Territorial Public Policies Evaluation Models (TPPEM)

Jorge ROCHA
Paulo MORGADO
Nuno COSTA
Eduarda COSTA
Centre of Geographic Studies, Lisbon University
Alameda da Universidade, 1600–214 Lisboa
Tel: (+ 351) 21 794 02 18,
e-mails: jrocha@fl.ul.pt, pms@fl.ul.pt, eduardamcosta@netcabo.pt, nunomcosta@netcabo.pt

Paulo NETO
University of Évora
Largo dos Colegiais 2, 7000-803 Évora
e-mail: neto@uevora.pt

Abstract

In this article we aim to define and build an evaluation model of territorial public policies. To pursue this goal it has been crucial the establishment of a permanent dialog between geography, economy, data analysis and Geographic information systems (GIS). In effect, it has been fruit of this interdisciplinary that the idea of a more stretched project has emerged: build an objected-oriented territory heritage data warehouse of public policies for regional development. The warehouse should allow decision making for data long series and a critical interpretation of the successful and unsuccessful knowing cases taking place on the territory. As results, we hope being able to appoint new relation and organization constellations, taking full advantage of new type of "proximities" between agents. For this article, we just will discuss and present the interactivity spatial data analysis and GIS methodologies that feed and support part of the project that concerns with these models.

Keywords: GIS, fuzzy logic, multicriteria evaluation model, self organizing feature map (SOFM), territorial public policies and decision support systems.

Resumo

Neste artigo pretende-se definir e construir um modelo da avaliação de políticas públicas territoriais. Para concretizar este objectivo foi crucial estabelecer uma ligação permanente entre a geografia, a economia, a análise de dados e os sistemas de informação geográfica (SIG). De facto, foi face a esta interdisciplinaridade que a ideia de um projecto mais abrangente emergiu: construir uma base de dados orientada para objecto da informação histórica do território e das políticas públicas para o desenvolvimento regional. A base de dados deve permitir a tomada de decisões com base em séries longas de dados e uma interpretação crítica dos casos conhecidos, bem sucedidos ou mal sucedidos que ocorrem no território. Como resultados, espera-se poder apontar as novas constelações da relação e da organização, fazendo uma análise completa da vantagem deste novo tipo de "proximidades" entre agentes. Neste artigo, apenas se discutirá e apresentará a interactividade, a análise de dados e as metodologias espaciais dos SIG que alimentam e suportam a parte do projecto que concerne aos modelos.

Palavras-Chave: SIG, lógica *fuzzy*, modelo de avaliação multicritério, *self organizing feature map* (SOFM), políticas públicas territoriais e sistemas de suporte à decisão.

Résumé

En cet article nous visons à définir et établir un modèle d'évaluation des politiques publics territoriaux. Pour poursuivre ce but il a été crucial l'établissement d'un dialogue permanent entre la géographie, l'économie, l'analyse de données et les systèmes d'information géographiques (SIG). En effet, c'a été fruit de cet interdisciplinaire que l'idée d'un projet plus étiré a émergé: construire une base de données objecte orienté de données d'héritage de territoire des aménagements publics pour le développement régional. Cette base de données devrait permettre la série de prise de décision longtemps et une interprétation critique des cas sachant réussis et non réussis ayant lieu sur le territoire. Comme résultat, nous espérons pouvoir nommer de nouvelles constellations de relation et d'organisation, profitant pleinement de nouveau type de "proximités" entre les agents. Pour cet article, nous juste discuterons et présenterons à l'interactivité l'analyse de données et les méthodologies spatiales de SIG qui alimentent et soutenons une partie du projet qui concerne par les modèles.

Mots-clés: SIG, logique floue, modèle d'évaluation multicritère, cartes auto- -organisatrices (SOFM), politiques publics territoriaux et systèmes interactifs d'aide à la décision.

1. Introduction

Territory competitiveness is a growing issue. Hence, a more precise and detailed knowledge of the markets, as well the strategic guidelines for the territory and the allocation options for business companies, or even the public policies and the individual strategies highlighted for the territory remain fundamental issues to a multicriteria evaluation model, capable to refine the decision making processes.

Since the adoption of the UE policies that new ways of territorial competitiveness as been stimulated, based on increased grid relationships between territories of course, but also between institutions (public, private and NGO's), business companies and people. Overcome for those UE policies, national and regional implications on several critical government policies issues have arise, such as: land use and planning, social and economic, environment, demographic, science and technologies, transports and mobility, among others.

To face this new collective territorial development challenge, national and regional policies have to be drawn, or, to be more precise about what have to be done: planning and models have to be define, discussed, build and re-build as changes occurs, so development strategies and decisions should be mores efficiently. Classic planning methodologies, based on a collection of master plans, lack an intergovernmental coordination that is fundamental to solve problems beyond the local level. To keep shortly, two things remain crucial: spatial data (or even non-spatial), that represents information, and science and technologies which are the tools for the implementation of new public policies. In fact, technologies can play an important role, as they integrate, treat and cross a great amount of information with ease and speed, an impossible task for traditional methodologies.

For integrated territory planning, spatial data analysis and GIS are the tools with more capacity to support, manage, and produce knowledge so give more accurate decision-making. In this article we will be focus on: the public policies (wish indicators should integrate the heritage warehouse of public policies for regional development); generalized concepts about data analysis (visualizing data, exploring and operating and modeling data); Spatial data analysis and GIS (the contribution of a more interactivity between GIS and spatial data analysis from statistics domains); the TPPEM (we present and explain the techniques and methods used to define and build the model) and finally the mains results and further developments.

