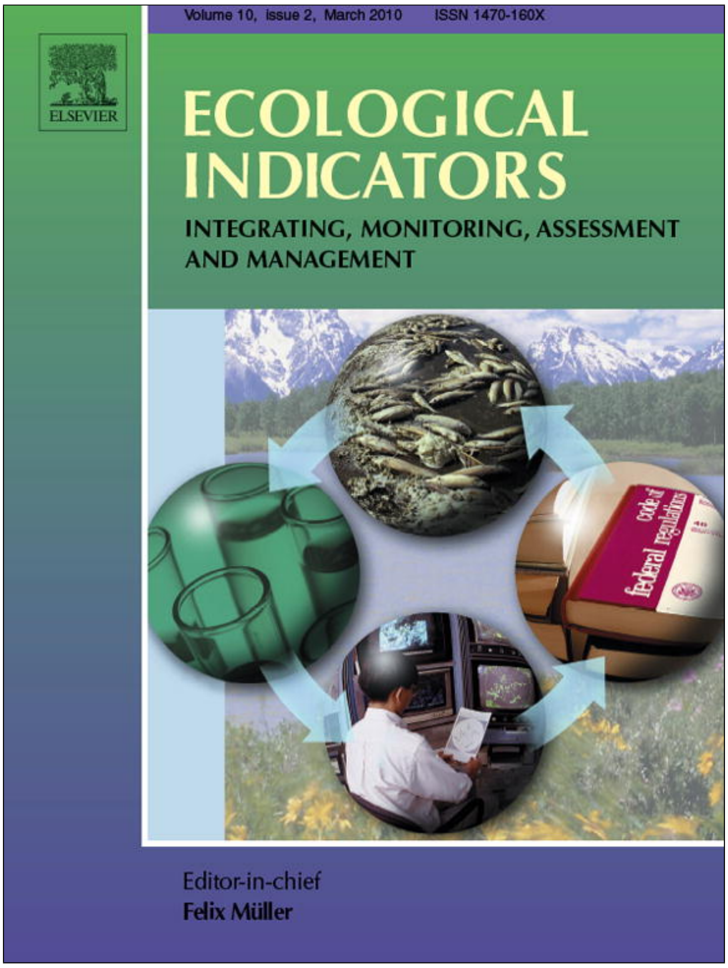
Volume 10, issue 2, March 2010
ISSN 1470-160X
ELSEVIER
ECOLOGICAL INDICATORS
INTEGRATING, MONITORING, ASSESSMENT AND MANAGEMENT
code of federal regulations
Editor-in-chief
Felix Müller

Ecological Indicators 10 (2010) 516–526

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Procrustes analysis as a tool for land management

Silvia D. Matteucci [a,*], Laura Pla [b]

[a] Consejo Nacional de Investigaciones Científicas y Técnicas, Grupo de Ecología del Paisaje (GEPAMA), Universidad de Buenos Aires, Washington 1821, C1430ETA Buenos Aires, Argentina
[b] Universidad Nacional Experimental Francisco de Miranda (UNEFM), Department of Agricultural Technology, Coro, Venezuela

ARTICLE INFO

Article history:
Received 28 February 2009
Received in revised form 21 September 2009
Accepted 24 September 2009

Keywords:
Ecological indicator
Matrix concordance space
Environmental policy
Land use
Concordance Class
Pampa Ecoregion
Argentina

ABSTRACT

Generalized Procrustes analysis (GPA) is a multivariate technique that involves transformations of data matrices to provide optimal comparability. We propose GPA to quantify the concordance among sets of variables that characterize natural, human and productive subsystems. When the land use fits in with the physical support of agricultural production, people's well-being should be evident in a high concordance between the land use and the social conditions. In a situation of instability each set of variables operates in diverse directions resulting in lower resilience and sustainability. Two GPA were performed, between physical support and land use data sets (concordance = 67.4%), and between land use and social conditions data sets (concordance = 65.3%). The interplay between the pair of concordance values constitutes a bi-dimensional index which serves as an ecological indicator. Based on bootstrap confidence interval, the 49 counties of the Pampa Ecoregion, Argentina, were classified in medium, high or low concordance. The lack of concordance is an indicator of imbalances which may contribute to guide environmental management.

1. Introduction

Environmental problems tend to accelerate at a faster rate than the capture and updating of biophysical and socioeconomic information, particularly in Argentina, where databases have always been scarce and outdated. This constitutes a serious drawback at a time when decision making is urgent. The use of statistical methods to minimize uncertainty in environmental management has become a common practice. Most of the multivariate methods are based on mean values taking into account variable association through covariance or correlation matrices (Jenerette et al., 2002; Caeiro et al., 2003; Jansen, 2006). The Generalized Procrustes analysis (GPA) is a multivariate technique that involves transformations (translation, rotation, reflection, isotropic rescaling) of individual data matrices to provide optimal comparability (Gower, 1975). This method has been routinely used in food science to analyze sensory data specially with free choice profiles for scoring the samples (Dijksterhuis, 1994), to investigate association between sets of site properties and biological communities through concordance between site classifications based on environmental factors and species assemblages (for example, Jackson and Harvey, 1993), and recently applied to characterize entries in a germplasm bank (Bramardi et al., 2005). However, GPA has not been applied in the classification of administrative entities on the basis of concordance among sets of variables that characterize each of their subsystems (natural, human and production).

In a stable system, these sets of variables should represent the entity's state in a similar fashion; that is, they should associate through stabilized mutual interactions. On the other hand, in a situation of instability each set of variables operates in diverse directions causing the reduction of resilience and sustainability. Each data set summarizes key attributes in the system's functioning. Whenever the land use data set concurs with the physical resources data set, a coincidence between the land use and the social well-being data sets is expected. Public policy for sustainable management should aim at maintaining and improving this consistency.

We hypothesize that, if the land use fits in with the physical support of agricultural production, people's well-being should be evident in a high concordance between the land use and social data sets. Since the smallest administrative unit is the county, in order to help decision makers at this level, we propose a statistical procedure to classify the counties according to the degree of concordance between pairs of data sets resulting from GPA. The interplay between the pair of concordance values constitutes a bi-dimensional index which serves as an ecological indicator to contribute in the guidance of sustainable management.

* Corresponding author.
E-mail addresses: smatt@arnet.com.ar (S.D. Matteucci), laura@reacciun.ve (L. Pla).

1470-160X/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ecolind.2009.09.005

2. Materials and methods

2.1. The study area

The area under study is located in Buenos Aires Province, Argentina, within the Pampa Ecoregion, which is a sedimentary basin filled with loess deposits during the Quaternary. The climate is temperate, milder than in other regions at the same latitude due to the temperature moderating effect of the ocean; thus, snowfall is absent, large diurnal or seasonal temperature amplitudes do not occur. This, together with evenly distributed rainfall allows for year-round cropping. It is divided in subregions, according to the general topography, which affects the present drainage system and soils, each of these subregions support different economic activities. Our study comprises 49 counties within two of the subregions: the Rolling Pampa and the Flooding Pampa (Fig. 1).

The Rolling Pampa, where the most productive agricultural lands of Argentina are encountered, is one of the five extensive loess areas of fertile soils of the world. It has enough rainfall to produce sustained high yields of soybeans, wheat, sunflower and corn, representing 52% of the national agricultural production value. The native grassland has been converted to croplands, and only very few and isolated relics remain. The capital city Buenos Aires and its metropolitan area are also located in this subregion. Urban growth, mainly metropolization, has triggered a fierce conflict with agriculture, and considerable extensions of farmland have been irreversibly lost (Morello et al., 2001).

