Active Learning Methods for Efficient Hybrid Biophysical Variable Retrieval

Jochem Verrelst, Sara Dethier, Juan Pablo Rivera, Jordi Muñoz-Marí, Gustau Camps-Valls, and José Moreno

Abstract—**Kernel-based machine learning regression algorithms (MLRAs) are potentially powerful methods for being implemented into operational biophysical variable retrieval schemes. However, they face difficulties in coping with large training data sets. With the increasing amount of optical remote sensing data made available for analysis and the possibility of using a large amount of simulated data from radiative transfer models (RTMs) to train kernel MLRAs, efficient data reduction techniques will need to be implemented. Active learning (AL) methods enable to select the most informative samples in a data set. This letter introduces six AL methods for achieving optimized biophysical variable estimation with a manageable training data set, and their implementation into a Matlab-based MLRA toolbox for semiautomatic use. The AL methods were analyzed on their efficiency of improving the estimation accuracy of the leaf area index and chlorophyll content based on PROSAIL simulations. Each of the implemented methods outperformed random sampling, improving retrieval accuracy with lower sampling rates. Practically, AL methods open opportunities to feed advanced MLRAs with RTM-generated training data for the development of operational retrieval models.**

Index Terms—**Active learning (AL), hybrid retrieval methods, kernel methods, machine learning regression algorithms (MLRAs), PROSAIL, Sentinel-3.**

I. INTRODUCTION

SPATIOTEMPORAL quantitative retrieval methods for the Earth's surface are a requirement in a variety of Earth system applications. Optical Earth observing satellites, endowed with high temporal resolutions, enable the retrieval and, hence, monitoring of biogeophysical variables [1]. With the forthcoming superspectral Copernicus Sentinel-2 and Sentinel-3 missions, as well as the planned hyperspectral EnMAP, HyspIRI, PRISMA, and European Space Agency (ESA) candidate FLEX missions, an unprecedented data stream for land monitoring will soon become available to a diverse user community. This vast data stream will require enhanced processing techniques that are accurate, robust, and fast.

One of the major challenges in this respect is to cope with the large amount of data that has to be processed. Over the last few decades, a wide diversity of biophysical variable retrieval methods have been developed, but only a few of them made it into operational use, and many of them are still in their infancy [2]. Generally, there are four main approaches to the problem of estimating biophysical variables from reflectance spectra, i.e., 1) parametric regression; 2) nonparametric regression; 3) physically based, i.e., inverting radiative transfer models (RTMs); and 4) hybrid methods. Hybrid methods combine the generic properties of physically based models combined with the flexibility and computational efficiency of nonparametric nonlinear regression models [2]. Hybrid retrieval schemes proved to be successful in operational processing chains. Current hybrid schemes for the generation of land products such as leaf area index (LAI) rely on neural networks (NNs), typically trained using a very large amount of RTM-simulated data [3], [4]. For LAI retrievals for instance, the most widely used RTM is PROSAIL [5].

Apart from NNs, for the last decade, various alternative nonparametric methods in the field of machine learning regression algorithms (MLRAs) have been introduced, many of them with interesting properties. In particular, kernel-based MLRAs, such as kernel ridge regression (KRR) [6], have proven to be simple for training and yield competitive accuracy [7]. Gaussian process regression (GPR) can even provide associated uncertainty intervals for the estimations [8], [9]. They both could become powerful alternatives to replace NNs in hybrid retrieval schemes [10]. However, these advanced MLRAs go at a computational cost, typically scaling poorly with the number of samples. This is currently the main shortcoming for their adoption in hybrid schemes for retrieval.

