



Nonlinear Time-Series Adaptation for Land Cover Classification

Adeline Bailly, Laetitia Chapel, Romain Tavenard, Gustau Camps-Valls

► **To cite this version:**

Adeline Bailly, Laetitia Chapel, Romain Tavenard, Gustau Camps-Valls. Nonlinear Time-Series Adaptation for Land Cover Classification. IEEE Geoscience and Remote Sensing Letters, IEEE - Institute of Electrical and Electronics Engineers, 2017, .

HAL Id: halshs-01515283

<https://halshs.archives-ouvertes.fr/halshs-01515283>

Submitted on 27 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Nonlinear Time Series Adaptation for Land Cover Classification

Adeline Bailly, Laetitia Chapel, Romain Tavenard, Gustau Camps-Valls, *Senior Member, IEEE*

Abstract—Automatic land cover classification from satellite image time series is of paramount relevance to assess vegetation and crop status, with important implications in agriculture, biofuels and food. However, due to the high cost and human resources needed to characterize and classify land cover through field campaigns, a recurrent limiting factor is the lack of available labeled data. On top of this, the bio-geo-physical variables exhibit particular temporal structures that need to be exploited. Land cover classification based on image time series is very complex because of the data manifold distortions through time. We propose the use of the kernel manifold alignment (KEMA) method for domain adaptation of remote sensing time series before classification. KEMA is nonlinear, semi-supervised, and reduces to solve a simple generalized eigenproblem. We give empirical evidence of performance through classification of biophysical (LAI, fAPAR, FVC, NDVI) time series at a global scale.

Index Terms—Land cover classification, time series, domain adaptation, manifold alignment, kernel methods

I. INTRODUCTION

Automatic land use and land cover classification from time series of satellite images is of paramount relevance to assess vegetation and crop status, with important implications in agriculture, biofuels and food. Producing an accurate and up-to-date land cover map is an important topic for remote sensing applications. The amount of remotely sensed data is constantly growing because of the new and upcoming satellite platforms available. However, due to the high cost of field campaigns in terms of money and human resources, only a very limited number of labeled data is available. Additionally, fast changes in landscape and land use would require a never-ending labeling, which is totally unfeasible. These problems cause a remarkable scarcity and heterogeneity of labeled data. From a machine learning point of view, this implies the need to design *semi-supervised learning* algorithms able to exploit both the little supervised information and the wealth of unlabeled Earth Observation (EO) data. Differences in EO acquisition conditions (*e.g.* atmosphere status and sensor field of view) for each satellite revisit time make the acquired data statistically different. Both observations and derived products thereof show up time and space distortions. In particular, the bio-geo-physical

variables exhibit distinct temporal structures. Actually, even for the same land cover class, time series show similar, yet distinct, properties. This makes the problem of land cover classification using EO time series very complex because of shifts, twists and warpings of the data manifold evolving through time. Therefore, the problem also calls for *domain adaptation* and *manifold alignment* techniques [1].

Adapting classifiers to (even slightly) shifted data distributions is an old problem in EO data processing, which started in the 1970s with the signature extension field [2], [3], and then evolved, due to the technological advances in both sensor and processing routines, into what is generally referred to as the *transfer learning* problem [4], [5]. Adapting the classifier to cope with the dataset shift among acquisitions has been recently introduced using semi-supervised strategies [6], [7] and active learning [8]. The related field of manifold alignment [1] aims at matching a set of data sources of potentially different dimensionality using feature extraction under pairwise proximity constraints [9]. Manifold alignment has been recently applied to remote sensing data to accommodate spatial and spectral distortions [10], [11]. However, previous approaches did not focus on the important problem of adaptation of time series, which imposes more challenging problems because of the higher dimensionality of the available data and because of the online and varying distortions. It goes without saying that the alignment of EO data should exploit both labeled and unlabeled data. In this paper, we use a kernel manifold alignment method for EO time series classification. We rely on the Kernel Manifold Alignment (KEMA) that has been proposed in [12]. KEMA is a particularly convenient method for our purposes: efficient with few labeled points of high dimension, robust to noise and varying distortions, and it boils down to solving a generalized eigenproblem. Here we apply KEMA in the time series adaptation context and evaluate its performance on a newly proposed EO dataset specifically tailored to assess quality of time series domain adaptation methods. This dataset features standard adaptation problems from the EO domain, including adaptation to different geographical areas, adaptation across years, satellites and vegetation indices. A detailed description of the dataset is given in the experimental section.

