A Multivariate Statistical Approach to the Analysis of Rural Development*

Mario Mazzocchi, Elisa Montresor*****

Abstract

The aim of this work is to contribute to the definition of an analytical approach for evaluating the dynamics in progress in the agricultural and rural development at a territorial level. For this purpose principal components analysis and cluster analysis were applied and the different methodological approaches reviewed. A two-stage method is also proposed. This could provide the analytical tools to simplify and interpret the results of the territorial analyses, also in order to supply a flexible instrument to policy-makers. A case-study on the Emilia-Romagna region, in Italy, is presented.

Keywords: *Agricultural and rural development, territorial analysis, principal components analysis PCA, cluster analysis, GIS mapping, rural indicators*

Introduction

In the last decade there has been a growing interest for research directed at the study of territorial differentiation of agricultural development in the EU countries, focusing on long-term changes [Kayser et al., 1994; OECD, 1994 and 1996]. Indeed, it became more and more urgent to understand how the single situations are adjusting to the deep structural changes in progress. The aim of this work is to contribute to the definition of an analytical approach for evaluating the dynamics in progress in agricultural and rural development at the territorial level. The analysis originated from the following considerations: (i) there is a growing need for analytical tools enabling the regional institutions to settle and monitor their rural development plans, as required by the bottom-up structure of the EU agricultural policy measures; (ii) the Italian agricultural and rural scenery is extremely heterogeneous and complex and the analytical tools to be employed need to be simple but exhaustive and should also take into account the social, economic and environmental issues; (iii) the multivariate statistical techniques, frequently employed in this field, are still characterised by significantly different approaches, that can lead to dissimilar results starting from the same data set.

Besides these basic considerations, many specific issues arise. Under the methodological perspective, the concerned multivariate statistical analysis consists of the joint application of Principal Components Analysis (PCA) and Cluster Analysis (CA). This paper aims to highlight the relevance of different application methods for the PCA. This technique can

* The authors wish to thank the anonymous referees for their careful reading and helpful suggestions.

** Dipartimento di Scienze Statistiche, Università di Bologna

*** Dipartimento di Economia, Società e Istituzioni, Università di Verona

E-mail: mazzocchi@stat.unibo.it

be either a simple instrument, preliminary to territorial mapping (achieved through CA) or it can become itself a useful methodology to "read" the territory and evaluate the relevance of the original indicators in explaining territorial differences. To this end, it can be interesting to evaluate the different choices in the application of the PCA, especially the two versions found in the cited literature, that we name the "per blocks" (of indicators) or "on the whole" (Anania and Tarsitano, 1995).

The traditional analysis "on the whole", employed (among others) by Cannata (1989), consists of the input of the whole set of indicators in a single passage. On the other hand, in Cannata and Forleo (1998) and in Boccafogli and Brasili (1998), the PCA "per blocks" is carried out in two steps: in the first one the set of indicators is subdivided into groups (blocks) which are homogeneous according to a given *a priori* classification scheme; in the second step the PCA is applied inside each block of indicators. The two approaches do not show substantial differences if the correlations between indicators belonging to different blocks are negligible, but Forleo (1998) points out that situations wherein an approach turns to be superior to the other, have still to be identified.

In the light of this issue, an application of the analysis at municipality level is proposed in this paper. The study will be based on data from the Italian region, Emilia-Romagna. This region is characterised by high levels of development, high agricultural and food productivity, relevant processes of concentration and specialisation in some territories, but also by the presence of less favoured areas or marked in some way by a consistent desertion of agricultural activities.

In what follows, section 2 describes the analytical approaches, whereas section 3 presents the indicators of agricultural and rural development. In the final section the results of the analyses carried out are interpreted and summarised.

Methodological issues and interpretation of the principal components

Through principal components analysis (PCA)[1], the initial set of p indicators observed on n statistical units can be transformed into a reduced set of variables able to explain a significant proportion of the original variability[2]. The variables obtained through PCA are a linear combination of the original indicators and, in contrast to the initial variables, they are uncorrelated between each other. The PCA was applied to the correlation matrix, in order to avoid any problems of different scale and measurement unit among the indicators[3]. Operating on the correlation matrix leads to the following formulation of the principal components scores, that can be computed for each statistical unit (in this case each municipality):

$$y_{i,j} = \sum_{l=1}^{p} a_{l,j}\hat{x}_{i,l} \quad \text{with} \quad i: 1, 2, \ldots, N \text{ and } j: 1, 2, \ldots, k \tag{1}$$

where $y_{i,j}$ is the j-th component score for municipality i

$\hat{x}_{i,l}$ is the standardised value of l-th indicator for municipality i

$a_{l,j}$ is the coefficient linking l-th indicator with j-*th* component

The scores for each municipality, contained in the vector $Y_i=\{y_{i,j}\}$ with $j:1,\ldots,k$ are the values to be employed in the cluster analysis. Before implementing the CA and obtaining the final geographic subdivision of territorial systems, the economic interpretation of the components and the comparison of scores assumed by the single municipalities could also supply some helpful indications for interpreting the agricultural territorial development. The explanation of the components derives from the sign and magnitude of coefficients $a_{l,j}$.

As the variables $\hat{x}_{i,j}$ are standardised, they all have variance equal to 1 and the weight of each variable on the component value is proportional to the absolute value of the coefficient. Hence, if one considers the squared coefficients, whose sum is constrained to be 1 by construction, these represent exactly the weight of each variable on the total component value. There are usually several problems in the interpretation of components: the number of indicators is often extremely large and the signs of the coefficients may return contradictory clues. Moreover, the interpretation is rarely univocal and it becomes more difficult when one tries to explain the last components, i.e. those with a lower proportion of explained variance.