The problem when using RTM-simulated data is that these data can easily scale up to several millions of samples corresponding to different variable selections, particularly when lookup tables (LUTs) are used and all input variables are systematically combined. Simulated data created in this way might be redundant, as slight variable variations might produce very similar spectra. Therefore, in order to have these kernel-based MLRAs trained by large data sets, we need intelligent sampling approaches that select the most informative samples from a large pool data set. In the spectral domain, we can use feature extraction techniques [11] to reduce these large data sets. In the sampling domain, we propose to use active learning (AL) techniques [12], [13] to obtain an optimized training set while keeping its size small enough. This letter analyzes the utility of those AL techniques for the benefit of developing efficient hybrid retrieval schemes.

This brings us to the following main objective: to introduce AL techniques that deal efficiently with RTM-simulated data to

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

2   IEEE GEOSCIENCE AND REMOTE SENSING LETTERS

optimize kernel-based MLRA hybrid retrieval strategies. The subobjectives are as follows: 1) to implement AL methods into a software framework that facilitates the use of these methods into hybrid retrieval strategies and 2) to evaluate the efficiency of AL methods in optimizing hybrid PROSAIL-MLRA models with a manageable training data set. This analysis is illustrated for the retrieval of LAI and leaf chlorophyll content (LCC) from simulated Sentinel-3 OLCI data.

II. AL Theory

Let us assume that we have a large (simulated) training data set composed of many samples, so large that it is unfeasible to use all the samples to obtain a regression model. In order to reduce its size, the typical naive approach is to select some samples from it randomly. This strategy does not necessarily optimize the selection since all samples, including redundant and noisy samples, are treated equally. AL [12] can be used to solve this shortcoming using smart sample selection criteria to improve data compactness, diversity and richness, and, hence, the model's estimation accuracy. AL techniques aim to select the most informative samples in a training data set to reach high accuracy using fewer labeled samples. While widely used in data classification [13]–[16], there are few experimental studies carried out for regression problems, and these are focused on the estimation of terrestrial [17] and ocean biophysical [18], [19] variables from remote sensing data. None of these aims to be generic at larger scales via the training of RTM simulations. Accordingly, the application of AL techniques to smartly sample large simulated training data sets in hybrid schemes creates the opportunity to improve the prediction accuracy of the resulting regression model.

AL methods start with a small training set of labeled data—in this case, reflectance-variable pairs—and use query strategies, or selection criteria, to select samples from a larger unlabeled data pool, i.e., one in which samples have no assigned variable value. The selected samples are labeled and are added to the training set until it becomes optimal [19]. In classification, the samples are typically labeled by a human expert. The use of simulated training data removes the need for a human expert as the labels are taken from the LUT, thus creating an automated optimized sampling solution. The optimal final training set should be small enough to substantially increase the computational efficiency of the final prediction model and large enough to accurately represent the original training set and give highly accurate predictions. The selection criteria select the most significant training samples from the pool, i.e., those which, if labeled, would improve the regression model the most.

The selected criterion algorithms can rank the samples according to the *uncertainty* of a sample or its *diversity* [15]. These criteria are sometimes used together within classification problems [20] and are here applied separately to regression. The separate use of these criteria in this letter already provides benefits, although it is possible to combine them as in classification problems. Selecting samples by uncertainty picks the most uncertain samples, i.e., those with the least confidence. Uncertainty criteria include variance-based pool of regressors (PAL) [19], entropy query by bagging (EQB) [13], and residual regression AL (RSAL) [21]. Selecting samples by diversity ensures that added samples are dissimilar from those already accounted for. Diversity criteria include Euclidean distance-based diversity (EBD) [19], angle-based diversity (ABD) [22], and cluster-based diversity (CBD) [23]. The algorithms are briefly described hereinafter.

A. Uncertainty Criteria Methods

1) *PAL*: This strategy [19] first generates k subsets by randomly choosing samples from the original training set. Each subset is then used to train a regressor and to obtain a prediction for each sample in the candidate set. This ends up with k different predictions for each candidate sample. Then, the variance of each prediction is estimated as

$$\sigma_y^2 = \frac{1}{k} \sum_{i=1}^{k} (y_i - \bar{y})^2 \qquad (1)$$

where $\bar{y} = 1/k \sum_{i=1}^{k} y_i$. The variance gives an indication of the spread of the estimations. The samples with the highest variance, i.e., greater disagreements between the different regressors, are added to the training set and removed from the candidate set.