2. Public Policies for a Regional Development

For public policies we understand the study's, plans, law's, actions and decisions that in a direct way influence the people and people's life. When we make an internet search on theme "public policies", we can find an amazing amount of domains that highlight it. From the social and economic, no matter the diverse sectors, to the environmental and technological domains, all stress public policies as the cause, effect or the leitmotiv to the successful, unsuccessful, growing, development, collapse or even stationery phase of them, as consequently of the territory/region compare to other's. Many author's, and academic experts from different scientific domains, eg. Lopes, Orsini, Peyrefitte, Gilbert, Hall, among many others, have already pointed the role and importance of public policies for regional development. For the project, our bigger concern lies on the selection for the model.

In fact the research, chosen and acquire of data that should integrate the objected-oriented territory heritage data warehouse of public policies for regional development is undoubted the phase that as consume most of time given to it. Nevertheless, concerning that the aim of this paper is about GIS and spatial data analysis, ie. not to much about what we have chose, but more about how we can operate it, public policies issues are just superficial presented, although we have already a very concise and theoretical support idea of what should integrate the warehouse of the territory in study.

At this moment, we think that should integrate the warehouse, in what concern public policies, the following issues:

- studies and an inventory of the actions that occurred in the territory for each type of problems, goals or evaluation of the results reported. The public policies evaluation should lie on different type of indicators (applied indicators - actions; indicators of behavior (operational); indicators of results (strategic goals); development indicators (general or sector goals). Behind the discussion about the efficiency of each public policy, it's also important to evaluate it pertinence and coherence.
- acquisition and analysis of other's public policies, with the same goals from the one's we want to implement – benchmarking - verified in other's territory's;
- appointment and observation of the on going public policies in the territory;
- appointment and observation of the on going public policies in the territory by the different sectors with direct or indirect relevance for other's sector's;
- frame each public policy that influence, as a cause or an effect the territory;
- find the exact scale for each territory intervention and each articulation model designed;

- evaluation of the vertical and horizontal articulation possibilities with different public policies and political instruments (define and appoint overlay areas, re-enforce holistic actions; autonomous interventions areas, define and appoint areas and themes not yet considered in the policies developed or applied);
- define and appoint, for each territory, the economic agents with territorial influence;
- evaluation of the sate of arte, in what concerns public policies;
- design possible socio-economic scenarios;
- validation method through the possible options of public, private or public--private policies;
- value and stimulate partnerships when it comes the moment to implement and execute public policies, but also be able to individualize the unity, the institution and the decision;
- appointment of the better political level and the administrative management unity, responsible for the policies and linking mechanic between it;
- define and appoint possible sponsor and stakeholders;
- design an evaluation model for territory public policies;
- scan for possible executants/beneficiates of the public policies developed in the territory.

In conclusion, we could say that an object-oriented territory heritage data warehouse, as well an evaluation model of territorial public policies could emerge as fundamental instruments to define long term decision for territorial development, built upon (build on) a cultural public participation.

From this moment our concern lays on two mains questions: how can we integrate and operate all this data? And, how can we territorialize this data and analyze it? Once again, the answer come from the technology industry; with have been through the years improving the functionality of software's and hardware, capable to deal with these problems. Nevertheless, the role of technologies in the planning and development issues is based in the "theory of iceberg": still are much more work to do, than the one we can see or expected. In fact, one main obstacle has already been overstep: politicians are already convinced the importance of technologies, and in our particular case the geographic information technologies (GIT), for the support decision making, as way to achieve the aimed sustainable development.

GIT have been assuming a crucial role in the regional development processes, as an instrument with the faculty to capture, acquire, save and manage knowledge, as well to implement over the territory ways of advanced inter--relationships of learning, experience and information between economic and social agents.

3. Spatial Data Analysis and GIS

Many author's, from different and honorable school's, have already study and write about the concept and the possible applications of spatial analysis and GIS. From Goodchild, to Batty, Longley, Dangermond, Tomlinson among many other's, no matter the fact of nonexistent common agreement about the concept of GIS, all community agrees that it's due to GIS that data as become spatial data analysis or geodata analysis, once data sets become spatially referenced. That's what makes de difference between statistical analysis and spatial analysis. In fact, GIS have allowed introduce new concepts and methods to geographical sciences – concepts of complex data models, interactive mapping, integrating data and spatial analysis, modeling, visualization and geoprocessing. But let us go by parts: first we should establish what we understand by spatial data analysis.

Spatial data analysis involves the accurate description of data relating to a process operating in space, as well the exploration of patterns and relationships among it. So, in what relies on GIS context, spatially data analyses are what come after: After overcome the initial phase of digitize, editing and validating geographic information and create the features that are now prepare to be into the GIS project, we start to make the information useful. Manage, cross, add, delete, modify, intersect, union, merge, overlay, measure, scripting, are only some of the functions that we have at our disposal. As Andy Mitchel, says, now is "...where the GIS rubber hits the road – where all the hard work of digitizing, building a database, checking for errors, and dealing with the details of projections and coordinate systems finally pays off in results and better decisions."

At this moment, we could ask the system to give me some figures, as well display's based on known basic questions: Where? ...is, or it stands. That is a question of basic curiosity, but fundamental to know what spatial objects/phenomenon's are in an exactly geographic position. Wish leads us to next question;

What? ...we could find in a particular place. Not less important than the first, this question allows us to make an inventory of the objects/phenomenon that exists in that particular place.