The remainder of the paper is organized as follows. Section II briefly reviews KEMA and its adaptation to time series. In Section III, we give empirical evidence of its performance in two real EO time-series classification problems that require adaptation. Section IV concludes the paper.

II. NONLINEAR ADAPTATION OF TIME SERIES

This section reviews the kernel manifold alignment method [12] for domain adaptation which is exploited in this paper for the time series adaptation problem.

Manuscript received March 13, 2017.

Adeline Bailly and Romain Tavenard are with LETG-Rennes COSTEL, Université Rennes 2, Place Recteur Henri le Moal, 35000 Rennes (France). E-mail: firstname.lastname@univ-rennes2.fr

Adeline Bailly and Laetitia Chapel are with IRISA, Université de Bretagne-Sud, Campus de Tohannic, 56000 Vannes (France). E-mail: firstname.lastname@irisa.fr

Gustau Camps-Valls is with the Image Processing Laboratory (IPL), Universitat de València, Catedrático A. Escardino - 46980 Paterna, València (Spain). E-mail: gustau.camps@uv.es

This paper has been supported by the ANR project ASTERIX (ANR-13-JS02-0005-01), by l'École des docteurs de l'UBL as well as by the Brittany Region. Gustau Camps-Valls acknowledges support by the European Research Council (ERC) under the ERC-CoG-2014 project 647423.

A. Semisupervised linear adaptation

We are given D domains \mathcal{X}_i , $i = 1, \dots, D$, and the corresponding time-series matrices defined therein, $\mathbf{X}_i \in \mathbb{R}^{d_i \times n_i}$, containing n_i examples (labeled, l_i , and unlabeled, u_i , with $n_i = l_i + u_i$) of dimension d_i , and $n = \sum_i n_i$. The Semi-Supervised Manifold Alignment (SSMA) method in [13] maps all the data to a latent space \mathcal{F} such that samples belonging to the same class become closer, those of different classes are pushed far apart, and the geometry of the data manifolds is preserved. Therefore, three entities have to be considered *per* domain: 1) a similarity matrix \mathbf{W}_s will have components $W_s^{\ell m} = 1$ if \mathbf{x}_ℓ and \mathbf{x}_m belong to the same class, and 0 otherwise (unlabeled included); 2) a dissimilarity matrix \mathbf{W}_d will have entries $W_d^{\ell m} = 1$ if \mathbf{x}_ℓ and \mathbf{x}_m are not from the same class, and 0 otherwise (unlabeled included); and 3) a similarity matrix that represents the topology of each given domain, \mathbf{W} , *e.g.* constructed using a k -nearest neighbor graph (k -NNG). The three different entities lead to three different graph Laplacians: \mathbf{L}_s , \mathbf{L}_d , and \mathbf{L} , respectively. Graph Laplacians are defined by $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where \mathbf{D} is a diagonal matrix with entries $D^{\ell\ell} = \sum_m W^{\ell m}$. Then, the embedding must minimize a joint cost function essentially given by the eigenvectors corresponding to the smallest non-zero eigenvalues of the generalized eigenvalue problem:

$$\mathbf{Z}(\mu\mathbf{L} + \mathbf{L}_s)\mathbf{Z}^\top \mathbf{v} = \lambda \mathbf{Z}\mathbf{L}_d\mathbf{Z}^\top \mathbf{v}, \quad (1)$$

where \mathbf{Z} is a block diagonal matrix containing the data matrices \mathbf{X}_i ; \mathbf{v} contains in the columns the eigenvectors organized in rows for the particular domain, $\mathbf{v} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_D]^\top$ and μ is a weight parameter between the labeled information and the intra-domain topology. The method allows the extraction of a maximum of $N_f = \sum_{i=1}^D d_i$ features, that serve for projecting the time series of domain \mathcal{X}_i into the common latent space \mathcal{F} by simply $P_f(\mathbf{X}_i) = \mathbf{v}_i^\top \mathbf{X}_i$. This method was successfully used in adapting remote sensing images [14].

