

Intelligent Sampling for Vegetation Nitrogen Mapping Based on Hybrid Machine Learning Algorithms

Jochem Verrelst¹, Katja Berger², and Juan Pablo Rivera-Caicedo³

Abstract—Upcoming satellite imaging spectroscopy missions will deliver spatiotemporal explicit data streams to be exploited for mapping vegetation properties, such as nitrogen (N) content. Within retrieval workflows for real-time mapping over agricultural regions, such crop-specific information products need to be derived precisely and rapidly. To allow fast processing, intelligent sampling schemes for training databases should be incorporated to establish efficient machine learning (ML) models. In this study, we implemented active learning (AL) heuristics using kernel ridge regression (KRR) to minimize and optimize a training database for variational heteroscedastic Gaussian processes regression (VHGPR) to estimate aboveground N content. Several uncertainty and diversity criteria were applied on a lookup table (LUT) composed of aboveground N content and corresponding hyperspectral reflectance simulated by the PROSAIL-PRO model. The best-performing AL criteria were Euclidian distance-based diversity (EBD) resulting in a reduction of the LUT training data set by 81% (50 initial samples plus 141 samples selected from a pool of 1000 samples). This reduced LUT was used for training VHGPR, which is not only a competitive algorithm but also provides uncertainty estimates. Validation against *in situ* N reference data provided excellent results with a root-mean-square error (RMSE) of 1.84 g/m² and a coefficient of determination (R^2) of 0.92. Mapping aboveground N content over an agricultural region yielded reliable estimates and meaningful associated uncertainties. These promising results encourage the transfer of such hybrid workflows into space and time within the frame of future operational N monitoring from satellite imaging spectroscopy data.

Index Terms—Active learning (AL), Gaussian processes (GP), hybrid retrieval methods, kernel ridge regression (KRR), nitrogen.

Manuscript received May 29, 2020; revised June 23, 2020; accepted August 1, 2020. This work was supported in part by the European Research Council (ERC) through the ERC-2017-STG SENTIFLEX Project under Grant 755617, in part by the Spanish Ministry of Economy and Competitiveness through the SynerTGE CGL2015-G9095-R (MINECO/FEDER, UE) Project, and in part by the ESA Fluorescence Explorer (FLEX) Calibration/Validation Campaign FLEXsense 2018 under Grant ESA RFP/3-15477/18/NL/NA. The work of Jochem Verrelst was supported by Ramón y Cajal Contract (Spanish Ministry of Science, Innovation and Universities). The work of Katja Berger was supported in part by the German Federal Ministry for Economic Affairs and Energy under Grant 50EE1623 and in part by the Space Agency of the German Aerospace Center (DLR) within the Research Project EnMAP Scientific Advisory Group Phase III: Developing the EnMAP Managed Vegetation Scientific Processor. (Corresponding author: Jochem Verrelst.)

Jochem Verrelst is with the Image Processing Laboratory (IPL), Universitat de València, 46010 València, Spain (e-mail: jochem.verrelst@uv.es).

Katja Berger is with the Department of Geography, Ludwig-Maximilians-Universität München, 80333 Munich, Germany (e-mail: katja.berger@lmu.de).

Juan Pablo Rivera-Caicedo is with the Secretary of Research and Postgraduate, CONACYT-UAN, Tepic 63155, Mexico (e-mail: jprivera@conacyt.mx).

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.2020.3014676

I. INTRODUCTION

WITH current and upcoming satellite imaging spectroscopy missions, unique data streams of hyperspectral measurements from the Earth surface will be provided in almost real time. Agriculture will be one of the key applications where up-to-date information about the crop status and development is required. Sufficient provision and subsequent uptake of nitrogen (N) by the plant influences crop growth and thus yield quality [1]. Crop N mapping from imaging spectroscopy data is considered as an efficient way to enable site-specific fertilization measures and consequently to assure sustainable management and production [2]. Within a plant, N is a major component of amino acids, the building blocks of proteins [3]. Hence, when it comes to N mapping, vegetation N should be derived from proteins rather than from the traditionally used chlorophyll–N relationship [4].

