

Unsupervised sub-pixelic classification using coarse resolution time series and structural information

Amandine Robin, Sylvie Le Hégarat-Mascle, and Lionel Moisan

Abstract—In this paper, a new method is presented for sub-pixelic land cover classification, using both high resolution structural information and coarse resolution temporal information. To that aim, the linear mixture model is used for pixel disaggregation. It enables to describe a coarse resolution time series in terms of the mixture of classes that are represented within each pixel. Then, the Bayes rule and the Maximum A Posteriori criterion lead to the definition of an energy function whose minimum corresponds to the researched optimal classification. A theoretical analysis of the labeling errors that may be obtained using this energy function is provided, raising the main parameters for labeling performance. The optimal classification is computed by combining linear regressions and simulated annealing, leading to an unsupervised algorithm. The method is illustrated with numerical results obtained on the agricultural scene of the ADAM database (Rumania).

Index Terms—sub-pixelic classification, Maximum A Posteriori, high resolution images, coarse resolution time series, land cover.

I. INTRODUCTION

In the last decades, the use of remote sensing data has proved to be efficient for monitoring the Earth surface. In particular, land cover maps provide essential information not only for the analysis of global or local changes but also for studies on geosphere-biosphere-atmosphere interactions that depend on reliable estimation of the terrestrial vegetation. Indeed, vegetation has a significant impact on surface processes involved in water or energy exchanges. Land cover maps are increasingly used to define environmental policies and they contribute to reduce the risks from natural disasters (floods, forest fires, etc.). Now, given the large size of satellite data, automatic classification techniques are required to generate the land cover maps.

As time evolution is one of the most discriminating criteria for vegetation, the classification of a scene in terms of land cover types requires high temporal frequency information. Nowadays, we distinguish between sensors with a high spatial resolution (e.g. SPOT/HRV, 1 pixel for $20m \times 20m$) but a two-monthly temporal acquisition frequency, and sensors with a medium or coarse spatial resolution (e.g. MERIS, 1 pixel for $300m \times 300m$, or SPOT-VGT, 1 pixel for $1km^2$) but daily or so temporal acquisition frequency. The interest of medium or coarse resolution time series is clear taking into account the cost of high resolution data and the fact that high

resolution data may be corrupted by clouds at the acquisition date. Moreover, medium or coarse resolution provide spectral information dedicated to vegetation applications. One of the key challenge for automatic land cover classification is hence the combination of information from different spatial resolutions to benefit from both a high discrimination between land cover types and an accurate spatial information.

Using coarse resolution (CR) time series, the size of the field of view of a pixel can be bigger than the size of the objects of interest (e.g. agricultural fields). Such pixels are called mixed pixels. Hence, the identification of land cover types requires an access to sub-pixelic information. In the literature, numerous methods have already been proposed to deal with the sub-pixelic problem, such as data fusion, disaggregation or classification approaches.

Multi-resolution data fusion methods aim at reconstructing high resolution images from high resolution spatial information and coarse resolution spectral information. Various approaches have been suggested, in particular for the fusion of panchromatic and multispectral data. Most of them suggest to handle separately spectral and spatial information, through a multi-scale decomposition based on wavelet decomposition [Sveinsson et Benediktsson, 2000], [Ranchin et Wald, 2000] or on pyramidal decomposition [Aiazzi et al., 1999], [Laporterie-Déjean et al., 2003]. Performances of these methods are strongly limited when the considered resolution ratio increases: except the morphological approach of [Laporterie-Déjean et al., 2003] that enables to fuse data with a resolution ratio higher than 10, these methods are restricted to the use of resolution ratio coarser than 4. Moreover, the use of these kind of methods as a classification preprocessing may be used in a monotemporal context but, in a multitemporal context, it would lead to a high amount of intermediate data and, hence, increase considerably the computation time.

Disaggregation of mixed pixels has been widely used to improve coarse resolution data interpretation, relatively to the objects of interest. Most disaggregation technics used for sub-pixelic analysis are fuzzy approaches related to fuzzy c-means or possibilistic methods ([Settle et Drake, 1993], [Foody, 1996], [Brown et al., 2000]) enabling an estimation of the proportions of each land cover type represented within a given pixel. Indeed, the measured signal results of a mix of several spectral responses corresponding to the different types of cover represented within the considered CR pixel. Most of the time, the linear mixture model ([Horwitz et al., 1971]) is used for land cover analysis. It models the expected intensity acquired in a mixed pixel as the mean intensity characterizing

A. Robin and L. Moisan are with Paris Descartes University, laboratory of Applied Mathematics (MAP5), France, e-mail: {robin,moisan}@math-info.univ-paris5.fr

S. Le Hégarat-Mascle is with Paris-Sud 11 University, fundamental electronic institute (IEF), France, e-mail: le-hegar@ief.psud.fr

each type of land cover, weighted by its occupation rate within the CR pixel. This model is used either to estimate class features knowing class proportions within each pixel ([Faivre et Fischer, 1997]) or to estimate class proportions knowing class characteristics ([Cardot et al., 2003]). The class proportions describe the pixel composition but do not bring any information on the spatial distribution of the land cover types within a pixel. Hence, this kind of decomposition model enables to handle with high resolution ratios but not to obtain directly sub-pixelic classifications. Now, the main difficulty with medium or coarse resolution images such as MERIS or SPOT/VGT is the size of the pixel field of view relatively to the size of *objects* to classify (numerous parcels are entirely contained in one CR pixel).

More recently, sub-pixelic classification approaches have been developed in the aim both to discriminate the types of land cover and to refine their spatial location in order to lead to a fine resolution land cover map. Some methods suggest to subdivide each CR pixel into sub-pixels and to classify the sub-pixels maximizing their spatial auto-correlation ([Atkinson, 1997], [Atkinson, 2001], [Kasetkasem et al., 2005]). These methods require a preliminary knowledge of class proportions and they are based on a spatial dependency hypothesis between pixels and sub-pixels. Hence, as soon as the object of interest is fully included in a pixel, its spatial location is disabled. Other works use complementary information (a high resolution image or a structural model) in order to refine the classes location ([Tatem et al., 2002]), restricting the space of solutions. The use of all these approaches is restricted to small resolution ratios as the spatial dependency hypothesis does not bring enough information when high resolution ratios are used.

The approach we propose is based on Bayesian theory, using the Maximum A Posteriori (MAP) criterion, and assumes that the spatial structure of the studied scene is invariant along the considered time period. With this hypothesis, the spatial information of the scene can be extracted at a given date by, for instance, segmenting a HR image through homogeneous segments (in this study, we use a multiscale pyramidal algorithm for image segmentation [Koepfler et al., 1994], based on Mumford and Shah functional [Mumford et Shah, 1989]). Such an assumption seems realistic for a time period such as one agricultural year, in particular if the HR image is over-segmented, enabling to distinguish all parcels. However, this stationarity hypothesis can be restrictive when changes occur along the considered time period. Parallel studies focus on the problem of land cover change detection [Robin et al., 2007] and may be used jointly to classification for a complete vegetation monitoring automatic process.

Assuming a segmentation image is available, the classification task refers to a labeling problem using CR radiometric temporal information. Section II describes the problem formulation and the image model leading to the definition of an energy function. Then, a probabilistic error analysis of mislabeling using this energy function is presented in Section IV, before a description of the algorithm is given in Section V. Finally, Section VII shows some numerical results and Section VIII gathers our conclusions.

II. METHODOLOGY

A. Problem formulation

Assuming the spatial information is invariant along time, we suggest to investigate the classification problem by exploiting on the one hand the spatial information from HR images for the location of objects of interest and on the other hand the temporal information of CR time series to discriminate the type of land cover represented within each segment. To that aim, the classification problem is decomposed in two steps: the HR segmentation of the scene and the labeling of each segment according to its land cover. In this paper, we focus on the labeling problem, assuming a HR segmentation is available.

Let us denote \mathcal{D}_{HR} and \mathcal{D}_{CR} , respectively, the HR and CR image domains. A HR segmentation of the \mathcal{D}_{HR} image domain into $|\mathcal{S}|$ homogeneous segments is defined as an application $s : \mathcal{D}_{HR} \rightarrow \mathcal{S} = \{1, 2, \dots, |\mathcal{S}|\}$ that maps a segment index $s(x)$ to each pixel $x \in \mathcal{D}_{HR}$ (the set $\mathcal{S} = \{1, 2, \dots, |\mathcal{S}|\}$ denotes the obtained segment indexes). Now, if the scene of interest contains $|\mathcal{L}|$ types of land cover where $\mathcal{L} = \{1, \dots, |\mathcal{L}|\}$ denotes the set of possible labels, we define a labeling process as an application $\lambda : \mathcal{S} \rightarrow \mathcal{L}$, that maps a label λ_k to each segment k of the HR segmentation. In practice, the segmentation and labeling process also refer, respectively, to the segmentation image and the label map resulting from these applications.

Let $\mathcal{T} = \{1, \dots, |\mathcal{T}|\}$ be the set of acquisition dates, the CR time series is denoted as the image vector $v = (v_1, \dots, v_{|\mathcal{T}|})$ where, for all $t \in \mathcal{T}$, the image v_t is a real-valued function defined on the CR image domain \mathcal{D}_{CR} . In Section II-B, the relationship between the HR segmentation and the CR time series is described.

