Journal of Applied Statistics, Vol. 30, No. 10, December, 2003, 1185–1199 Routledge Taylor & Francis Group

Functional approaches for predicting land use with the temporal evolution of coarse resolution remote sensing data

HERVÉ CARDOT, ROBERT FAIVRE & MICHEL GOULARD, *INRA*
Toulouse, Biométrie et Intelligence Artificielle, 31326 Castanet-Tolosan cedex, France

ABSTRACT *The sensor SPOT 4/Végétation gives every day satellite images of Europe with medium spatial resolution, each pixel corresponding to an area of 1 km × 1 km. Such data are useful to characterize the development of the vegetation at a large scale. The pixels, named 'mixed' pixels, aggregate information of different crops and thus different themes of interest (wheat, corn, forest, . . .). We aim at estimating the land use when observing the temporal evolution of reflectances of mixed pixels. The statistical problem is to predict proportions with longitudinal covariates. We compared two functional approaches. The first relies on varying-time regression models and the second is an extension of the multilogit model for functional data. The comparison is achieved on a small area on which the land use is known. Satellite data were collected between March and August 1998. The functional multilogit model gives better predictions and the use of composite vegetation index is more efficient.*

1 Land use and mixed pixels

On board SPOT 4, a satellite launched in March 1998, the Végétation sensor gives, at a high temporal resolution, daily images of Europe at a coarse spatial resolution, each pixel corresponding to a ground area of 1 km^2. The information given by this sensor are the reflectances, i.e. the proportion of reflected radiation in the four spectral bands, Blue, Red, Near Infra-Red and Short Infra-Red. This information allows one to characterize the development of vegetation and crops at the scale of a small country (Tucker, 1979). Because, in Europe and particularly

Correspondence: H. Cardot, INRA Toulouse, Biométrie et Intelligence Artificielle, 31326 Castanet-Tolosan cedex, France. E-mail: cardot@toulouse.inra.fr

ISSN 0266-4763 print; 1360-0532 online/03/0101185-15 © 2003 Taylor & Francis Ltd
DOI: 10.1080/0266476032000107187

in France, the size of plots is much less than 1 km^2, the observed reflectances are a mixture of different information since they contain different agricultural plots (maize, wheat, forest, ...); such pixels are named *mixed pixels*.

Our aim is to estimate the land use, i.e. the proportion of each type of culture or land cover inside each mixed pixel. Despite the medium spatial resolution, we take advantage of the high temporal resolution of such a sensor to derive estimations of the land use.

The classical land cover characterization deals with high spatial resolution data (Landsat data or SPOT 4/HRV sensor with a resolution of 30 m and 20 m respectively). The most well-known methods are classifications, supervised or not (Richards, 1994); the multispectral bands at two or three different dates during the season permit us to cluster pixels into predefined (or not) classes of the cover. At such a resolution, the problems related to mixed pixels concern only plot boundaries, edge detection, and classification methods give a binary yes or no affectation to classes that are natural classes (wheat, maize, sunflower, etc). When dealing with coarse spatial resolution, pure pixels are difficult to define, except if we consider composite classes. This is the main concern of land cover mapping of large areas, especially in the context of a global change study as an entry of the climate models (a special issue of the *International Journal of Remote Sensing* is dedicated to global and regional land cover characterization from satellite data, introduced by Defries & Belward, 2000). We are not interested in defining composite classes because they are difficult to use as entries of crop growth models and we aim at estimating directly the proportions of pure classes in each pixel of our images.

When observing mixed pixels, the reflectance is a function of the characteristic reflectances of each particular crop. In the considered visible and near infra-red wavelengths, we assume the combination is linear. The reflectance $Y_i(t)$ of a pixel i at date t is then written as:

$$Y_i(t) = \sum_{j=1}^{p} \pi_{ij}\rho_{ij}(t) + \varepsilon_i(t) \tag{1}$$

where $\rho_{ij}(t)$ and π_{ij} are the characteristic reflectance and the proportion of land use of crop j in pixel i, and $\varepsilon_i(t)$ is an error term. We assume that the land use is fixed during the observation period (about 6 months). With such a model, two questions of interest arise.

(1) How can we estimate the characteristic reflectance $\rho_{ij}(t)$ knowing the proportions π_{ij}?
(2) How can we estimate the land use π_{ij} in each pixel knowing the aggregated reflectances $Y_i(t)$?

Faivre & Fischer (1997) proposed a statistical approach for answering the first question. A random coefficient linear model was built to estimate the distribution of the characteristic reflectances assuming the land use was known. Unmixing or disaggregation was performed independently on each image, i.e. at each date, and Faivre *et al.* (2000) assimilated the predicted local reflectances of wheat into a crop growth model in order to predict the total production of winter wheat at a regional scale.

Our work deals with the second question; that is, to estimate the land use from the temporal evolution of the pixels. In fact, estimating acreage is the first step in

predicting regional crop productions. The two sources of available information are the measures in the frequency domain, but unfortunately we only have these measures at four different frequencies, and the temporal evolution of the reflectance of each pixel. We consider the latter to propose two non-parametric approaches to predict the land use.

The first approach is the most natural one. It consists of assuming that model (1) is the true model and the characteristic curves ρ_{ij} do not vary with the location (i.e. with pixel i). By means of varying-time regression models (Hastie & Tibshirani, 1993; Hoover *et al.*, 1998) we can obtain global non-parametric estimations of these characteristic curves. We are then able to estimate the vector of proportions in another region by considering a constrained least squares estimator.

The second approach is direct and is based on an extension of the generalized linear models (McCullagh & Nelder, 1989) for functional data (Ramsay & Silverman, 1997). The proportions are assumed to be drawn from a multinomial distribution whose parameters depend on the temporal evolution of the reflectance. If we suppose that the covariates are the observed reflectances along time, these covariates are too numerous and too highly correlated to be used directly in a multilogit model. We reduce the dimension by means of a functional principal components analysis (Deville, 1974; Ramsay & Silverman, 1997) and we use the principal components as covariates. Such a strategy has already been adopted by Cardot *et al.* (1999) in the setting of a functional linear model. Afterwards, the principal components are selected by means of a test based on the likelihood ratio.

