Various approaches for predicting land cover in mountain areas

Nathalie Villa [*†] Martin Paegelow [‡]
Maria T. Camacho Olmedo [§] Laurence Cornez [¶]
Frédéric Ferraty,[†] Louis Ferré,[†] Pascal Sarda,[†]

March 14, 2006

Abstract

Using former maps, geographers intend to study the evolution of the land cover in order to have a prospective approach on the future landscape; these simulations are usually done through the GIS (Geographic Information System). We propose here to confront this classical geographical approach with statistical approaches: a linear parametric model (polychotomous regression modelling) and a nonparametric one (multilayer perceptron). These methodologies have been tested on two real areas on which the land cover is known at various dates; this allows us to compare the different models in order to underline the benefit of each of them.

Key words: polychotomous regression modelling, multilayer perceptron, classification, prediction, comparison

1 Predicting land cover

From the sketch maps made by geographers or from the analysis of satellite images or aerial photographs, we can build land cover maps for a given country which can be rather precise: the studied area is then cut into several squared pixels whose sides are about 20 meters long and whose land cover is known on various dates. The type of land cover can be chosen from a pre-determined list: coniferous forests, deciduous forests, scrubs, ...

Here, we are not interested in making such maps (for satellite data analysis, see Cardot *et al.* (1993)). Our purpose is to analyse the land cover dynamics in order to estimate its future evolution; on a geographical point of view,

[*]Corresponding author's e-mail: villa@univ-tlse2.fr
[†]GRIMM, Equipe d'accueil 3686, Université Toulouse Le Mirail, France
[‡]GEODE UMR 5602 CNRS, Université Toulouse Le Mirail, France
[§]Instituto de desarrollo regional, Universidad de Granada, Spain
[¶]ONERA, Toulouse, France

prospective simulations have a great interest to help the local administrations to develop these mountain areas. The idea is then to compare different approaches in order to confront their ability to be generalized to various mountain areas.

For a given pixel, determined by its spatial coordinates, latitude (i) and longitude (j), the value of the land cover on date t, $c_{i,j}(t)$, is a categorical random variable depending on several variables:

- the land cover of this pixel on previous dates: $c_{i,j}(t-1), \ldots, c_{i,j}(t-T)$ (*time serie of length* T);
- the land covers of the neighbouring pixels on previous dates: $V_{i,j}(t-1), \ldots, V_{i,j}(t-T)$, where $V_{i,j}(t-\tau)$ is a set of values of land cover on date $t-\tau$ for the pixels in a neighbourhood of the pixel (i,j) (*vectorial time serie*);
- some environmental variables: for example, the elevation, the aspect, the proximity of roads and villages, ...: $Y^1_{i,j}, \ldots, Y^p_{i,j}$.

Finally, we face a problem of classification in which the predictors are both qualitative and quantitative and are also highly dependent (spatial time process). To solve this question, we propose and compare two statistical approaches with the empirical geographic method (namely the GIS, Geographic Information System). The first of these methods is a generalized linear model in which we estimate the parameters of the model by maximizing a log-likelihood type criterion. The second one uses a supervised multilayer perceptron. By confronting these various approaches, we expect to give ideas in order to improve the GIS approach.

A comparison of these two approaches has been made on two little areas: the "Garrotxes" ("Pyrénées Orientales", south west of France) and the "Alta Alpujarra Granaderia" (Sierra Nevada, Spain) where several surveys of the land cover have been made at various dates. We confront the various scenarii constructed with the real maps.

In the following, we describe the data more precisely (section 2) and present the two approaches (section 3). Then we apply these methodologies on these data sets (section 4) and finally, we compare the results obtained by analyzing the advantages and the limits of the models (section 5).

2 Description of the data sets

The areas under study stand in the moutains "Pyrénées" for the Garrotxes and Sierra Nevada "Alta Alpujarra". A big drift from the land has led to the desertion of the land under cultivation and the recovery of the fields by scrubs and forests. Human action on the land is then very low. Climatic differences are responsable for a very slow land cover evolution in south of Spain while the Garrotxes area has a faster evolution, at least on a geographical point of view, and it is then more difficult to predict the land cover dynamics.

