Bidirectional Recurrent Imputation and Abundance Estimation of LULC Classes With MODIS Multispectral Time-Series and Geo-Topographic and Climatic Data

José Rodríguez-Ortega, Rohaifa Khaldi, Domingo Alcaraz-Segura, and Siham Tabik

***Abstract*—Remotely sensed data are dominated by mixed land use and land cover (LULC) types. Spectral unmixing (SU) is a key technique that disentangles mixed pixels into constituent LULC types and their abundance fractions. While existing studies on deep learning (DL) for SU typically focus on single time-step hyperspectral or multispectral data, our work pioneers SU using MODIS MS time series, addressing missing data with end-to-end DL models. Our approach enhances a long-short-term-memory-based model by incorporating geographic, topographic (geo-topographic), and climatic ancillary information. Notably, our method eliminates the need for explicit endmember extraction, instead learning the input–output relationship between mixed spectra and LULC abundances through supervised learning. Experimental results demonstrate that integrating spectral-temporal input data with geo-topographic and climatic information significantly improves the estimation of LULC abundances in mixed pixels. To facilitate this study, we curated a novel labeled dataset for Andalusia (Spain) with monthly MODIS MS time series at 460-m resolution for 2013. Named Andalusia MultiSpectral MultiTemporal Unmixing, this dataset provides pixel-level annotations of LULC abundances along with ancillary information.**

***Index Terms*—Abundance estimation, bidirectional long-short term memory (LSTM), climatic data, deep learning (DL), geo-topographic data, land use and land cover (LULC), missing values, spectral unmixing (SU).**

Manuscript received 5 November 2023; revised 31 December 2023 and 14 January 2024; accepted 16 January 2024. Date of publication 29 January 2024; date of current version 15 February 2024. This work was supported in part by the project EarthCul PID2020-118041GB-I00 within the Spanish Research Projects Plan supported by MCIN/AEI/10.13039/501100011033 and by FEDER funds "Una manera de hacer Europa"; in part by the project ECOPOTENTIAL, supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant 641762; in part by the TED project under Grant TED2021-129690B-I00, supported by the Ministry of Science and Innovation; and in part by the project "Thematic Center on Mountain Ecosystem and Remote sensing, Deep learning-AI e-Services University of Granada-Sierra Nevada" under Grant LifeWatch-2019-10-UGR-01, which has been cofunded by the Ministry of Science and Innovation through the FEDER funds from the Spanish Pluriregional Operational Program 2014-2020 (POPE), LifeWatch-ERIC action line. *(Corresponding authors: José Rodríguez-Ortega; Domingo Alcaraz-Segura.)*

José Rodríguez-Ortega and Rohaifa Khaldi are with the Department of Computer Science and Artificial Intelligence, Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, 18071 Granada, Spain, and also with LifeWatch-ERIC, 41071 Seville, Spain (e-mail: e.jrodriguez98@go.ugr.es; rohaifa@go.ugr.es).

Domingo Alcaraz-Segura is with the Department of Botany, Faculty of Science and with the iEcolab, Andalusian Institute for Earth System Research (IISTA-CEAMA), University of Granada, 18071 Granada, Spain, and also with the Andalusian Center for the Assessment and Monitoring of Global Change, University of Almería, 04120 Almería, Spain (e-mail: dalcaraz@ugr.es).

Siham Tabik is with the Department of Computer Science and Artificial Intelligence, Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, 18071 Granada, Spain (e-mail: siham@ugr.es).

The dataset https://zenodo.org/records/7752348 and code https://github.com/jrodriguezortega/MSMTU are available to the public.

Digital Object Identifier 10.1109/JSTARS.2024.3359647

I. Introduction

LAND use and land cover (LULC) mapping is normally addressed by classifying each pixel in a satellite image into a LULC class, also known as semantic segmentation (SS) in RS images. Frequently, the spatial resolution of an image and the thematic resolution of its LULC legend do not match, which leads to the mixed pixel problem, where pixels are not pure but contain several LULC classes. Accordingly, many methods have tried to estimate the relative abundances of each LULC class in a pixel from the combined spectral signature [1]. Such estimation of the spectrum and the abundance of the LULC classes present within each pixel is known as spectral unmixing (SU) and is one of the most challenging areas of research in remote sensing (RS) [2]. Various unmixing approaches, including linear and nonlinear methods, have been developed [3], [4]. Many of these approaches require the use of the pure spectral signature (the endmember) of each LULC class. However, the acquisition of endmembers might be hard in areas dominated by mixed pixels [5]. To overcome this limitation, several methods have been introduced to avoid the need of endmembers extraction [2], [5], [6], [7], as depicted in Fig. 1.

In the last few years, modern deep learning (DL) models have been increasingly employed for addressing SU by directly learning the input–output mapping from the spectra of mixed pixels to their corresponding class abundances. Several studies explored the potential of DL methods for SU in LULC mapping using either single time-step hyperspectral (HS) data [2], [8], [9] or single time-step multispectral (MS) data [10]. Including temporal information could be a great opportunity to improve SU methods [4] and a few works (see Table I) have started exploring approaches with MS time-series data. However, to the best of our knowledge, none have explored an end-to-end DL solution, where recurrent neural networks (RNNs) and LSTM networks are a perfect fit.

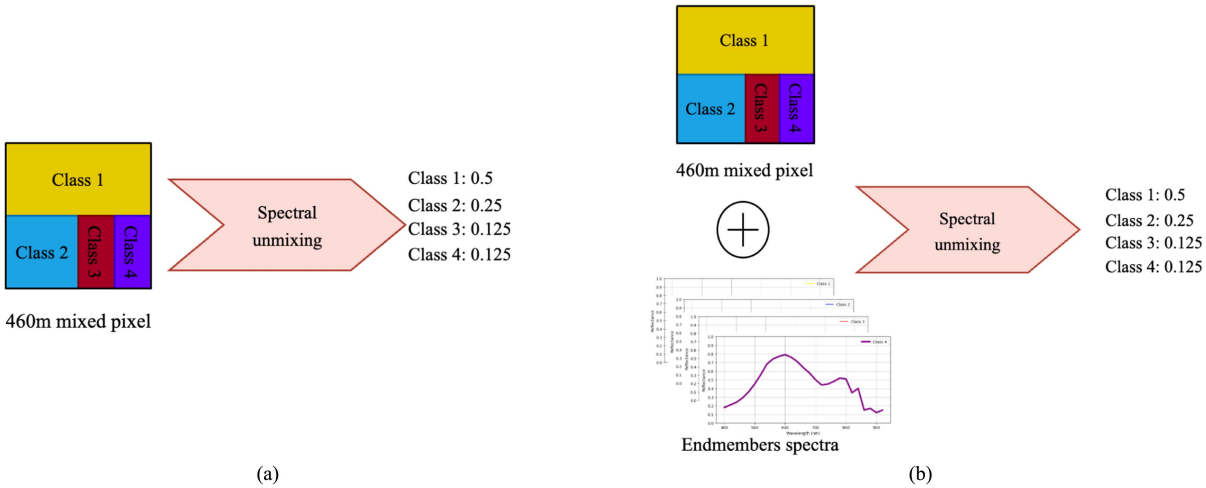

Fig. 1. Toy example illustrating (a) SU without endmember extraction versus (b) SU with endmember extraction in a 460-m mixed pixel. (a) SU without endmember extraction. (b) SU with endmember extraction.

TABLE I
LIST OF SU WORKS USING MS AND MULTITEMPORAL DATA

Ref. and year	Study area	Source	Spatial-temporal res.	Annotations	# classes	Public	Approach
[61] 2004	Yaqui Valley (Mexico)	MODIS	500m - 8 obs. (months)	HRM (Landsat-derived)	3	No	PTU
	Great Plains (US)	MODIS	500m - 8 obs. (months)	HRM (Landsat-derived)	3	No	PTU
[62] 2011	The Netherlands	MERIS	300m - 7 obs. (months)	HRM (LGN5 L1)	4	No	FCLSU
	The Netherlands	MERIS	300m - 7 obs. (months)	HRM (LGN5 L2)	12	No	FCLSU
[63] 2012	Japan	MODIS	250m - 365 obs. (days)	HRM (Landsat-derived)	2	No	TMA-based
[64] 2020	Broome County (US)	Landsat	30m - (2000 to 2014)	HRM (NLCD)	2	No	RF
[65] 2020	Rondônia (Brazil)	Landsat	30m - (1990 to 2013)	HRM (Rondônia)	5	No	OLS + RF
[5] 2021	Daxing (China)	MODIS	480m - 3 obs. (months)	HRM (Landsat-derived)	2	No	STSU
	Zibo (China)	MODIS	480m - 3 obs. (months)	HRM (Landsat-derived)	2	No	STSU
	Amazon (Brazil)	MODIS	480m - 3 obs. (months)	HRM (Landsat-derived)	2	No	STSU
[7] 2021	Daxing (China)	MODIS	480m - 2 obs. (months)	HRM (Landsat-derived)	2	No	RSTSU
	Zibo (China)	MODIS	480m - 2 obs. (months)	HRM (Landsat-derived)	2	No	RSTSU
	Lichuan (China)	MODIS	480m - 2 obs. (months)	HRM (Landsat-derived)	2	No	RSTSU
Ours	Andalusia (Spain)	MODIS	460m - 12 obs. (months)	HRM (SIPNA L1)	4	Yes	LSTM-based
	Andalusia (Spain)	MODIS	460m - 12 obs. (months)	HRM (SIPNA L2)	10	Yes	LSTM-based

HRM: Higher Resolution Map, PTU: Probabilistic Temporal Unmixing, FCLSU: Fully Constrained Least Squares SU, TMA: Temporal Mixture Analysis, RF: Random Forest, OLS: Ordinary Least Square, STSU: Spatio-Temporal SU, RSTSU: Real-Time STSU, LSTM: Long- Short Term Memory

In contrast to traditional methods, the application of DL in SU facilitates the exploitation of ancillary information such as geographic location, topography, and climate. For example, in the field of computer vision, ancillary data have been successfully used by DL models to improve the performance during image classification [11], [12], [13]. However, the introduction of ancillary information remains unexplored in SU methods. We hypothesize that injecting such ancillary information could boost the performance of the predictive model in SU. This information may help the model understand the spatial distribution and variations in climate of the different LULC types.

The primary problem addressed in this study is the SU of LULC classes using MS time-series data and ancillary information, and it faces several challenges.

1) Public labeled datasets with MS multitemporal data for SU of LULC classes are not available.
2) Creating a new dataset of MS time-series plus ancillary information together with LULC abundances annotations is complex, costly, and time consuming.
3) RS data usually contain missing values due to atmospheric conditions or sensors' errors, which requires applying robust processing techniques.
4) Feeding ancillary information to SU models is a promising direction but can be complex. Ensuring that these data improve the model robustness is a challenge and it is not explored yet.

Given the aforementioned challenges, the main objective of this study is twofold.

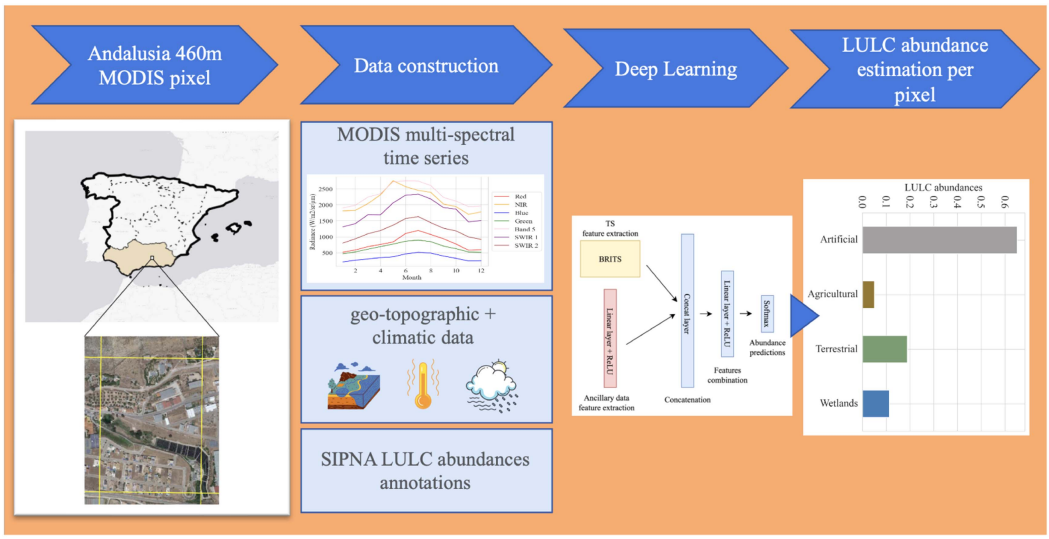

Fig. 2. Flowchart of our proposed method. First, Andalusia-MSMTU dataset is built using MODIS MS multitemporal data plus geo-topographic and climatic data together with the corresponding LULC abundances annotations extracted from SIPNA. Subsequently, the deep-learning-based model is designed to use both, MS times-series and geo-topographic and climatic data to estimate the LULC abundances.

1) To create a regional-scale dataset of more than 500 000 MODIS 460-m resolution pixels from Andalusia, Spain.
2) To develop a DL-based approach for SU, without the need of endmember extraction, which estimates the LULC abundances per pixel using MS time-series and ancillary data. This dataset provides for each individual pixel:

a) an MS time series of monthly observations during the year 2013 of the seven spectral bands of the MODIS sensor;
b) ancillary information containing geographic, topographic, and climatic variables;
c) their corresponding LULC class abundances at two different levels of the classes hierarchy, extracted Andalusia's official LULC map (SIPNA [14]).

