Miles, 2011; Cocianu and Grigoryan, 2015; Könnölä et al., 2011) and on the use of a quantitative method able to extract information from historical data with a good approximation capability, able to perform multivariable – and obviously multivariate – analysis and easy to interpret, we have delimited the methods to the quantitative ones and specifically to ANN[6] and Multivariate-Regression analysis. Particularly, among the various forecasting techniques of Multivariate-Regression analysis we chose to focus on the autoregressive moving average (ARMA) family, particularly on ARIMAX. In general, ARMA models use time series data in order to forecast future points. ARIMA[7] models are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied one or more times to eliminate the non-stationarity. ARIMAX, finally, is an ARIMA model with the addition of Exogenous Variables, capable of forecasting the future of a variable, based on the past values of

[5] http://ec.europa.eu/eurostat/web/nuts

[6] Neural Networks (NNs) are a set of modelling techniques which have a wide range of applications including statistical modelling, discrete classification, pattern recognition, control systems, etc. NNs mimic how biological NNs operate and learn. NNs are constructed from multiple layers of neurons connected by weights from each neuron to each neuron of the proceeding layer. Neurons are the base unit of the network. The main benefit of using the NN methodology over other techniques is that it is able to identify non-linear relationships between the independent and dependent variables (Bennett et al., 2014).

[7] The 'AR' part of ARIMA indicates that the evolving variable of interest is regressed on its own lagged (i.e. prior) values. The 'MA' part indicates that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various times in the past. The 'I' (for "integrated") indicates that the data values have been replaced with the difference between their values and the previous values.

the variable itself and on the variation and co-variation of other variables, properly selected for the purpose. In order to select the best method for our study, we based our assessment both on papers using ARIMAX and ANN and on some bibliographical data extrapolated by Scopus. Since not many comparative studies are present in literature about these two techniques in the context of our research, we also analyzed the researches carried out between ARMA-ARIMA model and ANN or ARIMAX and hybrid ANN. Cocianu and Grigoryan (2015) carried out research on the potential of artificial neural networks (ANN) applied to data forecasting and analyzed against the classical autoregressive integrated moving average (ARIMA) model; they performed a comparative analysis of these models revealing that results obtained using the neural approach provided better results. Adebiyi et al. (2014) confirmed that the performance of ANN model is better than ARIMA model in terms of forecasting accuracy on many occasions, analyzing them on Stock Price Prediction. Moreover, ANN techniques use high number of parameters and nonlinear processing that increases computational calculus exponentially and results in nonlinear filters with the better identification properties, allowing ANN to outperform ARIMA on volatile time series (Garcìa and Mendez, 2007). Besides, NNs models of forecasting are deprived of certain limitations of the classical methods of forecasting (e.g., Aleksandrova et al., 2007; Boychuk and Novakevych, 2014); for instance, they have the capabilty to deal with non-linearities thriought the management of the number of layers; monotony or periodicity of the future value, which is inherent to the numerous extrapolation methods; creation of complicated dependencies, because NNs are nonlinear by their nature (containing a great amount of incoming information, between which there exist no evident regularities and interrelations). Moreover, the forecasting made based on artificial neural networks does not have any limitations regarding incoming information (Yurynets, 2016). Concerning the use of ARIMAX on our data on Matlab, due to the small available time series per each region (around 10 years per region), the process of estimation of the parameters of ARIMAX returned with very high, not admissible, variance. Hence, we focused on ANN, carrying out a quantitative and a bibliometric analysis of papers found on Scopus by setting the search keywords on ANN and FORECASTING. A particular characteristic that is possible to deduce from the bibliometric map is the co-occurrence of neural networks with fuzzy systems and fuzzy logic. Analyzing the addition of Fuzzy logic and Fuzzy systems to ANN, we found that the combination of these two methods provides some advantages to the basic technique.

3.3 Data preparation

We have arranged the data in several ways in order to capture as many variations as possible and to understand which model better fits the data. The investigation has been split into two main parts:

1) Forecasting: we wanted to understand if, starting with the competitiveness drivers (CD), the model can predict the intermediate performances (IP) at time T+1; then, with these predicted ones, if we can predict the final GDP per capita at time T+1, T+2 and T+3 (the time lag is compared to the time of competitiveness drivers, which are at time T);
2) Classification: the first part (CD→IP) is the same of the forecasting analysis; concerning the second one, we wanted to investigate if the model can classify the various regions in the right zones of GDP per capita, in time T+1, T+2, T+3.

Table 2 shows a summary of the model constructed:

Table 2. Summary of the models

Without Spatial Variables			
Forecasting			**Classification**
Model E	Model D	Model R	K-means clustering Model
With Spatial Variables			
Forecasting			**Classification**
Model E	Model D	Model R	K-means clustering Model

where, Model E refers to the ‘exact values forecasting’ containing the original values of all the variables, without any arrangement; Model D refers to the ‘difference forecasting’ aiming at understanding first, the behavior of each variable with respect to itself (if it increased or decreased) and then, if the behavior of the inputs can influence the growth of the outputs and in which way; this is why Model D contains data of each variable taking its value at time T and subtracting its value time T-1[8]; finally, since we cannot know a priori which one among the difference and the ratio is the best data treatment for the purpose explained above, we decided to analyze both; the ratio was performed, for each variable, taking the value at time T and dividing the value of the same variable at time T-1; hence, Model R is about data of each variable being divided, taking its value at time T and subtracting its value time T-1[9]. It is worth noting that the competitiveness driver Public (government) R&D expenditure (%) was not included neither in the "difference dataset" nor in the "ratio dataset" since more than 80% of the data registrations do not change through the time. Finally, for the K-means clustering Model, we performed a cluster analysis using k-means algorithm: the GDP per capita has been divided into clusters, fixing the number of clusters (k) to 5; we chose this number in order to split the data in: very small, small, medium, high, very high. For all the models explained above, Latitude and Longitude have been added in input, alongside with the competitiveness drivers to investigate if the geographical collocation of the regions themselves can help to forecast the intermediate performances and therefore to forecast or classify GDP per capita. Once the data has been prepared, each dataset (Model E, Model D, Model R and K-means clustering Model) has been divided into three parts: Training set, Validation set, and Testing set to construct the final model (CD→predicted IP→predicted GDP).

[8] For instance, concerning GDP per capita: $GDPpercapita_{DIFF(T)}$=$GPDpercapita_{(T)}$-$GDPpercapita_{(T-1)}$
[9] For instance, concerning GDP per capita: $GDPpercapita_{RATIO(T)}$= $GPDpercapita_{(T)}$/$GDPpercapita_{(T-1)}$

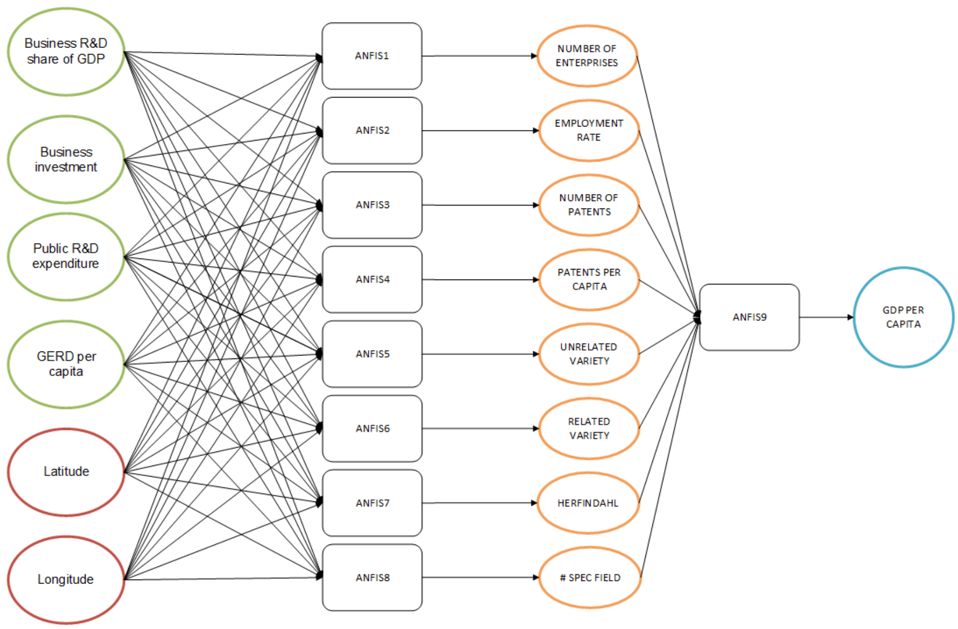

Figure 2. General procedure performed on Matlab

An example of this kind of rules (Takagi and Sugeno, 1983; Takagi and Sugeno, 1985) is:

If Input 1 is x and Input 2 is y, then Output is $z = ax + by + c$

According to the methodology and definitions advanced by the European Cluster Observatory[10], in this framework all the variables are represented:

- Competitiveness Drivers, divided into Firm Behavior and Business Environment, consist of indicators of public and private R&D investment: Business R&D share of GDP, business investment, public R&D expenditures, Gross R&D Expenditures (GERD) per capita;
- Intermediate Performance represents the technological development of a region and includes both the patent-based indicators (e.g., number of patents, patents per capita, unrelated variety, related variety, Herfindahl, number of specialized fields) and the socio-economic indicators (e.g., Number of Enterprises, Employment Rate);
- Outcome represents the growth of a region in terms of GDP per capita.

