AN «UNUSUAL» DATA FUSION

Elena Console (1), Paul C. Smits (2), Stefano M. Pagnotta (1,3), Alessandro Annoni (2)

(1) T..E.A.sas via Bruno Chimirri, 28 - 88100 Catanzaro (Italy)
ph.: +390961723634; fax: +390961794294; e-mail: tea@abramo.it
(2) Agricultural and Regional Information Systems Unit,
Space Applications Institute, European Commission DG JRC, Joint Research Center
TP262, 21020 Ispra (VA) (Italy); ph.: +390332786336; fax: +390332789936; e-mail: paul.smits@jrc.it
(3) Dipartimento di Economia e Territorio, Università degli Studi di Cassino; via Mazzaroppi, 03043-Cassino (FR).

ABSTRACT

The study of the relations between environmental changes and socioeconomic situations is of great relevance to the definition of the impact that national policies have on the environment.
In order to understand such relations, in this work, we have taken into consideration those changes taking place on a provincial scale in the period going from 1975 to 1992. To this end, both thematic and socioeconomic data have been taken into account, with the purpose of finding a correlation between the natural environment on the one hand and the anthropic environment on the other.

INTRODUCTION

In the widest sense of the word, the territory is defined as a dynamic and complex system, made up of such elements as natural resources and human factors, coexisting within the same space and connected with one another.
The description of such a system, entailing the study of the interdependence among economic development, environment exploiting and management of unrenewable resources, generate complex issues which are to be tackled on the basis of detailed information and suitable tools supporting the decision making processes.
The approach to the "Nature/Man" system needs:

- to go deeply into the knowledge of the "link" between the socio-economic and environmental situation;
- to apply methodologies of analysis and evaluation;
- to use all the available information in the planning activities.

This is the scope of this work, whose subject matter has been the study of the province of Reggio Calabria, an area of approx. 3.184 km^2, 97 towns mainly stretching along the coastline. The aim of this work is to explore the problems connected with the processing and fusion of different kinds of data (socio-economic and environmental data).
These problems are concerned with:

- the kind and selection of the data to be considered in the analysis regarding the phenomenon,
- the scale of the data in order to allow the comparison between the variables;
- the methodologies of analysis to be adopted for the processing, synthesis and "fusion" of the "Man/Nature" system.

The analysis is performed on the province of Reggio Calabria (Italy) and the data we have been using are:

- LACOAST (1975) and CORINE (1992/95), the European land-cover data produced by the E.U. on a scale 1:100.000;
- Vector data of the town boundaries;
- Socioeconomic indices computed from the official statistics produced by ISTAT (Italian National Institute of Statistics).

"NATURE" SYSTEM ANALYSIS

In order to analyse the "Nature" system we have been using database from two European projects, "Co-ordination Information on the Environment (CORINE) and LACOAST (1), starting from remote sensing images Landsat TM and MSS) and ancillary data concerning, respectively, the years 1992 and 1975. The database in question resort to the same nomenclature of land covering and exploiting, they are made on a scale of 1:100.000 and involve the coastal areas of the European Countries (with the exception of Great Britain) and their hinterland as deep as 10 Km. In our case, this means we can avail ourselves of information concerning 69 out of the 97 towns that make up the province of Reggio Calabria.
The other data we have used are the vectors representing the towns' administrative boundaries provided by the GIS of the European Commission (Gisco-Eurostat).
The method adopted aims at detecting those changes occurring within each municipal territory starting from the analysis of the changes in the texture of the land-cover classes and from the subsequent computation of the change rate in the natural environment. The natural covering taken into consideration consists of classes 2.4.3 (land mainly taken up by agriculture, with significant areas of natural vegetation), 2.4.4 (Agro-forestries areas) and 3 (Forest and semi-natural areas) of the CORINE nomenclature.

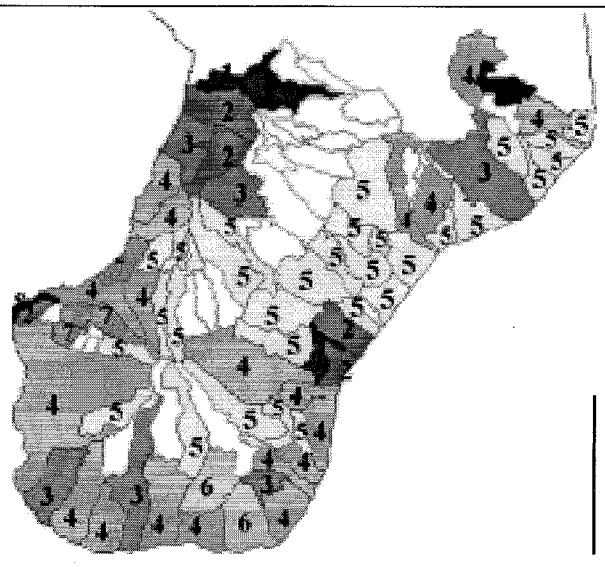
The first step of the analysis has required the calculation of the co-occurrence matrices for 1975 and 1992 [3]; starting from these values, in relation to both years considered in the second step of the analysis, we have calculated the value of the probability of finding 1 uninterrupted kilometre of the land cover class taken into consideration (PCN) [5]:

$$PCN \quad \frac{C_{i,d} \quad C_{ii,d}^{\ 2} \ C_{i,i,d}^{\ n\ 1}}{C_{i,d}^{n}}$$

where $C_{i,d} \quad \sum_{j} C_{i,j,d}$, C_{ij} stands for the probability of co-occurrence of i and j class and d for the distance selected for the sampling. The change index of the natural environment is given by the difference between the PCN computed for the two years taken into consideration and shows the increase or decrease of these classes in the period given:

$$C_d^{U,T0} \quad PCN^{T0} \qquad PCN^{U} \quad PCN^{U,T1} \quad PCN^{U,T0}$$
$$C_d^{U,T1} \quad PCN^{T1}$$

The figure 1 shows the obtained results. It is worth noting that the change index of the natural environment equals 0 (no change) for 30 towns out of 69, it is negative, between –72 and –0.2 (decrease), for 32 towns, while it is positive, between 0.6 and 20 (increase), only for 7 towns.



Legend: Values of index

Label	Values	Label	Values
1	< -60	5	0
2	-59.9 - -40	6	0.01 – 9.9
3	- 39.9 - -20	7	10 – 19.9
4	-19.9 - -0.01	8	> 20

Figure 1 – Change index of the natural environment: results

"MAN" SYSTEM ANALYSIS

In the "Man" system analysis, consisting in our case of the 97 towns of the province of Reggio Calabria, 34 indicators have been used, computed from the census data of 1981 and 1991 on a municipal scale and from other official papers issued by ISTAT (National Institute of Statistics). The indicators, set up for each town and for the years taken into account, have been arranged in increasing order and the corresponding values replaced with the rank filled in the classification so that the resulting data might be equalised. Then, the differences between the 1991 and the 1981 ranks have been computed and these differences (later on also described as "variables") have been processed and synthesised with the aid of two methods of multidimensional statistical analysis: principal components analysis, PCA, and cluster analysis.

PCA allows the setting out of a series of variables connected with one another (called factors or factorial axes) as linear combinations of the original variables. These factors allow a synthetic description of the system considered as a whole so that its peculiarities may be pointed out.

The principal components analysis has pointed out four statistically relevant factors, accounting for 34% of the system information structure which is being studied.

The first factor synthesises the variables concerning the variations of the activity rate in agriculture, of staff percentage and senility index contrasted with the variations in the percentage of employees and managers, of the activity rate in the services sector and in industry and of the youth index.

The second factor aggregates those variables concerning the variations in the percentage of vacant houses, of the general activity rate and of the illiteracy index contrasted with variations in the percentage of lived-in houses, density of population, percentage of houses built between 1971 and 1991 and the percentage of entrepreneurs and professionals.

After an in-depth investigation of the data most closely connected with one another and characterising the factors taken into consideration, on the whole, the first factor can be defined as the "variations in the production and demographic system" synthetic variable, expressing the contrapposition between the different components of the two systems: the variation concerning people employed in agriculture, along with changes occurring in the elderly population, is contrasted with the changes occurring among those employed in industry and the services and among young people. The second factor, on the contrary, shows the "variations of the housing system".

The research of the homogeneous areas was subsequently performed by means of hierarchic cluster analysis based on Ward distance. The result of such analysis are clusters of towns having the same socio-economic characteristics.

Cluster analysis has led to assigning the towns to eight classes (cfr. figure 2).