Furthermore, the DL-based method consists of a two-branch neural network (NN) where the first branch process the MS time series using an LSTM-based model capable of handling missing values, and the second branch process the ancillary information. A graphical illustration of the followed workflow in this study is shown in Fig. 2.

The following two assumptions are made in this study:

1) LULC changes within a one-year timeframe are limited at a 460-m pixel resolution, so our LULC abundance annotations are assumed to be static;
2) the selected MODIS time-series data, ancillary information, and LULC annotations adequately represent the land dynamics of Andalusia, since more than 500 000 pixels from Andalusia are collected, representing almost the whole region.

The constraint includes the challenge of dealing with missing values in RS data, which we solve by proposing a DL method capable of handling missing values.

The motivation behind this research is rooted in the need for improved methods to perform SU in complex, heterogeneous landscapes with MS time-series data. The absence of accessible labeled datasets, combined with the complexity of creating new datasets, underscores the significance of developing innovative approaches to advance the field.

The primary contributions of this research can be summarized as follows.

1) We built Andalusia-MSMTU dataset: a novel MS multitemporal labeled dataset with mixed pixels from Andalusia, a highly heterogeneous region in Spain. Each pixel is annotated with LULC abundances. In addition to the MS multitemporal information, each mixed pixel has its corresponding geo-topographic and climatic information. Such dataset will open the possibility for new explorations.
2) We designed and analyzed a DL-based approach that estimates the LULC abundances per pixel of LULC classes from MS time-series data with and without ancillary information.

The rest of this article is organized as follows. Section II presents the related work. Section III provides preliminaries and background. Section IV introduces the study area and the data construction process. Section V describes the used DL methodology. Section VI assesses the results obtained. Section VII provides a comprehensive discussion of the obtained results, comparing them with previous works. Finally, Section VIII concludes this article.

II. Related Work

First, general DL in RS methods is reviewed. Subsequently, related works on SU overall, with a specific focus on employing DL methodologies, are introduced. Finally, works that build labeled datasets designed for the unmixing approaches are reviewed and comprehensively summarized in Table I.

A. Deep Learning (DL) in Remote Sensing (RS)

Thanks to the recent success of DL methods in many learning tasks, tons of efforts have been made to bring DL to RS field [15].

Concretely, the LULC classification task is of paramount importance since many environmental applications rely on LULC maps, such as urban planning, forest monitoring, change detection, etc.

Traditionally, only one source of input data was used to perform the classification task, that is, only using HS [16], MS [17], LiDAR [18], or synthetic aperture radar (SAR) [19]. Recently, multimodal models have emerged with the promise to improve the LULC classification by combining the different input data types. Hong et al. [20] addressed challenges in the LULC classification using a multimodal DL (MDL) framework. They tackle limitations of traditional DL in complex scenes, introducing five fusion architectures and emphasizing applicability beyond pixel-wise classification to spatial information modeling. Also, Han et al. [21] introduced MUNet, a multimodal unmixing network for HS images, leveraging LiDAR data to enhance discrimination in complex scenes. MUNet uses an SE-driven attention mechanism, incorporating height differences from LiDAR for improved performance. Uezato et al. [22] presented IISU, an illumination invariant SU model addressing spectral variability caused by variable incident illuminations. Utilizing radiance HS data and a LiDAR-derived digital surface model, IISU provides explicit explanations for endmember variability, outperforming existing models, particularly in shaded pixels. The proposed model yields more accurate abundances and shadow-compensated reflectance. Hong et al. [23] built the C2Seg dataset for cross-city LULC classification, addressing limitations of DL models across diverse urban environments. Their proposed HighDAN network, employing high-resolution domain adaptation and adversarial learning, demonstrates superior segmentation performance and generalization abilities compared to existing methods. Following the modern self-supervised learning (SSL) paradigm, SpectralGPT [24] is proposed as a novel universal foundation model tailored for spectral RS data, utilizing a 3-D generative pretrained transformer. Trained on one million spectral RS images, it accommodates varied inputs, leverages 3-D token generation for spatial-spectral coupling, and achieves substantial performance gains across geoscience tasks like scene classification and semantic segmentation. Finally, Hong et al. [25] introduce a subpixel-level HS super-resolution framework, DC-Net, addressing the distribution gap between HS and high spatial resolution MS images. The novel decoupled-and-coupled network progressively fuses information from pixel to subpixel-level, mitigating spatial and spectral resolution differences. Employing an SSL module ensures material consistency for enhanced HS restoration.

B. Spectral Unmixing (SU)

The existing SU methods can be broadly categorized as linear mixture models (LMMs) and nonlinear mixture models (NLMM) according to the formulation describing the underlying mixing process of endmembers [26].

LMM consider that the spectral signature of a mixed pixel is a weighted sum of the endmember spectra and that the weights associated with the endmembers are given by their corresponding relative area abundance in the pixel. LMM-based methods have been widely developed in last decades including linear, geometrical, nonnegative matrix factorization, bayesian, and fuzzy models among others [3], [27], [28], [29], [30], [31]. LMM typically assumes that the spectrum of each LULC class is characterized by a single fixed endmember. However, pure pixels from the same LULC class may have different spectra, which is called intraclass variability [32]. To overcome this limitation, several multiple endmember spectral mixture analysis models have been developed [33], [34], [35], [36].

Since the extraction of a large number of pure endmembers is still a great challenge in areas dominated by mixed pixels, several works without assuming any prior knowledge about the mixing process were introduced. These methods, also known as blind spectral unmixing (BSU) methods, include independent component analysis [37], [38], [39], nonnegative matrix factorization [40], [41], [42], [43], sparse component analysis [44], or wavelet-based [45] methods.

Given the nonlinear mixing effects of endmembers, NLMM have been proposed through the years to overcome LMM limitations and enhance the SU performance. These include bilinear models [46], radial basis function networks [47], kernel-based models [48], NNs, and low-rank tensor [49] methods among others.

1) DL in SU: SU has also met DL models, which fall under the category of NLMM. One of the first DL approaches for SU was proposed by Foody [50], where three spectral bands values were introduced, and the NN predicts the abundances of three LULC classes. Atkinson et al. [51] compared NN, LMMs, and fuzzy c-means for SU of LULC classes, with NN being the best model given sufficient training samples. Then, Licciardi and Del Frate [6] proposed a two-stage NN architecture that first reduces the dimension of the input vector using an autoassociative NN, and then performs abundance estimation out of the reduced input using an MLP . Recently, Palsson et al. [8] evaluated autoencoders with different hyperparameters. Yu et al. [52] introduced MSNet, a multistage convolutional autoencoder network designed for linear HU, achieving this by capturing contextual relationships between pixels. Gao et al. [53] introduced CyCU-Net for HU, enhancing the performance by incorporating cycle consistency and self-perception loss. The network, leveraging cascaded autoencoders, preserves detailed material information and achieves high-level semantic preservation during unmixing. Yao et al. [54] introduced SeCoDe, a novel blind HS unmixing model designed for airborne and spaceborne HS imagery. Leveraging sparsity-enhanced convolutional decomposition, SeCoDe effectively addresses spectral variabilities and maintains continuous spectral components. Going beyond autoencoder-like architectures, Han et al. [55] introduced Deep HSNet, a novel siamese network for HU that considers diverse endmember properties from different extraction algorithms. Deep HSNet incorporates a subnetwork to effectively learn endmember information, enhancing the accuracy of the unmixing process. Following the success of transformers architecture [56], the authors in [9] and [57] introduced NN architectures with the attention mechanism for abundance estimation. Regarding SSL for SU works, Vijayashekhar et al. [58] proposed a two-stage fully connected SSL network for BSU, addressing challenges of

limited supervision and data requirements . The network jointly estimates endmembers and abundances in the first stage, and learns HS image acquisition physics in the second stage. Also, AutoNAS [59] explored neural architecture search (NAS) for determining the optimal network architecture in HU. Utilizing SSL and an affine parameter sharing strategy, it achieves optimal channel configuration. Furthermore, an evolutionary algorithm enables flexible convolution kernel search.

Regarding RNN-based works, the only work on SU using an LSTM-based network was introduced by Zhao et al. [60]. They proposed a nonsymmetric autoencoder network with an LSTM component to capture spectral correlation together with an attention mechanism to further enhance the unmixing performance. For a more detailed review of DL methods in SU, see [4] and [26].

In parallel, there exist few works that incorporate ancillary data to improve the performance of DL models. Most of these studies occur in the field of computer vision, such as high interclass similarity classification problems [11], plankton image classification [12], or crop type mapping [13].

C. Labeled Datasets Based on LULC Products

Supervised learning requires high amounts of ground-truth data to achieve better generalization. One of the biggest limitations in SU is the limited availability of ground-truth LULC maps [4], [66]. Some areas or regions, especially in western countries, have LULC ground-truth based on visual interpretation for specific fields of study. For example, SIPNA [14] was intended for territorial planing in Spain. Its annotation was carried out by experts during several years. This dataset can be used to annotate RS data.

In parallel, there exist several annotated MS multitemporal datasets prepared for supervised SU (see Table I). However, all of them are private.

Our work is the first to provide a public, good-quality MS multitemporal mixed-pixel labeled dataset, named Andalusia-MSMTU that includes not only spectro-temporal information but also geo-topographic and climatic ancillary data. Andalucía-MSMTU is organized into two hierarchical levels of classes with four and ten LULC types, and it is especially suitable for building umimixing DL-based models for LULC abundance estimation. The proposed methodology constitutes the state-of-the-art in Andalusia-MSMTU.

III. Preliminaries and Background

We define a multivariate time series as a sequence of observations $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$, where T is the number of observations or time steps. Each observation $\mathbf{x}_t \in \mathbb{R}^C$, where $t \in \{1, \ldots, T\}$ consists of C variables, such that $\mathbf{x}_t = \{x_t^1, x_t^2, \ldots, x_t^C\}$.

A. Recurrent Neural Network (RNN)

RNN [67] is an NN architecture specifically designed for handling sequential data. RNN considers the sequential relationship of inputs by using a shared function f to process each input. RNN processes the time series using a recurrence approach at every time step t, computing a hidden state $\mathbf{h}_t$ by considering the previous hidden state $\mathbf{h}_{t-1}$ and the current input $\mathbf{x}_t$ as

$$\mathbf{h}_t = f(\mathbf{h}_{t-1}, \mathbf{x}_t) \tag{1}$$

where $\mathbf{h}_0$ is normally, at the beginning, the zero vector, i.e., $\mathbf{h}_0 = \mathbf{0}$.

There are several choices on how to process sequential information. In this work, we focus on the LSTM network, which is an improvement of the normal RNN solving some of its biggest limitations [68].

B. Bidirectional Recurrent Imputation for Time Series (BRITS)

In time-series data and specifically in RS data, it is common to find missing values due to sensor errors, cloud cover, and more [69]. To handle this situation, there exists a type of RNN architecture that can learn to solve two tasks simultaneously: imputing missing values and classifying the input sequence data. This model is called recurrent imputation for time series (RITS) [70]. The RITS model perform the imputation algorithm to assist the classification task and obtain the final classification as

$$\hat{\mathbf{y}} = f_{\text{out}}(\mathbf{h}_T) \tag{2}$$

where $\hat{\mathbf{y}}$ is the final classification, f_{out} is the classification function, and $\mathbf{h}_T$ is the last hidden state.

In practice, considering only unidirectional forward dynamic is problematic due to slow convergence, inefficiency in training and bias exploding problem [70]. To overcome these issues, a bidirectional version named BRITS model is proposed also in [69] to learn forward and backward patterns by accessing information from past and future at any given time step. The final scheme of BRITS can be seen in Fig. 3.

IV. Study Area and Data Construction

This section describes the study area and provides full details on how the used dataset was built and processed.

A. Study Area

Andalusia is the second-largest, most populous, and southernmost autonomous community in Peninsular Spain (see Fig. 4). Andalusia is one of the most biodiverse and heterogeneous regions of Europe. It contains steep altitudinal gradients, and it has a wide variety of landscapes and climatic conditions, which results in a vast variety of vegetation types that hold the greatest diversity of plant and animal species in Europe. The long and dynamic history of human activities has also led to a complex landscape configuration with frequent mosaics of small patches of different types of natural, seminatural land covers and human land uses. Hence, Andalusia offers an ideal field laboratory to test the creation of detailed and fine-scaled LULC maps containing the abundance of each LULC class per pixel to monitor the socioeconomic and environmental dynamics in complex landscapes using DL and MS time series of satellite imagery.

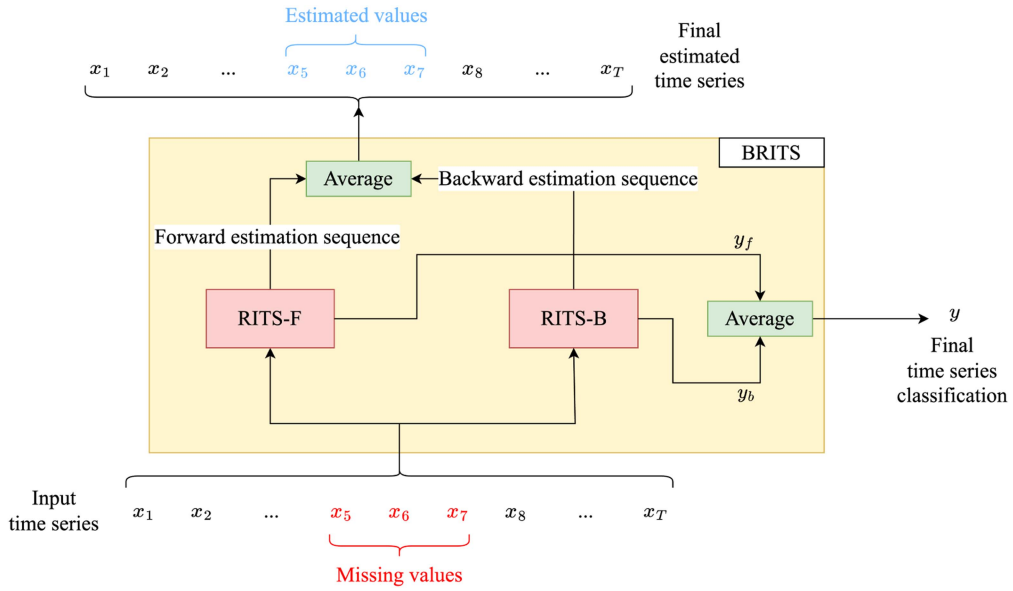

Fig. 3. BRITS architecture.