How? ... We can establish connection between the objects and their attributes, either they are more physical, either they are more abstract. If the first two questions only require simple queries functions, that one requires already basic principles of spatial analysis and computers knowledge about the software, as well the information.

When? ...did the phenomenon's occurred, or the transformations have started, or .... This question transports us to the time machine. As important the spatial dimension, temporal dimension allow us monitoring the territory and what is happening a crossing the time.

Ware if? ...this is the present-future of GIS spatial analysis and geo-computing. The spatial simulation based on complex models as cellular automated, neural network and artificial intelligence. That reality models are starting to have a fundamental role in the planning of risks environment, urban sprawl, epidemic diesis, transport and communication influences in territory configuration, etc.

Although all of this techniques and methods (besides theoretical contribution), are fundamental tools to create knowledge from a simple amount of data. Those techniques and methods are interactivity among statistical data analysis and GIS, so we could be able to operate and process chosen data, relate it to space and look for relationships in a way that we could fund new territorial constellations (strong and sustainable relationships) that facility decision support.

Nevertheless, in what concern methodologies taking place, a distinction have to be made between methods that are essentially for visualizing spatial data, to those which are exploratory, concerning measures and investigating relationships, and even to those rely on the modeling of reality. For us, there are a closely interplay of all three, as we can see further, but the modeling methods are the core of the project that we work on.

In conclusion we could say that GIS do not integrate must of the statistical functionality of some statistical programs nevertheless the improvement that have occur over the last decade on GIS software industry. Still, must of the GIS programs allow macro language, so technicians could introduce more functions into the program, namely in spatial analysis section.

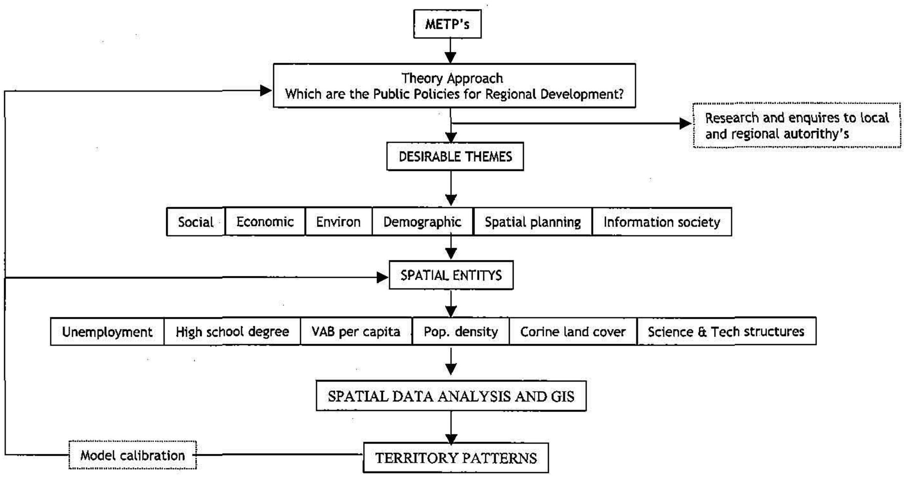
4. An Evaluation Model of Territorial Public Policies

A model is always a partial view of the reality; the result of a generalization of the world as we see it. Nevertheless, this are the way that we found to give a logical consistency and analytical form to some theory argued in different theses, as the one's we have study concerning to regional public policies of development. The model (Figure 1) always will be an imperfect representation of the reality, due to the artificial degree of the adopted considerations (no matter the multidisplinarity of the working team) which mask the real behavior of phenomena's.

Although, two main reasons could be accorded:

- the modeling exercise, today's improved by the high tech and scientific advances, come as the better way for multivariate and multicriteria reality evaluation. Is through him that we in fact could confront time (past with present, and even simulate future and make projections and prospective diagnosis) and space (multi scale exercise).

Figure 1 – TPPEM conceptual methodology diagram



- also due to the technology improvement, new emerged techniques and methods, as the one we use: spatial data analysis and GIS, model's are the more dynamic and flexible techniques and methods so we can monitoring reality systems and interview in real time, so damage could be avoid or minimize, along treading the path to development.

Therefore, the explanation of methodology define to build the TPPEM lies on the next questions: What data and time series have been chosen to integrate and test the model? What techniques and methods of spatial analysis and GIS have been used? How does it interact in the model?

4.1. Data and Time Series

By the study of the bibliography of the many author's referenced, and our own experience, five main topics are always stressed when it comes regional development: social, economic, land-use and land-cover, demographics and more recent, with the emerging knowledge and information society policies, science & technology. All data has been acquire from the census 1991 and 2001, from the National Institute of Statistics, with the exception of land-use and land-cover that has been provided from the project Corine Land Cover, from the Institute of Environment.

Social – in this topic we have select the education data: percentage of people of have an high degree of education. This chose lies by the strong linear regression between high developed regions and level of school education of people; and the unemployment rate, both for the year of 2001.

Economic – the Gross Added Value (GAV) per capita has been the indicator chosen because is the one that better reflects the intern economic dynamics; as well the changes occurred in the sector. For this, we have chosen the GAV for 2001 and the GAV per capita variation between the years 2001 and 1991, at 2003 prices;

Land-use and land-cover – give us the land dynamics, both natural and artificial. From this data we have used only the most land-uses changes in the period between 1991 and 2001 (Table 1), eg. Urban Land Use (class 1) and Transport Infrastructures Use (class 2), because it could help us understand mobility potential and relationships capability between regions. To achieve this we have applied a cross-table technique, between the two time series so we could not only know how much land cover changed, but what have have been change between land--cover classes.

