Sentinel-2 and Sentinel-3 Intersensor Vegetation Estimation via Constrained Topic Modeling

Ruben Fernandez-Beltran, Filiberto Pla, and Antonio Plaza, *Fellow, IEEE*

***Abstract*— This letter presents a novel intersensor vegetation estimation framework, which aims at combining Sentinel-2 (S2) spatial resolution with Sentinel-3 (S3) spectral characteristics in order to generate fused vegetation maps. On the one hand, the multispectral instrument (MSI), carried by S2, provides high spatial resolution images. On the other hand, the Ocean and Land Color Instrument (OLCI), one of the instruments of S3, captures the Earth's surface at a substantially coarser spatial resolution but using smaller spectral bandwidths, which makes the OLCI data more convenient to highlight specific spectral features and motivates the development of synergetic fusion products. In this scenario, the approach presented here takes advantage of the proposed constrained probabilistic latent semantic analysis (CpLSA) model to produce intersensor vegetation estimations, which aim at synergically exploiting MSI's spatial resolution and OLCI's spectral characteristics. Initially, CpLSA is used to uncover the MSI reflectance patterns, which are able to represent the OLCI-derived vegetation. Then, the original MSI data are projected onto this higher abstraction-level representation space in order to generate a high-resolution version of the vegetation captured in the OLCI domain. Our experimental comparison, conducted using four data sets, three different regression algorithms, and two vegetation indices, reveals that the proposed framework is able to provide a competitive advantage in terms of quantitative and qualitative vegetation estimation results.**

***Index Terms*— Constrained probabilistic latent semantic analysis (CpLSA), Sentinel-2, Sentinel-3, topic models, vegetation estimation.**

I. Introduction

THE Copernicus program is a joint initiative of the European Commission, the European Space Agency, and the European Environment Agency in order to provide operational monitoring information from space, useful for environment and security applications. In this context, five different Sentinel Earth observation missions have been planned to guarantee this operational provision [1]. Among all the program resources, Sentinel-2 (S2) and Sentinel-3 (S3) missions are focused on global monitoring services over terrestrial and aquatic surfaces, using for this purpose high-resolution and midresolution multispectral imagery [2]. More specifically, S2 [3] is a polar-orbiting mission, which comprises two identical satellites: S2A, launched on June 23, 2015, and S2B, which followed on March 7, 2017. Each satellite incorporates a multispectral instrument (MSI), which provides a versatile set of 13 spectral bands ranging from the visible and near infrared (VNIR) to the shortwave infrared (SWIR). Four of these bands (B02-B04, B08) are acquired at a spatial resolution of 10 m, six bands (B05-B07, B08A, B11, B12) at 20 m and the remaining three bands (B01, B09, B10) at 60 m. Analogously, S3 [4] includes a pair of satellites, called S3A and S3B, where the first one was launched on February 16, 2016 and the second one was successfully launched on April 25, 2018. Both satellites carry the Ocean and Land Color Instrument (OLCI), which provides 21 bands (Oa01–Oa21) spanning from 390- to 1040-nm VNIR spectral range with bandwidths from 2.5 to 40 nm. Regarding the spatial resolution of the sensor, OLCI has global resolution requirement of 300 m.

Although S2 and S3 missions have been designed to provide global data products of vegetation, soil and water cover, inland waterways and coastal areas, the spectral and spatial differences between MSI and OLCI sensors make each satellite more suitable for a particular application field. On the other hand, the higher spatial resolution in S2 enables the use of its products for characterization tasks, with the requirement of a high level of spatial details such as soil mapping or land use classification [5], S3 is able to capture imagery using smaller spectral bandwidths, which makes the OLCI data more convenient to highlight specific spectral responses that represent different features over the Earth's surface. Specifically, vegetation cover can exemplify this point [6]. In general, vegetation indices, such as the normalized difference vegetation index (NDVI) [7] and the soil-adjusted vegetation index (SAVI) [8], seek to exploit the correlation between the maximum chlorophyll absorption wavelength and the Red-Edge electromagnetic spectrum. As a result, the smaller VNIR spectral bandwidth of the OLCI sensor makes that fewer wavelengths are involved in the NDVI and SAVI computations. This fact generates an enhanced response for plant surfaces, which eventually increases the instrument sensitivity to detect those image areas with certain types of vegetation [4]. Precisely, these intersensor differences motivate the development of fused vegetation products to exploit MSI spatial resolution and OLCI spectral features.

