

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/274645402>

Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis

Article in *ISPRS Journal of Photogrammetry and Remote Sensing* · March 2015

DOI: 10.1016/j.isprsjprs.2015.02.005

CITATIONS

92

READS

394

3 authors:



Michele Volpi

ETH Zurich

64 PUBLICATIONS 2,614 CITATIONS

[SEE PROFILE](#)



Gustau Camps-Valls

University of Valencia

747 PUBLICATIONS 20,737 CITATIONS

[SEE PROFILE](#)



Devis Tuia

École Polytechnique Fédérale de Lausanne

278 PUBLICATIONS 9,837 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Project Advanced Learning in classification, change and target detection in remote sensing [View project](#)



Project SEDAL: Statistical Learning for Earth Observation Data Analysis [View project](#)

Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis

Michele Volpi^{a,*}, Gustau Camps-Valls^b, Devis Tuia^c

^aCALVIN, Institute of Perception, Action and Behaviour, the University of Edinburgh, EH8 9AB Edinburgh, United Kingdom.

<https://sites.google.com/site/michelevolpiresearch/>

^bImage Processing Laboratory, Universitat de València, C/ Catedrático Escardino, E-46980 Paterna (València), Spain.

<http://www.uv.es/gcamps/>

^cMultimodal Remote Sensing Group, Department of Geography, University of Zurich, 8057 Zurich-Irchel, Switzerland

<http://www.geo.uzh.ch/en/units/multimodal-remote-sensing>

Abstract

In this paper we propose an approach to perform relative spectral alignment between optical cross-sensor acquisitions. The proposed method aims at projecting the images from two different and possibly disjoint input spaces into a common latent space, in which standard change detection algorithms can be applied. The system relies on the regularized kernel canonical correlation analysis transformation (kCCA), which can accommodate nonlinear dependencies between pixels by means of kernel functions. To learn the projections, the method employs a subset of samples belonging to the unchanged areas or to uninteresting radiometric differences. Since the availability of ground truth information to perform model selection is limited, we propose a completely automatic strategy to select the hyperparameters of the system as well as the dimensionality of the transformed (latent) space. The proposed scheme is fully automatic and allows the use of any change detection algorithm in the transformed latent space. A synthetic problem built from real images and a case study involving a real cross-sensor change detection problem illustrate the capabilities of the proposed method. Results show that the proposed system outperforms the linear baseline and provides accuracies close the ones obtained with a fully supervised strategy. We provide a MATLAB implementation of the proposed method as well as the real cross-sensor data we prepared and employed at <https://sites.google.com/site/michelevolpiresearch/codes/cross-sensor>.

Keywords: Change detection, Cross-sensor, Relative spectral alignment, Kernel methods, Feature extraction, Canonical correlation analysis.

1. Introduction

The amount of Earth observation data available to study the Planet's surface is destined to grow exponentially, as dictated by the "big data" trends. Recent and forthcoming missions from both national and private agencies will further increase the amount of available images from remote sensing instruments. Consequently, the evolution in technologies and increased accessibility to data makes remote sensing a viable tool for many environmental studies. Moreover, image databases and derived products offer a variety of information sources to study the temporal evolution of the dynamical process occurring at the Earth's surface. However, versatile and generic methods able to extract relevant information automatically are needed. The ability of processing and integrating multi-temporal acquisitions from multiple sensors into standard remote sensing systems is of paramount importance for three main reasons. Firstly, integrating such series accommodates data complementarity: fused products extracted from complementary data possess an increased information content (Pohl and Van Genderen,

1998). Secondly, the ability of exploiting images from heterogeneous sources in multi-temporal scenarios drastically reduces constraints such as equal spectral resolution on same sensed wavelengths, possibly increasing the temporal resolution of such studies. Thirdly, response time of image processing systems, such as those required for disaster management and post-catastrophe assessment may be further reduced by allowing for integration of arbitrary multi-sensor and cross-sensor images (Römer et al., 2010; Chatelain et al., 2008).

In this paper we distinguish *multi-sensor* from *cross-sensor*. The former corresponds to two time series of acquisitions from different sensors with single pairs of images acquired within a short time interval. The latter denotes a single time series composed of acquisitions from multiple sensors. In this paper, we consider the *cross-sensor* setting. When dealing with time series composed of single-sensor images, it is common to observe gaps in the spatial or in the temporal sampling domains (for example due to clouds or sensor failures) (Gómez-Chova et al., 2013; Longbotham et al., 2012; Villa et al., 2013). The same issue affects also multi-sensor time series, but in this case images from the complementary sensor may be used to fill in the gap observed in the other time series (Röder et al., 2005; Desclée et al., 2013; Amorós-López et al., 2011a). However,

*Corresponding Author: Michele Volpi, michele.volpi@ed.ac.uk, Tel: +44 (0) 131-650-2733, Fax: +44 (0)131-651-3435. DT: devis.tuia@epfl.ch. GCV: gcamps@uv.es Web:

to model the time series one has to manually combine multiple sensors acquisitions or develop specific strategies of data fusion (Schmid and Koch, 2005). The systematic exploitation of optical cross-sensor data in real world systems is still limited by the complexity and lack of universality of current data fusion methods. Methods such as the STARFM (Gao et al., 2006) or the one by Amorós-López et al. (2011a) allow to fuse images from different sensors and to retrieve both spectrally and spatially enhanced images. However, these methods have been developed for precise pairs of sensors with specific characteristics. The change detection tasks may be greatly simplified by generic and possibly automatic cross-sensor image integration systems.

A question of interest when dealing with cross-sensor change detection problems is how to deal with differences in the spectral domain. If spatial resolutions may be easily adjusted by gridding and resampling, detecting spectral changes in inhomogeneous domains is more delicate. Most of the times, the common spectral channels across acquisitions are manually selected so that similar wavelength intervals are used for band-wise comparison (Wulder et al., 2008). However, if sensors provide *disjoint* representations of the spectrum, such a manual approach may not be applicable. A general framework to cross-sensor spectral alignment for change detection is still missing.

Standard approaches to perform cross-sensor change detection require pixel-level labels representing either the spectral classes or the type of transitions occurring across images. Post-classification comparison is usually employed in the former case (Singh, 1989; Mubea and Menz, 2012), while for the latter direct multi-date classification is often preferred (Singh, 1989; Turker and Asik, 2005; Qin et al., 2013; Volpi et al., 2013b). In the first case, thematic classification maps obtained independently for each acquisition are compared and the segmented multi-temporal information is summarized in a single map layer. However, this system does not account for temporal dependencies and the final results strongly depend on the quality of single *independent* classifications. In the second case, image channels are simply stacked and a single thematic classification is performed on the entire multi-temporal set, implicitly accounting for temporal dependencies. Compared to post-classification approaches it tends to provide more precise change detection maps, since not affected by the propagation of the errors of single maps. However, this procedure requires labeled multi-temporal signals describing all the transitions of interest, which are rarely available. Also, dealing with a high dimensional problem with few training samples will lead to the curse of dimensionality and robust supervised learners must be employed. This may occur easily when dealing with hyperspectral data or adding spatial-contextual descriptors in direct multi-date scenarios (Volpi et al., 2013b).

There are few techniques able to perform general cross-sensor change detection which not rely on the limiting assumption of having an exhaustive training set. Parametric models such as the one developed by Fernández-Prieto and Marconcini (2011) perform selective change detection on a subset of changes described by labels. Since relying on joint probabilities modeled independently on the images, the method is independent

of the spectral characteristics of the image. Alberga (2009) presented a method which exploits spectrally invariant measures to depict differences in the spatial arrangement of pixels between the multi-temporal acquisitions. The measures proposed consider only the relative local illumination differences and consequently the dimensionality of the spectral domain does not constraint the problem. This may be seen as comparing texture from local patches working on general single channel images. The last family of cross-sensor methods rely on joint transformations of the spectral domain of the images. These techniques perform a relative radiometric alignment of images by training a multi-output regression model to predict all the pixels of the post-event image on the basis of a subset of pairs of unchanged pixels across times (Yang and Lo, 2000; Healey et al., 2006). For change detection purposes, the residual image (difference between actual and estimated pixel values) are compared as if employing standard difference image analysis, e.g. change vector analysis (CVA) (Bovolo et al., 2011). A similar reasoning is behind the multivariate alteration detection (MAD), which relies on the canonical correlation analysis (CCA) transformation (Nielsen et al., 1998; Nielsen, 2007, 2002; Canty, 2010; de Carvalho Júnior et al., 2013). CCA is a linear rotation-based method that aims at finding an optimal linear combination of the two disjoint groups of features maximizing correlation (Hotelling, 1936). In this case, instead of fitting a least-squares regression on unchanged samples, MAD-based approaches transforms the pixels so that their projection maximizes the correlation. The multivariate difference of the projected data is used as a change indicator in standard change detection routines. Note that the multiple output regression schemes and CCA are intimately related, as illustrated in (de la Torre, 2012).

This latter family of methods estimates a rotation from the spectral channels directly (also known as the ‘primal formulation’). Despite the simplicity and generating light computational efforts, this assumption of linearity between spectral channels may be not sufficient to properly model cross-sensor data. The pair of acquisitions may show differences due to complex light interactions, seasonality, local atmospheric conditions, illumination changes and other possibly nonlinear spectral transformations affecting unevenly the image in both spatial and spectral domains but not representing ground cover changes (Tuia et al., 2014; Gómez-Chova et al., 2013). Nonlinear learning methods have been deeply studied for remote sensing applications, in particular for pixel classification tasks. Classical examples of employed classification models are support vector machines and kernel methods (Gómez-Chova et al., 2008; Maulik and Chakraborty, 2013) and neural networks (Li et al., 2014; Shao and Lunetta, 2012; Amorós-López et al., 2011b). In this paper, we focus on kernel methods, since offering sound theoretical guarantees and flexible solutions. Moreover, kernel methods are well suited for dealing with heterogeneous sources of information (Camps-Valls et al., 2008; Tuia et al., 2010), feature extraction and dimensionality reduction (Arenas-García et al., 2013) and change detection (Bovolo et al., 2008; Volpi et al., 2013b). The main intuition behind kernel methods is that a nonlinear problem in the original input space may be trans-

formed into a linear one by recasting it into a higher dimensional space. To achieve this, one usually has to find a mapping function of the original data samples to a space in which a selected linear method works. However, the optimal mapping function is not known in advance and the proper estimation directly from the data may be difficult and computationally unfeasible. To alleviate this issue kernel methods theory shows that a dot product between samples mapped into a higher dimensional space may be replaced by a valid kernel function, only taking as argument the samples in their input space. For more details on SVM and kernel methods, we refer to (Shawe-Taylor and Cristianini, 2004; Schölkopf and Smola, 2002).

In this paper, we propose an approach relying on the kernel extension of the canonical correlation analysis, the kernel CCA (kCCA), to perform relative spectral alignment of cross-sensor acquisitions. Kernel CCA is a nonlinear variant of the CCA, which aims at computing a projection of the samples maximizing the correlation in some higher dimensional space in which linear correlations depict the true structure of the data. When looking at these projections from the original input space they corresponds to nonlinear correlations between data samples. This transformation is independent of the number and spectral range of image channels of the single acquisitions. The images are mapped in a space in which component-wise comparisons are allowed, since corresponding to joint directions of maximal correlations. Closely related to the approach presented here, Nielsen and Vestergaard (2013) proposed a study of the kernel version of the MAD, also relying on the kCCA. In the settings presented in this paper, the method is able to properly highlight landcover changes in a pair of images from a same sensor. Similarly, in our previous work (Volpi et al., 2013a) the kernel CCA was studied for the alignment of data spaces in a pair of images from a single sensor. To study the robustness of the proposed method in cross-sensor situations, some channels were removed from one of the acquisitions considered. However, the model selection issue was not tackled directly and binary labels were used to select free parameters of the system minimizing the change detection error. Note that this is an important limitation towards the automatic application of the approach.