2) *EQB*: In this approach [13], the predictions of k different regressors are ranked according to their entropy

$$H(x) = - \sum_{i=1}^{k} p(x_i) \log p(x_i) \qquad (2)$$

where $p(x_i)$ is the probability of the sample x being predicted by the regressor i. Samples for which various regressors give similar values have lower uncertainties, and this is indicated by smaller or more negative entropy values. $H(x)$ is computed for each sample, and then, the samples with the greatest entropy are added to the training set.

3) *RSAL*: Adapted from [21], this method quantifies the systematic errors generated by a regression algorithm. It does so by training a second model (residual model), which estimates the prediction errors, $e(x) = y - \hat{y}$, where y is the actual observed value and $\hat{y} = \hat{f}(x)$ is the model prediction given the input x. The algorithm selects the samples that exhibit a high prediction error and adds these to the training set.

B. Diversity Criteria Methods

1) *EBD*: This method [19] selects the samples in the candidate set that are distant from the current training set using their squared Euclidean distance

$$d_E = \|x_u - x_l\|_2^2 \qquad (3)$$

where x_u is a sample from the candidate set and x_l is a sample from the training set. All distances between samples are computed, and then, the farthest are selected.

2) *ABD*: This strategy [22] measures the diversity between samples using the cosine angle distance, defined as

$$\angle(x_u, x_l) = \cos^{-1} \left(\frac{\langle x_u, x_l \rangle}{\|x_u\| \cdot \|x_l\|} \right) \qquad (4)$$

where $\langle x_u, x_l \rangle$ is the inner product between x_u and x_l. The learning samples showing the largest cosine angles with the training data are added to the training set.

TABLE I
RANGE AND DISTRIBUTION OF INPUT VARIABLES USED TO ESTABLISH THE SYNTHETIC CANOPY REFLECTANCE DATABASE FOR USE IN THE LUT

Model variables		Units	Range	Distribution
Leaf variables: PROSPECT-4				
N	Leaf structure index	unitless	1.3-2.5	Uniform
LCC	Leaf chlorophyll content	$[\mu g/cm^2]$	5-75	Gaussian ($\bar{x}$: 35, SD: 30)
C_m	Leaf dry matter content	$[g/cm^2]$	0.001-0.03	Uniform
C_w	Leaf water content	[cm]	0.002-0.05	Uniform
Canopy variables: 4SAIL				
LAI	Leaf area index	$[m^2/m^2]$	0.1-7	Gaussian ($\bar{x}$: 3, SD: 2)
α_{soil}	Soil scaling factor	unitless	0-1	Uniform
ALA	Average leaf angle	[∘]	40-70	Uniform
HotS	Hot spot parameter	[m/m]	0.05-0.5	Uniform
skyl	Diffuse incoming solar radiation	[fraction]	0.05	-
θ_s	Sun zenith angle	[∘]	22.3	-
θ_v	View zenith angle	[∘]	20.19	-
ϕ	Sun-sensor azimuth angle	[∘]	0	-

Similar variable ranges/values/distributions were used according to field configurations and related studies [27]. $\bar{x}$: mean, SD: standard deviation.

3) *CBD*: This method [23] first groups the data using a clustering algorithm, for instance, k-means. The number of clusters k is set to the number of samples to add in each iteration of the AL algorithm. For each cluster, the nearest sample to the cluster centroid is selected.

III. ARTMO TOOLBOX

This study was conducted within an in-house developed graphical user interface (GUI) software package named Automated Radiative Transfer Models Operator (ARTMO) [24]. ARTMO embodies a suite of leaf and canopy RTMs, including PROSAIL (i.e., the leaf model PROSPECT coupled with the canopy model SAIL [5]) and several retrieval toolboxes, i.e., a spectral index toolbox, a LUT-based inversion toolbox, and an MLRA toolbox.