B. Semi-supervised nonlinear adaptation

In order to cope with nonlinear distortions in the time series, the previous method was kernelized and was recently applied in change detection in [12]. The procedure first maps the data to a Hilbert space, apply the representer's theorem and replace the dot products therein with reproducing kernel functions. Let us first map the D different datasets with D mapping functions to D in principle different Hilbert spaces \mathcal{F}_i of dimension F_i , $\phi_i(\cdot) : \mathbf{x} \rightarrow \phi_i(\mathbf{x}) \in \mathcal{F}_i$, $i = 1, \dots, D$. It should be noted that, however, kernelization is only possible thanks to the use of the direct sum of Hilbert spaces, hence $F = \sum_i^D F_i$, a well-known property of Functional Analysis Theory [15]. We thus resort to the definition of D corresponding Riesz representation theorems so the corresponding eigenvectors $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_F]^\top$ can be expressed as linear combinations of mapped samples, which in matrix notation $\mathbf{U} = \Phi\mathbf{\Lambda}$, where Φ is a block diagonal matrix containing the mapped time series $\Phi_i = [\phi_i(\mathbf{x}_1), \dots, \phi_i(\mathbf{x}_{n_i})]^\top$. After replacing the dot products with the corresponding kernels, $\mathbf{K}_i = \Phi_i^\top \Phi_i$, a dual problem is obtained:

$$\mathbf{K}(\mu\mathbf{L} + \mathbf{L}_s)\mathbf{K}\mathbf{\Lambda} = \lambda \mathbf{K}\mathbf{L}_d\mathbf{K}\mathbf{\Lambda}, \quad (2)$$

where \mathbf{K} is a block diagonal matrix containing the kernel matrices \mathbf{K}_i . Now the generalized eigenproblem becomes of

size $n \times n$ instead of $d \times d$, and the number of extracted features becomes $N_f = \sum_{i=1}^D n_i$, which will be also referred as to $\dim(\mathcal{F})$. As for the linear case, now projection to the latent space requires first mapping the data \mathbf{X}_i to its corresponding Hilbert space \mathcal{F}_i , thus leading the mapped data Φ_i , and then to apply the projection vector \mathbf{u}_i defined therein, $P_f(\mathbf{X}_i) = \mathbf{u}_i^\top \Phi_i = \alpha_i^\top \Phi_i^\top \Phi_i = \alpha_i^\top \mathbf{K}_i$. Therefore, projection to the latent space is possible through the use of dedicated kernels. A simple linear discriminant analysis classifier (L-DAC) is used here for the classification step in \mathcal{F} , where a good representation of the data can be achieved by using the data projected onto the top eigenvectors.

C. Adaptation of time series

By aligning domains in a semi-supervised fashion, KEMA allows one to take into account both intra-domain similarity and cross-domain class membership information. In the context of EO time series considered here, this results in allowing temporal distortions and shifts across domains while being strict on the temporal behaviour inside each domain. In other words, as shown later in the experimental section of the paper, if one wants to use information from time series extracted in Europe in order to classify time series from South America, it is important to allow for time shift between domains (as Europe and South America are not located on the same hemisphere, vegetation growth will not happen at the same period of the year). On the contrary, inside a domain, time shifts should be strongly penalized as the date of growth is an important factor to be used for classification (*cf.* Fig. 1a in which grasslands and croplands mostly differ by the date at which their growth starts). As a consequence, we choose to compare time series using a standard Euclidean Distance to build the k -NNG that is used for the W matrix computation.

III. EXPERIMENTAL RESULTS

In this section, we present evaluation of the KEMA method for time series domain adaptation in EO applications. We rely on a new dataset we introduce that is specifically designed to assess performance of time series adaptation in several domain adaptation problems related to EO and land cover classification: Google Earth Engine - Time Series Domain Adaptation (GEE-TSDA). We also consider the BELMANIP¹ (BEenchmark Land Multisite ANalysis and Intercomparison of Products) dataset, allowing the study of adaptation of different biophysical parameters worldwide for a full year of *in situ* data acquisitions.