B. Pixel disaggregation

In this section, we describe the measurement (radiometric intensity, vegetation fraction cover, ...) which is observed within a CR pixel as a function of the class features that are represented within the pixel field of view. To that aim, let us assume that each HR pixel represents a unique class and that for all date t , a CR image v_t corresponds to the block average of the HR image u_t of the same scene. If \mathcal{D}_{HR} denotes the high resolution image domain, the coarse resolution intensity can be expressed as

$$v_t(y) = \frac{1}{N} \sum_{\{x \in y\}} u_t(x), \quad (1)$$

where $\{x \in y\}$ denotes the set of the HR pixels $x \in \mathcal{D}_{HR}$ that are enclosed by the CR pixel y and $N = |\mathcal{D}_{HR}|/|\mathcal{D}_{CR}|$ is the number of HR pixels that are included within a CR pixel (N is assumed to be integer). When the label map is known, notice that Equation (1) can be decomposed into

$$v_t(y) = \frac{1}{N} \sum_{l \in \mathcal{L}} \sum_{\substack{\{x \in y\} \\ \lambda_s(x) = l}} u_t(x), \quad (2)$$

where $\lambda_{s(x)}$ represents the label of the pixel x . The coarse resolution measurement can then be decomposed with respect to the labels. Let $N_l(y)$ be the number of HR pixels of label

l within the CR pixel y . The proportion of label l within the pixel y is then

$$\alpha_l(y) = \frac{N_l(y)}{N}, \quad (3)$$

with the condition that, for all $y \in \mathcal{D}_{\text{CR}}$,

$$\sum_{l \in \mathcal{L}} \alpha_l(y) = 1. \quad (4)$$

In general, CR image values result from a modulation transfer function which does not corresponds to a rectangle function as assumed by the bloc-average hypothesis. However, some previous studies show that, in practice, the linear mixture model is reasonable for numbers of applications. In particular, [Kerdiles et Grondona, 1995], [Small, 2001] showed that the use of a linear combination of NDVI values, even though they are not linear measurements, leads to minor errors.

Here, a probabilistic context is adopted in order to model the variability of the data. The HR and CR observations $u = (u_1, \dots, u_{|\mathcal{T}|})$ and $v = (v_1, \dots, v_{|\mathcal{T}|})$ are considered, respectively, as realizations of the label fields $U = (U_1, \dots, U_{|\mathcal{T}|})$ and $V = (V_1, \dots, V_{|\mathcal{T}|})$ and, for all $t \in \mathcal{T}$, $U_t = (U_t(y))_{y \in \mathcal{D}_{\text{HR}}}$ and $V_t = (V_t(y))_{y \in \mathcal{D}_{\text{CR}}}$. Moreover, a label map $\lambda = (\lambda_1, \dots, \lambda_{|\mathcal{S}|})$ is considered as a realization of a random field $\Lambda = (\Lambda_1, \dots, \Lambda_{|\mathcal{S}|})$. Then, the random variables $(V_t(y)|\Lambda = \lambda)$ satisfy, for all $t \in \mathcal{T}$ and $y \in \mathcal{D}_{\text{CR}}$,

$$(V_t(y)|\Lambda = \lambda) = \sum_{l \in \mathcal{L}} \alpha_l(y) \sum_{\substack{\{x \in y\} \\ \lambda_s(x) = l}} \frac{(U_t(x)|\Lambda = \lambda)}{N_l(y)} \quad (5)$$

This equation describes the observed measurement of a CR pixel as a function of the proportion of each label within the pixel of interest and the typical average intensity of each label.

In the literature, this model is usually applied to CR images, either to estimate unknown class proportions when an *a priori* knowledge of the class features is available ([Horwitz et al., 1971], [Settle et Drake, 1993]) or to estimate unknown class features knowing the class proportions ([Cross et al., 1991]). Some recent works enable to estimate both class proportions and features through an iterative algorithm such as EM or ICE ([Le Hégarat-Masclé et al., 2005]). In the following, the number of labels $|\mathcal{L}|$ is assumed to be known.

C. Maximum a posteriori

In this section, we derive the *a posteriori* probability of obtaining a label map knowing a CR time series observation and a HR segmentation. The maximization of this probability will lead to the reconstruction of a sub-pixelic label map.

Using the linear mixture model (Section II-B), the CR observed measurement can be expressed in function of the proportion of each label within the pixel of interest and of the corresponding class features. In order to map a label to each segment of the HR segmentation, let us consider the proportion of segment k within pixel y , denoted by $\beta_k(y)$. For all label $l \in \mathcal{L}$, the proportion of label l within pixel y is then equal

to the sum of proportions of all segments of label l that are present within pixel y , that is

$$\alpha_l(y) = \sum_{\substack{k \in \mathcal{S} \\ \lambda_k = l}} \beta_k(y), \quad (6)$$

where \mathcal{S} is the set of segments and λ_k is the label of segment k .

The probability of observing v knowing the label map λ describes the process of data observation and acquisition. Let us assume that, for all t , the measurement of label l is characterized, at high resolution, by a Gaussian distribution of mean $m_t(l)$ and $var_t(l)$ (*i.e.* for all $x \in \mathcal{D}_{\text{HR}}$, $U_t(x) \sim \mathcal{N}(m_t(l), var_t(l))$). This hypothesis is commonly used in the case of optical visible/infrared remote sensing data. We also assume that CR observations are spatially independent conditionnally to the label map. This hypothesis boils down to assuming that the CR variability is not spatially correlated. After computation (for details, see Appendix A), the random field V_t is then Gaussian conditionnally to the HR label map λ and the *a priori* probability density function of $V_t(y)$ conditionnally to $\Lambda = \lambda$ is

$$f_{V_t(y)|\Lambda}(v_t(y)|\lambda) = \frac{1}{\sigma_t(\lambda, y)\sqrt{2\pi}} e^{-\frac{(v_t(y) - \mu_t(\lambda, y))^2}{2\sigma_t^2(\lambda, y)}}, \quad (7)$$

i.e. a Gaussian density of mean

$$\mu_t(\lambda, y) = \sum_{l \in \mathcal{L}} \sum_{\substack{k \in \mathcal{S} \\ l_k = l}} \beta_k(y) m_t(l), \quad (8)$$

and variance

$$\sigma_t^2(\lambda, y) = \frac{1}{N} \sum_{l \in \mathcal{L}} \sum_{\substack{k \in \mathcal{S} \\ l_k = l}} \beta_k(y) var_t(l). \quad (9)$$

Let us underline that mean and variance depend on the considered CR pixel, through its composition in terms of the different types of land cover.

Moreover, let us assume temporal independency of observations conditionnally to the label map. This hypothesis is motivated by the fact that the estimation of cross-correlations between different dates is awkward (practically, it would require either a learning step or the definition of an *a priori* on the temporal evolution of each type of land cover). Note that from a theoretical point of view, it is not difficult to take into account temporal cross-correlations, through a non-diagonal covariance matrix. In the following, these hypotheses (normality, spatial and temporal independence) will be referred as hypotheses (H).

Under hypotheses (H), the probability density function of observing v is determined, conditionnally to the label map, by

$$f_{V|\Lambda=\lambda}(v|\lambda) = \prod_{t \in \mathcal{T}} \prod_{y \in \mathcal{D}_{\text{CR}}} f_{V_t(y)|\Lambda=\lambda}(v_t(y)|\lambda). \quad (10)$$

The mean of the random field V conditionnally to the label map is hence, for all pixel y , denoted $\mu(\lambda, y)$. It corresponds to the mean vector $(\mu_t(\lambda, y))_{y,t}$. Thanks to the temporal independency assumption, the covariance matrix is diagonal and can

be expressed as $\sigma^2(\lambda, y)I_{|\mathcal{T}|}$, where $\sigma^2(\lambda, y) = (\sigma_t^2(\lambda, y))_{y,t}$ and $I_{|\mathcal{T}|}$ stands for the identity matrix in dimension $|\mathcal{T}|$.

Considering the MAP criterion, let us estimate a label map as the optimal configuration of the label field Λ knowing the CR time series v and denote by $\lambda^* = (\lambda_1^*, \dots, \lambda_S^*)$ (where the k th coordinate stands for the label of the segment k) the optimal label map. In absence of contextual information related to the scene of interest, all label maps are assumed (as usual in such situation) to have the same probability of occurrence. From the Bayes rule, the *a posteriori* probability density function writes

$$f_{\Lambda|V=v}(\lambda|v) = \frac{f_{V|\Lambda}(v|\lambda) f_{\Lambda}(\lambda)}{f_V(v)}, \quad (11)$$

which has to be maximized with respect to λ . The density f_V does not depend on λ . Hence, without *a priori* information on label maps, all label configurations are considered as equiprobable. Then, the MAP boils down to the maximum likelihood, that is

$$\max_{\lambda \in \mathcal{L}^S} f_{\Lambda|V=v}(\lambda|v) = \max_{\lambda \in \mathcal{L}^S} f_{V|\Lambda=\lambda}(v|\lambda). \quad (12)$$

Under hypotheses (H), the optimal label map λ^* for MAP criterion is determined as the maximal argument over all label map configurations $\lambda \in \mathcal{L}^{|\mathcal{S}|}$ of $f_{V|\Lambda=\lambda}(v|\lambda)$, which is also

$$\operatorname{argmin}_{\lambda \in \mathcal{L}^S} \sum_{t \in \mathcal{T}} \sum_{y \in \mathcal{D}_{\text{CR}}} \left(\frac{(v_t(y) - \mu_t(\lambda, y))^2}{\sigma_t^2(\lambda, y)} + \ln(\sigma_t^2(\lambda, y)) \right). \quad (13)$$

Notice that the optimal label map is conditioned by the knowledge of mean and variance characterizing each type of land cover. In practice, when class means and variances are *a priori* known, a solution of Equation (13) can be obtained using an optimization process, leading to a supervised classification approach.