The Flooding Pampa is formed largely by a low plain, originally covered by natural grasslands in around 80% of its extension. Its flat topography, lack of a well developed drainage system, and low hydraulic conductivity of soils, determine the occurrence of floods in late winter and spring, while droughts are frequent in summer. Agriculture is impeded by flooding, soil salinity and low fertility. The main economic activity is ranching on natural grasslands. In a lower proportion, managed pastures and fodder are included in cattle raising practices. In the northern portion there is an important milking production. Crops are limited to the few hillocks sticking out the plain.

In a previous research (Matteucci, 2006) certain degree of imbalance between the social, natural and economic subsystems was uncovered in the total human system. The counties' vulnerability is higher in the neighborhood of the metropolitan area, showing a higher risk of agricultural lands loss. At a higher spatial scale, vulnerability decreases to the south and the north, and increases again in the tourist counties of the Atlantic coastal zone.

The influence of the external conditions, such as international market process, on the functioning of the agricultural systems in the Pampa Ecoregion has been pointed out (Morello and Matteucci, 1997), and this is valid for the analysis of the agricultural evolution in this region. However, there are local factors operating at smaller time and space scales that affect land use pattern, particularly in relation to urban growth. These factors and their interrelationships vary across the study area, which allows the comparison among the various production zones and counties within it.

The 49 counties of our study area were classified in three Land-use Groups, according to the main economic activity: agriculture, ranching, and mixed agriculture-ranching (Matteucci, 2006) (Fig. 1).

The 20 counties belonging to the Agriculture Group (2,992,000 ha) are located in the northern portion of the study area, within the Rolling Pampa. Their lands are mainly extended hillocks (51%) and slopes (24%), both filled with loess sediments. An average of 65% of their lands show high agricultural capacity and only 5% are unproductive. Croplands occupy 71% of their territory, mostly with annual crops (41%). In two counties there has been a recent exurban growth, mainly on agricultural lands (Matteucci and Morello, 2006).

The Agriculture-Ranching Group comprises 20 counties (3,039,000 ha) in the central portion of the study area. It represents a transition between the agricultural zone of the Rolling Pampa and ranching zone of the Flooding Pampa. An alluvial plain covers 13% of its area and 43% is occupied by low extended hillocks, similar to those of the Agricultural Group. Medium productivity capacity lands predominate. The mean number of farms per county is the highest in the study area, yet this difference among Land-use Groups is not statistically significant ($p = 0.999$). The average farm size is significantly smaller (304.58 ha) than in the ranching group ($p < 0.001$).

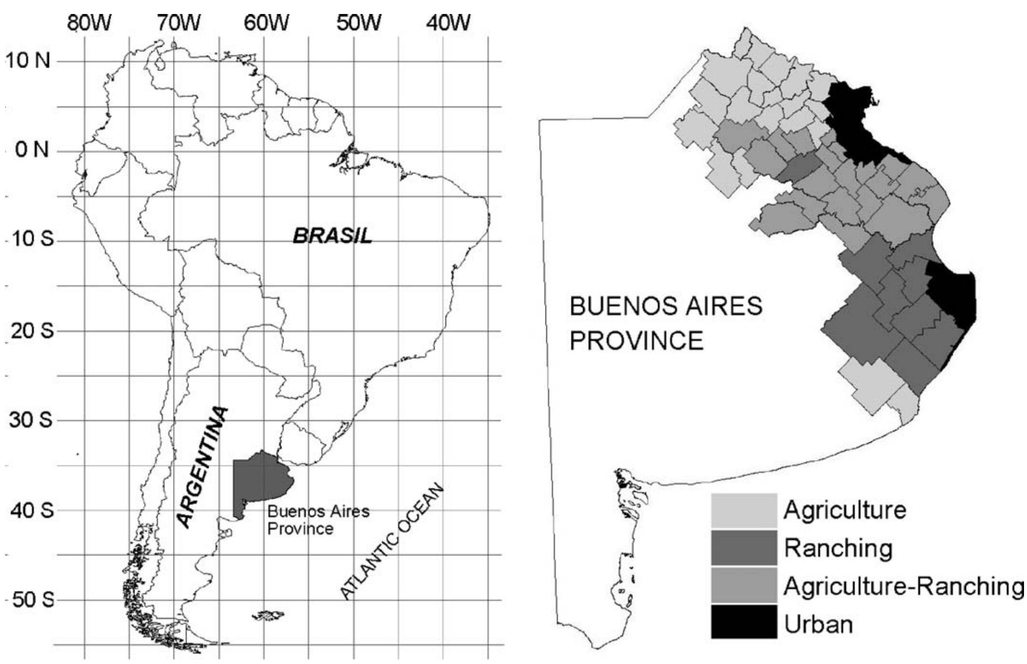

Fig. 1. Study area showing the spatial distribution of counties in the Land-use Groups. Adapted from Matteucci (2006).

The ranching group includes nine counties (2,689,000 ha). It occupies the lowlands extending along a river channel. It has a high proportion of flood prone lands (52% of the county). The farms are the largest in the study region, with an average of 749 ha. There is a high percentage of natural grassland coverage (70%) with a dominance of low herbaceous vegetation (78%), with no statistically significant differences with the other Land-use Groups. It has the highest proportion of natural ponds, with an average of 7% of the counties' extension.

2.2. The data

The combination of data from various disciplines poses difficulties because primary data sources are not homogeneous; either the data are not spatially explicit or they are at different temporal and spatial scales and resolutions. In ecological research, the spatial description is based on physical or biological criteria (geomorphology, soils, vegetation), which results in a classification of landscape units. Instead, social statistics relate to administrative units. Natural factors may be adjusted to administrative boundaries, but it is difficult to adjust the latter to the natural boundaries. Thus, our reference unit is the county, a subdivision of the province. The advantage of presenting the results at the administrative unit level is that the information, results and conclusions may be applied by local government officials in planning and management, and it may also be simplified for use at higher organization levels. In this paper the limitations of primary data sets were taken into account and spatial and temporal scales were unified as much as possible.

Three data sets were used (Table 1), and each one includes variables that bear on a particular subsystem. Each set of variables is taken from an independent source, but this does not mean that the sets are independent; because of this reason we used multivariate analysis.

2.2.1. Physical support of agricultural production

The information on soils was obtained from the Soils Atlas of the Argentine Republic (Maccarini and Baleani, 1995). Two thematic layers were extracted from the Atlas: landform units and cartographic productivity index. Each theme was cut out using the study area limits as a mask in order to simplify maps and tables handling. The 26 landform types in the Atlas were gathered into six landform types: plains, hillocks, alluvial plains, flooding lowlands, slopes, and non-productive lands (mostly sand deposits). The percentage cover of each landform type was calculated for each county.

The productivity index represents the relative capacity of lands to produce the main grain crops grown in a land unit under a consistent level of management, and it is expressed as a percentage of the potential productivity (the highest the value, the nearest the actual productivity to the potential productive capacity). It was developed by Brickman and Smith (1973) and adapted to the local conditions. The productivity index (IP) is calculated as a function of the most relevant factors for soil conditions, including climate, and soil drainage, texture, salinity, alkalinity, organic matter, depth, cationic exchange capacity and erosion. The continuous series of values obtained from the Atlas was ranked in four classes of productivity capacity: 0–30 (very low); 31–50 (low); 51–70 (medium) and 51–95 (high). For each county, the fraction of its territory occupied by each class was calculated.