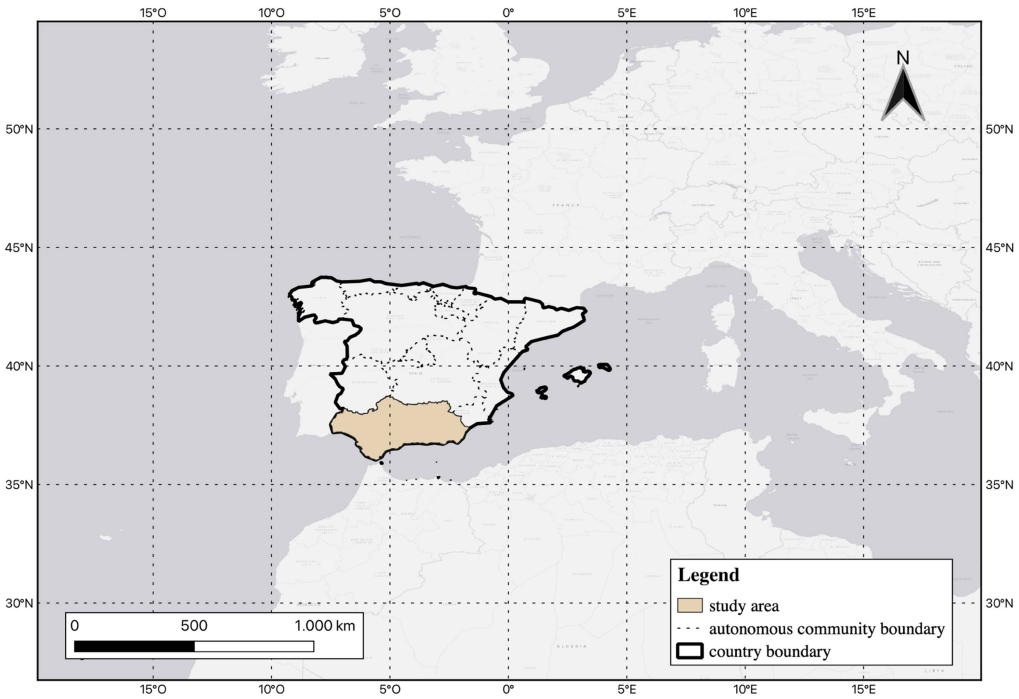

Fig. 4. Study area: Andalusia, Spain.

B. Andalusia-MSMTU Dataset Construction

To build Andalusia MultiSpectral, MultiTemporal Unmixing dataset, named Andalusia-MSMTU, several sources were utilized: MODIS, SRTM digital elevation data [71], REDIAM's environmental information [72], and SIPNA [14]. Herein, three different processes were used to create the dataset:) MODIS MS time-series extraction, ancillary data extraction, and LULC abundances annotations.

1) MODIS Pixel Time-Series Extraction: The time-series data were extracted from two satellites Terra and Aqua using MODIS sensor at 460-m spatial resolution and at monthly temporal resolution. As LULC changes during one year are very limited in a 460-m pixel, we assume that the LULC abundances are representative of the full year.

Spatio-temporal filtering was applied using MODIS 'quality assessment' (QA) flags and the "State QA" flags. Moreover, as the process of Terra and Aqua data filtering generates many

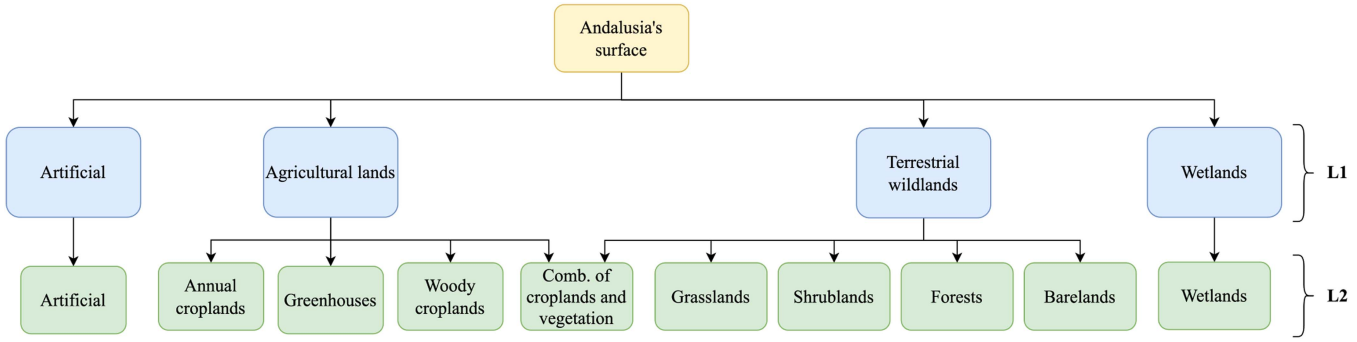

Fig. 5. Hierarchical structure of the SIPNA-based LULC classes. The blue boxes represent the Level 1 (L1) classes. The green boxes represents the Level 2 (L2) classes.

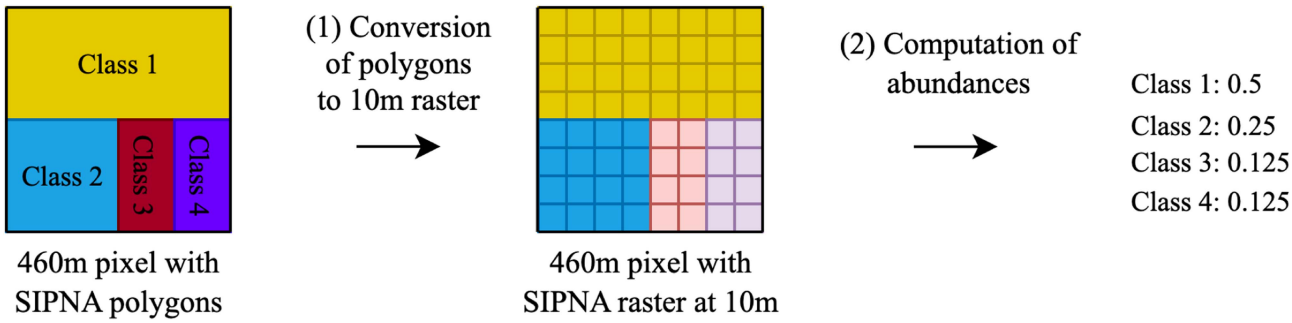

Fig. 6. Used scheme for extracting class abundances in every MODIS pixel of Andalusia. 1) Original SIPNA polygons were converted to a 10-m raster, then 2) LULC abundances were computed for each 460-m pixel.

missing values, to further reduce the amount of noise in the data, the following two solutions were employed:

1) the eight-days time-series data were transformed into monthly composites by computing the monthly mean from the individual observations;
2) the monthly data from the Terra and Aqua satellites were combined to generate a merged Terra+Aqua monthly dataset.

All this process was performed in Google Earth Engine (GEE) [73] and inspired by [74].

2) Ancillary Data Extraction for Each MODIS Pixel: In addition to the MODIS data, for every pixel, we included geographic, topographic, and climatic ancillary information. Pixel longitude and latitude were directly extracted from MODIS metadata. Pixel altitude was obtained using the SRTM 30 m/pixel digital elevation model [71]. MODIS pixel slopes were calculated using GEE slope calculation algorithm on the same 30-m elevation model. Finally, climatic data were downloaded from REDIAM's environmental information [72], including potential evapotranspiration, precipitation, mean annual temperature, mean of the maximum temperatures, and mean of the minimum temperatures. All types of ancillary data came in different resolutions or scale, so to match the resolution of our MODIS pixels we computed the average across all finer resolution pixels inside each MODIS pixels to obtain the value at 460-m resolution.

3) Pixels' LULC Abundances Annotation From SIPNA: To annotate each 460-m MODIS pixel with the abundance of each LULC class, the official LULC map of Andalusia for the year 2013 (SIPNA) [14] was used. Given the coarse resolution of MODIS pixels, we only considered Level 1 (four classes) and an adapted version of Level 2 (ten classes) of the classification hierarchy of SIPNA (see Fig. 5).

Given that SIPNA provides information at subpixel level, we calculated the exact abundances of all the LULC classes existing in each MODIS 460-m resolution pixel, as illustrated in Fig. 6, using QGIS software [75] as follows: the SIPNA polygons were first converted to raster format providing a LULC map at 10-m resolution. The rasterized map was then converted to match the spatial resolution of MODIS by counting the number of 10-m resolution pixels of each LULC class and dividing them by the total number of 10-m resolution pixels inside each 460-m resolution, resulting in the class proportions for each 460-m pixel of Andalusia. Finally, the MODIS pixels abundances annotations were coupled with their corresponding time-series and ancillary data to obtain the Andalusia-MSMTU dataset. With the help of several RS expert, we visually assess that the 10-m resolution was suitable for the rasterization. The proposed values were 100, 50, 10, and 5 m. The 100- and 50-m resolution pixels were too coarse to maintain the quality of the different polygon annotations. The 10- and 5-m resolution pixels were great to maintain the information and we finally decided to rasterize the

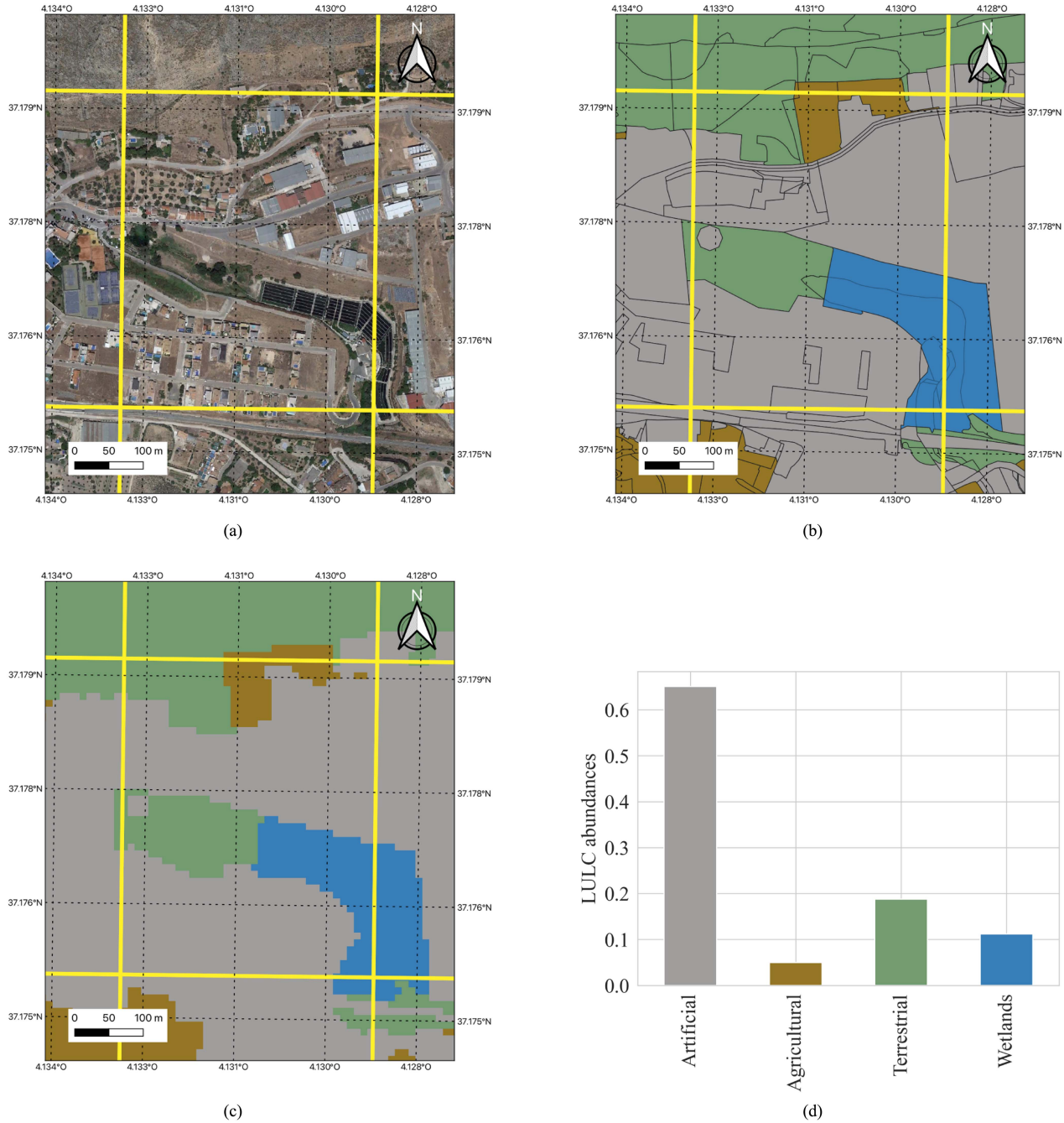

Fig. 7. Example of how class abundances are obtained for each MODIS pixel. (a) Satellite image of the Google Satellite corresponding to one MODIS pixel. (b) Annotated SIPNA polygons. (c) Rasterized LULC map at 10-m resolution. (d) Obtained abundance of Level-1 classes for that MODIS pixel. (a) Google satellite image of one MODIS pixel. (b) SIPNA LULC map (polygon format). (c) SIPNA LULC map (10-m raster format). (d) Level 1 LULC class abundances.

polygons to 10-m resolution because of computational and time convenience, since the 5-m raster was four times more expensive than the 10-m raster.

