Toward a Semiautomatic Machine Learning Retrieval of Biophysical Parameters

Juan Pablo Rivera Caicedo, Jochem Verrelst, Jordi Muñoz-Marí, *Member*, *IEEE*, José Moreno, *Member*, *IEEE*, and Gustavo Camps-Valls, *Senior Member*, *IEEE*

***Abstract*—Biophysical parameters such as leaf chlorophyll content (LCC) and leaf area index (LAI) are standard vegetation products that can be retrieved from Earth observation imagery. This paper introduces a new machine learning regression algorithms (MLRAs) toolbox into the scientific Automated Radiative Transfer Models Operator (ARTMO) software package. ARTMO facilitates retrieval of biophysical parameters from remote observations in a MATLAB graphical user interface (GUI) environment. The MLRA toolbox enables analyzing the predictive power of various MLRAs in a semiautomatic and systematic manner, and applying a selected MLRA to multispectral or hyperspectral imagery for mapping applications. It contains both linear and nonlinear state-of-the-art regression algorithms, in particular linear feature extraction via principal component regression (PCR), partial least squares regression (PLSR), decision trees (DTs), neural networks (NNs), kernel ridge regression (KRR), and Gaussian processes regression (GPR). The performance of multiple implemented regression strategies has been evaluated against the SPARC dataset (Barrax, Spain) and simulated Sentinel-2 (8 bands), CHRIS (62 bands) and HyMap (125 bands) observations. In general, nonlinear regression algorithms (NN, KRR, and GPR) outperformed linear techniques (PCR and PLSR) in terms of accuracy, bias, and robustness. Most robust results along gradients of training/validation partitioning and noise variance were obtained by KRR while GPR delivered most accurate estimations. We applied a GPR model to a hyperspectral HyMap flightline to map LCC and LAI. We exploited the associated uncertainty intervals to gain insight in the per-pixel performance of the model.**

***Index Terms*—Biophysical parameter retrieval, CHRIS, graphical user interface (GUI) toolbox, HyMap, leaf area index (LAI), leaf chlorophyll content (LCC), machine learning, nonparametric regression, Sentinel-2 (S2).**

I. Introduction

LEAF AREA INDEX (LAI) and leaf chlorophyll content (LCC) are essential land biophysical parameters retrievable from optical Earth observation (EO) data [1]–[3]. These parameters provide information about the phenological stage and health status (e.g., development, productivity, and stress) of crops and forests [4]. The quantification of these parameters from space over large areas has become an important aspect in agroecological, environmental, and climatic studies [5]. At the same time, remotely sensed observations are increasingly being applied at a within-field scale for dedicated agronomic monitoring applications [6]–[8]. With the forthcoming superspectral Sentinel-2 (S2) and Sentinel-3 missions and the planned EnMAP and PRISMA imaging spectrometers, the unprecedented data availability requires retrieval processing techniques that are accurate, robust, and fast to apply.

Biophysical parameter retrieval always requires an intermediate modeling step to transform the measurements into useful estimates [9]. This modeling step can be approached with either statistical, physical, or hybrid methods. In this paper, we will focus on the statistical approximation as this field has advanced largely over the last two decades [10], [11]. Statistical models can be categorized into either parametric or nonparametric approaches. *Parametric models* assume an explicit relation between the variables. They rely on the physical knowledge of the problem and build explicit parametrized expressions that relate a few spectral channels with the biophysical parameter of interest. Alternatively, *nonparametric models* are adjusted to predict a variable of interest using a training dataset of input–output data pairs, which come from concurrent measurements of the parameter and the corresponding reflectance/radiance observation. Several nonparametric regression algorithms are available in the statistics and machine learning literature, and recently they have been introduced for biophysical parameter retrieval [9], [12].