[10] www.clusterobservatory.eu

Table 3 shows the variables of the final dataset:

Table 3. Database variables and description

Variable	Description
Business R&D share of GDP (%)	The R&D conducted by private companies as a percentage of the total GDP in a region
Business investment (€ per employee)	Amount of euros invested in R&D by private companies on the total number of the employees
Public R&D expenditure (%)	The Research and Development activities conducted by public institutions as a percentage of the total GDP in a region
GERD per capita (€ per inhabitant)	The total intramural R&D expenditure (or Gross Domestic Expenditure on Research and Development) per inhabitant
Number of Enterprises	Number of operative enterprises in a region
Employment Rate	It is a measure of the extent to which available labour resources (people available to work) are being used. They are calculated as the ratio of the employed to the working age population in a region.
GDP per capita (€ per inhabitant)	The Gross domestic product - GDP at current market prices - is the result of the production activity of resident producer units
Latitude: Arithmetic mean of the latitude coordinates of Regions' boundaries	It is the angular distance north or south from the equator to the centroid of every Region
Longitude: Arithmetic mean of the longitude coordinates of Regions' boundaries	It is the angular distance east or west from the north-south line that passes through Greenwich, England, to the centroid of every Region
Number of Patents	Total number of EPO (European Patent Office) applications per year
Patents per capita	Number of EPO applications per year per million inhabitants
Unrelated Variety (UV) $UV = \sum_{g=1}^{G} S_g \log_2\left(\frac{1}{S_g}\right)$	S_g: share of the main class g in the regional technology portfolio. It can be obtained summing the Fraunhofer shares for all domains belonging to the main class considered: $S_g = \sum_{i \in T_g} S_i$, where with g= 1,…,5 indicates the main classes. UV measures the extent to which a region is diversified in very different types of technology ('main classes')
Related Variety (RV) $RV = \sum_{g=1}^{G} S_g H_g$	Where: $H_g = \sum_{i \in T_g} \frac{S_i}{S_g} \log_2\left(\frac{1}{S_i/S_g}\right)$, RV measures the extent to which a region is diversified in similar domains
Herfindahl $H = \sum_{i=1}^{N} s_i^2$	Where S_i: share of patenting of the technological domain i in the regional technology portfolio, i.e. in one of the 35 Fraunhofer domains. N: number of technology domains in which the respective region is active. This indicator analyses the level of concentration of applications by region by year
Number of Specialized Fields: Number of Fraunhofer domains in which RTAij >2	Where: $RTA_{ij} = \frac{P_{ij}}{\sum_i Pij} \Big/ \frac{\sum_j Pij}{\sum_{ij} Pij}$ with P the number of EPO applications, *i* = technological Fraunhofer domain and *j* = region grouping variable. It indicates the number of Fraunhofer domains in which a Region has an outspoken technological strength.

Table 4 shows an excerpt of the data and indicators calculated for a specific region:

Table 4. Excerpt of data

General information				Competitiveness drivers				Intermediate performance								Outcome		
NAT	REGCODE	REGNAME	YEAR	Business R&D share of GDP (%)	Business investment (thousand EUR/employee)	Public (government) R&D expenditure (%)	GERD	Number of Enterprises T+1	Empl. Rate T+1	Number of patents T+1	Patents per capita T+1	UV T+1	RV T+1	H T+1	Spec. Field T+1	GDP per cap. T+1	GDP per cap. T+2	GDP per cap. T+3
FR	FR10	Île de France	2002	2.3	20	0.5	1311.6	383377	66.2	5613	498.0446	2.044167	2.726391	0.045	2	40100	40900	42300

FR	FR10	Île de France	2003	2.2	19.05	0.4	1274.6	387445	66.1	6245	550.2062	2.075946	2.696163	0.044	1	40900	42300	43600
FR	FR10	Île de France	2004	2.1	18.61	0.5	1278	389156	65.5	6661	582.1462	2.089204	2.698759	0.044	1	42300	43600	46200
FR	FR10	Île de France	2005	2	18.96	0.5	1294.6*	400338	65.4	6520	565.3638	2.070298	2.688872	0.045	2	43600	46200	49200
FR	FR10	Île de France	2006	2	20.42	0.4	1311.2	399925	66.6	5344	460.7347	2.065828	2.641184	0.048	2	46200	49200	47500
FR	FR10	Île de France	2007	2	21.79	0.4	1358.8	404592	67.5	6108	523.8754	2.108597	2.666739	0.044	2	49200	47500	49700
FR	FR10	Île de France	2008	2	22.28	0.4	1409.3	415855	66.6	6679	569.4802	2.110661	2.711719	0.042	2	47500	49700	51200
FR	FR10	Île de France	2009	2	20.38	0.4	1440.8	413449	65.3	6865	582.4592	2.118975	2.691771	0.043	2	49700	51200	
FR	FR10	Île de France	2010	2	21.08	0.4	1492.4	413449	65.6	5100	430.2762	2.096831	2.667523	0.046	3	51200		
FR	FR10	Île de France	2011	2	21.08	0.4	1551.8	413449	66.2	1365	114.5425	2.075804	2.459317	0.057	5			

*Originally this cell was empty. This value comes from the 1-D linear interpolation

3.4 Data analysis

The analysis has been performed on Matlab using different combinations of membership functions (Gaussian and bell-shaped types), different number of training epochs and different optimization methods (backpropagation algorithm and hybrid learning technique) with the aim to investigate which one could provide the best results. To develop a Sugeno-type fuzzy inference system, that is the one necessary for ANFIS, we used grid partitioning algorithm if the number of inputs was ≤ 4, and subtractive clustering algorithm if the number of inputs was ≥ 4. Each variable of the intermediate performance has been forecasted using one ANFIS, having as inputs, in the analyses without the spatial variables the four competitiveness drivers, while in the analyses with the spatial variables both the four competitiveness drivers and latitude and longitude. Hence, at the end of the first phase, 8 ANFIS have been created. Once the 8 indicators of intermediate performance have been forecasted, we used them as inputs to another ANFIS model to forecast or classify the main outcome, GDP per capita. The general procedure is illustrated in Figure 2. The goodness of the models was calculated on the basis of the following indicators:

Forecasting

$$\text{Error: } RMS_H = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y'_H(x_i))^2}{n}} \qquad \text{Fitness Performance: } FP_H = \frac{L_E - RMS_H}{L_E} \cdot 100$$

Where, taking as an example the GDP per capita:

$$L_E = GDP_{\max} - GDP_{\min}$$

$$\text{Model E: } y'_E(x_i) \quad \text{Model D: } y'_D(x_i) = y_i + \Delta y'(x_i) \quad \text{Model R: } y'_R(x_i) = y_i + Ry'(x_i)$$

Classification

$$\text{Fitness Performance: } FP_H = \frac{\text{Number of forecasted values in the right cluster}}{\text{Total number of observations}} \cdot 100$$

4. Results

4.1 Identifying the best Decision Support Systems

After collecting the values of RMS and FP, we have been able to decide the combination to use during the building of the final models: 2 membership Gaussian function and 1,000 epochs of training and simple back propagation as optimization method (see Table 5).