In the literature, different kinds of regression algorithms have been successfully applied to conduct biophysical

Manuscript received November 6, 2018; revised January 10, 2019; accepted March 2, 2019. Date of publication March 26, 2019; date of current version September 25, 2019. This work was supported by Generalitat Valenciana (APOSTD/2017/007), the Spanish Ministry of Economy (ESP2016-79503-C2-2-P, TIN2015-63646-C5-5-R), Junta de Extremadura (Ref. GR18060) and the European Union under the H2020 EOXPOSURE project (No. 734541). *(Corresponding author: Ruben Fernandez-Beltran.)*

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

A. Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, PC-10003 Cáceres, Spain (e-mail: aplaza@unex.es).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2019.2903231

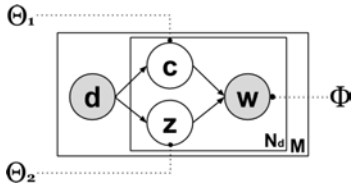

Fig. 1. CpLSA model.

parameter estimations within the context of Sentinel missions. Specifically, Verrelst *et al.* [9] review several state-of-the-art machine learning regression algorithms for S2 and S3 satellites, and Caicedo *et al.* [10] assess multiple linear and nonlinear regression algorithms with a range of remotely sensed data. Despite the value of these and other related works, the regression process is often conducted from a single-sensor perspective and, usually, they only consider simulated Sentinel data [11]. This letter is focused on a more general objective, where S2 and S3 operational products are combined to generate fused vegetation maps with MSI spatial resolution and OLCI spectral characteristics, that is, the objective of this letter is based on exploiting the existing synergy between S2 and S3 missions to generate improved vegetation estimates of the Earth surface. On the other hand, standard regression algorithms are able to generate such estimations by directly relying on low-level reflectance values, the proposed approach takes advantage of a newly proposed constrained probabilistic latent semantic analysis (CpLSA) topic model to uncover two different kinds of discriminating patterns in the S2 MSI spectral domain: 1) constrained-topics that are able to reproduce the vegetation detected by the S3 OLCI sensor and 2) standard-topics that represent the rest of the nonvegetation components in S2 MSI. In this way, image pixels are managed at a higher abstraction level as a dual mixture of spectral patterns and, hence, it is possible to infer more accurate high-resolution vegetation maps at S2 MSI spatial resolution using only the most S3 vegetation discriminating patterns. Our experiments considering four data sets and two different vegetation indices reveal the advantages of the proposed approach to generate intersensor vegetation estimations when compared to three different standard regression algorithms.

II. Methodology

A. Constrained Probabilistic Latent Semantic Analysis

Based on the incremental formulation of the asymmetric probabilistic latent semantic analysis model [12], we define a topic model extension, called CpLSA, which is specially designed to relate intersensor information throughout the high-level patterns uncovered by topics. Specifically, the proposed model (Fig. 1) considers two diverging hidden random variables, that is, c and z, to represent constrained-topics and standard-topics, respectively. Note that N_d is the number of words in d, M is the total number of documents in the collection, and shaded nodes represent the observable variables in the model, by analogy with the document analysis application field [13].

In this letter, $\Theta_1 \sim \{p(c|d)\}$, $\Theta_2 \sim \{p(z|d)\}$ and $\Phi \sim \{p(w|c), p(w|z)\}$ parameters are estimated by maximizing the complete log-likelihood function using the expectation–maximization (EM) algorithm [14], which performs two stages: 1) E-step where the likelihood expected values are computed given the current estimation of the parameters and 2) M-step where the new optimal values of the parameters are calculated according to the current settings. The E-step can be computed by using Bayes' rule and the chain rule as shown in (1) and (2). For the M-step, we calculate CpLSA likelihood partial derivatives, set them as equal to zero, and solve the equations in order to obtain (3)–(6)

$$p(c|w,d) = \frac{p(c,w,d)}{p(w,d)} = \frac{p(w|c)p(c|d)}{\sum_c p(w|c)p(c|d)} \tag{1}$$

$$p(z|w,d) = \frac{p(z,w,d)}{p(w,d)} = \frac{p(w|z)p(z|d)}{\sum_z p(w|z)p(z|d)} \tag{2}$$