D. A cost function for unsupervised classification

In general, class features are unknown : they depend on various parameters such as acquisition dates, atmospheric conditions or the local evolution of vegetation (the phenological cycle of a given type of crop also varies with its geographic location). Therefore an unsupervised classification approach seems more appropriate. Now, class features can be estimated using the linear mixture model, assuming the knowledge of classes proportions that are represented within each pixel. Practically, mean estimation can be processed in a reasonable time during the iterative classification process (*e.g.* through a linear regression algorithm run for the label map configuration obtained at each step of the iterative algorithm). However, variance estimation, for which there is no so simple robust estimator, would induce a high computation time cost.

Consequently, for an unsupervised approach, we overcome the class variance estimation issue by assuming that class variances ($\text{var}_t(l)$) are equal to a fixed value σ^2 . This hypothesis (denoted (H_σ) in the following) implies that the variance of each CR pixel is constant, that is $\sigma_t(\lambda, y) = \sigma^2/N$ and, hence, that the energy considered in Problem (13) boils down to

$$\sum_{t \in \mathcal{T}} \sum_{y \in \mathcal{D}_{\text{CR}}} ((v_t(y) - \mu_t(\lambda, y))^2). \quad (14)$$

Moreover, as class features are unknown in the supervised case, we consider the problem

$$\min_{\lambda \in \mathcal{L}^S} \min_{m \in (\mathbb{R}^{\mathcal{L}})^{\mathcal{T}}} \sum_{t \in \mathcal{T}} \sum_{y \in \mathcal{D}_{\text{CR}}} (v_t(y) - \mu_t(\lambda, y))^2. \quad (15)$$

The maximum *a posteriori* label map is hence determined by

$$\lambda^* = \operatorname{argmin}_{\lambda \in \mathcal{L}^S} \min_{m \in (\mathbb{R}^{\mathcal{L}})^{\mathcal{T}}} \sum_{t \in \mathcal{T}} \sum_{y \in \mathcal{D}_{\text{CR}}} (v_t(y) - \mu_t(\lambda, y))^2, \quad (16)$$

and, for a given label map λ , solving the problem (16) enables to estimate the class means as a simple least squares problem. According to some preliminary statistical analysis performed on actual data (SPOT/HRV), this hypothesis seems appropriate as all class variances are rather close, except the 'water' class whose variance is particularly weak. Moreover, an over-estimation of 'water' class variance is not really impacting as it boils down to attribute a weaker weight to errors on this class, which usually concern very few HR pixels.

III. INFLUENCE OF THE RESOLUTION RATIO

In this section, we focus on the loss of information due to the use of CR data instead of HR data, considering the problem (15). To that aim, we consider the monotemporal case and, for the sake of simplicity, we adopt the following matrix notations : U and V denote, respectively, the vector of HR and CR data ($(u(x))_x$ and $(v(x))_x$), B denote the matrix of size $|\mathcal{D}_{\text{CR}}| \times |\mathcal{D}_{\text{HR}}|$ representing the block average operator which enables to obtain CR images from HR ones, R is a matrix of size $|\mathcal{D}_{\text{HR}}| \times |\mathcal{S}|$ representing the segmentation (whose (i, j) term is 1 if the pixel i belongs to segment j , 0 otherwise), C is a matrix of size $|\mathcal{S}| \times |\mathcal{L}|$ representing the label application (whose (i, j) term is 1 if the label of the segment i is j , 0 otherwise) and $M \in \mathbb{R}^{|\mathcal{L}|}$ denotes the class means vector. Using these notations, the problem (15) can be written as

$$\min_{C, M} \|B(RCM - U)\|^2. \quad (17)$$

Restricting to the use of HR data, the analogous problem without pixel disaggregation can be defined as

$$\min_{C, M} \|RCM - U\|^2. \quad (18)$$

First, let us remark that, for a given segmentation and label map (R and C), the solutions of the least squares problems (17) and (18) are unique if, respectively, the rank of BRC and the rank of RC are equal to $|\mathcal{L}|$. As the matrix RC represents the application that enables to reconstruct a HR image where each pixel value is the mean value corresponding to its label, the application RC is generally injective.

For a given label map, if M and $M' \in \mathbb{R}^{|\mathcal{L}|}$ are, respectively, the solutions of the least squares problems (17) and (18), they satisfy the following inequalities (for details, see Annex B):

$$\|M' - M\|^2 \leq \| (I - {}^tBB)U \|^2 \sum_{i=1}^L \frac{1}{d_i^2} \quad (19)$$

and

$$\|M' - M\|^2 \leq \left(1 - \frac{1}{N}\right)^2 \|U\|^2 \sum_{i=1}^L \frac{1}{d_i^2} \quad (20)$$

where $\|\cdot\|$ is the Euclidean vector norm and $\|\cdot\|$ its subordinated matrix norm, and $(d_i)_{i=1\dots L}$ denotes the set of the singular values of RC .

The inequality (19) shows, in particular, that if all CR pixels correspond to *pure* pixels (*i.e.* they represent a unique class), the mean solutions obtained using HR or CR data are equal ($M = M'$). Indeed, the image tBBU represents the HR image U reduced to coarse resolution (by block-average with blocks of size $N \times N$) and zoomed again (with factor $N \times N$).

Moreover, these upper bounds can be computed of a given data set (a HR image U , a resolution ratio N and a label map RC), permitting an estimation of the information gain using coarse resolution data. These upper bounds become coarse when the singular values of RC are too small, that is, when a label proportion can be predicted from the linear combination of the other labels.

IV. ERROR ANALYSIS AND MULTITEMPORAL ASPECTS

Before using the labeling cost function (16) in an experimental context, we focus on an analysis of labeling errors occurring while solving Problem (16). To that aim, we consider the probability of mislabeling a given segment under the hypotheses (H, H_σ) . First, let us recall that $v(y) = (v_t(y))_{t \in \mathcal{T}}$ denotes the CR observation vector in a pixel y , modeled as a realization of the random process $V(y) = (V_t(y))_{t \in \mathcal{T}}$, and $m(\lambda_k) = (m_t(\lambda_k))_{t \in \mathcal{T}}$ is the temporal mean vector corresponding to the label associated to segment k . In this framework, the energy function considered in Problem (16) associated to the random variable V is denoted

$$E_V(\lambda) = \sum_{y \in \mathcal{D}_{CR}} \|V(y) - \sum_{k \in \mathcal{S}} \beta_k(y) m(\lambda_k)\|^2, \quad (21)$$

where $\|\cdot\|$ is the Euclidian norm (over time).

Assume the groundtruth label map is $\lambda = (\lambda_1, \dots, \lambda_{|S|})$. In the following, a label map $\lambda' = (\lambda'_1, \dots, \lambda'_{|S|})$ will be said 1-optimal if it verifies

- $\forall k \neq k_0, \lambda'_k = \lambda_k$,
- $\lambda'_{k_0} \neq \lambda_{k_0}$,

where the segment k_0 is fully included in the CR pixel y_0 . Denote $\beta_{k_0}(y_0)$ the occupation rate of segment k_0 in the pixel y_0 . As V is a Gaussian random field characterized by Equation (7), the probability for a 1-optimal label map λ' to be preferred to the optimal label map is

$$p = \mathbb{P}(E_V(\lambda) - E_V(\lambda') \geq 0) = \Phi\left(-\frac{\sqrt{N}\|\delta(y_0)\|}{2\sigma}\right), \quad (22)$$

where

$$\|\delta(y_0)\| = \beta_{k_0}(y_0) \|m(\lambda_{k_0}) - m(\lambda'_{k_0})\|, \quad (23)$$

and

$$\forall x \in \mathbb{R}, \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \quad (24)$$

is the cumulative distribution function of normal law $\mathcal{N}(0, 1)$ (for computation details, see Appendix C).

This probability depends on the considered HR/CR resolution ratio N , on the proportion of the erroneous segment within

the CR pixel, on the resulting error on mean estimation and on the inter-classes standard deviation σ (assumed equal for each class).

