© 2011 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Title: A Novel Domain Adaptation Bayesian Classifier for Updating Land-Cover Maps with Class Differences in Source and Target Domains

This paper appears in: IEEE Transactions on Geoscience and Remote Sensing

Date of Publication: May 16, 2012

Author(s): Kanchan Bahirat, Francesca Bovolo, Lorenzo Bruzzone, Subhasis Chaudhuri

Volume: 50, Issue: 7

Page(s): 2810 - 2826

DOI: 10.1109\TGRS.2011.2174154

A Novel Domain Adaptation Bayesian Classifier for Updating Land-Cover Maps with Class Differences in Source and Target Domains

Kanchan Bahirat[*], *Francesca Bovolo*[#], *Lorenzo Bruzzone*[#] *and Subhasis Chaudhuri*[*]

[*]*Dept. of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai 400076*

[#]*Dept. of Information Engineering and Computer Science, University of Trento, Via Sommarive, 14 I-38123, Trento, Italy*

Abstract—This paper addresses the problem of land-cover maps updating by classification of multi-temporal remote sensing images in the context of domain adaptation. The basic assumptions behind the proposed approach are two. The first one is that training data (ground reference information) are available for one of the considered multi-temporal acquisitions (source domain) whereas they are not for the other (target domain). The second one is that multi-temporal acquisitions (*i.e.*, target and source domains) may be characterized by different sets of classes. Unlike other approaches available in the literature, the proposed Domain Adaptation Bayesian classifier based on Maximum a Posteriori decision rule (DA-MAP) automatically identifies whether there exist differences between the set of classes in the target and source domains and properly handles these differences in the updating process. The proposed method was tested in different scenarios of increasing complexity related to multitemporal image classification. Experimental results on a medium resolution and a very high resolution multitemporal remote sensing data sets confirm the effectiveness and the reliability of the proposed DA-MAP classifier.

Index Terms—Domain adaptation, Partially supervised learning, Partially unsupervised learning, Bayesian Classifier, Maximum-a-Posteriori classifier, Multi-temporal images classification, Land-cover map updating, Remote sensing

I. INTRODUCTION

The objective of Domain Adaptation (DA) techniques (also known as transfer learning or partially supervised/unsupervised learning) is to take advantage of the available knowledge on a given source domain in order to infer a model/classifier suitable for the classification of a related (yet not identical) target domain for which a priori information is not available [1],[2]. This kind of techniques have proven to be effective in different applications mainly related to text analysis and natural language processing [1]-[4]. Moreover, few successful examples can be found also in remote sensing where DA

techniques become useful when there is a need of classifying images acquired: 1) on spatially disjoint areas that show similar characteristics; or 2) on the same geographical area at different times. In this work the attention is focused on the analysis of multi-temporal remote-sensing images acquired on the same area at different times and the related land-cover maps updating [5]-[8]. Due to the periodic and regular acquisition of remote sensing images on the same geographical area and to the difficulties in collecting the reference data on the ground with the same frequency, it is not possible to apply standard supervised classification algorithms to each available image. In this context, DA techniques have a high importance for the development of monitoring systems aimed at regularly mapping geographical areas of interest. In this specific scenario the source domain is identified as the first image to be classified for which a training set is assumed to be available, whereas the target domain is associated to each new available image to be classified (acquired on the same geographical area of the source domain) for which a training set is likely to be not available [7]-[12]. In this context, DA should be faced according to semisupervised strategies that take advantage of the training set available for the source domain and the unlabeled samples from target domain in order to derive a classification rule suitable for the target domain.

The few domain adaptation methods proposed in the remote sensing literature are based on the assumption that the set of land-cover classes that models the target domain should be the same as those included in the source domain. In other words, the differences between the two domains are only related to differences in the statistical parameters of land-cover classes due to differences in the acquisition conditions (*e.g.*, differences in the atmospheric conditions at the image acquisition dates, sensor nonlinearities, different levels of soil moisture). Under this assumption, in [9] a partially-unsupervised approach is proposed, which can update the parameters of an already trained parametric maximum-likelihood (ML) classifier on the basis of the distribution of a new image for which no ground reference information is available. In [10], in order to take into account the temporal correlation

between images acquired over the same area at different times, the partially-unsupervised ML classification approach is reformulated in the framework of the Bayesian rule for cascade classification. The basic idea in both approaches consists in modeling the observed spaces by a mixture of distributions, whose components are estimated through the employment of unlabeled data according to a proper inference applied to training samples of the reference image. In [11] and [12] partially-unsupervised classification approaches based on a multiple-classifier system and a multiple-cascade-classifier system (MCCS) have been defined, respectively. However, even if the above-mentioned methods resulted effective in many applications, in some real cases it is not possible to assume that the set of land-cover classes that models the target domain is the same as those included in the source domain. In these cases land-cover classes may not only change their statistical behaviors but may also appear and/or disappear. Differences in the set of classes that characterize source and target domains strongly affect the effectiveness of standard domain adaptation methods that cannot address this kind of problem.

In this work, we propose a domain adaptation technique that overcomes the abovementioned limit. This technique is based on the use of a maximum-a-posteriori (MAP) Bayesian classifier [13]. However, other kinds of classifiers could be adopted in the context of the proposed method, as for example machine learning based classification algorithms [7],[8]. The standard approach to domain adaptation based on MAP classifier considers the class statistical parameters estimated in a supervised way from an image for which training data are available (the source domain) as rough estimate of the statistical parameters of classes in each new image to be classified (target domain). These values need to be properly adapted to the target domain according to the characteristics of the new image that has to be analyzed. Unlike standard methods, the proposed technique identifies before performing adaptation whether the new classes have appeared and/or some of the existing classes disappeared in the new considered image by integrating well established change-detection techniques [15],[16],[17] with a

statistical distance measure in the analysis of the multi-temporal images. According to the specific detected situation on the relationship between classes present in the two images, the proposed approach properly initializes the domain adaptation procedure and updates the class statistical parameters relevant to the new image by using the Expectation Maximization (EM) algorithm. Before computing the final classification map, the Bayesian information criterion is used to validate the estimated model.

The manuscript is organized into seven sections. Section II introduces the notation and background of the addressed domain adaptation problem. Section III presents the proposed system for land-cover maps updating. Section IV details the proposed procedure for domain adaptation. The multitemporal Landsat and QuickBird datasets used in the experiments are described in Sections V and VI, respectively, together with the different scenarios used to test the effectiveness of the proposed approach and the achieved results. Finally, Section VII draws the conclusions of this work.

II. Definitions and Problem Formulation

Let $\mathbf{X}_1$ be a multispectral image made up of B spectral channels and M pixels acquired over a geographical area of interest at time t_1. Let $x_{1,j}$ be the feature vector associated with the jth pixel of image $\mathbf{X}_1$. Let us assume that reliable reference data on the ground (and thus a training set $\mathbf{T}_1$) are available for image $\mathbf{X}_1$ (source domain). Let $\Omega_1=\{\omega_1,\omega_2,\ldots,\omega_{K_1}\}$ be the set of land-cover classes that characterizes the considered geographical area at time t_1, where K_1 is the number of classes modeled in the training set $\mathbf{T}_1$.

In the context of the Bayes decision theory, the decision rule that maximizes the posterior probability (i.e., that minimizes the error probability in the sense of Bayesian theory) is expressed as follows [13],[14]:

$$x_{1,j} \in \omega_k \Leftrightarrow \omega_k = \underset{\omega_i \in \Omega_1}{arg\,max}\left\{P_1(\omega_i)\, p(x_{1,j}|\omega_i)\right\} \; with\; j=1,..,M \tag{1}$$

where $P_1(\omega_i)$ is the estimate of the *a priori* probability of class ω_i in image $\mathbf{X}_1$, $p(x_{1,j}|\omega_i)$ is the value estimated for $p(X_1|\omega_i)$ for feature vector $x_{1,j}$ given ω_i in $\mathbf{X}_1$, and $p(X_1|\omega_i)$ is the conditional density

function of random variable X_1 associated to image $\mathbf{X}_1$. According to (1), the training phase of a MAP classifier requires the estimation of $P_1(\omega_i)$ and $p(X_1|\omega_i)$ for each class $\omega_i \in \Omega_1$. Such estimates can be obtained by using classical supervised approaches which use the information present in the considered training set $\mathbf{T}_1$ [13],[14]. Let us assume (as usually done in the remote sensing literature) that classes in images acquired by multispectral passive sensors are Gaussian distributed (*i.e.*, $p(X_1|\omega_k)$, k=1,…,K_1, can be modeled by a Gaussian function) [14]. Under this assumption, the probability density function $p(X_1)$ of pixel values in $\mathbf{X}_1$ can be described as a mixture of multivariate Gaussian distributions with as many components as the number of classes in the training set $\mathbf{T}_1$, *i.e.*, $p(X_1)=\sum_{k=1}^{K_1} P_1(\omega_k)\, p(X_1|\omega_k)$. Therefore, the solution of the MAP supervised classification of $\mathbf{X}_1$ requires the estimation of the set of parameters $\theta_1 = \{\mu_{1,k}, \Sigma_{1,k}, P_1(\omega_k)\}$, k=1,…,K_1, where $\mu_{1,k}$ and $\Sigma_{1,k}$ are the mean vector and the covariance matrix of $\omega_k \in \Omega_1$, respectively, obtained according to maximum likelihood estimation in the feature space of image $\mathbf{X}_1$ according to the available training set $\mathbf{T}_1$. $P_1(\omega_k)$ is the estimation of the prior probability of $\omega_k \in \Omega_1$. It is computed, as usually done in the remote sensing literature, on the basis of the relative frequency of classes in $\mathbf{T}_1$.

Let us now assume that at time t_2, another B-dimensional multispectral image $\mathbf{X}_2$ of size M is acquired on the same geographical area as $\mathbf{X}_1$, and that the land-cover map of the area should be updated. Furthermore, let us assume that the reference data on the ground about $\mathbf{X}_2$ (target domain) is not available, as it often happens in real applications. In this situation, the few domain adaptation techniques presented in the remote sensing literature usually assume that the set of classes Ω_2 that characterizes image $\mathbf{X}_2$ is the same that describes $\mathbf{X}_1$ (*i.e.*, Ω_2=Ω_1). Despite this assumption, in general the classifier trained on the image $\mathbf{X}_1$ does not properly work on the image $\mathbf{X}_2$ because the estimates of statistical class parameters at t_1 do not provide an accurate approximation for the same terms at t_2. As stated in the introduction, this is due to several factors that alter the spectral signatures of land cover

classes at different acquisition times and consequently, the distributions of such classes in the feature space. However, usually the parameters in θ_1 estimated on $\mathbf{X}_1$ represent reasonable rough estimate for the ones on $\mathbf{X}_2$. Thus, the goal of the domain adaptation procedure is to adapt the parameters in θ_1 to the properties of $\mathbf{X}_2$. Adaptation can be carried out according to the unsupervised iterative Expectation-Maximization (EM) algorithm [9] but other approaches can be used also [7].

In real applications, despite Ω_2 and Ω_1 show a high correlation, it is not possible to guarantee that they are the same. According to this observation and differently to what commonly assumed in the literature [9], we consider the possibility that classes in Ω_2 are partially different from those in Ω_1. In other words, classes might have been appeared or disappeared between the two acquisition dates. Moreover, we consider that land-cover classes common to both $\mathbf{X}_1$ and $\mathbf{X}_2$ may change over time their spatial distribution within the geographical area of interest. This means that spatial changes might have occurred on the ground without necessarily implying differences between classes in Ω_1 and Ω_2. The following cases can be identified:

CASE A. No changes or only a spatial shift of classes have been detected: $\Omega_2=\Omega_1$.

