

Multisensor Temporal Unsupervised Domain Adaptation for Land Cover Mapping With Spatial Pseudo-Labeling and Adversarial Learning

Emmanuel Capliez¹, Graduate Student Member, IEEE, Dino Ienco², Member, IEEE, Raffaele Gaetano, Nicolas Baghdadi, Adrien Hadj Salah, Matthieu Le Goff³, and Florient Chouteau

Abstract—With the huge variety of Earth observation satellite missions available nowadays, the collection of multisensor remote sensing information depicting the same geographical area has become systematic in practice, paving the way to the further breakthroughs in automatic land cover mapping with the aim to support decision makers in a variety of land management applications. In this context, along with the increase in the volume of data available, the availability of ground-truth (GT) data to train supervised models, which is usually time-consuming and costly, may even be more critical. In this scenario, the possibility to transfer a model learned on a particular time span (*source domain*) to a different period of time (*target domain*), over the same geographical area, can be advantageous in terms of both cost and time efforts. However, such model transfer is challenging due to different climate, weather, or environmental conditions affecting remote sensing data collected at different time periods, resulting in possible distribution shifts between the *source* and *target* domains. With the aim to cope with the multisensor temporal transfer scenario in the context of land cover mapping, where multitemporal and multiscale information are used jointly, we propose a Multisensor, Multitemporal, and Multiscale Spatially Aware Domain Adaptation (M³SPADA) framework, a deep learning methodology that jointly exploits self-training and adversarial learning to transfer a multisensor land cover classifier (MSLCC) from a time period (year) to a different one on the same geographical area. Here, we consider the case in which each domain (*source* and *target*) is described by a pair of remote sensing datasets: a satellite image time series (SITS) of optical images and a single very high spatial resolution (VHR) scene. Experimental evaluation on a real-world study case located in Burkina Faso and characterized by operational constraints

shows the quality of our proposal to deal with the temporal multisensor transfer in the context of land cover mapping.

Index Terms—Data fusion, deep learning, multisensor land cover mapping, satellite image time series (SITS), temporal domain adaptation, very high spatial resolution (VHR) imagery.

I. INTRODUCTION

TODAY, an unprecedented multitude of satellite missions continuously acquires remotely sensed images, thereby opening up new opportunities to monitor the Earth surface at different spatial and temporal scales. Therefore, the same area can be effectively covered by a rich, multifaceted, multisensor, and diverse set of information [1]. For example, a study site can be described by very high spatial resolution (VHR) optical imagery that finely describes the underlying spatial patterns as well as high spatial resolution (HR) satellite image time series (SITS), such as the ones provided by the European Space Agency's Sentinel-2 mission [2], that allows to capture the temporal dynamic of spatiotemporal phenomena (i.e., crop phenology for agricultural land cover).

The remote sensing community has been focusing its efforts for a while now to demonstrate the benefit to leverage multisensor information provided by diverse sensors via recent deep learning approaches [3], [4]. With a particular emphasis on land use and land cover (LULC) mapping, several recent studies [5], [6], [7] have demonstrated superior performances through the combination of multiscale and multitemporal information via modern neural network approaches.

Nonetheless, the majority of the efforts done, to date, in the analysis of multisensor remote sensing data mainly cover the supervised machine learning setting [3], [4]. In this setting, a large amount of reference [or ground truth (GT)] data is required to train the classification model, hence posing serious challenges to its use in contexts where a reduced amount of, or no, reference data is available. For instance, when a land cover map (LCM) has to be updated from previous years, costs and resources related to new field campaigns can prevent the possibility to collect new reference data, thus hindering either the evolution or the train of an updated land cover classification model [8].

A straightforward approach would be to use the existing reference data collected by previous field campaigns or released by public/government agencies, to save time and money for the production/updating of LCMs. This possibility can, on the one hand, take advantage of prior endeavors and, on the other

Manuscript received 14 March 2023; revised 28 April 2023 and 24 May 2023; accepted 10 July 2023. Date of publication 24 July 2023; date of current version 14 August 2023. This work was supported by the French National Association of Research and Technology through the Convention Industrielle de Formation par la REcherche (CIFRE; referred as 2019/1810) Ph.D. Grant. (Corresponding author: Dino Ienco.)

Emmanuel Capliez is with the National Research Institute for Agriculture, Food and the Environment (INRAE), Joint Research Unit on Territory, Environment, Remote Sensing and Spatial Information (UMR TETIS), University of Montpellier, 34090 Montpellier, France, and also with Airbus Defence and Space, 31030 Toulouse, France (e-mail: emmanuel.capliez@inrae.fr).

Dino Ienco and Nicolas Baghdadi are with the National Research Institute for Agriculture, Food and the Environment (INRAE), Joint Research Unit on Territory, Environment, Remote Sensing and Spatial Information (UMR TETIS), University of Montpellier, 34090 Montpellier, France (e-mail: dino.ienco@inrae.fr; nicolas.baghdadi@inrae.fr).

Raffaele Gaetano is with the French Agricultural Research Centre for International Development (CIRAD), Joint Research Unit on Territory, Environment, Remote Sensing and Spatial Information (UMR TETIS), University of Montpellier, 34090 Montpellier, France (e-mail: raffaele.gaetano@cirad.fr).

Adrien Hadj Salah, Matthieu Le Goff, and Florient Chouteau are with Airbus Defence and Space, 31030 Toulouse, France (e-mail: adrien.hadjalah@airbus.com; matthieu.le-goff@airbus.com; florient.f.chouteau@airbus.com).

Digital Object Identifier 10.1109/TGRS.2023.3297077

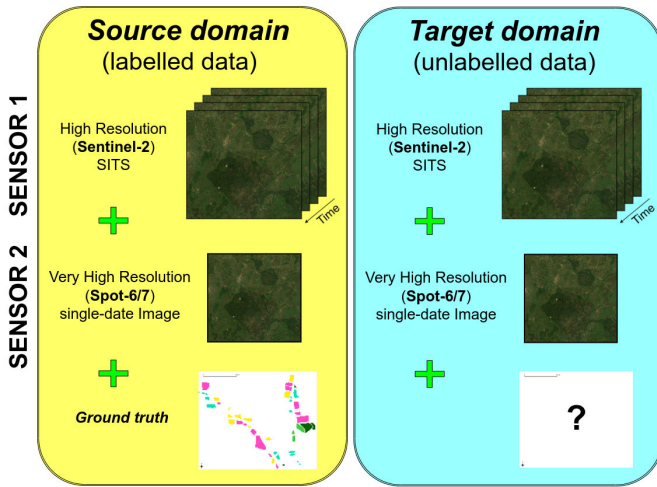


Fig. 1. Multisensor temporal domain adaptation setting. In our context, each domain is described by a multisensor dataset composed of a high-resolution optical (Sentinel-2) SITS and a VHR optical (Spot-6/7) single-date image. Only the source domain is associated with GT (labeled samples) data.

hand, mitigate the necessity of up-to-date reference data on a study area whose accessibility may be compromised.

Here, we start from the fact that directly transferring a multisensor land cover classifier (MSLCC) trained on a particular year (the source domain) to another period of time (the target domain) is challenging, since the two time periods can be affected by different climate, weather, or environmental conditions [9], [10]. This results in shifts or differences in the distributions of the underlying remote sensing data.

In the general field of computer vision and machine learning, addressing the distribution shift problem to adapt a model trained on a source domain to an unlabeled target domain is referred as unsupervised domain adaptation (UDA) [11]. Concerning the field of remote sensing, several approaches have been introduced to cope with UDA in the context of VHR imagery analysis [12], [13], and only few strategies are available for SITS data [10], [14], [15], while, to the best of our literature review, no approach has been proposed, as of now, to cope with the challenging UDA problem in the context of multisensor remote sensing data analysis for the LULC downstream task. More precisely, multiscale and multitemporal optical imagery is jointly exploited in order to take the most out of the complementary of these two sources of information.

Here, we consider the multisensor temporal UDA (MStUDA) problem where both source and target domains are characterized by multisensor information (i.e., a high-resolution optical SITS and an optical single-date VHR image, both available for each domain), while only the source domain is featured by GT (labeled samples) data. This setting is depicted in Fig. 1. The goal is to train an MSLCC from the labeled samples from the multisensor *source domain* and, subsequently, make inference on the multisensor remote sensing data describing the *target domain* where no GT reference is available at training time. More precisely, we focus on the specific case of temporal domain adaptation that is a special UDA case in which data shifts happen between

information coming from different periods of time over the same area. Other kind of UDA setting can involve shifts related to the use of different sensors between source and target domains [16] as well as source and target domains covering different geographical areas [10], but they are out of the scope of the research work proposed in this manuscript.

In addition, we focus our attention on the case of sparse GT data [17] where a limited quantity of annotated surface (with respect to the extent of the area under study) is available. This situation meets operational constraints related to field campaigns [18], [19], thus limiting the use of standard semantic segmentation approaches that require a fully dense reference annotation.

With the aim to cope with the MStUDA problem, in the context of land cover mapping from multitemporal and multiscale remote sensing data, we propose a Multisensor, Multitemporal and Multiscale SPatially Aware Domain Adaptation (M³SPADA) framework, a novel deep learning framework able to temporally transfer, over a particular geographical area, an MSLCC from a period of time (*source domain*) featured by reference data to a different period of time (*target domain*), still described by multisensor information, where no reference data are available.

M³SPADA combines both self-training and adversarial learning for MStUDA under sparsely annotated GT data. More precisely, we consider, as multisensor remote sensing data, the simultaneous availability, for both source and target domains, of optical SITS data at HRs (i.e., Sentinel-2 data at 10 m of spatial resolution) and a VHR imagery (i.e., Spot-6/7 image at 1.5 m of spatial resolution). Methodologically speaking, our framework exploits adversarial learning with the goal to learn domain-invariant features and gradually transfer the underlying MSLCC from the source to the target domain via pseudo-labeling. The pseudo-labeling process selects pseudo-labels on the target domain as stable spatial areas between the two considered years using them as tie points to connect the two domains, thus decreasing the distribution gap between them.

Experimental evaluation is carried out on a rural study site located in Burkina Faso, referred as *Koumbia* site and characterized by a mostly agricultural land cover nomenclature (crop types as well as natural and built-up classes). We consider multisensor data coming from two different years (2018 and 2021) and performing a transfer assessment considering both 2018 toward 2021 and vice versa. We underline that each year (domain) is characterized by multisensor information: 1) a high-resolution optical (Sentinel-2) SITS and 2) a VHR optical (Spot-6/7) single-date image. To assess the behavior of M³SPADA, we conduct both quantitative and qualitative investigations considering state-of-the-art UDA frameworks that were adapted to the case of multisensor land cover classification.

The rest of this manuscript is organized as follows. The related literature on multisensor land cover classification and domain adaptation is described in Section II. Section III describes the MStUDA problem and introduces M³SPADA. Study site and the associated multisensor remote sensing data are introduced in Section IV. The experimental evaluation is reported in Section V. Section VI concludes this article.