The kCCA-based scheme we are proposing aims at *automatically* align data spaces issued from cross-sensor acquisitions, thanks to an automatic model selection strategy. The proposed tuning process find automatically all the hyperparameters involved which define the best possible transformation of a set of unchanged samples from both acquisitions. This projection consequently enhances differences between pairs of transformed coregistered pixels, interpreted as changes. To evaluate the quality of the joint mapping, we perform change detection as a separate and *independent* step. Note that this last element of the pipeline is not a central contribution of the paper, but rather an instrument to evaluate the appropriateness of the transformation. We evaluate our proposed system on two cross-sensor dataset: a simulated one from single sensor images and a real case study involving 4 cross-sensor images. To foster reproducibility of our experiments, we release the latter dataset with the ground truth we defined and the MATLAB codes used in the paper. To our knowledge, this is the first publicly avail-

able dataset for cross-sensor change detection and relative spectral alignment problems.

The paper is organized as follows. Section 2 describes the CCA and its kernel extension, as well as the proposed model selection strategy. Section 3 presents the datasets employed and the setup of the experiments, which are detailed and discussed in Section 4. Finally, Section 5 concludes the paper.

2. Methodology

Canonical correlation analysis (CCA) is a multivariate feature extraction method that aims at finding the rotation of two sets of variables that maximizes their joint correlation (Hotelling, 1936). Its kernel extension implements exactly the same reasoning in reproducing kernel Hilbert spaces (RKHS), thus recasting the problem in some higher (possibly infinite) dimensional spaces (Bach and Jordan, 2002; Arenas-García et al., 2013). By doing so, dependencies and nonlinear correlations in the data may be retrieved by running the linear algorithm in the RKHS. In this section we detail the regularized kernel CCA starting from the linear multivariate CCA as well as the proposed model selection strategy.

Notation. Let us define two acquisitions X_1 and X_2 , which may belong to different data spaces $X_1 \in \mathcal{X}_1$, $X_2 \in \mathcal{X}_2$. From both images we select n *unchanged* pixels, corresponding spatially to the same coordinates. This means that we assume coregistered images. In the following derivations we focus only on those samples. Let also \mathbf{X}_1 and \mathbf{X}_2 be the corresponding mean-centered data matrices defined over $\mathbb{R}^{n \times d_1}$ and $\mathbb{R}^{n \times d_2}$ where d_1 and d_2 are the number of spectral channels of each image. We also define corresponding covariance matrices $\mathbf{C}_{11} = \mathbf{X}_1^\top \mathbf{X}_1 \in \mathbb{R}^{d_1 \times d_1}$ and $\mathbf{C}_{22} = \mathbf{X}_2^\top \mathbf{X}_2 \in \mathbb{R}^{d_2 \times d_2}$ and cross-covariance matrices $\mathbf{C}_{12} = \mathbf{X}_1^\top \mathbf{X}_2 \in \mathbb{R}^{d_1 \times d_2}$ and $\mathbf{C}_{21} = \mathbf{C}_{12}^\top$. If no subscript specifying the acquisition time is specified, we will refer to a general single data matrix / set.

2.1. Maximizing the correlation between two sets of variables

The CCA can be seen as a normalized extension of the PCA to paired data. PCA finds a linear transformation of a single set of variables \mathbf{X} (centered to zero mean) so that its projection maximizes the variance of the data across features. To this end, PCA aims at decomposing a covariance matrix \mathbf{C} into mutually orthogonal components depicting directions of largest variance:

$$\begin{aligned} \arg \max_{\mathbf{u}} & (\mathbf{X}\mathbf{u})^\top (\mathbf{X}\mathbf{u}) = \mathbf{u}^\top \mathbf{C}\mathbf{u} \\ \text{s.t. } & \mathbf{u}^\top \mathbf{u} = 1. \end{aligned} \quad (1)$$

The dataset is projected onto the axis of maximal variance by means of the vector $\mathbf{u} \in \mathbb{R}^{d \times 1}$, which solves the problem in Eq. (1). The projected data matrix is obtained by $\mathbf{X}' = \mathbf{X}\mathbf{u}$. The above system of equations may be solved by introducing the Lagrange multipliers λ and equating the derivative to zero. The solution is obtained by solving the following symmetric eigenvalue problem:

$$\mathbf{C}\mathbf{u} = \lambda\mathbf{u} \quad (2)$$

In the equation above, d different solutions exist, i.e. the pairs (λ, \mathbf{u}) , where \mathbf{u} is the eigenvector and λ the corresponding eigenvalue. To project the data into the new uncorrelated bases a linear transformation matrix is built as $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_q]$, where $q \leq d$ is the dimensionality of the latent space selected by the user. The first solution of the system $(\lambda_1, \mathbf{u}_1)$ correspond to the amount and direction of the axis of largest data variance. The remaining $(d - 1)$ solutions are ordered so that $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_d$ and $\mathbf{u}_i^\top \mathbf{u}_j = 0$ for $i \neq j$.

In contrast to PCA, which decomposes a single set, the CCA accounts simultaneously for two separate data spaces describing the same n objects. In our case, the data matrices \mathbf{X}_1 and \mathbf{X}_2 represent two sets of spectral channels representing the same geographical coordinates. It assumes that linear dependencies between the two sets exist and that they can be unveiled and summarized by two rotations of the sets to a common space. To this end, CCA finds two sets of projections \mathbf{U} and \mathbf{V} so that the correlation of the projected data is maximized. Denoting the vectors projecting the two sets into the common direction of maximal correlation as \mathbf{u} and \mathbf{v} , the CCA problem may be formulated starting from the definition of correlation:

$$\begin{aligned} \arg \max_{\mathbf{u}, \mathbf{v}} \text{corr}(\mathbf{X}_1 \mathbf{u}, \mathbf{X}_2 \mathbf{v}) &= \\ &= \frac{\text{cov}(\mathbf{X}_1 \mathbf{u}, \mathbf{X}_2 \mathbf{v})}{\text{var}(\mathbf{X}_1 \mathbf{u})^{1/2} \text{var}(\mathbf{X}_2 \mathbf{v})^{1/2}} \\ &= \frac{\mathbf{u}^\top \mathbf{X}_1^\top \mathbf{X}_2 \mathbf{v}}{(\mathbf{u}^\top \mathbf{X}_1^\top \mathbf{X}_1 \mathbf{u})^{1/2} (\mathbf{v}^\top \mathbf{X}_2^\top \mathbf{X}_2 \mathbf{v})^{1/2}} \\ &= \frac{\mathbf{u}^\top \mathbf{C}_{12} \mathbf{v}}{(\mathbf{u}^\top \mathbf{C}_{11} \mathbf{u})^{1/2} (\mathbf{v}^\top \mathbf{C}_{22} \mathbf{v})^{1/2}} \end{aligned} \quad (3)$$

Note that the correlation is invariant to the magnitude of \mathbf{u} and \mathbf{v} . Therefore, we may reformulate the problem by setting the norm of the projections to be unitary ($\mathbf{u}^\top \mathbf{C}_{11} \mathbf{u} = \mathbf{v}^\top \mathbf{C}_{22} \mathbf{v} = 1$):

$$\begin{aligned} \arg \max_{\mathbf{u}, \mathbf{v}} \mathbf{u}^\top \mathbf{C}_{12} \mathbf{v} \\ \text{s.t. } \mathbf{u}^\top \mathbf{C}_{11} \mathbf{u} = 1, \\ \mathbf{v}^\top \mathbf{C}_{22} \mathbf{v} = 1. \end{aligned} \quad (4)$$

This system can be solved by introducing Lagrange multipliers and equating partial derivatives to zero to obtain a generalized eigenvector problem (De Bie et al., 2004; Shawe-Taylor and Cristianini, 2004). The standard CCA is solved by:

$$\begin{pmatrix} \mathbf{0} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \rho \begin{pmatrix} \mathbf{C}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}, \quad (5)$$

where ρ is the eigenvalue corresponding to the correlation between the first pair of projections of the sets. This problem has $d_1 + d_2$ solutions, among which $\min(d_1, d_2)$ are nonzero. The correlations between the canonical variates are given as $\{\rho_1, -\rho_1, \rho_2, -\rho_2, \dots, \rho_p, -\rho_p\}$, where ρ is the correlation between $\mathbf{X}'_1 = \mathbf{X}_1 \mathbf{u}$ and $\mathbf{X}'_2 = \mathbf{X}_2 \mathbf{v}$ and $-\rho$ between $\mathbf{X}'_1 = \mathbf{X}_1(-\mathbf{u})$ and $\mathbf{X}'_2 = \mathbf{X}_2 \mathbf{v}$. The triplets $(\rho, \mathbf{u}, \mathbf{v})$ may be reorganized so that $\rho_1 \geq \rho_2 \geq \dots \geq \rho_p$. As in the PCA, the projections are uncorrelated ($\mathbf{u}_i^\top \mathbf{u}_j = 0, \forall i \neq j$) and may be used to transform the

original datasets into a lower p -dimensional dataset of maximally correlated pairs of variables by stacking projection vectors $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p]$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p]$, $\mathbf{X}'_1 = \mathbf{X}_1 \mathbf{u}$ and $\mathbf{X}'_2 = \mathbf{X}_2 \mathbf{v}$.

Nielsen et al. (1998) and Nielsen (2007) used this approach to define a measure of change. For this purpose, all the pixels in the pair of images are used no matter if they belong to changed or unchanged areas. By assuming that unchanged areas show the largest correlations, canonical variates containing information about the changed areas are the ones showing the lowest pairwise correlation (since pixels should be greatly scattered). Consequently, the corresponding difference would show the largest variance, $\text{var}(\mathbf{X}_1 \mathbf{u} - \mathbf{X}_2 \mathbf{v}) = \text{var}(\mathbf{X}_1 \mathbf{u}) - \text{var}(\mathbf{X}_2 \mathbf{v}) + \text{cov}(\mathbf{X}_1 \mathbf{u}, \mathbf{X}_2 \mathbf{v}) = 2(1 - \text{corr}(\mathbf{X}_1 \mathbf{u}, \mathbf{X}_2 \mathbf{v}))$. In these terms, the MAD difference image is defined as the $\mathbf{D}_{\text{MAD}} = \mathbf{X}_2 \mathbf{u} - \mathbf{X}_1 \mathbf{v}$ with eigenvectors corresponding to the lowest and non-zero absolute values of the correlations. Note that, by making use of all the pixels of the bi-temporal set (or a random subset of it), the MAD assumes that changes are present during the computation of the transformation. However, to be distinguished from noise, changes must either represent a significant but not predominant portion of the data cloud showing a strong deviation from the background signal. Authors in Nielsen (2007) proposed to iteratively reweight the output of the MAD, by assigning low values to changes. This way, the global statistics are iteratively reweighted so that the model obtained at convergence optimally aligns images based on unchanged samples. In contrast to this approach, our system only relies on a set of manually defined unchanged samples. On the one hand, it relaxes the assumptions made on the statistics of the changed classes, not required to be of any particular form, nor requires iterations to soften deviations. On the other hand, the manual selection of unchanged samples may be difficult. In our setting we consider pairwise spectral alignment problems and manual screening still feasible. However, we argue that employing the method of (Nielsen, 2007) to preselect a subset of samples may result in an optimal trade-off between the correctness of the samples used for learning and the time spent in screening images. Note that in the experiments we applied the MAD trained on the set of unchanged samples. This ensures a proper estimation of the projection and makes the method stable independently of the type and size of changes involved.

2.2. Estimating nonlinear correlation using kernels

To obtain the kernel version of the CCA, the samples are replaced with their RKHS maps defined by the functions $\phi_t : \mathcal{X}_t \rightarrow \mathcal{F}_t$ as $\mathbf{x} \mapsto \phi_t(\mathbf{x})$, where \mathcal{F}_t is a RKHS in which \mathbf{X}_t is mapped. The index t corresponds to the time instant. For simplicity, we drop the indices when referring to an arbitrary \mathcal{F} .