A. Experimental setup

For all experiments presented in the following, KEMA is compared to i) its non-kernelized counterpart SSMA and ii) the raw time series, denoted RD (Raw Data). In all cases, we use a L-DAC to perform the classification. For RD, we present results for two variants: RD-1 (for 1 domain) is trained using only labeled data from the target domain to be classified whereas RD-2 (for 2 domains) will use exactly the same set of training labeled data as SSMA and KEMA, regardless of which domain the data comes from. It is expected that, if both

¹Available at <http://calvalportal.ceos.org/web/olive/site-description>

TABLE I: Domains description of the GEE-TSDA dataset.

	# series	Length	Sat.	Temp. res.	Area	Yr.	VI
\mathcal{S}	311	46	MODIS	8 days	Europe	2011	NDVI
\mathcal{A}	338	46	MODIS	8 days	S. Amer.	2011	NDVI
\mathcal{B}	344	46	MODIS	8 days	N. Amer.	2011	NDVI
\mathcal{C}	389	46	MODIS	8 days	Europe	2003	NDVI
\mathcal{D}	355	41	LANDSAT	8 days	Europe	2011	NDVI
\mathcal{E}	339	91	MODIS	4 days	Europe	2011	LAI

TABLE II: Number of time series per class (GEE-TSDA).

	Total	Evergreen	Deciduous	Shrublands	Savannas	Grasslands	Croplands
\mathcal{S}	311	11	27	20	47	14	192
\mathcal{A}	338	48	60	58	94	46	32
\mathcal{B}	344	12	31	66	36	121	78
\mathcal{C}	389	16	29	33	42	27	242
\mathcal{D}	355	11	32	21	51	14	226
\mathcal{E}	339	9	31	21	58	19	201

domains are similar, RD-2 should outperform RD-1 as it uses more training data. When domains are very different however, a classifier trained on domain-specific data such as RD-1 is prone to be more reliable than RD-2. SSMA and KEMA have few parameters in common: the number k of neighbors for the k -NNG used for \mathbf{W} (here we fixed to $k = 5$), the μ parameter from equations (1)-(2) (here $\mu = 1$ to have the same weight between labeled information and intra-domain topology), and the maximum number of dimensions of the latent space used in the experiments (set to 10). The latter is chosen to be large enough to ensure the projection to have sufficient information. KEMA also uses an RBF kernel, whose bandwidth was set to the mean of pairwise distance between all time series per domain in order to enforce a domain-specific metric in each domain following prescriptions in [12]. We place ourselves in a typical EO context, in which little supervised information is available: for each of the following experiments, the training set is composed of $l_i = 5$ labeled time series per class, the remaining time series being equally distributed between the set of unlabeled time series and the test set. This dataset hence poses a difficult problem: a low number of labeled training data is available to perform classification in high dimension. Fully functional MATLAB code², as well as all datasets³ used in the experiments, are made available for download, in addition to the code to extract more time series from GEE framework.

B. Aligning global multisite, multisource time series

To assess performance of adaptation techniques for EO, we have gathered data for a wide set of different domains into a new dataset called GEE-TSDA. After presenting the main features of this dataset, we provide a detailed experiment on a specific adaptation problem, showing that KEMA naturally allows one to align noisy and shifted time series, leading to enhanced performances *w.r.t.* classifiers built on raw data. We finally provide results using KEMA for the different adaptation problems to serve as accuracy baselines for GEE-TSDA.