Hence, a useful method for simplifying the analysis and eliminating the temptation of getting into arbitrary interpretation is to "filter" the variables through the weight on the component value as measured by each squared coefficient. There are no standard criteria for such an operation, that leaves again some room to arbitrary choices made by researchers. In this work just the indicators whose squared coefficient was above or equal to 0,05 were considered, that is the ones whose impact on the component value was at least of 5%, defining such a proportion as:

$$s_{l,j} = a_{l,j}^2 \cdot 100 \qquad (2)$$

where $a_{l,j}$ (l=1,…, p) is the coefficient of the *l*–th indicator in the *j*-th component.

This allows us to give a first quick interpretation to the components, which can then be tested or revised on the grounds of the sign and relevance of other coefficients. In general the $s_{i,j}$ is an indicator of the intensity of the impact of a single variable on the component value, whereas the direction of this impact is indicated by the signs of the coefficients $a_{l,j}$. Another indication which can be derived from the PCA concerns the comprehensive contribution of each indicator to the *k* selected principal components. In practice, if the $s_{l,j}$ are a measure of the weight of each variable inside the single component, the simple average $\overline{s}_l$ returns a measure, although approximate, of the relevance of the variable on the principal components as a whole. Such an average assumes, however, that all the components are equally weighted, whereas a weighted average with respect to the proportion of variability explained by each component seems to be more appropriate:

$$w_l = \sum_{j=1}^{k} a_{l,j}^2 \cdot \frac{Var(Y_j)}{Var(Y)} \qquad (3)$$

where $Var(Y_j)$ is the variance of the *j*-th component (equal to the *j*-th eigenvalue)
$Var\ (Y)$ is the total variance of the first *k* components
(equal to the sum of the first *k* eigenvalues)

Following (3) it is possible to rank the most representative indicators in the description of the phenomenon, although a strong limit in such a classification is that the correlation among the original variables has not been taken into consideration.

The Principal Component Analyses "per blocks" and "on the whole"

As mentioned, the analysis "per blocks" starts with a preliminary subdivision of the indicators into "blocks" or macro determinants. Such a technique has the advantage to perform a series of PCA on a limited number of indicators belonging to each group, making

calculation and components' interpretation easier. However it involves a number of choices that increase the degree of arbitrariness of the researcher. Another advantage of utilising the approach "per blocks" could be to avoid that some indicators, considered relevant by the researcher, but correlated to other variables, are hidden in the interpretative step. However, choosing the option "per blocks" means assuming that each group or macro determinant is equally represented in the cluster analysis. The consequence is to amplify the influence of some variables significantly correlated between them but belonging to different groups. As already mentioned, previous methodological studies showed the equivalence of the two approaches when the correlation between indicators belonging to different groups is negligible. However, in such circumstances, there is no need to safeguard the representativeness of some indicators, as they do not risk to be hidden by any strong correlation with other indicators.

A potential solution to the problem, when correlation between important variables actually exists, is suggested by Anania and Tarsitano (1995) and it amounts to performing another PCA on the components extracted from the different blocks. This allows us to have uncorrelated components for the cluster analysis, but the final components are not directly linked to the original variables, so that the relevance of the original variables in explaining territorial variability is not computable any more according to (3).

Moreover the subdivision of indicators into each block becomes crucial to the analysis: the weight of each macro determinant will be proportional to the number of indicators classified into the same group. A key issue of the analysis "per blocks" is also the way the component scores derived by the PCA will be utilised in the cluster analysis. If no standardisation is carried out, each component will have a weight proportional to its own variance. As seen, such a variance is equal to the component eigenvalue, which is itself strictly linked to the number of indicators classified into the each block[4]. Hence, the approach "per blocks" introduces a bias, giving more relevance to the components extracted by the groups containing a larger number of indicators. On the other hand, if the scores are standardised, each single component will have the same weight, so each macro-determinant (group) will have an impact proportional to the number of components that it "produced". This latter option seems to be more coherent with the initial choice of operating on the correlation matrix, which just presupposes that all indicators have the same relevance. However, for any kind of analysis, the choice of indicators, not only in qualitative terms, but also in terms of the number of indicators related to single macro-determinants, plays a decisive role.

On the other hand, in the analysis "on the whole" set of indicators, each component is a linear combination of all the original variables. This approach allows us to give a general interpretation of the components, considering jointly all the original variables. The problems due to the large number of observations can be overcome through the method described in section 2. The case study in the Emilia-Romagna region, described in the next sections, also aims to provide an empirical comparison between the two approaches.

The indicators of agricultural and rural development

As it has been seen, the choice of the proper indicators is essential for a well specified analysis. However, the multivariate statistical approach described here can be also considered as a "positive" rather than "normative" model (Paris et al., 2000), employing a vast amount of information and reducing technically the dimension of the analysis.

Choosing proper indicators must take into account the objectives pursued in dividing the territory into homogeneous systems. In this case the aim is to detect the territorial differences with regard to agricultural and rural development. Generally the rural world presents

a wide range of shades and situations, that can be synthesised as follows: (i) the agricultural specialisation and concentration areas, with or without local integration with the food chain; (ii) the extended city areas, where the rural areas become important for their residential function; (iii) the rural areas in territories with a relevant presence of small and medium enterprises (industrial districts); (iv) other rural areas, subdivided according to the level of disadvantage-advantage, including the mountain areas. These territorial systems show strongly differentiated dynamics in the EU scenery and this will lead to the need for a different subdivision of measures at regional level. The aim is not to detect new institutional backgrounds, but to highlight the main issues to be focused on by regional and local planning authorities, in order to value the regional system of agricultural network management.