A comparison of the two approaches is done in a small region of the North-West of France, of size 1600 km^2, in which we know the land use. We considered images from the sensor Végétation between March and August 1998. We compare the predictions obtained for each channel separately and two vegetation indices in order to evaluate the potential of each channel to predict land use.

In the following, we first present the data and give notations. In the third and fourth sections, we describe the two different statistical approaches. These methods are compared in the last section. From a computational point of view, the procedures are written in MATLAB and are available on request. More details may be found in Cardot *et al.* (2000).

2 Notations and presentation of the data

The studied area is the county of Chartres whose size is about 40 km $\times$ 40 km. It is located in the North-West of France and it is a pilot site of the European project MARS (Monitoring Agriculture with Remote Sensing) in which we know the land use with good precision (see Fig. 1). The vectors of proportions of land use at the Vegetation pixel resolution in this area (Fig. 2) were computed thanks to the high resolution data.

Let us denote by π_{ij}, $j = 1, \ldots, p$, the proportion of land use of crop j in pixel i of 1 km^2. By construction, $\pi_{ij} \geqslant 0$ and $\Sigma_j \pi_{ij} = 1$ for $i = 1, \ldots, n = 1554$. Ten ($p = 10$) different classes of crops were present in this area and the mean value of their proportion is given in Table 1.

We also have satellite data given by the sensor Végétation at a high temporal frequency (theoretically each day) for four channels (Blue, Red, Near Infra-Red and Short Wavelength Infra-Red) between March 1998 and August 1998. We built vegetation indices that are frequently used in bioclimatology and remote sensing. These indices are the NDVI (Normalized Difference Vegetation Index, Tucker,

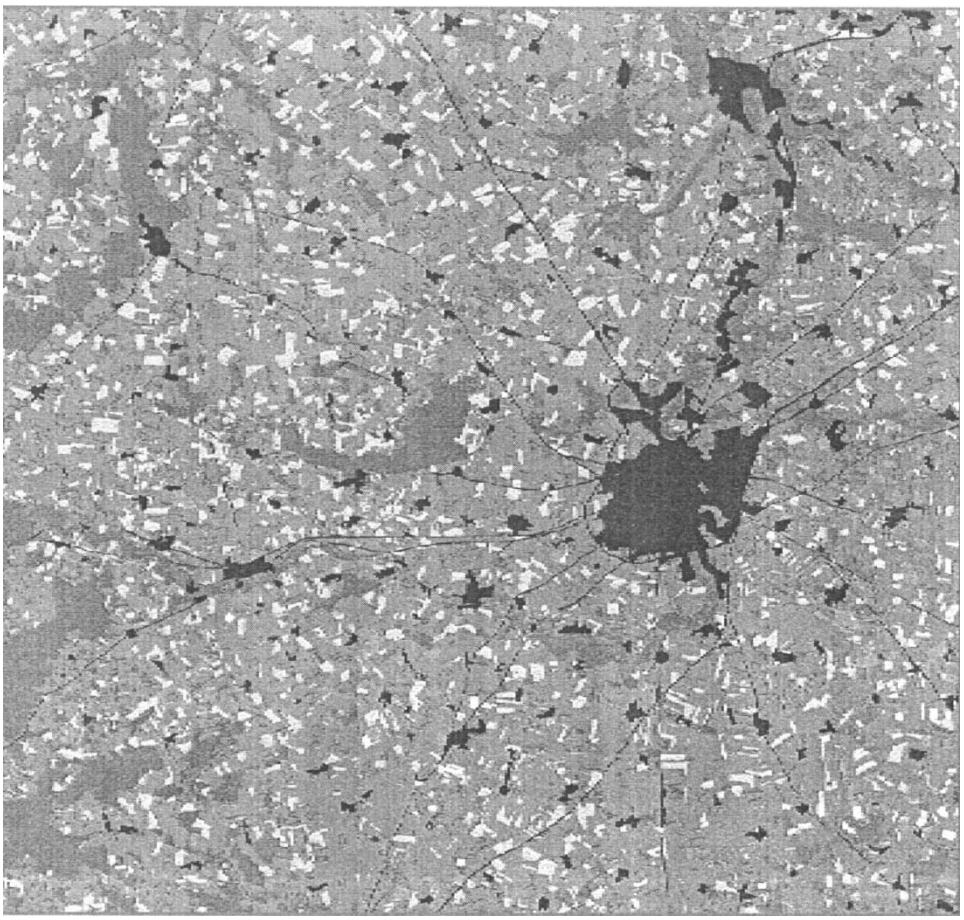

FIG. 1. Land use map in the county of Chartres at high resolution.

1979) and PVI (Perpendicular Vegetation Index, Richardson & Wiegang, 1977). They are obtained from the reflectances in the Red (R) and the NIR:

$$\text{NDVI}(t) = \frac{NIR(t) - R(t)}{NIR(t) + R(t)} \tag{2}$$

$$\text{PVI}(t) = \frac{NIR(t) - \alpha R(t)}{\sqrt{1+\alpha^2}} \text{ with } \alpha = 1.2 \tag{3}$$

The curves of reflectance in each channel for each pixel i are denoted by $\mathbf{Y}_i = [Y_i(t_1), \ldots, Y_i(t_K)]^{\mathrm{T}}$ where $t_1 < \ldots < t_k < \ldots < t_K$ are the instants of measure. The time is normalized so that $t_1 = 0$ and $t_K = 1$. The images in which the clouds were too important were removed to finally get 39 different discretization points ($K = 39$).