II. RELATED WORK

A. Multisensor Land Cover Classification

Multisensor land cover classification is getting more and more attention due to the unprecedented availability of heterogeneous sensors data covering the same study area [1]. Several recent surveys has discussed such growing topic [3], [4]. Hong et al. [3] provide a categorization and a baseline framework to cope with the multisensor analysis of remote sensing data via convolutional neural networks (CNNs). The research study focused on “what,” “where,” and “how” to fuse several pairs of remote sensing modalities mainly addressing mono-temporal imagery for land cover mapping. More recently, [4] analyzes the current trends in the field of multisensor remote sensing data fusion and discussed the different applications, among the other land cover mapping, that can benefit from the fusion of multisensor Earth observation data. Considering multisensor remote sensing land cover mapping, when both modalities are SITS, [20], [21], [22], and [23] combine together synthetic aperture radar (SAR) and optical SITS with the aim to leverage the complementarity between active and passive sensors. While the pioneer work presented in [20] adopts an early fusion mechanism (optical and radar data are first combined together, and subsequently, a deep learning architecture is deployed on the stack of multisensor data), the more recent strategies adopt a late fusion mechanism (an encoder per modality is used, and the combination is internally implemented by the neural network through an end-to-end process). In addition, the research works conducted in [21], [22], and [23] differ between each other by the type of the encoder they adopt to process the multitemporal information (recurrent, convolutional, or transformer neural network). Focusing on the combination of multitemporal and multiscale information via deep learning strategies, [5], [6], and [7] propose to combine optical SITS with a mono-temporal VHR optical imagery with the objective to jointly exploit multitemporal as well as multiscale information. While [5] and [6] exploit recurrent neural networks to model the time-series information recently, [7] adopts 1-D CNN to analyze multitemporal imagery, demonstrating that 1-D CNN seems more appropriate to manage multitemporal information in the fusion process. In all the proposed approaches, the VHR imagery is managed by means of a 2-D CNN encoder.

B. Multisensor UDA

UDA [11] methods belong to the family of transfer learning approaches [24], which has the main objective to transfer a model trained on a labeled source domain to an unlabeled target domain. Recent advances in UDA focus their efforts to align domains through data transformation and/or extract domain-invariant features to reduce the distribution gap between the source and target domains [11]. Ganin et al. [25] introduce the domain-adversarial neural network (DANN) model where a standard neural network model is augmented with a domain classifier that may distinguish between source and target samples in a multitask learning setting. The domain classifier is associated with a gradient reversal layer (GRL) that enforces the features extracted by

the encoder to be invariant with respect to the domains. The work described by Tzeng et al. [26] still considers an adversarial training setting: here, they define the adversarial discriminative domain adaptation (ADDA) method. Inspired by the concept of generative adversarial network (GAN), this approach set up a two-player game where a discriminator network tries to distinguish between source and target sample representations derived by the generator, while the generator tries to fool the discriminator network. Currently, adversarial learning is one of the main trends when it comes to UDA [27], [28], [29]. When multiple modalities come into the picture, still a limited amount of research work has been conducted in the general field of computer vision and machine learning that exploits the opportunities related to modern deep learning frameworks. Qi et al. [30] introduce a multisensor domain adaptation neural network based on the attention mechanism with the aim to learn domain-invariant features among modalities and domains. The proposed methodology is evaluated on cross-domain applications related to emotion recognition and cross-media retrieval. The involved modalities cover visual and acoustic information. With the aim to tackle fine-grained action recognition classification, [31] defines a deep learning-based framework that exploits both adversarial learning (GRL) and self-supervised alignment to cope with multisensor UDA. In this context, the multisensor information is characterized by RGB images and optical flows depicting human actions. The case of multisensor UDA for RGB and depth data has been tackled in the study proposed in [32]. Here, the authors have exploited collaboration among multiple modalities, pseudo-labeling, and generative models to capture modality-specific and modality-integrated information in order to transfer the neural network classification model from a multisensor source domain to a mono-sensor target domain. At the intersection of remote sensing analysis and UDA, early research focused on proposing UDA strategies for VHR imagery [33] with a major focus on semantic segmentation tasks [12] that require costly and time consuming dense annotations. While moving to the context of SITS data, only, in recent times, some strategies emerged, which address both spatial [10], [14] and temporal [9] transfer learning. Nonetheless, none of them cover the possibility to manage simultaneously multitemporal and multiscale remote sensing information in a multisensor data fusion setting. The extensive literature review we have performed clearly underlines that recent UDA approaches, especially the ones based on deep learning strategies, are still unexplored in the context of multisensor UDA for remote sensing analysis considering both spatial as well as temporal transfer tasks.

III. M³SPADA

In this section, we introduce our framework, M³SPADA, designed to cope with MStUDA for land cover mapping. First, we define the problem setting; then, we provide a general overview of our framework. Subsequently, we describe the different components on which M³SPADA is built on. Finally, we describe the multisensor classifier used in our framework.

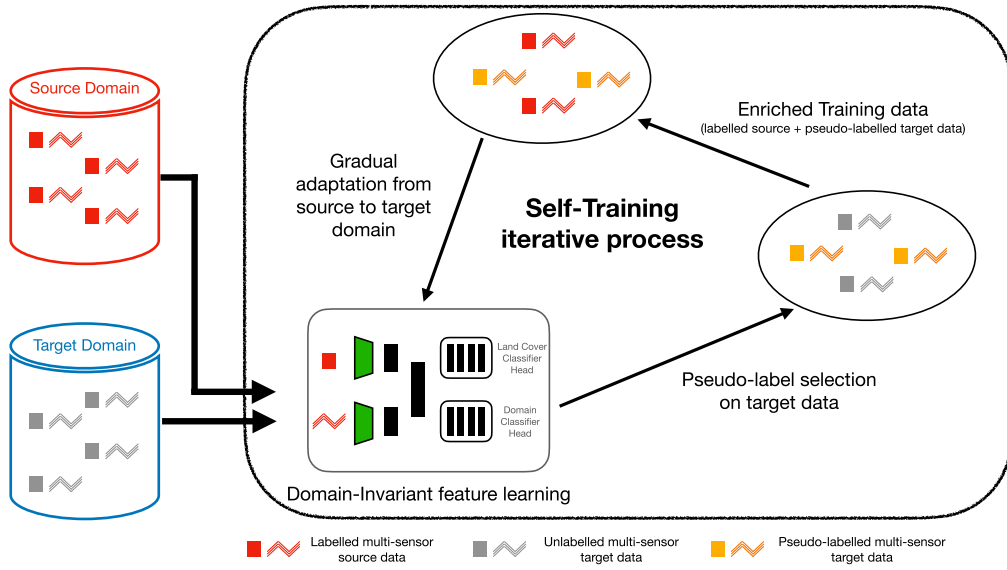


Fig. 2. Overview of our framework. M³SPADA takes as input multisensor (SITS and VHR) source data (\mathcal{D}^s) with its associated reference GT and multisensor (SITS and VHR) target data (\mathcal{D}^t) with the objective to learn a multisensor classification model to predict land cover labels for target data. During the self-training iterative process, M³SPADA alternates domain-invariant feature learning via adversarial training and pseudo-labeling with the objective to gradually adapt the underlying classification model from the source to the target multisensor data to cope with the MStUDA problem.

A. Problem Setting

In this work, we consider the problem of MStUDA. A multisensor source domain $\mathcal{D}^s = \{(m_i^s, y_i^s)\}_{i=1}^{n^s}$ and a multisensor target domain $\mathcal{D}^t = \{m_i^t\}_{i=1}^{n^t}$ are given, with n^s and n^t the number of samples for the source and target domains, respectively. We indicate with M^s , Y^s , and M^t the set of multisensor source samples, source labels, and multisensor target samples, respectively, $\mathcal{D}^s = \{M^s, Y^s\}$ and $\mathcal{D}^t = \{M^t\}$. In our case, each multisensor sample m_i is a pair of datasets describing the same geospatial location; more precisely, $m_i = (\text{sits}_i, \text{vhr}_i)$, where sits_i is a SITS pixel and vhr_i is a patch of a VHR imagery centered around the SITS pixel. The land cover information (y_i^s) is only available for the multisensor source domain. The number of land cover classes equals K , $y_i^s \in \{1, \dots, K\}$.

The set of multisensor samples belonging to the two domains cover exactly the same spatial extent but at different periods of time (i.e., different years). This means that the same geospatial location is covered by a multisensor sample coming from the source as well as one from the target domain. Thus, n^s is equal to n^t , and $\text{location}(m_i^s)$ is equal to $\text{location}(m_i^t)$, where $\text{location}(\cdot)$ supplies the geospatial location (geographical coordinates) of a multisensor sample.

Here, the goal is to train a multisensor land cover mapping model on labeled multisensor source data as well as unlabeled multisensor target data to overcome possible distribution shifts between the two domains (i.e., differences in environmental, weather, or climate acquisition conditions) in order to effectively classify samples belonging to the multisource target domain considering the same set of land cover classes, under a closed-set scenario [34].

B. General Overview of M³SPADA

Fig. 2 visually depicts how M³SPADA behaves. M³SPADA leverages both adversarial learning and self-training with the

goal to do the following: 1) learn a domain-invariant feature space with the aim to overcome possible data distribution gaps between the source and target domains and 2) gradually transfer the MSLCC from the source to the target domain via spatially aware pseudo-labeling.

Regarding the adversarial learning strategy, we adopt the GRL model proposed in [25]. Concerning the pseudo-labeling selection procedure, it is based on the fact that the multisensor remote sensing data from the source and the target domains are spatially aligned (i.e., they cover exactly the same geographical extent).

More precisely, given two multisensor samples m_i^s (coming from the source domain) and m_i^t (coming from the target domain), we indicate with $\text{location}(m_i^s) = \text{location}(m_i^t)$ the fact that they are spatially aligned. If the multisensor classifier provides the same decision for both samples [$\text{Cl}(m_i^s|W_F, W_L) = \text{Cl}(m_i^t|W_F, W_L)$] and the predicted class for the source sample (m_i^s) is the right one, then the target sample (m_i^t) is associated with the pseudo-label generated by the MSLCC. As the iterative training procedure goes on, more importance is attributed to pseudo-label information transferring the focus of the multisensor classifier from the source to the target data.

C. Learning Domain-Invariant Feature Space

With the aim to extract a domain-invariant feature space for multisensor remote sensing data, we leverage the strategy proposed in [25], namely, DANN, as backbone block of our framework.

The network structure has three parts, a multisensor encoder network \mathbf{F} , relying on the W_F parameters; a domain classifier head \mathbf{D} , with parameters W_D ; and a classifier head \mathbf{L} , for the land cover classification task, with parameters W_L . The multisensor encoder network is described later, in Section III-F.

The backbone network of our framework is a multitask network that has the objective to generate a new data representation via the multisensor encoder \mathbf{F} ensuring high land cover classification accuracy and, simultaneously, making difficult to distinguish between the domain each multisensor sample comes from. While the loss associated with the land cover classification task is defined as follows:

$$L_c(M^s, Y^s | W_F, W_L) \quad (1)$$

where $L_c(M^s, Y^s | W_F, W_L)$ is modeled with standard categorical cross-entropy loss, the adversarial loss function is defined as follows:

$$L_{\text{Adv}}(M^s, M^t | W_F, W_D) \quad (2)$$

where $L_{\text{Adv}}(M^s, M^t | W_F, W_D)$ is the loss related to the domain classifier head modeling a binary classification problem in which class label represents the possibility to belong exclusively to the source or the target domain. Also, in this case, the categorical cross-entropy function is employed. The two losses are combined together for the domain adaptation process as follows:

$$L_{\text{DA}}(M^s, Y^s, M^t | W_F, W_L, W_D) = L_c(M^s, Y^s | W_F, W_L) - \gamma L_{\text{Adv}}(X^s, X^t | W_F, W_D) \quad (3)$$

where the hyperparameter γ balances the contribution of the loss related to the domain classifier head in the domain-invariant feature learning process.

While the $L_c(M^s, Y^s | W_F, W_L)$ loss is optimized as commonly done for general neural network models, in order to leverage standard stochastic gradient descent to optimize the L_{DA} loss function, the $L_{\text{Adv}}(X^s, X^t | W_F, W_D)$ loss employs the GRL trick [25]. More in detail, the GRL acts as the identity transform during the forward propagation pass, while it reverses the gradient (the gradient is multiplied by -1) during the backward propagation (backpropagation) for the update of the multisensor encoder parameters \mathbf{F} . In this way, the GRL trick permits to implement the adversarial training strategy with standard backpropagation avoiding any extra parameter. More in detail, while the domain classifier head parameters are updated as usual, the reversed gradient is only applied to the multisensor encoder network with the aim to generate domain-invariant features and fool the domain classifier [14].

