Time-Resolved Sentinel-3 Vegetation Indices via Inter-Sensor 3D Convolutional Regression Networks

Ruben Fernandez-Beltran, *Senior Member, IEEE,* Damian Ibañez, Jian Kang, *Member, IEEE,* and Filiberto Pla

Abstract

Sentinel missions provide widespread opportunities of exploiting inter-sensor synergies to improve the operational monitoring of terrestrial photosynthetic activity and canopy structural variations using vegetation indices (VI). In this context, continuous and consistent temporal data are logically required to rapidly detect vegetation changes across sensors. Nonetheless, the existing temporal limitations inherent to satellite orbits, cloud occlusions, data degradation and many other factors may severely constrain the availability of data involving multiple satellites. In response, this paper proposes a novel deep 3D convolutional regression network (3CRN) for temporally enhancing Sentinel-3 VI by taking advantage of inter-sensor Sentinel-2 observations. Unlike existing regression and deep learning-based methods, the proposed approach allows convolutional kernels to slide across the temporal dimension in order to exploit not only the higher spatial resolution of the Sentinel-2 instrument but also its own temporal evolution to better estimate time-resolved VI in Sentinel-3. To validate the proposed approach, we built a database made of multiple day-synchronized Sentinel-2 and Sentinel-3 operational products from a study area in Extremadura (Spain). The conducted experimental comparison, including multiple state-of-the-art regression and deep learning models, shows the statistically significant advantages of the presented framework. The codes of this work will be made available at https://github.com/rufernan/3CRN.

R. Fernandez-Beltran, D. Ibañez and F. Pla are with the Institute of New Imaging Technologies, University Jaume I, E-12071 Castellón de la Plana, Spain. (e-mail: rufernan@uji.es; ibanezd@uji.es; pla@uji.es).

J. Kang is with the School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China (e-mail: jiankang@suda.edu.cn).

Index Terms

Sentinel-3 (S3), Sentinel-2 (S2), vegetation indices (VI), temporal resolution, inter-sensor regression.

I. INTRODUCTION

The increasing demand for remote sensing (RS) data strongly encourages the expansion of different Earth observation (EO) missions and programs to cover current societal needs as well as future challenges [1]. In this sense, large investments are constantly made to guarantee the operational provision of high quality RS imagery by means of different artificial satellites and constellations that pursue to satisfy the technical requirements of numerous application domains [2]. Being one of the most important intergovernmental organisations, the European Spacial Agency (ESA) aims at addressing present and future EO needs by providing continuous remotely sensed data useful for scientific purposes and operational applications that benefit worldwide citizens. To achieve this goal, the European Commission in partnership with ESA coordinate and manage the so-called Copernicus programme and additional Earth Explorer missions that focus on different aspects of EO, such as, atmosphere, biosphere, hydrosphere or cryosphere.

Within the Copernicus programme context, Sentinel-2 (S2) and Sentinel-3 (S3) share one of the most important synergies since both families of satellites make use of multi-spectral instruments. In the case of S2 [3], it includes a couple of satellites (S2A and S2B) that contain the Multi-Spectral Instrument (MSI), which is able to capture the Earth surface with a 10 to 60 m spatial resolution using 13 spectral bands (B01-B12) from the 443 to 2190 nm wavelength range. In the case of S3 [4], this mission also comprises two satellites (S3A and S3B) that carry, among other sensors, the Ocean and Land Colour Instrument (OLCI), which is able to provide a finer spectral resolution based on 21 bands (Oa01-Oa21) from the 390 to 1040 nm wavelength region but using a coarser spatial resolution of 300 m. Despite the fact that both MSI and OLCI instruments are able to provide operational data related to vegetation, land and water, their fundamental spatial-spectral differences make each mission particularly suitable for specific applications [5]. Whereas S2 becomes more convenient for land cover characterization tasks where the spatial precision is important [6], OLCI's spectral features make S3 better at capturing global information from oceans, inland waterways and coastal areas [7]. Nonetheless, the broad scope of these Sentinel missions still motivates the development of additional Earth

Explorers focused on key scientific challenges and breakthrough technologies for advancing the understanding of Earth systems and their functional interactions.