Table 5. Excerpt of Matlab-Errors routines and Fitness performance values

Variable	ENTERPRISES	ENTERPRISES	ENTERPRISES	ENTERPRISES	ENTERPRISES	ENTERPRISES	ENTERPRISES	ENTERPRISES
MF type	GBELL	GBELL	GBELL	GBELL	GAUSS	GAUSS	GAUSS	GAUSS
MF number	MF2	MF2	MF3	MF3	MF2	MF2	MF3	MF3
Opt. Method	MET 0	MET 1	MET 0	MET 1	MET 0	MET 1	MET 0	MET 1
RMS	3884.151	4636.312	3865.261	14981.21	3916.736	4700.872	3907.695	14659.73
FP	87.69007	85.30627	87.74994	52.52049	87.9868	85.10166	87.61546	53.53934
Variable	EMPLOYMENT RATE	EMPLOYMENT RATE	EMPLOYMENT RATE	EMPLOYMENT RATE	EMPLOYMENT RATE	EMPLOYMENT RATE	EMPLOYMENT RATE	EMPLOYMENT RATE
MF type	GBELL	GBELL	GBELL	GBELL	GAUSS	GAUSS	GAUSS	GAUSS
MF number	MF2	MF2	MF3	MF3	MF2	MF2	MF3	MF3
Opt. Method	MET 0	MET 1	MET 0	MET 1	MET 0	MET 1	MET 0	MET 1
RMS	1.238759	1.339755	1.255215	2.550508	1.238847	1.326425	1.262088	4.129635
FP	77.8793	76.07581	77.58545	54.45521	78.87773	76.31383	77.46271	26.25652
Variable	NUMBER OF PATENTS	NUMBER OF PATENTS	NUMBER OF PATENTS	NUMBER OF PATENTS	NUMBER OF PATENTS	NUMBER OF PATENTS	NUMBER OF PATENTS	NUMBER OF PATENTS
MF type	GBELL	GBELL	GBELL	GBELL	GAUSS	GAUSS	GAUSS	GAUSS
MF number	MF2	MF2	MF3	MF3	MF2	MF2	MF3	MF3
Opt. Method	MET 0	MET 1	MET 0	MET 1	MET 0	MET 1	MET 0	MET 1

RMS	234.1287	303.996	234.467	304.5156	233.6361	332.5206	234.6301	417.4127
FP	87.93151	84.3301	87.91407	84.30332	87.9569	82.85976	87.90566	78.48388

Final results have been collected and summarized in the Table 6. The values of RMS and FP of each model refer to the final value of the GDP per capita, forecasted using the indicators of the intermediate performance, which, in their turn, have been forecasted using the variables of investments in R&D of the competitiveness drivers (and latitude and longitude in the models with the spatial variables). Hence, the only real data used to build the final model have been the competitiveness drivers and the spatial variables, used as inputs to each model. According to the data in Table 6, we chose to interpret two models, building two different Decision Support Systems:

1) Model D without spatial variables at time T+1
2) K-means clustering with spatial variables at time T+1

Table 6. Error and Fitness Performance values of the final models (the two chosen models in light orange)

			CD → IP → GDP					
			Without spatial variables			With spatial variables		
			T+1	**T+2**	**T+3**	**T+1**	**T+2**	**T+3**
Forecasting	Model E	RMS	6.89E+03	7.57E+03	7.16E+03	9.02E+03	9.59E+03	9.20E+03
		FP	87.4171	83.0585	85.2112	81.4364	78.2774	80.4037
	Model D	RMS	2.08E+03	2.89E+03	2.67E+03	2.11E+03	2.98E+03	2.23E+03
		FP	96.1141	93.3670	94.5192	95.8075	93.1078	95.3845
	Model R	RMS	2.11E+03	2.97E+03	2.74E+03	2.19E+03	2.97E+03	2.98E+03
		FP	95.3817	93.1637	93.7792	94.2752	93.1871	93.1141
Classification	K-means clustering	% in cluster	59.4595 (44/74)	51.3514 (38/74)	47.2973 (35/74)	64.8649 (48/74)	54.054 (40/74)	52.7027 (39/74)

4.2 Decision Support System 1 (Model D without spatial variables)

We are going to interpret the rules of both the intermediate performance and the outcome using the fuzzy logic. Since used data refer to the difference, we are not going to use LOW, MEDIUM or HIGH as linguistic label, but rather the symbol "+" if the value has a positive sign and therefore it increases or the symbol "–" if the value has a negative sign and therefore it decreases. Besides the symbol ~ is used if the variation of the output, both negative and positive, is not significant. A variation can be classified "not significant" if its value is too small with respect to the range of the variable in question (i.e. regarding the variable “number of enterprises” a variation of 100 is too small to be taken into consideration). Hence, each rule has to be read as follows:

If competitiveness driver 1 decreases/increases and competitiveness driver 2 decreases/increases and competitiveness driver 3 decreases/increases,

Then intermediate performance decreases/increases/does not vary significantly

These are the results for the application of DSS1 (Model D without spatial variables[11]) to the relationships Competitiveness drivers → Intermediate performance:

If	***If***	***If***	***Then***
Competitiveness driver (1)	**Competitiveness driver (2)**	**Competitiveness driver (3)**	**Intermediate performances**
Business R&D share of GDP (%)	Business investment (€ per employee)	GERD per capita	Number of enterprises
–	–	–	~
–	–	+	~
–	+	–	~
–	+	+	~
+	–	–	~
+	–	+	~
+	+	–	–
+	+	+	+
Business R&D share of GDP (%)	Business investment (€ per employee)	GERD per capita	Employment rate
–	–	–	+
–	–	+	–
–	+	–	–
–	+	+	+
+	–	–	–
+	–	+	+
+	+	–	–
+	+	+	+
Business R&D share of GDP (%)	Business investment (€ per employee)	GERD per capita	Number of patents
–	–	–	~
–	–	+	~
–	+	–	–
–	+	+	+
+	–	–	~
+	–	+	~
+	+	–	~
+	+	+	~
Business R&D share of GDP (%)	Business investment (€ per employee)	GERD per capita	Patents per capita
–	–	–	~
–	–	+	~
–	+	–	–
–	+	+	+
+	–	–	~
+	–	+	+
+	+	–	~
+	+	+	~
Business R&D share of GDP (%)	Business investment (€ per employee)	GERD per capita	Unrelated Variety
–	–	–	–
–	–	+	~
–	+	–	+
–	+	+	+
+	–	–	+
+	–	+	–

[11] Public R&D expenditure (Competitiveness driver (4)) was excluded from this model because it does not vary significantly through the years

+	+	−	~
+	+	+	+
Business R&D share of GDP (%)	Business investment (€ per employee)	GERD per capita	Related Variety
−	−	−	+
−	−	+	+
−	+	−	−
−	+	+	~
+	−	−	−
+	−	+	+
+	+	−	~
+	+	+	−
Business R&D share of GDP (%)	Business investment (€ per employee)	GERD per capita	Herfindahl
−	−	−	~
−	−	+	−
−	+	−	+
−	+	+	−
+	−	−	+
+	−	+	+
+	+	−	~
+	+	+	+
Business R&D share of GDP (%)	Business investment (€ per employee)	GERD per capita	Number of specialized fields
−	−	−	~
−	−	+	−
−	+	−	~
−	+	+	−
+	−	−	~
+	−	+	+
+	+	−	~
+	+	+	+

When it comes to the relationships Intermediate performance → Outcomes, the rule becomes:

If intermediate performance 1 decreases/increases and intermediate performance 2 decreases/increases and intermediate performance 3 decreases/increases and intermediate performance 4 decreases/increases and intermediate performance 5 decreases/increases and intermediate performance 6 decreases/increases and intermediate performance 7 decreases/increases and intermediate performance 8 decreases/increases,

Then outcome decreases/increases/does not vary significantly

If	***If***	***If***	***If***	***If***	***If***	***If***	***If***	***Then***
Intermediate performance (1)	**Intermediate performance (2)**	**Intermediate performance (3)**	**Intermediate performance (4)**	**Intermediate performance (5)**	**Intermediate performance (6)**	**Intermediate performance (7)**	**Intermediate performance (8)**	**Outcome**
Number of enterprises	Employment rate	Number of patents	Patents per capita	Unrelated variety	Related Variety	Herfindahl	Number of specialized fields	GDP per capita
~	−	~	+	~	~	+	~	−
~	~	~	−	~	~	~	−	~
~	+	~	+	~	+	−	+	+

4.3 Decision Support System 2 (K-means clustering with spatial variables)

The interpretation of this model is completely different from the first one. First, because all the 4 competitiveness drivers are present since what matters here is their exact value, rather than the difference. Second, because of the presence of the spatial variables, latitude and longitude. First, we decided to represent the geographical zones descending from the subtractive clustering algorithm performed by Matlab graphically, using the European map below (Figure 3). As it is possible to observe, some regions are not considered in the model. The reasons are mainly two: some entire nations were not present in the cleaned database (e.g., United Kingdom) because of their missing data; other regions were deleted because ANFIS did not capture their values as significant or distinct from the other clusters (e.g., regions of Sweden, Romania). Once created all the colored circles in the map, we used the zones written in the table below to replace the numerical coordination during the interpretation, in order to simplify the reading of the rules.