D. Spatially Aware Self-Training

In order to further adapt the multisensor land cover model to effectively classify samples belonging to the unlabeled target domain, we exploit a pseudo-labeling strategy that introduces pseudo-labels for a subset of samples coming from \mathcal{D}^t . Commonly, in the context of self-training learning [35], given a set of unlabeled samples, the model output distribution is used to select high-confidence samples from which pseudo-labels are derived to enrich the current training set. The pseudo-labeling mechanism is usually implemented via thresholding on the model output *softmax* in such a way that all the samples on which the most probable prediction value is greater than the

defined threshold are retained as pseudo-labeled samples. The main issue related to this procedure is related to the way in which this hyperparameter is fixed [36], since it can negatively affect the performance of the underlying sampling process.

In our case, we leverage the specificity of the land cover mapping MStUDA problem defining a procedure based on the spatial consistency between the two samples m_i^s and m_i^t sharing the same geospatial location [$\text{location}(m_i^t) = \text{location}(m_i^s)$]. Such a strategy overcomes the standard issue related to pseudo-labeling procedure allowing our framework to avoid the definition of any threshold selection step. More in detail, the set of target samples to which pseudo-labels will be associated are chosen based on two criteria that need to be simultaneously met. Let $\text{Cl}(\cdot | W_F, W_L)$ be the prediction of the MSLCC; the first criterion is based on spatial consistency as described below: $\text{Cl}(m_i^s | W_F, W_L) = \text{Cl}(m_i^t | W_F, W_L)$, such that $\text{location}(m_i^t) = \text{location}(m_i^s)$, and the second criterion requires that the MSLCC provides the correct prediction for the source sample: $\text{Cl}(m_i^s | W_F, W_L) = y_i^s$.

The rationale behind this process is related to selecting target samples that exhibit a stable behavior, in terms of model output prediction, with respect to the corresponding source sample in terms of geospatial location, and, at the same time, we require that the MSLCC predicts the correct class on the source sample m_i^s . In this way, the procedure allows to choose pseudo-labeled samples that act as tie points between the source and target domains leveraging, on the one hand, the model output stability and, on the other hand, target samples that are, in principle, characterized by a small distribution gap; thus, they are more useful to pave the classification model transfer from the source to the target domain.

More formally, we can define the loss associated with the pseudo-labeled samples as follows:

$$L_p(M^s, M^t, Y^s, \hat{Y}^t | W_F, W_L) = \sum_{m_i^t \in M^t} \mathbb{1}_{\{\text{Cl}(m_i^s) = \text{Cl}(m_i^t) \text{ and } \text{Cl}(m_i^s) = y_i^s\}} H(\hat{y}_i^t, \text{Cl}_{\text{prob}}(m_i^t)) \quad (4)$$

where $\mathbb{1}_{\text{cond}}$ is an indicator function that returns 1 if the condition *cond* is verified and 0 otherwise, $\text{Cl}_{\text{prob}}(\cdot)$ provides the classifier output distribution over the possible set of land cover classes, $H(\cdot, \cdot)$ is the categorical cross-entropy loss, \hat{Y}^t is the whole set of pseudo-labels for the target domain, and \hat{y}_i^t is the pseudo-label land cover class with the highest model output probability with respect to $\text{Cl}_{\text{prob}}(m_i^t)$ for the sample m_i^t coming from the target domain.

E. Training Procedure of M^3 SPADA

Algorithm 1 depicts the pseudo-code of the training procedure for our framework. The inputs of the procedure are constituted by the data coming from the source [$\mathcal{D}^s = (M^s, Y^s)$] and the target [$\mathcal{D}^t = (M^t)$] domains and E , the number of epochs for the neural network model training.

At the beginning, the classification model is deployed on the target data M^t in order to obtain the set of pseudo-labels \hat{Y}^t (line 3). Then, (line 4) the trade-off value λ is computed as linear function of e (the current number of epochs) and E (the total number of epochs).

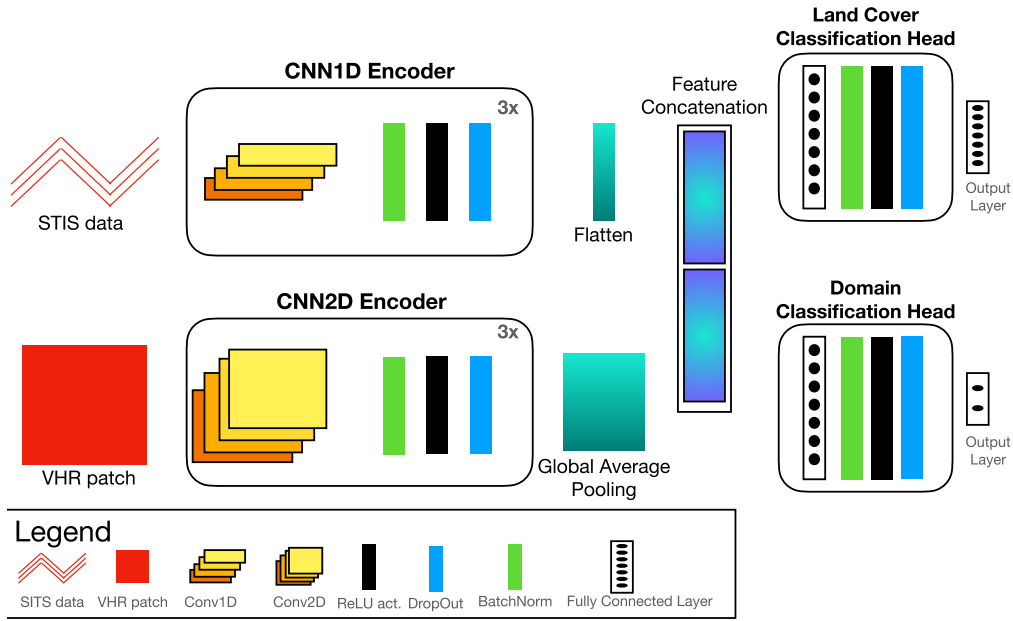


Fig. 3. MSLCC model we have employed as backbone of M^3 SPADA. The model as two input encoders, one for the SITS sensor and one for the VHR sensor; and two output heads, one for the land cover classification task and one for the domain classification task. This latter support the implementation of the adversarial learning strategy. At inference time, only the former head is retained, while the latter is ignored.

Algorithm 1 Training Procedure of M^3 SPADA

Require: M^s (the source samples), Y^s (the source labels), M^t (the target samples), E (the number of epochs).

Ensure: W_F (the encoder parameters), W_L (the land cover classifier parameters).

- 1: $e = 0$
- 2: **while** $e < E$ **do**
- 3: $\hat{Y}^t = Cl_{prob}(M^t)$
- 4: $\lambda = \frac{e}{E}$
- 5: $LOSS = (1-\lambda) \times L_{DA}(M^s, M^t, Y^s | W_F, W_L, W_D)$
 $\quad + \lambda \times L_p(M^s, M^t, Y^s, \hat{Y}^t | W_F, W_L)$
- 6: $\nabla_{W_F, W_L, W_D} LOSS$ with mini-batch SGD
- 7: $e = e + 1$
- 8: **end while**
- 9: **return** W_F, W_L

Subsequently (line 5), the λ value is employed to dynamically combine the L_{DA} and the L_p losses with the aim to vary their contribution during the learning procedure. More precisely, the λ value starts from zero and linearly increases with the objective to gradually provide more importance to the L_p term.

This is done, since at the early iterations, we force the model to exploit as much information as possible from the labeled source domain while learning samples' representations that are invariant with respect to the specific domain. The reason is that at the beginning of the training process, the model is not yet effective, so that the prediction on the target data could be highly biased. As the learning procedure advances on, the λ value increases, hence decreasing the importance of the first term (L_{DA}) while increasing the importance of the second one (L_p). This mechanism implements a gradual transfer from the

first to the second term during the learning procedure, allowing the underlying classification model to smoothly focus on the specificity of the target domain via the use of the pseudo-labels chosen as described in Section III-D. After that, the current loss $LOSS$ is computed (line 6); the network parameters W_F , W_L , and W_D are updated by mini-batch stochastic gradient descent. At the end of the training procedure (line 9), the network parameters W_F and W_L related to the multisensor encoder \mathbf{F} and the land cover classifier \mathbf{L} , respectively, are returned as a result of the training process of M^3 SPADA.

F. MSLCC Architecture

Fig. 3 sketches the architecture of the MSLCC we have used as backbone for the M^3 SPADA framework. We name such model as MSLCC. The model as two input branches, one for the SITS data and one for the VHR imagery; and two output heads, one for the land cover classification task (\mathbf{L}) and one devoted to discriminate between samples from the source and the target domains (\mathbf{D}) in order to implement the adversarial learning strategy. At inference time, only the former head is considered, while the latter is ignored.

Concerning the SITS encoder, it is inspired by the architecture of the TempCNN [37] model, a well-established 1-D CNN model especially tailored to cope with SITS land cover mapping. The SITS encoder is composed of three identical blocks followed by a flatten operation to extract the feature representation for this sensor. Each of the blocks is composed by a 1-D convolutional operator with 64 filters and a kernel equals 5, a batch normalization layer, a nonlinear activation function [rectifier linear unit (ReLU)], and a dropout layer.

Regarding the VHR encoder, this module is inspired by the one used in the architecture previously introduced in [5]. Such module is built on three identical blocks followed by a global average pooling operation. Each block is composed by

a 2-D convolution operator with 128 filters and a kernel with size equals 5×5 , a batch normalization layer, a nonlinear activation function (ReLU), and a dropout layer.

Once the per-sensor features are extracted, they are fused together by concatenation. Then, the fused multisensor representation is used to feed both the land cover classification (**L**) and the domain classification (**D**) heads. These two classification heads have the same structure, a fully connected layer of 256 neurons followed by batch normalization, ReLU activation function, dropout, and a final output layer with as many neurons as the numbers of land cover classes for the land cover classification head and two output neurons for the domain classification heads (source versus target).

According to recent advances in multisensor land cover classification [5], [23], we equip our backbone with auxiliary classifiers in order to take the most out of the fusion of different sensor data. More precisely, to enforce the multisensor deep learning model into fully leveraging the inter-sensor relationships, an additional classifier head per branch is used at training stage with the aim to strength as much as possible the discriminative power of the extracted per-source representations. At inference time, the per-source classification heads are discarded. Since the contribution of the per-source classification heads must be combined with the contribution of the main classification head, at training stage, we adopt the weighting schema proposed in [5].

IV. STUDY AREA, MULTISENSOR REMOTE SENSING DATA, AND GT REFERENCE

The study site is located around the commune of *Koumbia*, in the Province of Tuy, *Hauts-Bassins* region, in the southwest of Burkina Faso. This area has a surface of about 2338 km² and is situated in the subhumid Sudanian zone. The surface is covered mainly by forests and natural savannah (herbaceous and shrubby), interleaved with a large portion of land (around 35%) used for rainfed agricultural production, mostly small-holder farming. The main crops are cotton and cereals (maize, sorghum, and millet), followed by leguminous and oleaginous.

Fig. 4 presents the study site with the 2018 reference data (GT) superposed on a Sentinel-2 image of September 12, 2018. A more detailed view corresponding to the red box in the overview is also depicted on the bottom right of the figure. A specific analysis of the GT is provided in Section IV-B.

A. Multisensor Remote Sensing Data Description

Here, we describe the multisensor remote sensing data associated with both 2018 and 2021 years.