The first class, consisting of 27 towns, the second class, 4 towns, and the third class, 3 towns, are characterised by

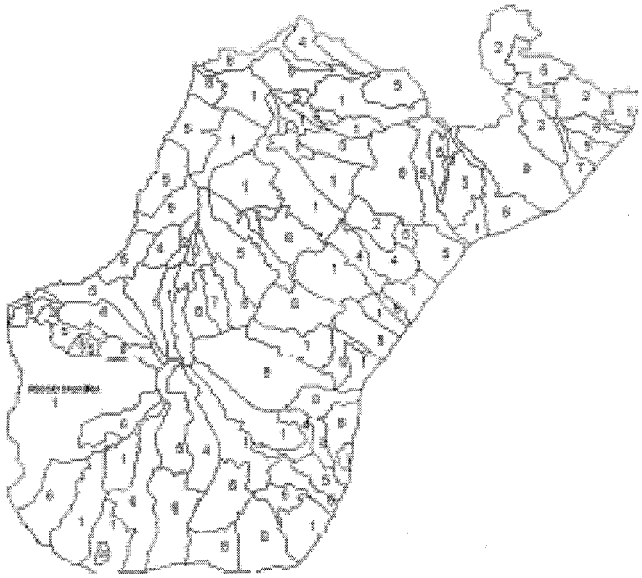
positive variations in the farming rate. Each of these 3 classes stand out for the following peculiarities: the first class for the negative variations of unemployment rates; the second class for the negative variations in the services activity rate; the third class, on the contrary, shows positive variations in the areas intended for woods matched by negative variations of the farmed areas and of the average size per farm, which might suggest a change in farming techniques and a shift towards an intensive rather than extensive farming.

The fourth class (8 towns) is characterised by positive changes of the housing property, by a resulting reduction in the wooded areas and by a negative variation of general activity rate and of youth index.

The fifth class (25 towns) is strongly marked by a positive variation in the employment rates, while the sixth class (10 towns) stands out for the positive changes of those variables which are linked with industry and handicraft, and for the negative changes in the density of population.

The seventh class (4 towns) is marked by remarkable positive changes of variables linked with the services and tourism (number of accommodation and people employed in trade), by the youth rate matched by a negative variation of the farmed areas.

The eighth class, too, (16 towns) is characterised by a change in those variables linked with tourism and the services, particularly as far as the proliferation of the holiday houses is concerned. It needs saying that the towns belonging to the last two classes are bordering on each other, that they are spatially contiguous.



Legend: Labels of classes

class 1	class 2	class 3	class 4
class 5	class 6	class 7	class 8

Figure 2: Cluster Analysis Results

CONCLUSIONS

Having separately analysed the two sets of data, we have then compared the results thus achieved for a combined interpretation aimed at integrating and, therefore, "fusing" the knowledge resulting from the investigation of such different data.

As a result, the reduction pointed out in the change rate of the natural environment is particularly evident in 3 classes, the first, fifth and eighth.

Specifically, the socioeconomic analysis shows a moderate increase of the farming activity index matched by a reduction in woodland for the first class; the fifth class, consisting of towns falling within the depressed areas with high unemployment rates, shows a remarkable reduction in those variables linked to farming (particularly the size of exploited farming land); finally, the eighth class, characterised by the development of tourism and proliferation of holiday houses and by a remarkable increase of the farming activity.

This enables us to point out that the variations occurring in the natural environment may take up different connotations and may be differently interpreted once they are integrated with socioeconomic data. In our case, the decrease detected in the index taken into consideration in the first and eighth class is matched by growth in the agricultural and tourist sectors, which does not necessarily imply a "worsening", a "degradation" of the system as a whole, while, on the contrary, it stands for a decline of farming in the fifth class.

These last considerations seem to support our main assumption: the integration of data coming from different sources allows a better knowledge of any system, leading to an all-round, articulate and full perspective of the system itself. This is why further research is needed to investigate into this matter, paying particular attention to methods and techniques of analysis that allow the setting up of more suitable and effective tools able to support the decision making processes.

REFERENCES

[1] A. Annoni, D. Ehrlich, P. C. Smits, L. Montanarella, "Data sets for sustainable management of Europe's regions", Proceedings 4th EC - GIS Workshop, Budapest, Hungary, 24 - 26 June 1998

[2] B. Escofier, J. Pagès, "Analyses factorielles simples et multiples. Objectifs, méthodes et intérprétation", Bordas, Paris, 1990

[3] R. Haralick, L. Shapiro, "Computer and robot vision" Reading, Massachusetts, Addison-Wesley, 1992

[4] H. Y. Roy, D. R. Green, S. H. Cousins, "Landscape ecology and GIS", London, Taylor & Francis, 1993

[5] P.C. Smits and A. Annoni, "Spatial Analysis of Land-Use Changes as Knowledge Tools in Support to European Spatial Policies and Ecosystem Health," Ecosystem Health, Vol. 5, No. 4, December 1999, pp. 275-284.