1) *GEE-TSDA dataset description*: The GEE-TSDA dataset has been built using data collected through the Google Earth Engine (GEE) framework and is made publicly available online. Land cover ground truth corresponds to International Geosphere-Biosphere Programme (IGBP) classification that

contains 17 different classes. We have aggregated some of them into 6 final representative global classes: evergreen forest, deciduous forest, shrublands, savannas, grasslands, and croplands. Labels are extracted from the MCD12Q1 MODIS product provided by NASA Land Process Distributed Active Archive Center⁴ which contains yearly land-cover information. We consider different domain adaptation scenarios, corresponding to the main sources of acquisition differences in the EO context. We consider a source domain \mathcal{S} where time series represent the evolution of the Normalized Difference Vegetation Index (NDVI) computed from images acquired by the MODIS satellite and collected across Europe at temporal resolution of 8 days in 2011. For a given class, profiles are likely to differ between acquisition areas; we provide two target domains with data collected across South America (domain \mathcal{A}) and North America (domain \mathcal{B}). We also consider different periods of acquisition (2003, domain \mathcal{C}), acquisition conditions (LANDSAT satellite, domain \mathcal{D}), and bio-geophysical variables (Leaf Area Index (LAI), domain \mathcal{E} with a temporal resolution of 4 days). TABLE I summarizes the six domains of the GEE-TSDA dataset, TABLE II gives the associated number of time series per class and Fig. 1 provides average profiles for each class in each domain, the x -axis and y -axis represent time stamps and their values respectively. In particular, we notice important temporal distortion and shift across domains.

2) *Classification of time series from different geographical areas*: We first consider the problem of adapting from one geographical area to another. This adaptation problem is of high importance as it is expected that pixels from a given class exhibit different temporal behaviours depending on their location on earth. To test this adaptation problem, we use domain \mathcal{A} of the GEE-TSDA dataset that is to be adapted from the source domain. The corresponding considered regions (Europe and South America) were selected as they present important differences. The most obvious shift is because of the season reversal; in addition, agricultural parcels are known to be smaller (on average) in Europe than in South America, which implies that South American pixels have higher chance to be pure pixels (*i.e.* corresponding to a single parcel) than European ones. Figure 2 presents the projection of the time series in the first two dimensions of the latent space defined by KEMA, distinguishing the labeled time series from each domain and the test set. We first notice that the different classes are well-separated in the latent space for the labeled datasets. We also note that, despite a large shift of the time series profiles of the two domains (see Fig. 1a & 1b), KEMA successfully aligns both target and source domains for the labeled set, leading to more discriminative data representation in the test set. Figure 3 presents accuracies obtained for the test time series of the South America area (domain \mathcal{A}) for discriminant analysis applied on the SSMA and KEMA projections for increasing dimensions of the latent space, together with both baselines RD-1 and RD-2. One should first notice that RD-1 outperforms RD-2: mixing the two domains leads to degraded performances as the data have shifted distributions. As soon as the dimension of the latent space is at least 2, both SSMA and KEMA outperform the RD variants: aligning the domains allows for reducing the shift and hence increasing accuracy. Also, KEMA consistently