CASE B. A new unknown class ω_u has been detected in the image $\mathbf{X}_2$: $\Omega_2=\Omega_1\cup\{\omega_u\}$.

CASE C. A class $\omega_k \in \Omega_1$ disappeared between $\mathbf{X}_1$ and $\mathbf{X}_2$: $\Omega_2=\Omega_1\setminus\{\omega_k\}$.

CASE D. A joint occurrence of CASEs B and C: $\Omega_2=\Omega_1\setminus\{\omega_k\}\cup\{\omega_u\}$.

CASE E. Complex situation in which combinations of CASEs A-D can be detected. This is often the case that may occurs in real applications, *i.e.*:

i) more than one new unknown class is detected: $\Omega_2=\Omega_1\cup\Omega_U$, where Ω_U is the set of unknown U added classes;

ii) more than one class disappeared: $\Omega_2=\Omega_1\setminus\Omega_K$, where Ω_K is the set of K deleted ω_k classes;

iii) some classes are appeared and some others are disappeared simultaneously:

$$\Omega_2 = \Omega_1 \setminus \Omega_K \cup \Omega_U.$$

As it will be explained in the next section, the adaptation procedure based on the EM algorithm [9] requires specific modifications in order to handle the abovementioned cases.

It is worth noting that, although Ω_2 may differ from Ω_1, a high correlation between them is expected. In other words, most of the classes present in the considered area at time t_1 (and represented in the training $\mathbf{T}_1$) are also present at time t_2. This is the case of many real applications in which land-cover maps updating is required. A reasonable assumption for the proposed DA technique is to consider a maximum of two added and/or two deleted classes (i.e., K and $U \leq 2$). If this is not the case, it would not be realistic to adapt the MAP classifier trained on $\mathbf{T}_1$ to image $\mathbf{X}_2$, as target and source domains would result in significantly different problems making senseless the adaptation process.

III. Proposed System for Land-Cover Maps Updating

The objective of the proposed domain adaptation method is to compute a reliable classification map of a given remote sensing image $\mathbf{X}_2$, for which reference data on the ground are not available. In this situation, it is not possible to estimate class statistical parameters only from the information carried by image $\mathbf{X}_2$. As standard domain adaptation methods, the proposed technique assumes that an image $\mathbf{X}_1$ previously acquired over the same geographical area is available together with a reliable training set for it. Furthermore it assumes that the information about land-cover classes present in the training set for $\mathbf{X}_1$ does not completely represents land-cover classes in image $\mathbf{X}_2$, as classes may be appeared or disappeared between the two acquisitions. In order to cope with this problem, a 4-step approach is proposed based on: i) hypothesis formulation on the differences between land-cover classes at the two acquisition dates; ii) domain adaptation (based on EM algorithm); iii) hypothesis validation; and iv) generation of the classification map of image $\mathbf{X}_2$. The block diagram in Figure 1 points out the general architecture of the proposed approach.

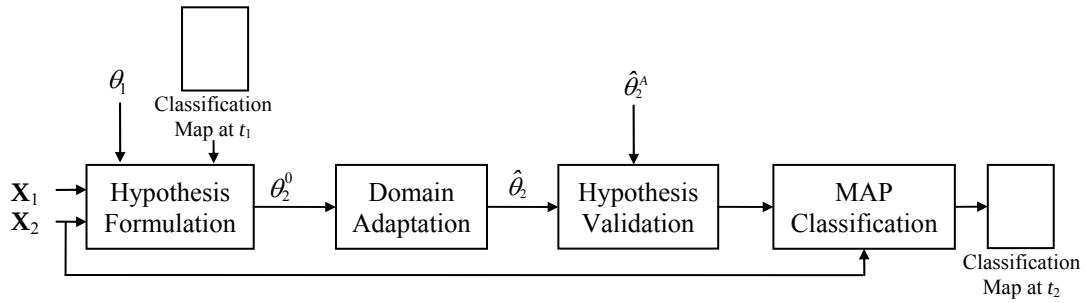


Figure 1. Block diagram of the proposed domain adaptation technique.

The first step aims at formulating the hypothesis on the classes present in Ω_2 and their relationship with the ones in Ω_1. The goal is to establish whether Ω_2 is equal to Ω_1 or not. If not, the procedure tries to identify the cause of differences between the two sets. This step takes advantage of unsupervised change-detection methods in order to formulate proper hypothesis. Accordingly, the set of class statistical parameters θ_2^0 for image $\mathbf{X}_2$ is initialized. Under the specific hypothesis formulated in the first step, the second step performs the adaptation of the Maximum-A-Posteriori classifier built for $\mathbf{X}_1$ to the statistical properties of $\mathbf{X}_2$. Once the new model $\hat{\theta}_2$ for $\mathbf{X}_2$ is estimated, adequate statistical criteria are used for validation purposes. In this step, the model $\hat{\theta}_2$ built according to the proposed method is compared to the one that could be achieved under the common assumption that the set of classes Ω_2 that characterizes image $\mathbf{X}_2$ is the same that describes $\mathbf{X}_1$ i.e., $\hat{\theta}_2^A$ (CASE A). If validated, the parameters in $\hat{\theta}_2$ are used to define the MAP classifier, and labels in $\mathbf{X}_2$ are assigned as follows

$$x_{2,j} \in \omega_k \Leftrightarrow \omega_k = \underset{\omega_i \in \Omega_2}{arg\,max}\left\{\hat{P}_2(\omega_i)\,\hat{p}(x_{2,j}\,|\,\omega_i)\right\} \; with\; j = 1,..,M \tag{2}$$

Where, Ω_2 is the set of classes that properly describes the new image $\mathbf{X}_2$, and $\hat{P}_2(\omega_i)$ and $\hat{p}(x_{2,j}\,|\,\omega_i)$ are the class prior probabilities and class conditional probabilities, respectively, defined according to the proposed method. K_2 is the cardinality of Ω_2 and depends on K_1 according to the formulated hypothesis.

In the following section our attention is devoted to steps i) to iii) of the proposed procedure, which represent the main novel contributions of this paper.

IV. DOMAIN ADAPTATION PROCEDURE

A. Hypothesis Formulation

The Hypothesis Formulation step aims at detecting whether $\mathbf{X}_2$ shares all the classes with the image $\mathbf{X}_1$ (except for their spatial position), *i.e.*, $\Omega_2=\Omega_1$, or not, *i.e.*, $\Omega_2\neq\Omega_1$. In the second case, the proposed procedure aims at identifying and modeling the differences that characterize the two sets. In order to achieve this goal in an unsupervised way (as no training set is available for $\mathbf{X}_2$), the unsupervised Change Vector Analysis technique [15] is applied to images $\mathbf{X}_1$ and $\mathbf{X}_2$ for change detection purposes. The vectors representing the two images are subtracted from each other pixel-by-pixel in order to build a multispectral difference image. The information present in the multispectral difference image is exploited according to the theoretical framework for unsupervised change detection based on Change Vector Analysis in polar domain proposed in [17]. According to [17], for simplicity two spectral bands out of B are selected such that the most informative features with respect to the specific considered problem are isolated excluding noisy and misleading spectral channels from the analysis. It is worth noting that, even if the assumption of working with a couple of spectral channels is reasonable in many change-detection problems [16],[17], the method can be also applied to all spectral channels. From the defined 2-dimensional feature space, the polar representation of the change-detection problem is built on the basis of the magnitude and the direction of spectral change vectors. In this feature space, no-changed pixels are concentrated close to the origin of the polar domain and fall within the *Circle of no-changed* (C_n) pixels, whereas changed pixels fall far from the origin within the *Annulus of changed* (A_c) pixels [17]. The threshold value T that separates C_n from A_c along the magnitude variable ρ can be computed according to any thresholding technique available in the literature [18]. Changed pixels belonging to different land-cover transitions (*i.e.*, different kinds of change) show along the direction variable ϑ in A_c different preferred directions and fall therefore in different *Annular sectors of change* (S_k). Each sector is bounded by a pair of angular thresholds ϑ_{k_1} and ϑ_{k_2} that can be detected according

to the method in [19]. It is worth noting that the identification of annular sectors S_k makes it possible to reduce the effect on the change-detection map of noisy pixels (*e.g.*, due to registration noise) that fall in A_c but are not associated with changes on the ground [17]. Figure 2 summarizes the decision regions according to the CVA framework.

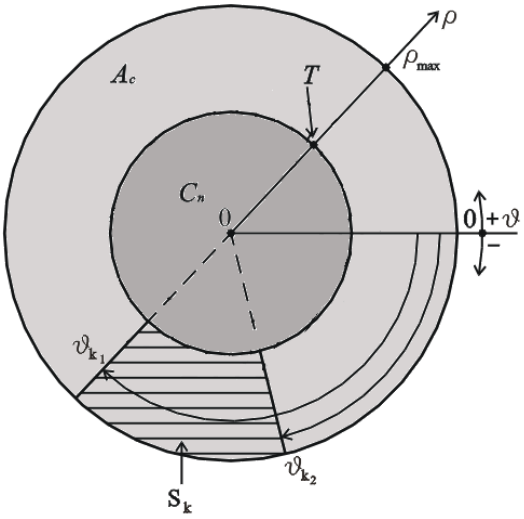


Figure 2. Representation of the regions of interest for the CVA technique in the Polar coordinate system.

Once the mentioned decision thresholds have been detected along both the magnitude and the direction variables, under the assumption that each identified annular sector corresponds to one (and only one) kind of change, each pixel of the considered scene can be associated to the class of no-changed pixels or to one of the possible kinds of change. Each annular sector isolates a cluster of pixels in the considered scene that experienced the same land-cover transition. The reader is referred to [15]-[19] for further theoretical details on the CVA framework and on how decision threshold values can be computed. It is worth stressing that this process is carried out in an unsupervised way. Therefore from the change-detection step, no explicit knowledge about the specific land-cover classes present in the two considered images and the specific land-cover transitions can be extracted.

According to the relationship between image $\mathbf{X}_1$ and $\mathbf{X}_2$, two main different situations can be detected in the polar representation: i) all the multispectral change vectors fall into C_n and none of them in A_c, *i.e.*, no changes occurred on the ground and $\Omega_2=\Omega_1$; and ii) a significant (from the statistical

point of view) number of multispectral change vectors fall into A_c. In this second case, changes are occurred between the two acquisitions. Such changes may correspond to a spatial shift of classes (*i.e.*, changes occurred on the ground but still $\Omega_2=\Omega_1$) or to the appearance/disappearance of some classes in the $\mathbf{X}_2$ image (*i.e.*, $\Omega_2\neq\Omega_1$). It is worth noting, that for each detected kind of change, the land-cover class at time t_1 is known, whereas the one at t_2 is unknown. On the basis of this observation, and in order to distinguish between these possible situations, the second step of the *Hypothesis Formulation* is a comparison between the computed change-detection map and the classification map obtained for $\mathbf{X}_1$. This comparison is performed in order to understand if the change event(s) affected: i) only a part of the pixels belonging to a given class ω_k in Ω_1 (CASE A and CASE B); or ii) all the pixels belonging to that class (CASE C and CASE D).