Particularly, the family of machine learning regression algorithms (MLRAs) emerged as a powerful nonparametric approach for delivering biophysical parameters. MLRAs have the potential to generate adaptive, robust relationships, and once trained, they are very fast to apply [13]. Typically, machine learning methods are able to cope with the strong nonlinearity of the functional dependence between the biophysical parameter and the observed reflected radiance. They may therefore be suitable candidates for operational applications. Effectively, algorithms such as neural networks (NNs) are already implemented in operational retrieval chains (e.g., CYCLOPES products) [14], [15]. It still remains to be questioned whether NNs offer the most flexible tools for parameter estimation, gaining insight in the retrievals, and evaluating retrieval performances. Besides, training NNs involve tuning several parameters that may greatly impact the final robustness of the model. In part, this why in the recent years alternative and simpler to train regression methods have started replacing NNs. Specifically, the family of *kernel methods* [10] has emerged as an alternative to NNs in many

Manuscript received June 14, 2013; revised December 13, 2013; accepted December 31, 2013. This work was supported in part by the Spanish Ministry of Science and Innovation under project AYA2010-21432-C02-01, in part by the Spanish Ministry of Economy and Competitiveness (MINECO) under project TIN2012-38102-C03-01 (LIFE-VISION), and in part by ESA PARCS project (4000105078/11/NL/AF).

The authors are with the Image Processing Laboratory (IPL), Universitat de València, València 46980, Spain (e-mail: juan.rivera@uv.es; jochem.verrelst@uv.es; jordi.munoz; jose.moreno@uv.es; gustavo.camps@uv.es).

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

Digital Object Identifier 10.1109/JSTARS.2014.2298752

scenarios. Kernel methods typically involve few and intuitive hyperparameters to be tuned, and can perform flexible input–output nonlinear mappings. Even though MLRAs are widely recognized as very powerful methods, some questions still remain open, e.g., how robust are these models in case of noisy situations, and how much they depend on changes between the training and testing data distributions. Perhaps more important is the fact that, for the broader remote sensing community, they are also perceived as complicated black boxes with several parameters to be tuned, which requires expertise. Further, until now no user-friendly graphical user interface (GUI) toolbox exists that brings several state-of-the-art MLRAs together. To facilitate and automate the use of MLRAs, in this work, we present a novel software package that allows systematically analyzing and applying MLRA-developed models. The so-called MLRA toolbox has been implemented within the in-house developed toolbox called Automated Radiative Transfer Models Operator (ARTMO). ARTMO is a scientific GUI toolbox dedicated to the retrieval of vegetation properties from optical imagery [16].

This brings us to the objectives of this work that are: 1) to present the novel MLRA toolbox for semiautomatic retrieval of biophysical parameters, 2) to evaluate the different MLRAs on their performance and robustness, and 3) to apply the best performing MLRA to EO imagery to test the robustness and accuracy in real scenarios.

The following sections will first briefly describe the considered nonparametric regression algorithms and then the latest status of ARTMO, followed by an introduction of the most important components of the new MLRA toolbox. The used data is subsequently described and an evaluation of six nonparametric regression methods is presented. A discussion on the use of these models for EO processing and a conclusion closes this paper.

II. Machine Learning Regression Algorithms

MLRAs learn the relationship between the input (e.g., reflectances) and output (e.g., biophysical parameters) by fitting a flexible model looking at the structure of the data. The hyperparameters of the model are typically adjusted to minimize the prediction error in an independent validation dataset. This way, one looks for the best *generalization* capabilities, not only a good performance in the training set that would give rise to an overfitted solution. In this paper, we compare several regression algorithms. A first family of linear methods follow a simple chained approach: first data dimensionality is applied to alleviate collinearity problems which is then followed by canonical least squares linear regression (LR). A second family of methods consists of building nonlinear functions of the data directly. Several state-of-the-art methods are considered here: 1) regression trees (RTs); 2) artificial NNs; 3) kernel ridge regression (KRR), also known as least squares support vector machine; and 4) Gaussian processes regression (GPR). All these regression techniques are popular in various application domains, thanks to its relatively fast training, good performance, and robustness to the overfitting problem. In the following sections, we briefly summarize them.