²<http://github.com/a-bailly/nonLinearAdaptationTimeSeries>

³http://github.com/a-bailly/time_series_data

⁴https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd12q1

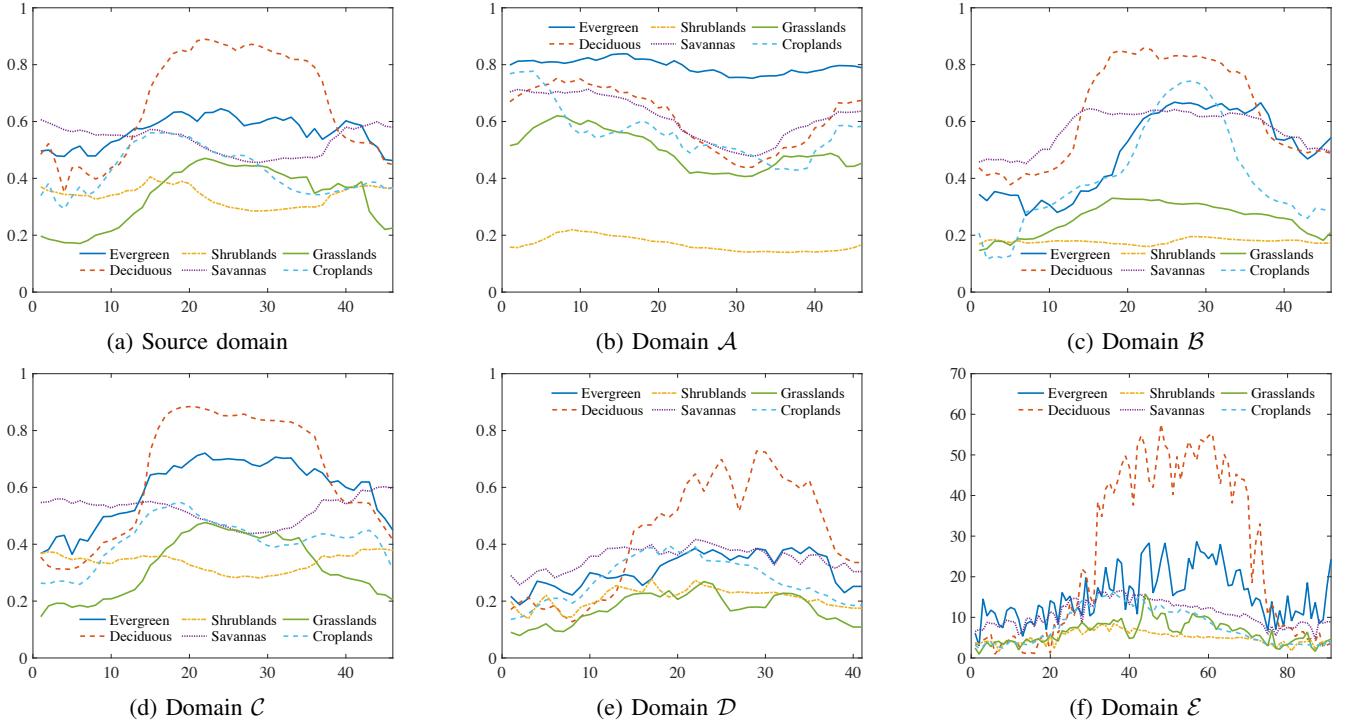


Fig. 1: Mean profiles per class and domain for GEE-TSDA dataset.

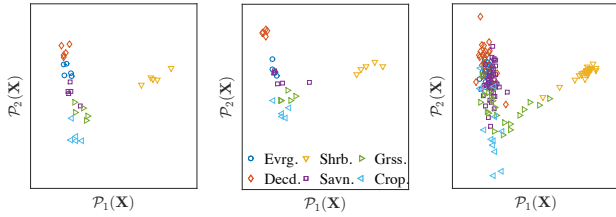


Fig. 2: KEMA projection of time series on the first two dimensions of the latent space \mathcal{F} . From left to right: labeled set from South America, labeled set from Europe, test set from South America

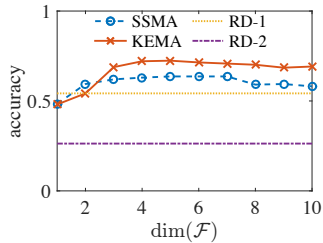


Fig. 3: Accuracy for the test set of South America time series as a function of the dimensionality of the latent space \mathcal{F} .

outperforms SSMA on this dataset, for almost all considered dimensions of the latent space. Finally, we observe that accuracy remains stable as long as the dimensionality of the latent space is sufficient to capture most of the information.

3) *Other domain adaptation problems:* We provide the results of the experiments of domain adaptation problems \mathcal{A}

TABLE III: Accuracies for the different domain adaptation scenarios for the GEE-TSDA dataset.

Target	RD-1	RD-2	SSMA	KEMA
\mathcal{A}	0.542	0.263	0.636	0.724
\mathcal{B}	0.554	0.411	0.627	0.631
\mathcal{C}	0.293	0.281	0.376	0.532
\mathcal{D}	0.188	0.169	0.265	0.412
\mathcal{E}	0.515	0.220	0.385	0.534

to \mathcal{E} in TABLE III. We choose to construct a latent space of dimension 5, as KEMA is known to need few dimensions to build an accurate latent space [12]. We also observed that in all the experiments we performed, the accuracies remain stable after 4 or 5 dimensions.