More generally, if every segment of the segmentation image represents a proportion upper than n/N of a CR pixel, the previous result implies the following inequalities:

$$p \leq \Phi\left(-\frac{\sqrt{N}}{2} \frac{n}{N} \frac{\min_{\lambda \neq \lambda'} \|m(\lambda) - m(\lambda')\|}{\sigma}\right), \quad (25)$$

and

$$p \geq \Phi\left(-\frac{\sqrt{N}}{2} \frac{\max_{\lambda \neq \lambda'} \|m(\lambda) - m(\lambda')\|}{\sigma}\right). \quad (26)$$

This result gives a reliability interval. In particular, the inequality (25) raises a ratio which can be seen as a contrast measurement, denoted

$$c = \frac{\min_{\lambda \neq \lambda'} \|m(\lambda) - m(\lambda')\|}{\sigma}. \quad (27)$$

Figure 1 represents this upper bound values when the image contrast c varies, for segments of size at least $\frac{n}{N}$ (with different values of $\frac{n}{N}$, ranging from about 0.4% to 75% of a CR pixel). For the regular case of a contrast level upper than 1, a 1-optimal label map has a probability less than 0.05 to be preferred to the correct label map when any segment represents more than 25% of a CR pixel. Moreover, the higher the contrast is, the weaker is the probability of mislabeling a segment contained in a CR pixel. Notice that

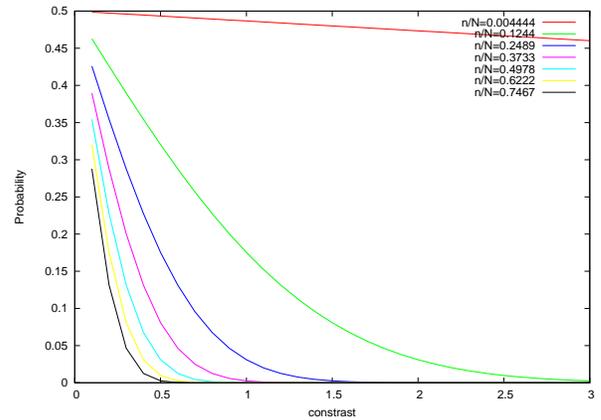


Fig. 1. Upper bound of the probability to prefer a 1-optimal label map to the correct label map when the mislabeled segment occupies more than $n/N\%$ of a CR pixel. The upper bound value is plotted in function of the image contrast. Each curve corresponds to a given proportion n/N of a CR pixel, ranging from 0.4% to 74.6%.

even though the mislabeling probability has been computed in the restricted case of a unique mislabeled segment, it raises the key parameters for labeling performance. In particular, the upper bound puts forward the two key characteristics that involve both the remote sensing sensor and the physical scene features, namely the ratio between segment and CR pixel sizes and the minimum distance between class means. An empirical analysis of sensitivity of the method to the resolution ratio is developed in Section VI-B, showing that the resolution ratio

impacts the performance only through the aforementioned parameter (ratio between segment and CR pixels).

As far as the minimum distance between classes is concerned, it can be used as a guide for the choice of a subset of dates in the time series. Assuming that we would handle only θ dates in the set of acquisition dates \mathcal{T} (e.g. for computational time reasons or for supervised class features robustness in a supervised approach), rather than choosing these dates from phenological cycle knowledge considerations, we suggest to choose θ dates in \mathcal{T} solving the following problem:

$$\max_{t_1 < \dots < t_\theta} \min_{\lambda \neq \lambda'} \sum_{i=1}^{\theta} |m_{t_i}(\lambda) - m_{t_i}(\lambda')|^2. \quad (28)$$

This criterion can be interpreted as a contrast measurement of the time series, joining up the idea of choosing the dates leading to the highest level of contrast (or class separability). Moreover, solving the problem (28) can be justified by the fact it enables to minimize labeling errors.

Note that the use of this criterion would be mainly attractive in a supervised case, where the class means are *a priori* known. Indeed, a solution of Problem (28) can be obtained exploring all sub-sets of θ dates contained in \mathcal{T} but this implies an exploration of $\binom{|\mathcal{T}|}{\theta}$ possibilities which rapidly requires a very costly computation time. In the unsupervised case, since the class means are *a priori* unknown, dates could be chosen *a posteriori* using (28) in order to refine the label map but that takes off most interest of the procedure. In this study, the choice of the dates was done according to an *a priori* phenological knowledge and on image availability and quality (mainly absence of clouds, high contrast).

V. ALGORITHM

Because of the size of the solution space, a systematic search of the minimum is not possible. As no heuristic seems justified for the considered problem, we decide to adopt a simulated annealing algorithm. The minimization problem with consider is composed of two correlated problems: a least squares problem which can be easily solved through a singular value decomposition and a combinatory optimization problem. For this latter, an exhaustive research is not possible hence we use a simulated annealing algorithm, which is classical for combinatory optimization problems. It has been introduced in image processing by [Geman et Geman, 1984] and used by [Le Hégarat-Masclé et al., 1996] for various applications.

This algorithm is based on an analogy between thermodynamics behaviour of solids and large combinatorial optimisation problems. Applied to minimization, simulated annealing can be seen as a general optimization process enabling to jump over energy barriers separating local minima. It allows to reach, (theoretically) the global optimum in a solution space having local optima. In our case, the presence of local minima has been confirmed in experiments, for instance, observing that for a given label configuration, changing only one segment increases the global energy whereas changing simultaneously several segments allows energy reduction. As far as the unsupervised approach is concerned, the algorithm takes as inputs the HR segmentation, the CR time series and

the number of labels. It returns the label field solution of (16) and the class features (means). In the following, we denote E_l the global energy to minimize in (16). It stands for the energy corresponding to a label field λ . Each step of the algorithm changes randomly one segment label in the label field and tests whether it makes the energy decreasing or not. Denote $E_v(\lambda_{prev})$ the energy corresponding to the previous label field, for a given time series v . The algorithm is the following.

- Compute proportions $(\beta_k(y))_{y,k}$ for all pixel y .
- Initialize randomly the label field.
- Initialize the temperature T to the graph diameter.
- While stop criterion is not verified, do
 - for $i = 0$ to $|\mathcal{S}|$
 - draw randomly a segment k ,
 - draw randomly a label c for segment k ,
 - estimate the label means,
 - compute $\Delta E_v = E_v(\lambda) - E_v(\lambda_{prev})$
 - if $\Delta E_v \leq 0$, accept the label change,
 - else reject it with probability $\exp(-\Delta E_v/T)$
 - decrease the temperature : $T \leftarrow T(n)$
 - $n \leftarrow n + 1$

The function $T(n)$ is called the cooling schedule, its role is to ensure the convergence of the algorithm to the global optimum of the energy E_v . Theroretically, the cooling schedule should have a logarithmic decreasing law such as $T(n) = h/\log(n)$, where h is a predefined constant. However, with such a law, the convergence is very slow and practically, a geometrical decreasing like $T(n) = Tq^n$ is often used. In our case, setting $q = 0.999$ provided good results in the experiments. Concerning the stop criterion, we stop the algorithm when no change is accepted during $n_r \times |\mathcal{S}|$ successive iterations (with $n_r = 400$). Once more, this value results from an empirical analysis of the stability of the obtained solution.

This algorithm can also be used in order to find the label field solution of Problem (13), taking a HR segmentation, a CR time series and the class features (means and variances) as inputs by simply removing the label means estimation step. Note that such a supervised version is much faster than the unsupervised one as each iteration does not require a time-consuming linear regression anymore.

VI. VALIDATION

In this section, the performance of the method is analysed both in a supervised and in an unsupervised context. The labeling method is, at first, assessed using simulated CR images in order to evaluate the method independently of the validity of data assumptions (e.g. Gaussian hypothesis) and on the preprocessing quality (e.g. image registration). However, these images have been realized using actual data features. Practically, HR images have been simulated using actual SPOT/HRV resolution characteristics and radiometric means and variances. From an actual HR time series of Danubian plain, 165 images of size 300×300 pixels have been extracted, segmented in 100 segments and finally randomly labeled using 5 classes. HR images have been simulated by random drawing of law $\mathcal{N}(m_t(l), var_t(l))$ where l is given by the corresponding 'groundtruth' label and parameters m_t

and var_t have been learnt from an actual data set. Secondly, CR images have been created averaging HR simulated images. The reference classification is the label map from which the HR images have been simulated. In this section, as we focus on the method and the algorithm behavior, experiments have been run using single CR images rather than CR time series. Global results are given in Table I for both supervised and unsupervised approaches. As far as the supervised version is concerned, results are very satisfactory since the average percentage of mislabeled pixels is less than 1%. The percentage of mislabeled segments may seem high but it is due to the presence of very small regions in the segmentation (as the percentage of pixels concerned remains very small). In fact, we made the choice not to constrain the segmentation so far. A minimum segment size may be imposed in the segmentation processing step, permitting to avoid that kind of artefacts. In the unsupervised case, the results are rather convincing since the average percentage of misclassified pixels is less than 5%, and the median percentage is 0.2%. This significant difference observed between median and average can be explained by the fact that the information contained in some simulated images is too poor for the unsupervised classification method to work (and CR images are very small, with 20×20 pixels), in addition to the fact that these tests were restricted to the use of monotemporal images.

TABLE I
AVERAGE RESULTS OBTAINED FOR A SET OF 165 SIMULATED IMAGES
WITH A RESOLUTION RATIO $|\mathcal{D}_{HR}|/|\mathcal{D}_{CR}| = 15 \times 15$.

	Mislabeled pixels (average %)	Mislabeled segments (average %)
Supervised version	0.87	23.6
Unsupervised version	4.35	31.5

We analyze more precisely the detail of these global errors in the next section.

A. Experimental error analysis

In this section, we aim at analyse the obtained errors in order to specify experimentally the cause of their occurrence : non-convergence of the algorithm (due to a too fast schedule), lack of information in the data, overlapping of the class features, etc. To that aim, the energy value E_0 corresponding to the reference label map is compared to the energy value E_k associated to a label configuration that is equal to the reference one except on the segment numbered k . This comparison is made successively for all erroneous segment k obtained by the algorithm. Such an experiment has been performed in order not to bias the analysis by the presence of errors concerning other segments.

The histogram of the gains due to local perturbations in the neighborhood of the reference label map is represented Figure 2. More precisely, it represents the energy differences histogram computed for all label configurations containing exactly one mislabeled segment among all erroneous ones we obtained. This histogram shows, in particular, that more than

one half of the erroneous label provide an improvement from the reference for the criterion E .