With respect to quantitative methods for retrieving biophysical and biochemical vegetation traits from Earth observation data [5], hybrid workflows have evolved as one of the most promising approaches [6], [7]. These approaches combine physics described by radiative transfer models (RTMs) with the speed and efficiency of machine learning (ML) algorithms. In such a scheme, lookup tables (LUT) are generated from RTM simulations. Then, the ML algorithm learns the (nonlinear) relationship between the pairs of reflectance and vegetation trait of interest. These training databases have to fulfill, on the one hand, the prerequisite of realistically representing the canopy structural and biochemical properties, and on the other hand, being small enough to avoid long computational times required by some ML regression algorithms. Hence, this approach demands a balanced training data set with a tradeoff between optimized accuracy and a minimum amount of samples.

When it comes to hyperspectral data analysis with ML, dimensionality reduction (DR) is a key issue. DR can be accomplished in both spectral and sampling domains [8]. Spectrally, feature engineering and feature extraction methods offer the possibility to reduce data space and thus to remove noise and redundant data. These techniques have been exhaustively analyzed in the context of vegetation properties retrievals from imaging spectroscopy data [8], [9]. Instead, reduction in the sampling domain has been rarely discussed for this purpose and is rather applied within classification [10]. Training ML regression over large randomly sampled data sets can lead to poor prediction performances due to the negative impacts of noise, redundancy, and outliers. Moreover, in particular, kernel-based ML algorithms go along with high computational costs if training data are too abundant, which limits their

applicability within hybrid retrieval schemes [11]. In order to reduce and optimize the available data pool for high training utility, active learning (AL) heuristics can be employed [10], [12]. AL is a subfield of ML seeking to optimize models to improve performance through intelligent sampling of training data sets [13]. In the context of solving regression problems with ML, different query frameworks have been proposed [12], [13] to be grouped into main AL methods of: 1) *diversity*, e.g., [14]; 2) *uncertainty*, e.g., [15]; and 3) *density*, e.g., [16]. So far, diversity and uncertainty selection strategies have been successfully tested by a few studies within terrestrial Earth observation analysis [11], [17], [18]. For instance, AL heuristics were investigated on simulated RTM data, but without evaluating the resulting models on *in situ* reference data [11].

In recent years, several new possibilities have evolved in ML regression with Gaussian processes (GPs) [19], being one of the main interesting kernel-based ML methods for vegetation properties retrievals; GPs excel other ML algorithms through delivering competitive prediction accuracy and their interesting property to provide associated uncertainty intervals of the estimates [20], [21]. In the study of [22], crop N content was estimated by training variational heteroscedastic GP models over a PROSAIL-PRO simulated database with optimized spectral band setting. Still, far too many training samples were used through missing optimization which impacted mapping speed. Previous studies with neural networks and LUT-based inversion even suggested LUT sizes from 8000 to 100 000 combinations of input parameters [23], [24], which represents an unfeasible size for kernel-based ML algorithms. Altogether, AL methods provide an efficient solution for optimizing RTM sampling in view of developing cost-effective kernel-based ML retrieval models. The objective of our study was therefore to propose an intelligent LUT sampling scheme exploiting available AL heuristics for the estimation of vegetation N content. By means of these methods, we expect to enhance mapping speed and, to the best, also N retrieval accuracy. This progress toward optimized training samples may permit to implement the established models within hybrid retrieval workflows in the frame of a future operational N monitoring system from satellite imaging spectroscopy data.

II. MATERIAL AND METHODS

A. Radiative Transfer Modeling

We used the PROSPECT-PRO leaf optical properties model [25] capable of separating leaf dry mass per unit leaf area (LMA) into protein content (Cp) and carbon-based constituents. PROSPECT-PRO was coupled with the 1-D canopy reflectance model Scattering by Arbitrarily Inclined Leaves, 4SAIL [26], to PROSAIL-PRO for generation of an LUT training database. The coupled model simulates reflectance at the canopy scale as a function of diverse biophysical [e.g., leaf area index (LAI) and average leaf inclination angle] and leaf biochemical input parameters (e.g., leaf chlorophyll content, leaf carotenoid content, or leaf equivalent water thickness). Leaf nitrogen content can be directly calculated from Cp with the protein-to-nitrogen conversion factor of 4.43 [27]. Furthermore, LAI is used to upscale from leaf to canopy level. Finally, “aboveground N content” in [g/m²] was calculated in the LUT, as suggested by Berger *et al.* [22]. This study also