1) *Sensor 1: HR Optical SITS*: We have collected Sentinel-2 SITS data spanning the years 2018 and 2021, amounting for a total of, respectively, 35 and 39 available scenes. According to the available acquisitions, we conducted a visual analysis and selected 24 images for each year accounting for an uniform temporal sampling among the two years. The main selection criteria were as follows: 1) discard images that were visually impacted by strong cloud coverage and 2) keep a sufficient amount of acquisitions over the rainy (cropping) season, occurring between May and October. Fig. 5 depicts the acquisition dates of the two Sentinel-2 SITS.

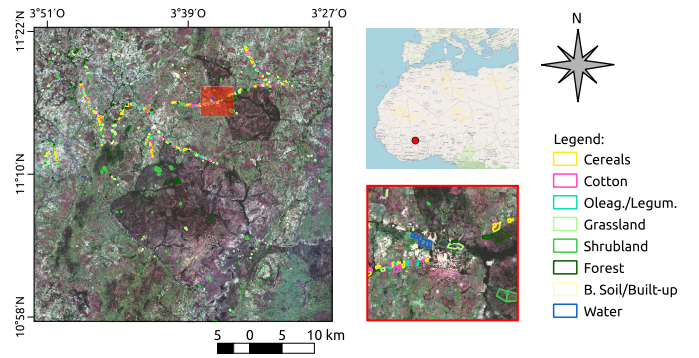


Fig. 4. Koumbia study area. The GT data coming from the year 2018 are superposed to a Sentinel-2 image covering the whole area. In the red box (bottom right), a detail of the study site is depicted, showing the complex structure of the underlying landscape.

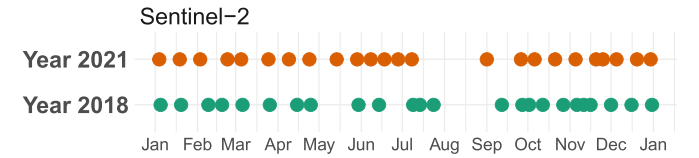


Fig. 5. Acquisition dates of sentinel-2 SITS.

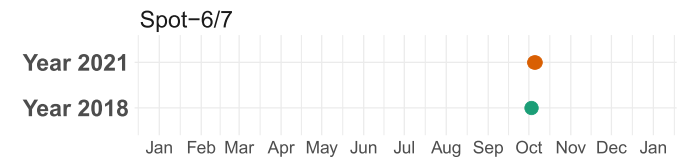


Fig. 6. Acquisition dates of Spot-6/7 satellite images.

All images were obtained by the value-adding products and algorithms for land surfaces (THEIA) Pole platform¹ at level 2A, which consist in atmospherically corrected surface reflectances via the Maccs-Atcor joint algorithm (MAJA) processing chain [38] with associated cloud/shadow masks. Only 10-m spatial resolution bands (blue, green, red, and near infrared spectrum) were considered in this analysis. Leveraging the method proposed in [39] and the available cloud masks, as preprocessing, cloudy pixels were imputed by means of linear interpolation considering precedent and posterior acquisitions.

2) *Sensor 2: Optical VHR Imagery*: The VHR imagery we consider in our research study is acquired via the Spot-6/7 satellite constellation, and they cover the whole study site. Fig. 6 illustrates the acquisition dates of the Spot-6/7 satellite images for 2018 and 2021. More precisely, for the year 2018, we exploit a Spot-7 image acquired on October 3, 2018, and for the year 2021, we exploit a composite image derived by the mosaicking of the Spot-6 and Spot-7 images acquired on October 5 and 6, 2021, respectively. All the Spot-6/7 images originally consisting of a 1.5-m panchromatic band and four multispectral bands (blue, green, red, and near infrared) at 6-m resolution were pansharpened to produce a single multispectral image at 1.5-m resolution. Images from both years are first orthorectified using ancillary data and a common

¹<http://theia.cnes.fr>

TABLE I
GT STATISTICS FOR YEARS 2018 AND 2021

Class Name	Class ID.	# Polygons	# Pixels	% Pixels
CEREALS	1	330	13056	16.4
COTTON	2	153	7672	9.6
OLEAGINOUS/LEGUMINOUS	3	161	3595	4.5
GRASSLAND	4	123	13108	16.4
SHRUBLAND	5	87	23121	28.9
FOREST	6	88	17369	21.7
BARE SOIL/BUILT-UP	7	46	835	1.0
WATER	8	10	1205	1.5
Total		998	79961	100

Class Name	Class ID.	# Polygons	# Pixels	% Pixels
CEREALS	1	268	11435	14.4
COTTON	2	121	6575	8.2
OLEAGINOUS/LEGUMINOUS	3	263	7316	9.1
GRASSLAND	4	113	11100	13.9
SHRUBLAND	5	90	24324	30.4
FOREST	6	82	16984	21.2
BARE SOIL/BUILT-UP	7	51	1022	1.3
WATER	8	10	1205	1.5
Total		998	79961	100

DEM and then manually co-registered using a rigid transform (translation) to enforce geographical colocation.

B. GT Data

GT data for 2018 and 2021 have been derived from a large agricultural land cover dataset available online [40], mainly consisting of field data collected by local experts on several sites all over the tropics. GPS waypoints were gathered, following an opportunistic sampling approach along the roads or tracks according to their accessibility, while ensuring the best representativity of the existing cropping practices in place. Records were also provided on different types of non-crop classes (e.g., natural vegetation, settlement areas, and water bodies) to allow differentiating crop and non-crop classes. Moreover, some additional non-crop reference polygons are also provided, obtained by photointerpretation.

Our final GT has been assembled in a geographic information system (GIS) vector file, containing a collection of polygons, each attributed with a land cover class based on information reported in the original database. Statistics about the yearly reference datasets used here are summarized in Table I. In order to ensure consistency with the proposed method, we kept the exact same surface for the two reference years by performing a year by year intersection of the polygons of the original database.

V. EXPERIMENTS

Here, we describe the experimental evaluation we have conducted on the study site presented in Section IV. With the aim to assess the behavior of M³SPADA, we investigate different dimensions: 1) we conduct a quantitative evaluation of the performances achieved by M³SPADA with respect to the UDA competitor we have adapted for the multisensor remote sensing scenario; 2) we perform a qualitative investigation of the LCMs generated by M³SPADA; and 3) we inspect the internal representation learned by our multisensor neural network model, and we visually compared them with some of the best performing competitors.

A. Competing Methods

With the goal to assess the performance of M³SPADA with respect to state-of-the-art UDA strategies, we consider the following.

- 1) The DANN method originally introduced in [25]. This is a well-known and largely employed UDA approach that exploits GRL with the aim to obtain data representations that are invariant to the particular domain they come from.
- 2) The conditional adversarial domain adaptation with entropy conditioning (CDAN + E) approach [27]. This method upgrades DANN by conditioning the domain discriminator on the classification output and minimizing an entropy loss on target data.
- 3) The margin disparity discrepancy (MDD) method introduced in [28]. This theory-inspired technique measures the distribution discrepancy between domains, and it internally leverages the DANN strategy.
- 4) The Wasserstein distance guided representation learning (WDGRL) method described in [41] that learn domain-invariant feature representations by using a domain critic neural network to estimate empirical Wasserstein distance between the two domains and optimizing the feature extractor network to minimize the estimated Wasserstein distance in an adversarial manner.
- 5) The adversarial-learned loss for domain adaptation (ALDA) method presented in [29]. ALDA combines domain-adversarial learning and self-training to minimize the domain gap and aligns the feature distributions by means of a noise-correcting domain discriminator.

We couple all the previous competing strategies with the same backbone neural network we have employed in our framework. On the one hand, this permits to directly extend all the competing frameworks to the scenario of multisensor land cover mapping, since no previous approach has been proposed for MStUDA in the context of remote sensing data analysis, and, on the other hand, it allows to fairly compare the framework performances between each other, since the same internal multisensor classifier is adopted.

Furthermore, we also consider two baseline strategies in which the following hold: 1) the multisensor backbone classifier is trained with only source data and directly deployed on target data, referred as “only \mathcal{D}_s ” and 2) the multisensor backbone classifier is trained on labeled target data and deployed on the rest of the target samples referred as “only \mathcal{D}_t .” The former constitutes a direct baseline that does not take into account the necessity to deal with temporal data distribution shifts, while the latter represents the performances we can (theoretically) achieve if we have labeled data associated with the target domain \mathcal{D}_t .

For the two baseline strategies, we consider, as supervised classification method, the multisensor deep learning classifier we have introduced in Section III (the same as the backbone of M³SPADA) as well as its per source ablations. We name such strategies MSLCC, MSCC_{SITS}, and MSLCC_{VHR} that indicate the multisensor deep learning classifier, the ablation considering only SITS data, and the ablation considering only VHR imagery, respectively.

Furthermore, we also consider the per-source ablations of M³SPADA in order to have a direct overview of the benefit related to the joint use of multitemporal and multiscale information. We indicate with M³SPADA_{SITS} the ablation that only leverages SITS and with M³SPADA_{VHR} the ablation that only leverages VHR imagery.

B. Experimental Settings

1) *Baseline Strategies Evaluation*: Regarding the first baseline (only \mathcal{D}_s), the supervised classification model is trained over all the source data and then deployed on the target data. Concerning the second baseline strategy (only \mathcal{D}_t), solely, the target data are involved. More precisely, target data are split into three sets: training, validation, and test sets with a proportion of 70%, 10%, and 20% of the whole target dataset, respectively. With the aim to avoid possible spatial bias in the evaluation procedure [42], we force that all the pixels belonging to the same object will be exclusively associated with one of the data partition (training, validation, or test). The splitting procedure is repeated ten times, and the average results are reported.

2) *UDA Competing Methods Evaluation*: All the UDA frameworks are trained exploiting the whole set of source and target samples with the sole access to label information coming from the source domain.

Regarding the evaluation tasks, according to the information presented in Section IV, we set up two temporal transfer tasks ($\mathcal{D}_s \rightarrow \mathcal{D}_t$) where the right arrow indicates the transfer direction from the source (\mathcal{D}_s) to the target (\mathcal{D}_t) domain: (2018 \rightarrow 2021) and (2021 \rightarrow 2018).

The SITS, both 2018 and 2021, are rescaled, per band, considering the 2nd and 98th percentile of the data distribution as minimum and maximum values. Similarly, we also rescale the values of the VHR imagery, per band, considering the same procedure adopted for SITS data.

For the VHR imagery, we consider image patches of size 15×15 pixels following the findings reported in the related literature [5]. Note that, as in this reference work, all images used are properly georeferenced by the provider and projected onto a common coordinate system; hence, a regular sampling through geographical coordinates ensures the correct superposition of data from different sensors, net of differences in the geometric precision of the respective orthorectification processes, which are, by the way, negligible in our case. More precisely, for each Sentinel-2 pixel sample, the coordinates of its centers are used to identify the corresponding pixel over the SPOT-6/7 image on which the patch sample is centered. A multisensor sample $m_i = (\text{sits}_i, \text{vhr}_i)$ can be, hence, built as a pair of Sentinel-2 pixel-level time series sits_i and SPOT-6/7 image patch vhr_i .

The assessment of the model performances was done considering the following metrics: *accuracy* (global precision) and *weighted F1 score*.