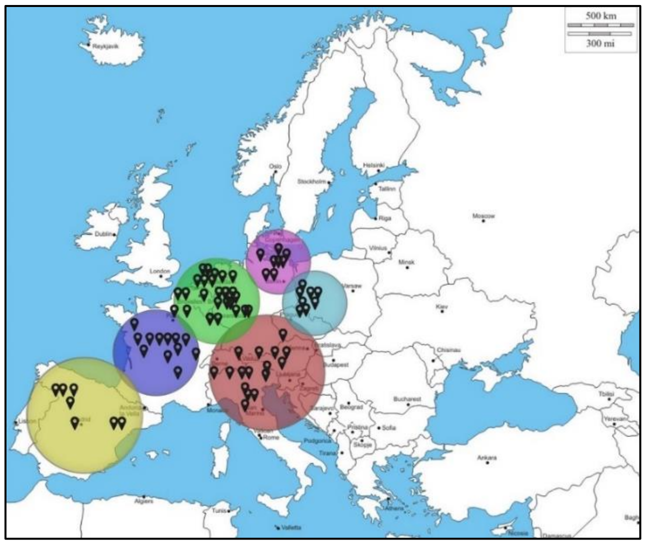

Figure 3. Geographical zones from ANFIS subtractive clustering (Yellow=Spain, Portugal; Dark blue=France; Green=Belgium, France, Germany, Luxembourg, The Netherlands; Light blue=Czech Republic, Poland; Red=Austria, Italy, Switzerland; Pink=Denmark, Germany)

Moreover, we created summarizing tables displaying the ranges of values corresponding to the labels LOW, MEDIUM, and HIGH used to interpret the rules following the fuzzy language (see Table 7). Because of the presence of the geographical zones, the language used for the interpretation of the first part of the DSS2 in object – Competitiveness Drivers + Spatial Variables → Intermediate Performance – has to be shown once. Taking as sample the first rule of the first variable below – Number of Enterprises – the rules is:

If we are in ZONE 3 and Input 1 is LOW and Input 2 is MEDIUM and Input 3 is LOW and Input 4 is MEDIUM, then output is LOW

Table 7. Range of values for LOW, MEDIUM, and HIGH labels

	Business R&D share of GDP (%)	Business investment (€ per employee)	Public R&D expenditure (%)	GERD per capita
LOW	<1	<10	<0.1	<300
MEDIUM	[1;2]	[10;17]	[0.1;0.25]	[300;700]
HIGH	>2	>17	>0.25	>700

	Number of enterprises	Employment rate	Number of patents	Patents per capita	Unrelated variety	Related Variety	Herfindahl	Number of specialized fields	GDP per capita
LOW	<100000	<65	<150	<100	<2	<2.3	<0.55	≤3	<20000
MEDIUM	[100000;500000]	[65;70]	[150;400]	[100;200]	[2;2.10]	[2.3;2.5]	[0.55;0.7]	[4;5]	[20000;30000]
HIGH	>500000	>70	>400	>200	>2.10	>2.5	>0.7	≥6	>30000

These are the results for the application of DSS2 (K-means clustering with spatial variables) to the relationships Competitiveness drivers + Spatial Variables → Intermediate performance (Table 8):

Table 8. Results for relationship Competitiveness drivers+Spatial variables →Intermediate performance

	If	*If*	*If*	*If*	*Then*
	Competitiveness driver (1)	**Competitiveness driver (2)**	**Competitiveness driver (3)**	**Competitiveness driver (4)**	**Intermediate performance**
Zone	Business R&D share of GDP (%)	Business investment (€ per employee)	Public R&D expenditure (%)	GERD per capita	Number of enterprises
33	LOW	MEDIUM	LOW	MEDIUM	LOW
22	LOW	MEDIUM	LOW	LOW	LOW
66	LOW	MEDIUM	HIGH	LOW	LOW
55	LOW	LOW	LOW	LOW	MEDIUM
4	MEDIUM	MEDIUM	MEDIUM	MEDIUM	LOW
4	HIGH	MEDIUM	MEDIUM	HIGH	LOW
4	LOW	HIGH	LOW	MEDIUM	HIGH
3	HIGH	MEDIUM	HIGH	HIGH	MEDIUM
1	LOW	HIGH	MEDIUM	LOW	MEDIUM
2	MEDIUM	MEDIUM	HIGH	MEDIUM	MEDIUM
Zone	Business R&D share of GDP (%)	Business investment (€ per employee)	Public R&D expenditure (%)	GERD per capita	Employment Rate
44	LOW	MEDIUM	MEDIUM	MEDIUM	MEDIUM
2	LOW	MEDIUM	LOW	LOW	LOW
3	MEDIUM	MEDIUM	LOW	MEDIUM	HIGH
4	LOW	HIGH	LOW	LOW	LOW
3	LOW	MEDIUM	HIGH	MEDIUM	HIGH
5	LOW	LOW	MEDIUM	LOW	LOW
2	MEDIUM	MEDIUM	MEDIUM	MEDIUM	LOW
6	LOW	MEDIUM	HIGH	LOW	MEDIUM
3	HIGH	MEDIUM	MEDIUM	HIGH	MEDIUM
1	LOW	HIGH	MEDIUM	LOW	LOW
Zone	Business R&D share of GDP (%)	Business investment (€ per employee)	Public R&D expenditure (%)	GERD per capita	Number of patents
3	LOW	MEDIUM	LOW	LOW	LOW
3	LOW	MEDIUM	MEDIUM	MEDIUM	HIGH
4	LOW	HIGH	LOW	MEDIUM	HIGH
5	LOW	LOW	MEDIUM	LOW	LOW
1	LOW	HIGH	MEDIUM	LOW	LOW
6	LOW	MEDIUM	HIGH	LOW	MEDIUM
4	MEDIUM	MEDIUM	LOW	MEDIUM	MEDIUM
3	LOW	MEDIUM	MEDIUM	MEDIUM	MEDIUM
4	HIGH	MEDIUM	MEDIUM	HIGH	HIGH
1	LOW	LOW	LOW	LOW	LOW
Zone	Business R&D share of GDP (%)	Business investment (€ per employee)	Public R&D expenditure (%)	GERD per capita	Patents per capita
3	LOW	MEDIUM	LOW	LOW	LOW
3	LOW	MEDIUM	LOW	MEDIUM	HIGH
4	LOW	HIGH	MEDIUM	MEDIUM	MEDIUM
5	LOW	LOW	MEDIUM	LOW	LOW
1	LOW	HIGH	MEDIUM	LOW	LOW
6	LOW	MEDIUM	HIGH	LOW	MEDIUM
2	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM
4	HIGH	MEDIUM	MEDIUM	HIGH	MEDIUM
3	HIGH	MEDIUM	HIGH	HIGH	HIGH
3	LOW	MEDIUM	HIGH	MEDIUM	MEDIUM
Zone	Business R&D share of GDP (%)	Business investment (€ per employee)	Public R&D expenditure (%)	GERD per capita	Unrelated Variety
3	LOW	MEDIUM	MEDIUM	MEDIUM	MEDIUM
4	LOW	HIGH	MEDIUM	MEDIUM	HIGH
3	LOW	MEDIUM	LOW	LOW	LOW
2	LOW	MEDIUM	MEDIUM	LOW	LOW
66	LOW	MEDIUM	HIGH	LOW	MEDIUM
3	HIGH	MEDIUM	MEDIUM	HIGH	MEDIUM
5	LOW	LOW	MEDIUM	LOW	LOW
3	LOW	MEDIUM	HIGH	HIGH	MEDIUM
1	LOW	HIGH	MIGIUM	LOW	LOW
2	MEDIUM	MEDIUM	HIGH	HIGH	HIGH
Zone	Business R&D share of GDP (%)	Business investment (€ per employee)	Public R&D expenditure (%)	GERD per capita	Related Variety

3	LOW	MEDIUM	MEDIUM	MEDIUM	HIGH
2	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM
4	LOW	HIGH	MEDIUM	MEDIUM	MEDIUM
6	LOW	MEDIUM	HIGH	LOW	MEDIUM
3	LOW	MEDIUM	LOW	LOW	HIGH
3	HIGH	MEDIUM	MEDIUM	HIGH	MEDIUM
3	LOW	MEDIUM	HIGH	HIGH	HIGH
5	LOW	LOW	MEDIUM	LOW	LOW
2	LOW	MEDIUM	MIGIUM	LOW	LOW
2	MEDIUM	MEDIUM	MEDIUM	HIGH	LOW

Zone	Business R&D share of GDP (%)	Business investment (€ per employee)	Public R&D expenditure (%)	GERD per capita	Herfindahl
3	LOW	MEDIUM	MEDIUM	MEDIUM	LOW
2	LOW	MEDIUM	MEDIUM	LOW	HIGH
4	LOW	HIGH	MEDIUM	MEDIUM	MEDIUM
6	LOW	MEDIUM	HIGH	LOW	MEDIUM
5	LOW	LOW	MEDIUM	LOW	HIGH
3	LOW	MEDIUM	HIGH	HIGH	LOW
4	MEDIUM	HIGH	MEDIUM	MEDIUM	HIGH
3	HIGH	MEDIUM	MEDIUM	HIGH	LOW
4	HIGH	MEDIUM	LOW	MEDIUM	MEDIUM
1	LOW	HIGH	MEDIUM	LOW	HIGH