provides full information about the generation of the training database (LUT), which was exploited in our study. Briefly, the LUT was established by randomly generating 1000 combinations of PROSAIL-PRO model input parameters (sampling and ranges, see [22]) and simulating corresponding spectral reflectance of ten specific bands. The band selection was based on the property of automatic relevance determination (ARD) covariance in a wrapper strategy using the GP-based band analysis [22]. By means of a sequential backward band removal (SBBR) algorithm [9], bands corresponding to the future satellite mission Environmental Mapping and Analysis Program (EnMAP) were selected, with central band positions at 786, 1556, 1568, 1579, 1623, 1656, 1667, 1762, 2124, and 2234 nm. Hence, with one exception, the best-performing bands are situated in the shortwave infrared (SWIR) domain, where protein absorption occurs [4].

B. AL Heuristics

AL heuristics from the two groups of uncertainty and diversity were applied in our study. Methods have been described before [11] and will be only shortly summarized here: uncertainty criteria select samples with greater disagreements between the different explanatory variables, including variance-based pool of regressors (PAL) [28], entropy query by bagging (EQB) [29], and residual regression AL (RSAL) [30]. In contrast, the group of diversity criterion is based on the principle to include samples that are distant from the available training samples. Hereby, Euclidean distance-based diversity (EBD) [18], angle-based diversity (ABD) [31], and cluster-based diversity (CBD) [32] were tested. In-depth explanation and equations are given in [11]. Moreover, all mentioned AL methods can be accessed and tested within the in-house software package ARTMO (<https://artmtoolbox.com/>). As part of this study, ARTMO’s AL module has been made more user-friendly by enabling: 1) running the AL algorithms against external validation data and 2) allowing users to distinguish between included and nonincluded samples selected by diverse AL methods (MLRA toolbox v.1.25).

C. Validation Data

Data from two different campaigns were pooled together to provide a larger variety of testing data, enhancing the validity of the established models in terms of retrieval accuracy. First, data of the German Munich-North Isar (MNI) campaigns in the North of Munich, in Southern Germany (N 48°16′, E 11°42′) were explored. There, extensive weekly field trials, including field spectroscopy and (non)destructive measurements on winter wheat (*Triticum aestivum*) and corn (*Zea mays*), were carried out during the growing periods of 2017 and 2018 at two test sites. Wheat plants covering an area of 0.25 m² aboveground were cut, weighed, and brought to the lab. In case of corn, three plants were cut and weighed, but only one plant was taken to the lab. Row distance and plants per meter were recorded. In the lab, the combustion method was applied using the vario EL cube device. Samples were oven-dried at 105 °C until the constant dry weight was reached (and determined) after 24 h [22]. Samples were grinded and N concentration (N%), referring to mass of absorbing materials (dry matter) per unit dry mass, given in [mg/g] or [%],

which was measured. N% of each plant organ (leaves, stalks, and fruits) was then converted into aboveground N content in $[\text{g}/\text{m}^2]$ by multiplying N% with plant organ-specific dry mass per unit ground area in $[\text{g}/\text{m}^2]$. From this campaign, 15 N samples were randomly chosen composed of leaves plus stalks N content of wheat and corn.