The French area was divided into about 241 000 pixels whereas the Spanish one is much bigger with about 560 000 pixels. For each pixel, we know:

- a categorical variable which is the land cover at different dates: 3 dates (1980, 1990 and 2000) are avalaible for the Garrotxes and 4 dates (1957, 1974, 1987 and 2001) for the Alta Alpujarra. As the land cover evolution is very slow in the Sierra Nevada, these dates can be considered as equidistant, according to geographers opinion. This categorical variable was taken from a list of several choices (8 for the Garrotxes and 9 for the Alta Alpujarra) which are of classical use in geography. These data have been used to make maps of the studied area (see Figure 1);

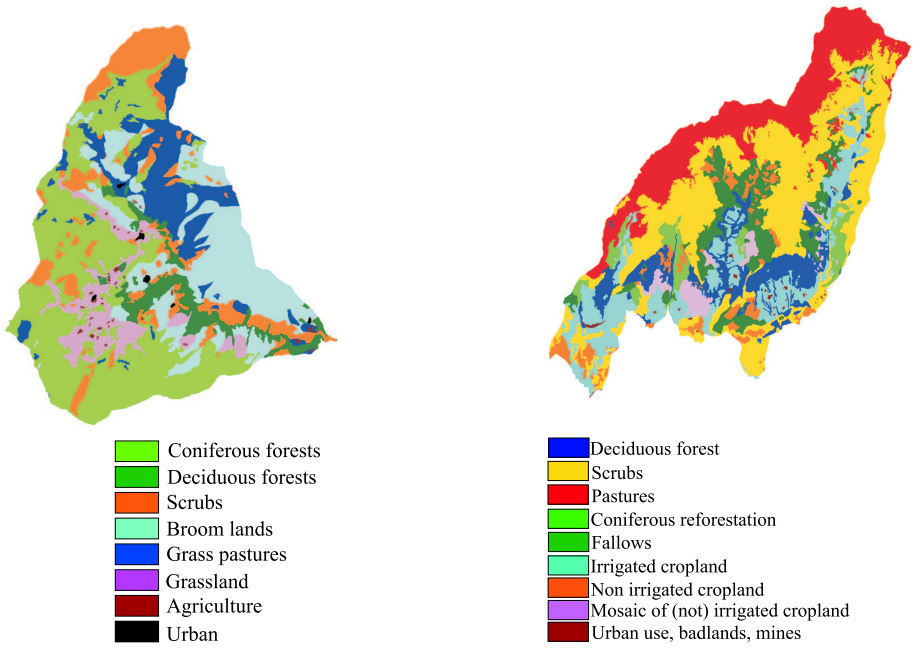

Figure 1: Land cover for the Garrotxes (1980 - left) and for the Alta Alpujarra (1957 - right)

- several environmental variables; some of them are of numeric type (the elevation, the slope, the aspect, the distance of roads and villages,...) and others are of categorical type (forest and pasture management: public or not ? ground geological type, ...). The environmental variables are not the same for the Garrotxes and the Alta Alpujarra but none of them has changed during the studied period (see Figure 2 for examples of environmental variables).

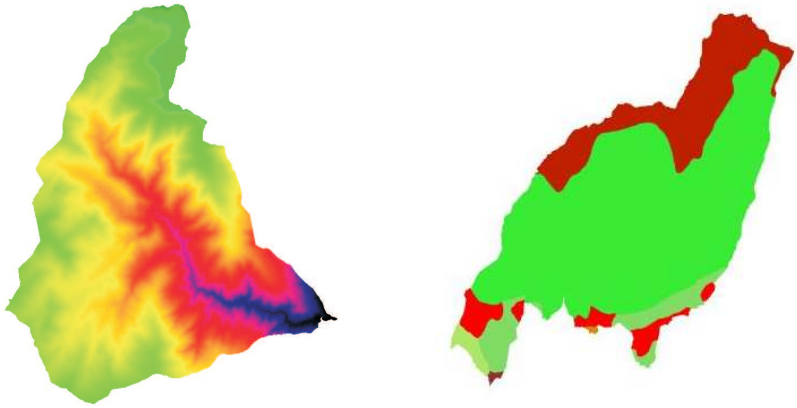

Figure 2: Examples of a numerical variable (elevation for the Garrotxes - left) and a categorical one (ground geological type for the Alta Alpujarra - right)

3 Presentation of the two approaches

Geographers usually estimate the land cover evolution by an empirical method which allows to introduce some expert knowledge. The so-called GIS (Geographic Information System) approach is time expensive and necessitates precise knowledge on the geographic constraints of the area under study. Roughly speaking, the method consists in two steps: at first one computes time transition probabilities for each land cover type whereas, in a second step, one uses spatial constraints (introduced by an expert) for smoothing the maps obtained at the first step (see Paegelow *et al.* (2004) or Paegelow & Camacho Olmedo (2005) for further details on GIS for these data sets). In order to propose automatic alternatives to the GIS, which can take in the same model the spatio-temporal nature of the problem, two approaches have been developed to estimate the evolution of the land cover: the first one, polychotomous regression modelling, is a generalized linear approach based on the maximum log-likelihood method. The second one, multilayer perceptron, is a popular method which has recently proved its great efficiency to solve various types of problems.

The idea is to confront a parametric linear model with a non parametric one: the polychotomous regression modelling is faster to train than multilayer perceptrons, especially in high dimensional spaces and does not suffer from the existence of local minima. On the contrary, multilayer perceptrons can provide nonlinear solutions and are then more flexible than the linear modelling; moreover, both methods are easy to implement even for non statisticians through the pre-made softwares.