Zone	Business R&D share of GDP (%)	Business investment (€ per employee)	Public R&D expenditure (%)	GERD per capita	Number of specialized fields
4	LOW	MEDIUM	MEDIUM	MEDIUM	LOW
3	LOW	MEDIUM	MEDIUM	MEDIUM	MEDIUM
3	LOW	MEDIUM	MEDIUM	MEDIUM	LOW
2	LOW	MEDIUM	MEDIUM	LOW	MEDIUM
5	LOW	LOW	MEDIUM	LOW	HIGH
6	LOW	MEDIUM	HIGH	LOW	LOW
3	HIGH	MEDIUM	MEDIUM	HIGH	LOW
3	MEDIUM	MEDIUM	LOW	MEDIUM	MEDIUM
4	LOW	HIGH	MEDIUM	LOW	MEDIUM
4	LOW	HIGH	MEDIUM	MEDIUM	LOW

When it comes to the relationship Intermediate performance → Outcomes we get (Table 9):

Table 9. Results for relationship Intermediate performance→ Outcomes

	If	***If***	***If***	***If***	***If***	***If***	***If***	***If***	***Then***
	Intermediate performance (1)	**Intermediate performance (2)**	**Intermediate performance (3)**	**Intermediate performance (4)**	**Intermediate performance (5)**	**Intermediate performance (6)**	**Intermediate performance (7)**	**Intermediate performance (8)**	**Outcome**
Rules	Number of enterprises	Employment rate	Number of patents	Patents per capita	Unrelated Variety	Related Variety	Herfindahl	Number of specialized fields	GDP per capita
1	LOW	HIGH	HIGH	HIGH	MEDIUM	HIGH	LOW	LOW	HIGH
2	MEDIUM	LOW	MEDIUM	LOW	MEDIUM	HIGH	MEDIUM	MEDIUM	LOW
3	LOW	HIGH	LOW	MEDIUM	HIGH	LOW	LOW	HIGH	HIGH
4	MEDIUM	LOW	HIGH	MEDIUM	HIGH	MEDIUM	MEDIUM	LOW	MEDIUM
5	LOW	MEDIUM	HIGH	HIGH	HIGH	MEDIUM	HIGH	LOW	LOW

5 Discussion and implications

Starting from the intermediate performance (representing the technological development of a region and including both the patent-based indicators and the socio-economic indicators), the Number of Enterprises increases only if all the competitiveness drivers simultaneously increase. Concerning the Employment Rate, the increasing of only one competitiveness driver – and the consequent decreasing of the others – causes its reduction. The Number of patents and Patent per capita behave in the same way, both increasing considerably if Business R&D share of GDP decreases and the other two competitiveness drivers increase; moreover, they decrease if Business R&D share of GDP and GERD per capita decrease and Business investment increases. Another important implication concerns the Unrelated Variety and Related Variety indicators: their behavior is nearly opposite. A correlation between these two indicators could be expected since they both are indicators of diversification. What is interesting is their opposite answers to the changings in the competitiveness drivers. This result let us suppose that, if in a region the EPO applications in one main class of the Fraunhofer domains increase – and then the Unrelated Variety decreases – it is likely that the number of domains within the main class, in which the EPO applications are deposited, increases, thus raising the value of the Related Variety.

The indicators of specialisation – Herfindahl and Number of specialized Fields – instead, react in a similar way to the changing in the competitiveness drivers: they both increase when Business R&D share of GDP and GERD per capita increase, independently from the changing of Business investment.

Concerning the second part of the Decision Support System – Intermediate performance and GDP per capita – the relevant variables with respect to the variation of the outcome have been: Employment Rate, Patents per capita, Related Variety, Herfindahl, Number of specialized fields. The other variables, as explained above in the steps of the methodology, presented overlapped Gaussian functions, thus their variation resulted no significant. Particularly, GDP per capita increases if Employment Rate, Patents per capita, Related Variety, and Number of specialized field increase while Herfindahl decreases. It is important to observe that the growth of Patents per capita is not a discriminant in the growth of GDP per capita: indeed, it increases both in the first and in the third rule, while GDP once decreases and once increases. This could mean that this indicator has to be considered with respect to the other patent-based indicators. Hence, if the Number of patents per capita increases together with the increasing of Herfindahl, GDP per capita tends to decrease; while, if the Number of patents per capita increases together with the decreasing of Herfindahl, GDP per capita tends to increase. Besides, concerning the positive influence of Patents per capita, it can be related also to the increase of Related Variety and Number of specialized fields, but this consideration does not apply to the negative influence, since these variables do not present substantial variations in the first rule. Finally, we investigated all the interpreted rules in order to discover if one or more combinations of all the variables – starting from the competitiveness drivers – existed, which could allow regions to increase their GDP per capita.

Overall, concerning the Decision Support System 1 (Model D without spatial variables at time T+1) we found that the competitiveness drivers which promote the technological development are Business investment and GERD per capita; increasing these two variables, Employment Rate, Patents per capita, and Related Variety increase as well. These latter variables are positively related to the growth of GDP per capita, enhancing the economic development of the regions. It is important to underline the relationship between Patent per capita and Herfindahl: if the number of patents per capita increases relating to a decrease of the specialisation index, GDP per capita increases, otherwise it decreases. For the DSS1, the best combination of variables is as follows:

	Business R&D share of GDP (%)	**Business investment (€ per employee)**	**GERD per capita**	**Employment rate**	**Patents per capita**	**Related Variety**	**Herfindahl**	**Number of specialized fields**
Rules	–	+	+	+	+	~	–	+

Concerning the Decision Support System 2 (K-means clustering with spatial variables at time T+1), regarding the first part namely, Competitiveness Drivers + Spatial Variables → Intermediate Performance, it is interesting to observe two things: first, the behavior of each output when in the same zone the inputs change: this occurrence mainly appears in zone 3 and zone 4 and sometimes in zone 2, for which several combinations of inputs per output are provided by the model; second, the different results, in terms of output, in different zones when the combination of inputs is almost the same. Analyzing the DSS from the first point of view, it is possible to highlight the influence of the inputs on the final results, independently from the geographical collocation: for example, in zone 3, a combination with most of the competitiveness drivers LOW corresponds to a low number of enterprises and low number of patents per capita, but if the competitiveness drivers become MEDIUM-HIGH, these values increase. Focusing, instead, on the second point of view, the reasoning is opposite: it is possible to evaluate the influence of the position of the various regions, basing on the zones, on the final results, comparing solutions with similar inputs. Concerning the second part of the Decision Support System, the Matlab Rule Viewer shows that Number of enterprises, Number of patents and unrelated variety have their Gaussian functions almost completely overlapped, hence their variations are not very relevant for our purpose. The overlapping of the functions for the variables Number of enterprises and Number of patents is due largely to the big variance, while for the Unrelated variety is due to the same mean. Hence, regarding Unrelated Variety we can affirm that it is irrelevant for the determination of the output, while the other two variables have a moderate influence on it. Employment rate, instead, has a positive correlation with GDP per capita: high values of the first correspond to high value of the second; other values are not particularly affecting the results. Concerning the patent-based variables, it is possible to verify an inverse correlation between the value of Herfindahl and the output; besides, the number of patents per capita has a positive impact on GDP only if a high number of patents per capita is related to a low value of Herfindahl. Vice versa, a high number of patents with a high value of Herfindahl causes a negative effect.

In the case of DSS2, because of the presence of different geographical zones, it is impossible to find a general combination to promote technological development and the growth of the regions the consecutive year. However, what is possible to observe is that the competitiveness drivers GERD is the most important in this sense: a high value of it leads to high value of technological development. However, the two DSSs can be used together as follows: by analyzing the results of DSS2 we can understand the structure of the competitiveness drivers in a region, understanding whether they are promoting the technological development or not. Then, thanks to the DSS1, it will be possible to take corrective measure, concentrating the forces on the right competitiveness drivers. In line with the rationale behind the Smart Specialisation program (RIS3 strategy) (European Commission, 2011) we provide a tool to better concentrate and link resources to a specific number of priorities through which regions can increasingly become competitive and innovative (e.g., Bartelsman et al., 1994; Krugman, 1991; McCann and Ortega-Argilés, 2015; Frenken et al., 2007). In this light, we were able to disentangle the role played by the competiveness drivers (consisting of the indicators of public and private R&D investment) and the intermediate performance (representing the technological development of a region and including both the patent-based indicators and the socio-economic indicators), by providing with an external validity check of the effectiveness of the indicators used within the framework of the European Cluster Observatory (European Commission, 2007). Differently from former studies (Xu and Xu, 2013; Van De Kaa et al., 2014), we propose an ANFIS-based DSS through which policy makers can find the optimal relationship between public and private R&D investments and indicators of technological development (both socio-economic and patent-based), improving the chances to better allocate scant resources and prioritize investments for the regional economic growth; we also extend our analyses to all industries and European countries.