A. Dimensionality Reduction and LR

Let us consider a supervised regression problem, and let $\mathbf{X}$ and $\mathbf{y}$ be the input and output centered matrices of sizes $n \times d$ and $n \times l$, respectively. Here, l is the number of training data points in the problem and d is the data dimension. The objective of standard LR is to adjust a linear model for predicting the output variable from the input features $\hat{\mathbf{y}} = \mathbf{X}\mathbf{w}$, where $\mathbf{w}$ contains the regression model coefficients (weights) and has size $d \times 1$. Ordinary least-squares (OLS) regression solution is $\mathbf{w} = \mathbf{X}^{\dagger}\mathbf{y}$, where $\mathbf{X}^{\dagger} = (\mathbf{X}^{\top}\mathbf{X})^{-1}\mathbf{X}^{\top}$ is the Moore–Penrose pseudoinverse of $\mathbf{X}$. Highly correlated input variables can result in rank-deficient covariance matrix $\mathbf{C}_{xx} = \frac{1}{n}\mathbf{X}^{\top}\mathbf{X}$, making the inversion unfeasible. The same situation is encountered in the small-sample-size case.

A common approach in statistics to alleviate these problems considers first in reducing the data dimensionality and then applying the OLS normal equations to the projected data or *scores* [17]. These scores reduce to a linear transformation of the original data $\mathbf{X}' = \mathbf{X}\mathbf{U}$, where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n_f}]$ is referred to as the projection matrix, $\mathbf{u}_i$ being the ith projection vector and n_f the number of extracted features. The best known linear dimensionality reduction method is principal component analysis (PCA) [18] which reduces to solve the eigenproblem

$$\mathbf{C}_{xx}\mathbf{u} = \lambda\mathbf{u}.$$

An alternative supervised method is partial least squares (PLS) [19] in which we have to solve

$$\begin{pmatrix} \mathbf{0} & \mathbf{C}_{xy} \\ \mathbf{C}_{xy}^{\top} & \mathbf{0} \end{pmatrix}\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \lambda\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}.$$

Note that PCA disregards the target data and exploits correlations between the input variables to maximize the variance of the projections, while PLS looks for projections that maximize the covariance between the features and the target data. In both cases, the user selects the dimensionality of the projected data n_f. After projection, the OLS equations are solved using $\mathbf{X}'$. The approaches respectively lead to the so-called principal component regression (PCR) [20] and the partial least squares regression (PLSR) [21] methods. Particularly, PLSR emerged as a popular regression technique for interpreting hyperspectral data with various experimental applications in vegetation properties mapping (e.g., [22]–[25]).

B. Regression Trees (RTs)

RTs build predictive models that take the observations as inputs and map them to the target variable. The model structure is made out of nodes (or leaves) and branches. Leaves represent output variable discrete values and branches constitute piecewise linear decisions. Decision tree (DT) learning can be done in several ways and using different algorithms, which mainly vary on the procedure used to determine where to split. In this paper, we focused on the standard Breiman's algorithm [26]. RTs have several advantages, among them: 1) they can manage a high number of features and examples in an easy way; 2) they are nonparametric flexible methods so they do not impose a specific functional form to the solution; and 3) the variables, or

combination of variables, used at each node to divide the samples into subgroups are the most discriminative features since they assure the lowest estimated error.

The main advantage for analysts in remote sensing applications is that RTs allow knowledge discovery and full interpretability by analyzing the surrogate and main splits of the tree. They have been successfully used to estimate land surface variables such as LAI, fraction of photosynthetically active radiation (FAPAR) and chlorophyll content from VEGETATION/SPOT4 [27], or broadband albedo from the Earth Observing 1 (EO-1) data [28], just to name a few applications.

C. Neural Networks (NNs)

The most common approach to develop nonparametric and nonlinear regression is based on artificial NNs [29]. An NN is a (potentially fully) connected structure of neurons organized in layers. A neuron basically performs a LR followed by a nonlinear function $f(\cdot)$. Neurons of different layers are interconnected with the corresponding links (weights). Therefore, in the limit case of using an NN with only one neuron, the results would be similar (or slightly better because of the nonlinearity) than those obtained with OLS regression. Training an NN implies selecting a structure (number of hidden layers and nodes *per* layer), initialize the weights, shape of the nonlinearity, learning rate, and regularization parameters to prevent overfitting. In addition, the selection of a training algorithm and the loss function both have an impact on the final model. In this work, we used the standard multi-layer perceptron, which is a fully connected network. We selected just one hidden layer of neurons. We optimized the NN structure using the Levenberg–Marquardt learning algorithm with a squared loss function. A cross-validation procedure was employed to avoid overfitting issues. NN weights were initialized randomly according to the Nguyen–Widrow method, and model regularization was done by limiting the maximum number of net weights to half the number of training samples. NNs have been vastly used in biophysical parameter retrieval (e.g., [14] and [30]–[33]), and are very useful in operational settings (e.g., [34]) because they scale well with the number of training examples.