$$p(w|c) = \frac{\sum_d n(w,d)p(c|w,d)}{\sum_w \sum_d n(w,d)p(c|w,d)} \tag{3}$$

$$p(w|z) = \frac{\sum_d n(w,d)p(z|w,d)}{\sum_w \sum_d n(w,d)p(z|w,d)} \tag{4}$$

$$p(c|d) = \frac{\sum_w n(w,d)p(c|w,d)}{\sum_c \sum_w n(w,d)p(c|w,d) + \sum_z \sum_w n(w,d)p(z|w,d)} \tag{5}$$

$$p(z|d) = \frac{\sum_w n(w,d)p(z|w,d)}{\sum_z \sum_w n(w,d)p(z|w,d) + \sum_c \sum_w n(w,d)p(c|w,d)} \tag{6}$$

where $n(w,d)$ represents the number of times the word w appears in the document d. The EM process is performed as follows. First, $p(w|c)$, $p(w|z)$, $p(c|d)$, and $p(z|d)$ are randomly initialized. Then, the E-step (1) and (2) and the M-step (3)–(6) are alternated until the model parameters converge. As default convergence settings, we use a 10^{-6} threshold in the log-likelihood or 1000 EM iterations.

B. Intersensor Vegetation Estimation Framework

The proposed S2 and S3 intersensor vegetation estimation framework is made up of a two-step process (see Fig. 2):

1) CpLSA-Tra: In the first step, the proposed model is used to learn coupled S2-S3 vegetation patterns at S3 spatial resolution (i.e., the sensor with the lowest spatial resolution). Specifically, the S2 input image (I_2) is initially downsampled

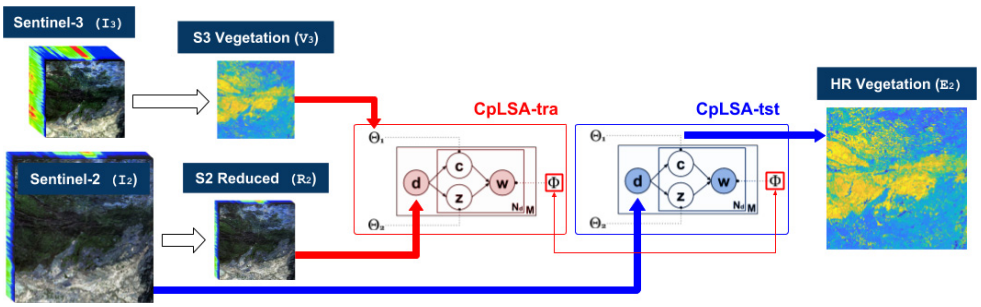

Fig. 2. Proposed Sentinel-2 and Sentinel-3 intersensor vegetation estimation framework.

to S3 nominal spatial resolution (R_2) using a bicubic kernel. Then, a vegetation index $\mathcal{V}$ is applied over the S3 input image (I_3) to generate the corresponding vegetation map $V_3 = \{V_3^1, ..., V_3^M\}$ with M pixel values. In addition, R_2 is vectorized in order to define topic model documents (d) as image pixels, words (w) as spectral bands, and document word-counts ($n(w, d)$) as pixel reflectance values. Then, CpLSA is used over R_2 by fixing $\Theta_1 \sim p(c|d)$ to a scaled and normalized version of V_3, as shown in (7) and (8), in order to learn $\Phi \sim \{p(w|c), p(w|z)\}$ parameter using C constrained-topics and Z standard-topics. Note that the conditional probability distribution $p(c|d)$ defines how image pixels are described by the target S3 vegetation map, $p(w|c)$ represents the reflectance patterns that generate this map and $p(w|z)$ contains the rest of the patterns that can be considered noise from a vegetation-based perspective

$$\hat{V}_3 = \left\{ \frac{V_3^i - \min(V_3)}{\max(V_3) - \min(V_3)} \right\}, \quad \forall\, i \in [1, M] \tag{7}$$

$$\Theta_1 = \left\{ \frac{\hat{V}_3^i}{\sum_i \hat{V}_3^i} \right\}, \quad \forall\, i \in [1, M]. \tag{8}$$

2) CpLSA-Tst: Once the Φ parameter has been estimated, the proposed model is again applied to infer the output vegetation map at S2 spatial resolution with the S3 spectral properties, that is, CpLSA is used over I_2 by fixing the Φ parameter in order to generate $\Theta_1 \sim p(c|d)$ and the resulting vegetation map as $E_2 = p(c|d)$.