6 Conclusion and future research

From this study, multiple conclusions can be drawn. The first concerns the appropriateness of the ANFIS method for the context of our study, based on the values of the errors and on the general goodness of all the models that have been built during the research. Therefore, ANFIS had provided adequate results with fitness scores above 90% in the forecasting models and around 2/3 of the regions classified in the right zone in the classification models. Concerning the results of the investigation, it has been possible to state that among the competitiveness drivers, Business Investment and GERD are the ones that enhance most of the technological development. The growth of a region, in terms of GDP per capita, is positively related to the Employment Rate regarding the socio-economic variables of the intermediate performance and to high values of Patents per capita, related to low values of the index of technological specialisation Herfindahl. In addition, the technological diversification indicator Related Variety seems to have a direct positive influence on GDP per capita. Hence, the investment on R&D in a region should be concentrated in different Fraunhofer domains of each main class, promoting a diversification strategy rather than a specialisation one. Indeed, if the number of patents per capita inside a region increases, with a view to the economic growth of the region, it is suggested that the additional patents (with respect to the preceding year) are in new or not consolidated technological domains.

This study does not come without limitations. First, although the model has been tested and created by leveraging on a relatively comprehensive database, it may lack generalizability and optimization capability; in order to improve on this point, future research could add to the ANFIS model a priori and a posteriori human knowledge, derived from the studies carried out on RIS3; this can be added in the form of hard coded rules. Second, a bigger time series is recommendable as carrying out a similar analysis on a larger number of years may allow for a better training and testing the models, (above all for what concerns the years between 2012 and 2016). Third, our analysis took considered the NUTS 2 level (i.e. basic regions for the application of regional policies) as a minimum level of analysis, while going deeper to the NUTS 3 level (i.e. small regions for specific diagnoses) can offer policy makers a finer grained view of the assessments and, eventually, a more powerful tool. Finally, other socio-economic and patent-based variables could be evaluated and put into the model in order to contribute with further evidence in supporting and refining the Smart Specialisation strategies.

Acknowledgements

We gratefully thank Dr. Chin-Wei Huang (Ecole Polytechnique Montréal) for the support with the methodological section.

References

Achiche, S., Shlechtingen, M., Raison, M., Baron, L. & Santos, I. F. (2016). Adaptive Neuro-Fuzzy Inference System Models for Force Prediction of a Mechatronic Flexible Structure. *Journal of Integrated Design & Process Science*, *19*, 77-94.

Achiche, S. & Ahmed-Kristensen, S. (2011). Genetic fuzzy modeling of user perception of three-dimensional shapes. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *25*, 93-107.

Adebiyi A. A., Adewumi A. O. & Ayo C. K. (2014). Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. *Journal of Applied Mathematics*, *2014*, 1-7.

Al-Hmouz, A., Shen, J., Al-Hmouz, R. & Yan, J. (2012). Modelling and simulation of an adaptive neuro-fuzzy inference system (ANFIS) for mobile learning. *IEEE Transactions on Learning Technologies*, *5*, 226-237.

Aleksandrova, V., Skrypnychenko, N., & Fedulova, L. (2007). Forecasting the impact of innovative factors on Ukraine's economy development. *Economy and the emergence of a new economic system*, *2*, 9-26.

Altuntas, S., Dereli, T., & Kusiak, A. (2015). Forecasting technology success based on patent data. *Technological Forecasting and Social Change, 96*, 202-214.

Bahar, D., Hausmann, R., & Hidalgo, C. A. (2014). Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion? *Journal of International Economics, 92*, 111–123.

Balazinski, M., Achiche, S. & Baron, L. (2000). Influences of Optimization Criteria on Genetically Generated Fuzzy Knowledge Bases. *Proceedings of the International Conference on Advanced Manufacturing Technology*, 159-164, Malaisia.

Bartels, F. L., Voss, H., Lederer, S., & Bachtrog, C. (2012). Determinants of National Innovation Systems: Policy implications for developing countries. *Innovation: Management, Policy & Practice*, *14*, 2-18.

Becker, H. S. (1983). Scenarios: a tool of growing importance to policy analysts in government and industry, *Technological Forecasting and Social Change*, *23*, 95–120.

Bennett, C., Rodney, A. S., & Lu, J. (2014). Autoregressive with Exogenous Variables and Neural Network Short-Term Load Forecast Models for Residential Low Voltage Distribution Networks. *Energies*, *7*, 2938-2960.

Benaroch, M., & Dahr, V. (1995). Controlling the complexity of investment decisions using qualitative reasoning techniques. *Decision Support Systems*, *156*, 115-131.

Boschma, R., Heimeriks, G., & Balland, P. (2014). Scientific knowledge dynamics and relatedness in biotech cities. *Research Policy, 43*, 107–114.

Boychuk, V. A., & Novakevych, V. Yu, V. (2014). Modern artificial neural network and approaches to their modeling. *Measuring and computing in technological processes,* 4, 216- 219.

Burinskiene, M., & Rudzkiene, V. (2004). Comparison of spatial-temporal regional development and sustainable development strategy in lithuania. *International Journal of Strategic Property Management*, *8*, 163-176.

Chen, Y. H., Chen, C. Y., & Lee, S. C. (2011). Technology forecasting and Patent Strategy of Hydrogen Energy and Fuel Cell. *International Journal of Hydrogen Energy, 36*, 6957-6969.

Chien, S.-C., Wang, T.-Y. & Lin, S.-L. (2010). Application of neuro-fuzzy networks to forecast innovation performance – The example of Taiwanese manufacturing industry. *Expert Systems with Applications*, *37*, 1086-1095.

Cocianu C.-L., & Grigoryan, H. (2015). An Artificial Neural Network for Data Forecasting Purposes. *Informatica Economică, 19*, 34-45.

Content, J. & Frenken, K. (2016). Related variety and economic development: a literature review. *European Planning Studies*, *24*, 2097-2112.

Cortinovis, N., & Van Oort, F. (2015). Variety, economic growth and knowledge-intensity of European regions: A spatial panel analysis. *Regional Studies*, *41*, 685–697.

Cuhls, K. (2003). From forecasting to foresight processes—new participative foresight activities in Germany. *Journal of Forecasting, 22*, 93-111.

De-Wang, C. & Zhang, J.-P. (2005). Time series prediction based on ensemble ANFIS. *Proceedings of the International Conference on Machine Learning and Cybernetics*, 6, 3552-3556, Guangzhou, China.

Doha, S., & Kimb, B. (2014). Government support for SME innovations in the regional industries: The case of government financial support program in South Korea. *Research Policy*, 43, 1557-1569.

Eerola, A., & Miles, I. (2011). Methods and tools contributing to FTA: A knowledge-based perspective. *Futures*, *43*, 265–278.

Ellison, G., & Glaeser, E. L. (1999). The geographic concentration of industry: does natural advantage explain agglomeration? *American Economic Review, 89*, 311-316.

European Commission (2007). European Cluster Observatory, http://www.clusterobservatory.eu/index.html

European Commission (2009). *Mapping Foresight: Revealing how Europe and Other World Regions Navigate into the Future*. European Commission, Brussels.

European Commission (2011). RIS3 strategy, smart specialization platform S3, http://s3platform.jrc.ec.europa.eu/

Fabbri, E. (2016). Strategic planning and foresight: the case of Smart Specialisation Strategy in Tuscany. *Foresight, 18*, 491-508.

Fernandez, E., Navarro, J., Duarte, A., & Ibarra, G. (2013). Core: A decision support system for regional competitiveness analysis based on multi-criteria sorting. *Decision Support Systems, 54*, 1417-1426.

Ferretti, V., & Montibeller, G. (2016). Key challenges and meta-choices in designing and applying multi-criteria spatial decision support systems. *Decision Support Systems*, *84*, 41-52.

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies, 41*, 685-697.

Garcìa, J. C. F., & Mendez, J. J. S. (2007). A Comparison of ANFIS, ANN and DBR systems on volatile Time Series Identification. *Proceedings of Fuzzy Information Processing Society, NAFIPS '07, Annual Meeting of the North American*, 319-324.

Gottschilich, J., & Hinz, O. (2014). A decision support system for stock investment recommendations using collective wisdom. *Decision Support Systems*, *59*, 52-62.

Harper, J. C., & Georghiou, L. (2005). The targeted and unforeseen impacts of foresight on innovation policy: the eFORESEE Malta case study. *International Journal of Foresight and Innovation Policy*, *2*, 84-103.

Hidalgo, C. A., Klinger, B., Barabasi, A. L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science, 317,* 482–487.

Hsiao, C.-T., & Liu, C.-S. (2012). Dynamic modeling of the development of the DRAM industry in Taiwan. *Asian Journal of Technology Innovation*, *20*, 277-293.

Intepe, G., & Koc, T. (2012). The Use of S Curves in Technology Forecasting and its Application On 3D TV Technology. *International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering, 6*, 11-28.

Jang, J. S. R. (1993). Adaptive-Network-Based Fuzzy Inference System, *IEEE Transactions on Systems, Man, and Cybernetics, 23*, 665-685.

Jun, S.-P., Kim, S.-G., & Park, H.-W. (2017). The mismatch between demand and beneficiaries of R&D support programs for SMEs: Evidence from Korean R&D planning programs. *Technological Forecasting and Social Change*, *116*, 286-298.

