DEEP LEARNING-BASED RECONSTRUCTION OF SPATIOTEMPORALLY FUSED SATELLITE IMAGES FOR SMART AGRICULTURE APPLICATIONS IN A HETEROGENEOUS AGRICULTURAL REGION

A. Htitiou [1,2] *, A. Boudhar [1,3], Y. Lebrini [1,2], T. Benabdelouahab [2]

[1] Water Resources Management and Valorization and Remote Sensing Team, Faculty of Sciences and Techniques, Sultan Moulay Slimane University, Beni Mellal, Morocco - (abdelaziz.htitiou, ab.boudhar, y.lebrini)@usms.ac.ma
[2] National Institute of Agronomic Research, Rabat, Morocco - tarik.benabdelouahab@inra.ma
[3] Center for Remote Sensing Applications (CRSA), Mohammed VI Polytechnic University, Ben Guerir, Morocco - Abdelghani.BOUDHAR@um6p.ma

KEY WORDS: Spatiotemporal data fusion, Normalized difference vegetation index (NDVI), High spatial and temporal resolution, Satellite image time series, Landsat 8, Sentinel-2

ABSTRACT:

Remote sensing offers spatially explicit and temporally continuous observational data of various land surface parameters such as vegetation index, land surface temperature, soil moisture, leaf area index, and evapotranspiration, which can be widely leveraged for various applications at different scales and contexts. One of the main applications is agricultural monitoring, where a smart system based on precision agriculture requires a set of satellite images with a high resolution, both in time and space to capture the phenological stages and fine spatial details, especially in landscapes with various spatial heterogeneity and temporal variation. These requirements sometimes cannot be provided by a single sensor due to the trade-off required between spatial and temporal resolutions and/or the influence of cloud cover. The data availability of new generation multispectral sensors of Landsat-8 (L8) and Sentinel-2 (S2) satellites offers unprecedented options for such applications. Given this, the current study aims to display how the synergistic use of these optical sensors can efficiently support such an application. Herein, this study proposes a deep learning spatiotemporal data fusion method to fill the need for predicting a dense time series of vegetation index with fine spatial resolution. The results show that the developed method creates more accurate fused NDVI time-series data that were able to derive phenological stages and characteristics in single-crop fields, while keeps more spatial details in such a heterogeneous landscape.

1. INTRODUCTION

One of the methods that scientists and decision-makers rely on for effective tracking and monitoring of agricultural landscapes is detailed spatial information about crop growth (DeFries et al. 2005). Currently, remote sensing satellite systems offer great potential to monitor and analyse agricultural crops at various spatial and temporal scales owing to its synoptic coverage and repetitive measurements (Benabdelouahab et al., 2019; Htitiou et al., 2020; Lebrini et al., 2020, 2019). A wide range of observation satellites can be used to produce vegetation indices such as the normalized difference vegetation index (NDVI) for monitoring crops, vegetation photosynthetic activities and biophysical properties by taking advantage of its high correlation with the leaf area index and leaf chlorophyll concentrations.

Nevertheless, this kind of application seems to be a challenging issue particularly in regions where high levels of both spatial and temporal vegetation heterogeneity are present and which constitute significant constraints that limit the efficiency of any vegetation studies.

For this reason, only the availability of NDVI time-series data with both high spatial and high temporal resolution could constitute a major asset for such application. However, these requirements cannot be fulfilled in most practical settings by a single sensor due to the trade-off between spatial and temporal resolutions. Moreover, cloud contaminations and poor atmospheric conditions can further increase the discontinuity and the fragmentation of satellite observations data and make their use seriously limited.

In response to these limitations, spatiotemporal fusion models are emerging as a powerful way to obtain an image that mitigates the individual limitations of input datasets and therefore produces a simultaneously high temporal and spatial resolutions product. So far, various spatiotemporal fusion models have been developed to blend remote sensing images such as the spatial and temporal adaptive reflectance fusion model (STARFM) (Gao et al. 2006) and its variants (Zhu et al. 2010).

However, these classical spatiotemporal methods are not always the preferred tool due to their main use in monitoring surface dynamics and phenological changes in areas without significant changes(Watts et al. 2011); while their prediction accuracies experience problems with surface heterogeneity and abrupt changes in land cover types and, therefore, cause vegetation studies errors (Emelyanova et al. 2013). Besides that, many fusion algorithms require plenty of parameters which are difficult to tune and carry a direct impact on prediction result (Zhu et al. 2010). Thus, there is an urgent need for the development of novel fusion methods that can achieve high prediction accuracy regardless of the change type, land-cover change, or phenological change.