3 The characteristic curves approach

We suppose that the reflectance of each crop does not depend on the location of the pixels so that model (1) becomes simpler,

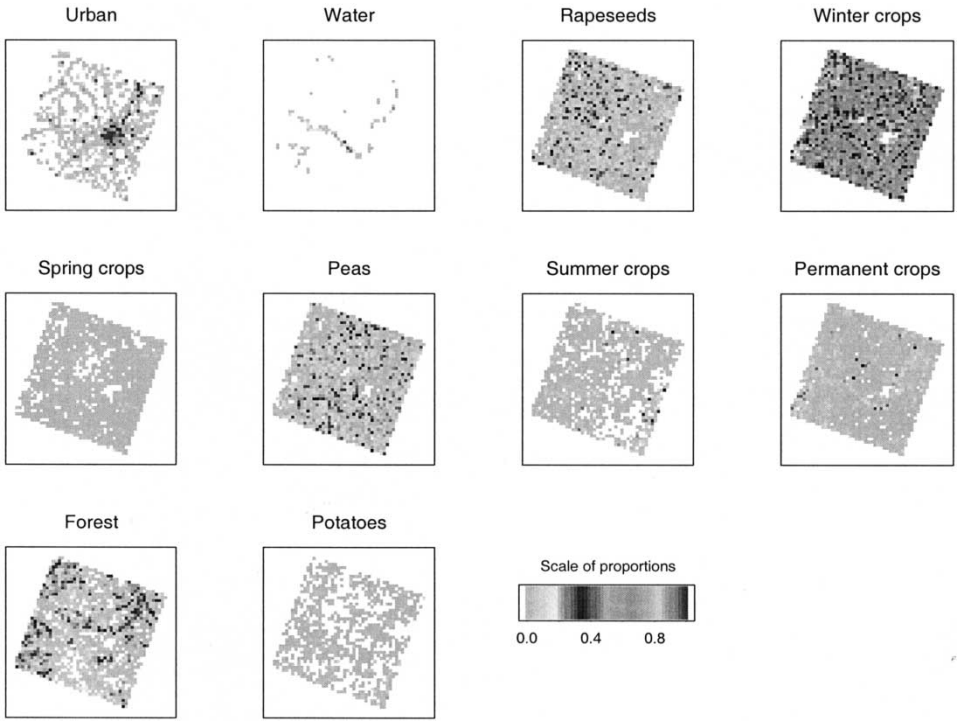

FIG. 2. Land use in the county of Chartres at the low resolution 1 km × 1 km.

TABLE 1. Mean proportion of each class

Land use	Percentage
Urban	6.07
Water	0.32
Rapeseed	11.81
Winter crops	49.13
Spring crops	0.60
Peas	11.51
Summer crops	2.01
Permanent crops	4.27
Forest	13.50
Potatoes	0.79

$$Y_i(t_K) = \sum_{j=1}^{p} \pi_{ij}\rho_j(t_k) + \varepsilon_i(t_k) \qquad \begin{matrix} i = 1, \ldots, n \\ k = 1, \ldots, K \end{matrix} \tag{4}$$

To predict the proportions π_{ij}, we will proceed in two steps:

Step 1. Estimation of the characteristic curves ρ_j of each crop j,
Step 2. Prediction of the land use with the help of the estimated curves obtained in the first step.

Let $\mathbf{Y}_i = [Y_i(t_1), \ldots, Y_i(t_K)]^T$ denote the discretized trajectory of the reflectance of the ith pixel.

3.1 Global estimation of the characteristic curves

We propose to estimate simultaneously all the characteristic curves with a varying time regression approach similar to the one studied by Hoover *et al.* (1998). We can suppose these curves are smooth, that is to say the biological phenomenons and the growth of crops do not occur too rapidly. Then we can approximate them by smooth functions such as B-splines (Dierckx, 1993). Let $\{B_1, \ldots, B_{q+d}\}$ denote a basis of B-splines functions of order q with d interior knots. We approximate the curves ρ_j as follows:

$$\tilde{\rho}_j(t_k) = \sum_{r=1}^{q+d} \tilde{\theta}_{rj} B_r(t_k) \tag{5}$$

and then estimate the vectors of coordinates by minimizing in $\boldsymbol{\theta}$

$$\mathcal{J}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \sum_{i=1}^{n} \sum_{k=1}^{K} w_k \left[Y_i(t_k) - \left(\sum_{j=1}^{p} \pi_{ij} \tilde{\rho}_j(t_k) \right) \right]^2 + \sum_{j=1}^{p} \lambda_j R(\tilde{\rho}_j) \tag{6}$$

The smoothing parameters λ_j allow us to control the smoothness of the solutions by means of the penalty

$$R(\tilde{\rho}_j) = \int_0^1 \left[\frac{\mathrm{d}^2}{\mathrm{d}t^2} (\tilde{\rho}_j(t)) \right]^2 \mathrm{d}t \tag{7}$$

The weights w_k are associated to a quadrature rule and we took $w_1 = (t_2 - t_1)/2$, $w_K = (t_K - t_{K-1})/2$, $w_k = (t_{k+1} - t_{k-1})/2$, $k = 2, \ldots, K-1$. Let $\mathbf{W}$ define the diagonal matrix with diagonal elements w_k. Let $\mathbf{B}$ denote the matrix with elements $[\mathbf{B}]_{kr} = B_r(t_k)$, $k = 1, \ldots, K$ and $r = 1, \ldots, k+d$, and $\mathbf{B}^{\mathrm{T}}(t) = [B_1(t), \ldots, B_{q+d}(t)]$. The penalty (7) may be expressed matricially (O'Sullivan, 1986),

$$R(\tilde{\rho}_j) = \tilde{\boldsymbol{\theta}}_j^{\mathrm{T}} \mathbf{P}_2 \tilde{\boldsymbol{\theta}}_j$$

where $\tilde{\boldsymbol{\theta}}_j = (\tilde{\theta}_{1j}, \ldots, \tilde{\theta}_{(q+d)j})^{\mathrm{T}}$. The solutions $\hat{\boldsymbol{\theta}}_j^{\mathrm{T}} = (\hat{\theta}_{1j}, \ldots, \hat{\theta}_{(q+d)j})$, $j = 1, \ldots, p$, of equation (6) are equivalently given by solving the system of equations

$$\sum_{j'=1}^{p} \mathbf{A}_{jj'} \hat{\boldsymbol{\theta}}_{j'} + \lambda_j \mathbf{P}_2 \hat{\boldsymbol{\theta}}_j = \mathbf{Q}_j, \quad j = 1, \ldots, p \tag{8}$$

with $\mathbf{A}_{jj'} = (\Sigma_{i=1}^{n} \pi_{ij} \pi_{ij'} \mathbf{B}^{\mathrm{T}} \mathbf{W} \mathbf{B})$ and $\mathbf{Q}_j = \Sigma_{i=1}^{n} \pi_{ij} \mathbf{B}^{\mathrm{T}} \mathbf{W} \mathbf{Y}_i$.