Kim, B. S. (2010). A case of forecast-based technology evaluation and its implications. *International Journal of Technology Intelligence and Planning*, *6*, 317-325.

Könnölä, T., Scapolo, F., Desruelle, P., & Mud, R. (2011). Foresight tackling societal challenges: Impacts and implications on policy-making. *Futures, 43*, 252–264.

Krugman, P. (1991). *Geography and Trade*. The MIT Press: Cambridge, MA.

Le Gallo, J. & Kamarianakis, Y. (2011). The Evolution of Regional Productivity Disparities in the European Union from 1975 to 2002: A Combination of Shift–Share and Spatial Econometrics. *Regional Studies*, *45*, 123-139.

Linstone, H. A., & Turoff, M. (2011). Delphi: A brief look backward and forward. *Technological Forecasting and Social Change*, *78,* 1712–1719.

Lourenço, J.C., Morton, A., & Bana E Costa, C.A. (2012). PROBE – A multicriteria decision support system for portfolio robustness evaluation. *Decision Support Systems*, *54*, 534-550.

Maranto-Vargas, D., & Rocío Gómez-Tagle, R. (2007). Development of internal resources and capabilities as sources of differentiation of SME under increased global competition: A field study in Mexico. *Technological Forecasting and Social Change, 74*, 90–99.

Marchais-Roubelat, A., & Roubelat, F. (2011). The Delphi method as a ritual: Inquiring the Delphic Oracle. *Technological Forecasting and Social Change, 78*, 1491–1499.

McCann, P., & Ortega-Argilés, R. (2015). Smart Specialization, Regional Growth and Applications to European Union Cohesion Policy. *Regional Studies*, *49*, 1291-1302.

Nayak, P.C (2010). Explaining Internal Behavior in a Fuzzy If-Then Rule-Based Flood-Forecasting Model. *Journal of Hydrogeologic Engineering, 15*, 20-28.

OECD (2013). *Innovation-driven growth in regions: the role of Smart Specialization*. Available online: https://www.oecd.org/innovation/inno/smart-specialisation.pdf

Okuwada, Y. Y. K., (2013). Identifying expected areas of future innovation by combining foresight outputs. *Foresight, 15*, 6-18.

Paliokaitė, A., Martinaitis, Z., & Sarpong, D. (2016). Implementing smart specialisation roadmaps in Lithuania: Lost in translation? *Technological Forecasting and Social Change, 110*, 143–152.

Paliokaitė, A., Martinaitis Z., & Reimeris, R. (2015). Foresight methods for smart specialisation Strategy development in Lithuania. *Technological Forecasting and Social Change*, *101*, 185–199.

Piirainen, K. A., Tanner, A. N., & Alkærsig, L. (2017). Regional foresight and dynamics of smart specialization: A typology of regional diversification patterns. *Technological Forecasting and Social Change*, *115*, 289-300.

Reichardt K. O., Negro, S. O., Rogge, K. S., & Hekkert, M. P. (2016). Analyzing interdependencies between policy mixes and technological innovation systems: The case of offshore wind in Germany. *Technological Forecasting and Social Change*, *106*, 11–21.

Robertson, A., Soopramanien, D., & Fildes, R. (2007). Segmental new-product diffusion of residential broadband services. *Telecommunications Policy, 31*, 265–275.

Robinson, D. K. R., & Propp, T. (2008). Multi-path mapping for alignment strategies in emerging science and technologies. *Technological Forecasting and Social Change, 75*, 517–553.

Saracoglu, B.O. (2013). Selecting industrial investment locations in master plans of countries. *European Journal of Industrial Engineering, 7*, 416-441.

Sarkkinen, M., & Tuomo Kässi, T. (2014). Strategic decision-making in the context of a regional innovation system. *International Journal of Business and Systems Research, 8*, 295-314.

Song, B. Y. B. (2014). A systematic approach of partner selection for open innovation. *Industrial Management & Data Systems, 114*, 1068-1093.

Sugheir, J., Phan, P. H., & Hasan, I. (2012). Diversification and innovation revisited: An absorptive capacity view of technological knowledge creation. *IEEE Transactions on Engineering Management, 59*, 530-539.

Takagi, T., & Sugeno, M. (1983). Derivation of fuzzy control rules from human operator's control actions. IFAC Proceedings Volumes, 16, 55-60.

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics, 15*, 116-132.

Van De Kaa, G., Van Heck, E., De Vries, H. J., Van Den Ende, J., & Rezaei, J. (2014). Supporting decision making in technology standards battles based on a fuzzy analytic hierarchy process. IEEE *Transactions on Engineering Management, 61*, 336-348.

Xu, E. & Xu, K. (2013). A Multilevel analysis of the effect of taxation incentives on innovation performance. *IEEE Transactions on Engineering Management, 60*, 137-147.

Wang, Y., Ning, L., & Prevezer, M. (2015). Technological diversification in China from 1986 to 2011: Evidence from patent data. *Technological Forecasting and Social Change, 99*, 54–66.

Weber, K. M., & Schaper-Rinkel, P. (2016). European sectoral innovation foresight: Identifying emerging cross-sectoral patterns and policy issues. *Technological Forecasting and Social Change*, *115*, 240-250.

Yücel, G., & Van Daalen, C. E. (2011). Exploratory analysis of the impact of information dynamics on innovation diffusion. *Technological Forecasting and Social Change, 78*, 358–372.

Yurimoto, S., & Masui, T. (1995). Design of a decision support system for overseas plant location in the EC. *International Journal of Production Economics, 41*, 411-418.

Yurynets, Z. (2016). Forecasting model and assessment of the innovative and scientific-technical policy of Ukraine in the sphere of innovative economy formation. *Investment Management and Financial Innovations, 13*, 16-23.

Reviewer 1

Throughout the entire text, a few sentence construction errors were observed. Authors are requested to rectify the errors.

We carefully revised the manuscript in order to free it from mistakes in sentence construction.

The Introduction section has been written quite well. The literature review has also been done with care and research gap has been identified. Recent research articles have also been cited.

Thanks for the positive remarks on literature review, research gap identification, and citations of proper extant research.

However, there are some issues which are required to be mentioned. First of all, the overall checking clearly shows that Table 3 is not present. After Table 2, the authors have referred Table 4. Besides, after Table 5, the subsequent tables have not been numbered.

Thanks for suggesting these modifications. Accordingly, we renamed all the tables and made sure that they were all cited into the text.

The calculations of RMS and FP have not been shown. The authors have directly started showing results after the calculations of RMS and FP which has resulted in confusion about the originality of the results. The authors have directly shown the 'final result' without showing the prior calculations which is not acceptable.

Thanks for asking this. We put an excerpt of the intermediate calculations through which we get to the identification of the best decision support systems.

The original data on which the calculations of RMS and FP have been done, have not been shown. Thus the authors should also show, at least, a sample of the original data set.

An excerpt of the original data and indicators has been put into the text (see Table 5). We avoided to introduce data concerning latitude and longitude for the region in question in order to increase the readability of the excerpt.

The authors could mention how competitiveness drivers and the spatial variables have been identified for the final results.

Spatial variables refer to longitude and latitude of the EU regions. This is publicly available data. Concerning the Competitiveness Drivers we derived them from the Cluster Observatory www.clusterobservatory.eu and these are also publicly available data. We mentioned this website in the text in order to make our procedure replicable and falsifiable.

The result section is not properly presented. More explanations for the tables are required. All of the tables have not been referred in the text.

Tables in the results section just show the if-then rules and relationships among the main variables. We do think they have been improved from the very first version of the paper. The introductory paragraph of each Tables in section 4.2 explain the content in the most basic way possible.

Reviewer 2

The authors have done changes in making the manuscript more readable and coherent. This is commendable. The research question is explicated better and the literature survey is presented better.

Thanks for these remarks.

At the same time, the flow between sections needs to be more logical and connected. The presentation may need an explicit theory development and research model section which can connect the literature survey and translate the research questions to specific hypotheses / propositions so that the specific parameters used in the methods section are seen as sufficient to for test the hypotheses. In this sense some of the previous round comments still need to be addressed. Some specific comments are given for below.

We see your point. We rewrote some sections in order to increase the logical flow among paragraphs. However, we would like to emphasize the need to keep the two research questions instead of formulating hypotheses. Our study is of an exploratory nature and tries to come up with an ANFIS-based DSS in order to support policy makers with their decision making processes. In this particular context, the core is not hypothesis testing; rather, it is providing proof that the DSS works and performs well. Our attention is more on the functioning of the tool rather than on theory building. This is why it was a reasoned (upfront) choice to not generate testable hypotheses, still keeping clear the two leading research questions.

Page 3 , Section 1: "…increasing globalization…" : Globalization is under threat in the current conditions of increased protectionism and trade tariffs, and in Europe, Brexit indicates a move to disaggregation. So the opening statement is going against the face of current realities. Authors can choose to reposition this statement.