2.2.2. Land use

The data source for agricultural productivity is the National Agriculture Census of 2002 (INDEC, 2006). Values per county were obtained directly from the census or derived from other census statistics. Land uses were classified in planted lands, annual and permanent crops, annual and permanent forages, forest plantations, and natural grasslands. The percentage cover of each land use was calculated for each county.

2.2.3. Social conditions

From the 2001 National Census (INDEC, 2001), those variables perceived as indicators of social conditions were selected. They include demographic (people and housing), social (basic infrastructure, education, well-being and poverty) and economic

Table 1
Data sets.

Physical support variables (% of the counties area)	
Ipc0-30	Very low productivity capacity index
Ipc31-50	Low productivity capacity index
Ipc51-70	Medium productivity capacity index
Ipc71-95	High productivity capacity index
TIN	Flooding lands
LO	Hillocks
PEN	Slopes
PLA	Plains
PAL	Alluvial plains
Land use statistics	
NEAP	Total number of farms
PI	Percentage of planted area per county
PPN	Percentage of the county area under natural grasslands
TPEAP	Mean size of farms (in hectares)
PFOR	Percentage of the county area under forestry
%cult	Percentage of the county area under annual and perennial crops
%forraj	Percentage of the county area under annual and perennial fodder crops
Social statistics	
Var%	Percentage of population variation between 1991 and 2001
Dpob	Population density (habitants per km^2)
NBI	Population with unsatisfied basic needs as percentage of total population
%Anal	Population 10 years old or older that is illiterate as percentage of total population
OS	Population in households with no social security income as percentage of total population
VD	Population living in houses with some deficiency as percentage of total population
ED	Population 10–14 years old that have never attended school as percentage of total population
JU	Population 70 years old or older not perceiving retirement pension as percentage of total population
PobRur	Percentage of rural population

indicators. All the variables are expressed as percentage of the county's population in the condition represented by the corresponding variable.

2.3. Statistical techniques

The variables are expressed in various measurement units, and most of them are in percentages. For all analysis, those values expressed as percentages were transformed to $\log(x + 0.05)$ to avoid the mixture of linear effects of absolute variables with multiplicative effects of percentages. Addition of 0.05 precludes indetermination with zero value.

Multiple comparisons between multivariate means of Land-used Groups were performed using re-sampling applied to data vector in order to automatically incorporate correlations between the multivariate observations. To detect mean differences in variable sets a test for each one was performed using 'proc.mult-test' (Westfall and Wolfinger, 2000) with *p*-adjusted by bootstrap. Results are presented in original scales using inverse transformation of variable means.

Canonical discriminant analysis using the 49 counties and 25 variables was performed using InfoStat (Di Rienzo et al., 1998). Significant dimensions that identify the variables that best discriminate between Land-use Groups were retained, based on the likelihood ratio for the hypothesis that the current canonical correlation and all smaller ones are zero in the population. Approximate F statistic based on Rao's approximation to the distribution of the likelihood ratio was used (Rao, 1973).

2.3.1. Detection of stabilized mutual interactions between data sets

In order to achieve the purpose, a multivariate technique is required to identify associations among data sets, each of which is composed by associated variables, and considering that the data sets act concurrently (Krzanowski, 2000). To interpret the relationships that lead to the partition of objects (counties) according to the patterns of variable behavior, an integrated approach as the following was applied.

Each data set is defined in a hyperspace with as many dimensions as variables, and, the number of dimensions of data sets differs from each other. Each object is represented by a point in each of the three hyperspaces of physical support, land use and social data sets. In each hyperspace the objects are present in a distinct configuration, and the distance (or dissimilarity) between them can be calculated. In our study area we have 49 points representing the location of the counties in each hyperspace (with dimensions 9, 7 and 9, respectively). These spatial configurations may or may not match, and the purpose of the analysis is to find a common joint hyperspace in which to represent the counties preserving as much as possible the original relationships between objects seen through data sets. In this new hyperspace, the counties may be projected through their data sets, and those that fit can be segregated from those that do not.

For this analysis, the relevant parameter is the dissimilarity between pairs of counties, and not their absolute condition, which was mainly taken into account to define the three Land-use Groups. Matches between configurations can be detected by overlaying the hyperspaces; however, this is not a direct process since the data sets are different. The analysis aims at finding the "best" hyperspace, in which the configurations can be reproduced. This new hyperspace, called concordance space, common mean space, or joint space is considered "better" because it preserves the relative distance between objects.

Let X_k be the data matrix, with the n rows or counties, and P_k columns or variables corresponding to the k-th data set. The joint space should fulfill the condition of multi-set minimization, such that

$$\sum_{k<k'}^{k} ||X_k Q_k - X_k Q_{k'}||$$

has a minimum value, and the transformation matrices Q_k are orthogonal. As a result of this transformation, a joint space can be derived as:

$$G = K^{-1} \sum_{k=1}^{K} ||X_k Q_k||$$

This transformation includes rotation, translation and scaling of the hyperspaces to be superimposed; the procedure is known as Generalized Procrustes analysis (Gower and Dijksterhuis, 2004). The object (county) may be represented in the joint space, together with the relative position of each original configuration. For some objects, their position in the joint and original configurations will not differ greatly, while for others, the difference in configurations may be large. These are distances in the joint space, and thus they can be quantified and compared. Through the Generalized Procrustes analysis (GPA), the fraction of variability represented in the joint configuration (JCV) can be known, and the lack of representation of each data set in each county can be quantified.

Two GPA were performed, between physical support and land use data sets, and between land use and social conditions data sets using Gower's distance as dissimilarity coefficient. GPA was done by pairs of variable sets because land use must be compared separately with the other two data sets in order to test the hypothesis. The ecological index results from the interplay between the pair of concordance values. Gower's distance was chosen because it is appropriate to handle variables expressed on various scales (Gower, 1971).

A 0.90 probability confidence interval was established applying bootstrap techniques for the joint configuration variability (JCV). We established a null model that preserves the covariation among variables of the same set and supposes independence between pairs of sets. This null model is a reference to compare the variability due to association of natural or human subsystems with the production subsystem. The JCV confidence interval was established with bootstrap samples ($B = 300$) of $n = 49$ counties for the complete vector data sets. The limits of the confidence intervals were found as 5% and 95% percentiles of the empirical distribution of the joint configuration obtained by re-sampling.

Each county was classified according to its own joint configuration (JCVc) in three cases: similar to the mean (m) if JCVc belongs to the JCV confidence interval; larger than the mean (H) if JCVc exceeds the higher limit of the confidence interval, and lower than the mean (L) if JCVc falls below the lower limit of the confidence interval. Each county is classified twice according to each pair of data sets based on the ecological index. Thus, the 49 counties were classified in nine clusters resulting of the intersection of the three cases. The most relevant clusters are those that show a bias from the mean; that is, the counties with a high concordance in both pair of data sets (HH), those with low concordance in both pairs of data sets (LL), and those with high concordance between physical support and land use, and low concordance between social conditions and land use (HL), and its inverse (LH). These four clusters are called Concordance Classes, since they represent common aspects in relation to environmental management.