Focusing on terrestrial vegetation, ESA is currently developing the Fluorescence Explorer (FLEX) [8] which has been designed to synergistically work with S3 mission. In this case, FLEX has a single satellite that will fly in tandem with S3 for remotely measuring the solar-induced chlorophyll fluorescence (SIF) emitted by plants as an accurate indicator of the actual photosynthetic activity [9]. Nonetheless, detecting SIF emissions from space is very challenging since the fluorescence signal is very weak with respect to the total acquired radiance. In response, FLEX will carry the fluorescence imaging spectrometer (FLORIS) which counts with a particularly ultra-fine spectral resolution (focused on the 500-780 nm wavelength range) and 300 m spatial resolution (likewise OLCI). Even though these specifications are convenient for decoupling SIF emissions from vegetation reflected light, they become rather limited to characterize other important aspects, like atmospheric features. Precisely, FLEX will orbit only few seconds before one of the S3 satellites in order to take advantage of the OLCI instrument for supporting FLORIS products and providing additional value [10].

Certainly, FLEX/S3 tandem mission will offer a great advancement in assessing vegetation dynamics while providing widespread opportunities of exploiting inter-sensor synergies within the Copernicus programme. In this regard, the growing development of vegetation products that involve multiple satellites exemplify this trend [11], [12]. Specifically, the majority of methods tackle this inter-sensor problem from a data fusion perspective where the temporal availability of data becomes a major challenge. Note that, when accounting for the accurate monitoring of terrestrial vegetation, even relatively small temporal deviations may produce relevant vegetation changes. Therefore, continuous and consistent inter-sensor data can be essential to effectively detect rapid changes in terrestrial photosynthetic activity and canopy structures across sensors. Within FLEX/Sentinel context, the particularly coarse temporal resolution of FLORIS (2 weeks) together with the open availability of Sentinel data highly motivate the development of time-resolved vegetation indices (VI) in S3 and the forthcoming FLEX which can serve as SIF proxies [13]. In other words, the temporal prediction of VI can play a fundamental role in actual FLEX/S3 exploitation environments for the analysis of the temporal dynamics of photosynthetic changes.

In the RS literature, it is possible to find different types of temporal prediction methods for VI [14]. Whereas traditional temporal replacement and temporal filter methods are able to provide positive results under rather constrained scenarios (e.g. homogeneous landscapes,

temporal stability, partially missing data...), the higher generalization capabilities of learning-based models make this paradigm more suitable for time-resolving operational RS data. For instance, Zeng *et al.* presented in [15] a linear regression-based framework for effectively reconstructing temporal information from the moderate resolution imaging spectroradiometer (MODIS). In [16], Zhao *et al.* showed the advantages of modeling the land surface temperature of MODIS via a random forest regression. Other authors opt for using deep learning techniques for dealing with the temporal reconstruction and product regression problems. It is the case of Zhang *et al.* who developed in [17] a novel convolutional neural network (CNN) which is able to reconstruct missing MODIS and Landsat 7 data by considering multi-source spatial, temporal or spectral information as input. Following a similar idea, Shao *et al.* proposed in [18] a novel generative adversarial CNN for reconstructing missing RS data from multiple data sources. In [19], Aptoula *et al.* also define a CNN-based regression architecture for effectively estimating chlorophyll-a concentration from S2 data.