3.1 The model

Let us now describe the statistical setting more formally. We note $X_{i,j}(t)$ the vector of variables that explain the value of the land cover for a given pixel (i, j) on date t. We suppose that the time dependance is of order 1; then, $X_{i,j}(t)$ contains:

- *for the time series:* the value of the land cover for the pixel (i, j) at the previous time $t - 1$;
- *for the spatial aspect:* the frequency of each type of land cover in the neighbourhood of pixel (i, j) on the previous date. Then we have to choose the shape and the size of the neighbourhood. For the shape, we have many choices: the simpler one is a square neighbourhood or a star-shaped neighbourhood around the pixel (i, j); the most sophisticated can use the slope to better take into account the morphological influences of the land. For the size of the neighbourhood, we have to find at which distance a pixel can influence the land use of pixel (i, j). Moreover, for the multilayer perceptrons, in order to respect the spatial aspect of the problem, we weight the influence of a pixel by a decreasing function of its distance to the pixel (i, j) (see Figure 3).

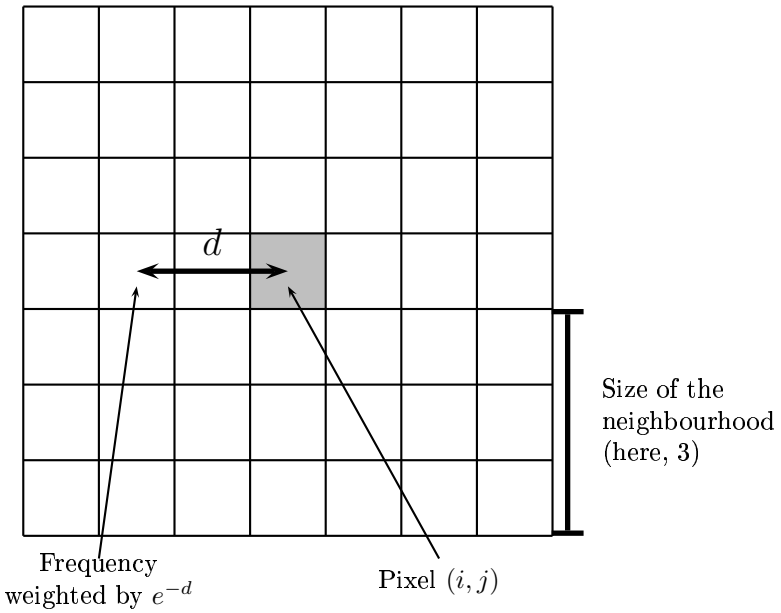

Figure 3: An example of neighbourhood

- environmental variables (slope, elevation, ...).

Let us repeat that $c_{i,j}(t)$ is the land cover for a given pixel on date t. We note $\mathcal{C}_1, \ldots, \mathcal{C}_K$ the different types of land cover. Then, for every $k = 1, \ldots, K$, we try to estimate the probability $P(c_{i,j}(t) = \mathcal{C}_k | X_{i,j}(t))$ that the pixel (i, j) has a land cover equal to $\mathcal{C}_k$ given the vector $X_{i,j}(t)$; thus, the model is of the following form :

$$P(c_{i,j}(t) = \mathcal{C}_k | X_{i,j}(t)) = f_k(X_{i,j}(t)). \tag{1}$$

Once this probability is estimated, the main idea consists in predicting the type of land cover, $c_{i,j}(t)$, by the quantity:

$$\text{argmax}_{k=1,\ldots,K} P(c_{i,j}(t) = \mathcal{C}_k | X_{i,j}(t)).$$

In both approaches, we estimate f_k thanks to a training sample. To that end, we have collected the values of the predictors and of the land cover for many pixels on various dates (see next section for more details); the observations are denoted by $(X^{(1)}, c^{(1)}), \ldots, (X^{(N)}, c^{(N)})$.

The time and spacial aspects are taking into account together both by the polychotomous regression modelling and by the multilayer perceptron. This is not the case for the usual GIS approach which first estimates the land cover probability by modelling a time serie and then introduce a spacial smoothing with environmental constraints. Finally, the difference between the two statistical approaches is in the way they try to modelize and then estimate the functions f_k.

3.2 Polychotomous Regression Modelling

When we wish to predict a categorical response given a random vector, a useful model is the *multiple logistic regression* (or *polychotomous regression*) model (Hosmer & Lemeshow (1989)). A smooth version of this kind of method can be found in Kooperberg *et al.* (1997). These statistical techniques have been applied to several situations. Their good behaviour both on theoretical and practical grounds have been emphasized. In our case, where the predictors are both categorical and scalar, we may think of generalizing this method. We describ the derived model below.