The actual complexity of various regional settings, highlights the fact that a comprehensive grid of indicators, referred to the municipality level and satisfactory for the whole region, does not seem to be practical at first glance, as the aspects to be investigated are very heterogeneous. The risk is to miss the specificity of single territories, especially with respect to the agricultural and rural perspective. Moreover, there is a relevant problem related to the possibility of acquiring the statistics necessary to derive the appropriate indicators. This could be solved through the settings of specific data banks, built through the cooperation with the regional institutions. This issue is seen as most relevant especially when the objective is to compare the different regional situations, not only at the national level, but also at the EU level.

Table 1. *Indicators for Emilia-Romagna utilised in the analysis*

First set of indicators	
Group	*Indicators*
Social and economic development	Per capita GDP, population density, % employees agriculture, unemployment ratio, food industry employees per firm
Agricultural structure	Avg. UAA per farm, % farms under 2 ha, % farms above 50 ha, % UAA of farms under 2 ha, % UAA of farms above 50 ha, tractors per ha UAA
Agricultural activities *crops: % of UAA* *animal production: heads for hectare of UAA*	Cereals, feeding crops, pastures, horticulture, fruits, vines, CDO vines, bovines, pigs, chickens, bovine heads per ha of pastures, pigs per cow
Productivity of agriculture	Workers per ha of UAA, SGM per ha of UAA, SGM per worker
Second set of indicators	
Social and demographic situation	Pop. change (81-91), ageing index, ratio of female workers, dependence ratio, graduating index, % employees industry, % employees services
Agricultural Structure dynamics (%)	UAA change (80-90), farms change (80-90), agricultural surface change (80-90)
Other indicators	Agr. Worker per food firm, % employees in large food firms, woodlands (% agr. surface), % farms with livestock

For these reasons the choice of proper indicators for the case studied here (Table 1) was carried out in two steps. The first series of indicators represents the minimum set to identify the dynamics in progress in agricultural and rural areas. It is a limited number of variables (26 in all), related to the social and economical background, the agricultural structure, the production specialisation and the productivity of the primary sector. These indicators are built utilising information at the municipality level from the Agricultural, Population and Industry censuses, with the exception of the GDP. This guarantees that the analysis is reproducible through space (for other regions)[5] and over time (for the same area in a different period).

This first set of indicators can be subdivided into four groups. The first group includes the minimum information necessary to understand the social and economic development level of a given area[6]; the second one, regarding the structures, was formed in order to understand the structural background where the agricultural and development dynamics are analysed. In this group there are 5 indicators related to the farms' dimensions, because the rural development path will be very different, depending upon whether the residential function is relevant to the territory or whether the agriculture of the area is efficient and productive. Moreover, the farm structures, according to their dimension, will respond in very different ways to environmental measures. The third group of indicators aims to evaluate the specialisation level for the main crops and livestock activities in the single territorial systems. Their choice was determined following the analysis of the main food chains existing in the region. The variables of this group are useful to evaluate, in successive analyses, how the specialisation level could change according to the adopted measures, but also to the level of integration with the local food processing firms and the actual implementation of environmental measures. Lastly, the fourth group reports three indicators related to the agricultural land and work productivity and intensity.

To this first set of 26 indicators, a second one was added, with 14 more variables. The aim of this second group is to evaluate in detail, some demographic and social dynamics, the agricultural structure and some indicators related to the structure of the food processing industry. This second set can be enlarged according to the specific objectives of the analysis. To widen the interpretation for the aims of economic policy, one can look particularly at the normative and institutional indicators.

Results of the analysis at regional level

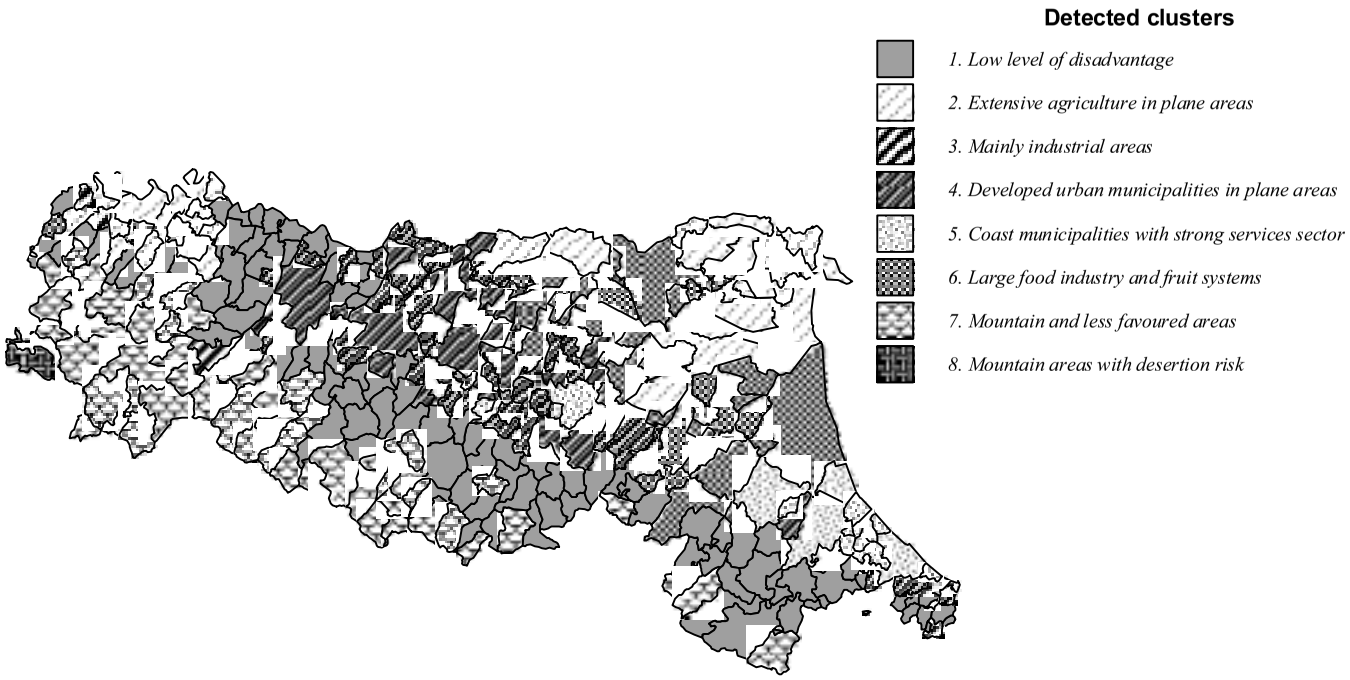
The analysis "on the whole"