The mapped dataset is defined as $\Phi = [\phi(\mathbf{x}_1) \phi(\mathbf{x}_2) \dots \phi(\mathbf{x}_n)]^\top$. By replacing directly \mathbf{X}_t with Φ_t , the CCA solution can be obtained into the higher dimensional space in which the Φ_t are embedded. However, projections \mathbf{U} and \mathbf{V} are now of dimensionality $\dim(\mathcal{F}) \times p$. Since $\dim(\mathcal{F})$ may be even infinite, solving the CCA in the RKHS via the primal formulation is unfeasible.

To avoid dealing with the estimation of explicit mappings the primal formulation in Eq. (4) is modified to obtain the dual (alternative) problem, which depends only on the number of samples involved. To do so, we make use of the representer's theorem which states that the solution may be expressed as a weighted linear combination of the training samples as $\mathbf{u} = \Phi_1^\top \alpha$ and $\mathbf{v} = \Phi_2^\top \beta$. The dual problem is obtained by plugging in \mathbf{u} and \mathbf{v} in Eq. (5). We also introduce a positive semi-definite *kernel function* $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$, generating a corresponding positive semi-definite kernel matrix \mathbf{K} with entries $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$. The kernel matrix corresponds to the dot product between all examples $\mathbf{K} = \Phi^\top \Phi$, and we assume it is centered to its origin. The dual problems is expressed as:

$$\begin{aligned} & \arg \max_{\alpha, \beta} \alpha^\top \mathbf{K}_{11} \mathbf{K}_{22} \beta \\ & \text{s.t. } \alpha^\top \mathbf{K}_{11}^2 \alpha = 1 \\ & \quad \beta^\top \mathbf{K}_{22}^2 \beta = 1, \end{aligned} \quad (6)$$

By introducing again Lagrange multipliers we can solve the dual by:

$$\mathcal{J}(\alpha, \beta) = \alpha^\top \mathbf{K}_{11} \mathbf{K}_{22} \beta + \rho_1(1 - \alpha^\top \mathbf{K}_{11}^2 \alpha) + \rho_2(1 - \beta^\top \mathbf{K}_{22}^2 \beta), \quad (7)$$

and by equating partial derivatives to 0 we obtain the two normal equations (note that $\rho_1 = \rho_2$):

$$\begin{aligned} \frac{\partial \mathcal{J}(\alpha, \beta)}{\partial \alpha} = 0 & \Leftrightarrow \mathbf{K}_{11} \mathbf{K}_{22} \beta - 2\rho \mathbf{K}_{22}^2 \alpha = \mathbf{0} \\ \frac{\partial \mathcal{J}(\alpha, \beta)}{\partial \beta} = 0 & \Leftrightarrow \mathbf{K}_{22} \mathbf{K}_{11} \alpha - 2\rho \mathbf{K}_{11}^2 \beta = \mathbf{0} \end{aligned}$$

In the case \mathbf{K}_t are invertible this system of equations leads to $n-1$ solutions equal to $\rho = 1$, $n-1$ solutions equal to $\rho = -1$ and 2 solutions equal to $\rho = 0$ no matter which samples are employed, thus becoming data independent as shown in (Gretton et al., 2005). To avoid falling in trivial situations the kCCA requires regularization. A possible solution to ensure well-posedness independently of the form of \mathbf{K} is to enforce smooth solutions in the weights \mathbf{u} and \mathbf{v} in the primal formulation by penalizing their norms $\|\mathbf{u}\|_{\mathcal{F}} = \alpha^\top \mathbf{K}_{11} \alpha$ and $\|\mathbf{v}\|_{\mathcal{F}} = \beta^\top \mathbf{K}_{22} \beta$, where $\|\cdot\|_{\mathcal{F}}$ denotes the norm of the projection functions in the RKHS. In the dual form of the problem, this regularization gives the following optimization problem:

$$\begin{aligned} & \arg \max_{\alpha, \beta} \alpha^\top \mathbf{K}_{11} \mathbf{K}_{22} \beta \\ & \text{s.t. } \alpha^\top \mathbf{K}_{11}^2 \alpha + \varepsilon_1 \alpha^\top \mathbf{K}_{11} \alpha = 1, \\ & \quad \beta^\top \mathbf{K}_{22}^2 \beta + \varepsilon_2 \beta^\top \mathbf{K}_{22} \beta = 1, \end{aligned} \quad (8)$$

where ε_1 and ε_2 are two regularization parameters. The above problem leads to the final formulation of the regularized kCCA solved by the following generalized eigenvector problem:

$$\begin{pmatrix} \mathbf{0} & \mathbf{K}_{11} \mathbf{K}_{22} \\ \mathbf{K}_{22} \mathbf{K}_{11} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \rho \begin{pmatrix} \mathbf{K}_{11} \mathbf{K}_{11} + \varepsilon_1 \mathbf{K}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{22} \mathbf{K}_{22} + \varepsilon_2 \mathbf{K}_{22} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \quad (9)$$

where, ρ corresponds to the amount of correlation of the sets in \mathcal{F}_t .

Finally, the kernel canonical variates are obtained by projecting test samples into the subspace spanned by the first q pairs $\mathbf{A} = [\alpha_1, \dots, \alpha_q]$ and $\mathbf{B} = [\beta_1, \dots, \beta_q]$, as $\mathbf{X}'_1 = \mathbf{K}_{11}^s \mathbf{A}$ and $\mathbf{X}'_2 = \mathbf{K}_{22}^s \mathbf{B}$, where \mathbf{K}^s denotes the centered test kernel. In this paper, we assume the number of components q to be a model parameter, to be selected. This term strongly depends on the data and influences heavily the informativeness and compactness of the latent space. In the following developments, we will use exclusively Gaussian RBF kernels, of the form $k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right)$, where σ is another hyperparameter that needs to be tuned during the model selection step.

2.3. Two-steps automatic model selection

The regularized kCCA depends on a series of hyperparameters that have to be specified by the user. These parameters are crucial to correctly model the cross-correlations, so that the projection of the images show maximal suppression of the no-change background. In this section, we propose an automatic procedure to find them without falling in under- or over-fitting. To flexibly encode data differences due to sensor characteristics into the transformation we allow for the selection of different hyperparameters for each image. The proposed model selection strategy consists of two subsequent steps, detailed below.

Step 1: Randomized association spectrum. In this step of the procedure the spectrum of the kCCA $\rho(\mathbf{X}|\Theta)$, i.e. the vector of eigenvalues sorted in decreasing order, is compared to the one issuing from the kCCA after randomization of the pixel associations, $\rho(\mathbf{X}_R|\Theta)$. In both cases, we test the same set of model hyperparameters Θ . To randomize the associations in \mathbf{X}_R we mix randomly \mathbf{X}_2 sample-wise (row-wise). In the literature this technique is known as scrambling (Nielsen and Vestergaard, 2013) or randomized surrogates (Hardoon et al., 2004; Biessmann et al., 2010). In this step, we search the hyperparameters of the system providing the largest deviation of $\rho(\mathbf{X}|\Theta)$ from $\rho(\mathbf{X}_R|\Theta)$. In principle, no structure should be present in the latter, since cross-relationships are destroyed by randomization. Thus, if both $\rho(\mathbf{X}|\Theta)$ and $\rho(\mathbf{X}_R|\Theta)$ result in large correlations values it is very likely that the current hyperparameter set leads to over-fitting, which means finding correlations in randomized associations. If both $\rho(\mathbf{X}|\Theta)$ and $\rho(\mathbf{X}_R|\Theta)$ are small, it is likely that no useful information is extracted by kCCA which underfits data relationships. In both cases, the corresponding fitness measure should return low values. The ideal situation is encountered when the randomized kCCA shows low and constant correlations, while the true kCCA shows large eigenvalues: in this case, we model the data structure correctly. The measure of fitness employed here simply compares the cross-correlations of the spectra by considering the ℓ_2 norm of the difference between corresponding eigenvalues. However, in order to make the hyperparameter selection procedure stable, we have to optimize them over a finite latent dimensionality. If employing too many latent dimensions the set of optimal hyperparameters will tend to accommodate the data correlations in too many components, possibly leading to low out-of-sample correlations. On

the contrary, by fixing the latent dimensionality to very few components, the set of hyperparameters is forced to compress all information in few components, possibly missing again important structures. In this paper we define this intermediate dimensionality as $P^{\text{temp}} = \max(d_1, d_2)$. This ensures that the size of the latent dimensionality of this step does not strongly influence the choice of the optimal set of hyperparameters Θ^* . Thus, the latter optimizes:

$$\Theta^* = \arg \max_{\Theta} \|\rho(\mathbf{X}|\Theta, P^{\text{temp}}) - \rho(\mathbf{X}_R|\Theta, P^{\text{temp}})\|_2 \quad (10)$$

where $\Theta = \{\varepsilon_1, \varepsilon_2, \sigma_1, \sigma_2\}$ is the set of free hyperparameters to be selected, given P^{temp} . Due to the nature of the data, random associations may result in some residual correlation. For instance, two pixels at different coordinates but corresponding to the same landcover class may indeed show some correlation. However, these situations are of no influence in the procedure, since the background correlation should be constant.

Step 2: kernel generalized variance. This second step of the model selection strategy aims at estimating the number of eigenvectors needed to map the original images into the latent space, independently of the previously fixed P^{temp} . To this end, we exploit a side product of the kernel CCA, the kernel Generalized Variance (kGV) $\psi(\mathbf{X}_1, \mathbf{X}_2|\Theta^*)$ (Bach and Jordan, 2002; Gretton et al., 2005). The kGV is a kernel measure of dependence between variables, tightly linked to the kernel mutual information. Our choice in adopting kGV instead of mutual information is the empirical superiority in small sample scenarios, as illustrated in (Gretton et al., 2005). Once determined the optimal hyperparameters Θ^* governing the kCCA transformation, we define the kGV as:

$$\psi(\mathbf{X}_1, \mathbf{X}_2|\Theta^*) = -\frac{1}{2} \log \left(\prod_{i=1}^P (1 - \rho_i^2) \right), \quad (11)$$

where ρ are the non-negative correlations from Eq. (9). $\psi(\mathbf{X}_1, \mathbf{X}_2|\Theta^*) = 0$ if the variables in the RKHS are independent and $\psi(\mathbf{X}_1, \mathbf{X}_2|\Theta^*) > 0$ for dependent variables. In our setting, we aim at *maximizing* the dependence of the projected pixels, since representing only unchanged areas. However, at the same time, we also want to limit the number of projections needed to reflect this dependence. To do so, we evaluate $\psi(\mathbf{X}_1, \mathbf{X}_2|\Theta^*)$ with growing P , up to the number of components showing non-null correlation. Then, we automatically detect where the measure saturates. Implicitly, the non-increase of this measure indicates that the relevant information is already represented in the retained components. Therefore, we set the number of retained projections for mapping the cross-sensor data as the corresponding P . This procedure can be seen as a “feature selection” in the RKHS, where only the transformed variables showing largest mutual dependence are retained. In our case, we take advantage of the implicit ordering provided by the value of the canonical correlations. In practice, we stop the search when the contribution to the kGV of a component i is smaller than a pre-selected threshold $\epsilon = 0.1$ (recall that correlations are in $[0, 1]$).

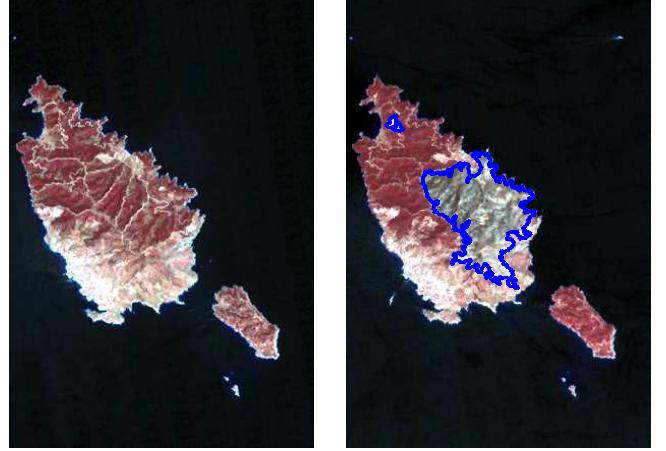


Figure 1: Spetses island pre- and post-event images. The burn scar is outlined in blue.