CASE A and CASE B

If only a part of the pixels of ω_k have been affected by the change event, it means that ω_k is still present in $\mathbf{X}_2$, but a further investigation is needed to understand whether the change event caused only a spatial shift of classes (*i.e.*, the transition occurred in the direction of a class already existing in Ω_1) or if a new class is present in $\mathbf{X}_2$. To this end, we first remove changed pixels from the processing. For the remaining set of pixels, the basic assumption made by standard domain adaptation procedure (*i.e.*, $\Omega_2=\Omega_1$) is satisfied (as they are not changed between the two dates) [9]. Under this assumption, a first rough estimate of the statistical parameters in θ_2^0 that characterize land-cover classes at time t_2 can be obtained by exploiting the Maximum Likelihood (ML) estimates at time t_1 performed on the available training set, *i.e.*,

$$\theta_2^0=\theta_1 \ . \tag{3}$$

The equivalence in (3) holds as we can assume that classes in image $\mathbf{X}_2$ are Gaussian distributed as the ones in $\mathbf{X}_1$, being $\mathbf{X}_2$ a multispectral passive sensor image as well as $\mathbf{X}_1$ and it is valid only for classes

ω_k that belong both to Ω_1 and Ω_2.[1] Being both $p(X_2)$ and $p(X_1)$ a mixture of as many Gaussian distributions as land-cover classes, equation (3) can be rewritten as

$$\mu_{2,k}^0 = \mu_{1,k}, \quad \Sigma_{2,k}^0 = \Sigma_{1,k}, \quad P^0(\omega_{2,k}) = P(\omega_{1,k}) \tag{4}$$

for each classes ω_k that belongs both to Ω_1 and Ω_2.

The adaptation of θ_1 to the properties of $\mathbf{X}_2$ is achieved in a robust way through the iterative Expectation-Maximization algorithm [9],[24]. It can be demonstrated that under the assumption of Gaussian distributed classes, the iterative equations to be applied for estimating the statistical terms associated with each class ω_k in $\mathbf{X}_2$ are as follows [22]-[24]:

$$P_2^{s+1}(\omega_k) = \frac{1}{M} \sum_{x_{2,j} \in \mathbf{X}_2} \frac{P_2^s(\omega_k)\, p(x_{2,j} \mid \omega_k)}{p^s(x_{2,j})} \tag{5}$$

$$[\mu_{2,k}]^{s+1} = \frac{\sum_{x_{2,j} \in \mathbf{X}_2} \frac{P_2^s(\omega_k)\, p(x_{2,j} \mid \omega_k)}{p^s(x_{2,j})} x_{2,j}}{\sum_{x_{2,j} \in \mathbf{X}_2} \frac{P_2^s(\omega_k)\, p(x_{2,j} \mid \omega_k)}{p^s(x_{2,j})}} \tag{6}$$

$$[\Sigma_{2,k}]^{s+1} = \frac{\sum_{x_{2,j} \in \mathbf{X}_2} \frac{P_2^s(\omega_k)\, p_2(x_{2,j} \mid \omega_k)}{p^s(x_{2,j})} \{x_{2,j} - [\mu_{2,j}]^{s+1}\}^2}{\sum_{x_{2,j} \in \mathbf{X}_2} \frac{P_2^s(\omega_k)\, p_2(x_{2,j} \mid \omega_k)}{p^s(x_{2,j})}} \tag{7}$$

where the superscripts s and s+1 refer to the values of the parameters at the current and next iterations, respectively. The iterative process ends when a local maximum of the log-likelihood function $L(\hat{\theta}_2)$ is achieved. $\hat{\theta}_2$ obtained at convergence is the new set of estimated parameters $\hat{\theta}_2 = \{\hat{\mu}_{2,k}, \hat{\Sigma}_{2,k}, \hat{P}_2(\omega_k)\}$, that describes in the feature space of $\mathbf{X}_2$ classes ω_k common to both Ω_1 and Ω_2.

Once the statistical parameters for image $\mathbf{X}_2$ have been estimated the attention is focused on the changed pixels. The similarity/dissimilarity of unknown class ω_u associated with changed pixels in $\mathbf{X}_2$ with respect to known (from the relationship with $\mathbf{X}_1$) classes $\omega_k \in \Omega_2$ in $\mathbf{X}_2$ is evaluated according to a statistical distance measure. It is expected that the distance between ω_u and a class $\omega_k \in \Omega_2$ is small if changed pixels in $\mathbf{X}_2$ belong to ω_k, whereas it is high if pixels do not belong to it. In other words, if for

[1] In this paper we are implicitly assuming that classes with the same subscript k correspond to the same land cover on the ground in both $\mathbf{X}_1$ and $\mathbf{X}_2$.

the set of changed pixels there exists a class $\omega_k \in \Omega_2$ such that the statistical distance measure is below a user defined threshold value T_L, then changed pixels in $\mathbf{X}_2$ are likely to belong to that class ω_k that is present also in $\mathbf{X}_1$ (*i.e.*, $\omega_u=\omega_k$). Accordingly, a spatial shift of classes is detected and $\Omega_2=\Omega_1$ (CASE A). On the opposite, if the set of changed pixels shows for all the classes ω_k a statistical distance measure higher than a user defined threshold value T_H, then ω_u is statistically different from all the classes in Ω_1 and an addition of class is detected, *i.e.*, $\Omega_2=\Omega_1 \cup \{\omega_u\}$ (CASE B). It is worth stressing that, even if it is possible to detect the existence of a new class, we do not know anything about the land-cover label associated to this specific class because we do not have ground truth information for $\mathbf{X}_2$.

In this work, we have selected the *Jeffreys-Matusita* (*JM*) distance as statistical distance measure which is defined between classes ω_k and ω_u as:

$$JM_{ku} = \sqrt{2\left(1-e^{-B_{ku}}\right)} \tag{8}$$

where B_{ku} is the Bhattacharyya distance between the two mentioned classes defined as

$$B_{ku} = -\ln\left\{\int_{X_2} \sqrt{p(X_2 \mid \omega_k)\, p(X_2 \mid \omega_u)}\, dX_2\right\} \tag{9}$$

Under the assumption of Gaussian distributed classes, B_{ku} can be rewritten as

$$B_{ku} = \frac{1}{8}(\mu_{2,k}-\mu_{2,u})^t \left(\frac{\Sigma_{2,k}+\Sigma_{2,u}}{2}\right)^{-1} (\mu_{2,k}-\mu_{2,u}) + \frac{1}{2}\log\left[\left|\frac{\Sigma_{2,k}+\Sigma_{2,u}}{2}\right| \Big/ \sqrt{|\Sigma_{2,k}||\Sigma_{2,u}|}\right] \tag{10}$$

where the statistical parameters $\mu_{2,k}$, $\Sigma_{2,k}$ of class ω_k are those obtained according to EM algorithm, whereas the statistical parameters $\mu_{2,u}$, $\Sigma_{2,u}$ of class ω_u are the ML-estimates of the mean vector and co-variance matrix evaluated only on changed pixels in the feature space defined by $\mathbf{X}_2$. Figure 3 shows the general behavior of the *JM* distance versus the Bhattacharyya (B) distance. The asymptotic properties of *JM* as a function of *B* [14] are exploited for the definition of the threshold values T_L and T_H. When the classes are highly separated (*i.e.*, they can be considered as being associated to two

different land-cover classes) the Bhattacharyya distance is high and the *JM* distance saturates to $\sqrt{2}$ [as it is clear from (8)]. It is worth noting that the *JM* distance is related to the Chernoff upper bound to the error probability [25]. Thus it is expected that the saturation is obtained when the classes are well separated in terms of Bayes classification rule. Here, we empirically consider *JM* distance as being saturated when it comes close to the 90% of $\sqrt{2}$. Accordingly, the upper threshold T_H is set equal to 1.27. On the other hand, two classes are highly similar to each other (*i.e.*, they have a high probability of being the same land-cover class), when the Bhattacharyya distance is small and the *JM* distance is far from the saturation. Here, to be conservative, this condition is considered to be satisfied when *JM* distance is smaller than 70% of $\sqrt{2}$, *i.e.*, $T_L = 0.99$. Even if T_L is set empirically, its value is not critical and different choices do not significantly affect the performance of the proposed method. The region between T_L and T_H is a region of uncertainty. For a pair of classes that would result in a distance between T_L and T_H it is not possible to conclude whether they are different or similar. Accordingly, a change that leads to this ambiguous situation deserves further non-automatic analysis.

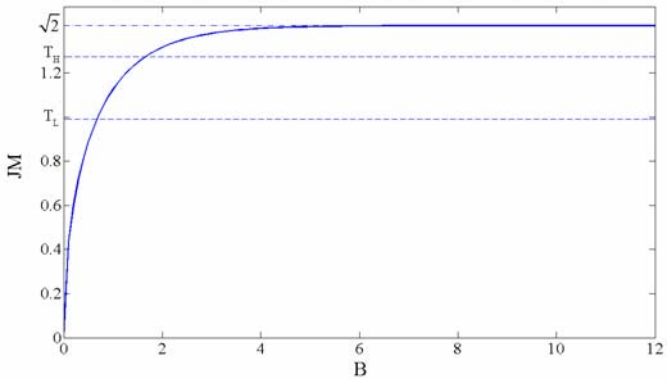


Figure 3. Behavior of the Jeffreys-Matusita (*JM*) distance versus the Bhattacharyya (*B*) distance.

CASE C and CASE D

If all the pixels belonging to a given class ω_k in Ω_1 are affected by a change event, the specific considered land-cover transition resulted in a deletion of class ω_k at t_2. Also in this situation, a further analysis is required in order to understand whether a class already existing in Ω_1 has replaced ω_k or if a

new class is present in $\mathbf{X}_2$. To this end, the same procedure already described before is applied without considering the deleted class ω_k (*i.e.*, the set $\Omega_1 \backslash \{\omega_k\}$ is considered). Once pair-wise *JM* distances have been computed, if changed pixels show a small distance (smaller than T_L) with respect to one of the classes in Ω_2 a simple deletion is detected: accordingly $\Omega_2 = \Omega_1 \backslash \{\omega_k\}$ (CASE C). Otherwise, if changed pixels show a high distance (higher than T_H) with respect to all the classes in $\Omega_1 \backslash \{\omega_k\}$, then a simultaneous deletion/addition of classes is detected and $\Omega_2 = \Omega_1 \backslash \{\omega_k\} \cup \{\omega_u\}$ (CASE D).

The described procedure is applied separately to each kind of change (*i.e.*, land-cover transition) detected in the change-detection step. Therefore different combinations of cases A-D can be detected, *i.e.*, simultaneous multiple additions and deletions of classes can be detected. Figure 4 gives a synthetic overview of possible hypothesis that can be formulated according to the proposed method about the relationship between Ω_2 and Ω_1.

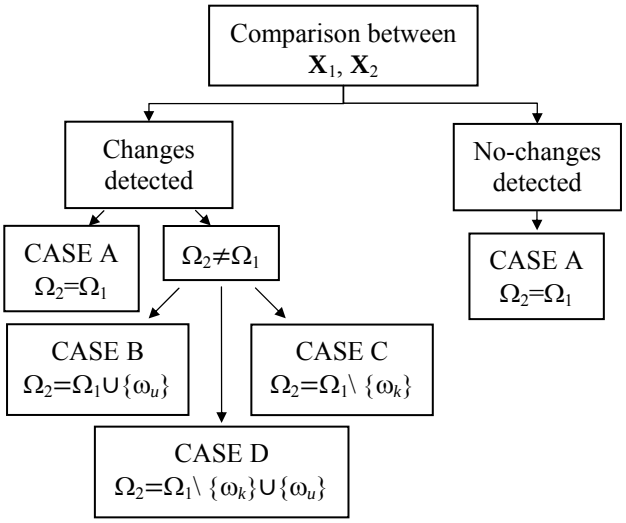


Figure 4. Overview of possible cases associated with the relation between Ω_1 and Ω_2.

B. Unsupervised Retraining and Hypothesis Validation

For each of the abovementioned cases an adequate initialization of the retraining process should be defined.

CASE A: $\Omega_2=\Omega_1$.

This is the case in which no changes (or only a spatial shift of classes) have been detected. In this situation, the hypothesis made in [9] holds and the retraining process can be carried out according to eq. (5)-(7). In the following, we will refer to this case as "standard" because this is the assumption commonly made by domain adaptation methods published in the literature [7],[9].