We agree on this remark. We completely rephrased the first sentences going straight to the point of Smart Specialisation.

Page 3, Section 1: "…aim of the work presented in this paper is to explore the extent to which technological specialization and technological diversification pay off…" Given that the aim statement mentions technological specialization and technological diversification specifically, these terms do not appear connected anywhere else in the document, and not even in the conclusion section.

Thanks for this remark, but this is not entirely correct. Perhaps are you making reference to other terms ? If I refer to technological specialization (measured through herfindahl and number of specialized fields) and technological diversification (measured through related variety and unrelated variety), we mentioned it in the Introduction; then, in the Research Setting of the Method section, where we clearly disentangle between specialization and diversification. We also highlight it in Figure 2 where specialization (Herfindahl and number of specialized fields) and diversification (unrelated and related variety) show up. They are also highlighted in the Results section. In the Discussion and Implications we discuss them, by calling them unrelated variety, related variety, herfindahl, and number of specialized fields in order to provide with a finer grained view. They also show up in the Conclusion section when pointing out the main outcome of our research. Technological specialization and diversification is the backbone of our paper and concrete elements concerning both of them are present in each section.

Page 3 , Section 1: "…framework advanced…" : What is this framework? It is not clear from the link provided in the reference.

You are right, how it was written created some confusion. We clarified as follows:

Specifically, we rely on the European Cluster Observatory initiative, …

Page 3 , Section 1, Footnote 2: "…as data on technology are more disaggregated (which may be considered as a combination of technologies)…" :

Still not clear what this means (with reference to comment in last review). Do you mean products are an aggregated combination of technology? Also, why is this point of relevance?

Yes, sure we are happy to clarify. A product can be considered a combination of technologies. Collected patent data go down to the level of technological components providing with a more fine-grained view of they impact on diversification/specialization and in turn on regional growth. It is the very nature of data that lukily allows us to focus on a lower level of analysis. On the contrary, running the same analysis with product data (which are potentially assemblies of technologies) can bias the calculation of specialization and diversification indicators.

Page 3 , Section 1: "…this study is the first one to build on to ANFIS-based DSSs…":

This is a rather bold statement. I am not able to verify the validity of the statement. The authors may consider rewording to avoid a possible denial of this singular assertion.

We rephrased as follows:

… this study is one of the few studies trying to build up ANFIS-based DSSs …

Section 2:

Overall, the relevance of this literature review to the research questions is not clear.

This somewhat contradicts what you stated at the beginning «The research question is explicated better and the literature survey is presented better. » but of course, we can improve it.

A section on theory development and research model may help to connect. At present, the connect is not clear. Further, the method section uses aspects which have not been explicated earlier.

Page 6: "…Grand challenges…":

The previous comments on inclusion of this remains. Not clarified. The reasoning for presenting it different from the RIS3 is not clear as it is also part of the shared vision of RIS3 steps. The authors have chosen global health specific definition from https://grandchallenges.org/#/map, where as RIS3 guide suggests a more broader description of "societal inclusive, environmental and sustainable economic development" in the RIS3 guide annex III".

Yes, we agree with your suggestion. Although this was something emerging from the literature review, we realized that the scope of RIS3 covers that of the GrandChallenges as well. We definitely deleted the paragraph since it creates confusion for the reader.

Page 6: "…comprehensive explanation of the four major emerging topics that form the …" The presentation does not explain, at least to my understanding, why these four are important to look at for the research and not others.

As we were highlighting in Figure 1, these three topics (we eliminated Grand Challenges since they are highily related to RIS3 strategy) emerge out of our literature review. Obviously, there may be other research topics which can connect, but it is not the case for what concerns extant literature we analyzed.

Page 10, Section 2.3: "…regional foresight seems to be less accurate than forecasting techniques…" This statement can be heavily contested. Cuhls 2003 does not mention the accuracy of either and indicates that foresight goes beyond forecast. The previous comment on bringing out how forecast is difference from foresight in the context of this particular study remains especially since policy making is tending more towards foresight techniques. Explicating forecast as a useful tool for future planning may be necessary for this paper. Also, do you think forecasting technique you are presenting here could be part of an overall foresight process?

Yes, we agree that our statement was too direct. We rephrased by taking into account the fact that the limitation of current RIS3 strategies and policy making actions is exactly the lack of sound forecasting techniques/tools. Foresight is already part of the RIS 3 formulation, but it is different from what we are trying to do. Of course, forecasting techniques can better inform the formulation of a strategic foresight plan, and we mention this in the Conclusion section. However, we focus on the forecasting technique, keeping foresight as a higher level/vision-oriented excercise. We rephrased as follows:

Research concerning Smart Specialisation Strategies (RIS3) is relatively recent. Authors mainly used empirical approaches and a combination of qualitative and quantitative techniques to perform their research. RIS3 envisages the concentration of resources on specific – limited – set of priorities; this approach, which is thought to be as objective and quantifiable as possible, emerged as a consequence of public resources misuses and lack of transparency in public funding decisions (Paliokaitė et al., 2015). Piirainen et al. (2017) argue that in the context of smart specialisation, regional foresight tends to be linked to the development of the policy-making agenda only at an higher level; indeed, while foresight is based on more qualitative information and data (e.g., Nayak, 2010; Cuhls, 2003) making it more prone to opinions when it comes to questioning how the future looks like in the specific area under investigation, forecasting techniques (by their very quantitative nature) can better inform policy makers in setting the R&D agenda leveraging on the most promising (emerging) technological opportunities (e.g., McCann and Ortega-Argilés, 2015; Paliokaitė et al., 2015; Harper and Georghiou,2005; Piirainen et al., 2017). Foresight and forecasting are not mutually exclusive; rather, results from forecasting techniques can inform a better formulation of foresight processes (and vice versa). Paliokaitė et al. (2015), with the purpose of identifying RIS3 priorities, have proposed a combination of analytical and participatory methods, combining both qualitative techniques (qualitative analysis of trend and challenges, surveys, Delphi Method, scenarios and roadmaps) and quantitative ones (statistical analysis, bibliometrics, multiple criteria analysis). Building on their first paper, Paliokaitė et al. (2016) performed another research by focusing on the implementation of roadmaps in Lithuania, built through the opinions of Experts panels, for the implementation of selected RIS3 priorities. Reichardt et al. (2016) have made an empirical research based on patent data focusing on the offshore wind service sector in four regions around the North Sea; they contributed to the smart specialization literature to the extent that "the empirically corroborated typology of diversification patterns that can be used as an analytical framework for both analysis and anticipation." Finally, Fabbri (2016) proposed an innovative methodological scheme for strategic planning decisions concerning the RIS3 strategies; the scheme concerned is based on foresight, road mapping and large participation processes of experts and stakeholders. However, in the context of our study the focus is on forecasting techniques since there is a lack of attention from extant literature on how such techniques can inform policy making within the RIS3 framework.

Page 11, Section 2.4 : " …. Contributions belonging to this field are homogeneously…" Which field?

As we were mentioning in Figure 1 and at the beginning of section 2 (Literature Review), we decided to categorize emerging contributions according to some recurrent topics like Technique, Field, and Approach. Field refers to the general argument of the article that the statistical techniques are applied to and they are RIS3 strategy, Resources allocation, Technology and Innovation development.

Page 12: "These four sections serve to understand the link between the two research questions, RQ1 and RQ2, previously stated in the introduction section." The connection is not self-evident from the presentation of literature survey.

Yes we agree. For that reason, we wrote a paragraph better clarifying this point. It goes like this:

Overall, the former three sections introduce the reader with the major emerging topics characterizing the theoretical background of our study. These three sections are instrumental to create the link with the two research questions (RQ1 and RQ2) stated upfront in the Introduction section. Concerning RQ1 (what is the best combination possible of competitiveness drivers, intermediate performance indicators, and regional economic growth?), the three sections highlight the need to link the regional economic growth with some antecedents; precisely, while RIS3 deal with the general rationale behind European actions for development and growth, Resources Allocation identifies the need to put emphasis on finding smart ways to allocate scant resources for specific actions able to stimulate and sustain the regional economic growth. The third section on Technology and Innovation development emphasizes the need to pay attention on the antecedents of the regional economic growth that can be inflected in terms of competitiveness drivers and intermediate performance. The competitiveness drivers are a collection of specialization indicators, firms behavior indicators, and business environment indicators; whereas intermediate performance mainly deal with indicators measuring the employment rate, number of enterprises, enterprises growth, number of patents, scientific publications, and labour productivity. Once selected the indicators, we can be able to answer RQ2 (can we design a DSS to help policy makers to better invest public financial resources driving the economic growth of their regions?) by building up an ANFIS-based DSS and testing it in order to check whether – and to what extent – the resulting forecasting can support policy makers in improving their investments decisions.