The solutions of equation (8) are obtained by applying the iterative Backfitting algorithm (Hastie & Tibshirani, 1993). In the implementation, we took $\lambda = \lambda_j$, $j = 1, \ldots, p$ and chose its value by cross-validation (Green & Silverman, 1994).

3.2 Prediction of land use by constrained regression

We want to achieve prediction on a new pixel u or, in other words, we want to estimate the land use for u given the evolution of its reflectance $\mathbf{Y}_u = (Y_u(t_{u,1}), \ldots, Y_\ell(t_{u,K_u}))$ at design points, $t_{u,1} < \ldots < t_{u,K_u}$. Note that these design points can be different than those used to estimate the characteristic curves.

The B-spline approximations allow us to extrapolate the values of the characteristic curves at these new design points and the prediction of the vector of proportion $\boldsymbol{\pi}_u = (\pi_{u1}, \ldots, \pi_{up})$ is obtained by minimizing

$$\underset{\boldsymbol{\pi}_u \in \mathbb{R}^p}{\text{Argmin}} \sum_{k=1}^{K_u} w_{u,k} \left(Y_u(t_{u,k}) - \sum_{j=1}^{p} \pi_{uj} \hat{\rho}_j(t_{u,k}) \right)^2 \tag{9}$$

under the constraints that $\pi_{uj} \geqslant 0$ and $\Sigma_{j=1}^{p} \pi_{uj} = 1$. The new quadrature weights $w_{u,k}$ depend on the new design points.

4 The multilogit model for functional data

We suppose now that the proportions π_{ij} given the temporal evolution of the reflectance $\{Y_i(t), t \in T = [0, 1]\}$ can be modelled as resulting from a multinomial distribution whose parameters satisfy

$$\mathbb{E}(\pi_{ij} \mid Y_i) = \frac{\exp\left(\delta_j + \int_T \beta_j(t)(Y_i(t) - \mathbb{E}(Y_i(t)))\, \mathrm{d}t \right)}{\sum_{l=1}^{p} \exp\left(\delta_l + \int_T \beta_l(t)(Y_i(t) - \mathbb{E}(Y_i(t)))\, \mathrm{d}t \right)} \tag{10}$$

We can give a sketch of interpretation on this approach. Indeed, each coarse resolution pixel is assumed to be composed of numerous small agricultural plots of similar area. Each of these plots of a pixel i is, with a probability $E(\pi_{ij} \mid Y_i)$, of the theme j of the land use. Thus, the number of plots of the themes exactly follows a multinomial distribution and we observe the proportions.

For identifiability reasons we took $\beta_p = 0$ and $\delta_p = 0$. Then, model (10) can be written, for $j = 1, \ldots, p-1$,

$$\log\left(\frac{\mathbb{E}(\pi_{ij} \mid Y_i)}{\mathbb{E}(\pi_{ip} \mid Y_i)} \right) = \delta_j + \int_T \beta_j(t) X_i(t)\, \mathrm{d}t \tag{11}$$

where $X_i(t) = Y_i(t) - \mathbb{E}(Y_i(t))$. Each functional coefficient β_j may have an interpretation by comparison to the reference function $\beta_p = 0$. For instance, if β_j is a positive function, then the ratio of the proportion will be higher than the mean value and thus the class j will be more important in the pixel i, if the centred reflectance curve is positive. Nevertheless, if the estimations have oscillating features, giving an interpretation may be a hard task.

We aim at estimating the vector $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_{p-1})^{\mathrm{T}}$ and the functional coefficients $\beta_j(t)$, $j = 1, \ldots, p-1$. The estimations are obtained by means of the maximum likelihood but we have to reduce the dimension in order to reduce the computation time and obtain sufficiently stable estimators. Another way to reach this is to penalize the likelihood, as was done in Marx & Eilers (1999). Nevertheless, it appears to us that this approach would need much more computation than we propose, in particular to choose the values of the smoothing parameters.

4.1 Maximum likelihood estimator

In the setting of generalized linear models, an extensive literature exists on the maximum likelihood principle (see, for example, McCullagh & Nelder, 1989, and

Fahrmeir & Tutz, 1994, for seminal references). In our functional setting, if we suppose that the pixels are independent given the $\mathbf{Y}_i$, we can give an approximation to the log-likelihood

$$\log \mathscr{L} = C + \sum_{i=1}^{n} \sum_{j=1}^{p} \pi_{ij} (\delta_j + \int_T \beta_j(t) X_i(t)\, \mathrm{d}t) \tag{12}$$

$$- \log \left[\sum_{j=1}^{p} \exp \left(\delta_j + \int_T \beta_j(t) X_i(t)\, \mathrm{d}t \right) \right] \tag{13}$$

where C is a constant that does not depend on the parameters. The likelihood is evaluated by approximating the integrals with quadrature schemes.

Let $\mathbf{X}_i = \mathbf{Y}_i - \frac{1}{\mathrm{n}} \Sigma_{i=1}^{n} \mathbf{Y}_i$ denote the ith centred discretized curve. An approximation of model (10) is given by

$$\int_T \beta_j(t) X_i(t)\, \mathrm{d}t \approx \sum_{k=1}^{K} w_k \beta_{jk} X_{ik} \tag{14}$$

where $X_{ik} = X_i(t_k)$, $\beta_{\ell k} = \beta_\ell(t_k)$ and the w_k are the weights associated with the quadrature scheme. Nevertheless, the number of parameters $(= K \times (p-1))$ of this model is too large to give satisfactory estimates.