CASE B: $\Omega_2=\Omega_1 \cup \{\omega_u\}$.

This is the case in which a new class has been detected in the image $\mathbf{X}_2$. For all classes common to $\mathbf{X}_1$ and $\mathbf{X}_2$, the same initialization approach as in CASE A can be adopted, whereas this is not possible for class ω_u. Since we can reasonably assume that all the pixels representing the class ω_u are those that in $\mathbf{X}_2$ are detected as being changed (see the previous section), the initial values for μ_{2u} and Σ_{2u} are the ML-estimation of the mean vector and co-variance matrix of changed pixels evaluated in the feature space of $\mathbf{X}_2$, and $P_2(\omega_u)$ is their relative frequency. The initial set of class statistical parameters to be used in the iterative adaptation procedure is defined as

$$\theta_2^0 = \theta_1 \cup \{\mu_{2u}^0, \Sigma_{2u}^0, P_2^0(\omega_u)\} \tag{11}$$

Where, the prior probabilities of classes in θ_1 are rescaled to take into account the prior probability of the new class ω_u. Once the initial values have been fixed, equations (5)-(7) can be iteratively applied until convergence including also ω_u in the updating process.

CASE C: $\Omega_2=\Omega_1 \setminus \{\omega_k\}$.

This is the case in which the class has disappeared between $\mathbf{X}_1$ and $\mathbf{X}_2$. For all classes common to $\mathbf{X}_1$ and $\mathbf{X}_2$ the same approach as for CASE A can be adopted, whereas this is not possible for class ω_k. As this class is no longer present in the new image to be classified, its statistical parameters are

removed from the estimation problem. Therefore θ_2^0 is defined as

$$\theta_2^0 = \theta_1 \setminus \{\mu_{1,k}, \Sigma_{1,k}, P_1(\omega_k)\} \tag{12}$$

where the prior probabilities of classes are properly rescaled for taking into account the lack of the class ω_k. Equations (5)-(7) can be iteratively applied until convergence.

CASE D: $\Omega_2 = \Omega_1 \setminus \{\omega_k\} \cup \{\omega_u\}$.

This case is a combination of CASE B and CASE C. Accordingly, θ_2^0 is defined as

$$\theta_2^0 = \hat{\theta}_1 \setminus \{\hat{\mu}_{1,k}, \hat{\Sigma}_{1,k}, \hat{P}_1(\omega_k)\} \cup \{\mu_{2,u}^0, \Sigma_{2,u}^0, P_2^0(\omega_u)\} \tag{13}$$

As the polar framework for change detection allows us to distinguish different sources of spectral change [15], each detected complex case of multiple addition/deletion of classes can be split in sub-cases ascribable to one of the basic cases A-D. In such complex cases, the definition of the initial set of statistical parameters θ_2^0 that describes $\mathbf{X}_2$ can therefore be performed according to straightforward combinations of (11)-(13).

According to the above analysis of all the possible situations that one may encounter, it is possible to define the number K_2 of classes in Ω_2 as a function of the number K_1 of classes in Ω_1, *i.e.*,

$$K_2 = K_1 + U - K \tag{14}$$

K and U are the number of deleted and unknown added classes, respectively.

C. Hypothesis Validation

Once the hypothesis about addition and deletion of classes has been formulated and the set $\hat{\theta}_2$ of statistical parameters of classes in $\mathbf{X}_2$ has been estimated according to the EM algorithm, given the complexity of the problem it is important to apply a procedure to further confirm the validity of the model in order to increase the reliability of the method. The objective of this step is to understand if the

class model defined for $\mathbf{X}_2$ according to the proposed domain adaptation procedure is better than the model that could be obtained under the standard hypothesis (*i.e.*, $\Omega_2=\Omega_1$). To this end, the Bayesian Information Criterion (BIC) proposed by Schwarz [21] is adopted. This criterion is used for the model selection among a set of parametric models with a different number of parameters. According to our notation, BIC can be written as[2]

$$BIC = -2ln\left[\sum_{k=1}^{K_2} p(X_2 \mid \omega_k)\right] + nln(M) \tag{15}$$

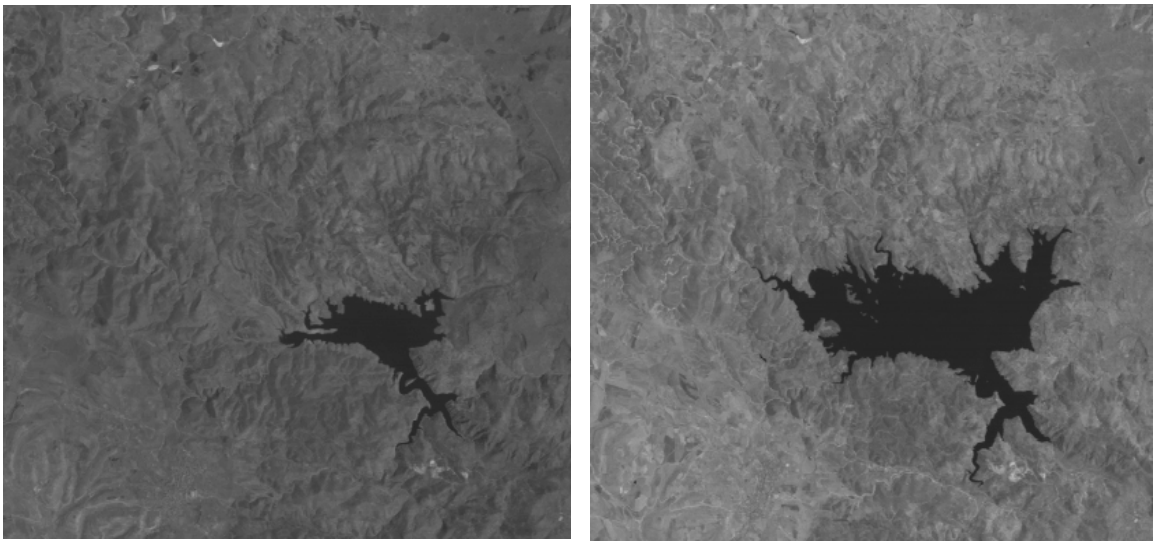
Where, K_2 is the number of classes in $\mathbf{X}_2$. The value of K_2 is a function of K_1 and depends on the number of added and deleted classes (14). The first term in (15) is the likelihood of data given the model, whereas the second term is a penalty term that avoids possible over-fitting that may arise from the increment in the number of parameters n that characterize the considered model. In our case, n is equal to the number of classes multiplied by the set of class parameters, and M is the number of samples (*i.e.*, the image size). In our architecture, BIC is computed for different models namely the one obtained according to the proposed domain adaptation procedure $\hat{\theta}_2$ and the one obtained according to the assumption that no changes are present between the classes of the two considered images, $\hat{\theta}_2^A$. As the smaller values of BIC indicate better fitting, if the proposed model results in the lowest BIC value then the hypothesis formulated according to the proposed domain adaptation approach is validated [21] and the MAP decision rule in (2) is applied to compute the desired classification map of $\mathbf{X}_2$ with parameters in $\hat{\theta}_2$. Else the model cannot be validated and we assume that the problem is too complex to be managed in a completely unsupervised way and thus external supervision (*e.g.*, the analysis of the problem from an expert) is required to confirm or change the assumption on the classes at t_2.

[2] This expression is an asymptotic result derived under the assumptions that the data distribution is in the exponential family. This assumption is verified in our case as classes are supposed to be Gaussian distributed.

V. EXPERIMENTAL RESULTS: LANDSAT DATA SET

A. Data Set Description

In order to test the effectiveness of the proposed DA-MAP classifier, we first considered a data set made up of two co-registered multispectral images acquired by the Thematic Mapper (TM) sensor of the Landsat-5 satellite. Bands 1-5 and 7 of the TM sensor were involved in the experiments (the thermal infrared band was neglected due to its lower geometrical resolution) for a total number of B=6 spectral channels. The selected test site is a section of 412×382 pixels of a scene including the area surrounding Lake Mulargias on the Island of Sardinia (Italy). The two images were acquired in September 1995 (**Y**) and July 1996 (**Z**), respectively. Figure 5 shows as an example of channels 4 of both images. The available ground truth information was used to derive a training set and a test set for each image. The two images share five land-cover classes [*i.e.*, pasture (ω_1), forest (ω_2), urban area (ω_3), water (ω_4), vineyard (ω_5)]. The image acquired in September 1995 also includes one additional class [*i.e.*, bare soil (ω_6)]. In other words, if we consider **Y** as $\mathbf{X}_1$ and **Z** as $\mathbf{X}_2$ (as temporally logic in the considered data set) then one class (*i.e.*, bare soil) has disappeared between the two acquisitions. It is worth noting that in previous works [7],[9] based on the same data set the presence of bare soil class



(a) (b)

Figure 5. Band 4 of the multispectral Landsat-5 TM images: (a) September 1995 (Y); and (b) July 1996 (Z).

was neglected by removing the related areas from the training and test sets. This choice practically allowed us to work under the simplifying assumption that $\Omega_2=\Omega_1$. In the following, this simple case will not be documented. The reader is referred to [7],[9] for details on the effectiveness of the method under this hypothesis.

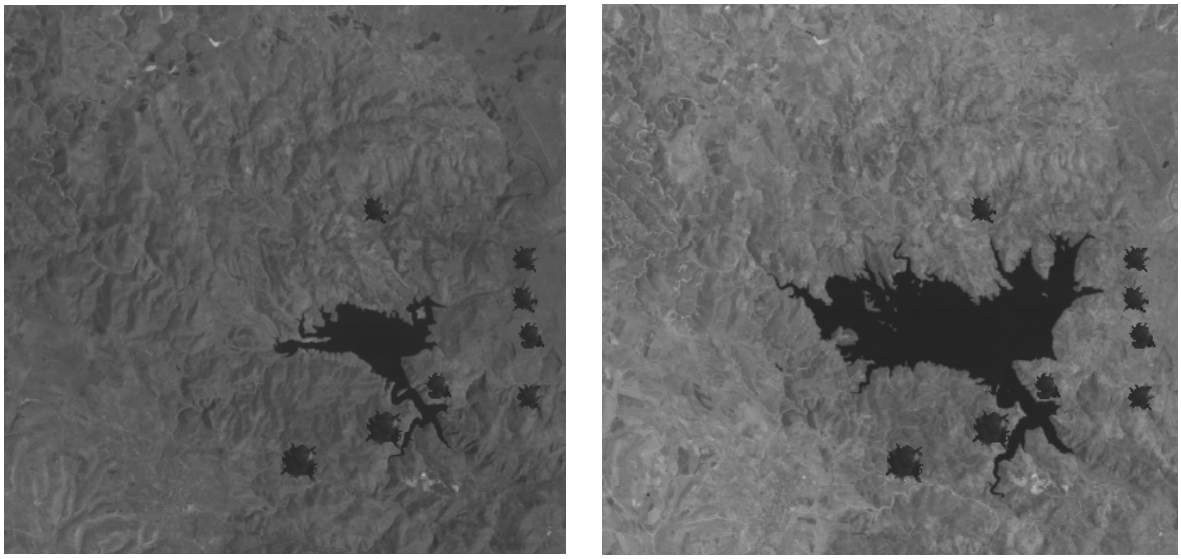
In order to test the proposed domain adaptation method in various situations and with growing complexity conditions, we have adopted two strategies. First, we observed that by reversing the role of images (using **Z** as $\mathbf{X}_1$ and vice versa), we can simulate also the case of addition of one class (*i.e.*, bare soil). Second, to simulate more complex situations with multiple addition/deletion of classes, the spectral signature of changes due to an event of forest fire was included in homogeneous regions of both images. In practice, this was done by considering a data set made up of two multispectral images acquired by the Landsat-5 TM multispectral sensor on the Island of Elba (Italy) in August 1994 and September 1994 [15]. Between these two acquisitions, a wildfire destroyed a large part of the vegetation in the north-west part of the island. On the basis of the available ground truth for Elba data set, areas affected from the fire were isolated and the digital numbers of the related spectral channels were inserted in the Sardinia Island one. Two different images were generated:

- Image **Y'** (Figure 6.a) is obtained by introducing the spectral signature of burned areas (ω_7), taken from the Elba dataset, in forest areas of image **Y**, and the corresponding reference information for training and test is also generated.
- Image **Z'** (Figure 6.b) is obtained by introducing the spectral signature of burned areas (ω_7), taken from the Elba dataset, in forest areas of image **Z**, and the corresponding reference information for training and test is also generated.