It seems that most errors are due to the stochastic nature of the problem, *i.e.* the accurate optimum leads to a label map that is often quite different from the reference classification. The positive difference errors (Figure 2) result mainly from two types of phenomenon. On the one hand, we observed that little segments may be misclassified if their occupation rate within a CR pixel is too low to have an effective contribution on the energy. In some cases, a little segment may be misclassified in order to compensate for the signed error due to the contribution of the other (well classified) segments of the CR pixel. On the other hand, as Gaussian distributions we consider partially overlap, an observation occurring in the distribution queue of its class may have a higher probability to belong to the adjacent class and, hence, to be misclassified. The use of multitemporal data, as performed in the actual case, may reduce this type of errors as classes become more severable with time series.

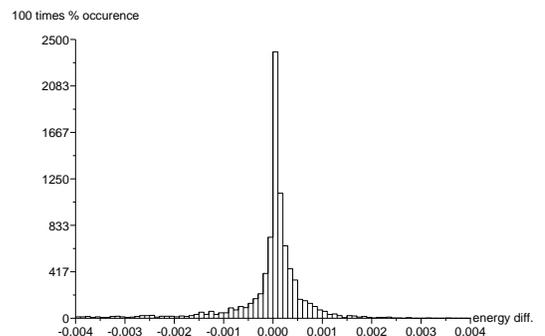


Fig. 2. Histogram of normalized energy variation $(E_0 - E_k)/|E_0|$ computed for all errors obtained using the supervised algorithm.

B. Sensitivity to resolution ratio

Since we aim at using high resolution ratios going up to 15^2 (*e.g.* for SPOT/HRV versus MERIS) or 50^2 (*e.g.* for SPOT/HRV versus SPOT/VGT), we now study the performance of the method when the resolution ratio varies. As Table I has shown that tiny segments (containing one or few HR pixels) tended to be misclassified, this study may also give some hints for the minimal acceptable segment size.

From a set of 165 simulated HR images, CR images have been created for different resolution ratios (15^2 , 30^2 and 50^2). For each mislabeled segment obtained using the supervised approach, the occupation rate within a CR pixel is computed. Table II presents some occupation rate statistics computed on the set of error segments. As we can see, the size of erroneous segments relatively to CR pixel size is almost independent from the resolution ratio. Moreover, about 70% of the error segments occupies less than 5% of a CR pixel, hence restricting to larger segments could increase strikingly the performance of the method. We also observe that 50% of label errors concern segments occupying less than 2% of the CR pixel. This observation could be used to constrain the segmentation process with a minimum segment size of 2% of the

CR pixel, which would reduce by about one half the labeling errors. In the literature, many methods using CR data for class proportion estimation deal with segments representing at least 5% of the CR pixel [Le Hégarat-Masclé et al., 2005].

TABLE II
STATISTICS ON THE OCCUPATION RATE (PERCENTILES) OF A MISLABELED SEGMENT WITHIN A CR PIXEL IN FUNCTION OF THE CONSIDERED RESOLUTION RATIO (N).

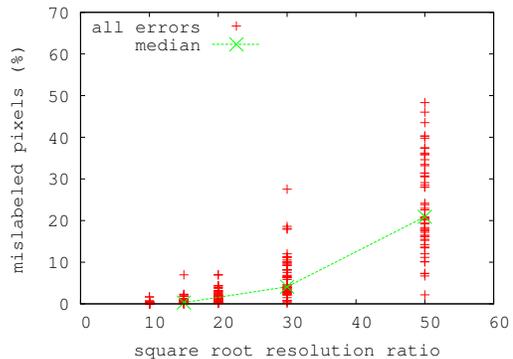
N	10 th ptl	50 th ptl	70 th ptl	90 th ptl	100 th ptl
15^2	0.44%	2%	5%	25.7%	100%
30^2	0.11%	1.77%	7.7%	41.4%	100%
50^2	0.08%	2%	8.13%	39.8%	100%

As far as the performance of the method versus the resolution ratio is concerned, Figure 3 shows that the error percentage values can reach rather high values. This can be attributed to the fact these experiments use monotemporal images while some class features may be very close depending on the considered date, leading to the observed quality variability of the results. Moreover, the median curve shows very good average performance for resolution ratios up to 30^2 , then performance decreases with resolution ratio 50^2 with about 20% of erroneous pixels. The dependency of the error percentage on the sensor resolution is manifest and non-linear. Notice that contrary to the previous case where the segment size was considered relatively to the CR pixel size, the error is here measured in terms of number of mislabeled HR pixels (without taking the size of objects relatively to the pixel size into account). Consequently, the decreasing performance of the results when N increases is essentially due to the fact that the minimal relative size of segments decreases (most errors are caused by the smallest segments).

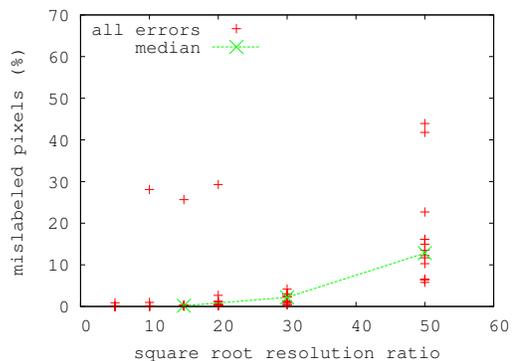
The coarser is the sensor resolution, the more multitemporal information is necessary to attain a good performance. Note also that the results we obtain using the supervised algorithm are very close in performance to those obtained using the unsupervised version.

Last, let us analyse the method performance in function of the resolution ratio in a simple case with pseudo-actual data and a monotemporal supervised context. To that aim, the 165 subparts of actual SPOT/HRV images of the ADAM database have been averaged by blocks of size 5×5 to 50×50 in order to build pseudo-actual CR images. Figure 4 represents, as in the case of simulated data (Figure 3), the percentage of mislabeled pixels as a function of the considered resolution ratio. As we can see, errors are significantly higher for pseudo-actual data: for instance, the median error for a 15×15 resolution ratio goes from 0.3% to 4.5%. This important difference may be attributed to the fact that, in the case of monotemporal actual images, the Gaussian hypothesis is not well justified and, moreover, the spectral response of some types of vegetation may not be spatially homogeneous (for instance, the wheat represented within a segment x or within a segment y may differently respond because of the irrigation, of the nature of the soil, ...)

Besides, these experiments runned on 165 pseudo-actual



(a) Supervised version



(b) Unsupervised version

Fig. 3. Each cross represents the percentage of misclassified pixels obtained for several simulated images and a given resolution ratio, using the supervised version (a) and the unsupervised version (b). The superimposed line represents the median of these percentage values.

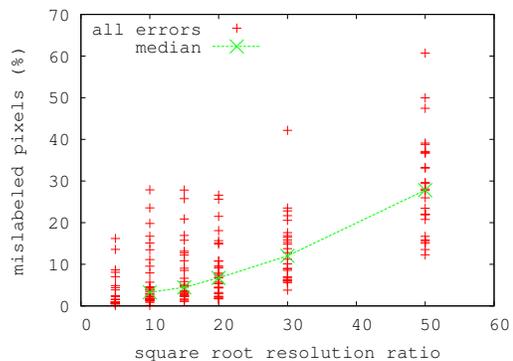


Fig. 4. Percentage of mislabeled pixels obtained by applying the supervised algorithm to monotemporal actual images as a function of the resolution ratio, and its median value.

images using the supervised labeling method show that, for a given HR segmentation, the label map quality is limited when the considered resolution ratio exceed 30×30 . In particular, for a ratio $N = 50 \times 50$, at least 30% of HR pixels are mislabeled for one half of the experiments. In the following, we will hence focus on lower resolution ratios, typically 15×15 as in the case of a joint use of SPOT/HRV and MERIS data.

VII. RESULTS

A. The ADAM database

The proposed approach has been applied to a subpart of an actual SPOT/HRV time series acquired over an agricultural site of the Danubian plain (Rumania) provided by the French spatial agency (CNES) in the framework of the ADAM European project. In this context, SPOT/HRV multispectral images have been specifically acquired about once every 2 weeks using satellites SPOT 1, 2 and 4. This constitutes an exceptional database even if the presence of clouds and some image quality issues disabled the use of some image parts.

A time series of 8 multispectral SPOT/HRV images has been selected from the whole data set (acquisition dates: 11/14/2000, 03/14/2001, 04/12/2001, 05/24/2001, 05/31/2001, 06/13/2001, 07/24/2001 and 07/31/2001), and pre-processed into vegetation fraction cover series using the SAIL+PROSPECT biophysical model [Jacquemoud et al., 1995], as this parameter is scale linear and known to be efficient for land cover discrimination. In section VI, the analysis concerning the sensitivity of the method to the resolution ratio showed good performance for resolution ratios less than 20×20 . Hence, as MERIS series were not available at the time, MERIS data (Figure 5 (e)) have been simulated from the actual SPOT/HRV time series (Figure 5 (b)), by averaging each image with a factor 16 in each direction. This approach enables to reproduce the intensity variability and distribution from actual data while discarding the registration issues which are beyond the scope of this study.

The HR segmentation that we used results from the multiscale algorithm [Koepfler et al., 1994]¹ processed on a SPOT/HRV image for a given number of segments (*e.g.* in Figure 5 (a), $|\mathcal{S}| = 100$). In the following, the results obtained using the unsupervised algorithm are presented for a given number of classes and for a fixed resolution ratio of 16×16 .