The MLRA retrieval toolbox [25] offers a suite of regression algorithms that enable the estimation of biophysical variables based on either experimental or simulated data with little user interaction. The toolbox is built around the SimpleR package [26] with over 15 MLRAs and additionally includes feature reduction techniques such as principal component analysis (PCA) and nonlinear alternatives (e.g., kernel PCA) along with cross-validation subsampling techniques. The option of applying AL methods was implemented for the first time in the latest MLRA toolbox version (v1.17). A GUI has been developed which allows selecting one or multiple AL methods, as well as the selection of a starting data set, the number of samples to add per iteration, and the stopping criterion. A goodness-of-fit validation overview table allows the user to select the best performing hybrid method, which then can be used to process remote sensing images efficiently. The ARTMO package runs in MATLAB and can be freely downloaded from http://ipl.uv.es/artmo/.

IV. EXPERIMENTAL SETUP

A. RTM Simulations

PROSAIL-simulated spectral and input variable data were used to test and compare the AL techniques on different kernel regression models. Sensor settings corresponding to S3-OLCI were chosen over 18 bands from 443 to 940 nm. Different selections of variables such as solar angles and leaf

characteristics were selected in order to create a generic data set by combining the ranges of all variables, as shown in Table I. The resulting data set contained 218 750 000 PROSAIL-simulated reflectance-parameter pairs and was stored in a LUT in ARTMO. This data set, which is overly big and largely redundant for regression, was randomly sampled into subsets of different sizes ranging between 1000 and 100 000 samples. The various subsets were used to investigate when the computational processing of the kernel regressions failed, using a random training/validation splitting of 50%.

B. MLRA Configuration

Two kernel-based MLRAs were selected: KRR and GPR. These methods have been amply described before (e.g., [2]) and are here only briefly summarized.

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 [28]. The linear regression model is defined in a Hilbert space $\mathcal{H}$ of very high dimensionality, where samples have been mapped through a mapping $\phi(\mathbf{x}_i)$. We have used the radial-basis-function kernel function.

GPR has recently been introduced as a powerful regression tool [8]. The model provides a probabilistic approach for learning generic regression problems with kernels. The GPR model establishes a relation between the input and the output variable in the same way as KRR. However, two additional advantages of GPR must be noted. First, not only a predictive mean but also a predictive variance can be obtained, which can serve as an uncertainty interval. The second advantage is that one can use very sophisticated kernel functions because hyperparameters can be learned efficiently by maximizing the marginal likelihood in the training set [8], [9]. In this letter, a standard RGB kernel function was used.

C. AL Sampling Design

Although KRR and GPR proved to deliver high accuracies and were fast in processing, they incurred a computational cost, thus limiting the number of input training samples. Initial tests revealed that the largest set that could be processed under a reasonable amount of time for both regression algorithms was
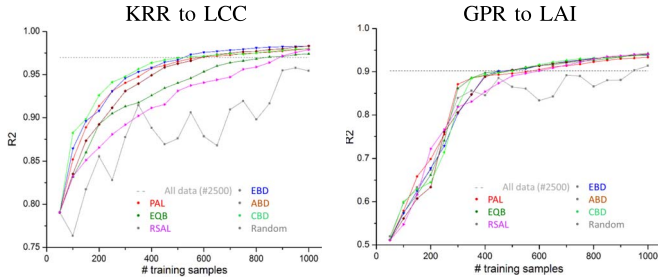
Fig. 1. Mean R^2 using AL methods with (left) KRR for LCC estimation and (right) GPR for LAI estimation.

Fig. 2. (Left) Number of samples needed to reach maximal accuracies, and (right) associated processing time for (top) LCC and (bottom) LAI.

the set containing 5000 samples of simulated RTM data; thus, this was used to analyze the AL methods.