Images **Y'** and **Z'** allowed us the simulation of addition and deletion of a further class represented by burned areas.

Table I lists the number of training and test patterns available for each original and simulated

image.



(a) (b)

Figure 6. Band 4 of the multispectral Landsat-5 TM images with simulated burned areas: (a) September 1995 (Y') and (b) July 1996 image (Z').

TABLE I. TRAINING AND TEST SETS FOR THE ORIGINAL AND THE MODIFIED IMAGES.

Land–cover classes	September 1995				July 1996			
	Y		Y'		Z		Z'	
	Train set	Test set	Train set	Test set	Train set	Test set	Train set	Test set
Pasture (ω_1)	554	589	554	589	554	589	554	589
Forest (ω_2)	304	274	128	159	304	274	128	159
Urban Area (ω_3)	408	418	408	418	408	418	408	418
Water (ω_4)	804	551	804	551	1120	867	1120	867
Vineyard (ω_5)	179	117	179	117	179	117	179	117
Bare soil (ω_6)	316	316	316	316	-	-	-	-
Burned Area (ω_7)	-	-	176	115	-	-	176	115
Total	2565	2265	2565	2265	2565	2265	2565	2265

In the following, images **Y** and **Z** and their modifications (**Y'** and **Z'**) are alternately used as $\mathbf{X}_1$ and $\mathbf{X}_2$ in order to simulate a variety of scenarios. In each experiment, we assume that only the training set associated with the image considered as $\mathbf{X}_1$ is available, whereas the training set for $\mathbf{X}_2$ is not available. Test sets for both $\mathbf{X}_1$ and $\mathbf{X}_2$ are only used for performance evaluation, but they are not involved in the processing. It is worth stressing that in this controlled experimental set up, we know a priori the land-cover classes that appeared and/or disappeared between the two acquisitions. Therefore, in the following, for the sake of readiness, we will refer to bare soil (ω_6) and burned areas (ω_7) classes

explicitly. However this kind of information is usually not available in real applications and is not used by the proposed method.

The quantitative evaluation of DA-MAP is documented in terms of user accuracy (or precision), producer accuracy (or recall) and overall accuracy computed according to the test set available for image $\mathbf{X}_2$.

B. Description of Experiments

To verify the robustness and effectiveness of proposed method, the following experiments are carried out on the medium resolution dataset.

Addition of Class(es)

Here we propose two different scenarios: the first one represents the simple case of the addition of one single class (CASE B), while the second one represents a more complex scenario in which two new classes have appeared between $\mathbf{X}_1$ and $\mathbf{X}_2$ (CASE E). These scenarios are simulated by considering $\Omega_1=\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5\}$ and by adding:

- *Experiment* 1: the bare soil class to Ω_1, *i.e.*, $\omega_u=\omega_6$, $\Omega_2=\Omega_1\cup\{\omega_6\}$, $U=1$;
- *Experiment* 2: the bare soil and burned area classes $\Omega_U=\{\omega_6,\omega_7\}$ to Ω_1, $\Omega_2=\Omega_1\cup\{\omega_6,\omega_7\}$, $U=2$;

Table II documents the image pairs involved for the simulations of the two scenarios.

Deletion of Class(es)

Also in this case, two scenarios are proposed. The first one is associated to the deletion of one single class (CASE C), whereas the second one is associated to the deletion of two classes (CASE E). The scenarios are simulated by:

- *Experiment* 3: considering $\Omega_1=\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6\}$ and deleting the bare soil class $\omega_k=\omega_6$ from Ω_1, *i.e.*, $\Omega_2=\Omega_1\setminus\{\omega_6\}$, $K=1$;

- *Experiment* 4: considering $\Omega_1=\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6,\omega_7\}$ and deleting the bare soil and burned area classes $\Omega_K=\{\omega_6,\omega_7\}$ from Ω_1, *i.e.*, $\Omega_2=\Omega_1\backslash\{\omega_6,\omega_7\}$, K=2;

Table II documents the image pairs involved for the simulations of the two scenarios. As one can observe, multi-temporal datasets for this case are generated by just swapping the images used for scenarios in CASE B.

Simultaneous Addition and Deletion of Classes

CASE E represents a more complex situation: a combination of scenarios taken from CASES *B* and C. In greater detail, we have the addition of one class and simultaneously also the deletion of one class between images $\mathbf{X}_1$ and $\mathbf{X}_2$. We simulated these scenarios by:

- *Experiment* 5: considering $\Omega_1=\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_7\}$ and by both adding the bare soil class ($\omega_u=\omega_6$) to Ω_1 and deleting the burned area class ($\omega_k=\omega_7$) from Ω_1, *i.e.*, $\Omega_2=\Omega_1\cup\{\omega_6\}\backslash\{\omega_7\}$, K=1 and U=1;
- *Experiment* 6: considering $\Omega_1=\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6\}$ and by both deleting the bare soil class ($\omega_k=\omega_6$) from Ω_1 and adding the burned area class ($\omega_u=\omega_7$) to Ω_1, *i.e.*, $\Omega_2=\Omega_1\backslash\{\omega_6\}\cup\{\omega_7\}$, K=1 and U=1;

Table II documents the image pairs involved for the simulations of the two scenarios and the number of classes involved for $\mathbf{X}_1$ and $\mathbf{X}_2$.

TABLE II. SUMMARY OF THE EXPERIMENTAL SETUPS FOR THE LANDSAT DATASET.

Case	Experiment	Case description	$\mathbf{X}_1$	$\mathbf{X}_2$	# of classes	
					K_1	K_2
B	1	Addition bare soil class	**Z**	**Y**	5	6
	2	Addition of burned area and bare soil classes	**Z**	**Y'**	5	7
C	3	Deletion bare soil class	**Y**	**Z**	6	5
	4	Deletion of burned area and bare soil classes	**Y'**	**Z**	7	5
E	5	Deletion bare soil class and addition of burned area class	**Y**	**Z'**	6	6
	6	Deletion burned area class and addition of bare soil class	**Z'**	**Y**	6	6

C. Results

To carry out the experiments described above, we assumed that for each multi-temporal image pair only training set associated with image considered as $\mathbf{X}_1$ is available, whereas, the reference ground data associated with the image considered as $\mathbf{X}_2$ is used only to evaluate the performance of the proposed technique.

Experiment 1*: Addition of One Class* (*CASE B*)*.* In order to apply the proposed method, at first the supervised MAP classification is obtained for image $\mathbf{X}_1$ (*i.e.*, $\mathbf{Z}$) and the set of parameters θ_1 which characterizes the probability density functions of the classes in $\Omega_1=\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5\}$ is estimated. The supervised classification of $\mathbf{X}_1$ leads to an overall accuracy of 93.64% on the test set. In order to compute the desired classification map of image $\mathbf{X}_2$ (*i.e.*, $\mathbf{Y}$), we need to understand how to properly initialize the set θ_2^0. To this end we use the proposed method.

We apply the CVA in polar coordinates to images $\mathbf{X}_1$ and $\mathbf{X}_2$ in order to formulate a proper hypothesis according to the procedure proposed in Sec. IV.A. For this data set, bands 4 and 7 of the TM sensors were used for change detection purposes. CVA resulted in a threshold value T along ρ equal to 31, and in 2 angular thresholds along the direction variable ϑ equal to 28° and 55°, respectively. The reader is referred to [17] for further details on how the change-detection step has been performed. The automatic comparison between the classification map of $\mathbf{X}_1$ and the change detection map pointed out that the change affected only partially the water class (ω_4). After that, standard adaptation of the MAP classifier to $\mathbf{X}_2$ is applied and the pairwise JM distances between each class in $\mathbf{X}_2$ and changed pixels in image $\mathbf{X}_2$ is computed according to (8). As one can see from Table III, all distance measures resulted in large values (much greater than the threshold value $\mathbf{T}_H$ =1.27 which is 90% of the saturation value) for each possible land-cover class already present in Ω_1. Based on this result, the hypothesis of addition of one class (CASE B) is formulated. θ_2^0 is initialized according to (11) and the domain adaptation procedure is applied. For comparison and validation

purposes also the hypothesis that $\Omega_2=\Omega_1$, (*i.e.*, neglecting the possibility that classes could change between the two acquisitions) is tested. Once the two models are available, the BIC [26] is used for hypothesis validation. The 6-class model resulted in a BIC_6 = -1.7776e+005 whereas the 5-class model resulted in BIC value greater than the 6-class model. Thus, the hypothesis of class addition is validated.

TABLE III. JM DISTANCES COMPUTED BETWEEN THE DETECTED CHANGED PIXELS AND CLASSES IN X_2 (EXPERIMENT 1, ADDITION OF BARE SOIL CLASS, LANDSAT DATASET).

Land-Cover Class	JM
Pasture (ω_1)	1.3709
Forest (ω_2)	1.4129
Urban Area (ω_3)	1.3052
Water (ω_4)	1.4142
Vineyard (ω_5)	1.3685

A further evidence of the correctness of the addition of class hypothesis and of the effectiveness of the proposed domain adaptation procedure comes from the analysis of the overall accuracy evaluated on the ground truth available for $\mathbf{X}_2$.[3] The overall accuracy obtained with the 6-class model is of about 87.95% (see Table IV). This value is very close to the one that can be obtained by the supervised MAP classification of $\mathbf{X}_2$ using available training samples, *i.e.*, 90.38% (which can be considered an upper bound for the proposed procedure).

TABLE IV. USER, PRODUCER AND OVERALL ACCURACY (%) ON THE TEST SET FOR IMAGE X_2 OBTAINED WITH THE PROPOSED DOMAIN ADAPTATION TECHNIQUE AND THE SUPERVISED MAP CLASSIFICATION (EXPERIMENT 1, ADDITION OF BARE SOIL CLASS, LANDSAT DATASET).

Land-cover class	Proposed Technique		Supervised MAP classification	
	User	Producer	User	Producer
Pasture (ω_1)	77.08	91.45	82.17	91.15
Forest (ω_2)	96.72	96.02	97.45	90.20
Urban Area (ω_3)	87.32	79.00	94.74	83.19
Water (ω_4)	100.00	100.00	100.00	100.00
Vineyard (ω_5)	70.09	59.42	62.39	77.66
Bare Soil (ω_6)	87.03	79.25	87.34	87.07
Overall Accuracy	87.95		90.38	

Experiment 2: Addition of Two Classes (*CASE E*). In order to better assess the effectiveness of the proposed domain adaptation procedure in the case of class addition, a more complex situation in which two new classes are present in the second image is considered. The same trials described for CASE B

[3] It is important to recall that available test patterns were used only for validation purposes, but they were not involved in the domain adaptation procedure.

were carried out in the simulated case of addition of both bare soil (ω_6) and burned area (ω_7) classes (see simulated dataset description in Sec. V.A). The supervised MAP classifier trained on $\mathbf{X}_1$ resulted in a set of parameters θ_1 that gives an overall accuracy on the test set of 93.64%. Using the proposed approach to find the classification map for $\mathbf{X}_2$, we applied CVA to images $\mathbf{X}_1$ and $\mathbf{X}_2$. The representation within the polar domain of the considered change-detection problem pointed out the presence of two clusters in A_c. Thus two annular sectors could be identified, each of them corresponding to a different kind of change. The first one corresponds to the appearance of the bare soil class (ω_6) due to the decrease of the extension of the lake surface. The reader is referred to Experiment 1 for an analysis of this kind of change and to Table III for the pair-wise distances between this changed area and classes in $\mathbf{X}_2$. The second cluster in A_c includes pixels with a magnitude higher than T=31 and direction values in $[323°,360°] \cup [0°,28°)$. As for the previous case, also in this one it was detected that only a part of the forest class experienced a change. The analysis of statistical distances of changed pixels from the already known classes present in $\mathbf{X}_2$ resulted in a set of high values (see Table V) suggesting the presence of a new class, namely the burned area class (ω_7). According to this analysis the final hypothesis is of addition of two classes.