Let us note, for $k = 1, \ldots, K$

$$\theta\left(\mathcal{C}_k | X_{i,j}(t)\right) = \log \frac{P\left(c_{i,j}(t) = \mathcal{C}_k | X_{i,j}(t)\right)}{P\left(c_{i,j}(t) = \mathcal{C}_K | X_{i,j}(t)\right)}.$$

Then, we get the following expression

$$P\left(c_{i,j}(t) = \mathcal{C}_k | X_{i,j}(t)\right) = \frac{\exp \theta\left(\mathcal{C}_k | X_{i,j}(t)\right)}{\sum_{k'=1}^{K} \exp \theta\left(\mathcal{C}_{k'} | X_{i,j}(t)\right)}. \tag{2}$$

Now, to estimate these conditional probabilities, we use the parametric approach to the polychotomous regression problem, that is the linear model

$$\theta\left(\mathcal{C}_k | X_{i,j}(t)\right) = \alpha_k + \sum_{c \in V_{i,j}(t-1)} \sum_{l=1}^{K} \beta_{kl} \mathbb{1}_{[c=\mathcal{C}_l]} + \sum_{r=1}^{p} \gamma_{kr} Y_{i,j}^{r}, \tag{3}$$

where we recall that $V_{i,j}(t-1)$ are the values of the land cover in the neighbourhood of the pixel (i,j) on the previous date $t-1$ and $(Y^r_{i,j})_r$ are the values of the environment variables. Let us call $\delta = (\alpha_1, \ldots, \alpha_{K-1}, \beta_{1,1}, \ldots, \beta_{1,K}, \beta_{2,1}, \ldots, \beta_{2,K}, \ldots, \beta_{K-1,1}, \ldots, \beta_{K-1,K}, \gamma_{1,1}, \ldots, \gamma_{1,K}, \ldots, \gamma_{K-1,1}, \ldots, \gamma_{K-1,p})$, the parameters of the model to be estimated. We have to notice that since $\theta(\mathcal{C}_K|X_{i,j}(t)) = 0$, we have $\alpha_K = 0$, $\beta_{K,l} = 0$ for all $l = 1, \ldots, K$, and $\gamma_{K,r} = 0$ for all $r = 1, \ldots, p$. We now have to estimate the vector of parameters δ. For that end, we use a penalized likelihood estimator which is performed on the training sample. Let us write the penalized log-likelihood function for model (3). It is given by

$$l_\varepsilon(\delta) \;=\; l(\delta) - \varepsilon \sum_{n=1}^{N} \sum_{k=1}^{K} u_{nk}^2, \tag{4}$$

where the log-likelihood function is

$$l(\delta) = \log \left(\prod_{n=1}^{N} P_\delta \left(c^{(n)} | X^{(n)} \right) \right). \tag{5}$$

In this expression, $P_\delta(c^{(n)}|X^{(n)})$ is the value of the probability given by (2) and (3) for the observations $(X^{(n)}, c^{(n)})$ and the value δ of the parameter.

In expression (5), ε is a penalization parameter and, for $k = 1, \ldots, K$, $u_{nk} = \theta_\delta(\mathcal{C}_k|X^{(n)}) - \dfrac{1}{K} \displaystyle\sum_{k'=1}^{K} \theta_\delta(\mathcal{C}_{k'}|X^{(n)})$. Our penalized likelihood estimator $\widehat{\delta}_\varepsilon$ satisfies:

$$\widehat{\delta}_\varepsilon = \text{argmax}_{\delta \in \mathbb{R}^M} l_\varepsilon(\delta),$$

where $M = K^2 + (K-1) * p - 1$ denotes the number of parameters to be estimated.

Traditionally, statisticians maximize the log-likelihood function to compute the estimators. But, as pointed out by Kooperberg *et al.* (1997), the introduction of a small penalty term allows numerical stability and guarantees the existence of a finite maximum. Moreover, for a reasonably small value of ε, the penalty term would not affect the value of the estimators that we could obtain without the penalty term. Numerical maximization of the penalized log-likelihood function is achieved by a Newton-Raphson algorithm.

3.3 Multilayer perceptron

Neural networks have a great adaptability to any statistical problems and especially to overcome the difficulties of non linear problems even if the predictors are highly correlated; thus it is not surprising to find them used in the chronological series prediction (Bishop (1995), Lai & Wong (2001) and Parlitz & Merkwirth (2000)). The main interest of neural networks is their ability to approximate any function with the desired precision (universal approximation): see, for instance, Hornik (1991).

Here we propose to estimate, in model (1), the function f_k in the form of a multilayer perceptron with one hidden layer (see Figure 4), ψ, which is a function from $\mathbb{R}^q$ to $\mathbb{R}$ that can be written: for all x in $\mathbb{R}^q$

$$\psi_w(x) = \sum_{i=1}^{q_2} w_i^{(2)} g\left(\langle x, w_i^{(1)}\rangle + w_{i,0}^{(1)}\right)$$

where q_2 in $\mathbb{N}$ is the number of neurons on the hidden layer, $(w_i^{(1)})_{i=1,\dots,q_2}$ (respectively $(w_i^{(2)})_{i=1,\dots,q_2}$, $(w_{i,0}^{(1)})_{i=1,\dots,q_2}$) are in $\mathbb{R}^q$ (resp. $\mathbb{R}$) and are called weights of the first layer (resp. weights of the second layer, bias) and where g, the activation function, is a sigmoïd; for example, $g(x) = \frac{1}{1+e^{-x}}$.