B. Accuracy of the classification

In this section, we suggest to measure the accuracy of a label map by comparing the result obtained using CR time series to the result obtained using the corresponding HR time series and the same energy function (be aware that pixels are then pure pixels, *i.e.* only 1 class per pixel). This performance measurement is based on the idea that an ideal sub-pixelic classification should be able to discriminate classes from CR data *as well as* from HR data.

In Figure 5, classifications (c) and (f) have been obtained from the HR segmentation (a) and, respectively, the HR and

CR time series (b) and (e) (only the first image of the time series is shown) using the unsupervised classification algorithm for 5 classes. Label errors between HR and CR classifications (c) and (f) are represented in (d), in white, showing about 3% of mislabeled HR pixels. Let us remark that some segments of size larger than the CR pixel size are mislabeled. This can be explained by the fact that the average intensity of this area takes a value intermediate between the two class means and, hence, is ambiguous. Another output of the method

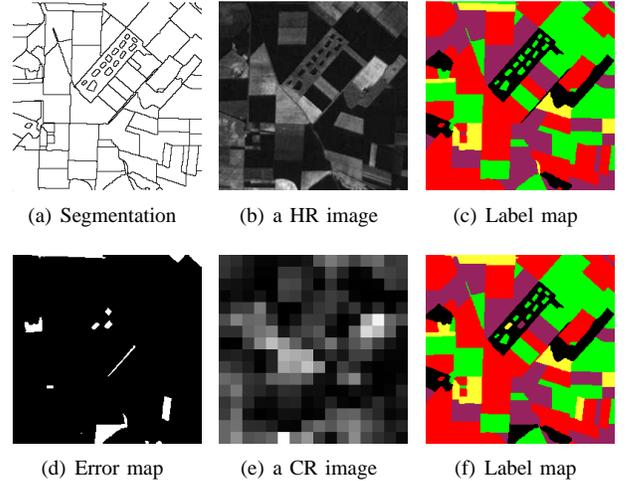


Fig. 5. The label map (c) has been obtained using the segmentation (a) and an actual time series of 8 vegetation fraction cover HR images (b), for 5 labels. The label map (f) represents the result obtained using the segmentation (a) and CR time series (e) (simulated from HR ones, with a resolution ratio 16×16). Image (d) shows the label differences between (c) and (f): only about 3% of the pixels are mislabeled (in white).

is the estimation of the class means. Figure 6 presents the vegetation fraction cover mean estimation corresponding to each label obtained from the same HR and CR time series. We can see that these estimated profiles are very close, which shows the accuracy of the unsupervised means estimation from CR data. A similar experiment, performed on another dataset,

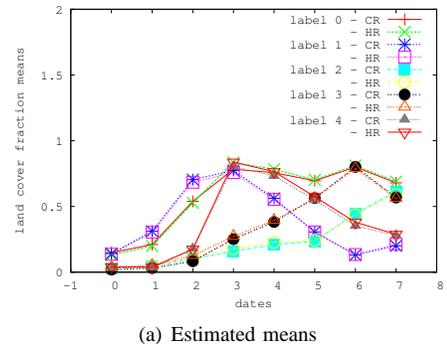


Fig. 6. For each class, corresponding vegetation fraction cover means are plotted in function of the acquisition date (estimation given by the unsupervised algorithm for 5 labels, using the data set Figure 5). For each class, two curves are plotted : one for the mean estimation using HR images and the other using CR images.

leads to the same conclusions (see Figure 7).

¹function `seget` of MegaWave2 image processing software

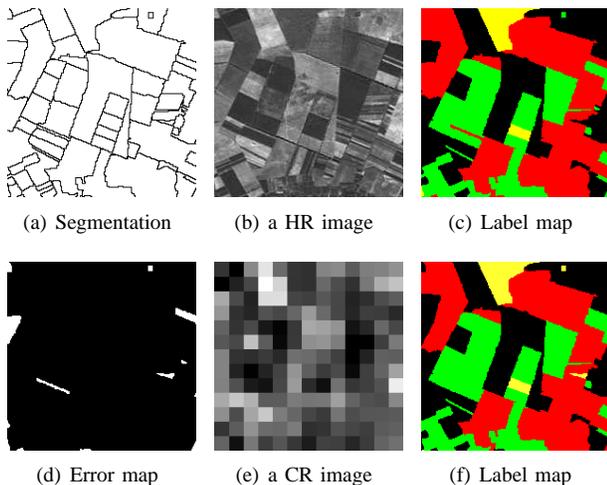


Fig. 7. Same experiment as Figure 5, performed on another dataset. Only about 2% of the pixels are mislabeled (in white on the error map (d)).

C. Robustness of the method

In order to estimate the robustness of the method with respect to the kind of data considered, the same experiment as in Figure 5 has been performed using NDVI images corresponding to the same dates. Let us remind that NDVI images are obtained directly from the red (R) and near infrared (NIR) bands of SPOT sensor through the relationship

$$NDVI = \frac{NIR - R}{NIR + R}, \quad (29)$$

whereas the vegetation fraction cover images are obtained from the inversion of a biophysical model. Using HR segmentation (Figure 5 (a)) and the NDVI CR time series (Figure 8 (a)), the unsupervised approach leads for 5 labels to the label map (Figure 5 (b)) which is very close to the label map previously obtained using vegetation fraction cover images. The error map (Figure 5 (c)) shows the mislabeled pixels, using the label map Figure 5 (c) (obtained using a HR time series of vegetation fraction cover) as a reference. The result is very close to the one obtained using CR time series of vegetation fraction cover. This remark confirms the

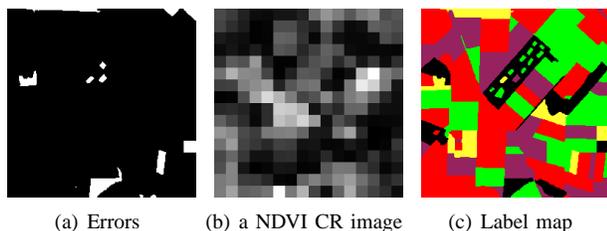


Fig. 8. Label map obtained using NDVI images: the label map (c) has been obtained using the segmentation 5 (a) and CR NDVI time series (b). Comparing the error map (a) to the one obtained from vegetation fraction cover (Figure 5 (d)) shows that the obtained results are very similar.

robustness of the method according to the type of data used for classification.

As we noticed earlier, the quality of the HR segmentation has a direct impact on the label map quality, as non-segmented

connected parcels are necessarily identically labelled even if their land cover is different. Some labeling errors can thus be explained by the fact that the segmentation was not fine enough, but they can also result from the lack of separability of the radiometric intensity measures for two types of land cover. Hence, an over-segmentation is generally preferable to an under-segmentation, in order to separate all parcels. In fact, a compromise has to be found since over-segmentations increase the dimension of the search space, yielding larger computation times and more possibilities of erroneous classifications.

If the obtained results seem satisfactory and reliable, they raise the difficulty of setting an *a priori* number of classes for the classification ($|\mathcal{L}|$). Indeed, in an unsupervised context, $|\mathcal{L}|$ is usually not known and, even if it were, there is no reason for it to correspond to the number of classes that can be distinguished from the CR time series. In the previous experiments, $|\mathcal{L}|$ was set arbitrarily but additional experiments showed some consistency between classifications of a same scene using different values of $|\mathcal{L}|$: increasing $|\mathcal{L}|$ rather led to splitting classes, raising some hierarchy property of the labeling process.

VIII. CONCLUSION

In this paper, a new approach enabling HR classification using a CR time series and additional structural information has been presented. It is based on a Bayesian model of the problem and on the linear mixture model for CR pixel disaggregation in terms of the characteristics of the different segments that are represented within a CR pixel. The Maximum A Posteriori criterion enables the definition of an energy function whose minimization leads to the desired classification.

A probabilistic analysis of the classification accuracy showed the crucial role of the temporal evolution of class features, and raised a criterion for optimizing the choice of a subset of acquisition dates. This criterion suggests that even if multitemporal data permit to improve the label map accuracy, the acquisition dates should be well chosen in order to find a compromise between the actual information gained and the increasing dimensionality of the search space. This idea, combined with experimental validations, could lead to interesting developments of the proposed approach.

A simulated annealing algorithm has been adopted to minimize the energy and provide a label map from a HR segmentation and a CR time series in an unsupervised context. An empirical analysis of the obtained labeling errors using simulated data showed that the algorithm convergence (delicate issue for this type of method) does not seem problematic, even though a geometrical temperature schedule has been considered.

The same experiments showed that the unsupervised approach remains reliable for resolution ratios less than 30×30 . Hence, the joint use of data such as SPOT/HRV and VGT (resolution ratio of 50×50) for labeling may not be expected as satisfying for the type of landscape we considered (typical European landscape). The joint use of SPOT/HRV and MERIS images, however, seems well suited to the HR/CR classification we presented, since their relative resolution ratio is (only) 15×15 . Experiments using actual SPOT/HRV

images and MERIS pseudo-actual images in the context of an agricultural application in the Danubian Plain (ADAM database) yielded satisfactory results according to HR data and available groundtruth elements. Moreover, some preliminary experiments showed the stability of the results through cross-validation with different kinds of data: in particular, very close label maps have been obtained using vegetation fraction cover CR images and NDVI CR images.