TABLE V. JM DISTANCES COMPUTED BETWEEN THE DETECTED CHANGED PIXELS AND CLASSES IN X_2 (EXPERIMENT 2, ADDITION OF BURNED AREA CLASS, LANDSAT DATASET).

Land Cover Class	JM distance
Pasture (ω_1)	1.4142
Forest (ω_2)	1.4137
Urban Area (ω_3)	1.4137
Water (ω_4)	1.4142
Vineyard (ω_5)	1.4142

According to the formulated hypothesis, (11) is applied for both kinds of change initializing θ_2^0 with $\Omega_U=\{\omega_6,\omega_7\}$, and the unsupervised adaptation is performed to obtain the new classifier for $\mathbf{X}_2$. Also in this case BIC is used to validate the hypothesis on the number of classes present in $\mathbf{X}_2$. As in the previous case, the model with 7 classes is compared with the one with 5 classes (*i.e.*, no new class

detected), and with two models with 6 classes (*i.e.*, only one new class out of two detected). Again the hypothesis formulated according to the proposed approach (*i.e.*, the 7-class model) resulted to be the best one as it was associated with the lowest BIC value.

The effectiveness of the proposed method is further confirmed by an analysis of the accuracies evaluated on the ground truth of $\mathbf{X}_2$ (see Table VI). The classifier obtained with the proposed method (7-class model) resulted in an 88.12% of overall accuracy, which is very close to the one obtained by the supervised MAP classifier trained on $\mathbf{X}_2$ (assuming the availability of a training set) which is equal to 90.33%. This is a relevant result, especially remarking that the proposed method did not use any ground truth information on $\mathbf{X}_2$.

TABLE VI. USER, PRODUCER AND OVERALL ACCURACY (%) ON THE TEST SET FOR IMAGE X_2 OBTAINED WITH THE PROPOSED DOMAIN ADAPTATION TECHNIQUE AND THE SUPERVISED MAP CLASSIFICATION (EXPERIMENT 2, ADDITION OF BARE SOIL AND BURNED AREA CLASSES, LANDSAT DATASET).

Land-cover class	Proposed Technique		Supervised MAP classification	
	User	Producer	User	Producer
Pasture (ω_1)	76.91	91 33	82.17	89.80
Forest (ω_2)	97.48	75.24	95.60	85.39
Urban Area (ω_3)	89.95	87.24	94.26	83.83
Water (ω_4)	100.00	100.00	100.00	100.00
Vineyard (ω_5)	61.54	70.59	62.39	78.50
Bare Soil (ω_6)	89.24	75.81	87.66	86.83
Burned Area (ω_7)	93.04	100.00	100.00	100.00
Overall Accuracy	88.12		90.33	

Experiment 3: Deletion of One Class (*CASE C*)**.** Experiments on class deletion are preformed similarly to the ones on class addition. The supervised MAP classifier trained on $\mathbf{X}_1$ (*i.e.*, $\mathbf{Y}$) resulted in a set of parameters θ_1 and an overall accuracy of 90.38%. The change-detection step applied to this specific case gives the same decision boundaries and thus change-detection map as the case of one-class addition. This is because the two trials involve the same images $\mathbf{Y}$ and $\mathbf{Z}$ by only exchanging their role, *i.e.*, $\mathbf{Y}=\mathbf{X}_1$ and $\mathbf{Z}=\mathbf{X}_2$ instead of $\mathbf{Y}=\mathbf{X}_2$ and $\mathbf{Z}=\mathbf{X}_1$ (see Table II), but CVA is not sensitive to image order. The comparison of the change-detection map with the classification map of $\mathbf{X}_1$ this time points out that all the area associated to the bare soil class (ω_6) is affected by the change event and therefore

ω_6 does not appear in $\mathbf{X}_2$. To understand whether the land-cover transition resulted in a new class or in a class already existing in Ω_2, statistical distances between classes in $\mathbf{X}_2$ (after standard retraining) and the changed area in image $\mathbf{X}_2$ are computed.

As one can see from Table VII, the area of change shows a high distance with respect to all classes but one: the water class. The JM distance with water class is much smaller than $\mathbf{T}_L$=0.99 (which is the 70% of the saturation value). This suggests that changed pixels belong to a class (namely water class) already existing in $\mathbf{X}_2$. Accordingly, the proposed method formulates the hypothesis of deletion of class ω_6. Therefore θ_2^0 is initialized according to (12) and unsupervised retraining is performed.

TABLE VII. JM DISTANCES COMPUTED BETWEEN THE DETECTED CHANGED PIXELS AND CLASSES IN X_2 (EXPERIMENT 3, DELETION OF BARE SOIL CLASS, LANDSAT DATASET).

Land Cover Class	JM distance
Pasture (ω_1)	1.4142
Forest (ω_2)	1.4142
Urban Area (ω_3)	1.4142
Water (ω_4)	0.6460
Vineyard (ω_5)	1.4142

TABLE VIII. USER, PRODUCER AND OVERALL ACCURACY (%) ON THE TEST SET FOR IMAGE X_2 OBTAINED WITH THE PROPOSED DOMAIN ADAPTATION TECHNIQUE AND THE SUPERVISED MAP CLASSIFICATION (EXPERIMENT 3, DELETION OF BARE SOIL CLASS, LANDSAT DATASET).

Land-cover class	Proposed Technique		Supervised MAP classification	
	User	Producer	User	Producer
Pasture (ω_1)	94.06	89.77	92.02	89.59
Forest (ω_2)	87.23	97.16	92.70	88.81
Urban Area (ω_3)	93.06	96.53	93.30	93.08
Water (ω_4)	99.89	100.00	99.89	100.00
Vineyard (ω_5)	64.10	55.97	58.97	77.53
Bare Soil (ω_6)	-	-	-	-
Overall Accuracy	93.73		93.64	

The validation of the hypothesis is carried out by comparing the BIC value obtained assuming that Ω_2=Ω_1 (6-class model) with the one obtained under the hypothesis formulated according to the proposed method (5-class model). The first model resulted in a higher BIC value than the second one, thus validating the assumption of the proposed method.

The confirmation of the effectiveness of the proposed procedure comes from the analysis of the

overall accuracy obtained with the supervised MAP classifier trained on $\mathbf{X}_2$ on the test set available for $\mathbf{X}_2$ (*i.e.*, 93.64%) and the one obtained at convergence by the proposed classifier, *i.e.*, 93.73%.

Experiment 4: Deletion of Two Classes (*CASE E*)**.** This experiment has been carried out by inverting the role of images of Experiment 2 (addition of two classes). The supervised MAP classification of $\mathbf{X}_1$ results in set of parameters θ_1 that achieve an overall accuracy of about a 90.33%. After that, the application of CVA to the two images pointed out the occurrence of two kinds of change. The first detected change corresponds to the one in Experiment 3; thus, according to Table VII, the hypothesis of class deletion is formulated. Concerning the second change, the comparison with the classification map of $\mathbf{X}_1$ pointed out that all the burned area experienced a change. Statistical distances between changed pixels in $\mathbf{X}_2$ and already known classes in $\mathbf{X}_2$ (see Table IX) show that changed pixels exhibit a high distance from all classes in $\mathbf{X}_2$ except one (*i.e.*, the forest class). The distance with the forest class is much smaller than T_L=0.99. Accordingly, the formulated hypothesis is of two-class deletion, and θ_2^0 is initialized as in (12) for both kinds of change.

TABLE IX. JM DISTANCES COMPUTED BETWEEN THE DETECTED CHANGED PIXELS AND CLASSES IN X_2 (EXPERIMENT 4, DELETION OF BURNED AREA CLASS, LANDSAT DATASET).

Land Cover Class	JM distance
Pasture (ω_1)	1.3561
Forest (ω_2)	0.5935
Urban Area (ω_3)	1.4106
Water (ω_4)	1.4142
Vineyard (ω_5)	1.4110

After unsupervised retraining, also in this case the analysis of the BIC values confirmed the validity of the formulated hypothesis.

Finally the effectiveness of the proposed method is also pointed out by the overall accuracies computed on the test set available for $\mathbf{X}_2$. The proposed method resulted in an accuracy of 94.22%, which is even slightly better than that of the supervised MAP classification of $\mathbf{X}_2$ which results in an overall accuracy of 93.64%. This can be explained by the capability of the EM algorithm to exploit the

information present in the unlabeled pixels of $\mathbf{X}_2$.

TABLE X. USER, PRODUCER AND OVERALL ACCURACY (%) ON THE TEST SET FOR IMAGE X_2 OBTAINED WITH THE PROPOSED DOMAIN ADAPTATION TECHNIQUE AND THE SUPERVISED MAP CLASSIFICATION (EXPERIMENT 4, DELETION OF BARE SOIL AND BURNED AREA CLASSES, LANDSAT DATASET).

Land-cover class	Proposed Technique		Supervised MAP classification	
	User	Producer	User	Producer
Pasture (ω_1)	95.42	89.21	92.02	89.59
Forest (ω_2)	88.32	96.41	92.70	88.81
Urban Area (ω_3)	91.87	97.46	93.30	93.08
Water (ω_4)	99.89	100.00	99.89	100.00
Vineyard (ω_5)	68.38	64.52	58.97	77.53
Bare Soil (ω_6)	-	-	-	-
Burned Area (ω_7)	-	-	-	-
Overall Accuracy	94.22		93.64	

Simultaneous Addition and Deletion of Classes: CASE E. The last set of experiments aims at validating the effectiveness of the proposed approach in a complex scenario in which one class disappeared and a new one appeared between the two acquisition dates. Under such a condition, the total number of classes remains the same at both the time instances but the set of classes are different. To study the scenario of simultaneous addition and deletion of classes, two experiments were carried out. Both of them are a combination of the four experiments described above. Therefore we will not discuss how hypotheses are formulated and validated.

Experiment 5: Addition of Burned area Class and Deletion of Bare Soil Class. Initially the MAP classifier is trained with the supervised technique on the $\mathbf{X}_1$ image. The overall accuracy obtained by using this classifier on test set of $\mathbf{X}_1$ is 90.38%. The comparison of the change detection map and the classification map of $\mathbf{X}_1$, and the study of the statistical JM distances (see Table XI) allow us to formulate the following hypotheses: i) one added class *i.e.*, burned area (Change 1), and ii) one deleted class *i.e.*, bare soil (Change 2). The initial estimates θ_2^0 of class statistical parameters were defined accordingly to this hypothesis and based on (11) and (12). After adaptation, a classifier for $\mathbf{X}_2$ was defined. The obtained statistical model of classes was validated using BIC. Here BIC values were computed: i) assuming that $\Omega_2=\Omega_1$; ii) considering that only the deletion of one class has been detected;

and iii) considering the model formulated with the proposed method. Also in this case the minimum BIC value was obtained for the hypothesis formulated with the proposed domain adaptation method. This confirms the validity of the method itself. The MAP classifier obtained at convergence resulted in a classification accuracy of 94.04% which is better than that obtained with a supervised MAP classifier trained on $\mathbf{X}_2$ (*i.e.*, 93.07%) (see Table XII).