III. Experiments

A. Data Sets

In this letter, four pairs of S2 MSI and S3 OLCI data products have been selected (Table I). The considered scenes include different European areas with multiple types of vegetation to increase the data heterogeneity. All the L1C products have been downloaded from the Copernicus Open Access Hub platform (https://goo.gl/uXmPxL) and they have been processed using the Sentinel Application Platform (SNAP) software as follows. The MSI products have been resampled to 20-m spatial resolution to manage the images as uniform data cubes while reducing the product size. Then, they have been atmospherically corrected using the Sen2Cor processor with the default settings. The OLCI products have been reprojected onto the corresponding S2 tiles. Besides, they have been corrected using the Rayleigh correction procedure since the complete atmospheric correction for land products is not still available in the last SNAP release. Finally, each image pair has been coregistered, obtaining a final image size of $5490 \times 5490 \times 13$ pixels in S2 and $366 \times 366 \times 21$ in S3.

TABLE I
Data Set Description

Dataset	Scene	Location	Sensing Date	Top-Left (Lat/Long) Bottom-Right (L/L)
AN	Natural park	Andujar (Spain)	10 Mar. 2017	(38.84°/ -4.15°) (37.85°/ -2.88°)
BR	Coastal area	Bourdeaux (France)	19 Apr. 2017	(45.14°/ -1.72°) (44.13°/ -0.37°)
ML	Mountain range	Milan (Italy)	16 Feb./Mar. 2017	(46.05°/ 8.99°) (45.05°/ 10.39°)
UT	Northern Europe	Utrecht (Netherlands)	27 Dec. 2016	(52.34°/ 4.46°) (51.32°/ 6.01°)

B. Experimental Protocol

The experimental part of this letter aims at validating the ability of the proposed approach to estimate S3 OLCI vegetation from S2 MSI data. More specifically, the proposed approach is compared to three standard regression algorithms, i.e., linear regression [15], Support Vector Regression (SVR) with radial basis function (RBF) [16], and Gaussian process regression (GPR) with squared exponential [17], when considering two different vegetation indices, i.e., NDVI [18] and SAVI [8]. Regarding the experimental procedure, all the methods have been trained for each image pair using the downsampled S2 image (R_2) and the corresponding S3 vegetation map (V_3). For the considered regression algorithms, we have used the corresponding MATLAB R2018b implementations with automatic scale, data standardization, and the default settings for the rest of the parameters. For the proposed approach, C and Z model parameters have been fixed to 1 and 3, respectively. Once the training process is complete, the full-resolution S2 product (I_2) is provided as a test image to estimate the corresponding S3 vegetation map at S2 spatial resolution (E_2). Since there is no S3 vegetation information available at the considered S2 pixel size (20 m), we adopt a reduced reference assessment protocol to validate the results [19]. In particular, this process consists of downsampling the original input images by the scaling ratio between S2 and S3 ($15\times$). Then, the output vegetation maps (E_2) are generated at the same spatial resolution than S3 OLCI (300 m), which allows using the original S3 vegetation maps (V_3) as a reference for a quantitative performance evaluation. As the evaluation metric, we use the mean squared error (MSE) index due to its simplicity and quadratic error computation, which penalizes predictions that substantially differ from the

TABLE II
QUANTITATIVE MSE ASSESSMENT

	Dataset	Sentinel-2	Linear	SVR	GPR	Proposed
NDVI	AN	0.0815	0.0201	0.0592	0.0203	**0.0103**
	BR	0.0847	0.0294	0.0302	0.0292	**0.0123**
	ML	0.0431	0.0417	0.0261	0.0216	**0.0115**
	UT	0.1127	0.0226	0.0212	0.0193	**0.0152**
	Avg.	0.0805	0.0284	0.0341	0.0226	**0.0123**
	Time (s)	-	0.01	0.29	0.61	1.93
SAVI	AN	0.0825	0.0201	0.0479	0.0102	**0.0094**
	BR	0.0460	0.0349	0.0342	0.0101	**0.0087**
	ML	0.0214	0.0178	0.0147	0.0118	**0.0081**
	UT	0.0701	0.0181	0.0175	0.0156	**0.0119**
	Avg.	0.0550	0.0227	0.0285	0.0119	**0.0095**
	Time (s)	-	0.01	0.29	0.59	1.89