Despite the good results achieved by these and other related models [20], [21], standard CNNs generally exhibit limitations on the temporal information exploitation since they are mainly focused on the spatial-spectral dimensions, being the temporal data often stacked inside the own spectra. With the evolution of deep learning technologies, additional models have been used for the temporal prediction of VI. In [22], Yu *et al.* developed a deep recurrent neural network (DRNN) which exploits spatio–temporal features to predict S2 and MODIS VI based on the long short-term memory (LSTM) architecture. Although the authors of this recent work concluded that recurrent models can be considered a state-of-the-art technology in VI temporal prediction, the inherent temporal limitations of operational FLEX/Sentinel data may constrain the resulting performance in actual production environments where discontinuous short-term data can be expected. Note that recurrent networks have shown to be particularly effective for modeling long-term temporal dependencies. However, the relatively recent availability in years of Sentinel data together with the expected short lifetime design of FLEX (3.5 years) motivate the investigation of other types of temporal models. In this scenario, this work presents a novel deep 3D convolutional regression network (3CRN) specially designed for temporally enhancing S3 VI from an inter-sensor perspective. That is, our objective consists of taking advantage of Sentinel synergies to generate time-resolved OLCI VI as SIF proxies from the MSI sensor. In this way, the temporal resolution of S3 can be improved to rapidly detect even small vegetation changes that can potentially be detected from S2 using also its substantially higher spatial resolution. To

achieve this goal, we define in Section II the newly proposed inter-sensor regression model based on 3D-CNNs, which jointly exploits S2 spatial-spectral features together with their own temporal evolution to better estimate time-resolved VI in OLCI. In Section III, we build a database made of multi-temporal OLCI and MSI data from a study area in Extremadura (Spain). Then, we conduct an extensive experimental comparison, including multiple state-of-the-art regression and CNN-based VI prediction models, to validate the proposed approach performance. Finally, Section IV provides the main conclusions of this work.

II. METHODOLOGY

The proposed CNN architecture aims at mapping multi-temporal S2 image products onto their corresponding S3 VI values for predicting inter-sensor vegetation data at unavailable timestamps in S3. Specifically, short sequences of S2 multi-spectral image patches around a target output time are considered as the input data whereas pixel-wise S3 VI estimates at the target timestamp correspond to the model output. To effectively exploit short-term time series data, the proposed architecture provides a novel inter-sensor regression network based on 3D-CNNs. In more details, 3D convolutions are defined over the spatio-spectral and temporal domain of S2 data, so that 4D kernels slide across height, width and time dimensions with the objective of finding temporal correlations between multi-spectral samples and allowing the extraction dynamic characteristics not only in the spatial domain, but also along the temporal dimension. Note that S3 VI, such as the OLCI Terrestrial Chlorophyll Index (OTCI), may change in time periods shorter than the own instrument temporal resolution. Thus, the multi-spectral evolution of S2 along relatively small time periods may also provide key information to generate better inter-sensor estimates.

Let $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ be a multi-temporal collection of S2 images sequentially acquired over a particular region of interest with a fix spatial-spectral size of $(x_1 \times x_2 \times B)$. Let $\mathcal{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$ be their synchronous inter-sensor S3 VI that cover the same area with a lower spatial resolution of $(y_1 \times y_2)$, representing R the corresponding scaling factor (i.e., $x_1 = y_1 R$ and $x_2 = y_2 R$). From each j-th pixel of the i-th temporal sample $\mathbf{y}_i^j \in \mathbb{R}$, it is possible to extract a multi-spectral S2 image patch $\mathbf{x}_i^j \in \mathbb{R}^{(P \times P \times B)}$ that includes the same Earth surface extent, being $P \geq R$. Additionally, a T multi-temporal window can also be considered centered at the i-th timestamp as $\mathbf{x}_{iT}^j \in \mathbb{R}^{(P \times P \times \times T \times B)}$. Let $\hat{\mathcal{Y}}$ identify the whole set of S3 pixel values in $\mathcal{Y}$ (i.e., $\mathbf{y}_i^j$) and $\hat{\mathcal{X}}$ their corresponding spatio-spectral and temporal volumes in $\mathcal{X}$ (i.e., $\mathbf{x}_{iT}^j$). In this scenario, the proposed model approximates a function $\mathcal{F} : \hat{\mathcal{X}} \rightarrow \hat{\mathcal{Y}}$ which takes as input multi-temporal
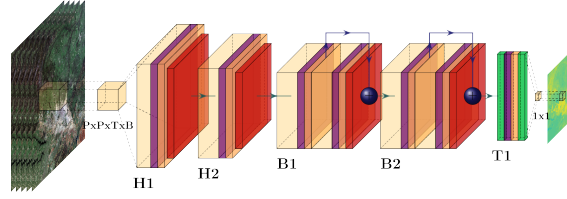
Fig. 1. Proposed architecture with two initial head blocks (H1 and H2), two body blocks (B1 and B2) and a final tail block (T1). In the diagram, 3D-CNN layers are shown in yellow, batch-normalization in purple, ReLU in orange, 3D pooling in red, residual connections in blue and dense layers in green.