D. Kernel Ridge Regression (KRR)

KRR minimizes the squared residuals in a higher dimensional feature space, and can be considered as the kernel version of the (regularized) OLS LR [35], [36]. The LR model is defined in a Hilbert space $\mathcal{H}$ of a very high dimensionality, where samples have been mapped to through a mapping $\phi(\mathbf{x}_i)$. In matrix notation, the model is given by $\hat{y}_i = \phi(\mathbf{x}_i)^{\mathrm{T}}\mathbf{w} + b$. Notationally, we want to solve a regularized OLS problem in Hilbert spaces

$$\sum_i (y_i - \hat{y}_i)^2 + \lambda \|\mathbf{w}\|^2. \tag{1}$$

Taking derivatives with respect to model weights $\mathbf{w}$ and b, and equating them to zero, leads to an equivalent problem depending on the unknown mapping function ϕ, which in principle is unknown. The problem can be solved by applying the Representer's theorem, by which the weights can be expressed as a linear combination of mapped samples, $\mathbf{w} = \sum_{i=1}^{n} \alpha_i \phi(\mathbf{x}_i)$. The prediction for a test sample $\mathbf{x}_*$ is obtained as a function of the dual weights $\alpha = [\alpha_1, \ldots, \alpha_n]^\top$ (one per sample), as follows:

$$\mathbb{E}[f(\mathbf{x}_*)] = \mathbf{k}_{*,:}^\top (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y} = \mathbf{k}_{*,:}^\top \boldsymbol{\alpha} \tag{2}$$

where $\mathbf{k}_{*,:}$ contains the (kernel) similarities between the test example and all training data points. Note that for obtaining the model, only the inversion of the Gram (or kernel) matrix $\mathbf{K}$ of size $n \times n$ regularized by λ is needed. We have used the RBF kernel function, whose components $[\mathbf{K}]_{ij}$ are

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / (2\sigma^2)). \tag{3}$$

Therefore, in KRR only the regularization parameter λ and the kernel parameter σ have to be selected. Both parameters were optimized via standard cross-validation. It is worth noting that KRR has been recently used in remote sensing applications [10], [37].

E. Gaussian Processes Regression (GPR)

GPR has been recently introduced as a powerful regression tool [38] and applied to remote sensing data [39]–[42]. 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 variables (biophysical parameter) in the same way as KRR (see Eq. 2). However, two main advantages of GPR must be noted.

First, not only a *predictive mean* but also a *predictive variance* can be obtained

$$\mathbb{V}[f(\mathbf{x}_*)] = \mathbf{k}_{*,*} - \mathbf{k}_{*,:}^\top (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{k}_{:,*}. \tag{4}$$

Note that the mean prediction in (2) is a linear combination of observations $\mathbf{y} = [y_1, \ldots, y_n]^\top$, whereas the predictive variance in (4) only depends on input data and can be taken as the difference between the prior kernel and the information given by observations about the approximation function.

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. See [38], [43], and [40] for further details. We used a scaled anisotropic Gaussian kernel function

$$K(\mathbf{x}_i, \mathbf{x}_j) = \nu \exp\left(-\sum_{b=1}^{B} \frac{(\mathbf{x}_i^{(b)} - \mathbf{x}_j^{(b)})^2}{2\sigma_b^2} \right) \tag{5}$$

where ν is a scaling factor, B is the number of bands, and σ_b is a dedicated parameter controlling the spread of the relations for each particular spectral band b.