The GPAs were done in R with FactoMineR (Husson et al., 2007), and the variables with higher loadings for the resulting concordance space were identified. The bootstrap test for JCV was programmed in R.

Table 2
Inference for multiple contrasts and multiple variables simultaneously in Land-use Group means.

		Land-use Group (number of counties)		
		Agriculture (20)	Agriculture-Ranching (20)	Ranching (9)
Geophysical support	Ipc0-30	12.59 b	22.91 ab	57.54 a
	Ipc31-50	0.11 b	3.55 a	7.76 a
	Ipc51-70	1.10 a	26.92 b	1.86 ab
	Ipc71-95	56.23 b	0.39 a	0.14 a
	TIN	7.94 a	0.78 b	50.12 a
	LO	30.90 a	40.40 a	0.95 b
	PEN	10.00 b	0.06 a	0.10 a
	PLA	0.42 a	5.13 a	7.76 a
	PAL	0.34 a	10.96 b	0.22 a
Land use	NEAP	381 a	464 a	365 a
	PI	70.79 b	21.38 a	14.13 a
	PPN	19.05 b	53.70 a	69.18 a
	TPEAP	320 a	305 a	749 b
	PFOR	0.44 a	0.31 a	0.12 a
	%cult	21.88 a	4.68 a	6.76 a
	%forraj	6.76 a	3.72 a	3.89 a
Social	Var%	4.47 a	13.18 a	2.69 a
	Dpob	57.82 a	62.10 a	5.21 a
	NBI	19.95 a	12.30 a	10.72 a
	%Anal	2.00 a	1.86 a	2.04 a
	OS	28.84 a	31.62 a	26.92 a
	VD	15.85 a	19.50 a	16.98 a
	ED	0.25 a	0.24 a	0.17 a
	JU	17.42 a	17.38 a	17.38 a
	PobRur	5.62 a	4.27 a	3.02 a

Different letters in the same raw indicate significant differences ($p < = 0.05$) valid only in the transformed scale. Code for variables is as in Table 1, means in original scale. The adjusted p-values for rejection in a 9 or 7 multivariate vectors, when covariance structure is taken into account, results in a conservative test in order to preserve Type I error rate.

2.3.2. The association between the original variables and the Concordance Classes

Canonical discriminant analysis based on the 25 original variables and the counties of the four Concordance Classes provides discrimination rules to allow the allocation of the rest of the counties. A two dimensional discriminant space between Concordance Classes was used to explore relationships with original variables, and to visualize the rest of the counties.

3. Results

3.1. Data sets within the Land-use Groups

The Land-use Groups show significant differences with regards to all the variables in the physical support data set, excepting the percentage of land covered by plains (PLA), which even though it is much lower in the Agricultural Group, the difference is not statistically significant. Among the variables of the land use data set, only percentage of planted lands (PI), of grasslands (PPN) and farm size (TPEAP) show significant differences among Land-use Groups. The social variables did not show significant differences among groups (Table 2).

Both discriminant canonical variates were significant ($p < 0.0001$). The classification error rate was 2.04 due to one county of the Agriculture-Ranching Group that was classified as Agriculture. The Land-use Group with high ranching production is well differentiated by the first canonical dimension while the second one discriminates among Agricultural and Agricultural-Ranching Land-use Groups (Fig. 2).

Social conditions variables are present in the first canonical variate as a contrast between percentage of illiterate (%Anal) and both population variation in the last 10 years (Var%), and percentage of population 70 years old or older not perceiving retirement pension (JU). Physical support variables show a contrast between soil classes (positive except for Ipc51-70) and percentage of hillocks (LO) and alluvial plains (PAL). From the set of land use variables, the mean size of farms (TPEAP) and the percentage of the county area under natural grasslands (PPN) are positively correlated with the first dimension while the percentage of the county area under forestry (PFOR) is negatively correlated with the first dimension (Fig. 2).

High values of the second canonical dimension (Fig. 2) are associated with low percentage of population variation between 1991 and 2001 (Var%) and with high percentage of slopes (PEN) and soils with low or high capacity index (Ipc31-50, Ipc71-95).

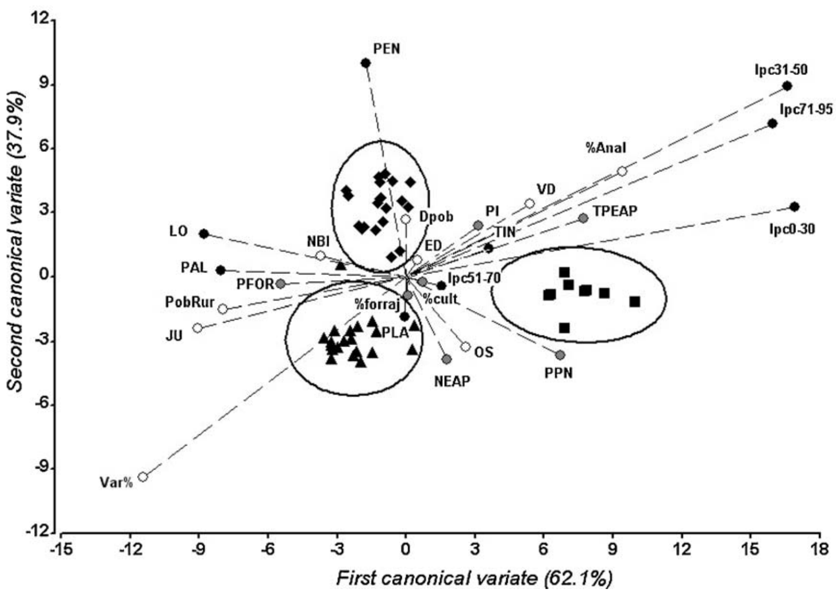

Fig. 2. Land-use Groups in canonical discriminant space. The variables are identified with the codes of Table 1. Black circles: variables of physical support data set; grey circles: land use data set; white circles: social statistics data set. The squares represent Ranching counties, triangles Agriculture-Ranching counties and diamonds Agriculture counties. Ellipses are for predicted 95% confidence.

Table 3
Concordance between data sets.

% Variability in the mean configuration	Data set	Variables that load heavily on the first component	Variables that load heavily on the second component
67.4 (63.9–73.7)	Physical support	% area with Ipc31-50 % area with Ipc51-70 % flooding lands % alluvial plains	% area with Ipc31-50, % area with Ipc71-95
	Land use	Number of farms	Farm size % planted lands % natural grasslands
65.3 (60.3–76.4)	Land use	Number of farms, farm size	Number of farms Farm size % forestry
	Social	Population density, % illiterate	% unsatisfied basic needs % houses with deficiency % population variation

The variability represented in the mean configuration is expressed as percentage of the total variability. Numbers between brackets indicate the bootstrap confidence interval with 0.90 probability.