S2 patches and produces as output their inter-sensor S3 VI values via 3D-CNN transformation blocks. Note that the considered 3D convolutions allow the kernel to slide across height, width and temporal dimensions. Hence, the whole S2 spectra is considered for uncovering spatio-temporal dynamics and generating S3 VI predictions. Since the network output is a pixel-wise prediction from limited temporal data, a sequential CNN structure will be chosen to efficiently address the inter-sensor mapping in contrast to other recurrent schemes which tend to produce data under-fitting when small number of time samples are considered [23]. Taking advantage of these components, the proposed regression architecture is defined according to the following data flow stages (Figure 1):

1) Head: The first part of the network consists of two building blocks (H1 and H2) which are made of four sequential layers: 1) 3D-CNN, 2) batch normalization (BN), 3) Rectified Linear Unit activation (ReLU), and 4) 3D pooling (3D-Pool). This configuration aims at processing the input images and generating an initial low-level characterization of S2 multi-spectral data. In this sense, the first head block (H1) extracts local features from the original S2 image domain to produce an over-complete characterization of elemental patterns. Then, the second head block (H2) increases the number of kernels to model a broader combination of patterns with the target of isolating the most relevant low-level features when predicting S3 VI values. In more details, the number of filters (K) is set to 32 in H1 and 64 in H2, respectively. Besides, all 3D-CNN layers are defined with a kernel size of $(3 \times 3 \times 3)$ together with a $(1 \times 1 \times 1)$ stride. Finally, H1 and H2 consider 3D-Pool layers with a $(2 \times 2 \times 1)$ pooling size and a $(1 \times 1 \times 1)$ stride.

2) Body: The second segment of the proposed architecture is also made of two twin blocks (B1 and B2) that have the following structure: 1) 3D-CNN, 2) BN, 3) ReLU, 4) 3D-CNN, 5) BN, 6) ReLU, 7) 3D-Pool and 8) residual connection ($\oplus$). This block design has the objective

of obtaining more complex and deeper spatio-temporal features from S2 data for progressively driving the uncovered deep embeddings towards pixel-wise S3 VI predictions. Note that the huge spatial resolution differences between S2 and S3 may generate important ill-posed effects during the back-propagation process since a large image patch in S2 corresponds to a single scalar value in S3. In order to relieve the degradation of the uncovered features after pooling the activation maps, the proposed body block includes a final residual connection. Besides, two 3D-CNN layers are considered before each 3D pooling operation with the goal of increasing the spatio-temporal receptive field within each block. Similarly to the network head, all 3D-CNN layers are defined with a $(3 \times 3 \times 3)$ kernel size and one-pixel stride, while considering $K = 64$. Additionally, 3D-Pool layers are defined with a $(2 \times 2 \times 1)$ pooling size, a $(1 \times 1 \times 1)$ stride and zero-padding for keeping the spatial dimension of the residual feature volume.

3) Tail: The last part of the network aims at projecting the extracted deep features onto the corresponding S3 VI values to achieve the required inter-sensor mapping. The following layers are considered within this building block: 1) a fully connected layer (FC), 2) BN, 3) RELU, and 4) FC. On the one hand, the first FC layer has a total of 256 fully connected neurons that globally associate all the spatio-temporal features uncovered before prediction. On the other hand, the last FC layer is connected to a single neuron which provides the estimated value of the target S3 pixel from the input S2 multi-temporal patch. Finally, the parameters of the proposed network are learnt by minimizing the mean squared error (MSE) reconstruction loss.