---

* Corresponding author

In this context, the numerous recent breakthroughs in deep-learning has emerged as an ideal tool for addressing and improving spatiotemporal fusion multi-sensors data, over existing algorithms (Ao, Sun y Xin 2020; Tan et al. 2018). Specifically, the convolution neural network (CNN), which is a type of deep learning technology that can automatically extract spatial features from source images, and has been shown to be successful for remote sensing image applications, such as image super-resolution and enhancement.

To date, there have only been a few studies that have investigated the use of CNN algorithms for the spatiotemporal fusion of earth-observing satellites data, particularly for the new ones such as Sentinel-2 and Landsat-8, and still lacks standard terminology on how they combined in detailed and frequent spatial and temporal resolution will improve the agricultural application and mainly crop monitoring. Therefore, the overarching goal of our study is to evaluate the composition of Landsat-8 and Sentinel-2 NDVI data to provide temporally dense synthetic NDVI in near real-time for improved agricultural management and monitoring all by using a deep learning model for spatiotemporal fusion.

2. STUDY AREA, SATELLITE DATA, AND SYNTHETIC DATA SETS

2.1 Study area

The irrigated perimeter of Tadla, with an area of 97 000 ha, is situated in the center of Morocco (Figure1), between the Atlantic coast in the north-west and the Atlas Mountains in the south-east (32°23' north latitude, 6°31' west longitude, 445 m above sea level). The area is marked by an arid to semi-arid climate that is characterized by a dry season (April to September) and a rainy season (October to March). The mean annual temperature is about 19 °C, and the mean annual precipitation is approximately 300mm.

2.2 Sentinel-2 and Landsat 8 data

The Sentinel-2A MSI (S2) and Landsat 8 OLI (L8) images covering the study area data were used in this process for further use. The L8 OLI provides 9 spectral bands at 30m spatial resolution, except the panchromatic band, which has a spatial resolution of 15 m (Irons, Dwyer y Barsi 2012). While, the S2 /MSI used in this study has spectral response functions quite different compared to its predecessor with thirteen spectral bands at 10 m, 20 m and 60 m spatial resolution (Drusch et al. 2012). In this work, a time series of 26 suitable images with less than 20% cloud cover from each sensor (L8 and S2) were acquired respectively via the United States Geological Survey (USGS) on-demand interface (ESPA) (https://espa.cr.usgs.gov), and the Theia land data center from Jun 2016 to August 2017 (Table 1). The L8 and S2 images were already processed (level 2A) by the Landsat Surface Reflectance Code (LaSRC) algorithm and the MAJA processor respectively. For each of the Sentinel-2 and Landsat 8 scenes, the NDVI layers were generated using red and near-infrared spectral bands according to the following equation (Tucker, 1979).
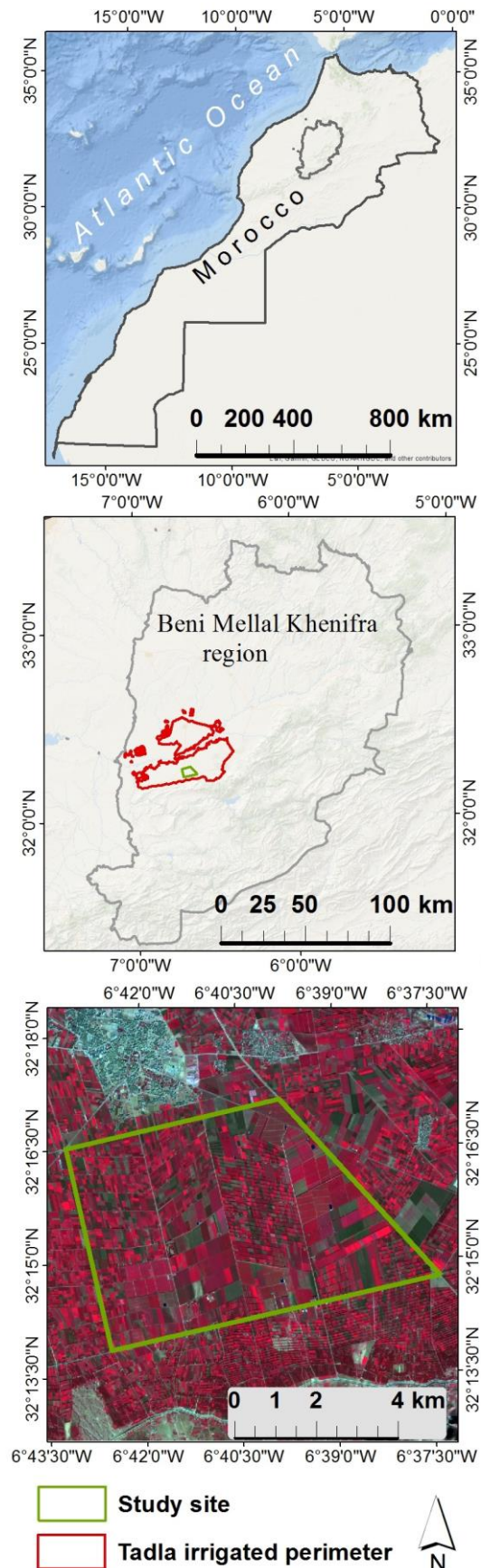
$$NDVI = (NIR-RED)/(NIR+RED) \qquad (1)$$