Table 1 – Cross-tabulation of corine Land Cover 1990 - review (columns) against corine Land Cover 2000 (rows) Transition area (cells) matrix

	1	2	3	4	5	6	7	8	9	10	11	12
1	1404	15	40	2	20	261	53	40	3	0	0	0
2	10	297	25	1	8	72	99	60	2	0	0	2
3	0	0	13146	185	62	138	55	56	0	1	3	0
4	0	0	30	338	0	1	0	2	0	0	0	0
5	0	0	100	0	5574	211	29	64	0	0	0	0
6	0	0	61	2	19	22877	101	102	0	0	0	0
7	0	1	92	8	8	135	21883	2488	106	0	0	0
8	0	3	163	2	25	288	2780	13937	340	0	9	0
9	0	0	0	0	0	0	73	235	1424	0	2	0
10	0	0	1	0	0	1	0	0	1	300	0	1
11	0	1	17	1	0	27	4	25	0	0	519	0
12	0	0	0	0	0	0	0	0	2	0	0	545

Classes:

1 - Urban fabric surfaces
2 - Industrial infrastructures and general facilities surfaces
3 - Agricultural areas with annual crops
4 - Agricultural areas with permanent crops
5 - Heterogeneous agricultural areas
6 - Pastures
7 - Forest
8 - Scrub and or herbaceous vegetations associations
9 - Open space with little or no vegetation

10 - Wetlands
11 - Continental water bodies
12 - Marine water bodies

The process occurs in grid GIS model for data analysis, with a spatial resolution of 1square kilometer. The performance of cross-tabulation process could be evaluated by Cramer's V measure, which indicates the correlation between the attributes crossed of the two images. In this case, Cramer's V indicate an high correlation (0.89) between Corine LandCover classes for 1990 reviewed image with Corine LandCover classes for 2000 image.

The results obtained from this cross table indicates that land uses trends to be more urban, and land cover more artificial despites the fact of high values for changes between agricultural, pastures and forests areas, that result mostly for fires (scrub's and herbaceous surfaces and pastures), speculation policy (soil that are uncultivated by owner's that play with the market and or expected for the soil to become classified in the master plans as for urbanize) rather that a real policy of land use legal affects.

Demographics – from this topic we have use the population density for 2001, it give us the demographic concentration, occurred in the territory.

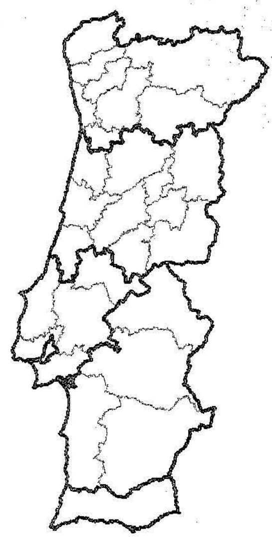
Science & technology – that has the most recent indicator introduced in regional development study's over the past century. The revolution (or evolution, as Castells prefer to say) of the society and consequently of the economy, at the end of XX century it's due to information and communication technologies, namely broadband internet. Nevertheless, and due the lack of significant information on this issue at the regional scale, mostly because of a hidden information policy, we have construct an composite indicator by the sum of universities, high institutes, national and state labs and techopoles existent in the study territory's.

In what concerns with the scale of study, we have chosen the sub-region area, which is equitable with statistical territorial unit III (NUT III) (Figure 2). That's the analysis scale that better correspond to the agreement between: data available, comparative territory scale and time computing for a standard desktop PC (Pentium IV, 2.0 GHz processor and 512 Mb RAM).

4.2. Multicritirea Analysis and GIS

Today we live in a rapidly changing world that is increasingly challenged, therefore we need be more alert and act more rapidly, as transformations occurs. The emergences of new technologies, like GIS, have been to give thus who have the responsibility to decide, face the challenge. In fact, has we could prove with this project, GIS involves much more than the just displayed facilities, or the desktop functionalities, and those faculties are ultimately far more important. Spatial analysis and GIS stays as one the examples still in progress.

Figure 2 – Main Portugal NUT III, 2001

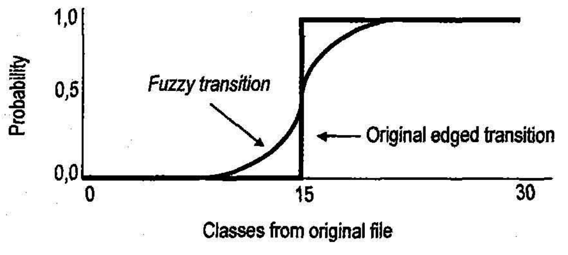


Nevertheless, has we said before, most of the times, from most of the commercial GIS software's existent, if we want to make some specific spatial analysis, eg. Multivariate, cluster, factorial, fuzzy, bayes, multi-regression and much more, we have to go outside GIS, so we could perform them and test different sort of hypothesis, relations and patterns. Evens if, in the past decade mostly, we have presence an improvement on the GIS industry sector to couple this kind of spatial analysis into GIS, there are a reduced number of programs and very limited sort of operations full integrated into GIS. Fortunately there are alternatives ways of attaching these functionalities to GIS, eg. Programming and calling a spatial analysis routine into GIS, or scripting, in proper language of the program, the routine in side the GIS.