3) *Details of the Implementation*: For the neural network approaches, the training stage has been conducted for 400 epochs, with a learning rate of 10^{-4} and a batch size of 32. Batch normalization layers has been inserted after each fully connected or convolutional layer (except for the

TABLE II
WEIGHTED *F1* SCORE AND ACCURACY OF THE COMPETING APPROACHES FOR THE TRANSFER TASK (2018 \rightarrow 2021). THE BEST SCORE, FOR UDA METHODS, IS HIGHLIGHTED IN BOLD

Strategy	Method	F1-score	Accuracy
Only \mathcal{D}_s	<i>MSLCC</i> _{VHR}	50.9	52.8
	<i>MSCC</i> _{SITS}	54.3	56.6
	<i>MSLCC</i>	60.6	62.2
w/ UDA	WDGRL	62.6	63.7
	MDD	63.5	65.3
	ALDA	64.4	66.7
	DANN	69.8	70.7
	CDAN+E	70.9	71.9
	M ³ SPADA _{VHR}	71.1	73.2
	M ³ SPADA _{SITS}	79.4	82.4
M ³ SPADA	83.0	84.5	
Only \mathcal{D}_t	<i>MSLCC</i> _{VHR}	62.5 \pm 1.5	62.8 \pm 1.5
	<i>MSCC</i> _{SITS}	74.2 \pm 2.5	73.9 \pm 2.5
	<i>MSLCC</i>	75.5 \pm 1.3	75.2 \pm 1.3

classification layer). The dropout rate is set to 0.5. For M³SPADA, we set the value of the hyperparameter γ as suggested by Ganin et al. [25]. In addition, domain-specific batch normalization [43] by processing the source and target mini-batches separately is implemented. This is similar to what commonly done in the related literature [10].

Concerning ALDA, we set the threshold of pseudo-labels to 0.9 according to the recent literature on self-training in the context of land cover mapping [10].

Experiments are carried out on a workstation with a dual Intel² Xeon² CPU E5-2667v4 (@3.20 GHz) with 256 GB of RAM and four TITAN X (Pascal) GPU. All the deep learning methods are implemented using the Python *TensorFlow* library except ALDA that was implemented in PyTorch based on the original open source implementation.³ The DANN, CDAN + E, MDD, and WDGRL competitors are implemented via the Python awesome domain adaptation python toolbox (*ADAPT*) library [44]. All the models run on a single GPU. The code implementation of M³SPADA is available at this link.⁴

C. Quantitative Results

The results, in terms of *F1* score and accuracy, are reported in Tables II and III for the (2018 \rightarrow 2021) and (2021 \rightarrow 2018) transfer tasks, respectively.

We note that the lowest performances are achieved for the models that are directly transferred from the source to the target multisensor domain for both transfer tasks (only \mathcal{D}_s scenario). Nevertheless, we observe that, in this direct transfer scenarios, the model exploiting the multisensor information outperforms, in terms of classification performances, the mono-sensor counterparts.

Regarding the UDA approaches, we see that M³SPADA clearly outperforms all the competitors by more than ten points in terms of both *F1* score and accuracy. This fact underlines the quality and the value of the proposed framework

²Registered trademark.

³<https://github.com/ZJULearning/ALDA>

⁴<https://github.com/ecapliez/M3SPADA>

TABLE III

WEIGHTED $F1$ SCORE AND ACCURACY OF THE COMPETING APPROACHES FOR THE TRANSFER TASK (2021 \rightarrow 2018). THE BEST SCORE, FOR UDA METHODS, IS HIGHLIGHTED IN BOLD

Strategy	Method	F1-score	Accuracy	
Only \mathcal{D}_s	$MSLCC_{VHR}$	52.8	55.9	
	$MSCC_{SITS}$	60.3	61.2	
	$MSLCC$	62.4	64.4	
w/ UDA	ALDA	61.4	63.7	
	WDGRL	69.1	70.0	
	MDD	69.8	71.5	
	DANN	72.7	73.1	
	CDAN+E	74.1	74.2	
	M^3SPADA_{VHR}	77.6	78.8	
	M^3SPADA_{SITS}	88.4	88.2	
	M^3SPADA	89.9	89.4	
	Only \mathcal{D}_t	$MSLCC_{VHR}$	72.3 \pm 1.2	72.8 \pm 1.2
		$MSCC_{SITS}$	78.3 \pm 1.3	78.2 \pm 1.3
$MSLCC$		80.9 \pm 1.2	80.9 \pm 1.2	

that effectively combines together both adversarial learning and pseudo-labeling in order to take the most out of the two approaches. We can further highlight that the multisensor information appears to be beneficial to the downstream land cover mapping task due to the fact that M^3SPADA exhibits higher classification performances than its mono-sensor ablations M^3SPADA_{VHR} and M^3SPADA_{SITS} , with the latter achieving, generally, the second best performances.

Finally, we observe that M^3SPADA achieves better performance results than the classifiers trained and tested on the target domain (only \mathcal{D}_s scenarios). This can be explained by the fact that our framework can exploit more training data (both labeled source and pseudo-labeled target samples) to learn its internal parameters as well as focus on domain-invariant characteristics. This last point allows M^3SPADA to overcome possible per domain biases and domain distribution shifts due to radiometric and gap-filling preprocessing and differences in data acquisition conditions related to environmental, climatic, and weather phenomena.

1) *Per-Class Analysis*: Here, we report and describe the per-class analysis regarding the competing methods on the two transfer tasks we have considered. We first report per-class $F1$ score, and subsequently, we discuss the different confusion matrices to understand possible interclass mistakes. For this analysis, we focus our attention on the supervised multisensor methods $MSLCC$ under the two scenarios (only \mathcal{D}_s and only \mathcal{D}_t) as well as M^3SPADA , its mono-sensor ablations, and the best competing UDA approach CDAN + E.

The per-class $F1$ score is depicted in Figs. 7 and 8 for the (2018 \rightarrow 2021) and (2021 \rightarrow 2018) transfer task, respectively.

In general, we observe that M^3SPADA achieves the best (or comparable to the best) per-class score for all the nonagricultural classes (grassland, shrubland, forest, bare soil/built-up, and water) no matter the transfer task.

Regarding agricultural classes (cereals, cotton, and oleaginous/leguminous), we note different behaviors depending from the transfer task.

Concerning the *cereals* land cover class (the agricultural class that is the most represented in terms of

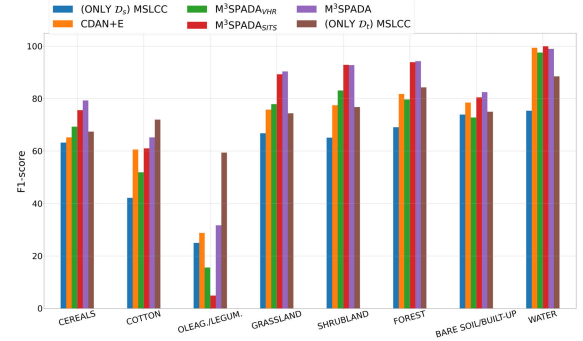


Fig. 7. Per land cover class $F1$ score of the different competing approaches considering the direct transfer strategy for the $MSLCC$ (only \mathcal{D}_s); the multisensor classification model trained on the target domain (only \mathcal{D}_t); and M^3SPADA , M^3SPADA_{SITS} , M^3SPADA_{VHR} , and the best UDA competitor on the transfer task (2018 \rightarrow 2021).

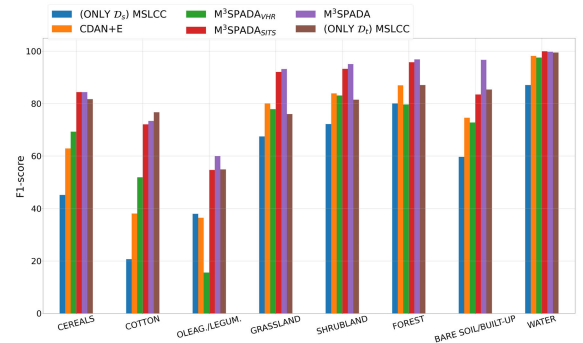


Fig. 8. Per land cover class $F1$ score of the different competing approaches considering the direct transfer strategy for the $MSLCC$ (only \mathcal{D}_s); the multisensor classification model trained on the target domain (only \mathcal{D}_t); and M^3SPADA , M^3SPADA_{SITS} , M^3SPADA_{VHR} , and the best UDA competitor on the transfer task (2021 \rightarrow 2018).

samples), M^3SPADA outperforms (or behaves comparable) to all the other methods. Conversely, for the cotton and oleaginous/leguminous land cover classes, M^3SPADA exhibits better performance results for the (2021 \rightarrow 2018) transfer task than for the (2018 \rightarrow 2021) one. This is probably due to the highly unbalanceness in class distribution characterizing 2018 with respect to 2021 related to these two agricultural classes. When 2018 is used as source domain (2018 \rightarrow 2021), the under representation of the oleaginous/leguminous land cover class negatively impacts the ability of M^3SPADA to transfer the multisensor land cover classification from the source to the target domain. Nevertheless, M^3SPADA still remains successful compared with the best competing UDA method, CDAN + E.

Focusing on the multisensor versus mono-sensor approaches (M^3SPADA versus M^3SPADA_{SITS} and M^3SPADA_{VHR}), we see that the joint exploitation of multitemporal and multiscale information systematically permits to improve the classification performances no matter the land cover class with respect to the exclusive use of either multitemporal or multiscale information. Despite the fact that the multitemporal information is clearly the most important remote sensing source for the considered land cover mapping task, introducing VHR imagery systematically ameliorates the results over all the land cover classes with some evident gain related to the

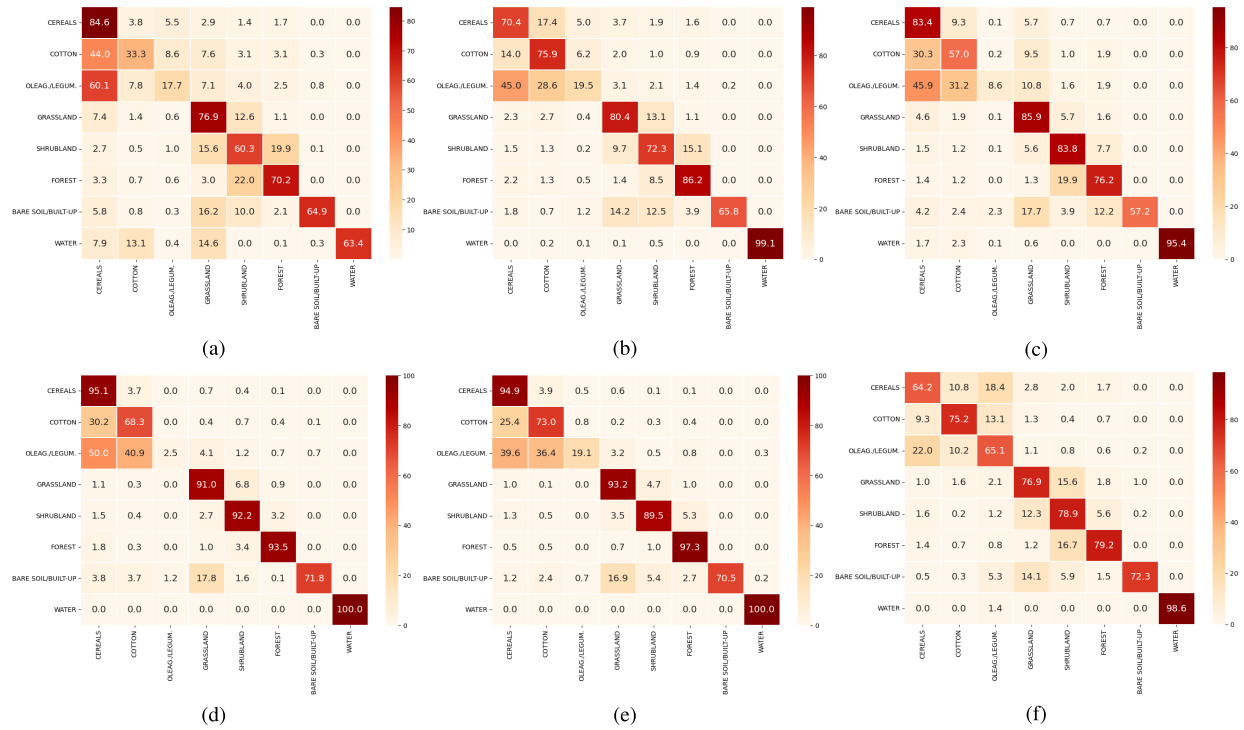


Fig. 9. Confusion matrices of the land cover classification for the transfer task (2018 \rightarrow 2021). True class (rows). Predicted class (columns). (a) Only \mathcal{D}_s MSLCC. (b) CDAN + E. (c) M^3SPADA_{VHR} . (d) M^3SPADA_{SITS} . (e) M^3SPADA . (f) Only \mathcal{D}_t MSLCC.