As discussed, this approach consists of feeding in all indicators, in a single step. Twelve principal components were extracted, in the region analysed, from the 40 intial indicators described in section 2. They explain 73% altogether of the total variability of all 40 indicators.

The most relevant indicators for interpreting the principal components were detected after setting a 5% benchmark for the $s_{i,j}$ statistic as described in (2),. In order to complete and improve the description of each component, other indicators less significant in terms of $s_{i,j}$, but relevant to the confirmation or revision of the profile description were considered[7]. The values w_l, computed as in (3) give an indication of the impact of the single indicators on the extraction of the principal components, the most significant are reported in table 2.

Once the component scores have been computed, the cluster analysis was carried out. In this work the non-hierarchical method of the k-means was employed. This technique re-

Figure 1. Analysis "on the whole" – Map of clusters identified for the Emilia-Romagna region



quires to set *a priori* the number of groups, *g,* chosen as the smallest among the ones that provide significant F-test on the components. The 341 municipalities of the region analysed were grouped in 8 clusters (see figure 1 and table 3). The data analysis allows to delimit the main territorial system and the prevalent dynamics in progress inside them.

Table 2. 10 most significant indicators

Variable	w_l
Agr. SGP / ha UAA	3,22
UAA per farm	3,03
Dependence ratio	2,99
Agr. SGP / AWU	2,93
Food Industry: empl./firms	2,92
% employees agriculture	2,91
AWU / ha UAA	2,89
% empl. large food firms	2,89
% employees services	2,85
% employees industry	2,84

Table 3. Analysis "on the whole" - Description of detected clusters and average of the most relevant indicators and other statistics at municipality level

Areas with high levels of social and economic development and high agricultural productivity	
Plane or hilly agricultural intensive system	*Cluster 4*
Extended city agricultural system in areas characterised by strong presence of tertiary sector	*Cluster 5*
Fruit and vineyard systems on the plane	*Cluster 6*
Agricultural systems in mainly industrial territories	*Cluster 3*
Areas characterised by high agricultural yields and by an average level of economic development	
Extensive systems with cereal crops and horticulture on the plane	*Cluster 2*
Rural Areas with a different level of disadvantage	
Hill and mountain agricultural system with a low level of disadvantage (cluster 1)	*Cluster 1*
Mountain and hill least favoured areas	*Cluster7*
Mountain areas being abandoned	*Cluster 8*

Table 3. Analysis "on the whole" - Description of detected clusters and average of the most relevant indicators and other statistics at municipality level *(con.)*