Back to our project, and the construction of the evaluation model of territorial public policies we have use as techniques and methods, multivariate analysis (principal components analysis) and grid GIS model. In what comes to the grid model chose, we have prefer the more analytical capacity and the facility of space formalization rather than the rigor and flexibility of vector GIS model. Normally, spatial continuous data is referred to geostatistics analysis rather than spatial data analysis, as often we see the attributes values relate only to a finite set of areas or regions, that partition the study area, instead a continuous territory analysis. But, we have agreed that it makes no sense to think of these indicators that we chose (social, economic, demographic, land-use and land-cover and S&T), that it's attributes measures, only affect a very well defined edge area. Therefore, we prefer to consider by continuous process in time and work this attributes in a grid data structure.

That implies we have to consider a set of observation Yi, i = 1..., n, on a spatially continuous attribute recorded at corresponding spatial locations, Si, in the study region R. For all the cases (ie all the attributes of the study data), we needed also to study aspects of local variability, as way of dissipate uncertainty, by an interpolation on prediction of the value of the attribute at points other then itself. The fuzzy (Figure 3) method, by linear process or sigmoidal process, has been chosen, so we may achieved a set of entities and their attributes in a continuous binary scale range of 0 to 255, in a bite, binary grid data structure.

Figure 3 – Fuzzy technique exemple



The spatial resolution of the grid it's an agreement between the original scale of the information, the objective(s) of the model and the processor and RAM pc capacity. The chosen lies on a grid with a resolution of 1 kilometer, for all the territory of Main Portugal.

In what concerns techniques and methods applied for spatial data analysis, we have opted by the multivariate methods, as the analysis of the relationships in the values of several variables or attributes treated simultaneously. The objective here is to identifying a small number of significant dimensions (combinations of attributes) which may then be examined from a spatial perspective, and use in spatial classification and discrimination.

The technique of multivariate method we used was principal components analysis. That is a multivariate exploration technique that transforms a number of correlated attributes into a small group of independent components, coming from linear combinations of the original ones, as we can see below:

$$C_1 = b_{11}V_1 + b_{12}V_2 + b_{13}V_3 + \dots + b_{1n}V_n$$
$$C_2 = b_{21}V_1 + b_{22}V_2 + b_{23}V_3 + \dots + b_{2n}V_n$$
$$\dots \tag{1}$$
$$C_n = b_{n1}V_1 + b_{n2}V_2 + b_{n3}V_3 + \dots + b_{nn}V_n$$

where C is the component, b the load adjust parameter and V the variable.

That resumes data, by eliminated redundancy, into independent components (Table 2). After we have used this first principal component, a second spatial analysis has to be put in scene: cluster analysis, so we could identify new potential regional constellations of relationships towards development.

Table 2 – Loading Matrix for the PCA

Variables	C1	C2	C3	C4	C5	C6	C7	C8
idri_cor2_trapfuzz	0,049	-0,164	-0,969	0,054	-0,169	0,020	0,000	0,000
idri_ctec_fuzz	0,252	-0,673	-0,059	-0,462	0,489	0,167	0,003	-0,001
idri_cor2_urbfa_fuz	0,239	-0,672	0,236	-0,107	-0,639	0,127	0,002	-0,001
idri_desmp01_fuzz	0,681	0,358	-0,045	-0,512	-0,149	-0,300	-0,177	-0,022
idri_dpop01_fuzz	0,541	-0,517	0,051	0,463	0,165	-0,428	-0,112	-0,003
idri_txesup_fuzz	0,962	0,108	0,000	0,041	0,010	-0,031	0,207	-0,134
idri_vab_01_fuzz	0,968	0,140	-0,002	0,000	-0,002	-0,013	0,121	0,166
idri_var_vab_fuzz	0,769	0,174	0,014	0,257	0,055	0,527	-0,178	-0,019
Eigenvalue	3,33	1,39	1,00	0,77	0,73	0,60	0,13	0,05
% Total variance	41,63	17,38	12,50	9,63	9,13	7,50	1,63	0,63
Cumulative %	41,63	59,01	71,51	81,14	90,27	97,77	99,40	100,00

The PCA extracts three main components that explain a little more than 71% of total variance. The first one is a social-economic component that is correlated with the high school education, GAV per capita and GAV per capita variation. The scores of this component shows areas where the population have a higher education level, a higher economic performance and, at same time, had a higher economic rise between 1991 and 2001. The second components have a negative correlation with urban land use areas and the presence of science and technologic labs and centers. High negative scores shows the urban areas and hi-tech territory. Finally, the third component is strongly correlated with the transport infrastructure. The scores show the areas with a higher level connection on transport network.

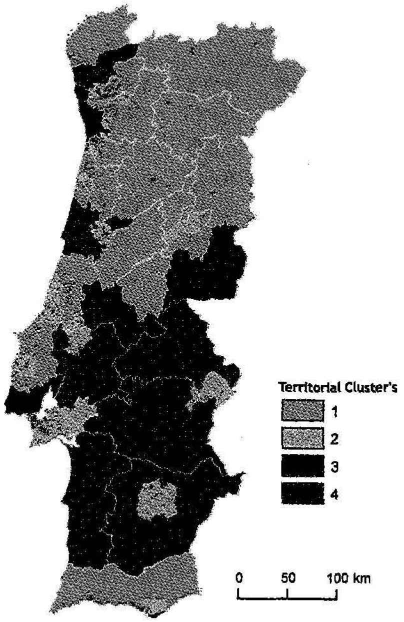
This multivariate analysis makes evidence and let to highlight the territorial differentiation on three potential vectors: economic and human resource potential;

urban and innovation and technical potential and areas with accessibility potential (Figure 4).