In Fig. 7, an example of the calculation of class abundances for a given pixel is shown. An illustrative example of the distribution of abundances of LULC classes in Level 1 of the classification hierarchy over the Andalusia territory is displayed in Fig. 8, being "agricultural lands" and "terrestrial lands," the classes that dominate the Andalusian territory. Andalusia-MSMTU dataset [76] is available in a public data repository hosted by Zenodo at: https://zenodo.org/records/7752348

V. Methodology

Given the success of DL methods in learning tasks, we propose an RNN-based model to learn the input–output relationship between the remotely sensed MS time-series and the corresponding LULC abundances in a pixel. We do this by building a

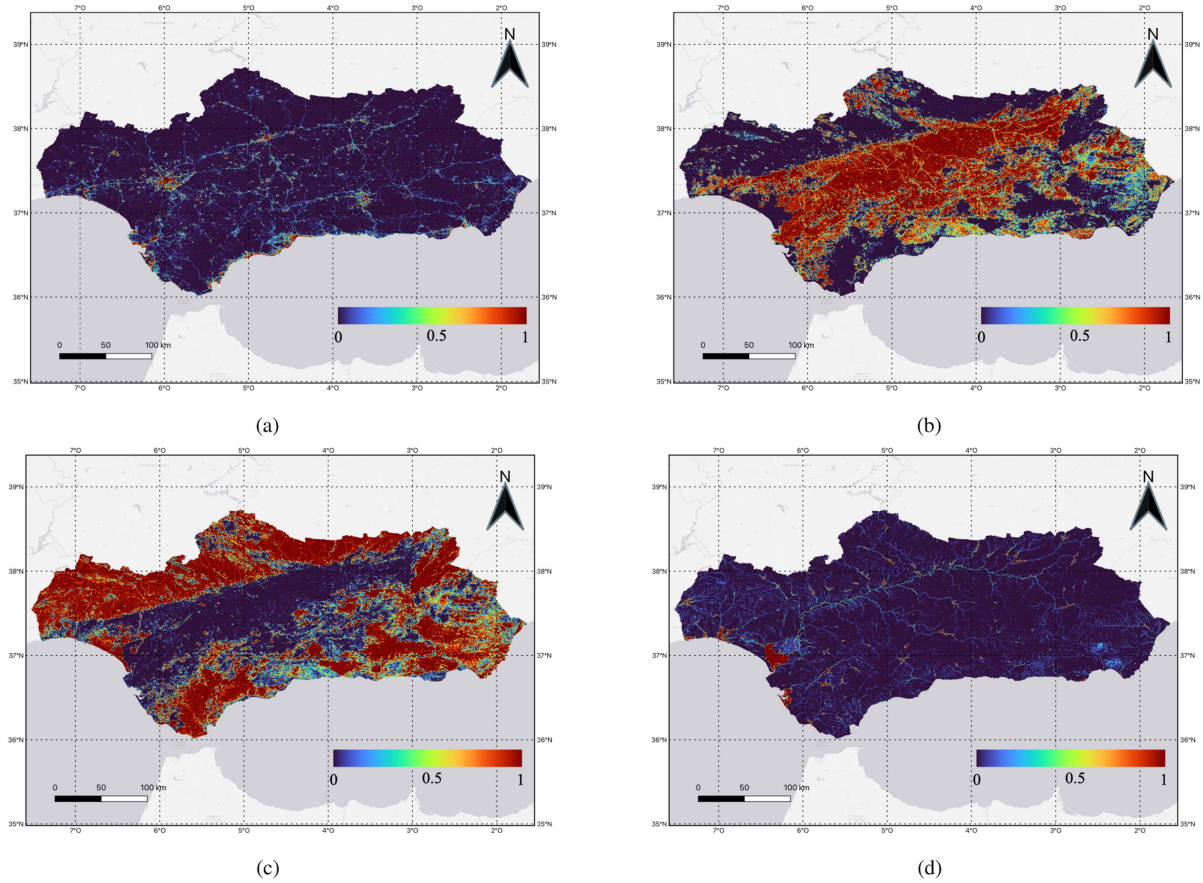

Fig. 8. Gradient map of abundances from Andalusia 460-m pixels for each LULC class in Level 1 of SIPNA. (a) Artificial, (b) Agricultural lands, (c) Terrestrial lands, and (d) Wetlands. Blue pixels represent low abundance and red pixels represent high abundance.

high-quality annotated dataset as described in Section IV-B and splitting it in training and test sets. We use the training dataset to train the DL model in a supervised manner. The test set is composed of unseen samples during the training and is used to evaluate the true performance of the model.

Formally, we have a set of n MS time-series pixels $\{X_1, X_2, \ldots, X_n\}$ with their corresponding class abundances $\{y_1, y_2, \ldots, y_n\}$ where $y_i \in S^C, i \in [1, n]$. S^C is the sample space of class abundances commonly referred to as the simplex [77]. In our case, C is equal to 4 and 10 for levels 1 and 2 of the hierarchy, respectively.

To enhance class abundance estimation further, in addition to using the MS multitemporal data, we also include ancillary information from the following two types.

1) *Geo-topographic data:* Geographical coordinates (longitude and latitude), altitude and slope. Incorporating geographic coordinates can help the model understand the spatial distribution of land cover types, which can be valuable in guiding the SU process and making it more contextually accurate. Similarly, adding topographic data (altitude and slope) provides useful information that complements the spectral characteristics of a pixel. In fact, terrain slope is known to influence surface reflectance, so incorporating it into the model can allow slope-related changes in reflectance to be taken into account, making its predictions more robust.
2) *Climatic data:* Precipitation, potential evapotranspiration, mean temperature, maximum temperature, and minimum temperature. Some land cover classes, such as agricultural lands, forests, and wetlands, respond differently to variations in climate. By using climatic variables, the DL model can distinguish between these climate-dependent classes more effectively.

In the following, we describe the architecture of the used model and the evaluation metrics.

A. Model Architecture

Our BRITS-based approach to estimate the class abundances using MS multitemporal data and ancillary information for each mixed pixel is depicted in Fig. 9. The proposed approach includes the following three components.

1) *Spectro-temporal feature extraction:* We use BRITS model [70] to extract the spectro-temporal patterns in the presence of missing values from our dataset.
2) *Ancillary data feature extraction:* To incorporate ancillary information, we process the external information using a linear layer with ReLU nonlinearity.
3) *Concatenation and features combination:* The output features of part (1) and (2) are concatenated and processed by a final dense layer that outputs C (the number of classes) scores.

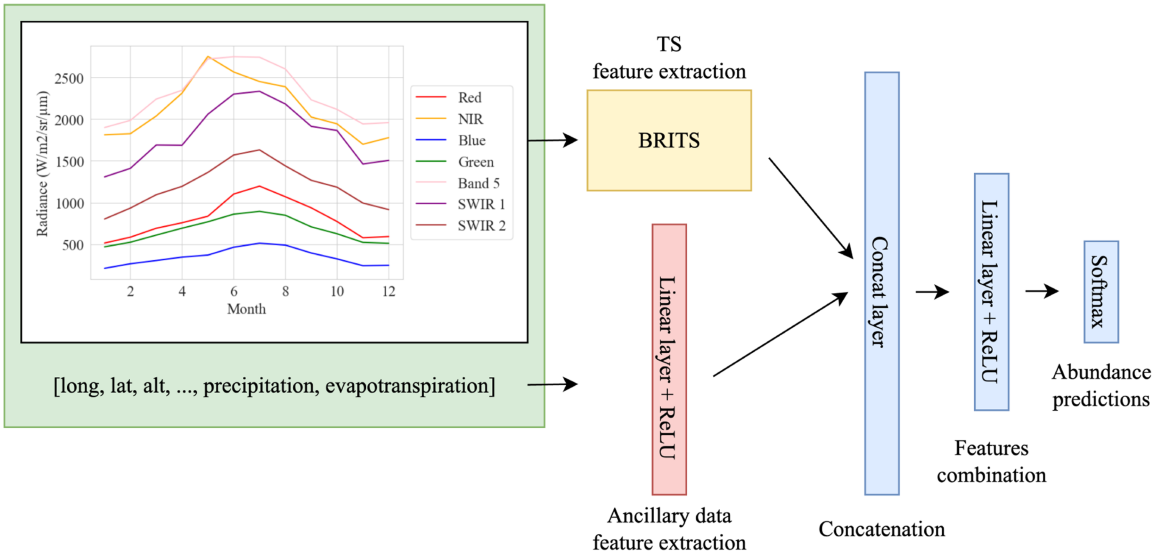

Fig. 9. Our proposed NN. The green box denotes the input data for a given pixel, i.e., MS time-series data + ancillary data. The yellow box denotes the BRITS model for MS time-series feature extraction, the red box denotes the ancillary data feature extraction layer and the blue boxes denote the final layers for features combination and softmax transformation of NN's output.

The final dense layer generates an unbounded outputs $\mathbf{o}$ where $\mathbf{o} \in \mathbb{R}^C$ with C being the number of classes. Following the work of [78], we applied the softmax transformation to obtain the final abundances predictions $\mathbf{a} \in S^C$ as

$$a_j = \frac{e^{o_j}}{\sum_{c=1}^{C} e^{o_c}} \tag{3}$$

where a_j denotes the abundance prediction for the jth class, o_j denotes the final layer's output associated with the jth class, and e denotes the exponential function.

Finally, the NN is optimized by minimizing the mean-square error (MSE) between the abundances predictions and the reference abundances:

$$\text{MSE} = \frac{\sum_{i=1}^{N}\sum_{c=1}^{C}(r_{ic} - a_{ic})^2}{N} \tag{4}$$

where r_{ic} and a_{ic} are the reference abundance and the predicted abundance, respectively, for the cth class in the ith sample, and N is the number of training samples.

B. Evaluation Criteria

To assess the effectiveness of the proposed unmixing model, the following four regression metrics are examined.

1) Pearson's correlation coefficient (CC):

$$\text{CC} = \frac{\sum_{i=1}^{N}(r_i - \overline{r})(a_i - \overline{a})}{\sqrt{\sum_{i=1}^{N}(r_i - \overline{r})^2 \sum_{i=1}^{N}(a_i - \overline{a})^2}}. \tag{5}$$

2) Root mean squared error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{N}(r_i - a_i)^2}. \tag{6}$$

3) Relative root mean squared error (RRMSE):

$$\text{RRMSE} = \sqrt{\frac{\sum_{i=1}^{N}(r_i - a_i)^2}{\sum_{i=1}^{N}(r_i)^2}} \tag{7}$$

4) Mean absolute error (MAE):

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{N}|r_i - a_i| \tag{8}$$

where r_i is the reference abundance, a_i the predicted abundance, and $\overline{r}$ and $\overline{a}$ are the mean of both variables. Finally, we also considered F1-score to evaluate how good is the model in predicting the majoritarian class in each mixed pixel as follows:

$$F1 = \frac{2 * \text{TP}}{2 * \text{TP} + \text{FP} + \text{FN}}. \tag{9}$$

C. Experimental Design

To analyze the effect of introducing ancillary data and using different levels of the LULC legend on the performance of our DL approach for spectro-temporal unmixing, we considered different input data combinations, that is, using the following:

1) only the MS time-series data;
2) time-series plus geo-topographic ancillary data;
3) time-series plus climatic ancillary data;
4) time-series plus geo-topographic and climatic ancillary data.

In order to avoid spatial autocorrelation of neighboring pixels, we used a block train test splitting [79], [80]. First, we divided the entire Andalusian territory in areas of equal size using blocks of 18×15 km, which means that each block contains 1250 of 460 m pixels approximately. Subsequently, 80% of the pixel blocks were assigned randomly to the training set, with the remaining 20% allocated to the test set. Fig. 10 illustrates the areas of

TABLE II
PERFORMANCE COMPARISON OF OUR MODEL TRAINED FROM SCRATCH (FIRST ROW) AND FINETUNED FROM TIMESPEC4LULC (SECOND ROW) USING ONLY MS MULTITEMPORAL INPUT DATA, BY ADDING GEO-TOPOGRAPHIC DATA ONLY (THIRD ROW), BY ADDING CLIMATIC DATA ONLY (FOURTH ROW), AND BY ADDING BOTH GEO-TOPOGRAPHIC AND CLIMATIC (FIFTH ROW)

Ancillary input data	MAE (%)	RMSE (%)	RRMSE (%)	CC	F1-score	RMSE (%) gain	CC gain	MFLOPs
None (from scratch)	7.25%	12.82%	43.73%	0.8447	0.8055	0.0000%	0.0000	2.3055
None (finetuned from TimeSpec4LULC)	6.70%	12.28%	42.40%	0.8570	0.8140	-0.5400%	0.0123	2.3055
Climatic	6.55%	12.24%	42.34%	0.8582	0.8107	-0.5800%	0.0243	2.3665
Geo-topographic	6.21%	11.80%	40.84%	0.8691	0.8195	-1.0200%	0.0244	2.3663
Geo-topographic + climatic	**6.15%**	**11.65%**	**40.34%**	**0.8723**	**0.8241**	**-1.1700%**	**0.0276**	2.3673

The performance is expressed in terms of average mean absolute error (MAE), root mean squared error (RMSE), relative root mean squared error (RRMSE), correlation coefficient (CC), F1-Score, RMSE gain, and CC gain with respect to baseline model for SIPNA Level 1 classes. Last column, "MFLOPs," indicates the model's computational complexity in terms of mega FLOPs.