Cluster	1	2	3	4	5	6	7	8	Region
Municipalities	97	36	6	79	20	46	54	3	341
% of regional population	11,8	9,6	0,7	30,7	24,6	19,7	2,9	0,0	100,0
% of regional surface	25,6	13,4	1,2	17,0	5,2	17,0	20,0	0,7	100,0
Per capita GDP	22,5	23,1	32,5	31,7	27,3	23,4	16,5	17,5	24,3
Pop. density (inh per km2)	89,8	129,5	103,1	267,6	910,2	200,9	28,0	8,4	188,0
% employees agriculture	18,4	18,4	8,7	8,8	7,0	18,1	30,1	41,2	17,4
% employees industry	39,0	38,1	60,6	54,8	28,8	39,7	24,5	15,1	39,9
% employees services	42,6	43,5	30,7	36,4	64,2	42,1	45,4	43,7	42,7
Unemployment ratio	6,8	10,0	7,4	5,4	11,0	7,1	7,6	9,8	7,3
Ageing index	186,4	192,2	186,7	148,2	144,7	209,4	367,7	1370,8	217,9
Female activity ratio	51,6	49,2	47,1	58,9	52,8	57,7	44,9	34,3	52,6
Dependence ratio	49,0	42,9	47,7	40,7	40,0	45,3	65,0	98,9	48,4
Graduated / school age	20,4	16,2	13,7	25,2	37,3	30,4	18,1	21,6	22,9
UAA per farm	8,4	15,6	18,4	8,2	3,7	8,5	7,7	5,0	8,9
% farms 0-2 ha UAA	34,2	21,4	18,2	35,5	55,9	30,3	31,9	61,3	33,5
% farms 50+ ha UAA	2,0	5,1	8,4	2,0	0,4	1,7	1,5	1,5	2,2
% UAA farms 0-2 ha	4,4	1,3	1,6	5,0	17,5	3,8	4,3	5,8	4,9
% UAA farms 50+ ha	16,3	33,6	37,8	18,5	10,5	20,7	13,0	22,6	18,8
Tractors per ha UAA	0,1	0,1	0,1	0,2	0,2	0,2	0,1	0,1	0,2
% woodlands on ag. sur.	18,5	2,7	11,4	5,4	3,4	4,7	45,8	44,4	15,5
% UAA - Cereals	22,8	42,5	36,6	29,0	34,3	29,3	8,7	1,9	25,7
% UAA - Feeding crops	50,5	20,2	38,3	30,2	14,7	10,4	41,4	3,4	33,0
% UAA - Pastures	14,1	1,6	11,9	6,0	2,2	1,8	45,1	90,1	14,1
% UAA – horticulture	1,2	11,6	5,2	1,5	7,4	2,1	0,1	0,2	2,7
% UAA - Fruits	1,6	3,7	0,1	6,2	15,5	24,8	0,2	0,0	6,6
% UAA - vineyards	3,5	0,9	0,7	9,6	8,4	11,1	1,5	0,7	5,6
% CDO vines on tot. vines	14,3	8,6	7,5	34,6	21,8	11,2	2,9	0,0	16,4
% aziende con allev.	55,6	56,4	63,3	53,6	44,3	44,6	57,0	32,0	53,2
Bovines / ha UAA	0,9	0,7	1,0	1,2	0,1	0,2	0,5	0,3	0,8
Pigs / ha UAA	1,5	0,8	0,7	3,8	0,7	0,6	0,4	0,0	1,6
Chickens / ha UAA	21,8	9,6	1,5	12,3	131,2	23,4	4,7	1,0	21,7
Bovines /ha UAA Pastures	1,1	3,4	0,1	5,4	0,3	3,6	0,0	0,0	2,4
AWU / ha UAA	0,1	0,1	0,1	0,1	0,3	0,1	0,1	0,1	0,1
Agr. SGP / ha UAA	1,8	3,6	2,1	3,8	4,4	4,4	0,7	0,4	2,8
Agr. SGP / ula	19,6	49,2	36,7	29,8	18,0	35,4	9,9	3,8	25,8
Food Industry: empl./firms	4,5	5,2	7,0	5,9	4,9	5,7	3,1	3,3	4,7
% empl. large food firms	4,6	11,3	67,1	12,2	24,2	26,9	0,9	0,0	11,7
Population change 81-91	2,2	-2,1	2,5	7,9	3,2	-0,5	-10,3	-23,9	0,6
UAA % change 81-91	-4,1	-0,2	3,0	-3,2	-0,8	-0,6	-10,8	NA	1,9
Farms n. % change 80-90	-12,0	-12,1	-13,0	-12,2	-3,3	-9,8	-20,9	-47,1	-13,0
Ag. surf. % change 80-90	-3,4	-1,1	-5,9	-3,1	2,5	-1,2	-11,3	-30,6	-4,0
% mountain municipalities	20,6	0,0	16,7	0	0	0	79,6	100	19,6

Table 4. Analysis "per blocks" - Description of detected clusters (territorial systems) and average of the most relevant indicators and other statistics at municipality level

Areas with high level of social and economic development and high agricultural productivity	
Intensive hill and plane agricultural systems	Cluster 1
Extended city agriculture in territories with relevant tertiary sector	Cluster 5
Areas with high agricultural productivity and an average level of development	
Extensive cereal crops and intensive horticulture and fruit production systems	Cluster 2
Agriculture with different level of disadvantage	
Hill and mountain systems with a low level of disadvantage	Cluster 4
Least favoured mountain systems	Cluster 3

Cluster	1	2	3	4	5	Emilia-Romagna
Municipalities	103	71	38	91	38	341
% of regional population	31,1	24,1	2,0	7,9	34,9	100
% of regional surface	21,8	27,0	15,0	25,1	11,2	100
Per capita GDP	30,1	23,9	16,7	20,4	26,1	24,3
Pop. density (inh / km2)	233,3	139,9	24,2	63,2	617,7	188,0
% employees agriculture	10,5	16,2	29,6	23,4	11,3	17,4
% employees industry	52,5	40,9	21,9	34,8	34,5	39,9
% employees services	36,9	42,9	48,6	41,8	54,2	42,7
Unemployment ratio	5,5	8,6	8,3	7,1	9,1	7,3
Ageing index	157,8	198,6	468,9	216,8	169,0	217,9
Female activity ratio	57,7	52,3	42,1	50,7	54,7	52,6
Dependence ratio	42,4	44,4	69,3	52,1	42,2	48,4
Graduated/school aged	23,5	21,8	17,5	19,5	37,0	22,9
UAA per farm	8,6	14,2	7,0	7,6	5,0	8,9
% farms 0-2 ha UAA	33,9	22,8	33,1	36,6	45,6	33,5
% farms 50+ ha UAA	1,9	4,8	1,0	1,7	0,5	2,2
% UAA farms 0-2 ha	4,9	1,7	5,7	5,0	13,1	5,3
% UAA farms 50+ ha	19,0	33,0	20,2	14,7	22,7	21,3
Tractors per ha UAA	0,2	0,2	0,1	0,1	0,2	0,2
% woodlands on ag. sur.	6,0	5,7	48,1	25,0	3,7	15,5
% UAA – Cereals	28,2	38,6	5,5	20,0	28,8	25,7
% UAA - Feeding crops	33,9	20,3	32,3	51,3	11,6	33,0
% UAA - Pastures	6,9	4,0	59,0	16,5	1,7	14,1
% UAA - horticulture	1,6	7,5	0,1	0,6	4,7	2,7
% UAA - Fruits	5,2	7,9	0,0	2,4	24,4	6,6
% UAA - vineyards	9,2	1,3	0,7	3,5	13,4	5,6
% CDO vines on vines	32,2	5,8	2,5	11,3	19,1	16,4
% farms with livestock	53,6	52,2	55,9	54,2	48,8	53,2
Bovines / ha UAA	1,2	0,6	0,4	0,7	0,2	0,8
Pigs / ha UAA	3,2	0,7	0,3	1,2	1,2	1,6
Chickens / ha UAA	12,8	8,9	3,9	20,0	91,5	21,7
Bovines / UAA Pastures	4,6	3,5	0,0	0,5	1,4	2,4
Pigs / Cows	5,5	44,0	7,8	13,6	82,7	24,5
AWU / ha UAA	0,1	0,1	0,1	0,1	0,2	0,1
Agr. SGP / ha UAA	3,5	3,4	0,6	1,5	4,9	2,8
Agr. SGP / AWU	29,1	43,4	8,8	16,2	23,6	25,8
Food Industry: empl./firms	8,2	13,8	3,5	4,3	10,3	4,7
AWU / food industry empl.	0,6	0,7	0,2	0,3	1,0	0,5
% empl. large food firms	10,1	25,3	0,0	1,8	26,1	11,7
Population change 81-91	6,5	-0,2	-12,5	-0,7	2,0	0,6
UAA % change 81-91	-2,1	-1,0	37,0	-5,1	-0,5	1,9
Farms n. % change 80-90	-11,9	-12,6	-28,5	-11,2	-5,2	-13,0
Ag. surf. % change 80-90	-2,6	-2,3	-15,9	-4,1	1,1	-4,0
% mountain municipalities	0,0	4,2	94,7	30,8	0,0	19,6