III. Experiments

A. Dataset

The dataset created for this work (https://github.com/rufernan/3CRN) consists of 21 coupled pairs of S2 MSI reflectance images and S3 OLCI vegetation products of the same day along the year 2019. In more details, we adopted the following three-step process: 1) study area selection, 2) operational data collection, and 3) final product generation. First (1), the Extremadura region was chosen for being the highest forest covered zone in Spain, with a 65% of forested area. The large amount of dehesas, an agrosilvopastoral system consisting of grassland and oak forests, makes this region highly relevant from a vegetation monitoring perspective, including different sorts of natural grasslands and agroforestry land. The exact location of the selected area is a 100 km^2 square bounded by (38.740370, -5.188652) and (39.697509, -6.498977) latitude-longitude coordinates, which corresponds to the T30STJ S2 tile. Second (2), daily synchronous S2 and S3

TABLE I

QUANTITATIVE ASSESSMENT FOR NDVI AND OTCI BASED ON RMSE ($\times 10^{-2}$) AND R^2 ($\times 10^2$) METRICS.

Methods	NDVI		OTCI		PSRI-NIR	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
LIN [15]	4.06±0.05	80.98±0.17	11.7±0.03	64.60±0.21	5.06±0.02	69.29±0.23
RFR [16]	3.23±0.04	87.94±0.18	11.5±0.05	65.57±0.32	3.43±0.04	85.92±0.34
MLP [20]	3.24±0.12	87.86±0.99	9.97±0.26	74.43±1.36	3.44±0.06	85.81±0.56
CNN2D[20]	3.11±0.15	88.83±1.13	9.57±0.12	76.47±0.61	3.48±0.07	85.53±0.64
CNNR [19]	3.07±0.24	89.08±1.90	9.84±0.09	75.12±0.46	3.47±0.06	85.59±0.51
DRNN [22]	2.88±0.04	90.42±0.30	8.80±0.16	80.10±0.70	3.27±0.08	87.15±0.64
CNN3D[20]	2.91±0.04	90.23±0.10	8.86±0.07	79.83±0.32	3.24±0.08	87.40±0.67
3CRN (ours)	**2.79±0.05**	**90.92±0.26**	**8.73±0.15**	**80.43±0.66**	**3.18±0.08**	**87.85±0.61**

TABLE II

STATISTICAL T-TEST WITH A SIGNIFICANCE LEVEL OF $\alpha = 0.05$.

Null hypothesis	p-value (NDVI)		p-value (OTCI)		p-value (PSRI-NIR)	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
CNN2D $\sim$ 3CRN	0.03727	0.04964	0.00062	0.00075	0.00526	0.00558
CNNR $\sim$ 3CRN	0.01000	0.00111	0.00002	0.00003	0.00174	0.00184
DRNN $\sim$ 3CRN	0.02275	0.02935	0.01434	0.01339	0.03295	0.03266
CNN3D $\sim$ 3CRN	0.04095	0.02482	0.00643	0.00801	0.04336	0.04643

data products from 2019 covering this region were selected and downloaded from the Copernicus Open Access Hub. To avoid cloud contamination, we initially filtered S2 products by considering only those images with less than a 1% of cloud coverage. Once the cloudless S2 products were obtained, a S3 product of the same day was selected for pairing a total of 21 coupled S2/S3 images, with a maximum inter-platform sensing difference time of an hour. Then, S2 images were downloaded as MSI Level-2 bottom-of-atmosphere products and S3 images as OLCI Level-1 top-of-atmosphere products. Third (3), the downloaded images were processed to generate the final S2 and S3 collections (i.e. $\mathcal{X}$ and $\mathcal{Y}$). On the one hand, S2 images were spatially re-sampled to 20 m in order to produce uniform data volumes of $(5490 \times 5490 \times 12)$. On the other hand, S3 images were rectified via the Rayleigh correction, projected and cropped to match the region of interest with a $(366 \times 366 \times 21)$ data size. Finally, Normalized Difference Vegetation Index (NDVI) and OLCI Terrestrial Chlorophyll Index (OTCI) were computed to build the target VI.