Table 1: Tested spectral channel configurations for the Spetses Island dataset.

| Exp. | Domains | \mathcal{X}_1 (band #) | \mathcal{X}_2 (band #) |
|---------------|--|--------------------------|--------------------------|
| FULL vs. FULL | $\mathcal{X}_1 = \mathcal{X}_2$ | [1 to 7] | [1 to 7] |
| FULL vs. RGB | $\mathcal{X}_1 \supset \mathcal{X}_2$ | [1 to 7] | [1 to 3] |
| RGB vs. IR | $\mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset$ | [1 2 3] | [4 5 7] |

3. Data and experimental setup

In this work, we validated the proposed cross-sensor spectral alignment system by employing two different datasets, dubbed ‘Spetses Island’ and ‘Bastrop Fires’. The first dataset is simulated from a pair of single-sensor Landsat 5 TM images, while the second is composed of a series of 4 real images: two Landsat 5 TM, a EO-1 ALI and a Landsat 8 OLI. We used the CVA (Bovolo et al., 2011) as the change detection method, on both the original data (when possible) or after projection with the MAD (Nielsen et al., 1998; Nielsen, 2007) and the proposed kCCA. Recall that the MAD is applied by maximizing the correlation between unchanged samples, as in the kCCA setting. This makes the system directly comparable to the one employing kCCA.

3.1. Spetses Island, Greece

This dataset is composed of two small subsets of Landsat 5 TM images of the Peloponnese, Greece, acquired respectively in 1987 and 1991. The scenes are 444×300 pixels with a spatial resolution of 30m, representing a small island named Spetses. The change detection problem consists in correctly delineating a forest fire burn scar on the west flank of the island (Fig. 1). For validation purposes, 26‘530 pixels have been labeled by photointerpretation. As there are few spectral classes in the dataset and the change is well defined and unique, we use the spectrally subsampled images as a synthetic example to analyze the behavior of the system under different channel associations, as depicted in Tab. 1.

We used as baseline the accuracy of the system when dealing with images from the same input space, i.e. $\mathcal{X}_1 = \mathcal{X}_2$ (FULL vs. FULL in Tab. 1). Then, to simulate a cross-sensor scenario,

we tested the setting in which the full spectrum of the first image is used in conjunction to the RGB space of the second one ($\mathcal{X}_1 \supset \mathcal{X}_2$, FULL vs. RGB in Tab. 1). In this case, the same information is contained in both dates, but there is less information available about the burned area since infrared channels which correlate well with vegetation are removed from the post-event image. A third case study involves two non-overlapping spectral spaces ($\mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset$, RGB vs. IR in Tab. 1), i.e. comparing a RGB with a NIR/MIR image. This way, we simulate the extreme case of cross-sensor change detection, in which the underlying information is represented by features of different nature. Note that in this setting we omitted the TIR channel since showing a slight spatial misalignment. In this extreme setting, the misaligned TIR channel overwhelmed the rest of the information making the spectral alignment fail. However, this channel is present in the first and second settings, where information from one or both set is considerably larger and the transformations were able to correctly depict correlation directions.

Experimental setup. For this dataset we evaluated performances for an increasing number of samples from the “unchanged” class used to estimate the alignment, as $n = \{50, 100, 250, 500\}$. The same samples are used for both the MAD and the kCCA projections. The search of the hyperparameters was performed in $\varepsilon = \{1e-5, \dots, 1e5\}$ and $\sigma = \{0.5\bar{\sigma}, 1\bar{\sigma}, 1.5\bar{\sigma}, \dots, 10\bar{\sigma}\}$, where $\bar{\sigma}$ is the median Euclidean distance of the samples from each acquisition (thus different for each time instant). We report the average and standard deviation of the estimated Cohen’s Kappa statistic (κ) over 10 runs, involving a fixed test set and random realizations of the training set. The threshold of each binary CVA model (original, MAD, kCCA) is fitted in a supervised way by exploiting 20 pixels randomly selected from the changed / unchanged areas. This step stabilizes the CVA by helping in selecting a proper threshold. These samples are varied along the experiments and excluded from the test set.

Supervised model selection. We provide a term of comparison for the two-step automatic model selection by using a fully supervised strategy which minimizes the error of the CVA over the 20 validation samples. We first select the set of hyperparameters by fixing $P^{\text{temp}} = \max(d_1, d_2)$ and minimizing the error of the CVA. Then, the optimal dimensionality of the subspace is retrieved by minimizing again the CVA error on subspaces of dimension 1 to 20. This setup simply re-implements the unsupervised strategy with the supervision brought by the CVA error over the validation set.

3.2. Bastrop county complex fire

This dataset is composed of a set of four images acquired by different sensors over the Bastrop County, Texas (USA). On September 4, 2011, the region has been struck by “the most destructive wildland-urban interface wildfire in Texas history”¹,

which counts 2 casualties, more than 1300 destroyed buildings and almost burned entirely the Bastrop county state park.

We use 4 images of size 1534×808 to validate the proposed cross-sensor relative spectral alignment strategy. Table 2 summarizes the data and Fig 2 shows the images along with the manually defined ground truth. We dispose of a pair of pre- and post-event images from the same sensor, the Landsat 5 TM (L5t1 and L5t2), which are completed by a post-event of another sensor, the Advanced Land Imager (ALI) from the Earth Observing (EO-1) mission, acquired very shortly after the L5t2 (ALIt2 hereafter). These three scenes are very similar to each other and no apparent changes between L5t2 and ALIt2 are visible, since images were acquired within 1 day interval. Differences between these pre- and post-event pairs are only due to burned forest since they were acquired at a 16 days interval during summer. We also dispose of a fourth image of the same area acquired one year and 9 months after the forest fire by the Landsat 8 Operational Land Imager (OLI, L8t2 hereafter). The differences between L5t1 and L8t2 are significant, due both to sun / sensor angles and the long temporal interval between acquisitions. A whole new series of building has been constructed in the burn scar and many cultivated crops are at a different stage of growth. Different change detection setting could have been implemented, but we opted for change detection between the pre-event image L5t1 and the post-event L8t2.

We prepared the ground truth for pairs of change detection problems by photo-interpretation and relying on the maps provided on the emergency response website¹. In the ground truth involving the L5t1-L8t2 problem we also included changes related to vegetation density and vegetation / bare soil transitions, since these changes are of the same spectral class as those related to the burned scar. For this last change detection problem we use angles between spectral change vectors as well, as $\omega_i = \arccos\left(\frac{\mathbf{x}_{2,i} - \mathbf{x}_{1,i}}{\|\mathbf{x}_{1,i}\|_2 \|\mathbf{x}_{2,i}\|_2}\right)$, where \mathbf{x} represent the i -th transformed pixel. The angle distribution $p(\omega)$ is exploited by the CVA to discriminate the burn scar from other types of changes.

As this data has been collected from the NASA LP DAAC Program², we are free to redistribute the data. For this reason we provide the Bastrop data and the ground truth we defined and used in these experiments (subset of original crops and ground truth) as well as the MATLAB implementation used to obtain the results on our webpage³. To our knowledge, this is the first annotated cross-sensor dataset which has been publicly released.

Experimental setup. For this dataset we perform a comparison across multiple binary alignments, as depicted in Tab. 3. As this case study involves larger images and a more complex spectral alignment, we employed 40 labeled random samples to validate CVA threshold. Otherwise, we explored same search ranges of hyperparameters as for the Spetses case study. The only difference is in the use of the CVA angles for the L5t1 vs L8t2, which uses the 40 validation samples to threshold the CVA angles.

¹A scientific report is available at (visited April 15, 2014)
<http://www.co.bastrop.tx.us/bcdisaster/index.php/wildfire-case-study>

²Land Processes Distributed Active Archive Center,
<http://eros.usgs.gov/about-us/data-citation>, visited on November 2014,

³<https://sites.google.com/site/michelevolpiresearch/codes/cross-sensor>

Table 2: Main characteristics of the cross-sensor Bastrop Fires data.

| | Pre-event (L5T1) | Post-event 1 (L5T2) | Post-event 2 (ALIτ2) | Post-event 3 (L8τ2) |
|-----------------------|-------------------------------|--------------------------------|---|--|
| Sensor | Landsat 5 TM | Landsat 5 TM | EO-1 ALI | Landsat 8 OLI |
| Date | 26.08.2011 | 11.09.2011 | 12.09.2011 | 28.6.2013 |
| # channels | 7 | 7 | 9 [†] | 7 [‡] |
| Pixel size [m] | 30, 120 | 30, 120 | 30 | 30 |
| Spectral range | [0.45 - 2.35, μm] | [0.45 - 2.35, μm] | [0.4 - 2.4] | [0.43 - 2.25] |
| | [10.40 - 12.50] | [10.40 - 12.50] | | |

[†] The panchromatic was not included.

[‡] The panchromatic and “cirrus” bands were not included.

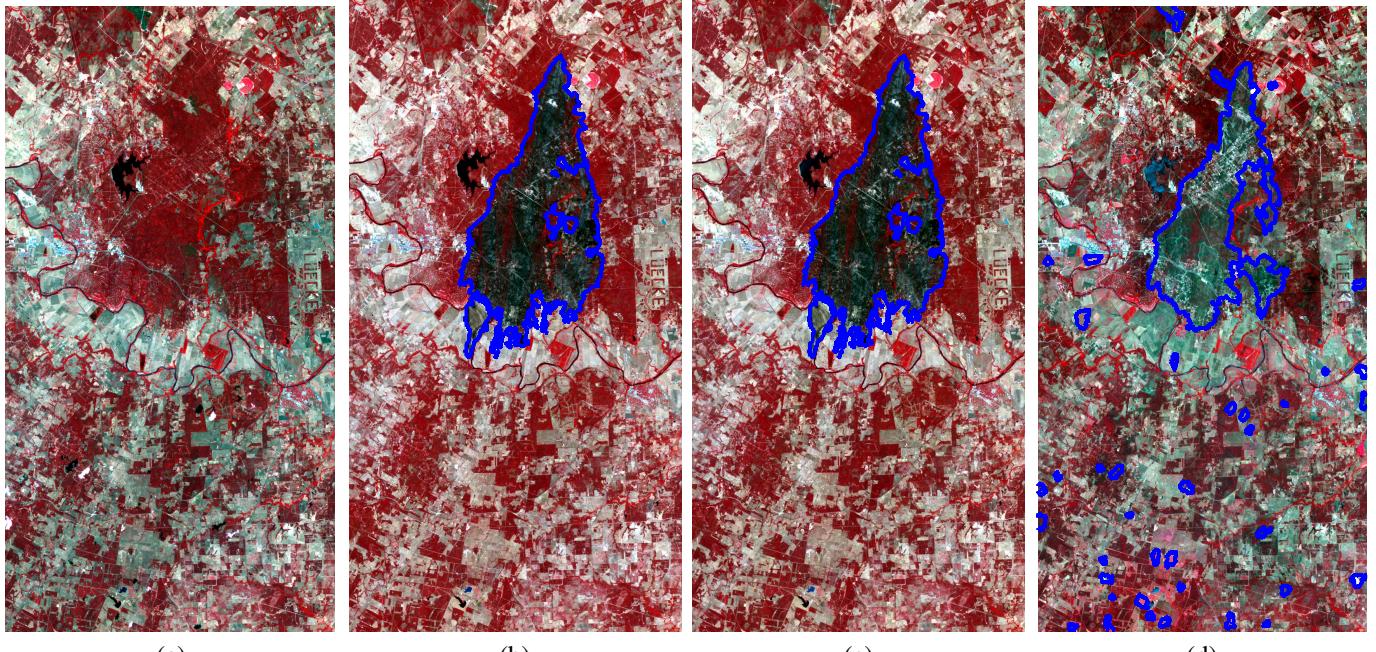


Figure 2: False color composite of the images used in the experiments. For the image details refer to Tab. 2. The burn scar is in blue over (b) the Landsat 5 TM (c) the ALI and (d) the Landsat 8 images.