4.2 Reducing the dimension

In the setting of the functional linear regression, Hastie & Mallows (1993) have proposed adapting the principal components regression to the functional case and Cardot *et al.* (1999) have shown the consistency of this estimator. Thus, in a first step, we perform a functional principal components analysis (see Deville, 1974, for a seminal work and also Ramsay & Silverman, 1997) which permits us to reduce the dimension of the data in an optimal way according to the variance. This procedure relies on the spectral decomposition of the covariance operator of the curves $Y_i(t)$, which is approximated by the covariance matrix

$$\Gamma_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^{\mathrm{T}} \mathbf{W} \text{ with metric } \mathbf{W}$$

The eigenelements $(\tilde{\lambda}_\ell, \tilde{\mathbf{v}}_\ell)$ satisfy

$$\Gamma_n \tilde{\mathbf{v}}_\ell = \tilde{\lambda}_\ell \tilde{\mathbf{v}}_\ell$$

with $\tilde{\lambda}_1 \geqslant \tilde{\lambda}_2 \geqslant \ldots$ and $\tilde{\mathbf{v}}_{\ell_1}^{\mathrm{T}} \mathbf{W} \tilde{\mathbf{v}}_{\ell 2} = 1$ if $\ell_2 = \ell_1$ and 0 if $\ell_2 \neq \ell_1$. The best linear approximation of the discretized trajectories in a q dimensional subspace is obtained by projection on the first q eigenvectors and we may write:

$$\int_T X_i(t) \beta_j(t)\, \mathrm{d}t \approx \sum_{\ell=1}^{q} \tilde{c}_{i\ell} \boldsymbol{\beta}_j^{\mathrm{T}} \mathbf{W} \tilde{\mathbf{v}}_\ell \tag{15}$$

The new covariates are now the principal components $\tilde{c}_{i\ell} = \mathbf{X}_i^{\mathrm{T}} \mathbf{W} \tilde{\mathbf{v}}_\ell$ and the parameters to be estimated are the coordinates $\alpha_{j\ell} = \boldsymbol{\beta}_j^{\mathrm{T}} \mathbf{W} \tilde{\mathbf{v}}_\ell$ of the functional coefficients β_j on the eigenfunctions $\tilde{v}_\ell$:

$$\beta_j(t) \approx \sum_{\ell=1}^{q} \alpha_{j\ell} \tilde{v}_\ell(t) \tag{16}$$

the functions $\tilde{v}_\ell(t)$ being derived by spline interpolation of the eigenvectors $\tilde{\mathbf{v}}_\ell$.

Afterwards, we can express the part of log likelihood that depends on the parameters as follows:

$$\begin{aligned} \log \mathscr{L} = & \sum_{i=1}^{n} \sum_{j=1}^{p} \pi_{ij} \left(\delta_j + \sum_{k=1}^{q} \alpha_{jk} \tilde{c}_{ik} \right) \\ & - \sum_{i=1}^{n} \log \left[\sum_{l=1}^{p} \exp \left(\delta_l + \sum_{k=1}^{q} \alpha_{lk} \tilde{c}_{ik} \right) \right] \end{aligned} \tag{17}$$

and maximizes it in α_{jk} and δ_j, $j = 1, \ldots, p-1$, $k = 1, \ldots, q$.

The number of parameters may still be large and we decided to select the most significative ones by means of the likelihood ratio test (Lehmann, 1986) with an ascendant procedure.

4.3 Prediction of land use

The prediction of the vector of proportions of a new pixel u is done in three steps. The trajectory Y_u is first interpolated to get values of the reflectance at instants $t_1, \ldots, t_K$. Then, the mean trajectory $n^{-1} \Sigma_{i=1}^{n} \mathbf{X}_i$ is subtracted in order to obtain the centred discretized trajectory $\mathbf{X}_u$ and the principal components $\tilde{c}_{uk}$. The prediction for the proportion of the different crops is done by

$$\hat{\pi}_{uj} = \frac{\exp \left(\hat{\delta}_j + \sum_{k=1}^{q} \tilde{c}_{uk} \hat{\alpha}_{jk} \right)}{\sum_{l=1}^{p} \exp \left(\hat{\delta}_l + \sum_{k=1}^{q} \tilde{c}_{uk} \hat{\alpha}_{lk} \right)} \quad j = 1, \ldots, p \tag{18}$$

5 Application to satellite data

The data consist in the observation of $n = 1554$ pixels at $K = 39$ distinct instants in four channels. We drew a sample, named the *learning* sample, which contains 1055 pixels. It was used to estimate the parameters and to select the 'best' models. The other 499 pixels were used to make the comparisons. We chose the following criterion to evaluate the ability of each method to give good predictions:

$$C_{ij} = \frac{|\, \pi_{ij} - \hat{\pi}_{ij} \,|}{\pi_{.j}} \tag{19}$$

where $\pi_{.j} = \frac{1}{499} \Sigma_{i=1}^{499} \pi_{ij}$ is the mean proportion of class j in the test sample. We also considered the most simple model, named M_0, as a benchmark to indicate if it is worth building sophisticated statistical models. It consists of predicting the land use of one crop by its empirical mean in the learning sample. This is a particular case of the multilogit model with no covariates.

The comparisons were done for each channel, Blue (B), Red (R), Near Infra-Red (NIR) and Short Wave Infra-Red (SWIR) and for the vegetation indices NDVI and PVI.

5.1 Characteristic curves

The order of the B-splines was $q = 4$ and we considered $d = 5$ interior knots located at the quantiles of the design points.

The characteristic curves obtained for the NDVI are presented in Fig. 3. We notice that curves associated with the crops 'rapeseed' and 'winter crops' are very similar. Let us also notice the stability along time of the curve associated with the class 'urban'.

The median errors are given in Table 2 and we can immediately see that predictions are better with the indices NDVI and PVI than with the original channels. For instance, the error is divided by three, compared with the reference model M_0, for the classes 'summer crops', 'urban' and 'forest'. Nevertheless, some important classes, such as 'winter crops', 'rapeseed' and 'peas', are really badly predicted with this approach.