Summarizing, three important properties of the method are worth stressing here. First, the obtained weights α_i after optimization give the relevance of each spectrum $\mathbf{x}_i$. Second, the inverse of σ_b represents the relevance of band b. Intuitively, high values of σ_b mean that relations largely extend along that band, hence suggesting a lower informative content. Finally, a GPR model provides not only a pixel-wise prediction for each

spectrum but also an uncertainty (or confidence) level for the prediction.

The previous methods are implemented from the simple regression toolbox [44], simpleR, freely available at http://www.uv.es/gcamps/code/simpleR.html. The simpleR toolbox contains simple educational code for LR, DTs (TREE), NNs, support vector regression (SVR), KRR, aka least squares SVM, Gaussian process regression (GPR), and variational heteroscedastic Gaussian process regression (VHGPR). The toolbox is not explicitly included in ARTMO, but may be of interest for the reader, as it provides more regression and analysis tools.

III. ARTMO

ARTMO brings multiple leaf and canopy radiative transfer models (RTMs) together along with essential tools required for semiautomatic retrieval of biophysical parameters in one GUI toolbox. In short, the toolbox permits the user: 1) to choose between various invertible leaf and canopy RTMs of a low to high complexity (e.g., PROSPECT-4, PROSPECT-5, DLM, 4SAIL, and FLIGHT); 2) to specify or select spectral band settings specifically for various existing air- and space-borne sensors or user defined settings, typically for recently developed or future sensor systems; 3) to simulate large datasets of top-of-canopy (TOC) reflectance spectra for sensors sensitive in the optical range (400–2500 nm); 4) to generate look-up tables (LUTs), which are stored in a relational SQL database management system (MySQL, version 5.5 or higher; local installment required), and finally; 5) to configure and run various retrieval scenarios using EO reflectance datasets for biophysical parameter mapping applications. ARTMO is developed in MATLAB (2009 version or higher) and does not require additional MATLAB toolboxes. Fig. 1 presents ARTMO v3's main window and a systematic overview of the drop-down menu below. To start with, in the main window, a new project can be initiated, a sensor chosen and a comment added, whereas all processing modules are accessible through drop-down menus at the top bar.

A first rudimentary version of ARTMO has been used in LUT-based inversion applications [45], [46]. ARTMO v3 is formally presented in this paper. The software package is freely downloadable at: http://ipl.uv.es/artmo. Its most important novelties are briefly listed below.

1) ARTMO v3 is designed modularly. Its modular architecture offers the possibility for easy addition (or removal) of components, such as RTM models and post-processing modules.
2) The MySQL database is organized in such a way that it supports the modular architecture of ARTMO v3. This avoids redundancy and increases the processing speed. For instance, all spectral datasets are stored as binary objects.
3) New retrieval toolboxes are incorporated. They are based on parametric and nonparametric regression as well as physically based inversion using a LUT. This has led to the development of a: 1) "Spectral Indices assessment toolbox" [47]; 2) "MLRA toolbox"; and 3) "LUT-based inversion toolbox" [48].

This paper introduces the "MLRA module." Its general architecture is outlined in Fig. 2.

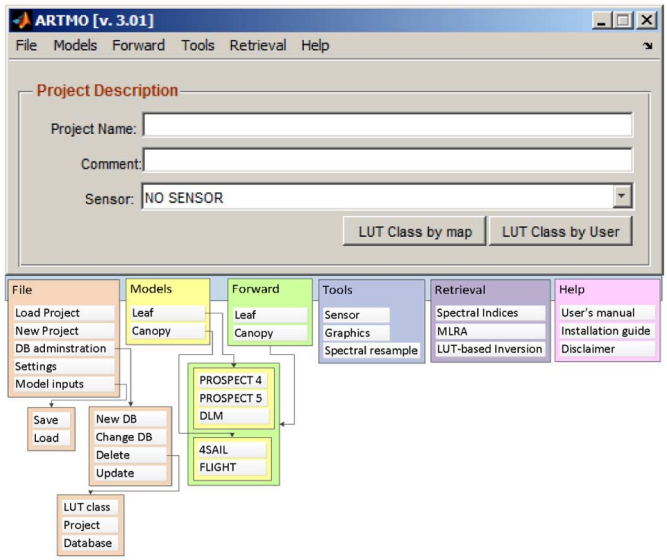

Fig. 1. Screenshot of ARTMO's main window and schematic overview of its drop-down menu.