B. Experimental settings

To validate the performance of the proposed approach, we conduct multiple regression experiments based on the mapping between multi-temporal MSI patches and their associated OLCI VI values. Specifically, the experimental comparison includes the following models available in the RS literature: linear regression (LIN) [15], random forest regression (RFR) [16], multi-
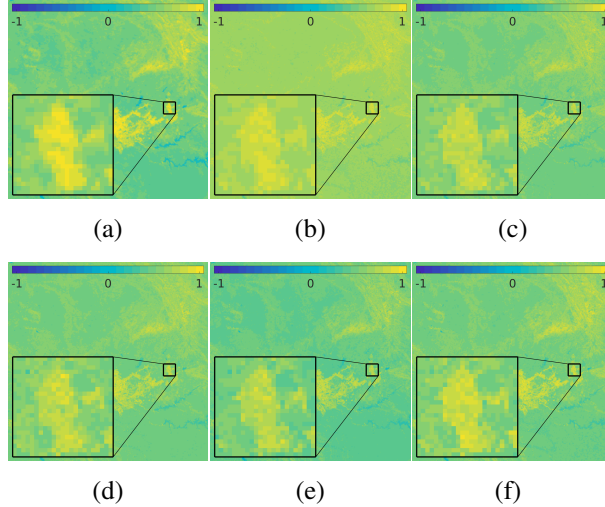
Fig. 2. NDVI qualitative assessment: (a) Ground-Truth, (b) CNN2D, (c) CNNR, (d) DRNN, (e) CNN3D and (f) 3CRN.

layer perceptron (MLP) [20], two-dimensional CNN (CNN2D) [20], CNN-based chlorophyll regression (CNNR) [19], deep recurrent neural networks (DRNN) [22] and three-dimensional CNN (CNN3D) [20]. All the methods use the configuration described in the corresponding papers, while considering analogous settings in terms of hidden units, filters, kernel sizes and activation functions. Regarding the considered data, they have also been trained and tested using the same data partitions extracted from the created collection. For computational reasons, S2 images were reduced to a $1/3$ ratio with respect to S3 ($R = 3$). Besides, a couple of S2/S3 products was excluded (from the available 21 image pairs) for generating a test qualitative output map. The rest of the data were divided into multi-temporal patches with size $P = 5$ and temporal window $T = 5$. Under this scheme, 5 different random partitions with balanced VI values were generated for cross-validating all the models, using a 60% of the data for training, another 20% for testing and a 20% for validation. All CNN-based models were trained using the ADAM optimizer with a 10^{-3} learning rate with plateau decay, a 128 batch size and 100 epochs. The experimentation has been carried out using Python 3.6, Keras and Scikit-learn on a Ubuntu 16.04 x64 machine with Intel(R) Core(TM) i7-6850K, NVIDIA GeForce GTX 1080Ti and 64 Gb of RAM.

C. Results

Table I presents the quantitative evaluation of the results based on three regression metrics: root mean square error (RMSE) and coefficient of determination (R^2). As it is possible to see,