TABLE III
STATISTICAL TEST ANALYSIS. (a) FRIEDMAN'S TEST. (b) POST HOC HOLM'S METHOD

(a)

p-value $= 3.47e-6$	
Algorithm	Ranking
Proposed	1
GPR	2.0909
SVR	3.3636
Linear	3.5455

(b)

$\alpha = 0.10$			
Hypothesis	z	p	Holm
Linear vs. Proposed	4.624048	0.000004	0.016667
SVR vs. Proposed	4.293759	0.000018	0.02
GPR vs. Proposed	1.981735	0.047509	0.05
SVR vs. GPR	2.312024	0.020776	0.033333

corresponding reference values. In addition, two statistical tests, that is, the tests in [20] and [21], have been applied for detecting statistical differences among the methods' results. It should be mentioned that both vegetation indices have been scaled and normalized, as shown in (7) and (8), to unify their corresponding value ranges for assessment purposes.

C. Results

Table II presents the quantitative evaluation of the estimated vegetation results for the considered indices, data sets, and methods in terms of the MSE metric. For each vegetation index (i.e., NDVI and SAVI), the four considered data sets are provided in rows, whereas columns represent the tested methods, i.e., Sentinel-2, Linear, SVR, GPR, and Proposed. Note that the first column measures the differences between the vegetation captured by the S2 MSI sensor with respect to the vegetation detected by S3 OLCI. The last column reports the quantitative assessment of the proposed CpLSA-based vegetation estimation framework and the last two rows in each vegetation index provide the average MSE values and test computational times. In addition, Table III presents a summary of Friedman's and Holm's statistical tests. Regarding the qualitative evaluation, Figs. 3 and 4 display the estimated vegetation maps.

One of the first noteworthy points arises when comparing the performance of the considered regression functions to the vegetation result obtained by Sentinel-2: the quantitative results reported in Table II reveal that all the tested regression functions, that is, Linear, SVR, and GPR, are able to approximate the reference OLCI vegetation better than the S2 MSI sensor, that is, using a regression function from the original S2 data to the corresponding S3 vegetation indices allows combining MSI spatial resolution and OLCI spectral characteristics. This fact is also supported by the vegetation maps displayed in Figs. 3 and 4, where it is possible to see that directly computing NDVI and SAVI over S2 data

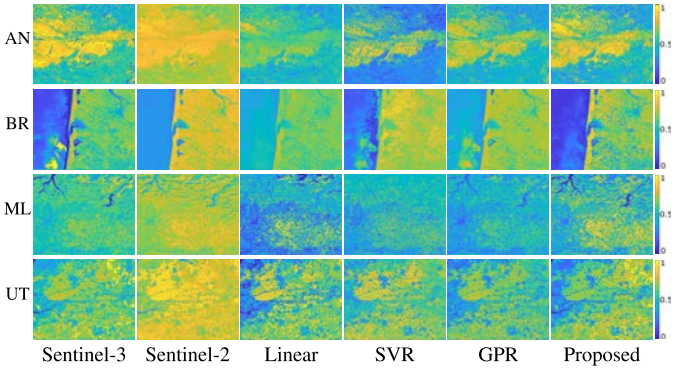

Fig. 3. NDVI qualitative evaluation results.

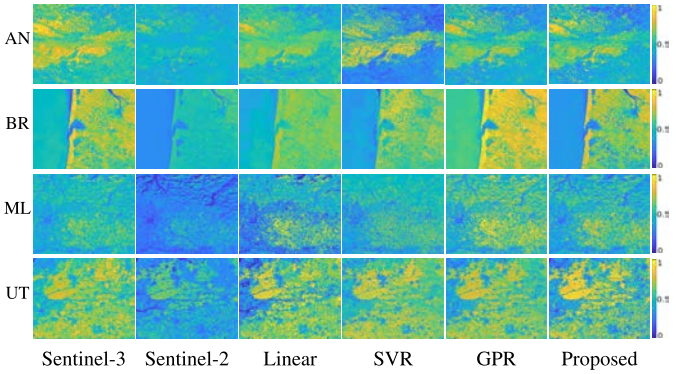

Fig. 4. SAVI qualitative evaluation results.