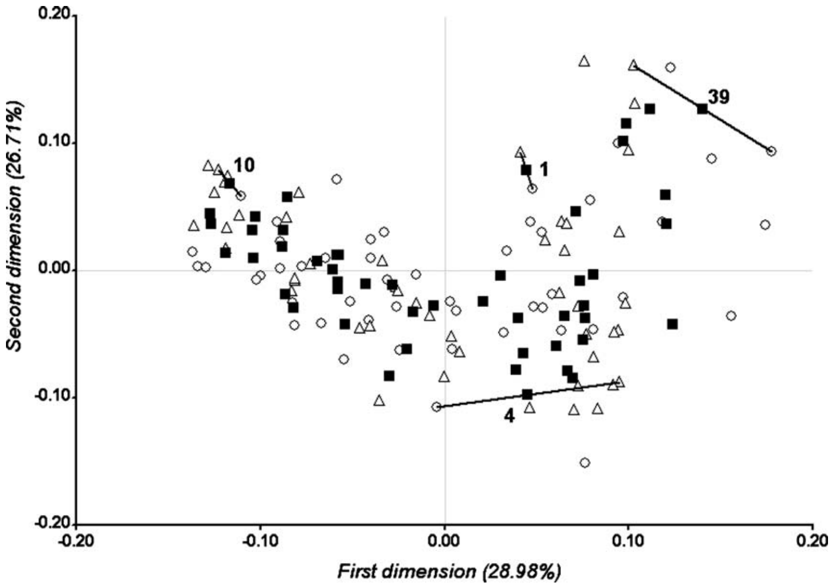

Fig. 3. Consensus plot between physical support and land use. Each county is marked by a filled square in the mean configuration, an open triangle in the physical support hyperspace, and an open circle in the land use hyperspace. The numbers correspond to the four counties exemplified in the text.

3.2. Stability of the mutual interactions between data sets

The concordance between physical support and land use, which is of 67.4% (Table 3), is summarized by the first two dimensions, which capture 55.7% of the total variability. The relative position of the 49 counties in the joint space can be seen in Fig. 3, together with the configuration of each data set. Four counties are shown to exemplify the graph, two of them with high (counties 1 and 10) and two with low (4 and 39) concordance values. The joint configuration (filled squares) resembles those of the land use (open circles) and the physical (open triangles) data sets.

The concordance between the land use and social data sets, which amounts to 65.3% (Table 3), is summarized in the two first dimensions accounting for almost all (62.5%) the variability. The relative position of the 49 counties is shown in Fig. 4, together with the configuration of each data set. The same four counties of Fig. 3 are marked, one of them shows low concordance (county 39), two show medium concordance (counties 1 and 10), and one has very high concordance (county 4). A relatively high level of global concordance is observed, since the joint configuration (filled squares) matches those of the land use (open circles), and social (open diamonds) data sets.

The concordance between land use and social data sets does not differ significantly from the concordance between physical support and land use (Table 3, see bootstrap confidence intervals).

Each county was classified twice using the bi-dimensional index with the pairwise concordance and assigned to one of the nine clusters based on its concordance degree (Fig. 5). The four Concordance Classes characterized by high (H) and low (L) concordance between the two pairs of variable sets comprise the 29 counties identified in Fig. 5.

3.3. Description of the Concordance Classes

When the concordance is studied using physical and land use data, significant differences between concordance clusters are found in the proportion of medium and high land productivity capacity. Significant differences are also found in farm size when the concordance is studied with land use and social data sets. The lack of significant differences among mean values is to be expected, since the concordance was obtained through GPA, in which mean differences are adjusted by translation.

Concordance Classes obtained by cross classification are based on differences among variable interactions which may act in

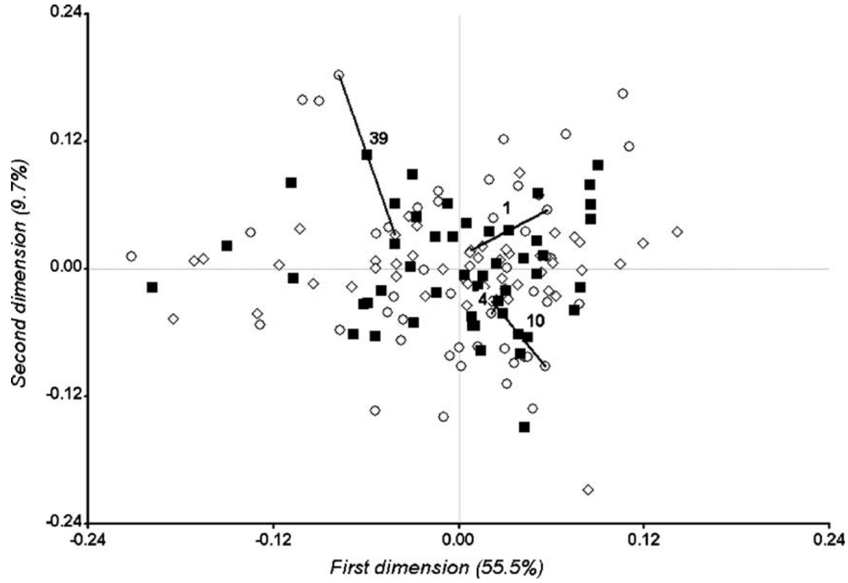

Fig. 4. Consensus plot between land use and social statistics. Each county is marked by a filled square in the mean configuration, an open diamond in the social hyperspace, and an open circle in the land use hyperspace. The numbers correspond to the same four counties exemplified in Fig. 3.

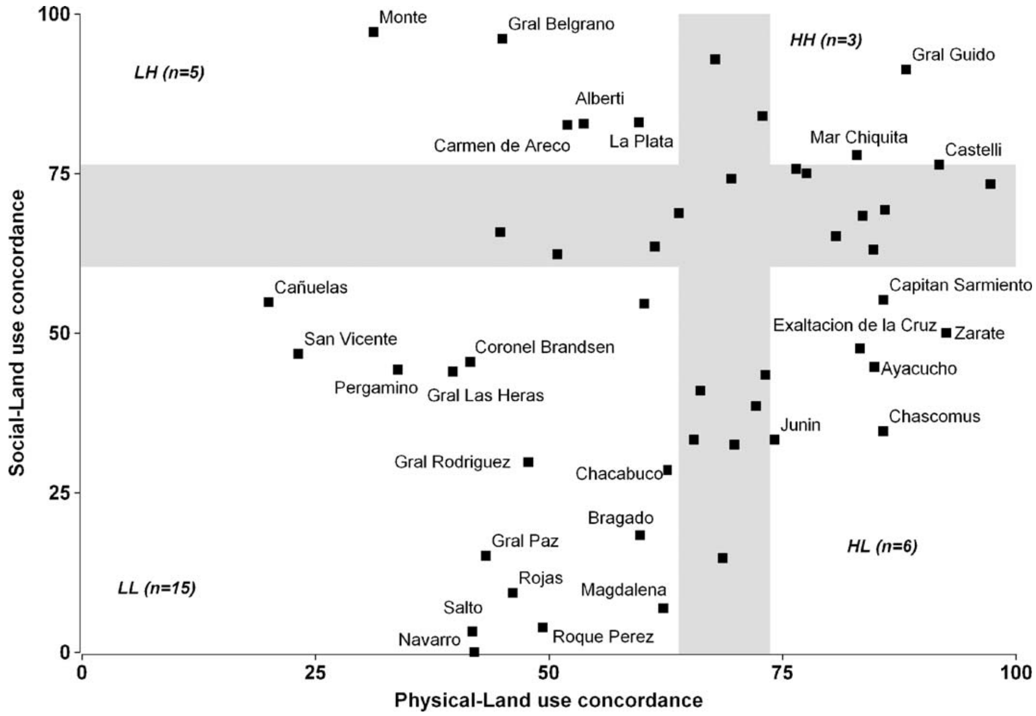

Fig. 5. Confidence intervals for joint configuration variability. Confidence intervals are shown as shadow bands. Each county is plotted using its two concordance values: physical and land used variable sets (*x* axis) versus social and land used sets (*y* axis). Those belonging to the four Concordance Classes are identified with their names. In brackets the total number of counties in each Class.

opposite directions. A projection of the mean vector of each Concordance Class into the first two canonical dimensions may help organize the information in such a way that it is useful for environmental management (Fig. 6). The variables that segregate the Concordance Classes may be identified in the first canonical dimension, which synthesizes 92.6% of the variation, and those whose contributions are near zero may be discarded. The variables with 'raw coefficients' within the −0.25 to 0.25 interval, were discarded on the basis of the slope change in a scatter plot. Raw coefficients are expressed in the transformed scale for percentage variables.