Figure 1. Location of the study area and a corresponding Sentinel-2 image (False Color Combination: NIR, red, green band as RGB).

Sensor	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
Sentinel-2A	-	-	-	-	-	2	2	1	2	-	2	2	2016
Landsat 8	-	-	-	-	-	2	2	2	2	2	2	1	2016
Sentinel-2A	2	2	-	2	3	3	3	-	-	-	-	-	2017
Landsat 8	2	2	2	2	1	1	2	1	-	-	-	-	2017

Table 1. Acquisition dates (2016 and 2017) of Sentinel-2 and Landsat 8 satellite imagery used in this analysis.

3. METHODOLOY

3.1 Convolutional Neural Networks (CNNs)

With the rapid development of deep learning techniques in recent years, various deep learning-based methods have been proposed to deal with complex tasks in several applications. Among these architectures, convolutional neural networks (CNN) are drawing considerable attention and popularity due to their simple structure, as well as the applicability and efficiency in various fields of image processing (Dong et al., 2016; Jiao et al., 2017; Masi et al., 2016). Generally, a CNN is a multilayer feed-forward neural network that mainly consisted of one input layer, convolution layers, pooling layers, reLU (Rectified Linear Unit) layers, fully connected layers and one output layer. Additional layers can be optionally added to the network depending on the requirement of more complex problems.

3.2 Super-resolution convolutional neural network (SRCNN) for spatiotemporal image fusion

Although, CNN has been initially proposed for the prediction of categorical variables such as class labels; it has recently been modified to produce continuous output values (Dong et al. 2016). This has opened the door for a wide CNN-based image restoration super-resolution and fusion. In this context, (Dong et al. 2016) proposed the milestone Super Resolution CNN (SRCNN) with three layers of neural networks to downscale the coarse images to high spatial resolution. In our experiment, we employ SRCNN as a remote sensing image spatiotemporal fusion method itself. Its network architecture is shown in Figure 2. The SRCNN architecture comprises three-layer, where the filter sizes of each layer are $64 \times 1 \times 9 \times 9$, $32 \times 64 \times 5 \times 5$ and $1 \times 32 \times 5 \times 5$. After each convolution filter in every layer, an activation function, rectified linear unit (RELU), is applied except last layer. Last layers of the network are again convolutional layers, which provide the final learned mapping from an input low-resolution image patch to a high-resolution image pixel value. Hence, the high-resolution image is constructed at the output of the CNN network.

The proposed SRCNN framework needs three inputs: a coarse- and fine-resolution NDVI images from different satellite sensors observed at the same date (t1) and one coarse-resolution image acquired on the prediction date (t2). The output is the high-resolution image on the prediction date. The proposed framework is summarized in the flowchart of Figure 2.

The proposed fusion SRCNN model is divided into training phase and testing phase. In the training phase, we used the two pairs of Landsat–Sentinel data that were imaged on 11 September 2016 and 18 February 2017 (base dates), respectively (as indicated by Table 2). On the other hand, in the testing phase, the Landsat NDVI data acquired on 09 May 2017 (prediction date) is used for prediction of Sentinel-like NDVI data.