TABLE XI. JM DISTANCES COMPUTED BETWEEN THE DETECTED CHANGED PIXELS AND CLASSES IN X_2 (EXPERIMENT 5, ADDITION OF BURNED AREA CLASS AND DELETION OF BARE SOIL CLASS, LANDSAT DATASET).

Land Cover Class	JM distance	
	Change 1	Change 2
Pasture (ω_1)	1.4142	1.4142
Forest (ω_2)	1.4142	1.4142
Urban Area (ω_3)	1.4142	1.4142
Water (ω_4)	0.6460	1.4142
Vineyard (ω_5)	1.4142	1.4142

TABLE XII. USER, PRODUCER AND OVERALL ACCURACY (%) ON THE TEST SET FOR IMAGE X_2 OBTAINED WITH THE PROPOSED DOMAIN ADAPTATION TECHNIQUE AND THE SUPERVISED MAP CLASSIFICATION (EXPERIMENT 5, ADDITION OF BURNED AREA CLASS AND DELETION OF BARE SOIL CLASS, LANDSAT DATASET).

Land-cover class	Proposed Technique		Supervised MAP classification	
	User	Producer	User	Producer
Pasture (ω_1)	96.94	87.85	91.85	87.68
Forest (ω_2)	85.54	86.62	81.13	80.12
Urban Area (ω_3)	91.15	97.19	93.30	92.64
Water (ω_4)	99.65	100.00	99.65	100.00
Vineyard (ω_5)	60.68	74.74	58.97	79.31
Bare Soil (ω_6)	-	-	-	-
Burned Area (ω_7)	93.04	100.00	100.00	100.00
Overall Accuracy	94.04		93.07	

Experiment 6: Deletion of Burned Area Class and Addition of Bare Soil Class. This experiment is the complementary of the previous one. Supervised MAP classification on $\mathbf{X}_1$ resulted in an overall accuracy of 93.07%. According to the proposed method and in accordance with JM distances in Table XIII, the hypotheses of addition of one class, *i.e.*, bare soil class (Change 3) and deletion of one class, *i.e.*, burned area class (Change 4) is formulated. The set of parameters θ_2^0 are initialized using (11) and (12), and adaptation is performed. The effectiveness of the new class model is confirmed by BICs analysis. Accuracies in Table XIV also confirm the effectiveness of the proposed technique.

TABLE XIII. JM DISTANCES COMPUTED BETWEEN THE DETECTED CHANGED PIXELS AND CLASSES IN X_2 (EXPERIMENT 6, DELETION OF BURNED AREA CLASS AND ADDITION OF BARE SOIL CLASS, LANDSAT DATASET).

Land Cover Class	JM distance	
	Change 3	Change 4
Pasture (ω_1)	1.3700	1.3996
Forest (ω_2)	1.4109	0.6955
Urban Area (ω_3)	1.3084	1.4097
Water (ω_4)	1.4142	1.4142
Vineyard (ω_5)	1.3641	1.4098

TABLE XIV. USER, PRODUCER AND OVERALL ACCURACY (%) ON THE TEST SET FOR IMAGE X_2 OBTAINED WITH THE PROPOSED DOMAIN ADAPTATION TECHNIQUE AND THE SUPERVISED MAP CLASSIFICATION (EXPERIMENT 6, DELETION OF BURNED AREA CLASS AND ADDITION OF BARE SOIL CLASS, LANDSAT DATASET).

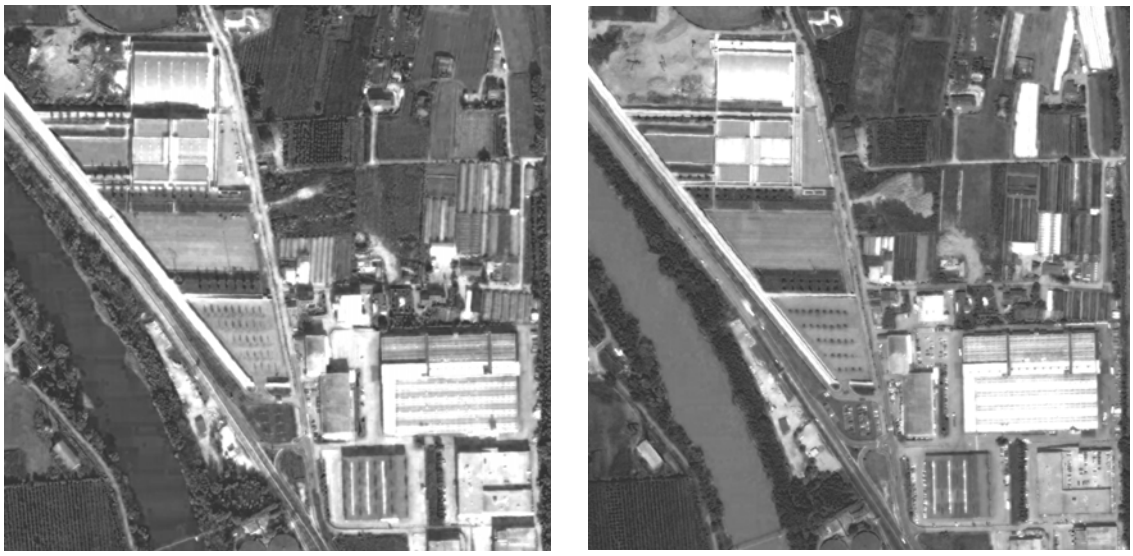
Land-cover class	Proposed Technique		Supervised MAP classification	
	User	Producer	User	Producer
Pasture (ω_1)	76.57	91.48	82.17	91.15
Forest (ω_2)	96.72	96.015	97.44	90.20
Urban Area (ω_3)	87.32	79.52	94.74	83.19
Water (ω_4)	100.00	100.00	100.00	100.00
Vineyard (ω_5)	70.09	59.42	62.39	77.66
Bare Soil (ω_6)	87.03	79.02	87.34	87.07
Burned Area (ω_7)	-	-	-	-
Overall Accuracy	87.82		90.38	

VI. EXPERIMENTAL RESULTS: QUICKBIRD DATA SET

A. Data Set Description and Experimental Set-Up

The second data set is made up of two co-registered and pansharpened multispectral Very High geometrical Resolution (VHR) images acquired by the QuickBird satellite. All the 4 spectral bands of QuickBird were considered in the experiments. The selected test site is a section of 1520×1504 pixels of a scene including an area on the south part of the city of Trento (Italy). The two images were acquired in October 2005 (**Y**) and July 2006 (**Z**), respectively (see Figure 7). The available prior information about the considered area was used to build a training set and a test set for each image. Five main land-cover classes common to the two considered dates were identified, *i.e.*, water (ω_1), red roof (ω_2), asphalt (ω_3), fields (ω_4), and bare soil (ω_5). For the image acquired in July 2006 one additional class has been detected, *i.e.*, plastic-mulched fields (ω_6). Therefore, if we consider **Y** as $\mathbf{X}_1$ and **Z** as $\mathbf{X}_2$ (as temporally logic in the considered data set) a new class (*i.e.*, plastic-mulched fields) appeared between the two acquisitions. Table XV lists the number of training and test patterns

available for each original and simulated image. In addition the deletion of the plastic-mulched fields class was simulated by reversing the role of images (using $\mathbf{Z}$ as $\mathbf{X}_1$ and vice versa). To verify the robustness and effectiveness of proposed method, the following experiments are carried out on the VHR dataset.



(a) (b)

Figure 7 Band 1 of the multispectral QuickBird VHR images: (a) image acquired in October 2005 (Y); and (b) image acquired in July 2006 (Z).

TABLE XV. TRAINING AND TEST SETS FOR THE QUICKBIRD DATASET.

Land–cover classes	October 2005 (Y)		July 2006 (Z)	
	Training set	Test set	Training set	Test set
Water (ω_1)	1099	1104	1099	1104
Red roof (ω_2)	449	469	449	469
Asphalt (ω_3)	673	474	673	474
Fields (ω_4)	647	534	647	534
Bare soil in field (ω_5)	1017	773	600	483
Plastic-mulched field (ω_6)	-	-	417	290
Total	3885	3354	3885	3354

Addition of One Class: CASE B

For this data set, only one scenario is proposed for CASE B: addition of a single class between $\mathbf{X}_1$ and $\mathbf{X}_2$. In this scenario, $\Omega_1=\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5\}$ and the new plastic-mulched field class is present in $\mathbf{X}_2$, *i.e.*, $\omega_u=\omega_6$, $\Omega_2=\Omega_1\cup\{\omega_6\}$ (Experiment 7).

Deletion of One Class: CASE C

Also in this case, one single scenario is proposed associated to the deletion of a class. In this

scenario, $\Omega_1=\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6\}$ and the plastic-mulched field class is deleted between the two acquisitions, *i.e.*, $\omega_k=\omega_6$ and $\Omega_2=\Omega_1\backslash\{\omega_6\}$ (Experiment 8).

Table XVI summarizes the image pairs involved for the simulations of the two scenarios and the number of classes for $\mathbf{X}_1$ and $\mathbf{X}_2$.

TABLE XVI. SUMMARY OF THE EXPERIMENTAL SETUPS FOR QUICKBIRD DATASET.

Case	Experiment	Case description	$\mathbf{X}_1$	$\mathbf{X}_2$	# of classes	
					K_1	K_2
B	7	Addition plastic-mulched field class	**Y**	**Z**	5	6
C	8	Deletion plastic-mulched field class	**Z**	**Y**	6	5

B. Results

As in the experiments for the Landsat dataset, also here we assumed that only the training set associated with image considered as $\mathbf{X}_1$ is available, whereas the training set associated with the image $\mathbf{X}_2$ is used only for validation.

Experiment 7: Addition of One Class (*CASE B*). Initially, the supervised MAP classification is applied to image $\mathbf{X}_1$ (*i.e.*, **Y**) and the set of parameters θ_1 which characterizes the probability density functions of the classes in Ω_1 is estimated. The supervised classification of $\mathbf{X}_1$ leads to an overall accuracy of 97.29% on the test set. In order to adapt the classifier obtained for $\mathbf{X}_1$ to the statistical properties of $\mathbf{X}_2$ (*i.e.*, **Z**), we applied the proposed method. Channels 3 and 4 of the QuickBird images were selected for the change-detection step. CVA resulted in the threshold value T along ρ equal to 300, and in 2 angular thresholds along the direction variable ϑ equal to 0° and 63°, respectively. (The reader is referred to [17] and [27] for further details on how CVA can be applied to VHR data). The unsupervised and automatic analysis of the information in the polar domain led to the identification of a single kind of change. The comparison with the supervised classification map of $\mathbf{X}_1$ pointed out that only part of the fields class has changed. The pair-wise JM distances between the spectral signatures of changed pixels in $\mathbf{X}_2$ and each class in $\mathbf{X}_2$ already present in Ω_2 (after standard adaptation of the MAP classifier) are computed according to (8). All the distance measures result in large values, much greater

than the threshold value $\mathbf{T}_H$ (see Table XVII). Accordingly, the hypothesis of addition of one class (CASE B) is formulated. The set θ_2^0 is initialized using (11) and the domain adaptation procedure is applied. As for the Landsat dataset, the hypothesis that $\Omega_2=\Omega_1$ (*i.e.*, neglecting the possibility that classes could change between the two acquisitions) is also tested. Once the two models are available, the BIC [26] is used for hypothesis validation. The 6-class model resulted in BIC value smaller than the 5-class model. Thus, the hypothesis of class addition is validated.