The contribution of the physical support variables may be represented by a contrast between flooding lands (TIN) and alluvial plains (PAL) together with medium productivity capacity lands (Ipc51-70) and a mayor contribution of hillocks (LO) (Fig. 6a):

$$-0.30 * \mathrm{TIN} + 0.30 * \mathrm{PAL} + 0.32 * \mathrm{Ipc51\text{-}70} + 0.85 * \mathrm{LO}$$

Within the land use variable set, mean size of farms (TPEAP), percentage of forage lands (%forraj) and of natural grasslands (PPN)

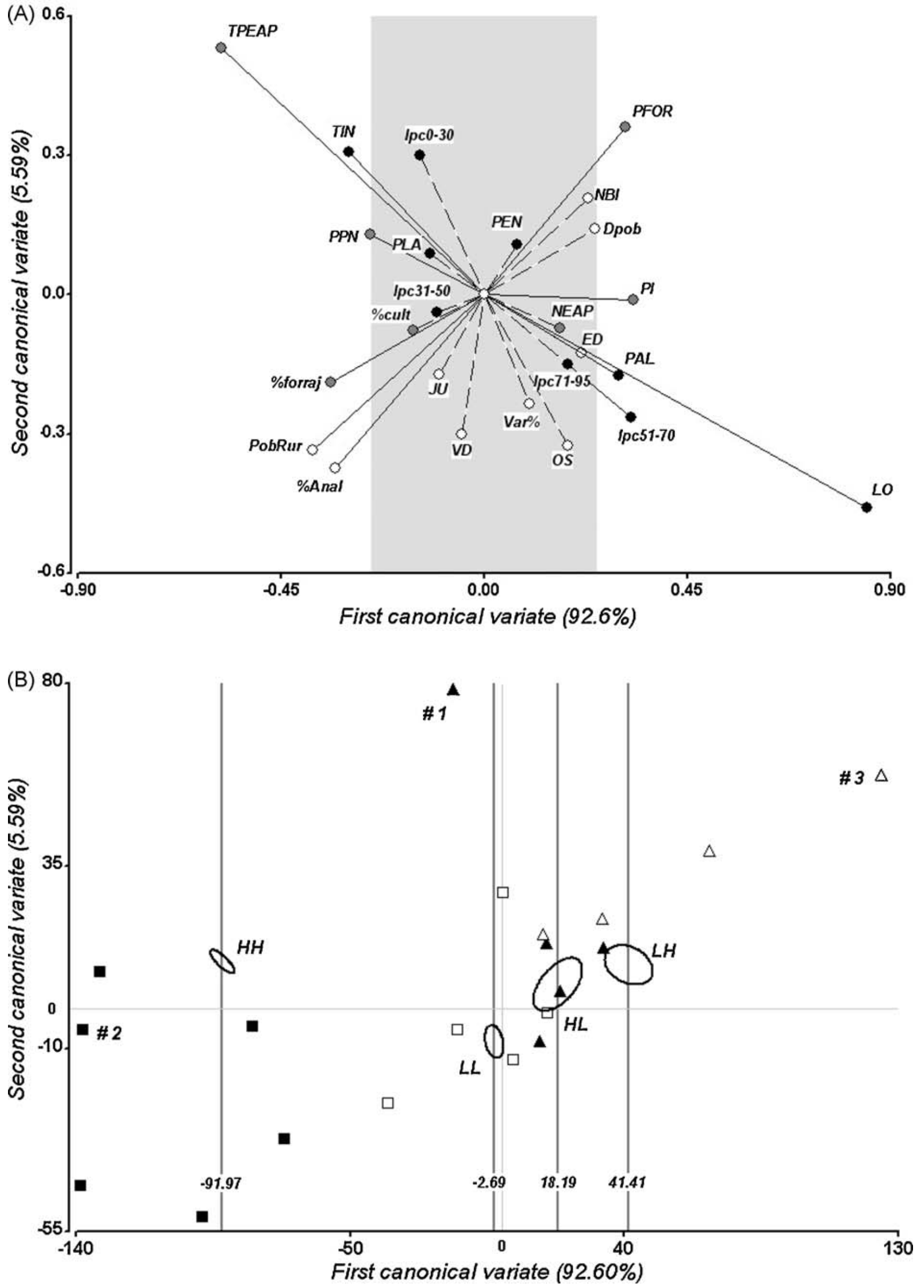

Fig. 6. Original variables and Concordance Classes in the canonical discriminant space. (a) Raw coefficient for original variables. Black circles: variables of physical support data set; grey circles: land use data set; white circles: social statistics data set. The variables are identified with the codes of Table 1. (b) Mean values for each Concordance Class are shown with vertical lines at the first canonical variate. The 99% confidence ellipses of prediction are shown. Counties not used for discrimination are assigned based on discriminant function: filled squares for HH, open squares for LL, filled triangles for HL and open triangles for LH. The numbers correspond to the three counties exemplified in the text.

contrasts with percentage of planted lands (PI) and percentage of forest plantations (PFOR) (Fig. 6a):

$$-0.58 * \text{TPEAP} - 0.34 * \%\text{forraj} - 0.25 * \text{PPN} + 0.33 * \text{PI} + 0.31 * \text{PFOR}$$

Among the social data set, both the percentage of rural population (PobRur) and the percentage of illiterate (%Anal) contribute negatively (Fig. 6a):

$$-0.38 * \text{PobRur} - 0.33 * \%\text{Anal}$$

The three counties clustered in the HH Concordance Class, in the negative end of the first dimension (Fig. 6b), are located in the Flooding Pampa, on plains lacking hillocks, with soils of very low productive capacity. Farm mean size is the largest in the study area, and prevailing land uses are natural grasslands and forage crops. In these counties, illiterate population and rural population are high (Table 4). It is worth mentioning that the three counties in this Concordance Class differ from those in the mH and Hm groups for their economic diversification, since cattle raising is performed on the lowest lands and, crops and forages are grown in patches of agricultural lands. This should be taken into account in the management policies for the mH and Hm counties.

The 15 counties in the LL Concordance Class, at the center of Fig. 6b, are associated to values near zero in the first canonical dimension, as a result of variables with opposing signs in each contrast, that cancel each other. Their characteristics are highly

Table 4
Mean values by Concordance Class and raw coefficient for first canonical dimension.