Acquisition date	Landsat 8	Sentinel-2A
11 September 2016	Model input (start date)	Model input (start date)
18 February 2017	Model input (end date)	Model input (end date)
09 May 2017	Model input (for prediction date)	For model validation

Table 2. The acquisition dates of every image used for the downscaling process and their function

3.3 Qualitative and quantitative assessment of the fusion results

The results were compared to the actual S2 image on May 09, 2017, using both qualitative and quantitative assessments to check the fused NDVI products and summarize their performance differences. The qualitative assessment is performed based on visual analysis of observed (actual) and predicted (synthetic) NDVI in terms of change in reflectance and spatial pattern. On the other hand, a quantitative assessment is performed two main statistical metrics, such as root mean square error (RMSE), and the correlation coefficient (r). It is noted that a smaller RMSE indicates better prediction, and an r close to 1 indicates a high similarity between the compared images.
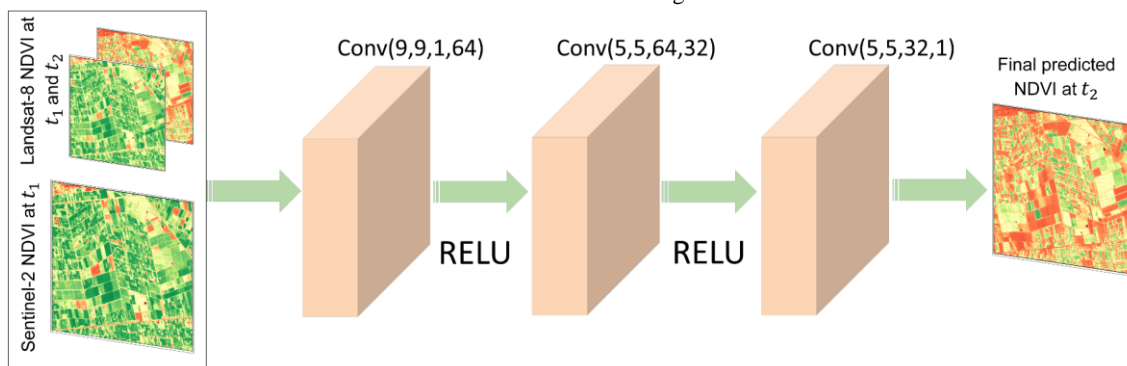

Figure 2. The architecture of the proposed deep convolutional spatiotemporal fusion network method.

4. RESULTS AND DISCUSSIONS

4.1 Assessment of the fusion performance

As was mentioned earlier, the pairs of Sentinel-2 and Landsat 8 NDVI images of the base dates were fused with a Landsat NDVI image of the prediction date to generate synthetic Sentinel-like NDVI on May 9, 2017. Figure 3 demonstrates the visual comparison of the predicted NDVI result obtained from the developed method and its corresponding reference Landsat and Sentinel NDVI images for two sub-areas of A and B that represent farmland sites with great heterogeneity, noticeable and rapid phenological change.
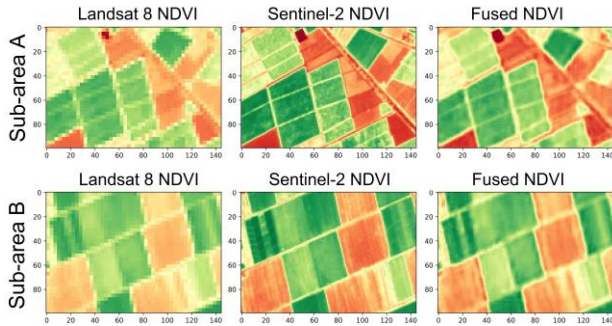

Figure 3. Visual comparisons between the actual and the predicted Sentinel-like NDVI with the corresponding Landsat NDVI on May 09, 2017. Note that the sub-areas A and B are selected and cropped from the study site

In generally, it is observed that the proposed method produces accurate and satisfactory results regarding the fact that the actual and predicted images are comparable and exhibit spatial and spectral similarities.

To further confirm the performance and great potential of our method, we randomly selected pixels as samples from four sites in the study area and compared (as shown in Figure 4) their NDVI values of the predicted Sentinel-like image against that of the true image for the prediction date on 09 May 2017 using scatter and residuals plots, which will provide an intuitive comparison in explaining the approximate extent of the distribution between the fusion and the actual NDVI.