the tested models are arranged in rows whereas the considered vegetation indices and metrics are presented in columns. Note that the displayed values correspond to the average and standard deviation results from 5 different data partitions. Besides, each metric is multiplied by a constant factor for a better visualization, i.e. $\times 10^2$ for RMSE and $\times 10^{-2}$ for R^2. The obtained results reveal that the proposed model (3CRN) is able to outperform every other considered method for all the analyzed settings. Among traditional regression methods, the models with the best performances are RFR and MLP with a considerable improvement with respect to LIN. Regarding CNN techniques, it is possible to see that all models (except CNN2D and CNNR with PSRI-NIR) are able to achieve higher performances than those of RFR and MLP, which indicates that CNNs are generally better technologies to relieve the gap between multi-temporal S2 measurements and S3 VI values. In this sense, DRNN and CNN3D have shown to be the best performing competitors since their architectures are tailored for effectively working with multi-temporal observations, being DRNN better with NDVI/OLCI and CNN3D with PSRI-NIR by a relatively small margin. Nonetheless, the proposed approach is able to provide consistent improvements across all the considered VI and metrics. In Table II, the corrected re-sampled t-test [24] reveals the statistical significance of the obtained average improvements with a significance level of $\alpha = 0.05$. The qualitative results displayed in Figure 2 also support these observations. According to the magnified details, the NDVI map estimated by the proposed model is the one that provides the most distinguishable predictions for dense vegetation areas while inferring inter-sensor estimates closer to the corresponding ground-truth values.

With respect to the other methods, the presented architecture has two main advantages: (i) multi-temporal feature patterns and (ii) short-term data adaptability. On the one hand, the use of spatio-temporal kernels that slide across time allows the proposed network to uncover dynamic patterns which become more meaningful than those convolutional features directly extracted from staked S2 data. This can be noticed in the experimental comparison, where the best performing competitors are certainly those that are able to exploit temporal features. On the other hand, the proposed network topology has been specially designed to avoid temporal over-fitting when accounting for the specific nature of S2 and S3 data. In contrast to other successful temporal models, like the LSTM used by DRNN, the presented regression framework embeds the multi-temporal data into a new convolutional dimension which allows extracting discriminating temporal features even from short time series that may over-fit other schemes. Besides, the configuration of layers and blocks has been adapted to retain as much as possible the inter-sensor

spatial information taking into account the important spatial resolution differences between S2 and S3 instruments. All in all, these two advantages contribute to the better performance of the proposed model to time-resolve S3 VI from multi-temporal S2 imagery.

IV. CONCLUSIONS

This paper proposes a novel 3D-CNN architecture (3CRN) to generate temporally enhanced S3 VI from multi-temporal S2 data. The proposed model not only takes advantage of inter-sensor spatial-spectral data, but also short-term temporal information with the objective of finding temporal correlations that may help to uncover S3 VI as SIF proxies at unavailable timestamps. The experimental comparison, conducted over a collection of 21 coupled MSI and OLCI products, validates the performance of the presented architecture with respect to multiple traditional and CNN-based regression methods. The main conclusion that arises from this work is the importance of considering multi-temporal data for generating more accurate inter-sensor predictions and how the proposed 3D-CNN may help to relieve important spatial differences between instruments. In the future, we plan to extend this work to other multi-modal platforms.

REFERENCES

[1] M. Chi, A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen, and Y. Zhu, "Big data for remote sensing: Challenges and opportunities," *Proc. IEEE*, vol. 104, no. 11, pp. 2207–2219, 2016.

[2] A. S. Belward and J. O. Skøien, "Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites," *ISPRS J. Photogramm. Remote Sens.*, vol. 103, pp. 115–128, 2015.

[3] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort *et al.*, "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, 2012.

[4] C. Donlon, B. Berruti, A. Buongiorno, M.-H. Ferreira, P. Féménias, J. Frerick, P. Goryl, U. Klein, H. Laur, C. Mavrocordatos *et al.*, "The global monitoring for environment and security (GMES) sentinel-3 mission," *Remote Sens. Environ.*, vol. 120, pp. 37–57, 2012.

[5] Z. Malenovskỳ, H. Rott, J. Cihlar, M. E. Schaepman, G. García-Santos, R. Fernandes, and M. Berger, "Sentinels for science: Potential of Sentinel-1-2-3 missions for scientific observations of ocean, cryosphere, and land," *Remote Sens. Environ.*, vol. 120, pp. 91–101, 2012.