The next step lies on making a cluster analysis from scores of the three principal components. The result shows us, different potential constellations territories:

- the more central urban territories, with human resource and economic potential (red-3),
- the urban areas with innovation and technical potential (yellow-2),
- and a slight differences in terms of economic and human resource potential between the remaining territories (dark blue -4, slight better then light blue - 1).

Figure 4 – Regional Development Clusters for Main Portugal



4.3. Self organizing feature map

Self-organizing feature maps (SOFM) transform the input of arbitrary dimension into a one or two dimensional discrete map subject to a topological (neighborhood preserving) constraint. The feature maps are computed using Kohonen unsupervised learning. The output of the SOFM can be used as input to a supervised classification neural network such as the multi layer perceptron (MLP).

This network's key advantage is the clustering produced by the SOFM which reduces the input space into representative features using a self-organizing process. Hence the underlying structure of the input space is kept, while the dimensionality of the space is reduced.

This model stems from Kohonen [16] and builds upon earlier work of Willshaw and von der Malsburg [30]. The model is similar to the (much later developed) neural gas model since a decaying neighborhood range and adaptation strength is used. An important difference, however, is the topology which is constrained to be a two-dimensional grid and does not change during self--organization.

The SOFM [9,16,17,20] is a neural network algorithm that has been used for a wide variety of applications, mostly for engineering problems but also for data analysis [2,28]. A comprehensive treatment of the topic is provided by Kohonen [17]; here only aspects relevant for data exploration and aspects needed for understanding the relationships between the SOM and the other algorithms will be presented.

Data mining tools have been divided into two categories, clustering methods and projection methods. The SOFM is a special case in that it can be used at the same time both to reduce the amount of data by clustering, and for projecting the data nonlinearly onto a lower-dimensional display.

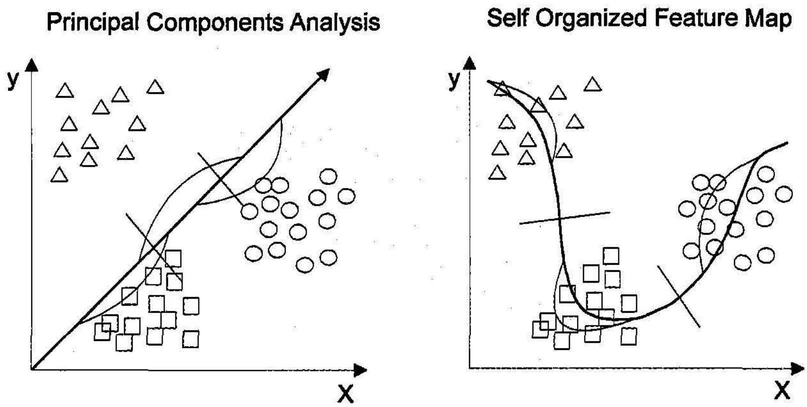
There are two particularly useful purposes for this: visualization and cluster analysis. Visualization has typically been a difficult matter for high-dimensional data. SOFM can be used to explore the groupings and relations within such data by projecting the data on to a two-dimensional image that clearly indicates regions of similarity. Even if visualization is not the goal of applying SOFM to a dataset, the clustering ability of the SOFM is very useful.

There are many methods that can be used to reduce the dimensionality of the data. The Self-Organizing Feature Map works both as a projection method and a clustering algorithm. High-dimensional data is mapped to a low-dimensional space, typically to one or two dimensions. In clusterization similar data vectors are mapped to near-by neurons and therefore the topology of the original data is preserved. These properties are very useful in data analysis.

The number of neurons in a SOFM needs to be large enough so that there are a few neurons to represent each larger group of the data. These small sets of neurons provide a subsymbolic representation to the data groups. In the training process these automatically forming small sets of neurons become the prototypes, which are ordered according to their similarity. Figure 5 depicts how the SOM reduces data dimensionality from two to one compared to the use of a principal component.

If we just use the first principal component, no clear decision boundaries can be found, and the reduction of dimensionality does not improve the analysis. The

Figure 5 – Dimensionality reduction in a two dimensional case



Self-Organizing Map is able to provide a "curve" to which the data points can be projected appropriately and grouping is easy. In this sense the SOM behaves in a similar way as the principal curves [4, 7].

After the neurons are trained, each data vector can be classified to one of the neurons. Once the original data is grouped in this manner, statistical properties of the training (and also background) variables can be calculated and visualized.

The competitive learning, related with this process, is an adaptive process in which the neurons in a neural network gradually become sensitive to different input categories, sets of samples in a specific domain of the input space [1,15,16,29]. A kind of a division of labor emerges in the network when different neurons specialize to represent different types of inputs.

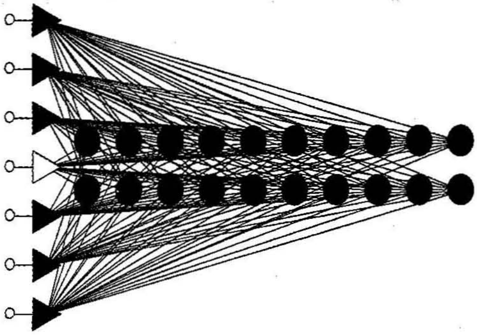
The specialization is enforced by competition among the neurons: when an input arrives, the neuron that is best able to represent it wins the competition and is allowed to learn it even better, as will be described below.