a) *Multi-location classification:* We consider another problem of adapting from different geographical areas, considering North America (domain \mathcal{B}). Despite Europe and North America areas are both located in the same hemisphere, some classes exhibit different behavior (e.g. croplands) in the two domains. This observation is confirmed by the fact that RD-1 outperforms RD-2 to a great extent. As for previous experiment, adaptation methods (SSMA and KEMA) improve on RD baselines, and exhibit comparable accuracies.

b) *Multi-year classification:* In scenario \mathcal{C} , we study the adaptation between time series acquired at different years that are likely to induce small temporal shifts due to the variability of meteorological conditions from one year to another. We observe that RD-1 and RD-2 have similar performance, which indicates a moderately difficult adaptation problem. Again, KEMA lead to the best performances, followed by SSMA.

c) *Multi-satellite classification:* Our following set of experiments concerns adaptation between different satellites (domain \mathcal{D}). Both domains have a time resolution of 8 days,

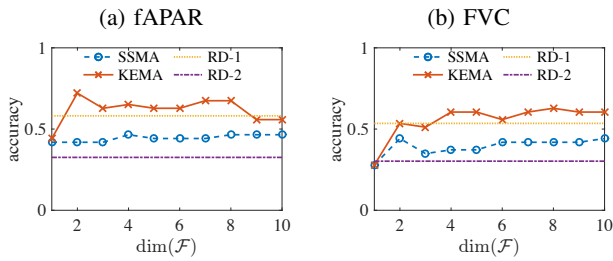


Fig. 4: Accuracy obtained on dataset BELMANIP for (a) fAPAR and (b) FVC time series.

however time series are not of the same length (46 time-instants per year for MODIS, 41 for LANDSAT). Linear interpolation is hence applied when comparing time series from LANDSAT and MODIS with RD-2. Again, KEMA outperforms both SSMA and the RD baselines for this adaptation problem.

d) *Multi-modality classification*: We focus in this last experiment on the adaptation between time series of different vegetation indices (domain \mathcal{E}). NDVI time series have 8-day time resolution whereas LAI ones have 4-day time resolution. The relationship between NDVI and LAI vegetation parameters is well-known, [16] stated that NDVI increases almost linearly with LAI until a full cover is reached. Beyond that point, an increasing LAI leads to a small and asymptotic increase of NDVI. We first note here that RD-2 does not perform well at classifying LAI time series. However, the use of the target domain to classify the source domain \mathcal{E} allows the improvement of the classification results when using the KEMA method.

C. Aligning time series across modalities

To further evaluate performances of KEMA for EO time series adaptation, we experiment on the BELMANIP dataset, composed of 445 sites where measurements and ground validation have been performed in 2012. For each site, we have gathered time series for two different kinds of land biophysical modalities: Fractional Vegetation Cover (FVC) and Fraction of Absorbed Photosynthetically Active Radiation (fAPAR). As for the GEE-TSDA dataset, land cover ground truth corresponds to the IGPB classification aggregated into 6 classes (evergreen forest, deciduous forest, shrublands, savannas, grasslands, and croplands). In this study, we focus on North America sites, as they feature the highest class diversity into a limited space. Resulting dataset is composed of 68 time series per modality. The domain adaptation problem we consider here is that of classifying time series from one modality (either FVC or fAPAR) using information from both modalities. Figure 4 presents accuracy obtained for the test time series of each domain. One should notice that, in both cases, RD-1 outperforms RD-2: mixing the two domains leads to degraded performances as the data have shifted distributions. As the difference is rather significant, it suggests that the shift is strong. Finally, KEMA outperforms compared methods on this dataset for most dimensions of the latent space.