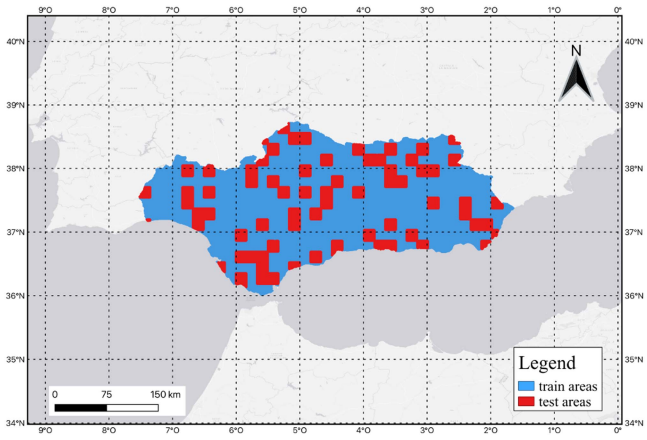

Fig. 10. Train (blue) and test (red) areas.

pixels designated for the training and testing sets. The source code to run these experiments will be available after acceptance at https://github.com/jrodriguezortega/MSMTU.

1) Implementation Details: Our models undergo training using the Adam optimizer [81] for a total of 200 epochs with a batch size of 2048. We initialize the learning rate at 0.003 and progressively reduce it via the cosine learning rate decay scheduler. All experiments were conducted utilizing the PyTorch DL framework [82].

2) TimeSpec4LULC [74] Pretraining: TimeSpec4LULC is an open-source dataset comprising MS time-series data for 29 LULC classes, designed for training machine learning models. This dataset is constructed using the seven spectral bands from MODIS sensors, providing data at a 460-m resolution, spanning the time period from 2000 to 2021. We found that pretraining the BRITS model on TimeSpec4LULC dataset and fine tuning it on Andalusia-MSMTU provides better results that training it from scratch, mainly because of the similarity between both datasets.

VI. EXPERIMENTAL RESULTS

This section provides the experimental results of the proposed model at SIPNA Level 1 and Level 2.

A. SIPNA Level 1

We evaluated the proposed model in Section V on different combinations of spectro-temporal data and ancillary data. In particular, we considered these combinations: (spectro-temporal data), (spectro-temporal data + geo-topographic data), (spectro-temporal data + climatic data), and (spectro-temporal data + geo-topographic and climatic data). Besides, we also include the results of a baseline model trained from scratch on spectro-temporal data only to show how the pretraining on TimeSpec4LULC dataset is highly beneficial. The results of these five models in terms of the average MAE, RMSE, RRMSE, CC, F1-score, RRMSE gain, and CC gain across the four classes of Level 1 are provided in Table II. In addition, the computational complexity of each model is expressed in terms of MFLOPs in the last column. First, we can see in the first two rows that by just pretraining our model in TimeSpec4LULC dataset improves the results in every metric, proving the value of pretraining DL models in similar tasks to achieve better performance. Second, it can be seen that including ancillary information always improves the SU performance with respect to the baseline model (using MS time series only and trained from scratch). The highest performance is achieved when including both, geo-topographic and climatic data together with the MS time series showing the lowest MAE, RMSE, and RRMSE, with 1.10%, 1.17%, and 3.39% of improvement, respectively, and highest CC and F1-score with 0.0276 and 0.0216 of improvement, respectively, with respect to the baseline model.

A further analysis of the five metrics for each class is depicted in Fig. 11. In general, including the geo-topographic and climatic information improves the abundance predictions of all the classes of Level 1. The "terrestrial lands" and "agricultural lands" achieve better performance in terms of CC, F1-score, and RRMSE. However, the classes that benefit the most from adding the ancillary data are "artificial" and "wetlands" since the relative improvement is greater in these classes.

It is worth noting that the RMSE and MAE metrics are not fair for comparisons between classes as they do not take into account the range of abundance values within each class. The most appropriate metric for these comparisons in this case is the RRMSE.

To better illustrate the reasons behind these differences in performance between observed and predicted abundances in each

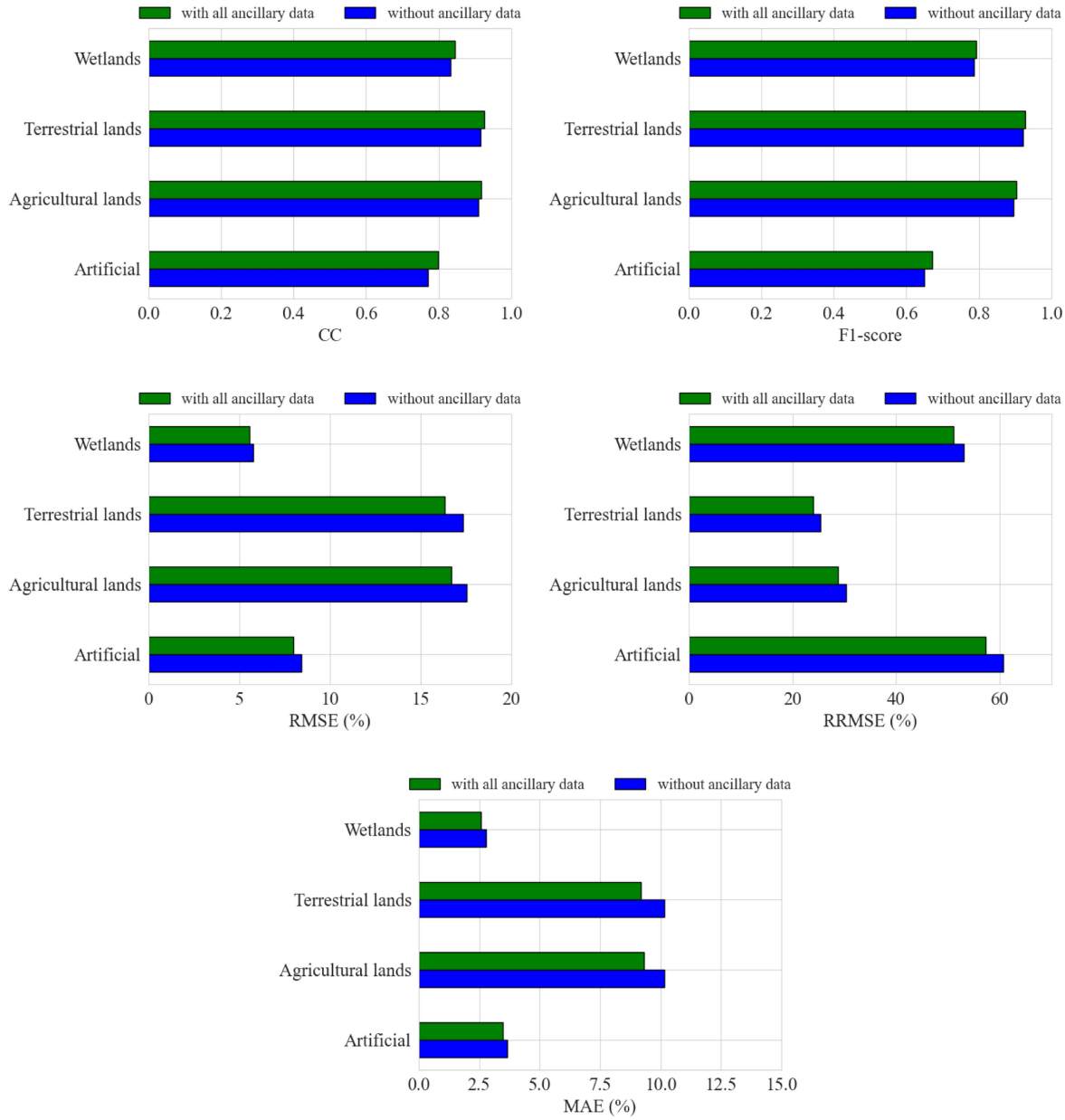

Fig. 11. Test results for the four SIPNA Level 1 classes obtained by including all ancillary information (green) and without ancillary information (blue): CC values (top left), F1-score values (top right), RMSE values (middle left), RRMSE values (middle right), and MAE values (bottom).

of the four LULC classes, Fig. 12 shows a density scatter plot for each class. The scatter plots of "artificial" and "wetlands" pixels showed a less aligned distribution along the 1:1 straight line than terrestrial and agricultural lands. In artificial and wetlands plots, most points are concentrated in the lowest abundances, while in terrestrial and agricultural lands, points tend to concentrate in both the extremes of the abundance gradient but also along the 1:1 line. This proofs that the model works reasonably good for both abundant (terrestrial and agricultural lands) and scarce (artificial and wetlands) classes.

Finally, Fig. 13 shows the results achieved by the best model on three test areas (top row) with their corresponding RMSE (middle row) and RRMSE (bottom row) per pixel maps. As we can observe, most of the pixels are in blue tones, which indicates a low RMSE and RRMSE and a great LULC abundances predictions. A reduced number of pixels with red tones in the RRMSE maps indicates an important prediction error relative to the scale of the reference abundance. These pixels mainly correspond to small heterogeneous rural areas with a large diversity of urban, crop and even forest areas, which makes the task of correctly predicting each and every LULC class abundances very difficult.

B. *SIPNA Level 2*

Similarly, we evaluated the proposed model on these input combinations: (spectro-temporal data), (spectro-temporal data + geo-topographic data), (spectro-temporal data + climatic data),

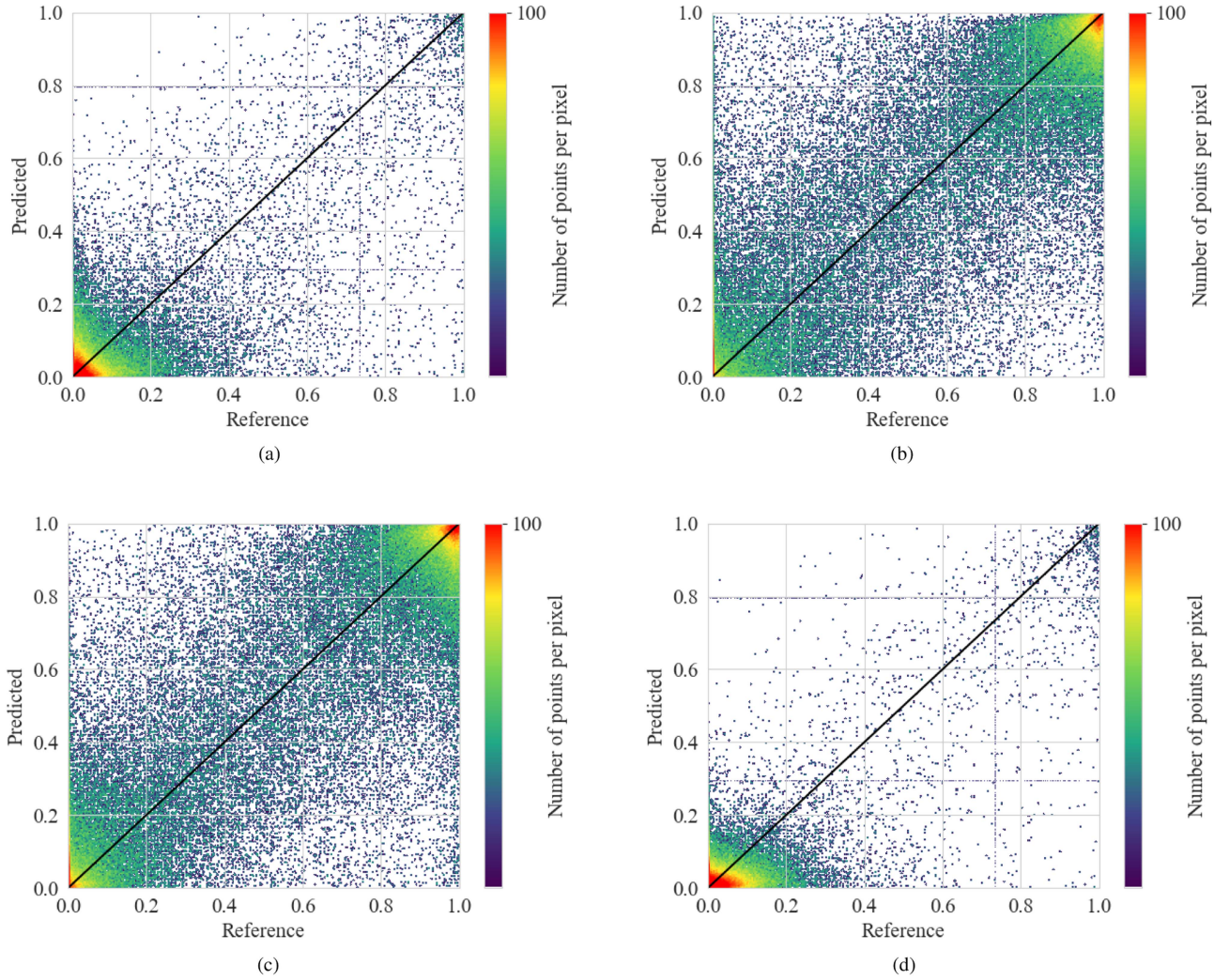

Fig. 12. Density scatter plots of every Level 1 class abundances (predicted versus reference) for the best model (including all ancillary data). (a) Artificial. (b) Agricultural lands. (c) Terrestrial lands. (d) Wetlands.