These poor results may have the following explanation. Indeed, we noticed that some transfer between crops occurs with this method. This is clearly seen on the mean prediction of each crop on the test sample (see Table 3). There is an underestimation of the class 'winter crops' and an overestimation of the class

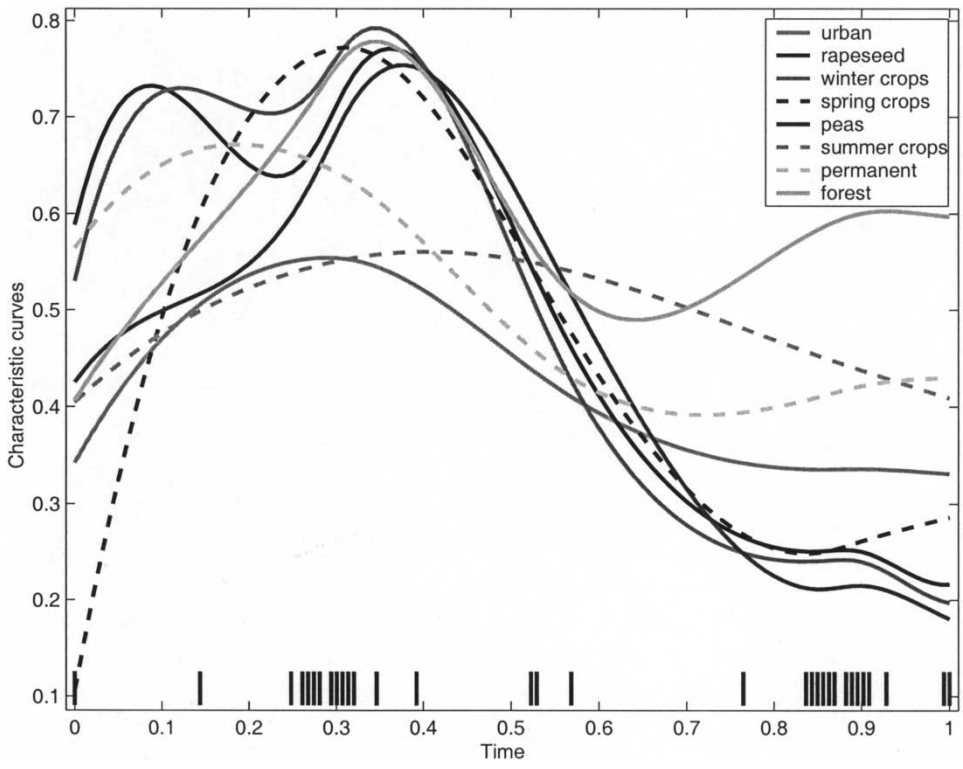

FIG. 3. Characteristic curves for the NDVI index.

TABLE 2. Median errors on the test sample for the characteristic curves approach and the reference model M_0. Bold face numbers correspond to the best predictions and underlined numbers indicate that M_0 gives better prediction

Themes	NDVI	PVI	Blue	Red	NIR	SWIR	M_0
Urban	0.19	0.18	0.18	**0.13**	0.48	0.18	0.86
Water	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	1.30
Rapeseed	0.90	1.15	**0.82**	1.19	1.01	1.26	0.59
Winter crops	**0.35**	0.49	0.36	0.51	0.41	0.57	0.30
Spring crops	0.49	0.85	0.49	0.43	0.67	**0.43**	0.69
Peas	**0.77**	0.94	0.81	0.98	0.84	0.86	0.63
Summer crops	0.21	0.23	3.41	0.34	**0.19**	0.21	0.88
Permanent crops	0.79	**0.72**	1.01	0.78	0.91	0.95	0.61
Forest	0.36	**0.33**	0.48	0.54	0.37	0.34	0.98
Potatoes	0.37	0.24	**0.23**	0.30	0.72	0.30	1.30

TABLE 3. Mean predicted proportions by the characteristic curves approach on the test sample for the different vegetation indices

Themes	NDVI	PVI	Blue	Red	NIR	SWIR	True proportion
Urban	4.5	4.7	7.2	4.1	6.6	3.4	6.7
Water	3.9	2.5	4.6	4.0	1.2	3.8	0.3
Rapeseed	16.1	18.1	15.2	10.8	21.7	21.1	12.3
Winter crops	36.8	35.5	27.6	31.8	32.7	32.1	48.3
Spring crops	2.6	2.6	4.0	2.1	2.2	2.2	0.7
Peas	12.3	14.3	7.7	21.2	14.0	14.1	11.6
Summer crops	2.4	2.0	10.5	3.2	4.4	3.9	2.1
Permanent crops	5.7	4.06	9.8	4.8	3.8	5.8	4.5
Forest	13.0	13.8	9.4	14.8	10.6	11.4	13.0
Potatoes	2.7	2.5	4.0	3.2	2.9	2.3	0.7

'rapeseed' that is certainly due to the similarity of the characteristic curves. With our data, the matrix of the characteristic curves is badly conditioned and, consequently, despite the imposed constraints, the estimation of the parameters is unstable.

5.2 The multilogit approach

The class 'urban' was chosen as the reference class because of its stability along time. For almost all channels, the first ten principal components explain more than 98% of the variability of the reflectance curves. The principal components were selected with a high nominal level of 15% because of the low power of the likelihood ratio test. Even if the first principal component generally explains more than 50% of the variance, it was not systematically selected by this test.

The estimated functional coefficients for the NDVI are drawn in Fig. 4. For classes corresponding to agricultural plots, one can recognize some biological cycles. On the other hand, for the stable classes such as 'forest' or 'permanent crops', the functional coefficients are rather flat. The prediction errors are given in Table 4. We can see that these errors are systematically better for each channel, than the errors of model M_0. The predictions seem to be better with the PVI.