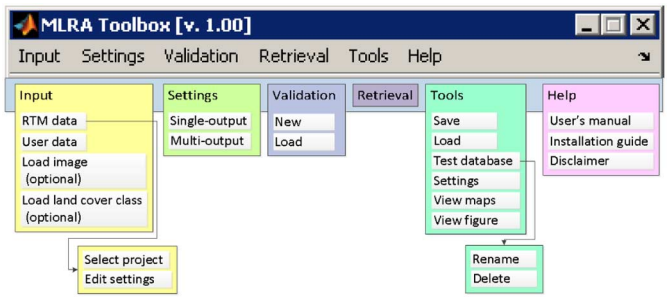

Fig. 2. Screenshot of MLRA's toolbox and schematic overview of its drop-down menu.

A. MLRA Settings Module

The following step addresses the analysis of multiple MLRA-based retrieval strategies. A first step to do is inserting input data (i.e., a plain text file), which refers to retrievable biophysical parameters and associated spectra. This is done in "Input" and can be either RTM-simulated data or can be ground truth data as measured in the field. The GUI will guide the user through the data selection steps and checks if data is properly read (not displayed for brevity). Once data is inserted the "MLRA settings" module can be configured (Fig. 3). It can be opted to select either single-output or multi-output regression algorithm. Currently, only PLSR, NN, and KRR encompass multi-output capabilities. Obviously, these models can also be used for single-output applications.

The "MLRA settings" module configures the regression algorithms given various options. First, if a land cover map in ENVI format (Exelis Inc.) has been provided then retrieval strategies can be configured *per* land cover class. Second, multiple regression algorithms at once can be selected, which means that they will be analyzed one-by-one. Third, options to add Gaussian noise are provided. Noise can be added both on the parameters to retrieve and on the spectra. A range of noise level can be configured, so that multiple noise scenarios can be evaluated.

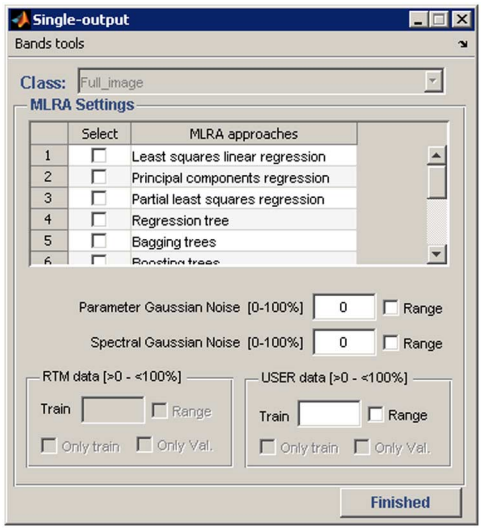

Fig. 3. MLRA's setting window.

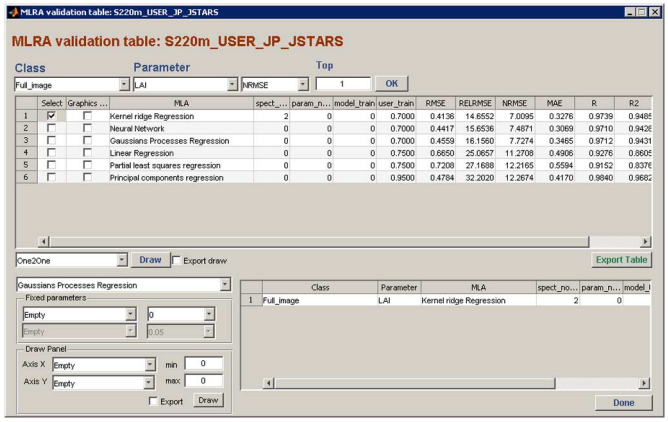

Fig. 4. MLRA's validation window.