If there exists an ordering between the neurons, i.e., the neurons are located on a discrete lattice, the self-organizing map, the competitive learning algorithm can be generalized: if not only the winning neuron but also its neighbors on the lattice are allowed to learn, neighboring neurons will gradually specialize to represent similar inputs, and the representations will become ordered on the map lattice. This is the essence of the SOFM algorithm.

The network is created from a 2D lattice of 'nodes', each of which is fully connected to the input layer. Figure 6 shows our small Kohonen network representing a two dimensional vector.

During the training period, each neuron with a positive activity within the neighborhood of the winning neuron participates in the learning process. A winning

Figure 6 – SOFM - Kohonen network



processing element is determined for each input vector based on the similarity between the input vector and the weight vector.

Let us denote the input vector, X, as follow:

$$X=[x_1,x_2,\ldots,x_p]t \tag{2}$$

The weight vector, W_j, corresponding to output layer neuron j can be written:

$$W_j=[w_{j1},w_{j2},\ldots,w_{jp}]t \qquad j = 1, 2, \ldots, n \tag{3}$$

The neurons represent the inputs with reference vectors m_i, the components of which correspond to synaptic weights. One reference vector is associated with each neuron called unit in a more abstract setting. The unit, indexed with c, whose reference vector is nearest to the input is the winner of the competition:

$$c = c(x) = \arg\min_i \left\{ \|x - w_i\|^2 \right\} \tag{4}$$

Usually Euclidean metric is used, although other choices are possible as well.

The winning unit and its neighbors adapt to represent the input even better by modifying their reference vectors towards the current input. The amount the units learn will be governed by a neighborhood kernel h, which is a decreasing function of the distance of the units from the winning unit on the map lattice. If the locations of units i and j on the map grid are denoted by the two-dimensional vectors r_i and r_j, respectively, then $h_{ij}(t) = h(\|r_i - r_j\|;t)$, where t denotes time.

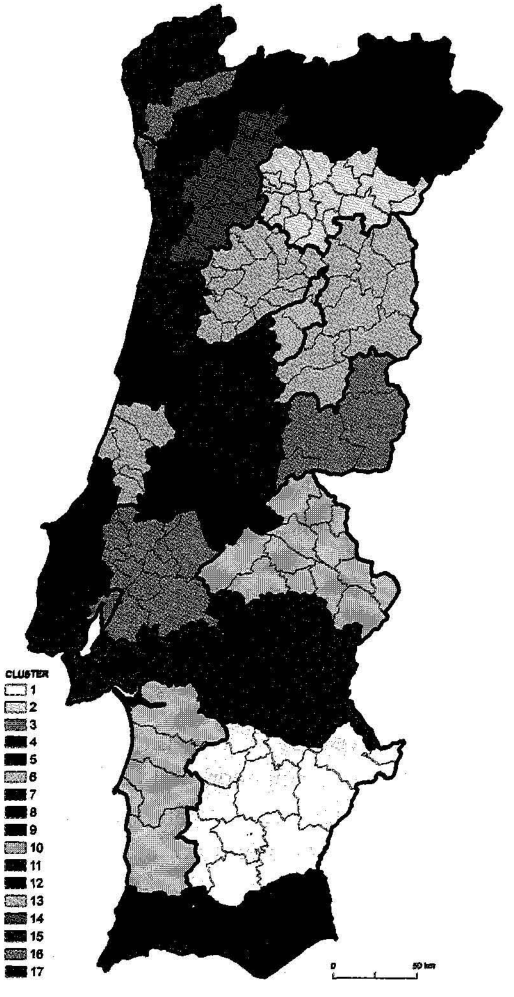
During the learning process at time t the reference vectors are changed iteratively according to the following adaptation rule, where $x(t)$ is the input at time t and $c = c(x(t))$ is the index of the winning unit:

$$w_i(t+1) = w_i(t) + h_{ci}(t)[x(t) - w_i(t)] \quad (5)$$

In practice the neighborhood kernel is chosen to be wide in the beginning of the learning process to guarantee global ordering of the map, and both its width and height decrease slowly during learning.

The learning process consisting of winner selection by Equation 4 and adaptation of the synaptic weights by Equation 5, can be modeled with a neural network structure, in which the neurons are coupled by inhibitory connections [8,14]. The resultant Portugal Map for our set of data is depicted in Figure 7.

Figure 7 – Portugal Regional Development Clusters



5. Conclusions

Some considerations could be pointed related to this approach. First the possibility to use and integrate statistical data from different geographical levels (NUT3, Corine Land cover, municipalities and so on). Second, avoid the limitation of boundaries limits (continuous variation by interpolation processes). In fact from the integration of current statistical analysis into GIS spatial analysis, a lot of barriers and limitations to a more accuracy and precision results on the territory as been dissipated. From this exercise, that represents an experience of integrated Spatial data analysis and GIS that we wish to execute in a more stretched project, we could truly say that with this improved techniques and methods - and after achieved more data from regional and local authority's (that should be active membership in the project) -, a regional competitively intelligent system could be draw for what we as usually call the sustainable development, in a designated information and knowledge society.