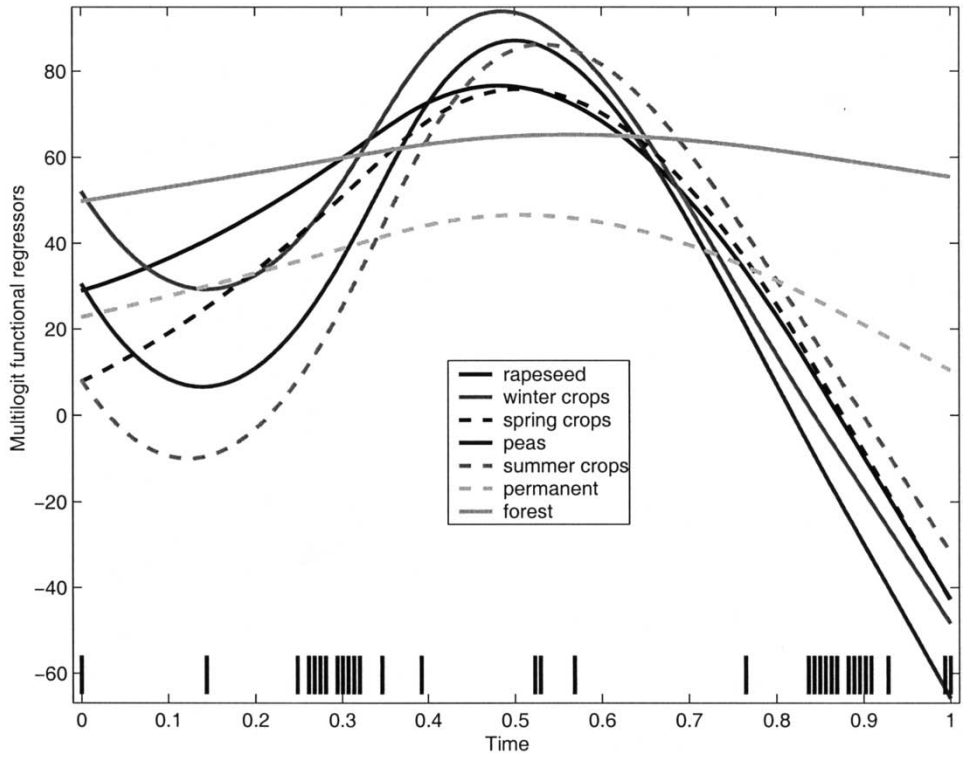

FIG. 4. Functional coefficients of the multilogit approach for the NDVI index.

TABLE 4. Median errors on the test sample for the functional multilogit model and the reference model M_0. Bold face numbers correspond to the best predictions

Themes	NDVI	PVI	Blue	Red	NIR	SWIR	M_0
Urban	0.49	**0.36**	0.47	0.54	0.41	0.51	0.86
Water	0.43	**0.29**	0.78	0.62	0.61	0.31	1.30
Rapeseed	0.48	0.46	**0.45**	0.50	0.47	0.47	0.59
Winter crops	0.20	0.21	**0.19**	0.20	0.22	**0.19**	0.30
Spring crops	0.58	**0.56**	0.60	0.61	0.65	0.61	0.69
Peas	0.50	**0.43**	0.45	**0.43**	0.48	0.46	0.63
Summer crops	0.61	0.68	0.61	0.60	0.76	**0.53**	0.88
Permanent crops	0.47	**0.46**	0.52	0.49	**0.46**	0.50	0.61
Forest	0.34	0.36	0.34	**0.31**	0.45	0.35	0.98
Potatoes	0.90	0.93	0.94	0.90	1.06	**0.85**	1.31

5.3 Comparison and discussion

First, let us notice that the functional multilogit approach, even if it is less natural, gave better predictions for all channels and indices than the characteristic curves method. The main drawback of the latter was the variability of its predictions, even when they are as effective as the multilogit approach, as seen in Fig. 5, which can be mainly due to the identifiability problem of the parameters (recall the discussion about transfer between crops). The satellite Spot 4 was launched in March 1998 and we only had data for the end of the growth period. Moreover, there was a long

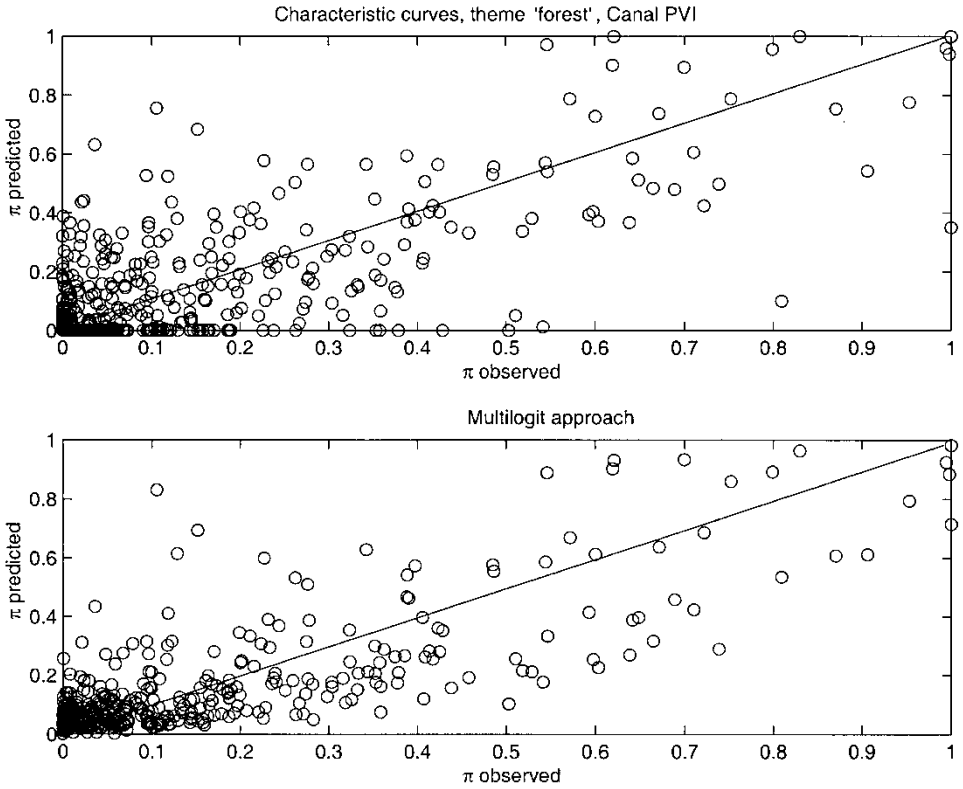

FIG. 5. Predicted proportions of the theme 'forest' for both methods with the channel PVI.

rainy period during this spring and that is why we only kept $K = 39$ images to perform the analysis. This can explain, for instance, that the 'rapeseed' and 'winter crops' curves were very similar and we had problems predicting the proportions of these crops. With better data, the results would probably have been better.