Second, a data set from the Majadas de Tiétar research station, a Mediterranean tree-grass ecosystem located in central Spain (N $39^\circ 56'$, W $5^\circ 46'$), was exploited. Grass samples were collected in June and July 2018 following the field protocols generated in the SynerTGE project [33], [34]. Nitrogen analyses were performed at the Department of Environment, Spanish National Institute for Agricultural and Food Research (INIA). Destructively sampled grass biomass was analyzed by the dry combustion method to obtain N%. The strategy to obtain canopy N content was different compared to the MNI site. In Majadas, LMA $[\text{g}/\text{cm}^2]$ and LAI $[\text{m}^2/\text{m}^2]$, estimated by manual scanning of grass samples and using gravimetric methods [35], were used to upscale leaf-based N content to the canopy to obtain aboveground N content in $[\text{g}/\text{m}^2]$. Information about the campaign was provided through personal communication with Dr. M. Pilar Martín from SpecLab Laboratory, Spanish National Research Council. From this campaign, 36 N samples were used. Therefore, a total number of 51 N samples was available for testing the validity of the developed method against ground-based *in situ* data. During both campaigns, hyperspectral signatures of the canopy within the 350–2500-nm range were collected at each date of biomass sampling using the Analytical Spectral Devices Inc. (ASD; Boulder, CO, USA) FieldSpec FR3. Spectral settings corresponding to the ten abovementioned selected EnMAP spectral bands (Section II-A, [22]) were configured out of these measurements. In this way, LUT and *in situ* measured spectral samplings were equivalent.

For mapping, an airborne imaging spectroscopy acquisition with the HyMap sensor over Barrax agricultural region, La Mancha, Spain (coordinates $30^\circ 3' \text{N}$, $2^\circ 6' \text{W}$), was used. This flight line is described in various earlier studies [9], [36].

D. Experimental Setup

In a first step, six AL heuristics and random sampling (RS) were tested on the LUT focusing on performance and differences between the AL methods. For this procedure, a kernel ridge regression (KRR) [37] algorithm was applied, which presents the optimal kernel-based method to perform costly simulations [5]. KRR minimizes the squared residuals in a higher dimensional feature space and can be considered as the kernel version of the regularized ordinary least-squares linear regression [38]. The linear regression model is defined in a Hilbert space, \mathcal{H} , of very high dimensionality, where samples were mapped to \mathcal{H} via a mapping function $\phi(\mathbf{x}_i)$. Due to this simplicity, KRR is not only a competitive ML regression method (see [36]), and it is also very fast. Hence, KRR is an ideal ML method to combine with AL to seek for an optimal number of samples. From the pool of 1000 labeled samples (pairs of simulated reflectance and N content), 5% of the data were randomly selected as the initial training data set and the process was repeated for up to 1000 iterations to ensure a low impact of the initial choice and reach statistically reliable results. Following each iteration, a new sample was added to

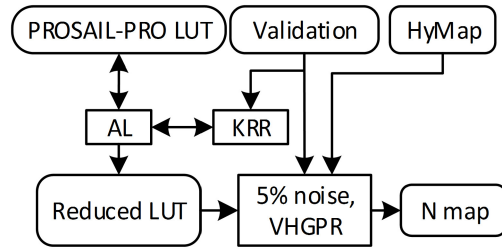


Fig. 1. Workflow of AL mapping strategy for efficient vegetation N mapping.

the training data selected according to diversity or uncertainty criteria of the respective AL method. For instance, in the case of EBD, squared Euclidean distance is calculated. When all distances between samples are computed, the farthest is selected [11]. In our procedure, this new sample is only added when performance improves as evaluated by root-mean-square error (RMSE) against the validation data. The whole process was repeated until all samples of the LUT were evaluated. Finally, the added samples and corresponding goodness-of-fit statistics (e.g., RMSE, R^2) were recorded for each AL method.

The optimal performing samples were then used to train a variational heteroscedastic GP regression (VHGP) model. These models deal with heteroscedastic noise using a marginalized variational approximation [39] and thus may provide most realistic estimates of uncertainty [22], [40]. For both kernel-based ML algorithms, we have used the radial basis function (RBF) kernel. Before training the final VHGP model, we added 5% noise to the simulated spectra, as proposed for hybrid retrieval procedures to generalize the model and to prevent overfitting on the pure (ideal) RTM outputs [7]. Following optimization of the LUT with the AL methods, in a next step, the reduced LUT was used to demonstrate mapping applications. For this, we subsequently trained VHGP and retrained KRR. The applied workflow is conceptualized in Fig. 1.