From these results, we can observe that scatter and residuals plots also confirmed that the proposed SRCNN method yielded revolutionary results and good performance for downscaling Landsat-8 NDVI to a higher spatial resolution. Figure 4a indicates a high agreement between the actual Sentinel-2 NDVI and synthetic Sentinel-2 NDVI from the proposed method. This is explained by the fact that all the points of predicted data were scattered along the 1:1 line and correlated well with the referenced data for the four sites. Figure 4b shows also that the residuals are non-randomly and more closely distributed around the zero line which means a perfect fit and good linear relationship between predicted and actual images. From a statistical point of view it is evident that SRCNN method can produce fused images that more closely resemble the real Sentinel-2 NDVI even in areas experiencing land use and land cover changes through the season as confirmed by a lower RMSE (0.109) and a higher correlation r (0.918) for the whole synthetic NDVI image.
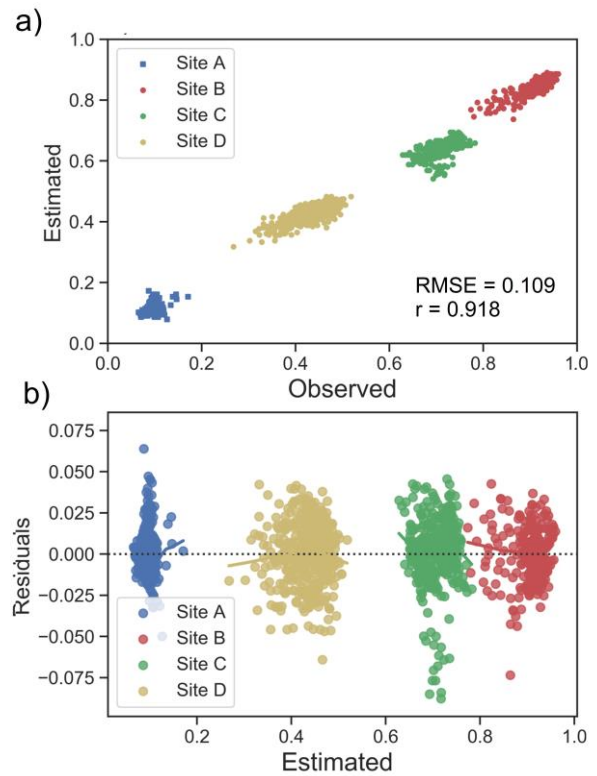

Figure 4. Correlation analysis between the actual and synthetic Sentinel-2 NDVI on May 09, 2017 in four sites over the study area

4.2 Phenological Analysis

Agricultural monitoring was selected as a relevant remote sensing application to demonstrate the potential of the proposed data fusion approach. Based on the above analysis of the SRCNN fusion scheme, we decided to extend the developed method for each time-series L8 image during the growing season and served as predicted images. As a result, a total of 26 Sentinel-like NDVI scenes were predicted and were added to 26 existing Sentinel-2 NDVI observations to generate a very dense time series.

However, noise (clouds etc.) usually makes it difficult to generate reliable time series of vegetation indices. For that reason, Savitzky-Golay filter (Chen et al., 2004) is applied to reconstruct continuous dense time series by eliminating the remaining noises in the time-series besides that preserving the trend of the original NDVI data fairly well; in a way that better depicted the seasonal variation in vegetation characteristics; it, therefore, provides reliable data for the extraction of phenological information.