Analysis "per blocks"

The 40 indicators were subdivided in 5 macro-determinants: social and economical development (including the indicators related to food processing industry), agricultural structure, demographic and social indicators, specialisation and dynamic indicators, collecting all the variables related to the demographic and structural dynamics with respect to the previous decade. In this case the components extracted from the 40 indicators are 14 altogether. The cluster analysis on the (non standardised) scores resulted in 5 significant clusters, as described in table 4.

The "two stages" approach

In this part of the work a third approach was attempted, carried out through an analysis in two stages. In the first stage an analysis "on the whole" was carried out on the first grid of 26 indicators (see par. 2), considered to be essential for detecting a limited number of macro-areas of agricultural and rural development, whereas in the second stage the remaining specific indicators were utilised. Here we report the application to the agricultural system characterised by different level of disadvantage[8]. The aim is to test the possibility to carry out a general analysis of the rural and agricultural development of a region, then further analyse the territorial systems that the policy-makers assume to be a priority for intervention.

Plugging in the 26 indicators resulted in 8 principal components with eigenvalues above 1, that explain altogether 71% of the total variance. The 341 municipalities were then clustered into 5 groups (table 5).

As stated above, in the second step of the analysis, the PCA and cluster analysis were carried out only on the 112 municipalities belonging to the less favoured areas, in order to better understand the causes of the different levels of disadvantage. The resulting territorial sub-systems are presented in Table 6.

This further subdivision allows, for example, to implement with better precision the policy instruments of the compensatory payments, according to the different disadvantage level.

A comparison between the different approaches

The results of the cluster analysis carried out on the original indicators, without employing the PCA, are summarised below, in order to derive some indication about the differences between the approaches applied here. The local "optimal" mapping according to the pseudo F statistic produced 5 clusters, with 86% of the municipalities concentrated in one single cluster, whereas 97% of the municipalities were grouped into two clusters. This result can be interpreted as the extreme consequence of ignoring the correlation between the indicators.

Hence the analysis "per blocks" appears to be exposed to significant biases. The "two-stages" approach offers a solution to conciliate the research need to give relevance to a specific set of variables, often pursued through the analysis "per blocks", and the respect of the statistical requirements. The need for deriving uncorrelated components for the CA can be satisfied by applying the approach "on the whole" in two or more successive stages. Such a technique would not eliminate the possibility to value the "a priori" knowledge of the phenomenon being studied and isolate groups of indicators considered especially relevant under an economic point of view. At the same time another advantage of dividing the analysis into successive stages, would be to avoid the statistical problems internal to the analysis "per blocks", making more transparent the effects of the researches' choice.

Table 5. Two-stages analysis (first stage) - Description of detected clusters (territorial systems) and average of the most relevant indicators and other statistics at municipality level

Areas with high level of social and economic development and high agricultural productivity	
Intensive hill and plane agricultural systems	Cluster 4
Hill and plane vineyard and fruit production system	Cluster 2
Extended city agriculture, in territories with strong tertiary sector	Cluster 3
Areas with high agricultural productivity and an average level of development	
Plane cereal crops and intensive horticulture and fruit production systems	Cluster 1
Agriculture in less favoured and less developed areas (cluster 5)	