III. RESULTS

A. AL-Based LUT Selection

Applying AL criteria on the 1000 pool samples required only a few minutes using fast KRR. The AL selection criterion was based on the reduction of RMSE, as can also be observed in Fig. 2 (left). Note that AL algorithms work in a way that a sample is only accepted if it leads to RMSE improvement against the validation data; otherwise, the procedure is stopped. Optimal results were obtained by the EBD method after adding 141 samples to the 50 starting samples (RMSE: $2.81 \text{ g}/\text{m}^2$ and R^2 : 0.76). EBD also belongs to the fastest methods. In the figure, it can also be noted that other methods found even a fewer number of samples, but they were not able to reach the low errors as obtained by EBD. It should hereby be remarked that a lowering of the RMSE does not necessarily go along with an improvement of R^2 , as can be viewed in Fig. 2 (right). Although the R^2 patterns are more irregular, it follows the general trends as RMSE, with the highest R^2 obtained by the EBD method. The second best result was achieved by PAL that added the same amount of samples (141) but with a slightly poorer accuracy (RMSE: $3.00 \text{ g}/\text{m}^2$ and R^2 : 0.77). Yet, PAL needs the longest runtime, which renders

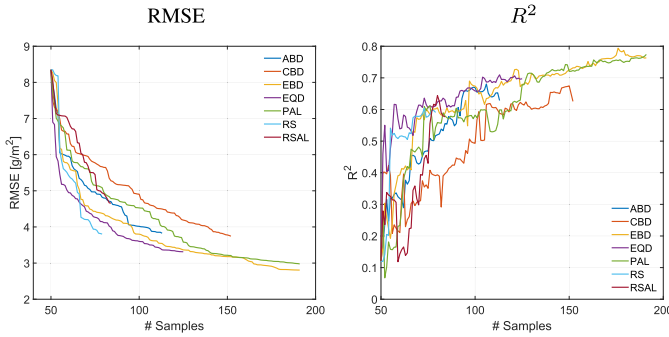


Fig. 2. (Left) RMSE and (Right) R^2 for N retrieval using six different AL methods and RS applied on a PROSAIL-PRO LUT with KRR.

TABLE I

GOODNESS-OF-FIT STATISTICS OF KRR AND VHGPB MODELS (AS TRAINED WITH EBD-REDUCED LUT AND 5% NOISE ADDED) AGAINST VALIDATION DATA AND CPU TIME FOR TRAINING AND TESTING (SECONDS)

Model	RMSE (g/m^2)	NRMSE(%)	R^2	train (s)	test (s)
KRR	4.00	20.67	0.90	0.14	0.0006
VHGPR	1.84	9.55	0.92	2.48	0.0024

it less interesting for operational retrieval. For instance, PAL needed 4.7 times longer for sample selection than the EBD method. To demonstrate the efficiency of the AL algorithms, RS procedure was applied for comparison. Hereby, each time a random sample was added, which initially led to rapid improvement of prediction accuracy but was unable to improve further after only 29 samples were added (RMSE: 3.80 g/m^2 and R^2 : 0.59). Except for RSAL and ABD, optimized LUTs by all AL methods achieved increased N prediction accuracy.

B. Application of Optimized Sampling for N Sensing

Following optimization of the LUT with the EBD method, the next step was to use LUT for mapping applications. To render the LUT more suitable for processing real images, 5% noise was added to the simulated reflectance. Goodness-of-fit results and processing times are provided in Table I. A first observation is the substantial improvement using VHGPB as opposed to the earlier KRR results without adding noise (see Section III-A). This is in line with [7], where the role of noise was found crucial in optimizing hybrid approaches. Another interesting observation is the significant improvement of error measures [RMSE or Normalized RMSE (NRMSE)] of VHGPB as opposed to KRR with noise (RMSE: 4.00 versus 1.84 g/m^2), although R^2 results of KRR and VHGPB are alike (0.90 versus 0.92). Hence, these results underline the gain in accuracy that can be achieved by using a more advanced ML algorithm. However, the improved performances of VHGPB go along with some extra computational cost: training needed about 18 times longer, and image processing (mapping) with VHGPB was four times slower than using KRR. Still, VHGPB runs very fast with the EBD-reduced LUT, generating maps in the order of seconds. In comparison, in our previous study using the same approach without implementing AL heuristics, a slightly lower accuracy was obtained with a similar data set (only MNI, RMSE of 2.1 g/m^2) and a longer runtime [22].