TABLE III
PERFORMANCE COMPARISON OF OUR MODEL TRAINED FROM SCRATCH (FIRST ROW) AND FINETUNED FROM TIMESPEC4LULC (SECOND ROW) USING ONLY MS MULTITEMPORAL INPUT DATA, BY ADDING GEO-TOPOGRAPHIC DATA ONLY (THIRD ROW), BY ADDING CLIMATIC DATA ONLY (FOURTH ROW), AND BY ADDING BOTH GEO-TOPOGRAPHIC AND CLIMATIC (FIFTH ROW)

Ancillary data added	MAE (%)	RMSE (%)	RRMSE (%)	CC	F1-score	RMSE (%) gain	CC gain	MFLOPs
None (from scratch)	6.13%	11.91%	60.23%	0.7372	0.6072	0.0000%	0.0000	2.3055
None (finetuned from TimeSpec4LULC)	5.88%	11.54%	59.11%	0.7504	0.6216	-0.3700%	0.0132	2.3055
Climatic	5.73%	11.47%	58.57%	0.7566	0.6317	-0.4400%	0.0233	2.3665
Geo-topographic	5.60%	11.31%	57.92%	0.7635	0.6396	-0.6000%	0.0312	2.3663
Geo-topographic + climatic	**5.57%**	**11.26%**	**57.43%**	**0.7664**	**0.6404**	**-0.6500%**	**0.0320**	2.3673

The performance is expressed in terms of average mean absolute error (MAE), root mean squared error (RMSE), relative root mean squared error (RRMSE), correlation coefficient (CC), F1-Score, RMSE gain, and CC gain with respect to baseline model for SIPNA Level 2 classes. Last column, "MFLOPs," indicates the model's computational complexity in terms of mega FLOPs.

and (spectro-temporal data + geo-topographic and climatic data) considering SIPNA Level 2. We also include the results of a baseline model trained from scratch on spectro-temporal data only to show how the pretraining on TimeSpec4LULC dataset is beneficial for Level 2 as well. The results of these five models in terms of the average MAE, RMSE, RRMSE, CC, F1-score, RRMSE gain, and CC gain across the ten classes of Level 2 are provided in Table III. In addition, the computational complexity of each model is expressed in terms of MFLOPs in the last column. Again, we can see in the first two rows that by just pretraining our model in TimeSpec4LULC dataset improves the results in every metric, proving the value of pretraining DL

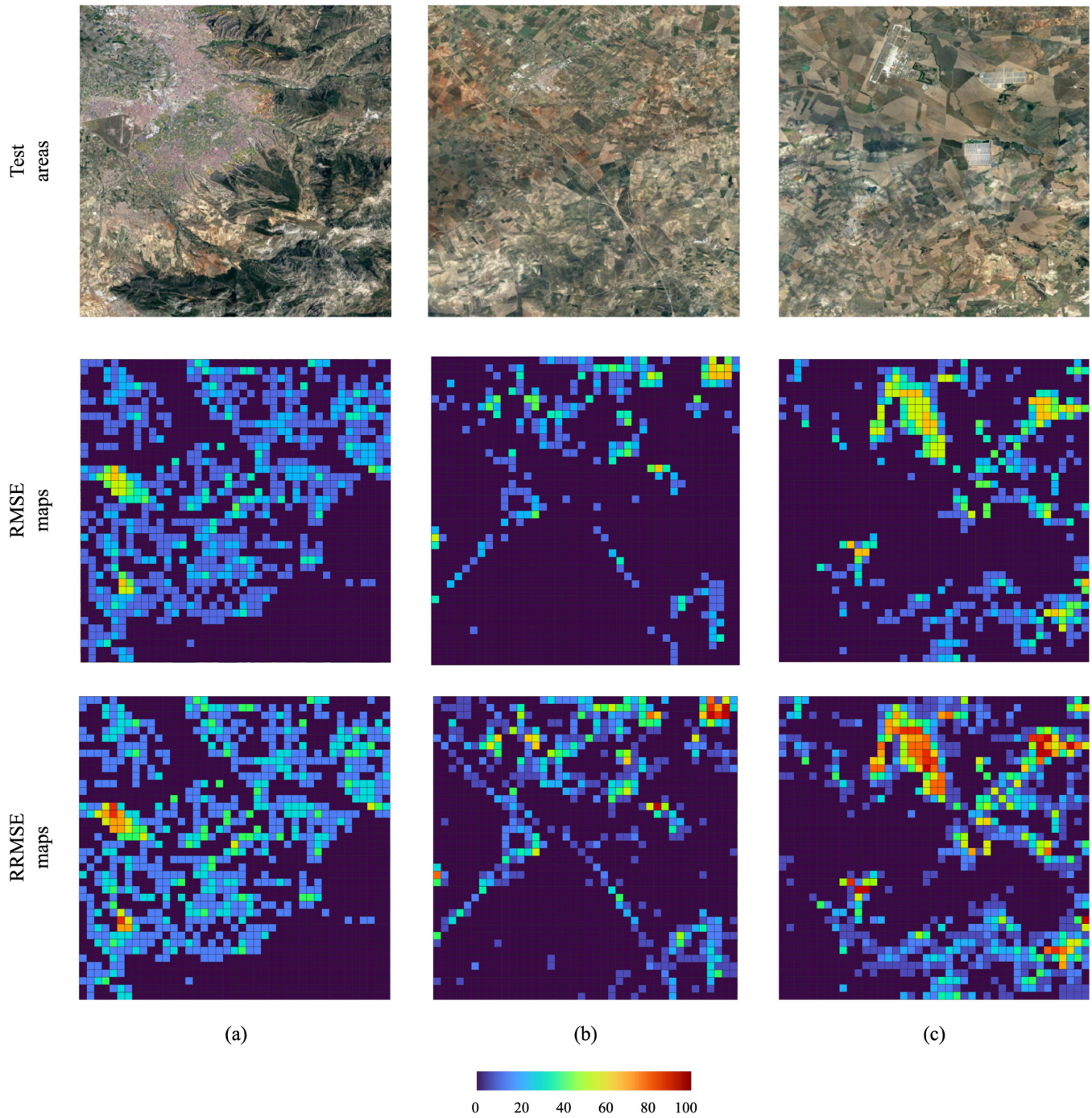

Fig. 13. Three test areas (top row) with their corresponding RMSE (middle row) and RRMSE (bottom row) per pixel maps in Level 1. (a) Granada, Granada. (b) La Carlota, Córdoba. (c) El Coronil, Sevilla.

models in similar tasks to achieve better performance. Similarly, including ancillary information improves the SU task even in a much more difficult SU setting (see Table III). Compared to the baseline, the best performing model (including all the ancillary data) decreases the MAE, RMSE, and RRMSE by 0.56%, 0.65%, and 2.80%, respectively, and increases CC and F1-score by 0.0320 and 0.0332, respectively.

In the same way as in Level 1, Fig. 14 shows a comparison between the baseline model and the model including geo-topographic and climatic data for every class in each of the five metrics used for evaluation. In general, adding ancillary information improves the abundances predictions of all the classes. The best performance is achieved in "woody crops" and "annual crops" classes in terms of CC, F1-score, and RRMSE. Besides, adding the ancillary information to the model achieves a greater improvement for the classes with the worst results like "combinations of croplands and vegetation," "barelands," and "artificial."

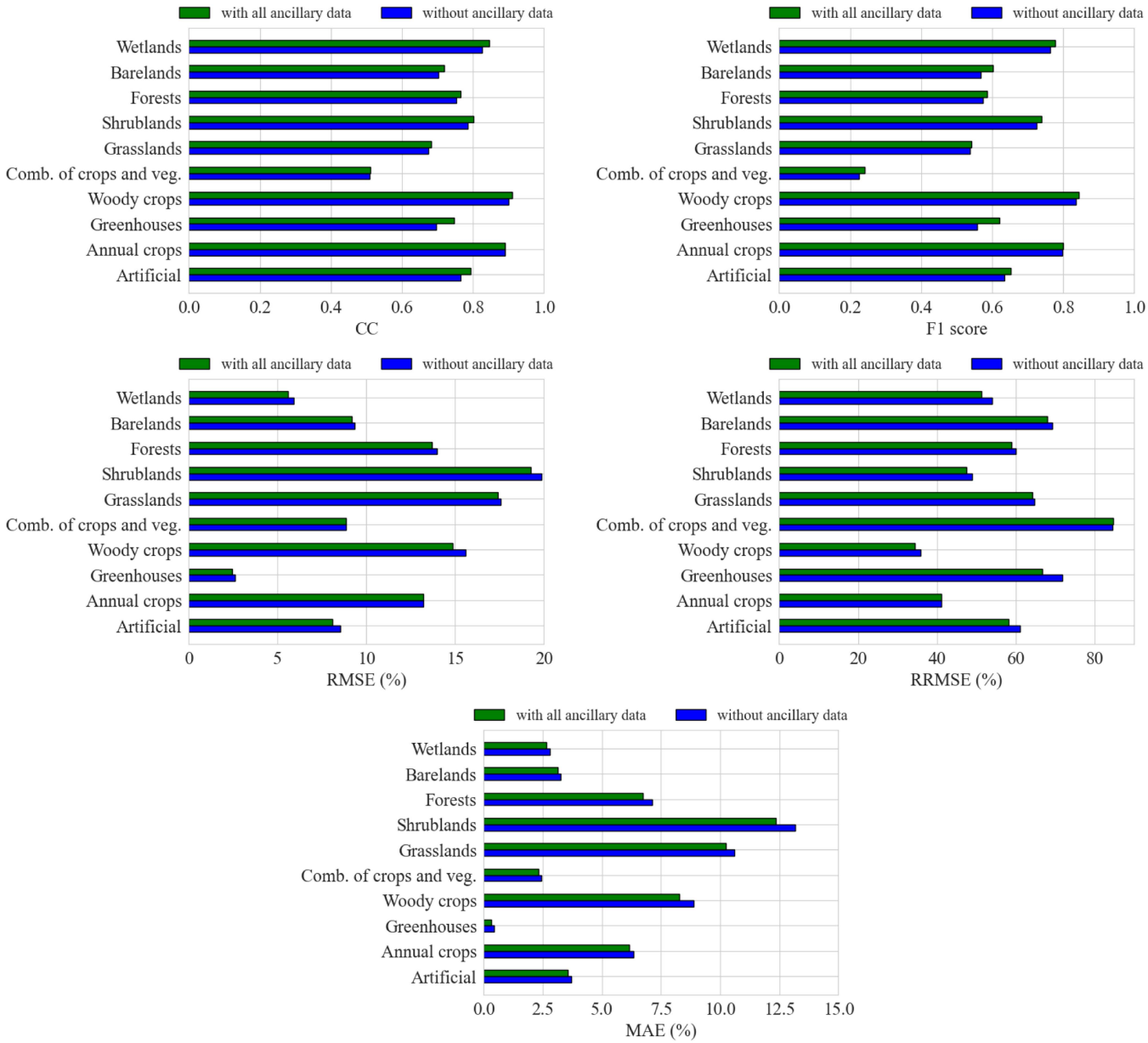

Fig. 14. Test results for the ten SIPNA Level 2 classes obtained by including all ancillary information (green) and without ancillary information (blue). CC values (top left), F1-score values (top right), RMSE values (middle left), RRMSE values (middle right), and MAE values (bottom).

Looking at the density scatter plot for each Level 2 class in Fig. 15, we see that the correlation between the reference and the predicted abundances is generally good, except for "combinations of croplands and vegetation" and "barelands" classes since they show a large dispersion. It is worth emphasizing the strong performance of the model for the class "greenhouses." Despite having so few representations of middle-range values of abundance in the pixels of Andalusia, the correlation in this class between the reference and predicted abundances is similar to the classes with a good representation. We argue that the reason for that could be due to their very high albedo, i.e., high reflectance in all bands. Finally, the worst performance metrics were obtained for "combinations of croplands and vegetation" class, which may be due to the mixed nature of this class definition itself. By combining crop and vegetation, this class is a mixture of some of the other classes, and hence, it is complicated for the model to predict the correct abundances.