Finally, let us remark that the applicability of this method to actual CR time series will highly depend on the accuracy of the registration away CR images, and between CR images and the HR segmentation. The bounds on the resolution ratio we pointed out while assuming perfect registrations hence should be considered as upper limits, that would require some adjustments to take into account the limited registration accuracy.

APPENDIX A A PRIORI PROBABILITY

Using Equation (5) and assuming that, for all t , the random field U_t defined on the HR domain \mathcal{D}_{HR} is Gaussian conditionally to the HR label map, with mean $m_t(l)$ and variance $\text{var}_t(l)$, the variable $(V_t(y)|\Lambda = \lambda)$ is then Gaussian as a mixture of Gaussian variables. Its mean is, for any CR pixel y and any date t ,

$$\begin{aligned} & \mathbb{E}[V_t(y)|\Lambda = \lambda] \\ &= \mathbb{E}\left[\sum_{l \in \mathcal{L}} \alpha_l(y) \sum_{\substack{\{x \in y\} \\ \lambda_{s(x)} = l}} \frac{(U_t(x)|\Lambda = \lambda)}{N_l(y)}\right] \\ &= \sum_{l \in \mathcal{L}} \alpha_l(y) \sum_{\substack{\{x \in y\} \\ \lambda_{s(x)} = l}} \frac{m_t(l)}{N_l(y)} \\ &= \sum_{l \in \mathcal{L}} \alpha_l(y) m_t(l), \end{aligned} \quad (30)$$

as $N_l(y) = |\{x \in y, \lambda_{s(x)} = l\}|$. Finally, using equation (6),

$$\mathbb{E}[V_t(y)|\Lambda = \lambda] = \sum_{l \in \mathcal{L}} \sum_{\substack{k \in \mathcal{S} \\ l_k = l}} \beta_k(y) m_t(l). \quad (31)$$

The variance is also computed from equation (5) and, thanks to the variables $(U_t(x))_{x \in \mathcal{D}_{\text{HR}}}$ independence hypothesis, we have, for any date t and pixel y ,

$$\begin{aligned} \sigma_t^2(\lambda, y) &= \sum_{l \in \mathcal{L}} \alpha_l^2(y) N_l(y) \left(\frac{1}{N_l(y)}\right)^2 \text{var}_t(l) \\ &= \frac{1}{N} \sum_{l \in \mathcal{L}} \alpha_l(y) \text{var}_t(l) \\ &= \frac{1}{N} \sum_{l \in \mathcal{L}} \sum_{\substack{k \in \mathcal{S} \\ l_k = l}} \beta_k(y) \text{var}_t(l). \end{aligned} \quad (32)$$

APPENDIX B INFLUENCE OF THE RESOLUTION RATIO

The matrix RC can be decomposed in singular values as

$$RC = O D {}^t W. \quad (33)$$

where O is an orthogonal matrix of size $|\mathcal{D}_{\text{HR}}| \times |\mathcal{L}|$, W is an orthogonal matrix of size $|\mathcal{L}| \times |\mathcal{L}|$ and $D = (d_i)_i$ is a diagonal matrix of size $|\mathcal{L}| \times |\mathcal{L}|$ containing the singular values of RC on the diagonal. As the rank of RC is $|\mathcal{L}|$, all singular values of RC are non-null. Let us denote $O = (O_1, \dots, O_{|\mathcal{L}|})$ and $W = (W_1, \dots, W_{|\mathcal{L}|})$ where for $i = 1 \dots |\mathcal{L}|$, the terms O_i and W_i are, respectively, the column vectors of the matrices O and W . Hence, for all $X \in \mathbb{R}^{|\mathcal{L}|}$,

$$RCX = \sum_{i=1}^{|\mathcal{L}|} d_i \langle X, W_i \rangle O_i. \quad (34)$$

Consequently, for a given matrix C , the solution of the problem (17) is defined as

$$M = \sum_{i=1}^L \frac{1}{d_i} \langle {}^t BBU, O_i \rangle W_i. \quad (35)$$

and the solution of the problem (18) is

$$M' = \sum_{i=1}^L \frac{1}{d_i} \langle U, O_i \rangle W_i. \quad (36)$$

Hence,

$$\begin{aligned} \|M' - M\|^2 &= \left\| \sum_{i=1}^L \frac{1}{d_i} \langle U - {}^t BBU, O_i \rangle W_i \right\|^2 \\ &\leq \sum_{i=1}^L \frac{1}{d_i^2} \|(I - {}^t BB)U\|^2 \\ &\leq \sum_{i=1}^L \frac{1}{d_i^2} \|I - {}^t BB\|^2 \|U\|^2. \end{aligned} \quad (37)$$

(38)

Moreover, the matrix ${}^t BB$, with size $|\mathcal{D}_{\text{HR}}| \times |\mathcal{D}_{\text{HR}}|$, can be written as $\frac{1}{N^2} J_N$ where $J_N = \begin{pmatrix} 1 & \dots & 1 \\ \vdots & & \vdots \\ 1 & \dots & 1 \end{pmatrix}$ and N is the resolution ratio ($|\mathcal{D}_{\text{HR}}|/|\mathcal{D}_{\text{CR}}|$). Hence,

$$\|I - {}^t BB\| = 1 - \frac{1}{N^2} + \frac{N-1}{N^2} = 1 - \frac{1}{N}$$

and

$$\|M' - M\|^2 \leq \left(1 - \frac{1}{N}\right)^2 \|U\|^2 \sum_{i=1}^L \frac{1}{d_i^2}. \quad (39)$$

APPENDIX C ERROR ANALYSIS

The energy difference between the correct label map λ and a 1-optimal label map λ' is null everywhere but on the segment k_0 , which is fully represented within the pixel y_0 . Hence

$$\begin{aligned} E(\lambda) - E(\lambda') &= \|V(y_0) - \sum_{k \in \mathcal{S}} \beta_k(y_0) m(\lambda_k)\|^2 \\ &\quad - \|V(y_0) - \sum_{k \in \mathcal{S}} \beta_k(y_0) m(\lambda_k) + \delta(y_0)\|^2, \end{aligned}$$

where $\delta(y_0) = (\delta_t(y_0))_{t \in \mathcal{T}}$ and, for all $t \in \mathcal{T}$,

$$\delta_t(y_0) = \beta_{k_0}(y_0) (m_t(\lambda_{k_0}) - m_t(\lambda'_{k_0})). \quad (40)$$

For all date $t \in \mathcal{T}$ and for all pixel $y \in \mathcal{D}_{\text{CR}}$, let us define the random variable

$$\bar{V}_t(y) = V_t(y) - \sum_{k \in \mathcal{S}} \beta_k(y) m_t(\lambda_k). \quad (41)$$

It is Gaussian as a linear combination of the Gaussian random variables $V_t(y)$, with mean

$$\begin{aligned} \mathbb{E} [\bar{V}_t(y)] &= \mathbb{E} \left[V_t(y) - \sum_{k \in \mathcal{S}} \beta_k(y) m_t(\lambda_k) \right] \\ &= \mathbb{E} [V_t(y)] - \sum_{k \in \mathcal{S}} \beta_k(y) m_t(\lambda_k) \\ &= 0 \text{ using Equation (8),} \end{aligned} \quad (42)$$

and variance

$$\begin{aligned} \mathbb{V} [\bar{V}_t(y)] &= \mathbb{V} \left[V_t(y) - \sum_{k \in \mathcal{S}} \beta_k(y) m_t(\lambda_k) \right] \\ &= \mathbb{V} [V_t(y)] \\ &= \frac{\sigma^2}{N}. \end{aligned} \quad (43)$$

Consequently, the random vector $\bar{V}(y) = (\bar{V}_t(y))_{t \in \mathcal{T}}$ is a Gaussian central vector with covariance matrix $\frac{\sigma^2}{N} I_T$, where I_T represents the identity matrix in dimension T (under hypotheses of dates independence and class variance equality). With these notations, the energy difference between the label maps λ and λ' is

$$E(\lambda) - E(\lambda') = \|\bar{V}(y_0)\|^2 - \|\bar{V}(y_0) + \delta(y_0)\|^2. \quad (44)$$

As $\delta(y_0)$ is constant, the random vector $\bar{V}(y_0) + \delta(y_0)$ is Gaussian, with mean $\delta(y_0)$ and covariance matrix $\frac{\sigma^2}{N} I_T$. The probability for a given 1-optimal label map to be preferred to the correct label map is hence

$$\begin{aligned} &\mathbb{P}(E(\lambda) - E(\lambda') \geq 0) \\ &= \mathbb{P}\left(\|\bar{V}(y_0)\|^2 - \|\bar{V}(y_0) + \delta(y_0)\|^2 \geq 0\right) \\ &= \mathbb{P}\left(\frac{2 \langle \bar{V}(y_0), \delta(y_0) \rangle}{\|\delta(y_0)\|^2} \leq -1\right). \end{aligned} \quad (45)$$

Now,

$$\mathbb{E} [\langle \bar{V}(y_0), \delta(y_0) \rangle] = \sum_{t \in \mathcal{T}} \delta^t(y_0) \mathbb{E} [\bar{V}^t(y_0)] = 0 \quad (46)$$

and

$$\begin{aligned} \mathbb{V} [\langle \bar{V}(y_0), \delta(y_0) \rangle] &= \delta^t(y_0) \mathbb{V} [\bar{V}(y_0)] \delta(y_0) \\ &= \delta^t(y_0) \frac{\sigma^2}{N} I_T \delta(y_0) \\ &= \frac{\sigma^2}{N} \|\delta(y_0)\|^2. \end{aligned} \quad (47)$$

Then,

$$\langle \bar{V}(y_0), \delta(y_0) \rangle \sim \mathcal{N}\left(0, \frac{\sigma^2}{N} \|\delta(y_0)\|^2\right) \quad (48)$$

and the random variable

$$X = \frac{2 \langle \bar{V}(y_0), \delta(y_0) \rangle}{\|\delta(y_0)\|^2}, \quad (49)$$

follows a Gaussian central law with variance $\frac{4\sigma^2}{N\|\delta(y_0)\|^2}$, so that

$$Z = \frac{\sqrt{N}\|\delta(y_0)\|}{2\sigma} X \sim \mathcal{N}(0, 1). \quad (50)$$

Hence, the probability for a given 1-optimal label map to be preferred to the correct label map is

$$\begin{aligned} \mathbb{P}(E(\lambda) - E(\lambda') \geq 0) &= \mathbb{P}\left(Z \leq \frac{-\sqrt{N}\|\delta(y_0)\|}{2\sigma}\right) \\ &= \Phi\left(-\frac{\sqrt{N}\|\delta(y_0)\|}{2\sigma}\right), \end{aligned} \quad (51)$$

where $\|\delta(y_0)\| = \beta_{k_0}(y_0) \|m(\lambda_{k_0}) - m(\lambda'_{k_0})\|$ and $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$.