Indicator	1	2	3	4	5	Region
Municipalities	64	48	16	101	112	341
Per capita GDP	23,25	25,43	26,67	29,72	19,20	24,30
Population density	146,30	215,92	750,84	270,49	45,15	188,03
% employees agriculture	16,75	16,53	7,70	10,63	25,50	17,36
Unemployment ratio	8,93	7,06	11,94	5,44	7,37	7,26
% farms under 2 ha	23,25	40,51	59,58	30,78	35,10	33,51
% farms above 50 ha	4,51	1,46	0,24	2,08	1,58	2,20
% UAA of farms under 2 ha	1,73	6,22	26,50	3,71	5,20	5,25
% UAA of farms above 50 ha	31,80	13,06	13,67	24,98	16,65	21,32
Tractors per ha UAA	0,17	0,21	0,23	0,16	0,12	0,16
Woodlands % on total ag. Surf.	4,03	5,39	2,69	7,95	34,94	15,47
Cereals % on total UAA	39,92	24,92	35,64	28,83	13,65	25,70
Feeding crops % on UAA	16,43	18,39	16,09	38,99	45,86	33,04
Pastures % on total UAA	2,12	2,95	1,63	8,34	32,69	14,10
Horticulture % on total UAA	7,52	1,47	8,82	2,11	0,27	2,74
Fruits % on total UAA	9,70	19,97	12,03	3,45	1,10	6,58
Vines % on total UAA	1,48	18,29	7,78	5,41	2,24	5,56
CDO Vines % on vines UAA	6,73	40,73	20,10	21,64	6,10	16,35
% Farms with livestock	49,85	52,80	47,95	53,88	55,38	53,18
Bovines per ha UAA	0,51	0,59	0,34	1,25	0,59	0,76
Pigs per ha UAA	0,70	2,71	1,23	2,69	0,67	1,59
Chickens per ha UAA	10,27	40,94	147,39	8,35	14,04	21,69
Bovine heads / ha pastures	3,79	5,58	0,62	2,92	0,12	2,43
Pigs per cow	43,44	25,40	144,80	5,21	13,65	24,55
AWU per ha of UAA	0,08	0,17	0,28	0,11	0,08	0,11
SGM per ha of UAA	3,71	4,50	4,61	3,06	1,04	2,79
SGM per AWU	46,01	27,36	17,17	28,30	12,41	25,75
Food industry empl. per firm	11,29	8,83	8,21	9,35	4,72	8,07
Population change 81-91	-0,90	4,31	5,30	4,56	-4,45	0,58
Ageing index	198,05	170,52	124,21	168,99	307,17	217,94
Ratio of female workers	52,92	56,52	51,88	55,96	47,94	52,64
Dependence ratio	43,68	44,65	39,74	43,53	58,22	48,36
Graduating index	21,99	24,52	25,11	27,10	18,75	22,94
% employees industry	41,62	45,13	30,56	48,48	30,41	39,94
% employees services	41,63	38,34	61,74	40,88	44,09	42,70
UAA change (80-90)	-0,14	-1,85	-1,45	-1,71	8,27	1,85
Farms change (80-90)	-11,26	-7,47	-3,72	-13,71	-16,95	-12,97
Agr. surface change (80-90)	-1,35	-0,76	0,94	-2,64	-8,81	-3,99
Agr. worker per food firm	0,59	0,35	1,25	0,79	0,27	0,54
% empl. in large food firms	21,11	15,42	21,92	12,87	2,22	11,70
Share of mountain municip.	1,5%	0%	0%	1%	65%	67%

Table 6. Two-stages analysis (second stage) - Description of detected clusters (territorial sub-systems) and average of the most relevant indicators and other statistics at municipality level

Territorial system with low agricultural productivity and an average development disadvantage (presence of Industrial activities and craftmanship)	*(cluster 2)*
Territorial system with higher agricultural productivity and an average disadvantage in the development	*(cluster 4)*
Areas with extreme agricultural marginality and a low development level	*(cluster 1)*
Areas with extreme agricultural and development marginality	*(cluster 3 and 5)*

Indicator	1	2	3	4	5	Total
Municipalities	24	26	16	42	4	112
% of regional surface	9,8	6,7	6,7	12,4	0,9	36,7
% of regional population	1,4	2,1	1,0	3,6	0,1	8,1
Per capita GDP	17,97	22,05	15,65	19,87	15,13	19,20
Population density	25,11	62,22	27,51	55,96	11,38	45,15
% Employees agriculture	26,06	14,92	32,91	27,13	44,16	25,50
% Employees services	54,73	40,98	42,41	40,93	40,45	44,09
Unemployment ratio	7,93	6,66	7,66	7,18	9,43	7,37
Ageing index	348,67	223,93	396,56	211,92	1241,63	307,17
Avg. UAA per farm	7,70	8,12	7,71	7,21	5,52	7,54
% Farms under 2 ha	31,95	39,67	25,87	35,76	54,30	35,10
% Farms above 50 ha	1,31	2,24	1,27	1,48	1,37	1,58
% UAA of farms under 2 ha	4,22	5,84	3,62	4,65	19,13	5,20
% UAA of farms above 50 ha	14,70	17,37	13,35	13,02	4,93	16,65
Woodlands % on total surface	50,45	33,77	43,53	23,03	40,16	34,94
Pastures % on total UAA	52,90	27,18	41,90	17,64	68,33	32,69
Feeding crops % on total UAA	34,84	45,68	48,69	53,75	18,98	45,86
Bovines % on total UAA	0,42	0,47	0,64	0,76	0,33	0,59
Bovine heads/ pastures ha	0,02	0,26	0,03	0,13	0,04	0,12
SGM per ha of UAA	0,62	1,05	0,83	1,40	0,44	1,04

Some final considerations

The analysis allowed some final consideration, both under the analytical profile and under the perspective of the detection of the most appropriate tools to face the complex and heterogeneous situation of the European agricultural and rural scenery.

With regard to the analytical aspect, the study especially highlights the extremely strong influence of the methodological approach. Starting from the same data set, the analysis “on the whole” and “per blocks” return very different mapping of the agricultural development. Of course it is still an open question whether these results could be generalised and it is a difficult task to be investigated, given the arbitrary elements inner to the subdivision in macro-determinants on which the analysis “per blocks” is based. Such a subdivision does not influence the results only if there is no correlation between indicators belonging to different macro-determinants, a situation that rarely arises in practice. On the other hand, the analysis “on the whole” allows to obtain, albeit with extreme care, some indications on the most relevant indicators to explain the differences in agricultural development, according to the approach indicated in section 2. Hence, the analysis “on the whole” seems to be preferable, for a lower degree of arbitrariness and for the interpretative advantages. The “two stages” approach, despite being more elaborated, seems to be richer in terms of results, also for widening the analysis of the agricultural and rural development. Under the

analytical point of view the advantages of the analysis "on the whole" also emerge in this approach: in the first stage only 2 out of the 5 territorial systems detected appeared to be less defined than in the analysis on the whole, leading to the second stage analysis.