Internet, and it's implications in the live of the people, are the future. We can access information and communicate, anywhere, and anytime we want. That is a divine capacity, of omnipresence that are changing the way we interact with the territory. Network analysis, based on telecommunications infrastructures namely as optic fiber for broadband internet, power line, cable, and even other already installed infrastructures that are able to support this news ways of communicate that redefined distance. With the recent UE aims based on the Lisbon Agenda, the government authority's of Portugal could not lost any time more: to become competitive, information should be store and organize, territory and public policies remotely observe and study, private and public agents have to work together, people have to be technological educated and decisions take it according to local/ /regional spot culture's and political mimetic decisions avoid.

References:

AMARI, S., Topographic organization of nerve fields. Bulletin of Mathematical Biology, 42, 1980, 339-364.

Back, B., Sere, K., and Vanharanta, H., Data mining accounting numbers using self-organizing maps. In Alander, J., Honkela, T., and Jakobsson, M., editors, Proceedings of STeP'96, Finnish Artificial Intelligence Conference, pages 35-47. Finnish Artificial Intelligence Society, Vaasa, Finland, 1996.

BAILEY, T., GATRELL, A, Interactive spatial data analysis, Prentice Hall, England, 1995, pp.409.

Cherkassky, V. and F. Mulier. 'Self-Organizing Networks for Nonparametric Regression', In Cherkassky et al. pp. 188-212, 1994. (7)

GILBERT, G., Le Fédéralism Financier, Revue Economique, n° 2, 1996.

GOODCHILD, M., Geographical information science. International Journal of Geographical Information Systems 6(1), 1992, 31–45.

HASTIE, T. and W. STUETZLE. 'Principal Curves', Journal of the American Statistical Association. 84, pp. 502-516, 1989. (6)

KASKI, S. and KOHONEN, T., Winner-take-all networks for physiological models of competitive learning. Neural Networks, 7, 1994, 973-984.

KOHONEN, T. et al, Engineering applications of the self-organizing map. Proceedings of the IEEE, 84:1358--1384, 1996b.

KOHONEN, T. et al., SOM_PAK: the self-organizing map program package. Technical Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland, 1996a.

KOHONEN, T., Construction of similarity diagrams for phonemes by a self-organizing algorithm. Report TKK-F-A463, Helsinki University of Technology, Espoo, Finland, 1981.

KOHONEN, T., Emergence of invariant-feature detectors in self-organization. In Palaniswami, M., Attikiouzel, Y., Marks II, R. J., Fogel, D., and Fukuda, T., editors, Computational intelligence. A dynamic system perspective, pages 17-31. IEEE Press, New York, NY, 1995b.

KOHONEN, T., Emergence of invariant-feature detectors in the adaptive-subspace self-organizing map. Biological Cybernetics, 75, 1996, 281-291.

KOHONEN, T., Physiological interpretation of the self-organizing map algorithm. Neural Networks, 6, 1993, 895-905.

KOHONEN, T., Self-Organization and Associative Memory. (3rd edition 1989). Springer, Berlin, 1984.

KOHONEN, T., Self-organized formation of topologically correct feature maps. Biological Cybernetics, 43:-69, 1982..

KOHONEN, T., Self-Organizing Maps. Springer, Berlin, 1995c.

KOHONEN, T., Self-organizing maps: optimization approaches. In Kohonen, T., Mäkisara, K., Simula, O., and Kangas, J., editors, Artificial Neural Networks. Proceedings of ICANN'91, International Conference on Artificial Neural Networks, volume II, pages 981-990, North-Holland, Amsterdam, 1991.

KOHONEN, T., The adaptive-subspace SOM (ASSOM) and its use for the implementation of invariant feature detection. In Fogelman-Soulié, F. and Gallinari, P., editors, Proceedings of ICANN'95, International Conference on Artificial Neural Networks, volume 1, pages 3-10. EC2 & Cie, Paris, 1995a.

KOHONEN, T., The Self-Organizing Map. Proceedings of the IEEE, 78, 1990,1464-1480.

LONGLEY, P.; GOODCHILD, M. et al., Geographic information systems ad science. John Willey & Sons., London, 2001.

MARTIN, D., Geographic information systems. Socioeconomic applications. 2ª Ed. Routledge, London, 1996.

MITCHELL, W. J., e-topia. "URBAN LIFE, JIM – BUT NOT AS WE KNOW IT". MIT Press. London, 2000.

NIOCHE, J.P., De l'Évaluation à l'Analyse des Politiques Publiques, Revue Française de Science Politique, vol. 32, nº1, Fevereiro, 1982.

ORSINI, A., "La Problématique Générale de L'Évaluation", Revue d'Économie Régionale et Urbaine, nº 3, 1998.

PEYREFITTe, M., La Prise en Compte des Logiques d'Acteurs dans L'Évaluation d'une Politique Publique Territoriale", Revue d'Économie Régionale et Urbaine, nº 3, 1998.

SERRANO, A. et al., Cidades e Territórios do Conhecimento. Um novo referencial para a competitividade. Edições Sílabo. APGC. Lisboa, 2005, 177pp.

SERRANO-CINCA, C., Self-organizing neural networks for financial diagnosis. To appear in Decision Support Systems, 1996.

SWINDALE, N. W., A model for the formation of ocular dominance stripes. Proceedings of the Royal Society of London, B, 208, 1980, 243-264.

WILLSHAW, D. J. and C. von der Malsburg., How patterned neural connections can be set up by self--organization. In Proceedings of the Royal Society London, volume B194, pages 431-445, 1976.