[6] P. Thanh Noi and M. Kappas, "Comparison of random forest, k-nn, and support vector machine classifiers for land cover classification using Sentinel-2 imagery," *Sensors*, vol. 18, no. 1, p. 18, 2018.

[7] X. Wang, F. Ling, H. Yao, Y. Liu, and S. Xu, "Unsupervised Sub-Pixel Water Body Mapping with Sentinel-3 OLCI Image," *Remote Sens.*, vol. 11, no. 3, p. 327, 2019.

[8] J. Vicent, N. Sabater, C. Tenjo, J. R. Acarreta, M. Manzano, J. P. Rivera, P. Jurado, R. Franco, L. Alonso, J. Verrelst *et al.*, "FLEX end-to-end mission performance simulator," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4215–4223, 2016.

[9] G. H. Mohammed, R. Colombo, E. M. Middleton, U. Rascher, C. van der Tol, L. Nedbal, Y. Goulas, O. Pérez-Priego, A. Damm, M. Meroni *et al.*, "Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress," *Remote Sens. Environ.*, vol. 231, p. 111177, 2019.

[10] D. Arnas, P. Jurado, I. Barat, B. Duesmann, and R. Bock, "FLEX: A Parametric Study of Its Tandem Formation With Sentinel-3," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 12, no. 7, pp. 2447–2452, 2019.

[11] R. Fernandez-Beltran, J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and F. Pla, "Multimodal probabilistic latent semantic analysis for sentinel-1 and sentinel-2 image fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 9, pp. 1347–1351, 2018.

[12] R. Fernandez-Beltran, F. Pla, and A. Plaza, "Sentinel-2 and Sentinel-3 Intersensor Vegetation Estimation via Constrained Topic Modeling," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 10, pp. 1531–1535, 2019.

[13] S. Bandopadhyay, A. Rastogi, S. Cogliati, U. Rascher, M. Gabka, and R. Juszczak, "Can Vegetation Indices Serve as Proxies for Potential Sun-Induced Fluorescence (SIF)? A Fuzzy Simulation Approach on Airborne Imaging Spectroscopy Data," *Remote Sens.*, vol. 13, no. 13, 2021.

[14] H. Shen, X. Li, Q. Cheng, C. Zeng, G. Yang, H. Li, and L. Zhang, "Missing information reconstruction of remote sensing data: A technical review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 61–85, 2015.

[15] C. Zeng, H. Shen, M. Zhong, L. Zhang, and P. Wu, "Reconstructing MODIS LST based on multitemporal classification and robust regression," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 512–516, 2014.

[16] W. Zhao, H. Wu, G. Yin, and S.-B. Duan, "Normalization of the temporal effect on the MODIS land surface temperature product using random forest regression," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 109–118, 2019.

[17] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4274–4288, 2018.

[18] M. Shao, C. Wang, T. Wu, D. Meng, and J. Luo, "Context-Based Multiscale Unified Network for Missing Data Reconstruction in Remote Sensing Images," *IEEE Geosci. Remote Sens. Lett.*, 2020.

[19] E. Aptoula and S. Ariman, "Chlorophyll-a retrieval from sentinel-2 images using convolutional neural network regression," *IEEE Geosci. Remote Sens. Lett.*, 2021.

[20] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 279–317, 2019.

[21] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on convolutional neural networks (cnn) in vegetation remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 24–49, 2021.

[22] W. Yu, J. Li, Q. Liu, J. Zhao, Y. Dong, C. Wang, S. Lin, X. Zhu, and H. Zhang, "Spatial-temporal prediction of vegetation index with deep recurrent neural networks," *IEEE Geosci. Remote Sens. Lett.*, 2021.

[23] N. Arslan and A. Sekertekin, "Application of Long Short-Term Memory neural network model for the reconstruction of MODIS Land Surface Temperature images," *J. Atmos. Sol. Terr. Phys.*, vol. 194, p. 105100, 2019.

[24] C. Nadeau and Y. Bengio, "Inference for the generalization error," *Machine learning*, vol. 52, no. 3, pp. 239–281, 2003.