The 5000 samples were separated into an initial learning set of 2500 samples and a validation set of 2500 samples. The GPR and KRR models were first trained and validated on those data sets, further referred to as full training models, and were used as reference models. Subsequently, an initial random training set of 50 samples of spectral data was selected from the learning set, followed by MLRA training and validation. Samples from the learning set were then added to the training set using the AL methods in batches of 50 samples per iteration until no more samples were left in the learning set. The labels at each iteration were taken from the LUT containing the simulated reflectance-parameter pairs. This was repeated ten times with a different initial training set each time to reach statistically reliable results. The number of pools used in the PAL and EQB methods (k) was determined by comparing their predictive capability and processing time using 5 and 10 pools in the initial testing phase. Since they both gave similar results, the number of pools was fixed to 5 in order to be faster. Finally, the averaged coefficients of determination (R^2) values of the ten runs over the first 19 iterations, i.e., to reach 1000 training samples, were plotted. This procedure was completed using KRR and GPR to retrieve LAI and LCC values, although because of leading to similar results, only LCC by KRR and LAI by GPR results are addressed here.

V. RESULTS

Fig. 1 shows the average R^2 performances of the AL methods over ten runs against validation data, each time until a training set of 1000 is reached. The dashed line represents the R^2 accuracy of the full reference training data set containing 2500 samples. The gray line represents the performance of the random sampling sequence, which is added to appreciate the benefit of AL. The colored lines represent the six AL methods considered. For both LCC and LAI, the random sampling technique reduces the uncertainty in the prediction model as more samples are added, but it is unstable as it leads to irregular convergence curves and is outperformed by the AL methods in terms of convergence rate. All AL methods surpass the reference training data set before reaching 1000 training samples.

Regarding LCC retrieval with KRR, the diversity AL methods EBD, ABD, and CBD, along with the uncertainty method PAL, led to similarly good results. These methods reached the accuracy of the full training with about 500 samples, a fifth of the original data set size. Moreover, these methods led to better accuracies than with the training of the whole 2500 learning
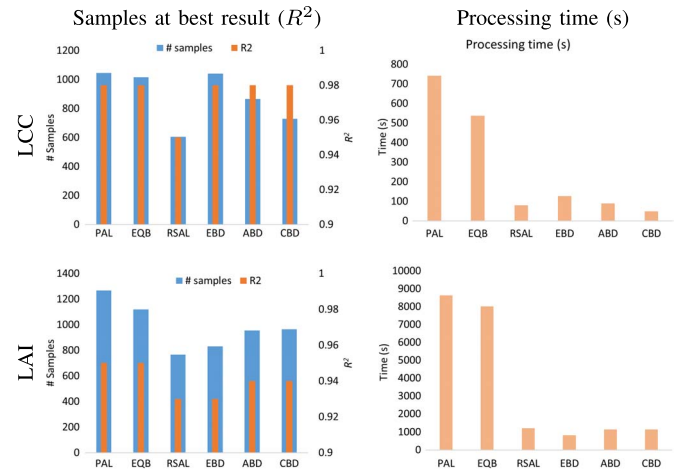
samples. The uncertainty methods RSAL and EQB resulted in poorer results.

Regarding LAI retrieval with GPR, similarly to the LCC case, the use of AL methods led to retrieval accuracies superior to that resulting from the full training data set. The results converged faster here, with 400 samples needed to reach the full training accuracy. Random sampling led to similar results to the AL methods until 400 samples, but it then became unstable. Contrary to the aforementioned LCC case, RSAL performed better, leading to R^2 values similar to that of the other AL methods. In fact, all AL techniques showed similar accuracy results. The fastest converging method was PAL until 400 samples, but past this, CBD and EBD led to slightly superior accuracy results, although the overall difference between them is negligible.