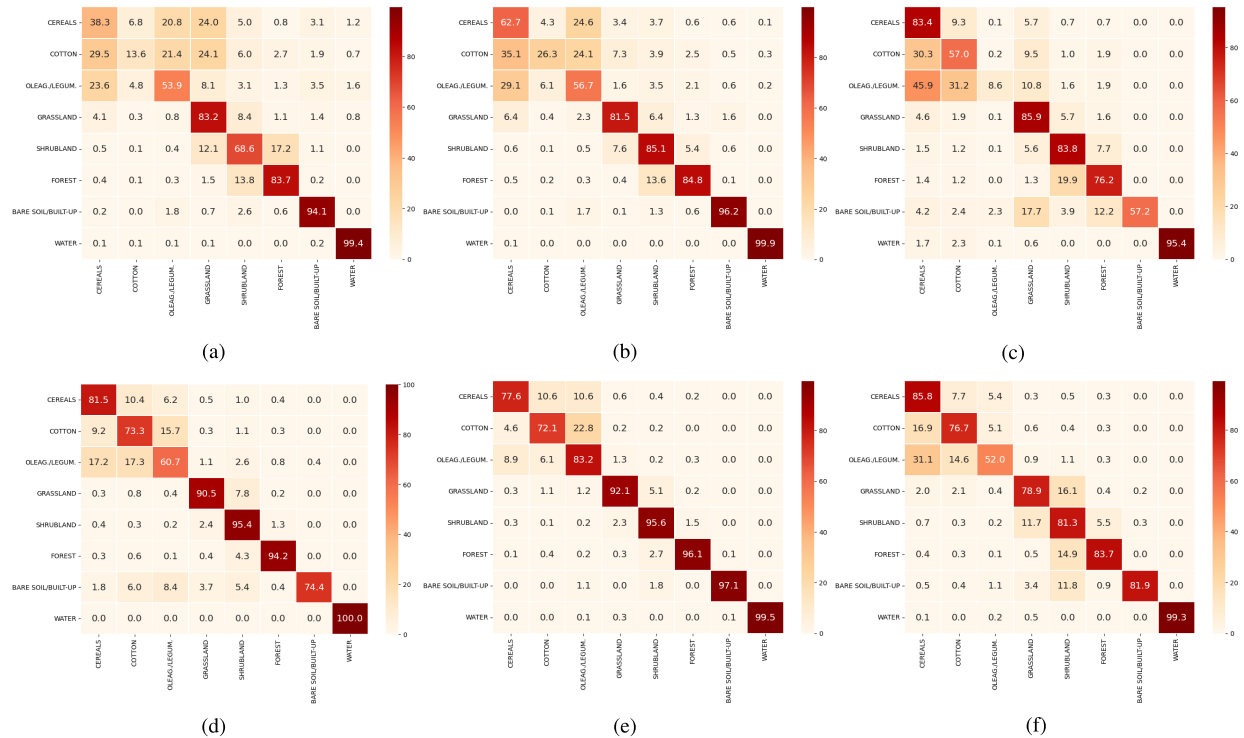


Fig. 10. Confusion matrices of the land cover classification for the transfer task (2021 \rightarrow 2018). True class (rows). Predicted class (columns). (a) Only \mathcal{D}_s MSLCC. (b) CDAN + E. (c) M^3SPADA_{VHR} . (d) M^3SPADA_{SITS} . (e) M^3SPADA . (f) Only \mathcal{D}_t MSLCC.

oleaginous/leguminous land cover class for the transfer task (2018 \rightarrow 2021) and for the baresoil/built-up land cover class for the transfer task (2021 \rightarrow 2018).

Figs. 9 and 10 depict confusion matrices for the (2018 \rightarrow 2021) and (2021 \rightarrow 2018) transfer tasks, respectively.

Globally, the confusion matrices confirm the trend observed in the per-class $F1$ -score analysis. All the methods have some issues in discriminating among the different agricultural classes. As discussed before in Section V-C1, the UDA approaches suffer from class unbalanceness related to the

TABLE IV
F1 SCORE AND ACCURACY OF THE M³SPADA ABLATIONS FOR THE
TRANSFER TASK (2018 → 2021)

Method	F1-score	Accuracy
DANN	69.8	70.7
M ³ SPADA _{Th}	63.2	67.5
M ³ SPADA _{OnlyC1}	81.3	83.7
M ³ SPADA	83.0	84.5

transfer task (2018 → 2021). We also note that all the methods exhibit some confusions between grassland/shrubland and shrubland/forest land cover classes. This is not surprising, since these three classes refer to three different degrees of density of woody vegetation in natural areas, which vary in a continuous way over the site, making a neat discrimination more challenging. However, M³SPADA (and its ablations) seems more robust to this issue compared with all the other evaluated approaches.

Nevertheless, M³SPADA provides a more visible diagonal structure (the dark red blocks concentrated on the diagonal) than the other methods alleviating some of the major confusions exhibited by the competitors.

2) *Ablation Analysis*: Here, we inspect the added value of the different components on which M³SPADA is built on. Table IV summarizes the behavior of the different ablations of M³SPADA for the (2018 → 2021) task. More precisely, we consider three specific ablations of M³SPADA.

- 1) The DANN approach previously introduced as standard competitor. This approach can be seen as a direct ablation of M³SPADA, since it is the first method that introduces the GRL strategy.
- 2) M³SPADA_{Th}, an ablation of our framework where the selection of pseudo-labeled samples is achieved by the traditional thresholding approach [45]. More precisely, during the iterative process, samples from the target domain are associated with pseudo-label if the most confident class predicted by the MSLCC has an associated confidence greater than a specific threshold θ . We set θ equal to 0.9 similar to what done for the ALDA approach.
- 3) M³SPADA_{OnlyC1}, an ablation of the proposed method that removes the pseudo-label condition requiring that the predicted class on source domain is equal to the true class ($\text{Cl}(m_i^s) == y_i^s$) for the pseudo-labeling selection stage. Here, the L_p loss is redefined as follows:

$$L_p(M^s, M^t, Y^s, \hat{Y}^t | W_F, W_L) = \sum_{m_i^t \in M^t} \mathbb{1}_{\{\text{Cl}(m_i^s) = \text{Cl}(m_i^t)\}} H(\hat{y}_i^t, \text{Cl}_{\text{prob}}(m_i^t)). \quad (5)$$

We first notice that M³SPADA provides by far better behaviors than DANN, its baseline ablation. Second, we observe that choosing pseudo-labels based on a traditional thresholding mechanism (M³SPADA_{Th}) results in a dramatic decrease in terms of classification performances. This is due to the fact that too many samples are pseudo-labeled, and, among the selected samples, spurious ones negatively affect the

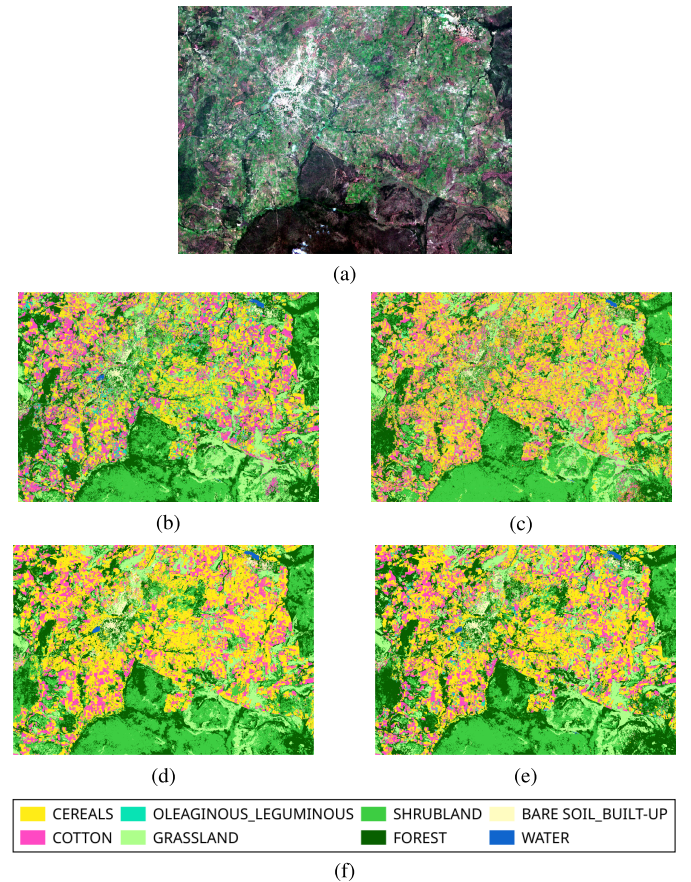


Fig. 11. Qualitative investigation of LCMs produced by CDAN + E, M³SPADA_{VHR}, M³SPADA_{SITS}, and M³SPADA for the transfer task (2018 → 2021): zoomed-in view on the Koumbia city. (a) Sentinel-2 image acquired on September 26, 2021. (b) CDAN + E. (c) M³SPADA_{VHR}. (d) M³SPADA_{SITS}. (e) M³SPADA. (f) Land cover classes legend.

transfer process. The ablation related to the spatial consistency (M³SPADA_{OnlyC1}) still exhibits lower performances with respect to the complete framework. As conclusion, the ablation analysis underlines that the pseudo-labeling mechanism we have introduced allows to focus on reliable pseudo-labeled samples selected from the target domain, thus positively enhancing the current training set. M³SPADA always outperforms all its ablations, highlighting that the interplay among the different components on which it is built on provides an effective strategy for the MStUDA land cover mapping problem.

D. Visual Analysis

In this section of the experimental assessment, we provide qualitative analyses to further evaluate the behavior of M³SPADA in the case of the transfer task (2018 → 2021). More precisely, we first examine some extracts related to the LCMs provided by M³SPADA and some of the competing approaches, and then, we visually examine the internal representations learned by our framework and the most competitive approaches.

1) *Land Cover Maps*: In Fig. 11(b)–(e), maps corresponding to the (2018 → 2021) transfer task are compared with respect to the scene subset shown in Fig. 11(a). The maps

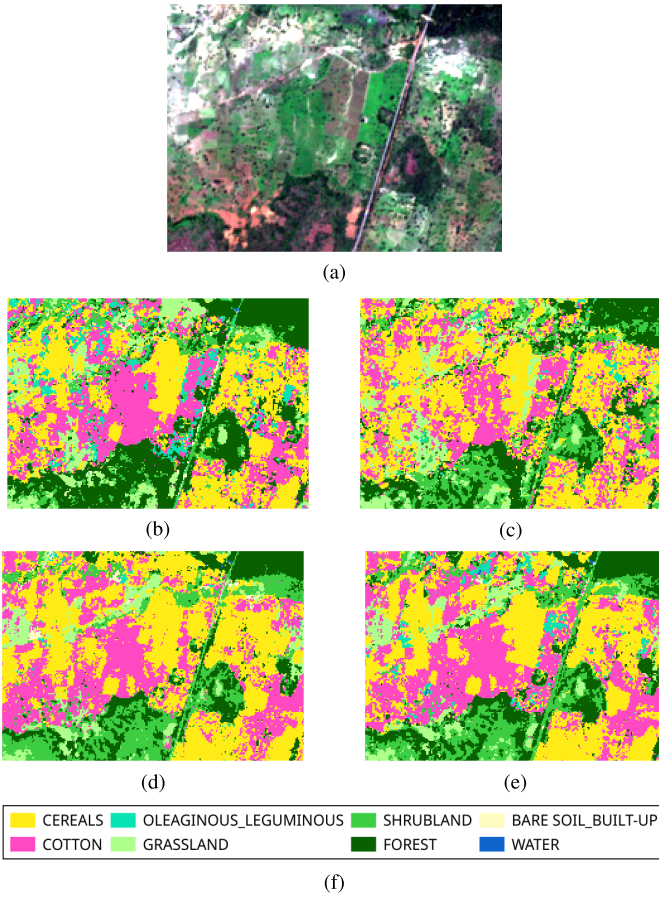


Fig. 12. Qualitative investigation of LCMs produced by CDAN + E, M^3SPADA_{VHR} , M^3SPADA_{SITS} , and M^3SPADA for the transfer task (2018 \rightarrow 2021): zoomed-in view on area $n^{\circ}1$. (a) Sentinel-2 image acquired on September 26, 2021. (b) CDAN + E. (c) M^3SPADA_{VHR} . (d) M^3SPADA_{SITS} . (e) M^3SPADA . (f) Land cover classes legend.

shown here are those obtained using the following methods: CDAN + E, M^3SPADA_{VHR} , M^3SPADA_{SITS} , and M^3SPADA .