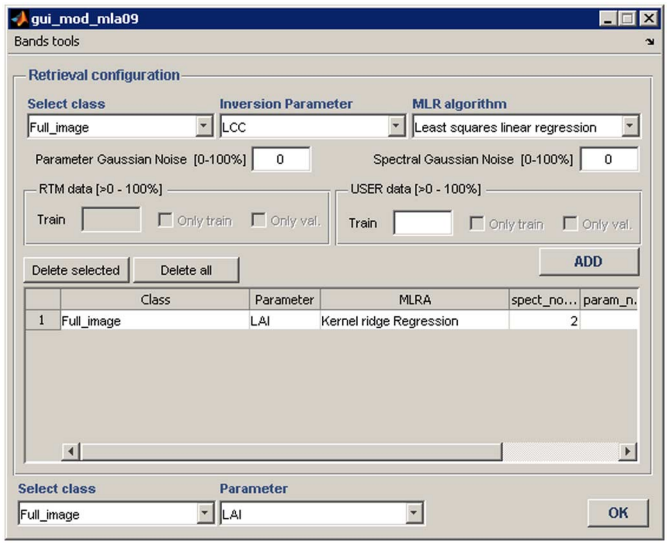

Fig. 5. MLRA's retrieval window.

The injection of noise can be of importance to account for environmental and instrumental uncertainties when synthetic spectra from RTMs are used for training. Fourth, the training/validation data partition can be controlled by setting the percentage of how much data from an RTM or user-defined is assigned to training or to validation (i.e., split-sample approach). Thereby, the user can evaluate the impact of ranging training/validation partitioning by entering a range of training/validation partitions. For each training/validation partition, the MLRA toolbox internally divides the defined training set into k subsets using a k-fold cross-validation strategy in order to tune the free parameters of the model.

B. Validation Module

Once that the training/validation data splitting has been defined and MLRA settings configured, a range of scenarios can be run, tested and their performance assessed. This is done by naming a validation set in the "Validation" module. Each regression model strategy over the configured ranges are one-by-one analyzed through goodness-of-fit measures and validation results are stored in a MySQL database. As such, a large number of results can be stored in a systematic manner, so that they can be easily queried and compared. Validation results are presented in the "MLRA validation table" (Fig. 4). The table shows the best performing validation results according to a selected land cover class (if loaded), parameter, and statical goodness-of-fit measure. Various options to display the results are provided, e.g., 1:1-line, plotting the band relevance as given by the GPR model, and 2-D matrices of performances along ranges of noise and varying training/validation distribution (see Section V). Finally, by clicking on "Retrieval," an analyzed regression function can be selected for each retrievable parameter (e.g., the best one). Such regression function will be accessed in the "Retrieval" GUI and can then be applied to a remote sensing image.

C. Retrieval Module

The "Retrieval" module enables to run an evaluated model or directly configure a model and apply it to an image (provided in standard ENVI file format) to map a parameter (Fig. 5). Hence, the user can select the required land cover class (if available), the retrievable parameter, the regression algorithms and training/validation splitting. Similarly, noise can be added to the spectra or parameters and the size of the training data can be selected. The user will then be invited to select one or multiple remote sensing images to which the developed model will be applied. Generated maps are stored in ENVI format.

IV. Mapping Applications

Having the MLRA toolbox presented, it is subsequently applied for evaluating the performance of the six presented nonparametric regression techniques to achieve optimized biophysical parameters estimation. Used data is first outlined, followed by the experimental setup. Results are then presented and a mapping application is shown.

A. Used Data

A diverse field dataset, covering various crop types, growing phases, canopy geometries, and soil conditions was collected