Table 4: Estimated κ statistic (and standard deviation) for the Spetses Island case study after employing CVA, MAD, and the kCCA on simulated cross-sensor data for different sizes of the training set. For the kCCA, we compare the proposed automatic ℓ_1 model selection with the ℓ_2 counterpart and the fully supervised strategy (kCCA^s).

| | <i>n</i> | Original | MAD | kCCA | kCCA^s |
|--|------------|--------------------------|-------------|-------------|-------------------------|
| FULL vs. Full $\chi_1 = \chi_2$ | 50 | 0.44 (0.06) | 0.46 (0.11) | 0.68 (0.07) | 0.63 (0.08) |
| | 100 | | 0.55 (0.06) | 0.73 (0.04) | 0.66 (0.11) |
| | 250 | | 0.53 (0.06) | 0.71 (0.05) | 0.72 (0.13) |
| | 500 | | 0.59 (0.04) | 0.66 (0.07) | 0.75 (0.04) |
| FULL vs. RGB $\chi_1 \subset \chi_2$ | 50 | 0.34 (0.09) [†] | 0.17 (0.06) | 0.49 (0.11) | 0.56 (0.13) |
| | 100 | | 0.22 (0.03) | 0.49 (0.10) | 0.60 (0.12) |
| | 250 | | 0.20 (0.06) | 0.61 (0.09) | 0.63 (0.10) |
| | 500 | | 0.21 (0.04) | 0.61 (0.07) | 0.68 (0.10) |
| RGB vs. IR $\chi_1 \cap \chi_2 = \emptyset$ | 50 | 0.48 (0.04) [‡] | 0.58 (0.09) | 0.57 (0.13) | 0.69 (0.08) |
| | 100 | | 0.57 (0.07) | 0.59 (0.11) | 0.70 (0.06) |
| | 250 | | 0.63 (0.04) | 0.63 (0.06) | 0.72 (0.08) |
| | 500 | | 0.60 (0.07) | 0.65 (0.10) | 0.74 (0.09) |

[†] χ_1 has been downsampled to contain RGB only.

[‡] χ_1 matches χ_2 (IR channels only).

Table 3: Set of experiments for the Bastrop county dataset.

| Exp. ID | X_1 (# bands) | X_2 (# bands) |
|----------------|-----------------|-----------------|
| L5T1 vs. L5T2 | 7 | 7 |
| L5T1 vs. ALIr2 | 7 | 9 |
| L5T1 vs L8T2 | 7 | 7 |

4. Results

4.1. Spetses Island artificial case study

Table 4 shows the numerical accuracies of the change detection performed on the original single-sensor images and after relative spectral alignment, for all the tested settings. The baseline CVA accuracy (column ‘Original’) is always surpassed by all the tested methods, excepted for the MAD on the FULL vs. RGB ($X_1 \subset X_2$). For the two scenarios in which the two domains are composed of different sets (FULL vs. RGB and RGB vs. IR) we apply the original CVA by modifying the input spaces: in the first case we downgrade the pre-event acquisition to match RGB channels and in the second case we use same IR channels. Although practically unrealistic, these simulated situations provide useful baselines.

Globally, increasing the number of unchanged pixels for learning the transformation improves the numerical accuracy of the maps. The differences in accuracy among the three experiments are caused by the dependency between the spectral channels corresponding to unchanged areas. For instance, RGB components only explain little about the variation of the unchanged vegetation, while IR channels provide more valuable information on which estimating the relative spectral alignment. This may explain why the setting FULL vs RGB provides less accurate results compared to the other two scenarios: the IR bands are available only for the pre-event image and the post-event image only carries visible bands, that make the alignment more difficult.

By comparing the kCCA results with automatic model selection and those supervised model selection (kCCA^s column in Tab. 4) we observe that the latter are in general more accurate. However, the proposed automatic strategy does show a significantly higher accuracy if compared to either the original CVA or to the linear MAD. The availability of binary supervision may help in particular when only a small subset of unchanged samples is available for learning the transformation. The hyperparameters selected by both automatic and the supervised kCCA are illustrated in Tab. 5.

By observing hyperparameters obtained with the automatic selection, three main observations may be drawn. First, regularization parameters ϵ may strongly differ depending on the information content of the feature sets. For instance, the FULL vs. FULL scenario requires very small regularizations, equally distributed between pre- and post-event image. In the second case, the FULL vs. RGB experiment, strong regularization is required for both sets, with accent on the ‘FULL’ feature set. In the third situation, RGB vs. IR, the regularizers seem to penalize mostly the RGB feature set, probably since the IR contributes

massively to the overall alignment by providing features less correlated between each other (in the ‘IR’ feature set).

Secondly, the scaling of the median Euclidean distance employed as the kernel hyperparameter $\sigma/\bar{\sigma}$ is more or less stable for each separate scenario. This indicates that the degree of nonlinearity required to perform the automatic relative spectral alignment does not vary much depending on the feature sets. The degree of nonlinearity learned by the kCCA is proportional to the size of the kernel: the smallest, the more nonlinear the solution is.

Finally, the dimensionality of the latent space q is also relatively stable along each experiment. Two tendencies are observed: on one hand the dimensionality of the latent space is related to the size of the original feature sets to be aligned and on the other hand the amount of information required to describe the feature sets increases with the number of samples employed in the transformation. This is intuitively linked to the dual formulation of the kCCA that upper bounds the number of canonical variates by the number of samples and searches correlations as linear combinations of the samples (rather than directly between variables).

An interesting observation may be drawn by comparing the hyperparameters obtained by the unsupervised model selection strategy with those obtained directly minimizing the CVA error. With the exception of some particular cases, the regularizers and the final dimensionality of the set are comparable, with the latter being only slightly larger on the average. This indicates that the information contained in the original sets requires roughly the same degree of regularization and can be summarized in a compact number of latent dimensions. The only significant difference resides in the kernel scaling: the supervised strategy tends to select much smaller scalings. This is also reflected in the optimal number of component retained by the supervised strategy, which is slightly larger due to the nonlinearity brought by smaller kernel parameters. Nevertheless, recall that the proposed automatic strategy does not make use of any labeled pixel during model selection. Moreover, the fully supervised strategy directly minimizes the CVA cost. Although being more accurate, the supervised strategy lacks completely in generality.

Figure 3 illustrates the projections into the first 6 canonical correlations for the pre- and post-event images for the kCCA and MAD approaches respectively (FULL vs. FULL experiment). Note that for illustration purposes the first 6 projections have been rescaled to a common range. The last row of the figure shows the difference of the corresponding components. Homogeneous unchanged area having large contrast with the changed one should correspond in a better relative spectral alignment and thus in more accurate change detection maps. It may be observed that both the kCCA and MAD, in the first 3 components, are able to align well the images (similar colors for same spectral class) as illustrated in Fig. 3(a),(e) (kCCA) and Fig. 3(c),(g) (MAD). The resulting differences consequently enhance well the changes. For the first 3 canonical correlations, the kCCA Fig. 3(i) results in a clear and well contrasted difference, with a homogeneous background, while the MAD Fig. 3(k) in a less homogeneous but still clear difference. In this last case, it is

Table 5: Estimated set of hyperparameters (and standard deviation) via the automatic model selection for the Spetses Island scenarios. As a term of comparison, we show hyperparameters obtained with supervised model selection (kCCA^s). The values of the hyperparameters in the table correspond to the average over the 10 experiments. ε_1 , ε_2 column show the magnitude of the regularizer, $\sigma/\bar{\sigma}$ the scaling of the kernel and P the retained latent dimensionality.

| n | kCCA | | | | kCCA ^s | | | | |
|----------------------------|-----------------|-----------------|-----------------------|-------------|-------------------|-----------------|-----------------------|-------------|-------------|
| | ε_1 | ε_2 | $\sigma/\bar{\sigma}$ | P | ε_1 | ε_2 | $\sigma/\bar{\sigma}$ | P | |
| $X_1 = X_2$ | 50 | -5 | -5 | 4.60 (1.02) | 5.0 (0.82) | 0 | -5 | 1.70 (0.45) | 8.8 (2.68) |
| | 100 | -5 | -5 | 5.05 (0.76) | 5.3 (0.48) | 0 | -5 | 2.60 (0.96) | 11.4 (6.07) |
| | 250 | -5 | -5 | 4.90 (0.62) | 5.8 (0.63) | -5 | -5 | 2.10 (0.74) | 6.8 (4.71) |
| | 500 | -5 | -5 | 5.15 (0.71) | 5.9 (0.57) | -5 | -5 | 2.30 (0.84) | 6.0 (3.87) |
| $X_1 \subset X_2$ | 50 | -1 | -5 | 5.80 (0.35) | 2.8 (0.42) | -1 | -1 | 2.35 (1.03) | 2.9 (0.88) |
| | 100 | -1 | -2 | 5.75 (0.35) | 3.1 (0.32) | -1 | -1 | 3.50 (1.25) | 2.9 (0.88) |
| | 250 | -1 | -1 | 5.65 (0.47) | 3.8 (0.78) | -5 | -1 | 1.80 (0.89) | 4.0 (0.10) |
| | 500 | -1 | -1 | 5.95 (0.16) | 4.2 (0.63) | -1 | -1 | 2.50 (1.31) | 3.4 (1.17) |
| $X_1 \cap X_2 = \emptyset$ | 50 | -1 | -5 | 4.95 (1.01) | 2.1 (0.32) | -1 | -1 | 1.90 (0.74) | 3.6 (1.08) |
| | 100 | -1 | -1 | 5.20 (1.03) | 2.3 (0.68) | -1 | -1 | 1.95 (1.11) | 3.1 (1.10) |
| | 250 | -1 | -5 | 4.10 (1.00) | 2.9 (0.57) | -1 | -1 | 2.60 (1.58) | 3.4 (1.65) |
| | 500 | -1 | -3 | 3.80 (1.00) | 3.4 (0.70) | -1 | -1 | 2.05 (1.67) | 3.1 (1.37) |

visible a noisy pattern caused by differences in illuminations over the village in the south flank of the island. These temporally incoherent and abrupt reflections cannot be aligned well by the linear approach. In this case, the nonlinearity offered by the kCCA is able to cope with such deviations offering a set of projections less prone to false alarms. When computing the difference with the 3 next canonical correlations (4th to 6th), the kCCA Fig. 3(j) still able to provide a well-defined and homogeneous representation of the changed / unchanged areas. For the MAD components Fig. 3(l), sensor noise and striping starts to overwhelm true changes and thus less information can be exploited. Similar observations can be made for the projections of the subsampled images for the other Spetses tests, as illustrated in Fig. 4. It is worth underlying the homogeneous projections obtained by the kCCA in both FULL vs. RGB and RGB vs. IR cases. In the same situations the MAD is still able to provide a meaningful difference image, but with projections showing poorly aligned data spaces.