In line with what was reported in the quantitative analysis, the visual investigation confirms that the transfer of knowledge from 2018 to 2021 is a challenging task, probably due to longer term changes in seasonal vegetation dynamics that appear after a three-year delay, as well as a redistribution of proportions between the different crop classes. The main observation concerns the strong underestimation of the oleaginous/leguminous class, regarding all the maps, using as transfer approach the mono-sensor strategies: M^3SPADA_{VHR} and M^3SPADA_{SITS} [see Fig. 11(c) and (d)], to the benefit of the cereals and cotton classes. In this case, the M^3SPADA_{VHR} method is particularly destructive. However, it seems quite evident that both CDAN + E and M^3SPADA [Fig. 11(b)–(e)] appear to effectively restore the extent of the oleaginous/leguminous class. Furthermore, the M^3SPADA map is less noisy than those processed by the other approaches, further demonstrating that it has greater potential to recover spatial structure than its competitor.

With the aim to evaluate the spatial precision of the M^3SPADA maps compared with its competitors, we also report some zoomed-in areas in Figs. 12 and 13. In both cases, the spatial details provided by the M^3SPADA are of higher

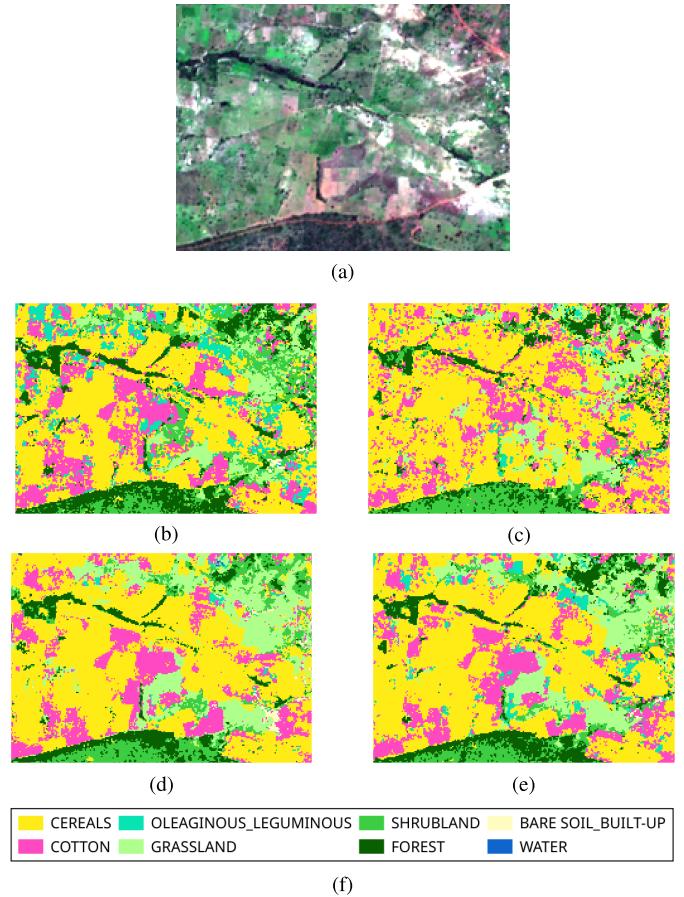


Fig. 13. Qualitative investigation of LCMs produced by CDAN + E, M^3SPADA_{VHR} , M^3SPADA_{SITS} , and M^3SPADA for the transfer task (2018 \rightarrow 2021): zoomed-in view on area $n^{\circ}2$. (a) Sentinel-2 image acquired on September 26, 2021. (b) CDAN + E. (c) M^3SPADA_{VHR} . (d) M^3SPADA_{SITS} . (e) M^3SPADA . (f) Land cover classes legend.

quality than the ones supplied by the competitors, both over agricultural fields, with more structured and less noisy plots (in terms of salt and pepper errors), especially over the cotton and cereals classes, and over natural areas.

The underestimation phenomena related to the oleaginous/leguminous class can be related to the imbalance of the crop classes that characterizes the 2018 reference data. More precisely, as shown in Section IV, the 2018 reference data collected for oleaginous/leguminous are much smaller in terms of surfaces than those collected for the cereals class. This apparent imbalance between crop classes affects both the direct transfer and the domain adaptation strategies, thus introducing a class distribution bias with respect to the source domain into the target domain.

2) *Visualization of Internal Model Representations*: In this last stage of our experimental assessment, we provide a visual inspection of the internal feature representation learned by CDAN + E, M^3SPADA_{VHR} , M^3SPADA_{SITS} , and M^3SPADA on the transfer task (2018 \rightarrow 2021). To this end, we randomly chose 200 samples per land cover class from the target domain, and we extracted the corresponding feature representations, per method. Subsequently, we applied truncated stochastic neighbourhood embedding (t-SNE) [46] to reduce the feature dimensionality for visualization purposes. Results are depicted in Fig. 14.

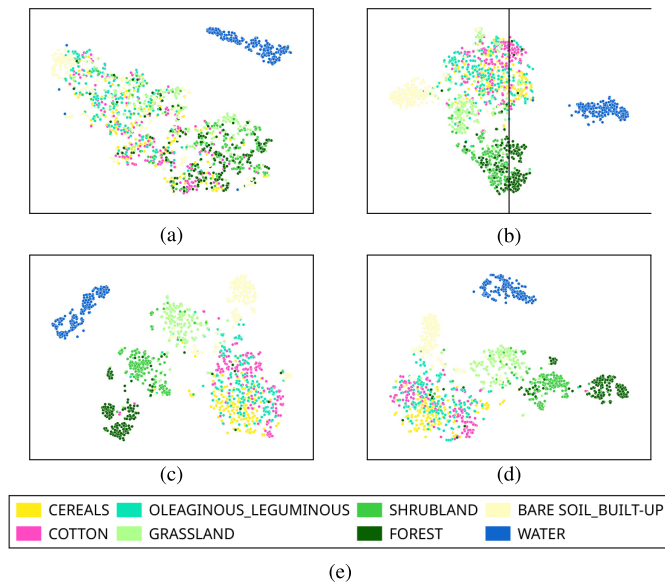


Fig. 14. t-SNE visualization of internal feature representation learned by (a) CDAN + E, (b) M³SPADA_{VHR}, (c) M³SPADA_{SITS}, and (d) M³SPADA over 200 randomly selected samples per class from the target domain considering the transfer task (2018 → 2021). (e) Land cover classes legend.

We note that the CDAN + E method mainly separates the water class from all the other ones. This is consistent with results described in Figs. 7 and 9(b). Conversely, the M³SPADA_{VHR} approach makes a step further also distinguishes the bare soil/built-up class in addition to the water samples from the rest of the data. These two methods clearly mix samples from all the other land cover classes together. Considering M³SPADA_{SITS} and M³SPADA, they partially alleviate the visual clutter problems of the remaining classes with the latter providing slightly better visual behavior. This can be observed, for instance, for the grassland, scrubland, and forest land cover classes.

Overall, the visualization of internal features representation is coherent with the quantitative as well as qualitative findings we previously discussed.

VI. CONCLUSION

We have presented M³SPADA, a new framework to cope with temporal UDA for multisensor land cover classification from multitemporal and multiscale remote sensing data. More precisely, we consider as multisensor input data, for both source and target domains: 1) optical HR Sentinel-2 SITS and 2) VHR Spot-6/7 optical imagery. Our framework jointly exploits adversarial learning and pseudo-labeling with the aim to gradually transfer/adapt a multisensor neural network model from a source domain (a specific year featured by GT data) to a target domain (an unlabeled different year) with the aim to produce an LCM on the latter. While domain-invariant features are extracted by means of adversarial learning, a spatially aware pseudo-labeling procedure is conceived and implemented to effectively transfer the multisensor classification model from the labeled source to the unlabeled target domain.

The obtained results on the *Koumbia* study site have underlined that our framework, taking explicitly advantage of the

spatiotemporal features related to the underlying multisensor remote sensing data, clearly outperforms all the UDA competitors we have considered in the quantitative evaluation.

Several future works can be planned as possible extensions of our M³SPADA. Regarding the multisensor aspect, we can go a step further involving additional remote sensing sources as, for instance, radar SITS information coming from the open Sentinel-1 mission.

Concerning methodological extensions of M³SPADA, we can consider a scenario in which multiple labeled source domains are available (i.e., multiple years of previous field campaign over the same study area), and the goal is still to transfer a classification model toward a single target domain. This kind of setting is referred as multisource UDA. The main challenge related to this setting will be how to make values of multiple labeled source dataset in order to enhance the extraction of invariant features from multiple domains.

Another possible methodological future works can be related to make the reasonable and realistic assumption to have limited, but still worthy of interest, reference data on the target domain to integrate in the transfer learning process and move our research effort to extend M³SPADA toward a semi-supervised domain adaptation scenario.

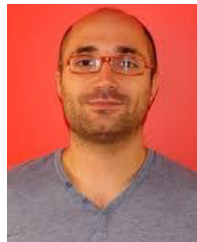
REFERENCES

- [1] M. Schmitt and X. X. Zhu, "Data fusion and remote sensing: An ever-growing relationship," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 4, pp. 6–23, Dec. 2016.
- [2] M. Berger, J. Moreno, J. A. Johannessen, P. F. Levelt, and R. F. Hanssen, "ESA's sentinel missions in support of Earth system science," *Remote Sens. Environ.*, vol. 120, pp. 84–90, May 2012.
- [3] D. Hong et al., "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.
- [4] J. Li et al., "Deep learning in multimodal remote sensing data fusion: A comprehensive review," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, Jan. 2022, Art. no. 102926.
- [5] P. Benedetti, D. Ienco, R. Gaetano, K. Ose, R. G. Pensa, and S. Dupuy, "M³Fusion: A deep learning architecture for multiscale multimodal multitemporal satellite data fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4939–4949, Dec. 2018.
- [6] K. K. Gadiraju, B. Ramachandra, Z. Chen, and R. R. Vatsavai, "Multimodal deep learning based crop classification using multispectral and multitemporal satellite imagery," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2020, pp. 3234–3242.
- [7] L. Bergamasco, F. Bovolo, and L. Bruzzone, "A dual-branch deep learning architecture for multisensor and multitemporal remote sensing semantic segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 2147–2162, 2023.
- [8] B. Tardy, J. Inglada, and J. Michel, "Fusion approaches for land cover map production using high resolution image time series without reference data of the corresponding period," *Remote Sens.*, vol. 9, no. 11, p. 1151, Nov. 2017. [Online]. Available: <https://www.mdpi.com/2072-4292/9/11/1151>
- [9] B. Tardy, J. Inglada, and J. Michel, "Assessment of optimal transport for operational land-cover mapping using high-resolution satellite images time series without reference data of the mapping period," *Remote Sens.*, vol. 11, no. 9, p. 1047, May 2019. [Online]. Available: <https://www.mdpi.com/2072-4292/11/9/1047>
- [10] J. Nyborg, C. Pelletier, S. Lefèvre, and I. Assent, "TimeMatch: Unsupervised cross-region adaptation by temporal shift estimation," *ISPRS J. Photogramm. Remote Sens.*, vol. 188, pp. 301–313, Jun. 2022.
- [11] G. Wilson and D. J. Cook, "A survey of unsupervised deep domain adaptation," *ACM Trans. Intell. Syst. Technol.*, vol. 11, no. 5, pp. 51:1–51:46, 2020.