Under the perspective of economic policy, the two stages analysis allows to choose the second stage indicators for the specific local situations. These could be also defined directly by the policy-maker, according to the various problems arising in the regional planning. Finally, the two-stage analysis seems to be more useful in making comparisons between different regional situations, but also for the same region in different time periods. In fact, the second stage indicators could be monitored ex post, in order to evaluate the effectiveness of the measures adopted.

Notes

[1] For a more formalised description of the PCA see Krzanowski (1988) and Mignani and Montanari (1993). Fanfani and Mazzocchi (1999) analyse in detail the issues of the application to the territorial analysis of rural development.

[2] In the case study described hereafter, the method of Guttman-Kaiser was adopted to choose the number of principal components to be considered. The method advises to retain the components whose eigenvalue is equal or larger than 1. Generally this criterium led to the selection of a number of components explaining between 65% and 70% of the original data variance.

[3] Working on the correlation matrix means in practice standardising the original data matrix and avoiding that the indicators with larger variability exercise a distorting influence on the principal components extraction.

[4] As a matter of fact, acting on the correlation matrix implies that the sum of the eigenvalues is equal to the number of considered indicators.

[5] Most of this information is also available in the Eurostat data bank REGIO, so a comparison in the whole EU territory should be possible.

[6] These indicators have been utilised as benchmarks for the detection of the 5b municipalities and, according to the AGENDA 2000 regulation proposals, will also be utilised to define the areas included in the new objective 2. A new variable was added to these indicators, reporting the number of employees per local food processing unit, in order to evaluate if the presence of the food industry is significant.

[7] For reasons of synthesis the full interpretation of the components emerged from the case study is not reported here. Roughly summarising, it returned the following components: level of economic and agricultural development, level of fragmentation of the farm structure, relevance of extensive cereal crops and intensive horticulture, level of rurality, level of integration with the food industry, level of specialisation of agriculture, presence of integrated agro-food districts, presence of poultry breeding, presence of large farms with wine and animal production, level of services and farm concentration, presence of very large farms, intensity of bovine breeding.

[8] For the social and demographic dynamics the indicators are the variation of the resident population between 1981 and 1991, of the employment shares for industry and services, the ageing index, the female activity index, the dependence index, the graduation index. In order to highlight the structure and dynamics of the primary sector the indicators were: dynamics of the total agricultural surface, UAA variation, change in the number of farms, relevance of the woodlands areas and of the livestock farming.

References

Anania, G., Tarsitano, A. (1995) "Tecniche di analisi statistica multivariata per l'individuazione dei sistemi agricoli territoriali in Italia". In G. Cannata (ed.), *I sistemi territoriali agricoli italiani degli anni '90. Contributi metodologici*. Cosenza: Rubbettino. 105-242.

Boccafogli F., Brasili C. (1998) "L'articolazione territoriale dello sviluppo agricolo in Emilia-Romagna", *La Questione Agraria*, 70: 67-93.

Brun A., Cavailhes J., Perrier-Cornet P., Schmitt B. (1992), "Les Espaces Ruraux Revisités", *Revue d'Economie Régionale et Urbaine*, 1: 37-66.

Cannata G. (ed, 1989), *I sistemi agricoli territoriali Italiani*, Milano: Franco Angeli.

Cannata G., Forleo, M.B. (eds. 1998), *I sistemi agricoli territoriali delle regioni italiane* CNR-IPRA

Fanfani, R., Mazzocchi, M. (1999), "I metodi statistici per l'analisi dei sistemi agricoli territoriali", *Serie Ricerche, n.2, Dipartimento di Scienze Statistiche, Università di Bologna*

Forleo, M.B. (1998), "I sistemi agricoli territoriali delle regioni italiane. Aspetti metodologici ed operativi" in Cannata and Forleo (eds.), *I sistemi agricoli territoriali delle regioni italiane* CNR-IPRA.

Kayser B., Brun A., Cavailhes J., Lacombe P. (1994), *Pour une Ruralite Choisie,* Paris: Editions de l'Aube.

Krzanowski (1988), *Principles of multivariate analysis*, Oxford: Clarendon Press.

Mignani S., Montanari A. (1997), *Appunti di analisi statistica multivariata*, 2nd edition, Bologna: Esculapio.

Montresor, E. (1998) "Gli effetti regionali della politica comunitaria agricola", *Proceedings of the Jean Monnet Seminars*

OECD (1994), *Creating rural indicators for shaping territorial policy,* Paris: OECD.

OECD (1996), *Territorial indicators of Employment - focusing on Rural Development*, PARIS: OECD.

Paris, Q., Montresor, E., Arfini, F., Mazzocchi, M. (2000) "An Integrated Multi-phase Model for Evaluating Agricultural Policies Through Positive Information", *Paper presented at the 65th EAAE Seminar,* Bonn, 29-31 March.

Roux B. (1992), "Marginalisation et Développement Local dans les Espaces Ruraux de l'Europe du Sud", *Revue d'Economie Regionale et Urbaine*, 4:

Sotte F. (ed., 1995), *The Regional dimension in Agricultural and Economics and Policies*, Ancona: CNR-RAISA.

Stilwell F. (1991), "Regional Economic Development: an Analytical Framework", *Revue d'Economie Regionale et Urbaine*, 1: 107-115.