Finally, the number of samples needed to reach maximal retrieval accuracy and their corresponding processing time were recorded (see Fig. 2). The diversity methods ABD and CBD processed fastest to reach maximized accuracies. Since these two diversity methods also led to high accuracies with relatively small training sets, they seem to be the best methods for operational applications. EBD and RSAL were also fast, but EBD performed unstably when comparing LCC against LAI and RSAL led to the poorest accuracies. PAL and EQB, although yielding high accuracies, took the longest time to compute since these methods created five regression models to be compared.

VI. CONCLUSION AND OUTLOOK

Kernel-based MLRAs allow for fast and accurate variable predictions, and they possess attractive properties to become the core of next-generation operational hybrid retrieval schemes [2]. However, most of them cannot process large training databases. For instance, here, it was found that standard implementations of GPR cannot typically cope with more than 5000 samples, i.e., a biophysical variable associated with superspectral or hyperspectral data, under reasonable time, which hampers its direct application toward generic retrieval models. To mitigate this problem, the use of six AL methods was explored in this letter. AL methods are promising to crystallize

an optimal training sampling set that is manageable by kernel-based MLRA methods.

For two kernel-based MLRA methods, i.e., GPR and KRR, we demonstrated the utility of AL methods as an intelligent sampling step that selects the best possible samples from a large (simulated) training data set. By using a data pool of PROSAIL-simulated LAI, LCC, and S3-OLCI spectra, each AL method converged faster to the lower error bound than a random sampling strategy. Diversity criteria (EBD, ABD, and CBD) performed generally fastest, both in terms of reaching high accuracies as well in processing time. Random sampling converged in the long run to similar accuracies but did not offer a stable error minimization over few samples. The reason that each of the AL methods smoothly converged to the full training error bound is that the methods only allow the addition of samples that improve the overall accuracy of the model. These observations using simulated training data confirm earlier results that were obtained in experimental regression problem studies [17], [19]. This demands for a follow-up study to combine uncertainty and diversity criteria. Another attractive path to pursue is to select samples according to a density criterion [29].

This work led to an updated ARTMO MLRA toolbox with the ability to apply AL methods into hybrid retrieval models in a semiautomatic manner. Combining both AL methods with MLRAs makes the toolbox particularly powerful in processing optical remote sensing images into vegetation products. It opens possibilities toward the development of powerful (kernel-based) and generic hybrid retrieval schemes, e.g., based on RTM simulations. However, to pursue with developing hybrid strategies for an operational setting would imply introducing a larger training data set, which requires more samples to be added per AL iteration. A generic validation data set would then be required to ascertain the minimum amount of samples that stabilizes the retrieval performance. Another recommendation is to feed the AL methods not only with simulated data but also with data collected in the field or coming from earlier products. For instance, ESA's current LAI product comes from an NN that is trained by earlier LAI products [4]. Also, while the initial training set was chosen in a random way, a more sophisticated initialization strategy may further improve the AL sequence. Eventually, the freely available MLRA toolbox will allow users to develop their own robust, compact, accurate, and generically applicable retrieval algorithms with little user interaction.

REFERENCES

[1] M. E. Schaepman et al., "Earth system science related imaging spectroscopy-An assessment," Remote Sens. Environ., vol. 113, pp. S123–S137, 2009.

[2] 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, 2015.

[3] F. Baret et al., "LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. Part 1: Principles of the algorithm," Remote Sens. Environ., vol. 110, no. 3, pp. 275–286, 2007.

[4] F. Baret et al., "GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part1: Principles of development and production," Remote Sens. Environ., vol. 137, pp. 299–309, 2013.

[5] S. Jacquemoud et al., "PROSPECT+SAIL models: A review of use for vegetation characterization," Remote Sens. Environ., vol. 113, pp. S56–S66, 2009.

[6] J. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," Neural Process. Lett., vol. 9, no. 3, pp. 293–300, 1999.