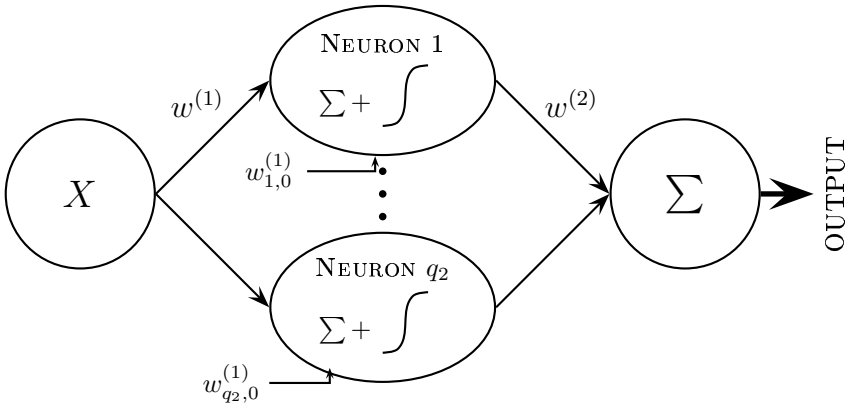

Figure 4: Multilayer perceptron with one hidden layer

Then, the output of the multilayer perceptron is a smooth function (here it is indefinitly continuous and derivable) of its input. This property ensures that the neural network takes into account the spatial aspect of the data set, since two neighbouring pixels have "close" values for their predictor variables.

To determine the optimal value for weights $w = ((w_i^{(1)})_i, (w_i^{(2)})_i, (w_{i,0}^{(1)})_i)$, we usually minimize the quadratic error on the training sample: for all $k = 1, \dots, K$, we choose

$$w_{opt}^k = \text{argmin}_{w \in \mathbb{R}^{q_2(q+2)}} \sum_{n=1}^{N} \left[c_k^{(n)} - \psi_w^k(X^{(n)})\right]^2, \tag{6}$$

where $c^{(n)}$ and the categorical data in $X^{(n)}$ are written on a disjonctive form. This can be performed by classical numerical methods of the first or the second order (such as gradient descent or conjugate gradients, ...) but faces local minima problems. We explain in section 4 how we try to overcome this difficulty. Finally, White (1989) gives many results that ensure the convergence of the optimal empirical parameters to the optimal theoretical parameters.

4 Practical application on the data sets

In order to compare the two approachs, we apply the same methodology: we first determine the optimal parameters for each approach (training step, see below) and then, we use the first maps to predict the last one and compare the errors to real map (comparison step, see section 5).

As usual in statistical methods, there are two stages in the training step: the *estimation step* and the *validation step*.

- The *estimation step* consists in estimating the parameters of the models (either for the polychotomous regression or the neural network);
- The *validation step* allows us to choose, for both methodologies, the best neighbourhood, for polychotomous regression, the penalization parameter and, for neural network, the number of neurons on the hidden layer. Concerning the neighbourhood, we only consider square shapes so choosing a neighbourhood leads us to determine its size.

For the Sierra Nevada, we saw that large areas are constant, thus we only used the pixels for which one neighbour, at least, has a different land cover. These pixels will be called "frontier pixels"; the others are considered as constant (see Figure 5). For the generalized linear model, we used the whole frontier pixels of the 1957/1974 maps for the estimation set and the whole 1974/1987 maps for the validation set. We then construct the estimated 2001 map from the 1987 one. For the multilayer perceptron, we need to reduce the training set size in order not to have huge computational times when minimizing the loss function. Then, estimation and validation data sets are chosen randomly in the frontier pixels of the 1957/1974 and 1974/1987 maps.

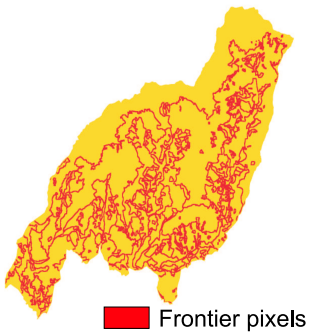

Figure 5: Frontier pixels (order 4) for the 1957 map (Sierra Nevada)

For the Garrotxes data set, due to the fact that we only have 3 maps and much less pixels, we had to use the 1980/1990 maps for the estimation step (only their frontier pixels for the MLP) and the whole 1990/2000 ones for the

validation step. This leads to a biased estimate when constructing the 2000 map from the 1990 map but, as our purpose is to compare two models and not to make significant the error rate, we do not consider this bias as important.