4.2. Bastrop complex fires case study

Analysis of the real cross-sensor Bastrop complex fire case study confirms the observations made for Spetses synthetic experiments. The main difference resides in the baseline CVA performance for single sensor images ('Original' column). The standard CVA outperforms the linear method and ranges in the kCCA transformation accuracy. The kCCA on the L5t1 vs. L5t2 returns a projection allowing accurate change detection, but only slightly better than the baseline CVA. However, MAD is always outperformed. For the L5t1 vs. ALIt2 cross-sensor experiment, kCCA provides accuracies very close to the ones obtained with the same-sensor CVA and only slightly inferior to the kCCA over the same-sensor set. Again, the MAD is outperformed significantly by the kCCA. This suggests that the relative spectral transformation is able to provide valuable information only when accounting for data nonlinearities. For

Table 6: Estimated κ statistics (and standard deviation) for the Bastrop complex fire case study after employing standard CVA, MAD, and the proposed automatic kCCA for different sizes of the training set.

| n | Original | MAD | kCCA |
|------------|------------|-------------|-------------|
| L5 vs. L5 | 50 | 0.46 (0.10) | 0.66 (0.13) |
| | 100 | 0.50 (0.07) | 0.61 (0.14) |
| | 250 | 0.46 (0.13) | 0.64 (0.08) |
| | 500 | 0.48 (0.07) | 0.65 (0.09) |
| L5 vs. ALI | 50 | 0.38 (0.08) | 0.56 (0.10) |
| | 100 | 0.32 (0.06) | 0.62 (0.07) |
| | 250 | - | 0.62 (0.10) |
| | 500 | 0.41 (0.06) | 0.65 (0.08) |
| L5 vs. L8 | 50 | 0.16 (0.03) | 0.39 (0.10) |
| | 100 | 0.17 (0.04) | 0.44 (0.11) |
| | 250 | - | 0.45 (0.08) |
| | 500 | 0.17 (0.02) | 0.51 (0.06) |

the last case we already mentioned that L5t1 vs. L8t2 is a more complex experiment and consequently provide lower numerical accuracies. In this case, the linear MAD is not able to provide a representation in which changes can be detected. This is caused by the large amount of pervasive differences which are not well modeled by the MAD. In this case, a nonlinear solution seems more appropriate. The kCCA provides an alignment of unchanged areas in which changes can be detected with an acceptable accuracy only when using large training sets. This observation is in line with the complexity of the problem, requiring more samples to be properly represented. Recall that, in order to obtain more accurate results, the angle and magnitude of changes are jointly used in the CVA to exclude transitions not related to vegetation (after both the MAD and kCCA).

By observing the projections and the change maps obtained after the kCCA spectral alignment (Fig. 5) we can visually confirm the relative performances presented above. To be fair, we

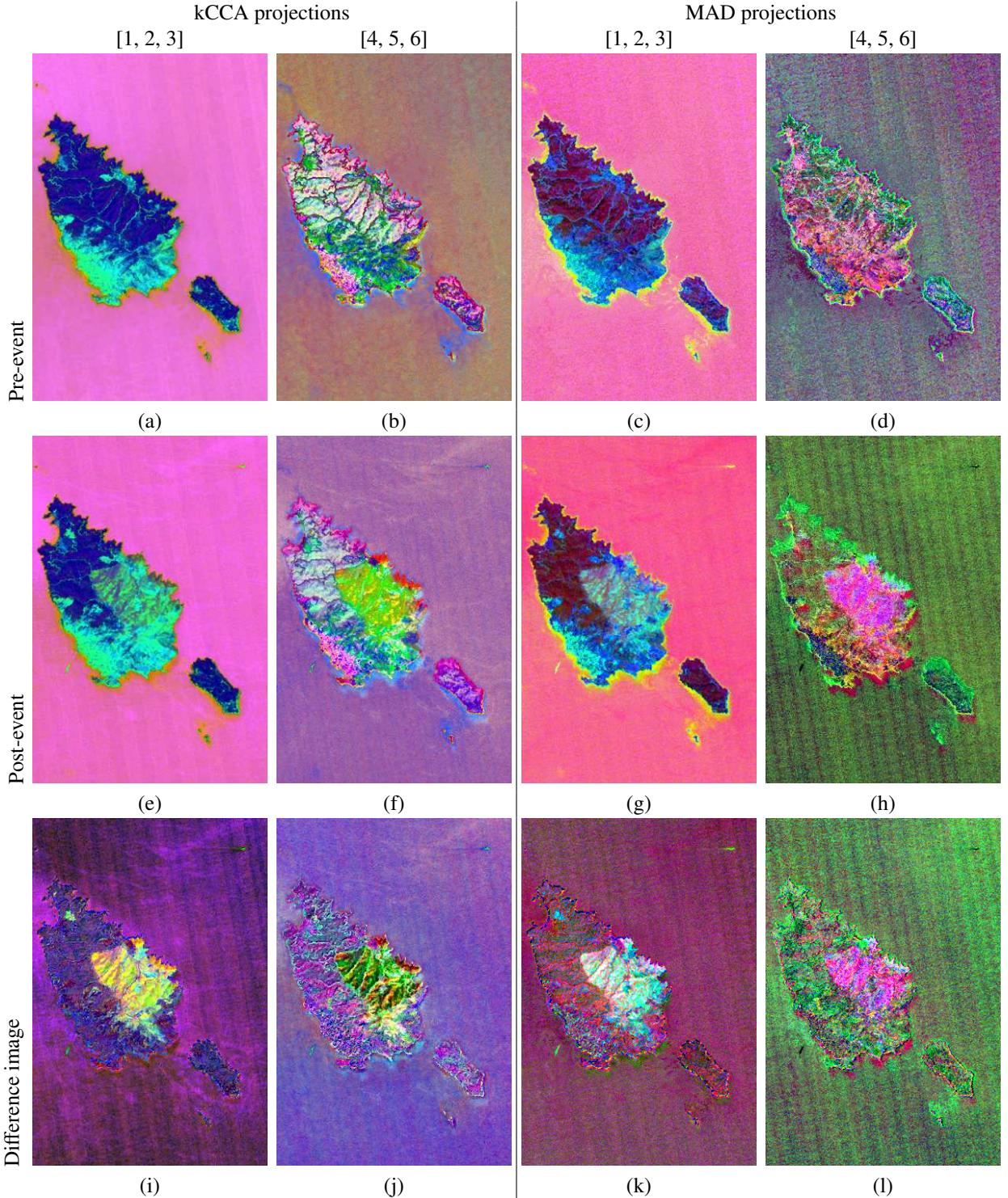


Figure 3: The Spetses dataset projected images for the FULL vs. FULL setting. Each figure is a RGB composition of three projections in decreasing correlation order. First row: pre-event projected images, (a)-(b) first 6 projections by kCCA and (c)-(d) MAD. Second row: post-event projected images by kCCA (e)-(f) and MAD (g)-(h); last row: difference of projected images: kCCA (i)-(j) and MAD (k)-(l). For illustration purposes, the projections have been rescaled to a common range.

show the maps for the second best performing run. In the first two experiments, the change detection problem is easier since the pair of images were acquired within a small time interval, with the post-event image acquired shortly after the fire. In this case, the kCCA alignment is able to provide very smooth un-

changed areas in the difference image. Consequently, changes are correctly enhanced and detected. This appears also clear by observing the projection of the single pre- and post-event change detection maps.

For the L5T1 vs. L5T2 case (first row, Fig. 5(a)-(d)), the

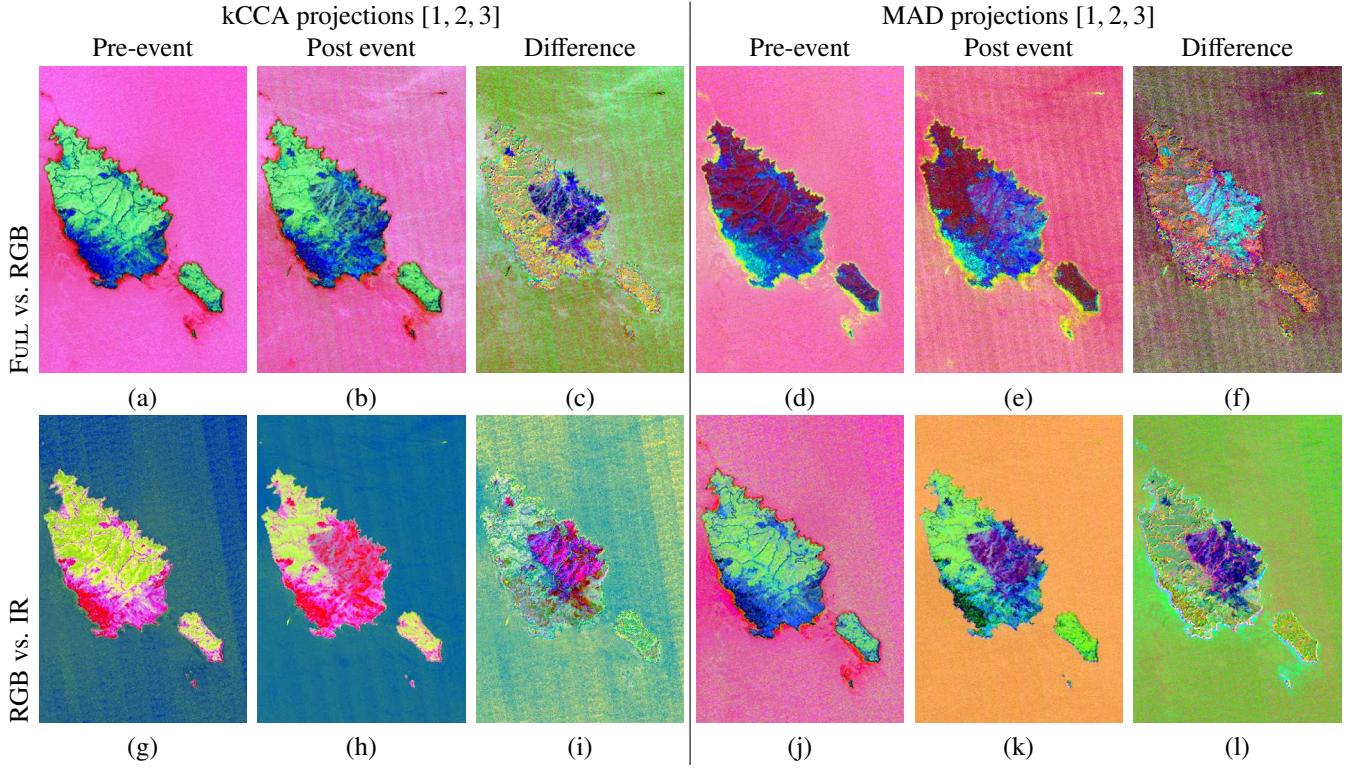


Figure 4: The projected images obtained for the Spetses dataset for the FULL vs. RGB and RGB vs IR case studies (one for each setting). Each figure is a RGB composition of three projections in decreasing correlation order. (a)-(c) First 3 kCCA projections for the pre- and post-event image and their difference and same for MAD (d)-(f). For the RGB vs. IR case study, projections and differences in (g)-(i) for the kCCA and in (j)-(l) for the MAD.

final change maps correctly delineates the burned area, and additionally includes some other changes related to differences over water bodies caused by sun reflections (both the river and other small water basins). For the L5r1 vs. ALI case (second row, Fig. 5(e)-(h)), the projections are very similar to the previous ones, but the final change map is more noisy. This is mainly caused by different range of values in the projections and a larger class overlap, probably caused by the larger dimensionality of the ALI image. This squeezes and reduces the range in which a proper CVA threshold may be discovered. Still, the changes are accurately delineated in the final change map. The last case study, L5r1 vs L8r2 (third row, Fig. 5(i)-(l)), shows a completely different projection for the pre- and post-event images, resulting in a difference in the relative importance of the components. This may be observed by the different colors in both single time image projections and in the corresponding difference image. The final map, however, is mostly coherent with the observed differences, but shows a large quantity of missed detections.

Figure 6 reports the correlations obtained during model selection and the corresponding kGV values. All the curves correspond to the average of 10 runs of independent training set realizations. It appears that the amount of information that can be extracted is directly dependent on the number of samples used in the transformation. For each sample size, the model selection strategy returns a set of kernel canonical variates describing the associations between images with as little features as possible. This is underlined by the tendency of the spectrum to reach null correlation within the first 20 components. Inter-

estingly, the gap between randomized associations and actual correlation increases for larger sets, i.e. the solid lines move towards the right, while the dotted ones towards the left / bottom. This observation links to the theoretical tenet stating that the larger the sample size, the less overwhelming overfitting is. In our case this overfitting is directly related to models which discover many large correlations in randomized associations. However, note that in practice random associations return some correlations, since the number of spectral classes is limited and consequently by randomly permuting samples it may happen that different but similar pixels are associated by chance. Overall, this effect is limited but possibly exacerbated when using small sample sizes.