		Concordance Class (number of counties)					Raw coefficient, dim1
		HH (3)	HL (6)	LH (5)	LL (15)	General mean (29)	
Geophysical support	Ipc0-30	45.66	23.39	32.31	20.37	25.67	−0.14
	Ipc31-50	2.4	0.69	0.84	1.36	1.24	−0.10
	Ipc51-70	0.3	1.3	25.65	13.75	11.83	0.32
	Ipc71-95	0.1	4.42	0.84	2.29	2.25	0.18
	TIN	51.24	12.54	1.77	1.81	9.14	−0.30
	LO	0	31.57	46.72	44.62	37.67	0.85
	PEN	0.15	3.5	0.06	0.15	0.83	0.07
	PLA	10.67	0.9	5.45	2.35	3.45	−0.12
	PAL	0.19	0.37	8.86	3.04	3.20	0.30
Land use	NEAP	286	468	453	451	438	0.17
	PI	13.75	39.76	33.83	28.13	30.03	0.33
	PPN	72.39	32.31	39.76	37.1	40.22	−0.25
	TPEAP	769	337	382	283	362	−0.58
	PFOR	0.14	0.67	0.55	0.14	0.32	0.31
	%cult	14.74	6.41	4.22	9.28	8.38	−0.16
	%forraj	14.08	3.93	0.93	6.41	5.75	−0.34
Social	Var%	1.53	6.26	2.83	8.08	6.12	0.10
	Dpob	4	32	132[a]	35	48	0.24
	NBI	10.18	11.7	9.5	11.97	11.30	0.23
	%Anal	1.99	1.65	1.53	1.99	1.84	−0.33
	OS	25.07	29.46	28.79	31.57	29.98	0.19
	VD	15.8	15.44	15.44	19	17.32	−0.05
	ED	0.05	0.2	0.13	0.16	0.15	0.22
	JU	16.93	15.44	15.8	17.33	16.63	−0.10
	PobRur	12.83	1.65	2.77	9.28	6.95	−0.38

[a] Biased towards the right due to one county. Code for variables as in Table 1.

variable, and the relationships among the data sets are not stabilized. In average, positive values for the physical support contrast are due to the low proportion of flooding lands (TIN) and the dominance of hillocks (LO) with high proportion of medium productive capacity soil (Ipc51-70). The mean farm size (TPEAP) is the lowest of the study area, it lacks forest plantations (PFOR) and it has a higher proportion of forage production (%forraj) than in the HL and LH classes, so their contribution is negative. The social variables also contribute to cancel out the positive values of the physical support data, since the percentages of rural population and of illiterate population are high (Table 4).

The counties within the HL and LH Concordance Classes are on the positive end of the first canonical dimension. In the six counties of the HL class, the physical support variables that contribute negatively are higher than the general mean value, and most of those that contribute positively are lower, resulting in a negative contribution of the physical data set. Within the land use data set, the percentage of forest plantation (PFOR) is higher than the general mean, as well as the percentage of planted lands (PI), resulting in a positive contribution. Also natural grasslands (PPN) and forage production (%forraj) contribute positively because their mean values are below the general mean. The social variables are lower than the general mean, contributing also positively to partially canceling out the negative contribution of the physical support variables (Table 4).

In the five counties belonging to the LH Concordance Class, the contribution of the physical support data set is reversed with respect to the HL class. Within the land use variables the dominance of large farms (TPEAP) contributes negatively while the proportion of forest plantation (PFOR) and the percentage of forage lands (%forraj) contribute positively. Both the illiterate population and the rural population are lower than the general mean resulting in a positive contribution (Table 4).

Those counties classified according to its own fraction of variability represented in the joint configuration (JCVc) as medium (m) in any of the two GPA, can be assigned to a Concordance Class on the basis of their probability of belonging to one of them (Table 5). In Fig. 6b, six counties have been identified as potentially HH, five as potentially HL, four as potentially LH, and five as potentially LL.

A detailed comparison of selected variables for each county with the mean of each Concordance Class permits the diagnostic of individual counties. Three divergent cases (marked with numbers in Fig. 6b), laying within the confidence interval in one pair of data sets (m) may exemplify the application of the results.

County #1, belonging to the Lm class may be classified as HL (Fig. 6b). It has medium productivity capacity soils, with agriculture (mainly perennial fodder and potato production) as its main activity, showing an appropriate use of physical resources in some areas. Social conditions are poor; rural population is low, which may indicate that the county is being urbanized. Population with unsatisfied basic needs is high, as well as illiterate population and population with no social security income. The latter may mean that workers are hired under an informal basis. In this county, management policies should point to improving social conditions, with a better distribution of the county's income. If the present situation persists, it may tend towards the LL Concordance Class, with most of the poor gathered around cities to exploit urban waste as resources to make a living.

Table 5
Distribution of counties in the Concordance Classes.

Concordance from GPA	Predicted Concordance Class				Total
	HH	HL	LH	LL	
H-m	3	2	1	1	7
L-m	1	1	–	1	3
m-H	1	–	1	–	2
m-L	1	2	–	3	6
m-m	–	–	2	–	2
Total	6	5	4	5	20

Figures indicate the number of counties within each class. First letter identifies the concordance between physical support and land use data sets, second letter identifies the concordance between land use and social data sets; low (L), medium (m), high (H).

A different situation is encountered in county #2, belonging also to the Lm class but that may be classified as HH (Fig. 6b). The agricultural productivity capacity of soils is high but there is a low fraction of lands under annual and perennial crops. The social conditions are not as good as they should be but they are not bad. The aim of management polices should be to improve land use assigning the most productive lands to crop production. In the coastal fringe of this county, tourist activities prevail; tourism is highly profitable, and urbanization may be indirectly promoted to the detriment of the rural way of life as well as of food production. The tendency is towards the LL Concordance Class, unless an impulse is given to rural activities and food production.

County #3, classified as mH may be assigned to LH (Fig. 6b). It is occupied by large farms, some of them established during XIX century. Soils are of low production capacity and there is a large proportion of flooding lands. Cattle raising on natural grasslands prevails. Population density is very low, and this may be the cause for the relatively good social conditions. These large farms probably persist due to family tradition. If population grows, the county is doomed to social disaster unless land use is diversified according to land use capacity.

4. Discussion

In this paper we have presented a method to analyze a complex system in terms of data sets from various disciplines, and to identify associations and imbalances between data sets within each county in Buenos Aires Province, Argentina. We propose as ecological indicator the bi-dimensional space formed by two pairwise concordance values.

The county as a unit of analysis is justified by the fact that, in a federal administration system, such as that of Argentina, each administrative unit (county or municipality at the provincial level) is responsible for the management of its territory, people and economy. According to the law, each county is in charge of designing its budget, on the basis of the available economic resources and the expected expenses. Since a few years ago, the county budget includes, besides the income from its own municipal taxes, a co-participation from provincial gross income taxes. The former are levied on local people, property and business, and are proportional to incomes and profits. Recently, part of the tax collection at provincial level has been decentralized through the "Program of Tax System Decentralization", as for example, the rural real-estate taxes and the gross income taxes of high income tax payers. The county government is responsible for the administration of the tax income and for distributing the income in public policies that satisfy the people's needs (infrastructure, public health, education, social development). Each county is in charge of designing its own Master Plan for resources and land assignment, as well as enforcing the plan. The fate of the counties and its people depends on the national and local policies; a lack of action or an incorrect decision may worsen the social and economic situation.