(Sentinel-2 column) generates a substantially different result than the corresponding reference (Sentinel-3 column). Note that the number of bands considered in NDVI and SAVI computations is rather limited; however, the regression functions are applied over the whole spectra, which allows for a better estimation of the actual vegetation.

Regarding the overall performance of the tested methods, GPR yields a remarkable MSE average performance (0.0173) when compared to the linear (0.0256) and SVR (0.0314) regression results. The reduced spatial resolution of S3, together with the straightforward nature of the NDVI and SAVI indices, makes that the SVR regression function with the RBF kernel is unable to achieve satisfactory results, with the linear regression function obtaining an even better average result. In the case of GPR, it has shown to be the most robust regressor among the three tested ones (i.e., linear, SVR, and GPR). Nonetheless, the proposed approach is able to provide a competitive advantage from both quantitative and qualitative viewpoints. On the one hand, the proposed CpLSA-based framework achieves a consistent metric improvement when estimating S3 vegetation from S2 data, reaching the best MSE average performance (0.0109) over all the considered methods. On the other hand, the vegetation maps provided by the proposed approach are certainly the most similar to the S3 data, providing NDVI and SAVI vegetation details that are not estimated by any of the other considered methods. For instance, it is possible to see in the UT row shown in Fig. 3 that the proposed approach is the only method able to retrieve the NDVI vegetation detected at the top-right image corner. A similar example can be found in the BR row shown in Fig. 4, where all the considered regression functions struggle at

capturing the coastal vegetation. These results are supported by the conducted statistical analysis. In particular, Friedman's test [Table III(a)] ranks CpLSA in the first place, GPR in the second, SVR in the third, and linear in the last position. In addition, the computed p-value provides a high level of significance to conduct a post hoc multiple comparison test. Considering a confidence level of $\alpha = 0.10$, Holm's method [Table III(b)] rejects the statistical equality hypotheses when comparing the proposed approach to linear, SVR, and GPR. Note that those hypotheses with unadjusted p-values (p) that are smaller than the adjusted Holm's values are rejected. The conducted analysis reveals that the performance improvement of the proposed approach is statistically relevant. Nevertheless, it should also be mentioned that the proposed approach is a computationally demanding model and further research should be conducted for its operational deployment.

In general, estimating the vegetation captured by the OLCI sensor from the MSI raises the challenge of uncovering information that is not present in the original S2 spectra, due to the different spectral resolution of the S3 instrument. Note that the smaller spectral bandwidths in the VNIR wavelength allow the OLCI sensor to enhance those image areas with more vegetation. Standard regression algorithms attempt to directly map the S2 spectra onto the vegetation values detected by S3. However, this straightforward approach only relies on the low-level reflectance values acquired by the coarser spectral resolution sensor, which eventually limits the resulting performance under the most challenging scenarios. The proposed approach uses the CpLSA semantic characterization space to relieve this lack of spectral information by uncovering reflectance latent patterns in the S2 spectral domain and their relationship with the S3 spectral values extracted during the training stage. In particular, CpLSA has been specifically designed to uncover two kinds of patterns: c (constrained-topics) that aim at reproducing S3 vegetation and z (standard-topics) that represent the rest of the nonvegetation components in S2. Then, it is possible to isolate the S2 reflectance patterns that are able to represent the S3 vegetation (c) from noisy patterns that do not help to map vegetation. In other words, each S2 image pixel is managed as a dual composition of spectral patterns instead of a collection of raw reflectance values, which represents the input data using the most discriminative patterns of vegetation from S3.