The amount of dependency extracted by the kCCA is better quantified by the cumulative kGV, shown in the right column of Fig. 6. This measure saturates where the amount of dependency extracted by the projected features reaches a maximum. This saturation level is easily discovered by a threshold, as the derivative tends to zero as the number of components grows. As for the canonical correlation, kGV shows that the larger the set of unchanged pixels, the most information is contained in the transformed feature sets. As a side note, it is also possible to appreciate the complexity of the third cross-sensor case study by looking at the total amount of kGV that can be extracted using 500 unchanged samples. The values at which the curve saturates is significantly lower if compared to the one of the first two experiments, which in turn show similar values.

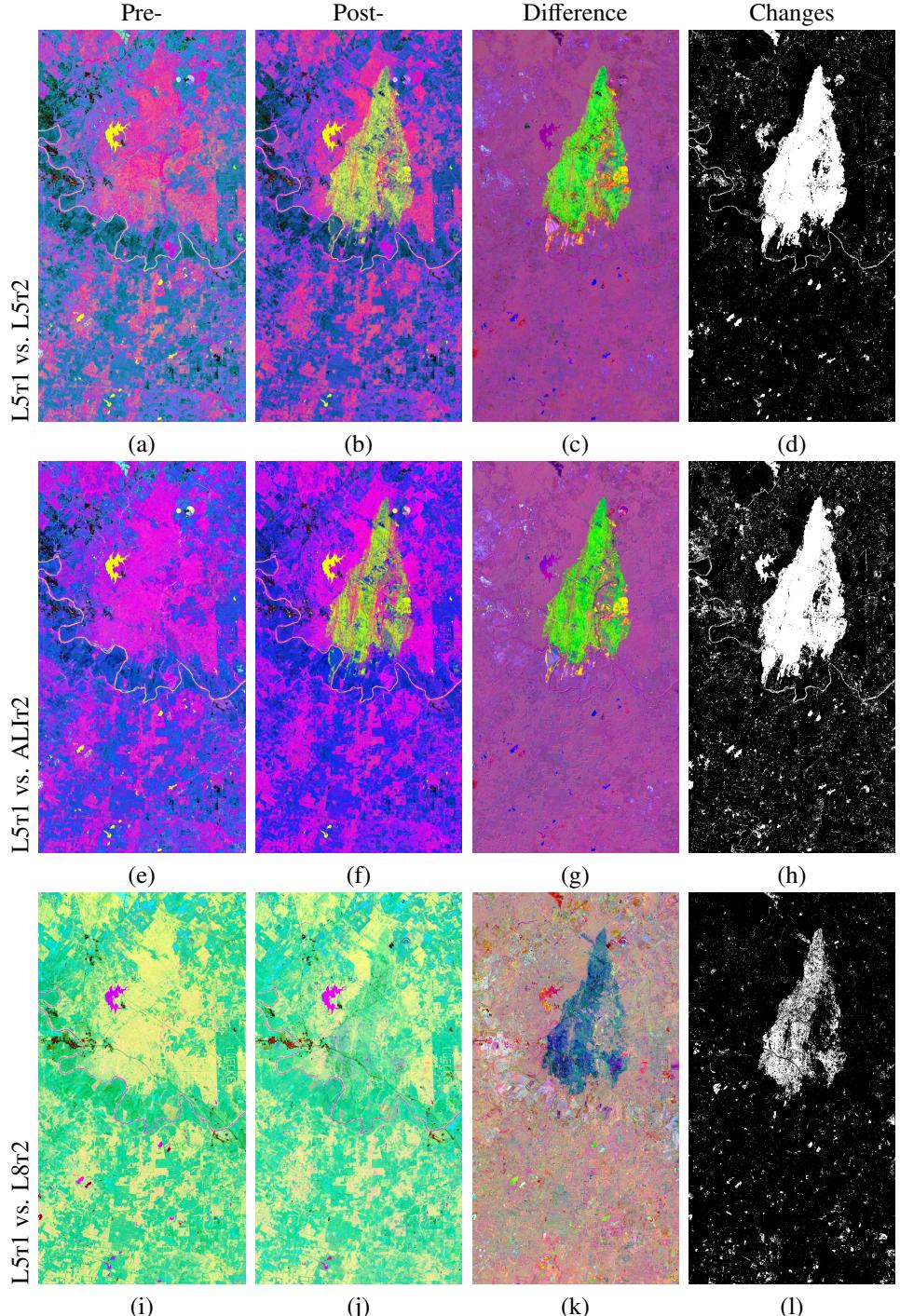


Figure 5: The projected images obtained by kCCA for the Bastrop dataset for the 3 cross-sensor settings. Each figure represent a false color composition of the 3 kernel canonical variates in decreasing correlation order. First row: L5T1 vs L5T2 (a)-(b) projections, (c) difference and (d) CVA map. Second row: L5T1 vs ALIr2 (e)-(f) projections, (g) difference and (h) CVA map. Third row: L5T1 vs L8T2 (i)-(j) projections, (k) difference and (l) CVA map.

4.3. Discussion

In both the tested experiments we can successfully carry out accurate cross-sensor change detection with a very simple change detection strategy. The observations made on the experiments agree on both datasets and we can summarize some general considerations.

The most accurate results are obtained by employing the

kCCA to perform relative spectral alignment. This is achieved in situations in which the original images are either from the same sensor or from cross-sensor acquisitions. This is achieved also in situations in which the feature sets represent disjoint information (Spetses RGB vs. IR case study) or showing external changes strongly distorting the probability distribution of the unchanged sample (Bastrop Fires L5T1 vs. L8T2 case

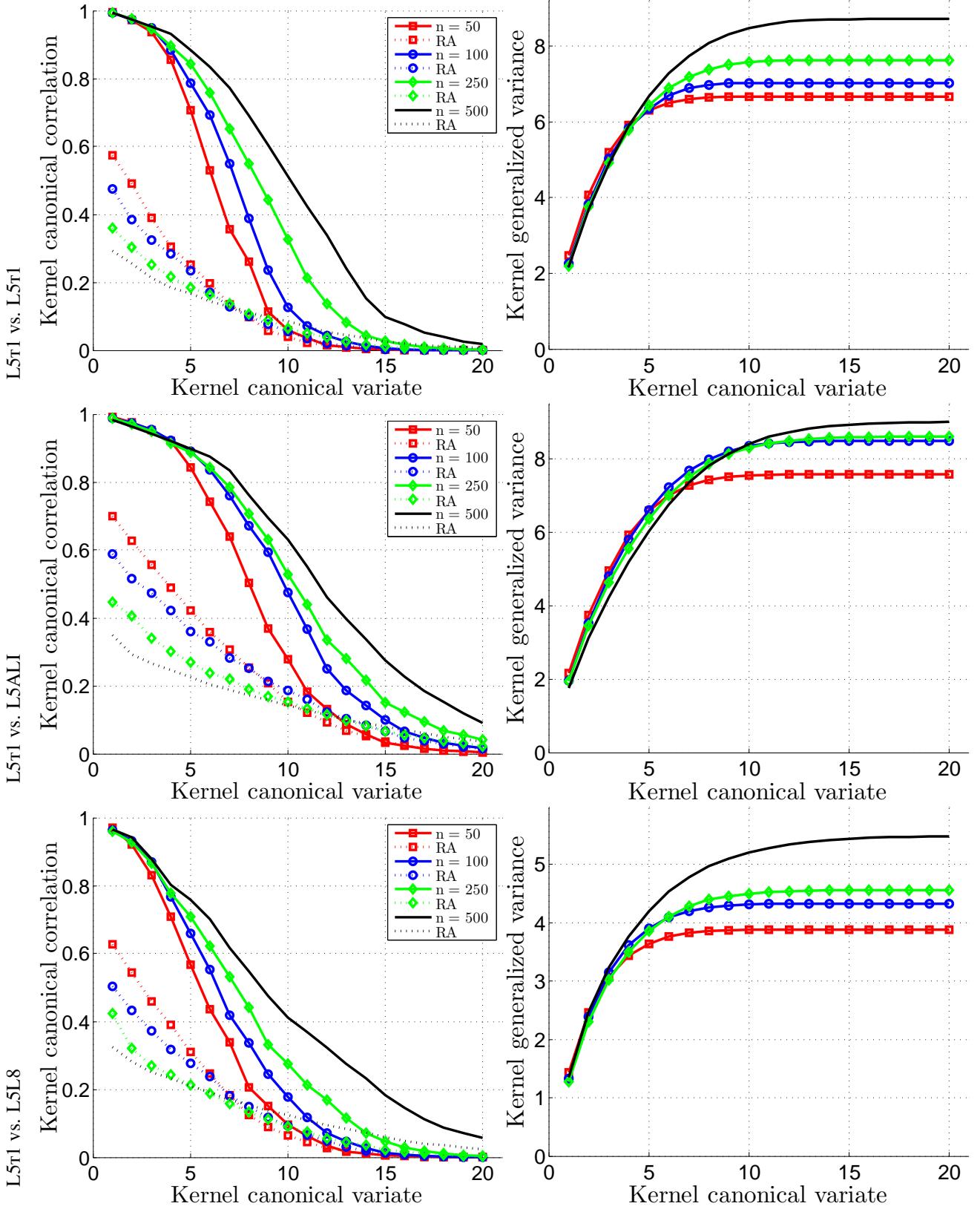


Figure 6: Left column: average kernel canonical correlations *after* the model selection step. The solid lines represent actual correlations, while the dotted ones refers to correlations between random associations (RA), both averaged over 10 runs. Right column: the cumulative kernel Generalized Variance, explaining the dependence between feature sets in the RKHS, also averaged over the 10 runs of the experiments.

study). By comparing with its linear variant (the MAD), the improvements brought by employing nonlinear kernel methods are evident, although the computational load grows. The latter changes from a scaling proportional to the dimensionality of the original datasets to one depending on the number of training samples used. The informativeness of the input space has much less influence on the final change detection results when using the kCCA rather than the MAD.

When performing the relative spectral alignment, we rely on a set of unchanged samples selected from the pair of images to learn the optimal transformation. However, particular attention has to be paid to this task, although the selection of unchanged samples is relatively easy compared to provide a full labeling for all the change transitions. Ideally, the training set should cover the variability of the unchanged class to guarantee stability of the results. However, this requirement can be greatly relaxed if the unchanged areas only include few spectral classes. Learning the projections from a training set guarantees that the projections of each image are maximally aligned for unchanged samples. Consequently, pixel-based change detection can discover these deviations from the aligned pixels.

One of the main drawback of the kCCA resides in the selection of the free hyperparameters. In this paper, we proposed a completely automatic strategy which relies on “side products” of the kCCA transformation. The selection of hyperparameters forcing the kCCA to extract proper data correlations is revealed to be a good strategy. Moreover, we adopted a very simple yet effective way to select the dimensionality of the latent space. The kGV offers a principled measure to select the number of components which contributes to the total amount of extracted information. We are able to separate the components which carry the most of the information by estimating this measure in a cumulative manner over the latent dimensions. Since canonical variates are ordered following their correlation, once the measure saturates the remaining components does not improve the dependency between data. Finally, as depicted by the fully supervised experiments over the Spetses data, labeled samples can only improve the relative alignment strategy. If such samples are available, we advocate their use to refine the model selection strategy.

In this work we employed the CVA as a change detection model since it is a very basic yet powerful approach, which allows to evaluate the goodness of the alignment by working directly on the distribution of the transformed difference images. Obviously, CVA has its own limitations and nothing prevents the use of more complex change detection approaches. Particularly beneficial would be approaches also modeling the spatial relationships of samples (e.g. Markov random fields). In this case, we would expect a global improvement in the numerical accuracies, and a reduced difference between MAD and kCCA results. However, this would have made the discussion of the properties of the relative spectral alignment approach more difficult, since it would have been more complex to disentangle its contribution from the one of the change detection system. We leave the choice of the final change detection model to the user.

5. Conclusions

In this paper, we presented a framework to perform a nonlinear relative spectral alignment between optical cross-sensor acquisitions. This step allows to perform subsequent change detection with any method. The presented approach may be seen as a preprocessing technique which de facto enables any existing change detection method to operate on cross-sensor data. By exploiting a set of samples belonging to unchanged area the kernel canonical correlation analysis transformation (kCCA) allows to align in a relative manner the spectrum of the two datasets. The proposed approach does not need full supervision, which may be a major constraint in real change detection scenarios.