Finally, Fig. 16 shows the results achieved by the best model on three test areas (top row) with their corresponding RMSE (middle row) and RRMSE (bottom row) per pixel maps. In general, most pixels are in dark blue tones (low error) in the RMSE maps, which at first glance may seem better than the results achieved for Level 1. However, when looking at the RRMSE maps, we can notice a slightly higher number of pixels with red tones than in the Level 1 RRMSE maps, located mainly in heterogeneous rural areas. Given that in Level 2, we have 12 LULC classes, there are more heterogeneous pixels, and consequently, the unmixing task is harder. It is important to note that although the error at Level 1 is lower in absolute terms, when it is relativized by the scale of the reference abundances it

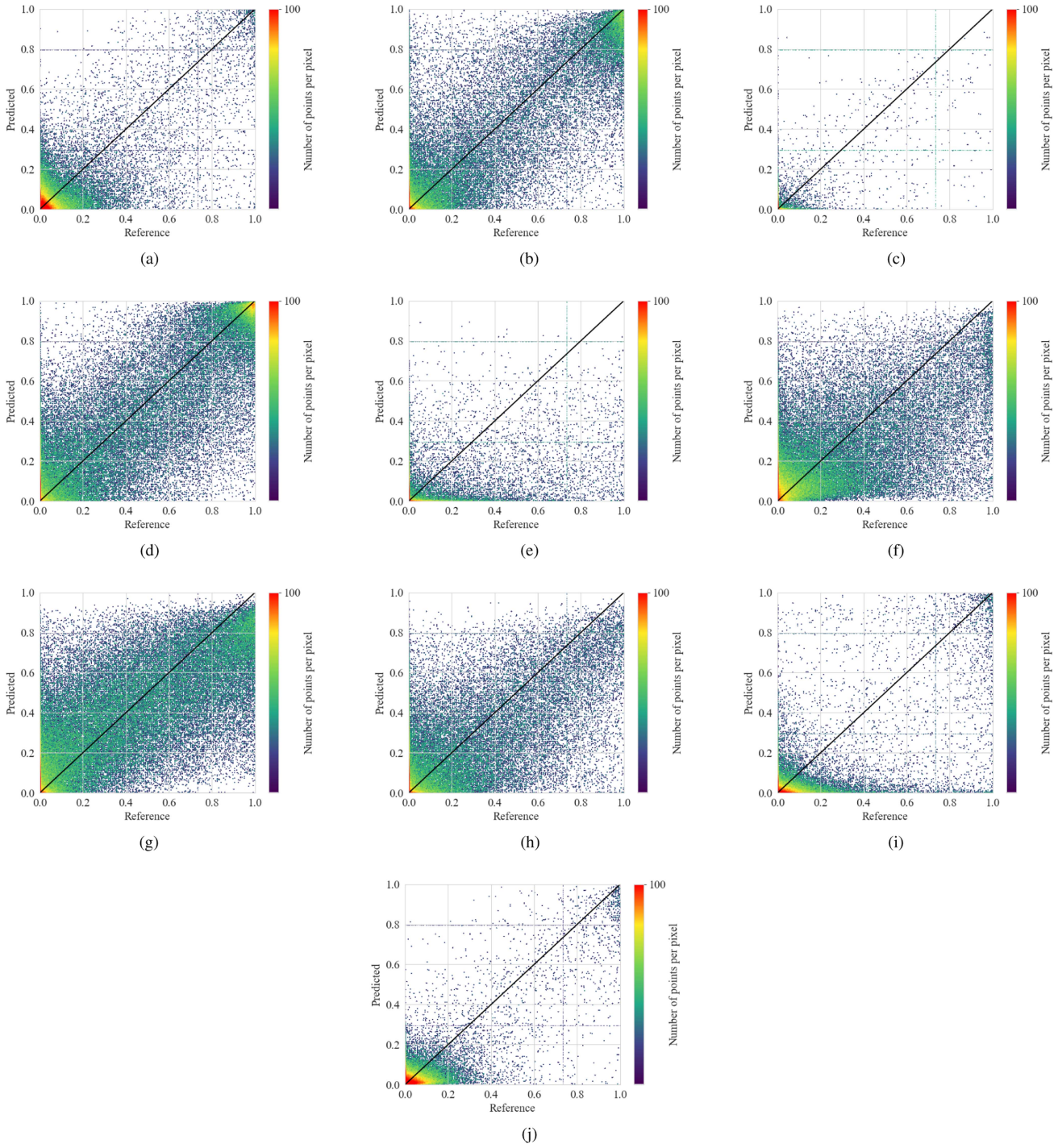

Fig. 15. Density scatter plots of every Level 2 class abundance (predicted versus reference) for the best model (including all ancillary data). (a) Artificial. (b) Annual croplands. (c) Greenhouses. (d) Woody croplands. (e) Combinations of croplands and vegetation. (f) Grasslands and grasslands with trees. (g) Shrublands and shrublands with trees. (h) Forest. (i) Barelands. (j) Wetlands.

becomes higher than in Level 1, indicating only moderate results compared to the good results obtained at Level 1. For this reason, it is recommended to evaluate not only the RMSE but also the RRMSE to get better conclusions.

VII. Discussions

SU of LULC classes is a challenging problem commonly addressed by physical models with the need for endmember extraction [3], [4]. Moreover, the variability naturally present in the spectral signature for a given LULC class (spectral variability) makes this problem even more difficult [83], [84]. DL methods represent a great solution to eliminate the need for endmember extraction and they are known for their robustness against noise given sufficient amounts of training data. Most previous works focus on HS or MS data and do not exploit temporal information to estimate the abundance of mixed pixels. Obtaining MS time series of large territories can be prohibitive

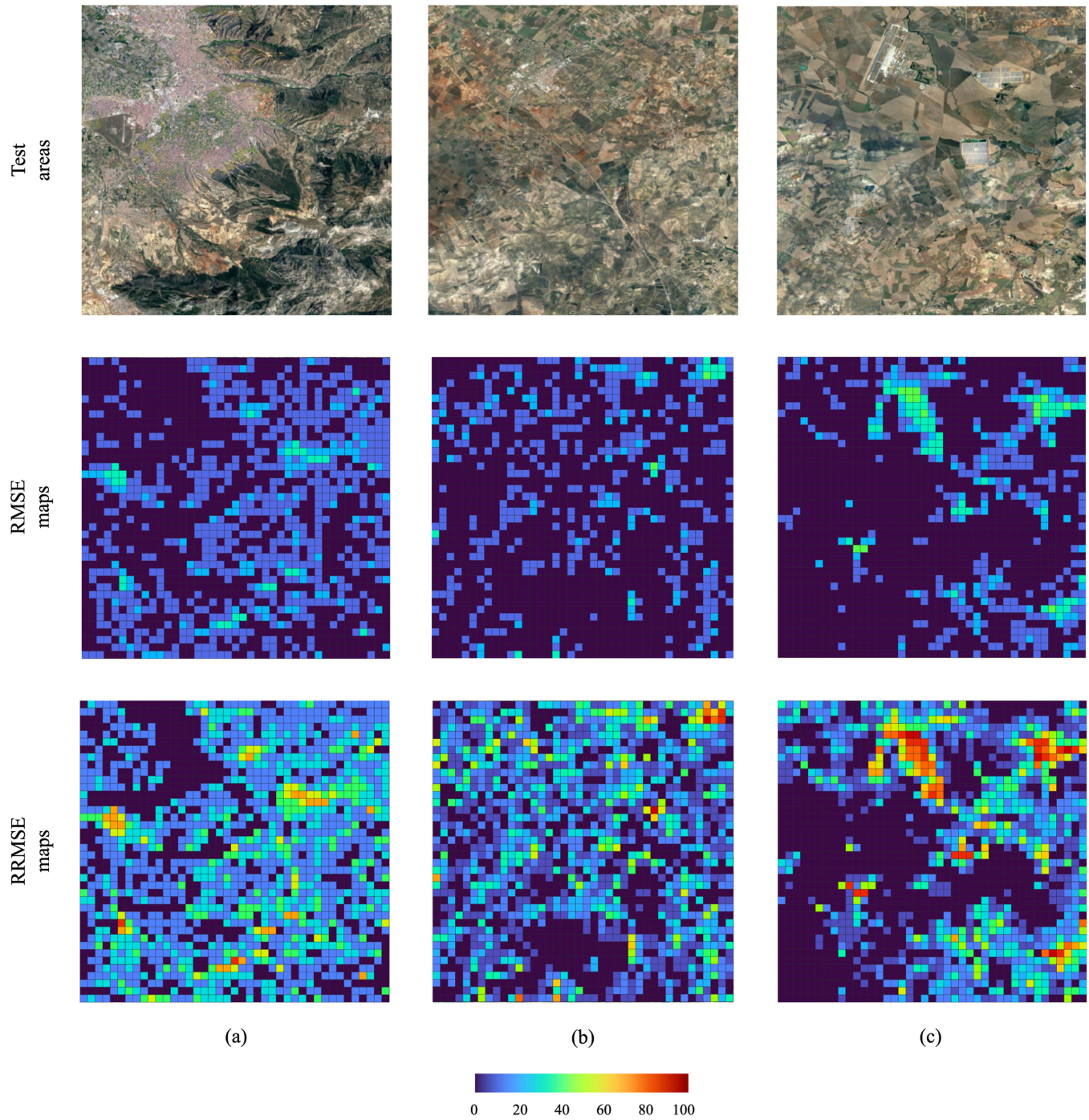

Fig. 16. Three test areas (top row) with their corresponding RMSE (middle row) and RRMSE (bottom row) per pixel maps in Level 2. (a) Granada, Granada. (b) La Carlota, Córdoba. (c) El Coronil, Sevilla.

due to the cost and time required to acquire them [85], [86]. In addition, no previous work has explored the possibility of adding ancillary data to enhance the SU results, which are used successfully in other computer vision tasks [11], [12], [13].

In this article, we tried to solve the mentioned constraints of previous works by the following.

1) Developing Andalusia-MSMTU, a high-quality MS time-series dataset of mixed pixels labeled with LULC class abundances at two classification levels and making it publicly available so other researchers can develop new approaches in the field of SU. This dataset followed several data preprocessing steps as explained in Section IV-B1 in order to smooth spectral variabilities associated with the imaging process.
2) Proposing and analyzing DL-based approaches for SU without the need of endmembers extraction. Moreover, we intentionally included pixels well distributed around our study area in models' training, which implies a high

number of diverse pixels with different spectral variations. This way, the DL models will be robust against the spectral variations of pixels in the test areas.

Our results showed that our DL-based method achieved good results for SU of LULC classes by using MS data and it can be used in areas with similar features such as the rest of Spain and mediterranean countries. Besides, by including ancillary information, the model improved in terms of every metric used for evaluation, showing that adding external data is an interesting avenue to explore in the SU problems.

Finally, one significant limitation still exists in our work. Although DL models have shown great performance in mapping complex input–output relationships and have demonstrated promising results for SU of LULC classes, they lack physical interpretability. This means that it is difficult to understand how the model arrived at its decision, and it may not be clear why certain input features were given more weight than others [87]. In the context of SU, physical interpretation may be desirable [88] because it allows us to understand the underlying physical processes that govern the interaction of electromagnetic radiation with land surface materials.

VIII. Conclusion

In this article, we introduced and made publicly available Andalusia-MSMTU dataset, a new DL-ready dataset to explore SU approaches on MS time-series data. Furthermore, we introduced ancillary information to improve the SU performance consisting on two geographic, two topographic, and five climatic variables. Our experiments show that the use of MS time-series data for LULC abundance estimation achieves good results, which are further improved by including ancillary information.

For future work, we would like to explore taking advantage of spatial autocorrelation between neighboring pixels, which provides useful information for the SU task [89], by arranging the MODIS pixels in images and using a convolutional-LTSM network with a BRITS-like approach to deal with missing values. Moreover, given the recent availability of higher spatial resolution sensors like Sentinel-2, data fusion between MODIS long-term data and Sentinel-2 higher resolution data is another avenue to improve SU performance. Finally, since common DL-based models lack physical interpretation and it is sometimes important in the context of SU, an effort to make DL-based methods physically aware is worthwhile.

References

[1] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*. New York, NY, USA: Elsevier, 2006.

[2] X. Zhang, Y. Sun, J. Zhang, P. Wu, and L. Jiao, "Hyperspectral unmixing via deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 11, pp. 1755–1759, Nov. 2018.

[3] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[4] J. S. Bhatt and M. V. Joshi, "Deep learning in hyperspectral unmixing: A review," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 2189–2192.

[5] Q. Wang, X. Ding, X. Tong, and P. M. Atkinson, "Spatio-temporal spectral unmixing of time-series images," *Remote Sens. Environ.*, vol. 259, 2021, Art. no. 112407.

[6] G. A. Licciardi and F. Del Frate, "Pixel unmixing in hyperspectral data by means of neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4163–4172, Nov. 2011.

[7] Q. Wang, X. Ding, X. Tong, and P. M. Atkinson, "Real-time spatiotemporal spectral unmixing of modis images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, Sep. 1, 2022.

[8] B. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Blind hyperspectral unmixing using autoencoders: A critical comparison," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1340–1372, Jan. 1, 2022.

[9] P. Ghosh, S. K. Roy, B. Koirala, B. Rasti, and P. Scheunders, "Hyperspectral unmixing using transformer network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, Aug. 1, 2022.

[10] J. F. Mustard, L. Li, and G. He, "Nonlinear spectral mixture modeling of lunar multispectral data: Implications for lateral transport," *J. Geophysical Res., Planets*, vol. 103, no. E8, pp. 19419–19425, 1998.

[11] T. Berg, J. Liu, S. Woo Lee, M. L. Alexander, D. W. Jacobs, and P. N. Belhumeur, "Birdsnap: Large-scale fine-grained visual categorization of birds," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2011–2018.

[12] J. S. Ellen, C. A. Graff, and M. D. Ohman, "Improving plankton image classification using context metadata," *Limnol. Oceanogr.: Methods*, vol. 17, no. 8, pp. 439–461, 2019.

[13] G. Tseng, H. Kerner, and D. Rolnick, "TIML: Task-informed meta-learning for agriculture," 2022, *arXiv:2202.02124*.

[14] SIOSE, 2023. Accessed: Jun., 2023. [Online]. Available: https://www.juntadeandalucia.es/medioambiente/portal/landing-page-%C3%ADndice/-/asset_publisher/zX2ouZa4r1Rf/content/sistema-de-informaci-c3-b3n-sobre-el-patrimonio-natural-de-andaluc-c3-ada-sipna-/20151

[15] Q. Yuan et al., "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, 2020, Art. no. 111716.

[16] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.

[17] R. Kemker, C. Salvaggio, and C. Kanan, "Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 60–77, 2018.

[18] H. Guan, Y. Yu, Z. Ji, J. Li, and Q. Zhang, "Deep learning-based tree classification using mobile LiDAR data," *Remote Sens. Lett.*, vol. 6, no. 11, pp. 864–873, 2015.

[19] Q. Lv et al., "Urban land use and land cover classification using remotely sensed SAR data through deep belief networks," *J. Sensors*, vol. 2015, 2015.

[20] 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.

[21] Z. Han, D. Hong, L. Gao, J. Yao, B. Zhang, and J. Chanussot, "Multimodal hyperspectral unmixing: Insights from attention networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, Mar. 1, 2022.

[22] T. Uezato, N. Yokoya, and W. He, "Illumination invariant hyperspectral image unmixing based on a digital surface model," *IEEE Trans. Image Process.*, vol. 29, pp. 3652–3664, Jan. 1, 2020.