IV. Conclusion

This letter has presented an intersensor vegetation estimation framework based on topic models to effectively estimate Sentinel-3 (S3) vegetation from Sentinel-2 (S2) data. On the one hand, the S3 OLCI sensor allows obtaining low-resolution vegetation estimations that highlight those areas with more vegetation, due to its smaller spectral bandwidths. On the other hand, the S2 MSI is able to generate higher spatial resolution vegetation maps, but with a different sensitivity to the near-infrared wavelength. On the other hand, standard regression algorithms make use of low-level S2 reflectance values to directly estimate the S3 vegetation, the proposed approach takes advantage of the CpLSA model to discriminate those S2 reflectance patterns that are useful to retrieve S3 vegetation at S2 spatial resolution. Our experiments, conducted using four coupled S2 and S3 data products, reveal that the presented framework provides competitive advantages, from both quantitative and qualitative perspectives, with respect to other regression functions available in the literature. The main conclusion that arises from this letter is the potential of probabilistic topic models to uncover intersensor patterns, useful to estimate S3 vegetation from S2 data. Although our results are quite encouraging, more research work is required in future developments. Specifically, our future work is aimed at extending this work to different sensors, biophysical parameters, and deep fusion architectures.

References

[1] J. Aschbacher and M. P. Milagro-Pérez, "The European Earth monitoring (GMES) programme: Status and perspectives," *Remote Sens. Environ.*, vol. 120, pp. 3–8, May 2012.

[2] Z. Malenovský *et al.*, "Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land," *Remote Sens. Environ.*, vol. 120, pp. 91–101, May 2012.

[3] M. Drusch *et al.*, "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, May 2012.

[4] C. Donlon *et al.*, "The global monitoring for environment and security (GMES) Sentinel-3 mission," *Remote Sens. Environ.*, vol. 120, pp. 37–57, May 2012.

[5] 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, Sep. 2018.

[6] A. Viña, A. A. Gitelson, A. L. Nguy-Robertson, and Y. Peng, "Comparison of different vegetation indices for the remote assessment of green leaf area index of crops," *Remote Sens. Environ.*, vol. 115, no. 12, pp. 3468–3478, Dec. 2011.

[7] J. W. Rouse, R. H. Haas, J. A. Scheli, and D. W. Deering, "Monitoring vegetation systems in the great plains with ERTS," presented at the 3rd ERTS Symp., NASA SP-351, 1973.

[8] A. R. Huete, "A soil-adjusted vegetation index (SAVI)," *Remote Sens. Environ.*, vol. 25, no. 3, pp. 295–309, 1988.

[9] J. Verrelst *et al.*, "Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3," *Remote Sens. Environ.*, vol. 118, pp. 127–139, Mar. 2012.

[10] J. P. R. Caicedo, J. Verrelst, J. Muñoz-Marí, J. Moreno, and G. Camps-Valls, "Toward a semiautomatic machine learning retrieval of biophysical parameters," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1249–1259, Apr. 2014.

[11] J. Verrelst *et al.*, "Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods—A comparison," *ISPRS J. Photogramm. Remote Sens.*, vol. 108, pp. 260–272, Oct. 2015.

[12] R. Fernandez-Beltran and F. Pla, "Incremental probabilistic latent semantic analysis for video retrieval," *Image Vis. Comput.*, vol. 38, pp. 1–12, Jun. 2015.

[13] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Mach. Learn.*, vol. 42, no. 1, pp. 177–196, Jan. 2001.

[14] R. Fernandez-Beltran and F. Pla, "Prior-based probabilistic latent semantic analysis for multimedia retrieval," *Multimedia Tools Appl.*, vol. 77, no. 23, pp. 16771–16793, 2018.

[15] B. D. Craven and S. M. Islam, *Ordinary Least-Squares Regression*. Thousand Oaks, CA, USA: Sage Publications, 2011.

[16] D. Basak, S. Pal, and D. C. Patranabis, "Support vector regression," *Neural Inf. Process. Lett. Rev.*, vol. 11, no. 10, pp. 203–224, 2007.

[17] L. Pasolli, F. Melgani, and E. Blanzieri, "Gaussian process regression for estimating chlorophyll concentration in subsurface waters from remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 3, pp. 464–468, Jul. 2010.

[18] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.*, vol. 8, no. 2, pp. 127–150, May 1979.

[19] R. Fernandez-Beltran, P. Latorre-Carmona, and F. Pla, "Single-frame super-resolution in remote sensing: A practical overview," *Int. J. Remote Sens.*, vol. 38, no. 1, pp. 314–354, 2017.

[20] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *J. Amer. Statist. Assoc.*, vol. 32, no. 200, pp. 675–701, Dec. 1937.

[21] S. Holm, "A simple sequentially rejective multiple test procedure," *Scand. J. Statist.*, vol. 6, no. 2, pp. 65–70, 1979.