TABLE XVII. JM DISTANCES COMPUTED BETWEEN THE DETECTED CHANGED PIXELS AND CLASSES IN $\mathbf{X}_2$ (EXPERIMENT 7, ADDITION OF PLASTIC-MULCHED FIELD CLASS, QUICKBIRD DATASET).

Land-Cover Class	JM
Water (ω_1)	1.4142
Red roof (ω_2)	1.3868
Asphalt (ω_3)	1.3878
Fields (ω_4)	1.4142
Bare soil (ω_5)	1.4132

TABLE XVIII. USER, PRODUCER AND OVERALL ACCURACY (%) ON THE TEST SET FOR IMAGE $\mathbf{X}_2$ OBTAINED WITH THE PROPOSED DOMAIN ADAPTATION TECHNIQUE AND THE SUPERVISED MAP CLASSIFICATION (EXPERIMENT 7, ADDITION OF PLASTIC-MULCHED FIELDS CLASS, QUICKBIRD DATASET).

Land-cover class	Proposed Technique		Supervised MAP classification	
	User	Producer	User	Producer
Water (ω_1)	96.52	100.00	96.52	100.00
Red roof (ω_2)	100.00	82.43	97.44	94.23
Asphalt (ω_3)	94.52	99.12	50.00	98.34
Fields (ω_4)	100.00	98.71	94.76	98.64
Bare soil (ω_5)	78.06	99.74	50.52	99.19
Plastic-mulched fields (ω_6)	100.00	83.33	100.00	36.12
Overall Accuracy	94.91		83.42	

A further evidence of both the correctness of the formulated hypothesis and the effectiveness of the proposed domain adaptation procedure come from the analysis of the overall accuracy evaluated on the test set available for $\mathbf{X}_2$. The overall accuracy obtained with the 6-class model is of about 94.91% (see Table XVIII). This value is higher compared to the one that can be obtained by the supervised MAP classification of $\mathbf{X}_2$ using available training samples, *i.e.*, 83.42%.

Experiment 8: Deletion of One Class (*CASE C*). Similar steps as in Experiment 7 were conducted for the case of deletion of a class. The supervised MAP classifier trained on $\mathbf{X}_1$ (*i.e.*, $\mathbf{Z}$) results in an

overall accuracy of 83.42%. The change-detection step applied to this specific case gave the same decision boundaries and thus change-detection map as the previous one. The comparison with the classification map obtained at $\mathbf{X}_1$ showed that all the pixels in ω_6 were affected by the change. The statistical distances (see Table XIX) between changed pixels in $\mathbf{X}_2$ and classes in $\mathbf{X}_2$ (after adaptation) were all high but the one related to the bare soil class was smaller than $\mathbf{T}_L$. This suggests that the changed pixels belong to a class (namely bare soil class) already existing in $\mathbf{X}_2$. Accordingly, the proposed method formulates the hypothesis of deletion of class ω_6. Therefore θ_2^0 is initialized according to (12) and adaptation is performed.

TABLE XIX JM DISTANCES COMPUTED BETWEEN THE DETECTED CHANGED PIXELS AND CLASSES IN X_2 (EXPERIMENT 8, DELETION OF PLASTIC-MULCHED FIELDS CLASS, QUICKBIRD DATASET).

Land Cover Class	JM distance
Water (ω_1)	1.4142
Red roof (ω_2)	1.4055
Asphalt (ω_3)	1.4142
Fields (ω_4)	1.4139
Bare soil (ω_5)	0.9328

The validation of the hypothesis is carried out by comparing the BIC value obtained assuming that $\Omega_2=\Omega_1$ (6-class model) with the one obtained under the hypothesis formulated according to the proposed method (5-class model). The first model resulted in a higher BIC value than the second one, thus validating our assumption.

TABLE XX. USER, PRODUCER AND OVERALL ACCURACY (%) ON THE TEST SET FOR IMAGE X_2 OBTAINED WITH THE PROPOSED DOMAIN ADAPTATION TECHNIQUE AND THE SUPERVISED MAP CLASSIFICATION (EXPERIMENT 8, DELETION OF PLASTIC-MULCHED FIELD CLASS, QUICKBIRD DATASET).

Land-cover class	Proposed Technique		Supervised MAP classification	
	User	Producer	User	Producer
Water (ω_1)	95.65	100.00	95.74	97.60
Red roof (ω_2)	99.79	65.00	99.36	92.10
Asphalt (ω_3)	55.06	100.00	100.00	98.55
Fields (ω_4)	100.00	98.34	95.51	98.65
Bare soil (ω_5)	98.45	98.32	97.80	98.57
Plastic-mulched fields (ω_u=ω_6)	-	-	-	-
Overall Accuracy	91.83		97.29	

The confirmation of the effectiveness of the proposed procedure comes from the analysis of the overall accuracy obtained with the supervised MAP classifier trained on $\mathbf{X}_2$ with the reference ground

data available for $\mathbf{X}_2$ (*i.e.*, 97.29%) and the one obtained at convergence after adaptation (*i.e.*, 91.83%).

VII. CONCLUSION

In this work a novel approach to domain adaptation in the context of classification of multi-temporal remote sensing images has been proposed. This approach addresses a very important problem related to the analysis of multitemporal images, which is connected with the need to update land-cover maps every time that a new image is available for a given geographical area. Unlike other techniques presented in the remote sensing literature, the proposed DA method can address problems in which the set of land-cover classes that model the target domain (image at t_1) is different from those that model the source domain (image at t_2). In greater detail, the method extends the use of the standard domain adaptation technique based on the Maximum-A-Posteriori classifier and Expectation-Maximization algorithm to cases in which the set of classes that characterize the two domains are potentially different. This is an important methodological and applicative improvement with respect to DA methods presented in the remote sensing literature that usually assume that the same classes are present in the source and target domains. Addition and/or deletion of class(es) is detected according to the joint use of change-detection methods and Jeffreys-Matusita (JM) statistical distance measure. The final model is validated according to the Bayesian Information Criterion. Proper strategies for initializing the adaptation step in the different cases are defined. The possibility to consider different sets of classes for source and target domains makes the proposed DA-MAP classifier highly flexible with respect to a number of possible (and common) applicative situations. The proposed method was tested within a set of scenarios showing different levels of complexity: i) addition or deletion of a single class; ii) addition or deletion of multiple classes; and iii) simultaneous addition and deletion of classes. For all setups and for different kind of remote sensing multispectral images (both medium and very high resolution) the proposed DA-MAP technique demonstrated to be highly effective resulting in classifications accuracies comparable with those achieved by a supervised MAP classifier without any training data

for the second image.

As future work, we plan to test the proposed method on image time series made up of more than two images for both further validating the presented DA classifier and for exploiting it in studies on the dynamic of land-covers of specific study areas. In addition, we plan to extend the proposed domain adaptation procedure to be used with kernel based classifiers and in particular with Support Vector Machines [7].

ACKNOWLEDGMENT

This work was carried out under the India-Trento Program for Advanced Research (ITPAR) during a stay of Miss Kanchan Bahirat at the University of Trento (Italy).

REFERENCES

[1] H. Daume` III and D. Marcu, "Domain Adaptation for Statistical Classifiers," J. Artificial Intelligence Research, vol. 26, pp. 101-126, 2006.

[2] H. Daumé III, "Frustratingly Easy Domain Adapatation," *Proceedings of the 2010 Workshop on DA for Natural Language Processing*, ACL 2010, Uppsala, Sweden, pp. 53-59, 15 July 2010.

[3] J. Jiang, and C. Zhai, "Instance weighting for Domain Adaptation in NLP," *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, Prague, Czech Republic, 2007..

[4] S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, "Analysis of representation for domain adaptation," in *Advances in Neural Information Processing Systems 19*, Vancouver, British Columbia, Canada, 2006.

[5] L. Bruzzone, D. Fernández Prieto, S. B. Serpico, "A neural statistical approach to multitemporal and multisource remote-sensing image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 1350–1359, May 1999.

[6] L. Bruzzone, "An approach to feature selection and classification of remote sensing images based on the Bayes rule for minimum cost," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 429–438, Jan. 2000.

[7] L. Bruzzone, M. Marconcini, "Domain Adaptation Problems: a DASVM Classification Technique and a Circular Validation Strategy," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 32, No. 5, pp. 770-787, 2010.

[8] L. Bruzzone, M. Marconcini, "Toward the Automatic Updating of Land-Cover Maps by a Domain-Adaptation SVM Classifier and a Circular Validation Strategy," *IEEE Trans. Geosci. Remote Sensing*, Vol. 47, No. 4, pp. 1108-1122, 2009.

[9] L. Bruzzone, D. Fernández Prieto, "Unsupervised Retraining of a Maximum Likelihood Classifier for the Analysis of Multitemporal Remote Sensing Images," *IEEE Trans. Geosci. Remote Sensing*, vol. 29, No.2, pp. 456–460, Jan. 2001.

[10] L. Bruzzone, and D. Fernàndez Prieto, "A partially unsupervised approach to the automatic classification of multitemporal remote-sensing images," *Pattern Recognition Letters*, vol. 33, no. 9, pp. 1063-1071, 2002.

[11] L. Bruzzone, R. Cossu, and G. Vernazza, "Combining parametric and non-parametric algorithms for a partially unsupervised classification of multitemporal remote-sensing images," *Information Fusion*, vol. 3, no. 4, pp. 289-297, 2002.

[12] L. Bruzzone, and R. Cossu, "A multiple-cascade-classifier system for a robust and partially unsupervised updating of land-cover maps," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 1984-1996, September 2002.

[13] J. T. Tou, R. C. Gonzalez, *Pattern Recognition Principles*. Reading, MA: Addison-Wesley, 1974.

[14] J. A. Richards, *Remote Sensing Digital Image Analysis*, 2nd ed. New York: Springer-Verlag, 1993.

[15] A. Singh, "Digital change detection techniques using remotely-sensed data," Int. J. Remote Sens., vol. 10, no. 6, pp. 989–1003, 1989.

[16] L. Bruzzone, D. Fernandez Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 38, No.3, 2000, pp. 1171-1182

[17] F. Bovolo, L. Bruzzone, "A Theoretical Framework for Unsupervised Change Detection Based on Change Vector Analysis in the Polar Domain," *IEEE Trans. Geosci. Remote Sensing*, vol. 45, No.1, pp. 218–236, Jan. 2007.
[18] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 294–307, 2005.
[19] F. Bovolo, S. Marchesi, L. Bruzzone, "A nearly lossless 2d representation and characterization of change information in multispectral images," Proc. of IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2010), pp. 3074 – 3077, 2010.
[20] J. A. Richards and X. Jia, Remote Sensing Digital Image Analysis: An Introduction. New York: Springer-Verlag, 1999.
[21] G. E. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, Vol. 6, No. 2, pp. 461–464, 1978.
[22] A. P. Dempster, N. M. Laird, D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.
[23] B. M. Shahshahani, D. Landgrebe, "The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 1087–1095, Sept. 1994.
[24] T. K. Moon, "The expectation-maximization algorithm," *Signal Processing Mag.*, vol. 13, no. 6, pp. 47–60, 1996.
[25] L. Bruzzone, F. Roli, S.B. Serpico, "An extension of the Jeffreys-Matusita distance to multiclass cases for feature selection," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 3, No. 6, pp. 1318-1321, 1995.
[26] Schwarz, G., "Estimating the dimension of a model," *The Annals of Stat.*, vol. 6, no. 2, pp. 461–464, 1978.
[27] S. Marchesi, F. Bovolo, L. Bruzzone, "A context-sensitive technique robust to registration noise for change detection in VHR multispectral images," *IEEE Transactions on Image Processing*, Vol. 19, No. 7, pp. 1877-1889, July 2010.