4.1 Polychotomous regression

- The *estimation step* produces the estimated parameter vector $\widehat{\delta}_\varepsilon$ of the parameters δ_ε of model (3) for fixed neighbourhood and penalization parameter ε. This step is repeated for various values concerning both neighbourhood and penalization parameter.

- *Validation step:* Once given an estimated parameter vector $\widehat{\delta}_\varepsilon = (\widehat{\alpha}_1, \ldots, \widehat{\alpha}_{K-1}, \widehat{\beta}_{1,1}, \ldots, \widehat{\beta}_{1,K}, \widehat{\beta}_{2,1}, \ldots, \widehat{\beta}_{2,K}, \ldots, \widehat{\beta}_{K-1,1}, \ldots, \widehat{\beta}_{K-1,K}, \widehat{\gamma}_{1,1}, \ldots, \widehat{\gamma}_{1,p}, \ldots, \widehat{\gamma}_{K-1,1}, \ldots, \widehat{\gamma}_{K-1,p})$, it is easy to estimate, for $k = 1, \ldots, K$, the quantities

$$\widehat{P}\left(c_{i,j}(t) = \mathcal{C}_k | X_{i,j}(t)\right) = \frac{\exp \widehat{\theta}\left(\mathcal{C}_k | X_{i,j}(t)\right)}{\sum_{k'=1}^{K} \exp \widehat{\theta}\left(\mathcal{C}_{k'} | X_{i,j}(t)\right)},$$

where

$$\widehat{\theta}\left(\mathcal{C}_k | X_{i,j}(t)\right) = \widehat{\alpha}_k + \sum_{c \in V_{i,j}(t)} \sum_{l=1}^{K} \widehat{\beta}_{kl} \mathbb{1}_{[c=\mathcal{C}_l]} + \sum_{r=1}^{p} \widehat{\gamma}_{kr} Y_{i,j}^r.$$

At each pixel (i, j) for the predicted map on date t, we affect the most probable vegetation type namely the $\mathcal{C}_k$ which maximizes

$$\left\{\widehat{P}\left(c_{i,j}(t) = \mathcal{C}_k | X_{i,j}(t)\right)\right\}_{k=1,\ldots,K}.$$

Programs have been made using R programm (see R Development Core Team (2005)) and are avalaible on request.

4.2 Multilayer perceptron

We use a neural network with one hidden layer having q_2 neurons (where q_2 is a parameter to be calibrated). The inputs of the neural network are:

- For the *time series*, the disjonctive form of the value of the pixel;
- For the *spatial aspect*, the weighted frequency of each type of land cover in the neighbourhood of the pixel;
- the environmental variables.

The output is the estimation of the probabilities (1).

The estimation is also made in two stages:

- The *estimation step* produces the estimated weights as described in (6) for a fixed number of neurons (q_2) and a fixed neighbourhood. For this step, the neural network has been trained with an early stopping procedure which allows to stop the optimization algorithm when the validation error (calculated on a part of the data set) is starting to increase (see Bishop (1995)).
  This step is repeated for various values of both neighbourhood and q_2.

- *Validation step:* once an estimation of the optimal weights is given, we choose q_2 and the size of neighbourhood, as for the previous model. Moreover, in order to escape the local minima during the training step, we trained the perceptrons many times for each value of neighbourhoud and of q_2 with various training sets; the "best" perceptron is then chosen according to the minimization of the validation error among both the values of the parameters (size of the neighbourhoud and q_2) and the optimization procedure results.

Programs have been made using Matlab (Neural Networks Toolbox, see Beale & Demuth (1998)) and are avalaible on request.

5 Comparison and discussion

The validation step leads to select the following parameters (Table 1):

	Garrotxes	Alta Alpujarra
Poly. regression		
Size of neighbourhood	9	1
ϵ	10	0.1
ML perceptron		
Size of neighbourhood	7	4
q_2	8	30
perceptron size	19-8-7	35-30-9

Table 1: Parameters selected by the validation step

After the two models have been trained, we build the predicting map on date 2000 (Garrotxes data set) and 2001 (Alta Alpujarra data set). The performances of the two models are compared with a GIS approach.

For the Garrotxes data set, the results are summarized in Table 2 and the frequency of error for each land cover type is made on the pixels which are really of this land cover type. We focus on the 6 more frequent land cover types, since the number of agriculture pixels tends to zero. In Figure 6, we can see the three predictive maps given by our approaches and the GIS approach that can be confronted with the real map.