[7] 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, 2012.

[8] C. E. Rasmussen and C. K. I. Williams, Gaussian Processes for Machine Learning. New York, NY, USA: MIT Press, 2006.

[9] 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.

[10] 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.

[11] J. Arenas-Garcia, K. Petersen, G. Camps-Valls, and L. Hansen, "Kernel multivariate analysis framework for supervised subspace learning: A tutorial on linear and kernel multivariate methods," IEEE Signal Process. Mag., vol. 30, no. 4, pp. 16–29, Jul. 2013.

[12] D. J. MacKay, "Information-based objective functions for active data selection," Neural Comput., vol. 4, pp. 590–604, 1992.

[13] 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. 5, no. 3, pp. 606–617, Jun. 2011.

[14] J. Muñoz Marí, D. Tuia, and G. Camps-Valls, "Semisupervised classification of remote sensing images with active queries," IEEE Trans. Geosci. Remote Sens., vol. 50, no. 10, pp. 3751–3763, Oct. 2012.

[15] M. Crawford, D. Tuia, and H. Yang, "Active learning: Any value for classification of remotely sensed data?" Proc. IEEE, vol. 101, no. 3, pp. 593–608, Mar. 2013.

[16] E. Pasolli, F. Melgani, D. Tuia, F. Pacifici, and W. Emery, "SVM active learning approach for image classification using spatial information," IEEE Trans. Geosci. Remote Sens., vol. 52, no. 4, pp. 2217–2223, Apr. 2014.

[17] E. Pasolli, F. Melgani, N. Alajlan, and Y. Bazi, "Active learning methods for biophysical parameter estimation," IEEE Trans. Geosci. Remote Sens., vol. 50, no. 10, pp. 4071–4084, Oct. 2012.

[18] Y. Bazi and F. Melgani, "Semisupervised PSO-SVM regression for biophysical parameter estimation," IEEE Trans. Geosci. Remote Sens., vol. 45, no. 6, pp. 1887–1895, Jun. 2007.

[19] Y. Bazi and F. Melgani, "Kernel ridge regression with active learning for wind speed prediction," Appl. Energy, vol. 103, pp. 328–340, Mar. 2013.

[20] Y. Gu, Z. Jin, and S. Chiu, "Active learning combining uncertainty and diversity for multi-class image classification," IET Comput. Vis., vol. 9, no. 3, pp. 400–407, 2015.

[21] F. Douak, N. Benoudjit, and F. Melgani, "A two-stage regression approach for spectroscopic quantitative analysis," Chemometrics Intell. Lab. Syst., vol. 109, no. 1, pp. 34–41, 2011.

[22] 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.

[23] S. Patra and L. Bruzzone, "A cluster-assumption based batch mode active learning technique," Pattern Recognit. Lett., vol. 33, no. 9, pp. 1042–1048, 2012.

[24] J. Verrelst, E. Romijn, and L. Kooistra, "Mapping vegetation density in a heterogeneous river floodplain ecosystem using pointable CHRIS/PROBA data," Remote Sens., vol. 4, no. 9, pp. 2866–2889, 2012.

[25] J. Caicedo, J. Verrelst, J. Munoz-Mari, 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.

[26] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, M. Lázaro-Gredilla, and J. Verrelst, simpleR: A simple educational Matlab toolbox for statistical regression, v2.1, Jun. 2013. [Online]. Available: http://isp.uv.es/

[27] J. Rivera, J. Verrelst, G. Leonenko, and J. Moreno, "Multiple cost functions and regularization options for improved retrieval of leaf chlorophyll content and LAI through inversion of the PROSAIL model," Remote Sens., vol. 5, no. 7, pp. 3280–3304, 2013.

[28] B. Schölkopf and A. Smola, Learning With Kernels—Support Vector Machines, Regularization, Optimization and Beyond. Cambridge, MA, USA: MIT Press, 2002.

[29] 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.