ACKNOWLEDGMENT

The authors would like to thank the Centre National d'Études Spatiales for providing ADAM images and EADS/ASTRIUM for enabling the conversion of images into vegetation fraction cover images.

REFERENCES

- [Pohl et Van Genderen, 1998] Pohl, C. et Van Genderen, J., Multisensor image fusion in remote sensing: concepts, methods and application. *Int. Journal of Remote Sensing* 19(5), pp. 823–854. 1998.
- [Ranchin et Wald, 2000] Ranchin, T. et Wald, L., Fusion of high spatial images: the ARSIS concept and its implementation. *Photogrammetric Engineering and Remote Sensing* 66(1), pp. 49–61. 2000.
- [Carper et al., 1990] Carper, W., Lillesand, T. et Kiefer, R., The use of intensity-hue-saturation transformations for merging spot panchromatic and multispectral image data. *Photogrammetric Engineering Remote Sensing* 56(4), pp. 459–467. 1990.
- [Laporterie-Déjean et al., 2003] Laporterie-Djean, F., Lopez-Ornelas, E. et Flouzat, G., Pre-segmentation of high-resolution images thanks to the morphological pyramid. In: *Proc. of Int. Geoscience and Remote Sensing Symposium, IGARSS'03*, Vol. 3, pp. 2042–2044. 2003.
- [Aiazzi et al., 1999] Aiazzi, B., Alparone, L., Barducci, A., Baronti, S. et Pippi, I., Multispectral fusion of multisensor image data by the generalized laplacian pyramid. In: *IGARSS*, Vol. 2, pp. 1183–1185. 1999.
- [Sveinsson et Benediktsson, 2000] Sveinsson, J. et Benediktsson, J., Data fusion and feature extraction using tree structured filter banks. In: *IGARSS*, Vol. 6, pp. 2617–2619. 2000.
- [Settle et Drake, 1993] Settle, J. et Drake, N., Linear mixing and the estimation of ground cover proportions. *International J. of Remote Sensing* 14(6), pp. 1159–1177. 1993.
- [Cross et al., 1991] Cross, A., Settle, J., Drake, N. et Paivinen, R., Sub-Pixel measure of tropical forest cover using avhrr data. *International Journal of Remote Sensing* 12, pp. 1119–1129. 1991.
- [Foody, 1996] Foody, G., Approaches for the production and evaluation of fuzzy land cover classifications from remotely sensed data. *International J. of Remote Sensing* 17, pp. 1317–1340. 1996.
- [Cardot et al., 2003] Cardot, H., Faivre, R. et Goulard, M., Functional approaches for predicting land use with the temporal evolution of coarse resolution remote sensing data. *Journal of Applied Statistics* 30, pp. 1185–1199. 2003.
- [Faivre et Fischer, 1997] Faivre, R. et Fischer, A., Predicting crop reflectances using satellite data observing mixed pixels. *Journal of Agricultural, Biological and Environmental Statistics* 2, pp. 87–107. 1997.
- [Brown et al., 2000] Brown, M., Lewis, H. et Gunn, S., Linear spectral mixture models and support vector machines remote sensing. *IEEE Trans. on Geoscience and Remote Sensing* 38, pp. 2346–2360. 2000.
- [Atkinson, 1997] Atkinson, P., Mapping sub-pixel boundaries from remotely sensed images. In: Z. Kemp (ed.), *Innovations in GIS IV*, Taylor and Francis, London, pp. 166–180. 1997.

- [Atkinson, 2001] Atkinson, P., Super-resolution target mapping from soft-classified remotely sensed imagery. In: Proc. of the 6th International Conference on Geocomputation, Brisbane, University of Queensland. 2001.
- [Tatem et al., 2002] Tatem, A., Lewis, H., Atkinson, P. and Nixon, M., Super-resolution land cover pattern prediction using a hopfield neural network. *Remote Sensing of Environment* 79(1), pp. 1–14. 2002.
- [Kasetkasem et al., 2005] Kasetkasem, T., Arora, M. et Varshney, P., Super-resolution land cover mapping using a markov random field based approach. *Remote Sensing of Environment* 96, pp. 302–314. 2005.
- [Robin et al., 2007] Robin, A., Moisan, L. and Le Hégarat-Masclé, S., An unsupervised approach for sub-pixelic land cover change detection. In: Proc. of the 4th International Workshop on the Analysis of Multitemporal Remote Sensing Images, Leuven, Belgium. 2007.
- [Van Der Meer, 1999] Van Der Meer, F., Iterative spectral unmixing. *International Journal of Remote Sensing* 20(17), pp. 3431–3436. 1999.
- [Horwitz et al., 1971] Horwitz, H., Nalepka, R., Hyde, P. et Morgenstern, J., Estimating the proportions of objects within a single resolution element of a multispectral scanner. In: Proceedings of the 7th International Symposium on Remote Sensing of Environment., Ann Arbor, Michigan, pp. 1307–1320. 1971.
- [Kerdiles et Grondona, 1995] Kerdiles, H. et Grondona, M., Noaa-avhrr ndvi decomposition and sub-pixel classification using linear mixing in the argentinean pampa. *International Journal of Remote Sensing* 16(7), pp. 1303–1325. 1995.
- [Small, 2001] Small, C., Estimation of urban vegetation abundance by spectral mixture analysis. *International Journal of Remote Sensing* 22(7), pp. 1305–1334. 2001.
- [Geman et Geman, 1984] Geman, D. et Geman, S., Stochastic relaxation, gibbs distribution, and the bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 6, pp. 721–741. 1984.
- [Le Hégarat-Masclé et al., 1996] Le Hégarat-Masclé, S., Vidal-Madjar, D. and Olivier, P., Applications of simulated annealing relaxation to SAR image clustering and classification problems. *Int. J. of Remote Sensing* 17(9), pp. 1761–1776. 1996.
- [Koepfler et al., 1994] Koepfler, G., Lopez, C. et Morel, J., A multiscale algorithm for image segmentation by variational method. *SIAM Journal on Numerical Analysis* 31(1), pp. 282–299. 1994.
- [Mumford et Shah, 1989] Mumford, D. et Shah, J., Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.* 42, pp. 577–685. 1989.
- [Jacquemoud et al., 1995] Jacquemoud, S., Baret, F., Andrieu, B., Danson, F. et Jaggard, K., Extraction of vegetation biophysical parameters by inversion of the prospect+sail models on sugar beet canopy reflectance data. application to tm and aviris sensors. *Remote Sensing of Environment* 52(3), pp. 163–172. 1995.
- [Le Hégarat-Masclé et al., 2005] Le Hégarat-Masclé, S., Ottlé, C. et Guérin, C., Land cover change detection at coarse spatial scales based on iterative estimation and previous state information. *Remote Sensing of Environment* 95, pp. 464–479. 2005.

Lionel Moisan studied mathematics and computer science at the Ecole Normale Supérieure in Paris. He received the PhD degree in 1997 from Paris-Dauphine University and the HDR degree in 2003 from Paris XI University. He was a CNRS researcher at Ecole Normale Supérieure de Cachan from 1998 to 2003, and is currently Professor of Applied Mathematics at Paris Descartes University. His research activities are mainly focused on variational models for image processing (restoration, smoothing) and on a-contrario statistical models for image analysis (gestalt and structure detection).

Amandine Robin received the PhD degree in Applied Mathematics from Paris Descartes University (France) in 2007. Her thesis was dedicated to sub-pixelic change detection and classification from remote sensed images. Her research interests include image processing, data fusion, a-contrario statistical modeling. It is particularly dedicated to change detection and classification for environmental applications.

Sylvie Le Hégarat-Masclé received the PhD degree in Signal and Image from ENST Paris (France) in 1996 and the "Habilitation à Diriger des Recherches" at Versailles University in 2006. She was associated professor at Versailles University from 1998 to 2005. Since 2006, she is professor at Paris XI University. Her interests include statistical pattern recognition, image analysis, and data fusion. She works with the Bayes theory and the Dempster-Shafer evidence theory. In remote sensing domain (SAR and optical sensor), concerned applications are: land cover / land-use classification, vegetation and soil moisture monitoring.