Land cover types	Frequency in the area	Poly. Regression error rate	ML perceptron error rate	GIS error rate
Coniferous forests	40.9 %	11.9 %	10.6 %	11.4 %
Deciduous forests	11.7 %	51.7 %	45.8 %	55.3 %
Scrubs	15.1 %	57.1 %	54.5 %	51.9 %
Broom lands	21.6 %	14.4 %	16.2 %	17.1 %
Grass pastures	5.7 %	59.2 %	59.4 %	54.4 %
Grasslands	4.8 %	25.6 %	19.3 %	30.4 %
Overall		**27.2 %**	**25.7 %**	**27.2 %**

Table 2: Missclassification rates for the Garrotxes

For the Alta Alpujarra data set, the results are summarized in Table 3 (land cover type under 5 % of the area have been omitted). Predicted maps and real maps are compared in Figure 7.

Land cover types	Frequency in the area	Poly. Regression error rate	ML perceptron error rate	GIS error rate
Deciduous forests	10.9 %	3.5 %	2.6 %	14.3 %
Scrubs	33.0 %	3.1 %	1.4 %	15.2 %
Pasture	20.8 %	0.6 %	0 %	12.5 %
Coniferous refor.	9.23 %	3.5 %	16.3 %	1.9%
Fallows	18.8 %	32.5 %	41.4 %	46.8%
Irrigated cropland	5.8 %	8.9 %	6.8 %	38.9%
Overall		**9.0 %**	**11.28 %**	**21.1 %**

Table 3: Missclassification rates for the Alta Alpujarra

First of all, the predictive maps provided by the two statistical methods are coherent, smooth and close to reality. This can also be shown through the good error rates (about 25 % - 27 % for the Garrotxes data set and 9 % - 12 % for the Alta Alpujarra) which are clearly a good performance considering the poverty of the data (we only have 3 or 4 dates to train the models).
Furthermore, the striking fact is that the "automatic" statistical approaches do as well (Garrotxes data set) or even much better (Alta Alpujarra) than the guided GIS approach. This is an interesting point in order to help improving the classical geographical approach to predicting land cover, and better understand the environmental changes in time and space. Moreover, the "automatic" statistical methods are much faster than the GIS as they do not use any expert knowledge which take a long time to be modelized and need to be remade for each area. On the contrary, the polychotomous regression modelling and the multilayer perceptron approach does not lead on these data sets to significant differences. The first method is much faster to train and it is then quite attractive to use it. However, we think that, on a general point of view, the greater

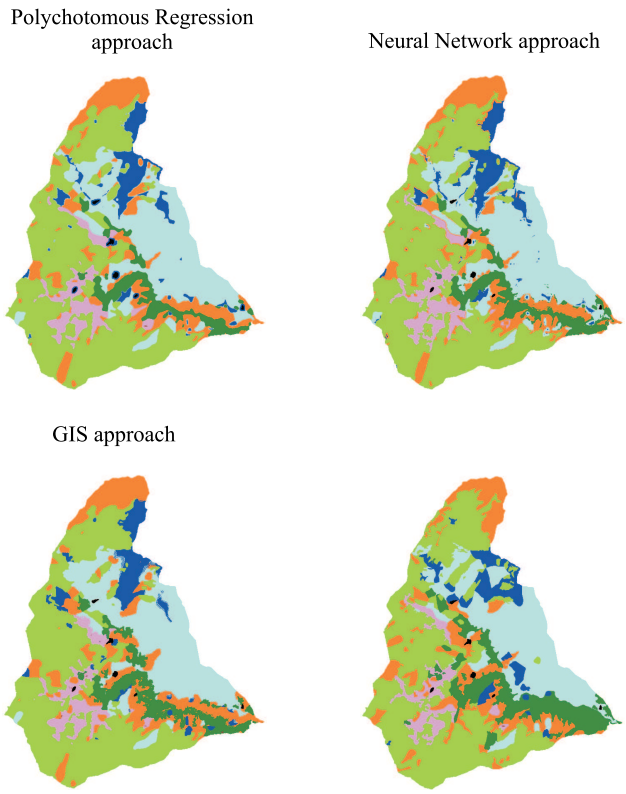

Figure 6: Predictive maps for the various approaches on date 2000 and real map (bottom right)

flexibility of multilayer perceptron can be usefulness to predict land cover for other data sets where a parametric model could fail.

The main advantage of the automatic statistical approaches is in the fact that they simultaneously take into account the spatio-temporal aspect of the problem and also the environmental variables. GIS works in two steps: it first predicts the number of pixels for each land cover type by a simple temporal model and then take into account the spatial aspect and the environmental variables to allocate these pixels spatially. This can partially explain, for the Alta Alpujarra data set, its bad performances as the coniferous reforestation used to be important in the 60's and has then be given up. This leads the GIS to predict, in the 2001 map, much more coniferous reforestation pixels than in the real map: 18.8 % of the pixels are predicted in the coniferous reforestation type against 7.9 % for the multilayer perceptron, 9.6 % for the polychotomous regression modelling and 9.2 % for the real map. Then GIS approach has a much lower error rate on the coniferous reforestation land cover type but a bigger one for the other ones.