IV. CONCLUSIONS

The classification of land cover and land use from time series of satellite observations is a challenging problem. The derived EO bio-geophysical time series exhibit complex characteristics:

missing points due to the presence of clouds, ice and snow alter time series parameters, complex manifold distortions due to the atmospheric effects and sensor acquisitions induce a wide variety of noise sources, while multiple sensor platforms acquire time series of remote sensing data in different spatial, temporal and spectral resolutions lead to unpaired time series difficult to harmonize and classify jointly. On top of this, labeling requires high cost and human resources through field campaigns hence supervised classification is not feasible. In order to address all these problems, we have used a semisupervised nonlinear manifold alignment method for time-series classification. The method can cope with strong distortions and misalignments of the time series manifolds, and it reduces to solve a simple generalized eigenproblem. To evaluate performance of the method, we have introduced a new dataset that covers the main EO-related domain adaptation problems and given empirical evidence of the excellent performance of KEMA. Results suggest that KEMA can natively blend information from heterogeneous and distorted EO time series domains effectively. As a future work, we plan to derive an online version of KEMA that will be able to deal with the intrinsic sequential nature of time series data, to control computational complexity, and reduce instability of eigenvectors in non-stationary settings.

REFERENCES

- [1] C. Wang, P. Krafft, and S. Mahadevan, "Manifold alignment," in *Manifold Learning: Theory and Applications*, Y. Ma and Y. Fu, Eds. CRC Press, 2011.
- [2] M. D. Fleming, J. S. Berkebile, and R. M. Hoffer, "Computer-aided analysis of LANDSAT-1 MSS data: a comparison of three approaches, including a "modified clustering" approach," Purdue University, LARS information note 072475, 1975.
- [3] I. Olthof, C. Butson, and R. Fraser, "Signature extension through space for northern landcover classification: A comparison of radiometric correction methods," *Remote Sensing Environ.*, vol. 95, no. 3, pp. 290–302, 2005.
- [4] S. J. Pan and Y. Qiang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [5] V. M. Patel, R. Gopalan, R. Li, and R. Chellappa, "Visual domain adaptation: a survey of recent advances," *IEEE Signal Proc. Mag.*, vol. 32, no. 3, pp. 53–69, 2015.
- [6] S. Rajan, J. Ghosh, and M. Crawford, "Exploiting class hierarchy for knowledge transfer in hyperspectral data," *IEEE Trans. Geosci. Remote Sensing*, vol. 44, no. 11, pp. 3408–3417, 2006.
- [7] L. Bruzzone and M. Marconcini, "Domain adaptation problems: A DASVM classification technique and a circular validation strategy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 770–787, 2010.
- [8] G. Matusci, D. Tuia, and M. Kanevski, "SVM-based boosting of active learning strategies for efficient domain adaptation," *IEEE J-STARS*, vol. 5, no. 5, pp. 1335–1343, 2012.
- [9] J. Ham, D. Lee, and L. Saul, "Semisupervised alignment of manifolds," in *Proc. AISTATS*, R. G. Cowell and Z. Ghahramani, Eds., London, UK, 2005, pp. 120–127.
- [10] D. Tuia, M. Volpi, M. Trollet, and G. Camps-Valls, "Semisupervised manifold alignment of multimodal remote sensing images," *IEEE Trans. Geosci. Remote Sensing*, vol. 52, no. 12, pp. 7708–7720, 2014.
- [11] H. Yang and M. Crawford, "Domain adaptation with preservation of manifold geometry for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations*, vol. 9, no. 2, pp. 543–555, 2015.
- [12] D. Tuia and G. Camps-Valls, "Kernel Manifold Alignment for Domain Adaptation," *PLoS ONE*, 2016.
- [13] C. Wang and S. Mahadevan, "Heterogeneous domain adaptation using manifold alignment," in *IJCAI*, Barcelona, Spain, 2011, pp. 1541–1546.
- [14] D. Tuia, M. Volpi, M. Trollet, and G. Camps-Valls, "Semisupervised manifold alignment of multimodal remote sensing images," *IEEE Trans. Geosci. Remote Sensing*, vol. 52, no. 12, pp. 7708–7720, 2014.
- [15] M. Reed and B. Simon, *I: Functional Analysis, Volume 1 (Methods of Modern Mathematical Physics)*, 1st ed. Academic Press, Jan. 1981.
- [16] T. N. Carlson and D. A. Ripley, "On the Relation between NDVI, Fractional Vegetation Cover, and Leaf Area Index," *Remote Sensing Environ.*, vol. 3, no. 62, pp. 241–252, 1997.