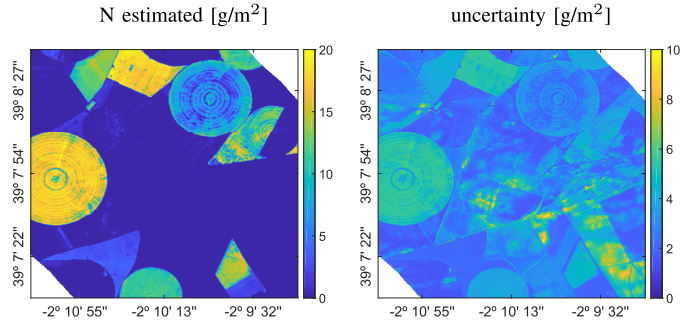


Fig. 3. (Left) Zoom-in of the HyMap flightline showing vegetation N content in $[\text{g/m}^2]$ over Barrax with estimates and (Right) absolute uncertainties in form of standard deviations.

C. Mapping Nitrogen Content Over Agricultural Areas

Finally, the VHGPB model was applied to the exemplary HyMap flight line over an agricultural area in Barrax, Spain. Since the model is so light, processing of the flight line took merely 30 s. When instead applying a VHGPB model trained over the full LUT, mapping took about four times longer. Moreover, VHGPB being developed in a Bayesian framework provides associated uncertainty estimates along with the mean estimates. As such, information on the per-pixel performance is obtained. A zoom-in of the estimates and uncertainties is provided to visualize the obtained maps (see Fig. 3). The green crops on irrigated parcels show a pronounced N content. Over the fallow lands, N content was only estimated where crop residues may have been present. The uncertainty estimates give a more nuanced impression, especially over bare soil suggesting to interpret bare soil areas with higher uncertainties. This is due to the fact that PROSAIL-PRO generated training data mainly consisted of vegetative spectra. Bare soil signatures present in the LUT may not completely reflect real-world conditions. To overcome this drawback, uncertainty maps could be used to reveal where bare soil spectra can be extracted and added to the VHGPB training data set. In comparison, when repeating the mapping with a randomly sampled subset of the same size, it not only led to poorer validation results (RMSE: 4.65 g/m^2 and R^2 : 0.87), but also the obtained map was characterized by higher uncertainties (results not shown).

IV. CONCLUSION

In this study, we showed that intelligent sampling using AL efficiently reduces the total number of LUT entries, which significantly speeds up the training process and running of kernel-based ML regression algorithms. Hence, this would lead to fast and highly efficient models, especially valid when it comes to process heavy imageries (either in spatial or spectral dimension). Variational heteroscedastic GP regression models trained on optimized RTM-generated LUTs can be the core of next-generation operational hybrid retrieval schemes using satellite imaging spectroscopy. In this context, the combination of DR in both the spectral and in the sampling domain is highly encouraged. In this way, LUT variability for training can be optimized and long processing times avoided. Within the ARTMO framework, a suite of leaf and canopy RTMs are available that can be used for simulating any input-output combination. Hence, abundant training data pools can be generated for a wide range of vegetation properties.

This work has extended this possibility by providing an updated AL module within ARTMO. It facilitates the development of hybrid retrieval models based on intelligent training data sets coming from RTM simulations and advanced ML regression algorithms. With the upcoming massive availability of hyperspectral imagery and the current tendency of data processing in cloud-computing platforms such as Google Earth Engine or Amazon Web Service, it becomes imperative to develop light models that make it possible to process big Earth observation data within a reasonable time.

ACKNOWLEDGMENT

The authors would like to thank the colleagues from SpecLab-CSIC, INIA, MPI-BGC, UEX, and CEAM who contributed to the acquisition and processing of the N data set for Majadas de Tiétar. They would also like to thank Matthias Woher and LMU team for organizing MNI campaigns.