Nevertheless, the effectiveness of the multilogit model may be moderated. Indeed, it relies on a precise estimation of the location parameters δ_j which represent a kind of mean value of each crop j. If we have to make predictions in an area whose partition of the crops is very different then we might get into trouble.

The best predictions seem to be obtained when using the PVI index. For instance, the errors were reduced by about 60% compared with the reference model M_0. Thus, combinations of the original wavelengths may be more appropriate to predict the land use and our future work will deal with finding optimal combinations of the available channels.

6 Concluding remarks

We demonstrated that the use of the temporal evolution of reflectances of mixed pixels allows us to recover information on land use. The procedure needs a learning sample in order to estimate the land use. On the application data we showed that a functional multilogit model gave better results than a more classical functional approach based on characteristic curves. We have built models on observed channel

reflectances and on composite indices. We noticed that the latter gave the better predictions.

In the application, we used each channel reflectance separately and certainly the use of several channel reflectances simultaneously may be better and should be studied. The fact that vegetation composite indices gave better results support this point. Two composite indices were used, PVI is a linear combination of two channel reflectances where NDVI is a non-linear combination. This leads to searching linear as well as non-linear combinations of channel information. This point needs further investigation and will be studied more deeply.

The data set was split into a learning sample in which the land use and the reflectances were known and a test sample for which we predicted the land use. The area where the information was taken is sufficiently homogeneous but, in a practical case, this may not be the case. We must study how to take into account prior information about the land use and the evolution of reflectance themes in the prediction set.

Other methods could also have been used for the prediction. Regression trees or neural networks are potential candidates (see Hastie *et al.*, 2001). These methods are known to be better if threshold effects or strong non-linear effects are suspected to exist. A preliminary work with CART, not presented in this paper, gave good results but this point needs further investigation.

Acknowledgements

The satellite data have been obtained thanks to the 'programme IUC-Végétation'. The high resolution land use has been established by the SCOT company. We would like to thank two master students, Heloise Vialard and Yasmine Yactine, for their preliminary work on these data.

REFERENCES

Cardot, H., Ferraty, F. & Sarda, P. (1999) Functional linear model, *Statistics and Probability Letters*, 45, pp. 11–22.

Cardot, H., Faivre, R. & Goulard, M. (2000) Estimation de l'occupation de sols à partir de l'évolution temporelle des images due capteur Végétation de SPOT, *Rapport Technique, INRA Toulouse, Biométrie et Intelligence Artificielle*, 2, p. 35.

Defries, R. S. & Belward, A. S. (2000) Global and regional land cover characterization from satellite data: an introduction to the Special Issue, *International Journal of Remote Sensing*, 21, pp. 1083–1092.

Deville, J. C. (1974) Méthodes statistiques et numériques de l'analyse harmonique, *Annales de l'INSEE*, 15, pp. 3–101.

Dierckx, P. (1993) *Curves and Surface Fitting with Splines* (Oxford, Clarendon Press).

Fahrmeir, L. & Tutz, G. (1994) *Multivariate Statistical Modelling Based on Generalized Linear Models* (Springer-Verlag).

Faivre, R. & Fischer, A. (1997) Predicting crop reflectances using satellite data observing mixed pixels, *Journal of Agricultural, Biological and Environmental Statistics*, 2, pp. 87–107.

Faivre, R., Bastié, C. & Husson, A. (2000) Integration of VEGETATION and HRVIR data into yield estimation approach. *In:* G. Saint (ed.) *Proceedings of Vegetation 2000, 2 Years of Operation to Prepare the Future*, Space Application Institute, Joint Research Center, Ispra, Varese, Italy, pp. 235–240.

Green, P. J. & Silverman, B. W. (1994) *Nonparametric Regression and Generalized Linear Models. A Roughness Penalty Approach* (Chapman and Hall).

Hastie, T. J. & Mallows, C. (1993) A discussion of 'A statistical view of some chemometrics regression tools' by I. E. Frank and J. H. Friedman, *Technometrics*, 35, pp. 140–143.

Hastie, T. J. & Tibshirani, R. J. (1993) Varying-coefficient models (with discussion), *Journal of the Royal Statistical Society, B*, 55, pp. 757–796.

HASTIE, T. J., TIBSHIRANI, R. J. & FRIEDMAN, J. (2001) *Elements of Statistical Learning: Data Mining, Inference and Prediction* (New York, Springer-Verlag)

HOOVER, D. R., RICE, J. A., WU, C. O. & YANG, L. P. (1998) Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data, *Biometrika*, 85, pp. 809–822.

LEHMANN, E. L. (1986) *Testing Statistical Hypotheses* (Wiley).

MARX, B. D. & EILERS, P. H. (1999) Generalized linear regression on sampled signals and curves: a *P*-spline approach, *Technometrics*, 41, pp. 1–13.

MCCULLAGH, P. & NELDER, J. A. (1989) *Generalized Linear Models* (Chapman and Hall).

O'SULLIVAN, F. (1986) A statistical perspective on ill-posed inverse problems, *Statistical Science*, 4, pp. 502–527.

RAMSAY, J. O. & SILVERMAN, B. W. (1997) *Functional Data Analysis* (Springer-Verlag).

RICHARDS, J. A. (1994) *Remote Sensing Digital Image Analysis* (Berlin, Springer-Verlag).

RICHARDSON, A. J. & WIEGANG, C. L. (1977) Distinguishing vegetation from soil background information, *Photogrammetric Engineering and Remote Sensing*, 43, pp. 1541–1552.

TUCKER, C. J. (1979) Red and photographic infrared linear combinations for monitoring vegetation, *Remote Sensing of Environment*, 8, pp. 127–150.