The aggregation of the counties in Concordance Classes guides the recommendations to select and prioritize the actions to promote a public policy for sustainable management. Those counties in the HH class appear as the most sustainable, and under the least risk. They could be maintained as in the present scenario. In the counties classified as HL, the causes of the present situation should be analyzed and the emphasis should be centered in improving resources distribution among people. The counties in the LH class show high vulnerability; even though social conditions are good, the low concordance between physical support and land use may be a symptom of failure to adjust economic production to environmental conditions, and this could lead to deterioration of the social conditions. The counties classed as LL appear as the most vulnerable, since they show imbalances between both pairs of data sets; in the first place, they should improve land use planning. Of the 15 counties in the LL class, eight are beginning to form a new urban fringe around the Metropolitan Area; two are next to expanding tourist cities in the coastal zone, and three of them, are suffering exurban expansion. The 15 counties are dissected by highways and roads prone to be converted to highways in a near future.

The chosen data sets do not include all factors. Cultural indicators would probably help explain some of the observations. The fact that some counties maintain their agricultural land uses and are not being converted to urban uses even though they are surrounded by counties with a dynamic exurban development, may be due to cultural peculiarities linked to the rural tradition of their people, or a conscious selection of a rural way of life.

Other variables, such as land tenure, and income from other economic activities (industry, tourism, harbor services for commercial and sports activities, and so on), should also be included; even though they are not the main source of income, they are important in a few counties.

The quantitative assessment of the quality of representation in the joint configuration using the generalized Procrustes transformation helps in the analysis of the differences among counties. The translation to a common centroid in the Procrustes analysis focuses on the interrelationships among the data sets and discards the differences between the mean values in which variables are expressed (a matter of traditionally studies applying multivariate analysis of variance). The lack of concordance between data sets in each county is an indicator of imbalances. The uncovering of this lack adjustment is an essential tool to guide environmental management.

Acknowledgements

This research was financed by the Agencia Nacional de Promoción Científica y Tecnológica de Argentina (PICT 13-8481), and the Consejo Nacional de Investigaciones Científicas y Técnicas (PIP 5921, 2005–2006), Argentina. The authors acknowledge the anonymous reviewers for their valuable suggestions.

References

Bramardi, S.J., Bernete, G.P., Azíns, M.J., Carbonell, E.A., 2005. Simultaneous agronomic and molecular characterization of genotypes via the generalized procrustes analysis: an application to cucumber. Crop Science 45, 1603–1609.

Brickman, R., Smith, A.J., 1973. Land evaluation for rural purposes: a summary of an expert consultation, Netherlands. Publication N° 17, International Institute for Land Reclamation and Improvement, Wageningen.

Caeiro, S., Goovaerts, P., Painho, M., Costa, M.H., 2003. Delineation of estuarine management areas using multivariate geostatistics: the case of Sado Estuary. Environmental Science and Technology 37, 4052–4059.

Dijksterhuis, G.B., 1994. Procrustes analysis in studying sensory-instrumental relations. Food Quality and Preference 5, 115–120.

Di Rienzo, J.A., Balzarini, M., Gonzalez, L., Casanoves, F., Tablada, M., Robledo, W., 1998. InfoStat Version 2007p. InfoStat Group, FCA, Universidad Nacional de Córdoba. Available via http://www.infostat.com.ar (accessed 18 February 2009).

Gower, J.C., 1971. A general coefficient of similarity and some of its properties. Biometrics 27, 857–874.

Gower, J.C., 1975. Generalized procrustes analysis. Psychometrika 40, 33–51.

Gower, J.C., Dijksterhuis, G.B., 2004. Procrustes Problems. Oxford Statistical Science Series 30. Oxford University Press, Oxford, 254 pp.

Husson, F., Lê, S., Josse, J., Mazet, J., 2007. FactoMineR Package, Version 1.05, running on R-2.5.0. Available via http://factominer.free.fr and http://www.agrocampus-rennes.fr/math (accessed 25 May 2007).

INDEC, 2001. 2001 National Census of Population, Households and Housing. Buenos Aires Province, 2nd Series. Ministry of Economy, of Economic Policy Bureau, National Institute of Statistics and Censuses, Buenos Aires.

INDEC, 2006. 2002 National Agricultural Census. Ministry of Economy, of Economic Policy Bureau, National Institute of Statistics and Censuses, Buenos Aires. Available via www.indec.mecon.ar (accessed 24–27 June 2006).

Jackson, D.A., Harvey, H.H., 1993. Fish and benthic invertebrates: community concordance and community-environment relationships. Canadian Journal of Fisheries and Aquatic Sciences 50, 2641–2651.

Jansen, L.J.M., 2006. Harmonization of land use class sets to facilitate compatibility and comparability of data across space and time. Journal of Land Use Science 1, 127–156.

Jenerette, G.D., Lee, J., Waller, D.W., Carlson, R.E., 2002. Multivariate analysis of the ecoregion delineation for aquatic systems. Environmental Management 29, 67–75.

Krzanowski, W.J., 2000. Principles of Multivariate Analysis: A User's Perspective. Oxford Statistical Science Series 23. Oxford University Press, Oxford, 612 pp.

Maccarini, G.D., Baleani, O. (Eds.), 1995. Atlas de Suelos de la República Argentina, Base cartográfica digital. ArgenINTA Foundation; Natural Resources Research Center, National Institute of Agricultural Technology, and Aeroterra, Buenos Aires.

Matteucci, S.D., 2006. La sustentabilidad del sistema humano-natural en el norte y noreste de la Provincia de Buenos Aires. In: Matteucci, S.D., Morello, J., Buzai, G.D., Baxendale, C., Silva, M., Mendoza, N., Pengue, W., Rodríguez, A. (Eds.), Crecimiento Urbano y Sus Consecuencias Sobre el Entorno Rural. Orientación Gráfica Editora, Buenos Aires, pp. 83–122.

Matteucci, S.D., Morello, J., 2006. Efectos ecológicos de los emprendimientos urbanísticos privados en la Provincia de Buenos Aires, Argentina. In: Matteucci, S.D., Morello, J., Buzai, G.D., Baxendale, C., Silva, M., Mendoza, N., Pengue, W., Rodríguez, A. (Eds.), Crecimiento Urbano y Sus Consecuencias Sobre el Entorno Rural. Orientación Gráfica Editora, Buenos Aires, pp. 197–221.

Morello, J., Matteucci, S.D., 1997. El modelo agrícola del Núcleo Maicero como sistema complejo. In: Morello, J., Solbrig, O.T. (Eds.), Argentina Granero del Mundo: Hasta Cuándo? Orientación Gráfica Editora, Buenos Aires, pp. 201–231.

Morello, J., Matteucci, S.D., Buzai, G.D., 2001. Urban sprawl and landscape perturbation in high quality farmland ecosystems, the case of Buenos Aires Metropolitan region. In: Solbrig, O.T., Paarlberg, R., Di Castri, F. (Eds.), Globalization Under Rural Environment, David Rockefeller Center Series on Latin American Studies. Harvard University Press, Cambridge, pp. 447–482.

Rao, C.R., 1973. Linear Statistical Inference and Its Applications. John Wiley & Sons Inc., New York, 658 pp.

Westfall, P.H., Wolfinger, R.D., 2000. Procedures and PROC MULTTEST. Observations (OBSwww23). SAS Institute, 23 pp. Available via www.sas.com/obs (accessed 23 November 2005).