Figure 5 shows the phenological growth profiles of the average NDVI generated by the SRCNN method of two fields of wheat and alfalfa, respectively in the study area during the 2016/2017 cropping season. The average NDVI extracted from the original Landsat-8 NDVI time series images and Sentinel-2 NDVI images were also presented.
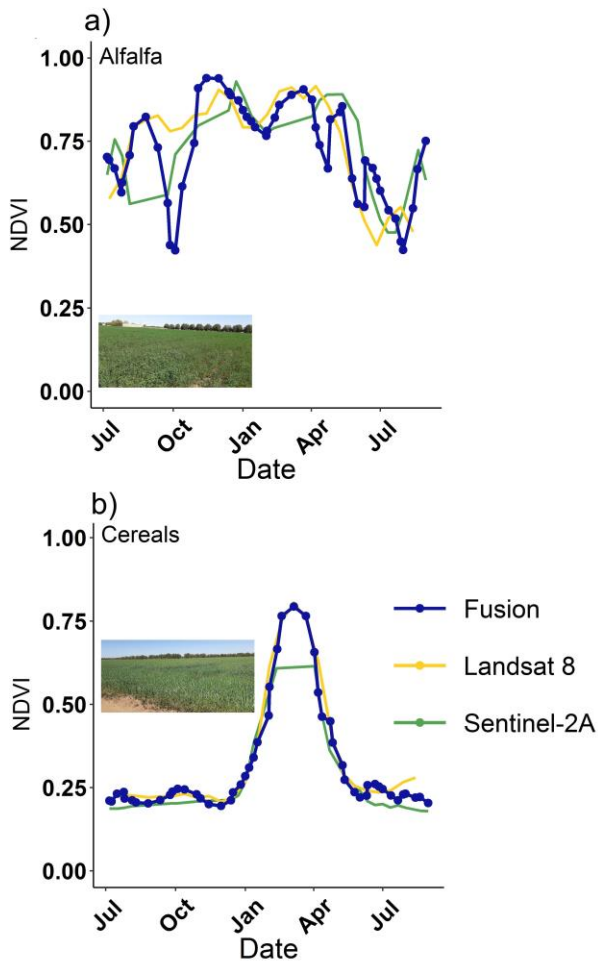
Figure 5. Comparison of NDVI temporal profiles obtained from Landsat, Sentinel-2, and their combined use by SRCNN spatiotemporal fusion method over an a) alfalfa and b) wheat field

In general, the smoothed phenological profiles of NDVI show a unique and well-defined phenological characteristic for each crop type. However, significant differences in crop growth status can be observed for each density of the NDVI time series. It is clear that combining S2 and L8 data made it possible to monitor perfectly and continuously the crop development during the season and help further to capture accurate temporal changes while preserving fine-spatial-resolution details, which was previously not achievable by the only use of L8 or S2 data as observed in Figure 5 and consistent with previous studies (Htitiou et al., 2019). This becomes clearer when a sudden change (wheat harvest, or alfalfa cut) in the temporal NDVI profile takes place since it cannot be estimated properly using only the Landsat or sentinel-2 time series. On the other hand, it can be perfectly satisfactory when it uses the fused L8/S2 NDVI time series.

This result in addition to previous findings about the quality assessment and consistency of the derived NDVI suggests that our spatiotemporal method is suitable to construct time-series images for continuous monitoring of vegetation change over areas that contrasting spatial and temporal heterogeneity. And it can perfectly estimate subfield-scale crop phenology and further provides new opportunities to increase the knowledge about environmental changes and to support many operational applications that require monitoring rapidly-varying phenomena and high spatial resolution, such as precision agriculture, irrigation advisory services, and near real-time change detection.

5. CONCLUSIONS

Precision agriculture requires detailed crop status information at high spatial and temporal resolutions. Remote sensing can provide such information, but single sensor observations are often incapable of meeting all data requirements. In recent years, many spatial and temporal satellite image fusion (STIF) methods have been developed to solve the problems of trade-off between spatial and temporal resolution of satellite sensors. In this study, we investigate the capacity of a deep learning-based approach to map NDVI at 10 m spatial resolution and at a high temporal resolution in an experimental site in the middle of Morocco. The obtained results indicate that even in a challenging study area with heterogeneous landscapes, the combined use of images from S2 and L8 satellites by our method can provide very useful information regarding spatial and temporal patterns of crops during their growing stages. Despite the promising results of the proposed fused approach in this study, some issues need to be studied further. The recent launch of the European Sentinel-2B sensor that will bring some improvement to our approach by providing global observations on average once every 2.9 days, which could make a very dense time series for each season and further better monitoring of the vegetation dynamics and temporal variations.