Furthermore, we have proposed a fully automatic model selection strategy bypassing the need for labeled pixels to validate both the hyperparameters and the latent dimensionality. The model selection strategy encodes a two-steps reasoning exploiting only theoretical features of the kCCA. It is based on the analysis of differences between actual and randomized correlations and on the extracted dependency from the cross-sensor acquisitions. We assumed that once the unchanged samples are well modeled, changes may be discovered as pixel-wise deviations from the background. In our case, simple image differencing with the change vector analysis (CVA) in the transformed domains was able to perform change detection in a satisfactory manner.

In this paper we dealt only with pixel-wise approaches for both the relative spectral alignment and change detection. Further research should address the use of the spatial context to further improve both the alignment as well as the discrimination of changes. This would open a large application area for very high resolution imagery, since in these cases the semantic classes cannot be modeled pixel-wise.

Acknowledgements

This work has been supported by the Swiss National Science Foundation (grants P2LAP2-148432 (MV, <http://p3.snf.ch/Project-148432>) and PZ00P2-136827 (DT, <http://p3.snf.ch/project-136827>); and by the Spanish Ministry of Economy and Competitiveness (MINECO) under project LIFE-VISION TIN2012-38102-C03-01 (GCV). Bastrop data available from the U.S. Geological Survey (<http://earthexplorer.usgs.gov/>).

6. References

References

- Alberga, V., 2009. Similarity measures of remotely sensed multi-sensor images for change detection applications. *Remote Sens.* 1, 122–143.
- Amorós-López, J., Gómez-Chova, L., Alonso, L., Guanter, L., Moreno, J., Camps-Valls, G., 2011a. Regularized multiresolution spatial unmixing for ENVISAT/MERIS and Landsat/TM image fusion. *IEEE Geosci. Remote Sens. Letters* 8, 844–848.
- Amorós-López, J., Izquierdo-Verdiguier, E., Gómez-Chova, L., Muñoz Marí, J., Rodríguez-Barreiro, J.Z., Camps-Valls, G., Calpe-Maravilla, J., 2011b. Land cover classification of VHR airborne images for citrus grove identification. *ISPRS J. Photogrammetry Remote Sens.* 66, 115–123.

- Arenas-García, J., Petersen, K., Camps-Valls, G., Hansen, L., 2013. Kernel multivariate analysis framework for supervised subspace learning: A tutorial on linear and kernel multivariate methods. *IEEE Sig. Proc. Magazine* 30, 16–29.
- Bach, F.R., Jordan, M.I., 2002. Kernel independent component analysis. *J. Mach. Learn. Res.* 3, 1–48.
- Biessmann, F., Meinecke, F.C., Gretton, A., Rauch, A., Rainer, G., Logothetis, N.K., Robert-Müller, K., 2010. Temporal kernel CCA and its application in multimodal neuronal data analysis. *Mach. Learn.* 79, 5–27.
- Bovolo, F., Bruzzone, L., Marconcini, M., 2008. A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure. *IEEE Trans. Geosci. Remote Sens.* 46, 2070–2082.
- Bovolo, F., Marchesi, S., Bruzzone, L., 2011. A framework for automatic and unsupervised detection of multiple changes in multitemporal images. *IEEE Trans. Geosci. Remote Sens.* 50, 2196–2212.
- Camps-Valls, G., Gómez-Chova, L., Muñoz-Marí, J., Rojo-Álvarez, J.L., Martínez-Ramón, M., 2008. Kernel-based framework for multi-temporal and multi-source remote sensing data classification and change detection. *IEEE Trans. Geosci. Remote Sens.* 46, 1822–1835.
- Canty, M.J., 2010. Image analysis, classification and change detection in remote sensing. 2nd ed., CRC Press, Taylor and Francis.
- de Carvalho Júnior, O.A.J., Guimarães, R.F., Correia-Silva, N., Gillespie, A.R., Trancoso-Gómes, R.A., Rosa-Silva, C., Ferreira de Carvalho, A.P., 2013. Radiometric normalization of temporal images combining automatic detection of pseudo-invariant features from the distance and similarity spectral measures, density scatterplot analysis, and robust regression. *Remote Sens.* 5, 2763–2794.
- Chatelain, F., Tourneret, J.Y., Ingla, J., 2008. Change detection in multi-sensor SAR images using bivariate Gamma distributions. *IEEE Trans. Im. Processing* 17, 249–258.
- De Bie, T., Cristianini, N., Rosipal, R., 2004. Eigenproblems in pattern recognition, in: Bayro-Corralchano, E. (Ed.), *Handbook of computational geometry for pattern recognition, computer vision, neurocomputing and robotics*. Springer Verlag, pp. 129–171.
- Desclée, B., Simonetti, D., Mayaux, P., Achard, F., 2013. Multi-sensor monitoring system for forest cover change assessment in Central Africa. *IEEE J. Sel. Topics Appl. Earth Observ.* 6, 110–120.
- Fernández-Prieto, D., Marconcini, M., 2011. A novel partially supervised approach to targeted change detection. *IEEE Trans. Geosci. Remote Sens.* 49, 5016–5038.
- Gao, F., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the landsat and MODIS surface reflectance: predict daily Landsat surface reflectance. *IEEE Trans. Geosci. Remote Sens.* 44, 2207–2218.
- Gómez-Chova, L., Camps-Valls, G., Muñoz-Marí, J., Calpe-Maravilla, J., 2008. Semisupervised image classification with Laplacian support vector machines. *IEEE Geosci. Remote Sens. Letters* 5, 336–340.
- Gómez-Chova, L., Izquierdo-Verduguer, E., Amorós-López, J., Muñoz Marí, J., Camps-Valls, G., 2013. Kernel change discriminant analysis for multitemporal cloud masking, in: *IEEE Int. Geosci. Remote Sens. Symp. IGARSS*, Melbourne (AUS), pp. 2974 – 2977.
- Gretton, A., Herbrich, R., Smola, A., Bousquet, O., Schölkopf, B., 2005. Kernel methods for measuring independence. *J. Mach. Learn. Res.* 6, 2075–2129.
- Haroon, D.R., Szedmak, S., Shawe-Taylor, J., 2004. Canonical correlation analysis: An overview with application to learning methods. *Neural Comp.* 16, 2639–2664.
- Healey, S.P., Yang, Z., Cohen, W.B., Pierce, J.D., 2006. Application of two regression-based methods to estimate the effects of partial harvest on forest structure using Landsat data. *Remote Sens. Env.* 101, 115–126.
- Hotelling, H., 1936. Relations between two sets of variates. *Biometrika* 28, 312–377.
- Li, X., Ling, F., Du, Y., Feng, Q., Zhang, Y., 2014. A spatialtemporal Hopeld neural network approach for super-resolution land cover mapping with multi-temporal different resolution remotely sensed images. *ISPRS J. Photogrammetry Remote Sens.* 93, 76–87.
- Longbotham, N., Glenn, T., Zare, A., Volpi, M., Tuia, D., Cristophe, E., Michel, J., Ingla, J., Chanussot, J., Du, Q., Pacifici, F., 2012. Multi-modal change detection, application to the detection of flooded areas: Outcome of the 2009-2010 data fusion contest. *IEEE J. Sel. Topics Appl. Earth Observ.* 5, 1–12.
- Maulik, U., Chakraborty, D., 2013. Learning with transductive SVM for semisupervised pixel classification of remote sensing imagery. *ISPRS J. Photogrammetry Remote Sens.* , 66–78.
- Mubea, K., Menz, G., 2012. Monitoring land-use change in Nakuru (Kenya) using multi-sensor satellite data. *Adv. Remote Sens.* 1, 74–84.
- Nielsen, A.A., 2002. Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data. *IEEE Trans. Im. Proc.* 11, 293–305.
- Nielsen, A.A., 2007. The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data. *IEEE Trans. Image. Proc.* 16, 463–478.
- Nielsen, A.A., Conradsen, K., Simpson, J.J., 1998. Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies. *Remote Sens. Env.* 64, 1–19.
- Nielsen, A.A., Vestergaard, J.S., 2013. A kernel version of multivariate alteration detection, in: *IEEE Int. Geosci. Remote Sens. Symp. IGARSS*, Melbourne (AUS), pp. 3451–3454.
- Pohl, C., Van Genderen, J.L., 1998. Multisensor image fusion in remote sensing: concepts, methods and applications. *Int. J. Remote Sens.* 19, 823–854.
- Qin, Y., Niu, Z., Chen, F., Li, B., Ban, Y., 2013. Object-based land cover change detection for cross-sensor images. *Int. J. Remote Sens.* 34, 6723–6737.
- Röder, A., Kuemmerle, T., Hill, J., 2005. Extension of retrospective datasets using multiple sensors. An approach to radiometric intercalibration of Landsat TM and MSS data. *Remote Sens. Env.* 9, 195–210.
- Roemer, H., Kaiser, G., Sterr, H., Ludwig, L., 2010. Using remote sensing to assess tsunami-induced impacts on coastal forest ecosystems at the Andaman Sea coast of Thailand. *Nat. Hazards Earth Syst. Sci.* 10, 729–745.
- Schmid, T., Koch, M., Gumizzio, J., 2005. Multisensor approach to determine changes of wetland characteristics in semiarid environments (Central Spain). *IEEE Trans. Geosci. Remote Sens.* 43, 2516–2525.
- Schölkopf, B., Smola, A.J., 2002. *Learning with Kernels: support vector machines, regularization, optimization, and beyond*. MIT press, Cambridge (MA).
- Shao, Y., Lunetta, R.S., 2012. Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data point. *ISPRS J. Photogrammetry Remote Sens.* 70, 78–87.
- Shawe-Taylor, J., Cristianini, N., 2004. *Kernel Methods for Pattern Analysis*. Cambridge University Press.
- Singh, A., 1989. Digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* 10, 989–1003.
- de la Torre, F., 2012. A least-squares framework for component analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 34, 1041–1055.
- Tuia, D., Camps-Valls, G., Matasci, G., Kanevski, M., 2010. Learning relevant image features with multiple-kernel classification. *IEEE Trans. Geosci. Remote Sens.* 48, 3780–3791.
- Tuia, D., Volpi, M., Trolliet, M., Camps-Valls, G., 2014. Semisupervised manifold alignment of multimodal remote sensing images. *IEEE Trans. Geosci. Remote Sens.* 52, 7708–7720.
- Turker, M., Asik, O., 2005. Land use change detection at the rural-urban fringe using multi-sensor data in Ankara, Turkey. *Int. J. Geoinfo.* 1, 27–36.
- Villa, G., Moreno, J., Calera, A., Amorós-López, J., Camps-Valls, G., Domenech, E., Garrido, J., González-Matesanz, J., Gómez-Chova, L., Martínez, J.À., Molina, S., Pece, J.J., Plaza, N. and Porcuna, A., Tejeiro, J.A., Valcárcel, N., 2013. Spectro-temporal reflectance surfaces: a new conceptual framework for the integration of remote-sensing data from multiple different sensors. *Int. J. Remote Sens.* 34, 3699–3715.
- Volpi, M., de Morsier, F., Camps-Valls, G., Kanevski, M., Tuia, D., 2013a. Multi-sensor change detection based on nonlinear canonical correlations, in: *IEEE Int. Geosci. Remote Sens. Symp. IGARSS*, Melbourne (AUS), pp. 1944 – 1947.
- Volpi, M., Tuia, D., Kanevski, M., Bovolo, F., Bruzzone, L., 2013b. Supervised change detection in VHR images using contextual information and support vector machines. *Int. J. Applied Earth Obs. Geoinfo.* 20, 77–85.
- Wulder, M.A., Butson, C.R., White, J.C., 2008. Cross-sensor change detection over a forested landscape: options to enable continuity of medium spatial resolution measures. *Remote Sens. Env.* 112, 796–809.
- Yang, X., Lo, C.P., 2000. Relative radiometric normalization performance for change detection from multi-date satellite images. *Photogramm. Eng. Remote Sens.* 66, 967–980.