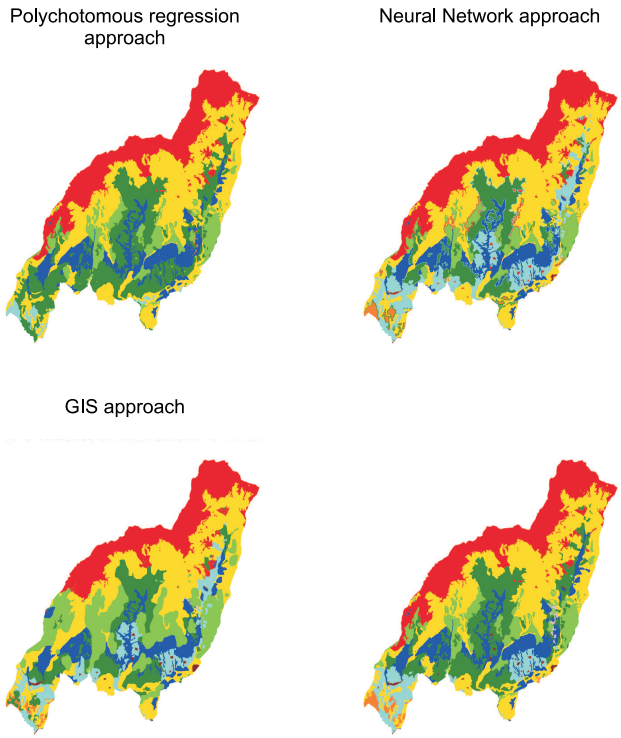

Figure 7: Predictive maps for the various approaches on date 2001 and real map (bottom right)

Finally, looking further in the missclassification rates for the various land cover types, we can see that the most dynamic land cover type are harder to train: this is the case, for instance, for the scrubs in the Garrotxes area where they tend to grow fast and become deciduous forests; this is also the case, in the Alta Alpujarra for the fallows and irrigated croplands because agricultural lands are tending to be left. These dynamics could be better predicted by adding pertinent informations for these kinds of land cover types (density of the scrubs, for example, can help knowing if they can, or not, become forests).

6 Conclusion

Finally, this work shows the great potential of the two statistical models in predictive prospection on geographical data. These models have as good performances as GIS approach and we can hope that a combination of the two points of view (statistics and GIS) can improve the land cover predictions: the empirical first step of the GIS could be improved by being replaced by one of these statistical approaches. This issue, that is of big interest for geographers, is

still under study as the GIS approach is performed though pre-made programms and has then to be totally re-though to that aim.

Another aspect that has to be worked on is the form of the data: for example, we underlined that an information on the density of the scrubs is needed to better understand their evolution. This helps geographers to better understand what is of interest for predicting the land cover evolution for their future studies.

7 Acknowledgements

The authors are grateful to the Ministerio de Cencia y Tecnologia who supports this research (Plan nacional de investigacion cientifica, Desarollo e innovacion tecnologica, BIA 2003_01499).

References

Beale, M. & Demuth, H. (1998). *Neural network toolbox user's guide.* The Matworks Inc. Version 3.

Bishop, C. (1995). *Neural Networks for Pattern Recognition.* Oxford University Press, New York.

Cardot, H., Faivre, R., & Goulard, M. (1993). Functional approaches for predicting land use with the temporal evolution of coarse resolution remote sensing data. *Journal of Applied Statistics*, **30**, 1185–1199.

Hornik, K. (1991). Approximation capabilities of multilayer feedfoward networks. *Neural Networks*, **4**(2), 251–257.

Hosmer, D. & Lemeshow, S. (1989). *Applied logistic regression.* Wiley, New York.

Kooperberg, C., Bose, S., & Stone, J. (1997). Polychotomous regression. *Journal of the American Statistical Association*, **92**, 117–127.

Lai, T. & Wong, S. (2001). Stochastic neural networks with applications to nonlinear time series. *Journal of the American Statistical Association*, **96**(455), 968–981.

R Development Core Team (2005). *R: a language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Paegelow, M. & Camacho Olmedo, M. (2005). Possibilities and limits of prospective GIS land cover modeling - a compared case study: Garrotxes (France) and Alta Alpujarra Granadina (Spain). *International Journal of Geographical Information Science*, **19**(6), 697–722.

Paegelow, M., Villa, N., Cornez, L., Ferraty, F., Ferré, L., & Sarda, P. (2004). Modélisations prospectives de l'occupation du sol. Le cas d'une montagne méditerranéenne. *Cybergéo*, **295**.

Parlitz, U. & Merkwirth, C. (2000). Nonlinear prediction of spatio-temporal time series. In *ESANN'2000 proceedings* 317–322, Bruges.

White, H. (1989). Learning in artificial neural network: a statistical perspective. *Neural Computation*, **1**, 425–464.