REFERENCES

Ao, Z., Sun, Y., Xin, Q., 2020. Constructing 10-m NDVI Time Series From Landsat 8 and Sentinel 2 Images Using Convolutional Neural Networks. IEEE Geosci. Remote Sens. Lett. 1–5. https://doi.org/10.1109/LGRS.2020.3003322

Benabdelouahab, T., Lebrini, Y., Boudhar, A., Hadria, R., Htitiou, A., Lionboui, H., 2019. Monitoring spatial variability and trends of wheat grain yield over the main cereal regions in Morocco: a remote-based tool for planning and adjusting policies. Geocarto Int. 1–20. https://doi.org/10.1080/10106049.2019.1695960

Chen, J., Jönsson, Per., Tamura, M., Gu, Z., Matsushita, B., Eklundh, L., 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. Remote Sens. Environ. 91, 332–344. https://doi.org/10.1016/j.rse.2004.03.014

DeFries, R., Pagiola, S., Adamowicz, W.L., Akcakaya, H.R., Arcenas, A., Babu, S., Balk, D., Confalonieri, U., Cramer, W., Falconí, F., 2005. Analytical approaches for assessing ecosystem condition and human well-being. Ecosyst. Hum. Well- Curr. State Trends 1, 37–71.

Dong, C., Loy, C.C., He, K., Tang, X., 2016. Image Super-Resolution Using Deep Convolutional Networks. IEEE Trans. Pattern Anal. Mach. Intell. 38, 295–307. https://doi.org/10.1109/TPAMI.2015.2439281

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. Remote Sens. Environ., The Sentinel Missions - New Opportunities for Science 120, 25–36. https://doi.org/10.1016/j.rse.2011.11.026

Emelyanova, I.V., McVicar, T.R., Van Niel, T.G., Tao Li, L., Van Dijk, A., 2013. Remote Sensing of Environment Assessing the accuracy of blending Landsat–MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection. Remote Sens. Environ. 133, 193–209.

Gao, F., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance. IEEE Trans. Geosci. Remote Sens. 44, 2207–2218.

Htitiou, A., Boudhar, A., Lebrini, Y., Hadria, R., Lionboui, H., Benabdelouahab, T., 2020. A comparative analysis of different phenological information retrieved from Sentinel-2 time series images to improve crop classification: a machine learning approach. Geocarto Int. 1–24. https://doi.org/10.1080/10106049.2020.1768593

Htitiou, A., Boudhar, A., Lebrini, Y., Hadria, R., Lionboui, H., Elmansouri, L., Tychon, B., Benabdelouahab, T., 2019. The Performance of Random Forest Classification Based on Phenological Metrics Derived from Sentinel-2 and Landsat 8 to Map Crop Cover in an Irrigated Semi-arid Region. Remote Sens. Earth Syst. Sci. 2, 208–224. https://doi.org/10.1007/s41976-019-00023-9

Irons, J.R., Dwyer, J.L., Barsi, J.A., 2012. The next Landsat satellite: The Landsat Data Continuity Mission. Remote Sens. Environ., Landsat Legacy Special Issue 122, 11–21. https://doi.org/10.1016/j.rse.2011.08.026

Jiao, L., Liang, M., Chen, H., Yang, S., Liu, H., Cao, X., 2017. Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 55, 5585–5599.

Lebrini, Y., Benabdelouahab, T., Boudhar, A., Htitiou, A., Hadria, R., Lionboui, H., 2019. Farming systems monitoring using machine learning and trend analysis methods based on fitted NDVI time series data in a semi-arid region of Morocco, in: Neale, C.M., Maltese, A. (Eds.), Remote Sensing for Agriculture, Ecosystems, and Hydrology XXI. Presented at the Remote Sensing for Agriculture, Ecosystems, and Hydrology XXI, SPIE, Strasbourg, France, p. 31. https://doi.org/10.1117/12.2532928

Lebrini, Y., Boudhar, A., Htitiou, A., Hadria, R., Lionboui, H., Bounoua, L., Benabdelouahab, T., 2020. Remote monitoring of agricultural systems using NDVI time series and machine learning methods: a tool for an adaptive agricultural policy. Arab. J. Geosci. 13, 796. https://doi.org/10.1007/s12517-020-05789-7

Masi, G., Cozzolino, D., Verdoliva, L., Scarpa, G., 2016. Pansharpening by convolutional neural networks. Remote Sens. 8, 594.

Tan, Z., Yue, P., Di, L., Tang, J., 2018. Deriving High Spatiotemporal Remote Sensing Images Using Deep Convolutional Network. Remote Sens. 10, 1066. https://doi.org/10.3390/rs10071066

Watts, J.D., Powell, S.L., Lawrence, R.L., Hilker, T., 2011. Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery. Remote Sens. Environ. 115, 66–75.

Zhu, X., Chen, J., Gao, F., Chen, X., Masek, J.G., 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. Remote Sens. Environ. 114, 2610–2623.