REFERENCES

- [1] F. Baret, V. Houles, and M. Guerif, "Quantification of plant stress using remote sensing observations and crop models: The case of nitrogen management," *J. Experim. Botany*, vol. 58, no. 4, pp. 869–880, Nov. 2006.
- [2] T. B. Hank *et al.*, "Spaceborne imaging spectroscopy for sustainable agriculture: Contributions and challenges," *Surv. Geophys.*, vol. 40, no. 3, pp. 515–551, May 2019.
- [3] R. F. Kokaly, "Investigating a physical basis for spectroscopic estimates of leaf nitrogen concentration," *Remote Sens. Environ.*, vol. 75, no. 2, pp. 153–161, Feb. 2001.
- [4] K. Berger *et al.*, "Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions," *Remote Sens. Environ.*, vol. 242, Jun. 2020, Art. no. 111758.
- [5] J. Verrelst *et al.*, "Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties—A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 108, pp. 273–290, Oct. 2015.
- [6] J. Verrelst *et al.*, "Quantifying vegetation biophysical variables from imaging spectroscopy data: A review on retrieval methods," *Surv. Geophys.*, vol. 40, no. 3, pp. 589–629, May 2019.
- [7] B. Brede *et al.*, "Assessment of workflow feature selection on forest LAI prediction with sentinel-2A MSI, landsat 7 ETM+ and landsat 8 OLI," *Remote Sens.*, vol. 12, no. 6, p. 915, Mar. 2020.
- [8] J. P. Rivera-Caicedo, J. Verrelst, J. Muñoz-Marí, G. Camps-Valls, and J. Moreno, "Hyperspectral dimensionality reduction for biophysical variable statistical retrieval," *ISPRS J. Photogramm. Remote Sens.*, vol. 132, pp. 88–101, Oct. 2017.
- [9] J. Verrelst, J. P. Rivera, A. Gitelson, J. Delegido, J. Moreno, and G. Camps-Valls, "Spectral band selection for vegetation properties retrieval using Gaussian processes regression," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 52, pp. 554–567, Oct. 2016.
- [10] M. M. Crawford, D. Tuia, and H. L. Yang, "Active learning: Any value for classification of remotely sensed data?" *Proc. IEEE*, vol. 101, no. 3, pp. 593–608, Mar. 2013.
- [11] J. Verrelst, S. Dethier, J. P. Rivera, J. Muñoz-Marí, G. Camps-Valls, and J. Moreno, "Active learning methods for efficient hybrid biophysical variable retrieval," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 7, pp. 1012–1016, Jul. 2016.
- [12] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *J. Artif. Intell. Res.*, vol. 4, pp. 129–145, Mar. 1996.
- [13] B. Settles, "Active learning literature survey," Dept. Comput. Sci., Univ. Wisconsin–Madison, Madison, WI, USA, Tech. Rep. 1648, 2010.
- [14] X. Lu, J. Zhang, T. Li, and Y. Zhang, "Incorporating diversity into self-learning for synergistic classification of hyperspectral and panchromatic images," *Remote Sens.*, vol. 8, no. 10, p. 804, Sep. 2016.
- [15] T. He *et al.*, "An active learning approach with uncertainty, representativeness, and diversity," *Sci. World J.*, vol. 2014, pp. 1–6, Aug. 2014.
- [16] B. Demir and L. Bruzzone, "A multiple criteria active learning method for support vector regression," *Pattern Recognit.*, vol. 47, no. 7, pp. 2558–2567, Jul. 2014.
- [17] D. Upreti *et al.*, "A comparison of hybrid machine learning algorithms for the retrieval of wheat biophysical variables from Sentinel-2," *Remote Sens.*, vol. 11, no. 5, p. 481, Feb. 2019.
- [18] F. Douak, F. Melgani, N. Alajlan, E. Pasolli, Y. Bazi, and N. Benoudjit, "Active learning for spectroscopic data regression," *J. Chemometrics*, vol. 26, no. 7, pp. 374–383, Jul. 2012.
- [19] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. New York, NY, USA: MIT Press, 2006.
- [20] J. Verrelst, L. Alonso, G. Camps-Valls, J. Delegido, and J. Moreno, "Retrieval of vegetation biophysical parameters using Gaussian process techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1832–1843, May 2012.
- [21] 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.