[23] D. Hong et al., "Cross-city matters: A multimodal remote sensing benchmark dataset for cross-city semantic segmentation using high-resolution domain adaptation networks," *Remote Sens. Environ.*, vol. 299, 2023, Art. no. 113856.

[24] D. Hong et al., "SpectralGPT: Spectral foundation model," 2023, *arXiv:2311.07113*.

[25] D. Hong, J. Yao, C. Li, D. Meng, N. Yokoya, and J. Chanussot, "Decoupled-and-coupled networks: Self-supervised hyperspectral image super-resolution with subpixel fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–12, Oct. 1, 2023.

[26] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, Jun. 2014.

[27] D. C. Heinz and Chein-I-Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.

[28] C. Quintano, A. Fernández-Manso, Y. E. Shimabukuro, and G. Pereira, "Spectral unmixing," *Int. J. Remote Sens.*, vol. 33, no. 17, pp. 5307–5340, 2012.

[29] J. M. Bioucas-Dias et al., "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[30] O. Ghaffari, M. J. Valadan Zoej, and M. Mokhtarzade, "Reducing the effect of the endmembers' spectral variability by selecting the optimal spectral bands," *Remote Sens.*, vol. 9, no. 9, 2017, Art. no. 884.

[31] J. S. Bhatt, M. V. Joshi, and S. Vijayashekhar, "A multitemporal linear spectral unmixing: An iterative approach accounting for abundance variations," in *Proc. 9th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, 2018, pp. 1–5.

[32] C. Zhang, L. Ma, J. Chen, Y. Rao, Y. Zhou, and X. Chen, "Assessing the impact of endmember variability on linear spectral mixture analysis (LSMA): A theoretical and simulation analysis," *Remote Sens. Environ.*, vol. 235, 2019, Art. no. 111471.

[33] D. A. Roberts, M. Gardner, R. Church, S. Ustin, G. Scheer, and R. Green, "Mapping chaparral in the Santa Monica mountains using multiple endmember spectral mixture models," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 267–279, 1998.

[34] C. A. Bateson, G. P. Asner, and C. A. Wessman, "Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 1083–1094, Mar. 2000.

[35] W. Li and C. Wu, "A geostatistical temporal mixture analysis approach to address endmember variability for estimating regional impervious surface distributions," *GIScience Remote Sens.*, vol. 53, no. 1, pp. 102–121, 2016.

[36] J. Degerickx, D. A. Roberts, and B. Somers, "Enhancing the performance of multiple endmember spectral mixture analysis (MESMA) for urban land cover mapping using airborne LiDAR data and band selection," *Remote Sens. Environ.*, vol. 221, pp. 260–273, 2019.

[37] J. Wang and C.-I. Chang, "Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2601–2616, Sep. 2006.

[38] S. Moussaoui et al., "On the decomposition of mars hyperspectral data by ICA and Bayesian positive source separation," *Neurocomputing*, vol. 71, no. 10-12, pp. 2194–2208, 2008.

[39] W. Xia, X. Liu, B. Wang, and L. Zhang, "Independent component analysis for blind unmixing of hyperspectral imagery with additional constraints," *IEEE Trans. Geosci. remote Sens.*, vol. 49, no. 6, pp. 2165–2179, Jun. 2011.

[40] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, Mar. 2007.

[41] Z. Yang, G. Zhou, S. Xie, S. Ding, J.-M. Yang, and J. Zhang, "Blind spectral unmixing based on sparse nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 1112–1125, Apr. 2011.

[42] F. Zhu, Y. Wang, S. Xiang, B. Fan, and C. Pan, "Structured sparse method for hyperspectral unmixing," *ISPRS J. Photogrammetry remote Sens.*, vol. 88, pp. 101–118, 2014.

[43] X.-R. Feng, H.-C. Li, R. Wang, Q. Du, X. Jia, and A. Plaza, "Hyperspectral unmixing based on nonnegative matrix factorization: A comprehensive review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4414–4436, May 1, 2022.

[44] Y. Zhong, X. Wang, L. Zhao, R. Feng, L. Zhang, and Y. Xu, "Blind spectral unmixing based on sparse component analysis for hyperspectral remote sensing imagery," *ISPRS J. Photogrammetry remote Sens.*, vol. 119, pp. 49–63, 2016.

[45] V. S. S. and J. S. Bhatt, "A blind spectral unmixing in wavelet domain," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10287–10302, Sep. 1, 2021.

[46] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4153–4162, Nov. 2011.

[47] Y. Altmann, N. Dobigeon, J.-Y. Tourneret, and S. McLaughlin, "Nonlinear unmixing of hyperspectral images using radial basis functions and orthogonal least squares," in *Prroc. IEEE Int. Geosci. remote Sens. Symp.*, 2011, pp. 1151–1154.

[48] J. Broadwater, R. Chellappa, A. Banerjee, and P. Burlina, "Kernel fully constrained least squares abundance estimates," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2007, pp. 4041–4044.

[49] L. Gao, Z. Wang, L. Zhuang, H. Yu, B. Zhang, and J. Chanussot, "Using low-rank representation of abundance maps and nonnegative tensor factorization for hyperspectral nonlinear unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, Mar. 1, 2022.

[50] G. M. Foody, "Relating the land-cover composition of mixed pixels to artificial neural network classification output," *Photogrammetric Eng. Remote Sens.*, vol. 62, no. 5, pp. 491–498, 1996.

[51] P. Atkinson, M. Cutler, and H. Lewis, "Mapping sub-pixel proportional land cover with AVHRR imagery," *Int. J. Remote Sens.*, vol. 18, no. 4, pp. 917–935, 1997.

[52] Y. Yu, Y. Ma, X. Mei, F. Fan, J. Huang, and H. Li, "Multi-stage convolutional autoencoder network for hyperspectral unmixing," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 113, 2022, Art. no. 102981.

[53] L. Gao, Z. Han, D. Hong, B. Zhang, and J. Chanussot, "CyCU-Net: Cycle-consistency unmixing network by learning cascaded autoencoders," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, Mar. 1, 2022.

[54] J. Yao, D. Hong, L. Xu, D. Meng, J. Chanussot, and Z. Xu, "Sparsity-enhanced convolutional decomposition: A novel tensor-based paradigm for blind hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, Apr. 1, 2022.

[55] Z. Han, D. Hong, L. Gao, B. Zhang, and J. Chanussot, "Deep half-siamese networks for hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 11, pp. 1996–2000, Nov. 2021.

[56] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017.

[57] C. Li, R. Cai, and J. Yu, "An attention-based 3 D convolutional autoencoder for few-shot hyperspectral unmixing and classification," *Remote Sens.*, vol. 15, no. 2, 2023, Art. no. 451.

[58] V. S. S., V. S. Deshpande, and J. S. Bhatt, "A practical approach for hyperspectral unmixing using deep learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Nov. 1, 2022, Art. no. 5511505.

[59] Z. Han, D. Hong, L. Gao, B. Zhang, M. Huang, and J. Chanussot, "AutoNAS: Automatic neural architecture search for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, Jun. 1, 2022.

[60] M. Zhao, L. Yan, and J. Chen, "LSTM-DNN based autoencoder network for nonlinear hyperspectral image unmixing," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 2, pp. 295–309, Feb. 2021.

[61] D. B. Lobell and G. P. Asner, "Cropland distributions from temporal unmixing of modis data," *Remote Sens. Environ.*, vol. 93, no. 3, pp. 412–422, 2004.

[62] R. Zurita-Milla, L. Gómez-Chova, L. Guanter, J. G. Clevers, and G. Camps-Valls, "Multitemporal unmixing of medium-spatial-resolution satellite images: A case study using MERIS images for land-cover mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4308–4317, Nov. 2011.

[63] F. Yang, B. Matsushita, T. Fukushima, and W. Yang, "Temporal mixture analysis for estimating impervious surface area from multi-temporal modis NDVI data in Japan," *ISPRS J. Photogrammetry Remote Sens.*, vol. 72, pp. 90–98, 2012.

[64] C. Deng and Z. Zhu, "Continuous subpixel monitoring of urban impervious surface using landsat time series," *Remote Sens. Environ.*, vol. 238, 2020, Art. no. 110929.

[65] E. L. Bullock, C. E. Woodcock, and P. Olofsson, "Monitoring tropical forest degradation using spectral unmixing and landsat time series analysis," *Remote Sens. Environ.*, vol. 238, 2020, Art. no. 110968.

[66] F. Zhu, "Hyperspectral unmixing: Ground truth labeling, datasets, benchmark performances and survey," 2017, *arXiv:1708.05125*.

[67] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[68] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[69] F. Gerber, R. de Jong, M. E. Schaepman, G. Schaepman-Strub, and R. Furrer, "Predicting missing values in spatio-temporal remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2841–2853, May 2018.

[70] W. Cao, D. Wang, J. Li, H. Zhou, L. Li, and Y. Li, "BRITS: Bidirectional recurrent imputation for time series," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018.

[71] T. G. Farr et al., "The shuttle radar topography mission," *Rev. Geophys.*, vol. 45, no. 2, 2007.

[72] *Red de información ambiental de andalucía*, 2023. Accessed: Jul. 2023. [Online.] Available: https://www.juntadeandalucia.es/medioambiente/portal/acceso-rediam

[73] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18–27, 2017, doi: 10.1016/j.rse.2017.06.031.

[74] R. Khaldi et al., "TimeSpec4LULC: A global multispectral time series database for training LULC mapping models with machine learning," *Earth System Sci. Data*, vol. 14, no. 3, pp. 1377–1411, 2022.

[75] QGIS Development Team, *QGIS Geographic Inf. Syst., QGIS Assoc.*, 2022. [Online]. Available: https://www.qgis.org
[76] J. Rodríguez-Ortega, R. Khaldi, D. Alcaraz-Segura, and S. Tabik, "Andalusia-MSMTU," Zenodo, Mar. 20, 2023, doi: 10.5281/zenodo.7752348.
[77] J. Aitchison, "The statistical analysis of compositional data," *J. Roy. Statist. Soc., Ser. B.*, vol. 44, no. 2, pp. 139–160, 1982.
[78] S. Heremans, J. A. Suykens, and J. Van Orshoven, "The effect of imposing 'fractional abundance constraints' onto the multilayer perceptron for subpixel land cover classification," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 44, pp. 226–238, 2016.
[79] D. R. Roberts et al., "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure," *Ecogr.*, vol. 40, no. 8, pp. 913–929, 2017.
[80] L. Uieda, "Verde: Processing and gridding spatial data using green's functions," *J. Open Source Softw.*, vol. 3, no. 29, 2018, Art. no. 957.
[81] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[82] A. Paszke et al., "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019.
[83] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.
[84] L. Ren, D. Hong, L. Gao, X. Sun, M. Huang, and J. Chanussot, "Orthogonal subspace unmixing to address spectral variability for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–13, Jan. 1, 2023.
[85] J. Cai, H. Chatoux, C. Boust, and A. Mansouri, "Extending the unmixing methods to multispectral images," in *Proc. Color Imag. Conf.*, 2021, pp. 311–316.
[86] J. Nalepa, "Recent advances in multi-and hyperspectral image analysis," *Sensors*, vol. 21, no. 18, 2021, Art. no. 6002.
[87] A. B. Arrieta et al., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, 2020.
[88] G. Camps-Valls et al., "Physics-aware machine learning for geosciences and remote sensing," in *Proc. IEEE Int. Geosci. remote Sens. Symp.*, 2021, pp. 2086–2089.
[89] C. Shi and L. Wang, "Incorporating spatial information in spectral unmixing: A review," *Remote Sens. Environ.*, vol. 149, pp. 70–87, 2014.

José Rodríguez-Ortega received the Computer Engineering degree in 2020 and the M.Sc. degree in data science in 2022, both from the University of Granada, Granada, Spain, where he is currently working toward the Ph.D. degree in computer science.

He is a Data Science and Artificial Intelligence Assistant with LifeWatch-ERIC, Seville, Spain. His research interests include the design of artificial intelligence models and remote sensing data processing methods for biodiversity conservation and global climate change.

Rohaifa Khaldi received the M.Sc. degree in mechanical engineering from the ENSEM College of Engineering, Hassan II University of Casablanca, Casablanca, Morocco, in 2015, and the Ph.D. degree in applied mathematics from the ENSIAS College of Engineering, Mohammed V University of Rabat, Rabat, Morocco, in 2021.

She is currently an AI Assistant and a Team Coordinator with the ICT-Core Office of the Research Infrastructure, LifeWatch-ERIC, Seville, Spain. She is also an AI Researcher with the Andalusian Research Institute in Artificial Intelligence (DaSCI), Granada, Spain. Her research interests include artificial intelligence, statistical machine learning, deep learning, computer vision, remote sensing, time-series forecasting, Big Data analysis, and data mining.

Domingo Alcaraz-Segura received the B.Sc. degree in environmental sciences and the Ph.D. degree in environmental sciences from the University of Almería, Almería, Spain, in 2000 and 2005, respectively.

He is currently a Professor with the Department of Botany, University of Granada, Granada, Spain. He is also an Associate Researcher with the Andalusian Center for the Assessment and Monitoring of Global Change, Almería, Spain. His research is based on fieldwork, remote sensing techniques, time-series analysis, and geographical information systems. His current research interests include environmental controls of biodiversity, the impact of land cover and land use changes on ecosystem functioning and services and on hydroclimate, and the development of monitoring and alert systems of global change effects on protected areas.

Siham Tabik received the B.Sc. degree in physics from the University Mohammed V, Rabat, Morocco, in 1998, and the Ph.D. degree in computer science from the University of Almería, Almería, Spain.

She is currently an Associate Professor with the Department of Computer Science and Artificial intelligence, University of Granada, Granada, Spain. Her research interests include the design of artificial intelligence models and remote sensing data processing methods for biodiversity conservation and global climate change.