- [12] L. Bai, S. Du, X. Zhang, H. Wang, B. Liu, and S. Ouyang, "Domain adaptation for remote sensing image semantic segmentation: An integrated approach of contrastive learning and adversarial learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5628313.
- [13] J. Chen, J. Zhu, Y. Guo, G. Sun, Y. Zhang, and M. Deng, "Unsupervised domain adaptation for semantic segmentation of high-resolution remote sensing imagery driven by category-certainty attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5616915.
- [14] M. Martini, V. Mazzia, A. Khaliq, and M. Chiaberge, "Domain-adversarial training of self-attention-based networks for land cover classification using multi-temporal Sentinel-2 satellite imagery," *Remote Sens.*, vol. 13, no. 13, p. 2564, Jun. 2021.
- [15] E. Capliez, D. Ienco, R. Gaetano, N. Baghdadi, and A. H. Salah, "Temporal-domain adaptation for satellite image time-series land-cover mapping with adversarial learning and spatially aware self-training," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 3645–3675, 2023.
- [16] J. Wang, A. Ma, Y. Zhong, Z. Zheng, and L. Zhang, "Cross-sensor domain adaptation for high spatial resolution urban land-cover mapping: From airborne to spaceborne imagery," *Remote Sens. Environ.*, vol. 277, Aug. 2022, Art. no. 113058.
- [17] Y. J. E. Gbodjo, O. Montet, D. Ienco, R. Gaetano, and S. Dupuy, "Multisensor land cover classification with sparsely annotated data based on convolutional neural networks and self-distillation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 11485–11499, 2021.
- [18] S. Liu, Q. Shi, and L. Zhang, "Few-shot hyperspectral image classification with unknown classes using multitask deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5085–5102, Jun. 2021.
- [19] Y. Dong, T. Liang, Y. Zhang, and B. Du, "Spectral-spatial weighted kernel manifold embedded distribution alignment for remote sensing image classification," *IEEE Trans. Cybern.*, vol. 51, no. 6, pp. 3185–3197, Jun. 2021.
- [20] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 778–782, May 2017.
- [21] D. Ienco, R. Interdonato, R. Gaetano, and D. H. T. Minh, "Combining Sentinel-1 and Sentinel-2 satellite image time series for land cover mapping via a multi-source deep learning architecture," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 11–22, Dec. 2019.
- [22] S. Ofori-Ampofo, C. Pelletier, and S. Lang, "Crop type mapping from optical and radar time series using attention-based deep learning," *Remote Sens.*, vol. 13, no. 22, p. 4668, Nov. 2021.
- [23] V. Sainte Fare Garnot, L. Landrieu, and N. Chehata, "Multi-modal temporal attention models for crop mapping from satellite time series," *ISPRS J. Photogramm. Remote Sens.*, vol. 187, pp. 294–305, May 2022.
- [24] F. Zhuang et al., "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021.
- [25] Y. Ganin et al., "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2030–2096, 2017.
- [26] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2962–2971.
- [27] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," in *Proc. NeurIPS*, 2018, pp. 1647–1657.
- [28] Y. Zhang, T. Liu, M. Long, and M. Jordan, "Bridging theory and algorithm for domain adaptation," in *Proc. 36th Int. Conf. Mach. Learn.*, vol. 97, Jun. 2019, pp. 7404–7413.
- [29] M. Chen, S. Zhao, H. Liu, and D. Cai, "Adversarial-learned loss for domain adaptation," in *Proc. AAAI*, 2020, pp. 3521–3528.
- [30] F. Qi, X. Yang, and C. Xu, "A unified framework for multimodal domain adaptation," in *Multimedia*. New York, NY, USA: ACM, 2018, pp. 429–437.
- [31] J. Munro and D. Damen, "Multi-modal domain adaptation for fine-grained action recognition," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3723–3726.
- [32] W. Zhang, D. Xu, J. Zhang, and W. Ouyang, "Progressive modality cooperation for multi-modality domain adaptation," *IEEE Trans. Image Process.*, vol. 30, pp. 3293–3306, 2021.
- [33] D. Tuia, C. Persello, and L. Bruzzone, "Recent advances in domain adaptation for the classification of remote sensing data," *Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 41–57, 2016.
- [34] J. Nath Kundu, N. Venkat, A. Revanur, M. V. Rahul, and R. Venkatesh Babu, "Towards inheritable models for open-set domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12373–12382.
- [35] D.-H. Lee et al., "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2013, vol. 3, no. 2, p. 896.
- [36] Y. Li, D. Huang, D. Qin, L. Wang, and B. Gong, "Improving object detection with selective self-supervised self-training," in *Proc. ECCV*, 2020, pp. 589–607.
- [37] C. Pelletier, G. Webb, and F. Petitjean, "Temporal convolutional neural network for the classification of satellite image time series," *Remote Sens.*, vol. 11, no. 5, p. 523, Mar. 2019.
- [38] O. Hagolle, M. Huc, D. Villa Pascual, and G. Dedieu, "A multi-temporal and multi-spectral method to estimate aerosol optical thickness over land, for the atmospheric correction of FormoSat-2, LandSat, VEN μ S and Sentinel-2 images," *Remote Sens.*, vol. 7, no. 3, pp. 2668–2691, Mar. 2015.
- [39] J. Inglada, A. Vincent, M. Arias, and B. Tardy, "iota2-a25386," *Tech. Rep.*, Jul. 2016, doi: 10.5281/zenodo.58150.
- [40] A. Jolivot et al., "Harmonized in situ JECAM datasets for agricultural land use mapping and monitoring in tropical countries," *Tech. Rep.*, 2021, doi: 10.18167/DVNI/P7OLAP.
- [41] J. Shen, Y. Qu, W. Zhang, and Y. Yu, "Wasserstein distance guided representation learning for domain adaptation," in *Proc. AAAI*, 2018, pp. 4058–4065.
- [42] N. Karasiak, J.-F. Dejoux, C. Monteil, and D. Sheeren, "Spatial dependence between training and test sets: Another pitfall of classification accuracy assessment in remote sensing," *Mach. Learn.*, vol. 111, pp. 2715–2740, Apr. 2021.
- [43] Y. Li, N. Wang, J. Shi, J. Liu, and X. Hou, "Revisiting batch normalization for practical domain adaptation," in *Proc. ICLR*, 2017, pp. 1–12.
- [44] A. de Mathelin et al., "ADAPT : Awesome domain adaptation Python toolbox," 2021, *arXiv:2107.03049*.
- [45] W. Shi, Y. Gong, C. Ding, Z. Ma, X. Tao, and N. Zheng, "Transductive semi-supervised deep learning using min-max features," in *Proc. ECCV*, vol. 11209, 2018, pp. 311–327.
- [46] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.



Emmanuel Capliez (Graduate Student Member, IEEE) received the Engineering degree in computer science with specialty in computer vision and remote sensing from the ENSTA Paris IP Engineering School, Paris, France, in 2004. He is currently pursuing the Ph.D. degree in computer science with the Joint Research Unit on Territory, Environment, Remote Sensing and Spatial Information, National Research Institute for Agriculture, Food and the Environment (INRAE), University of Montpellier, Montpellier, France, with a focus on temporal domain adaptation approaches devoted to manage remote sensing data for land cover mapping.



Dino Ienco (Member, IEEE) received the M.Sc. and Ph.D. degrees in computer science from the University of Torino, Turin, Italy, in 2006 and 2010, respectively.

In 2011, he joined the TETIS, National Research Institute for Agriculture, Food and the Environment (INRAE), University of Montpellier, Montpellier, France, as a Junior Researcher. His main research interests include machine learning, data science, graph databases, social media analysis, information retrieval, and spatiotemporal data analysis with a particular emphasis on remote sensing data and Earth observation data fusion.

Dr. Ienco served in the program committee of many international conferences on data mining, machine learning, and database, including the IEEE International Conference on Data Mining (ICDM), European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), Asian Conference on Machine Learning (ACML), and International Joint Conference on Artificial Intelligence (IJCAI). He served as a reviewer for many international journals in the general field of data science and remote sensing.



Raffaele Gaetano received the Laurea (M.S.) degree in computer engineering and the Ph.D. degree in electronic and telecommunication engineering from the University of Naples Federico II, Naples, Italy, in 2004 and 2009, respectively.

From 2009 to 2015, he conducted post-doctoral research on fundamental image processing mainly applied to remote sensing for different institutions (National Institute for Research in Digital Science and Technology (INRIA), France; Institute for Computer Science and Control (MTA SZTAKI), Budapest, Hungary; Telecom ParisTech, Paris, France; and University of Naples Federico II). Since 2015, he has been a Permanent Researcher with the French Agricultural Research Centre for International Development (CIRAD), TETIS Research Unit, University of Montpellier, Montpellier, France. His research interests include machine learning for remote sensing image analysis and processing, mainly oriented to the design of operational methods for information extraction from multisensor and multitemporal imagery (e.g., land cover and land use mapping).



Nicolas Baghdadi received the Ph.D. degree from the University of Toulon, Toulon, France, in 1994.

From 1995 to 1997, he was a Post-Doctoral Researcher with the INRS Ete Water Earth Environment Research Centre, Quebec University, Montreal, QC, Canada. From 1998 to 2008, he was with the French geological Survey (BRGM), Orleans, France. From 2013 to 2022, he was the Scientific Director of the French Land Data Center, France. Since 2008, he has been the Research Director with National Research Institute for Agriculture, Food and the Environment (INRAE), University of Montpellier, Montpellier, France. His research interests include the analysis of remote sensing data (mainly radar and lidar) and the retrieval of environmental parameters (e.g., soil moisture content, soil roughness, canopy height, and forest biomass).

Dr. Baghdadi is an Editor of two series of books: *Land Surface Remote Sensing* set and *QGIS* in remote sensing set.



Adrien Hadj Salah received the Engineering degree from the Ecole Nationale de l'Aviation Civile (ENAC) Engineering School, Toulouse, France, in 2012, and the master's degree in air and space operations technology from the Technical University of Berlin, Berlin, Germany, in 2012.

He is currently a Computer Vision and Advanced Studies Team Leader with Airbus Defence and Space, Toulouse.



Matthieu Le Goff received the Engineering degree in computer vision from École Centrale Paris, Gif-sur-Yvette, France, in 2014, and the Ph.D. degree in artificial intelligence from IRT Saint Exupéry, Toulouse, France, in 2017.

After his Ph.D. degree, he joined Airbus Defence and Space, Toulouse. He first worked on deep learning-based information extraction techniques for Earth observation projects and gradually switched to vision-based navigation projects, including the ESA's Jupiter Icy Moons Explorer (JUICE) image processing algorithm.



Florient Chouteau received the degree from the École nationale supérieure de l'aéronautique et de l'espace (SAE-SUPAERO), Toulouse, France, in 2016.

He has worked in artificial intelligence applied to remote sensing—both in information extraction and image processing. He is currently occupying the role of artificial intelligence expert for Earth observation systems at Airbus Defence and Space, Toulouse.