- [22] K. Berger *et al.*, "Retrieval of aboveground crop nitrogen content with a hybrid machine learning method," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 92, Oct. 2020, Art. no. 102174.
- [23] M. Weiss, F. Baret, R. B. Myneni, A. Pragnère, and Y. Knyazikhin, "Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data," *Agronomie*, vol. 20, no. 1, pp. 3–22, Jan. 2000.
- [24] B. Combal, F. Baret, and M. Weiss, "Improving canopy variables estimation from remote sensing data by exploiting ancillary information. Case study on sugar beet canopies," *Agronomie*, vol. 22, no. 2, pp. 205–215, Mar. 2002.
- [25] J.-B. Féret, K. Berger, F. de Boissieu, and Z. Malenovsky, "PROSPECT-PRO for estimating content of nitrogen-containing leaf proteins and other carbon-based constituents," 2020, *arXiv:2003.11961*. [Online]. Available: <http://arxiv.org/abs/2003.11961>
- [26] W. Verhoef and H. Bach, "Coupled soil–leaf–canopy and atmosphere radiative transfer modeling to simulate hyperspectral multi-angular surface reflectance and TOA radiance data," *Remote Sens. Environ.*, vol. 109, no. 2, pp. 166–182, Jul. 2007.
- [27] H.-H. Yeoh and Y.-C. Wee, "Leaf protein contents and nitrogen-to-protein conversion factors for 90 plant species," *Food Chem.*, vol. 49, no. 3, pp. 245–250, Jan. 1994.
- [28] F. Douak, F. Melgani, and N. Benoudjit, "Kernel ridge regression with active learning for wind speed prediction," *Appl. Energy*, vol. 103, pp. 328–340, Mar. 2013.
- [29] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Muñoz-Marí, "A survey of active learning algorithms for supervised remote sensing image classification," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 3, pp. 606–617, Jun. 2011.
- [30] F. Douak, N. Benoudjit, and F. Melgani, "A two-stage regression approach for spectroscopic quantitative analysis," *Chemometric Intell. Lab. Syst.*, vol. 109, no. 1, pp. 34–41, Nov. 2011.
- [31] B. Demir, C. Persello, and L. Bruzzone, "Batch-mode active-learning methods for the interactive classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1014–1031, Mar. 2011.
- [32] S. Patra and L. Bruzzone, "A cluster-assumption based batch mode active learning technique," *Pattern Recognit. Lett.*, vol. 33, no. 9, pp. 1042–1048, Jul. 2012.
- [33] M. P. Martín *et al.*, "Estimation of essential vegetation variables in a dehesa ecosystem using reflectance factors simulated at different phenological stages," *Revista de Teledetección*, vol. 55, no. 55, pp. 31–48, 2020. [Online]. Available: <https://polipapers.upv.es/index.php/raet/article/view/13394>
- [34] J. Melendo-Vega *et al.*, "Improving the performance of 3-D radiative transfer model FLIGHT to simulate optical properties of a tree-grass ecosystem," *Remote Sens.*, vol. 10, no. 12, p. 2061, Dec. 2018.
- [35] G. Mendiguren, M. Pilar Martín, H. Nieto, J. Pacheco-Labrador, and S. Jurdao, "Seasonal variation in grass water content estimated from proximal sensing and MODIS time series in a Mediterranean Fluxnet site," *Biogeosciences*, vol. 12, no. 18, pp. 5523–5535, Sep. 2015.
- [36] 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.
- [37] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, Jun. 1999.
- [38] B. Schölkopf and A. Smola, *Learning With Kernels—Support Vector Machines, Regularization, Optimization & Beyond*. Cambridge, MA, USA: MIT Press, 2002.
- [39] M. Lazaro-Gredilla, M. K. Titsias, J. Verrelst, and G. Camps-Valls, "Retrieval of biophysical parameters with heteroscedastic Gaussian processes," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 4, pp. 838–842, Sep. 2013.
- [40] G. Camps-Valls, J. Verrelst, J. Muñoz-Marí, V. Laparra, F. Mateo-Jimenez, and J. Gomez-Dans, "A survey on Gaussian processes for Earth-observation data analysis: A comprehensive investigation," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 58–78, Jun. 2016.