during SPARC (Spectra bARrax Campaign). The SPARC-2003 campaign took place during 12–14 July in Barrax, La Mancha, Spain (coordinates $30°3'\mathrm{N}$, $28°6'W$, 700 m altitude). Biophysical parameters were measured within a total of 110 elementary sampling units (ESUs) among different crops (garlic, alfalfa, onion, sunflower, corn, potato, sugar beet, vineyard, and wheat). ESU refers to a plot size compatible with pixel dimensions of about $20\ \mathrm{m} \times 20\ \mathrm{m}$. LCC was derived by measuring within each ESU about 50 samples with a calibrated CCM-200 Chlorophyll Content Meter [49] Green LAI was derived from canopy measurements made with a LiCor LAI-2000 digital analyser. Each ESU was assigned to an LAI value, which was obtained as a statistical mean of 24 measures ($8\ \mathrm{data\ reading} \times 3\ \mathrm{replications}$) with standard errors between 5% and 10% [50]. In total, LAI varies between 0.4 and 6.3 and LCC between 2 and $55\ \mu\mathrm{g/cm}^2$. Additionally, 60 random spectra over bare soils, man-made surfaces, and water bodies were added to broaden the dataset to nonvegetation samples (i.e., with an LCC and LAI value of zero), leading to a total of 170 samples.

During the campaign airborne hyperspectral spaceborne CHRIS images and airborne HyMap flightlines were acquired. CHRIS provides high spatial resolution hyperspectral data over the VNIR spectra from 400 to 1050 nm. It can operate in different modes, balancing the number of spectral bands, size of the covered area, and spatial resolution because of on-board memory storage reasons [51]. We made use of nominal nadir CHRIS observation in Mode 1 (62 bands, maximal spectral information), which were acquired during the SPARC campaign (12 July 2003). CHRIS Mode 1 has a spatial resolution of 34 m at nadir. The spectral resolution provides a bandwidth from 6 to 33 nm depending on the wavelength. CHRIS imagery was processed using ESA's CHRIS-Box available in VISAT/BEAM, which includes radiometric recalibration, coherent-noise reduction, geometric correction and atmospheric correction [52], [53]. HyMap was configured with 125 bands between 430 and 2490 nm with bandwidths varying between 11 and 21 nm and a pixel size of 5 m. The same geometric and atmospheric preprocessing as for CHRIS was applied, but given a superior signal-to-noise ratio this sensor provides a better radiometric quality than CHRIS.

B. Experimental Setup

SPARC field data were used for training and validation, and associated spectral data came from CHRIS and HyMap. In view of ESA's forthcoming S2 mission, also S2 data at a spatial resolution of 20 m were additionally generated. S2 satellites capitalize on the technology and the vast experience acquired with SPOT and Landsat over the past decades. It provides a set of 13 spectral bands spanning between 443 and 2190 nm, four bands at 10 m, six bands at 20 m and three bands at 60 m spatial resolution [54]. Because of being spaceborne and providing similar pixel size, CHRIS data were resampled to the band settings of S2. Nearest neighbor was used for the spatial resampling and a Gaussian model with full width at half maximum (FWHM) spacings was used for spectral resampling. Constrained by the spectral range of CHRIS, a dataset of eight bands at 20 m (4 bands at 20 m plus 4 bands at 10 m coarse-grained at 20 m) was prepared, hereafter referred as "S2-20 m."

The MLRA toolbox was used to evaluate the performance of the different regression algorithms along gradients of changing training/validation distributions (from 5% to 95% training, with steps of 5%; the remaining data go to validation) and increasing Gaussian noise levels (from 0% to 20% with steps of 2%). By systematically evaluating the performance along those two dimensions in a 2-D matrix format, an indication about the robustness of these regression methods can be obtained. Models were developed both for LCC and LAI. The predictive power of the developed models was evaluated with the absolute root-mean-squared error (RMSE), the normalized RMSE ($\mathrm{NRMSE}\ [\%] = \mathrm{RMSE}/\mathrm{range\ parameter\ measurements} * 100$) and the coefficient of determination (r^2) to account for the goodness-of-fit. Here, validation results are presented in the form of NRMSE, which allows accuracy comparison across different parameters. Typically, remote sensing end users require an error threshold below 10%.

V. Results

A. Regression Method Evaluation

For each parameter, sensor type and regression algorithm, NRMSE results along varying training/validation distribution and increasing noise levels are presented in 2-D matrices (Fig. 6). The best performing scenario for each matrix is also shown in Table I. When comparing these matrices, the following observations can be made.

Starting with PCR, this method proved to perform rather stable within the matrix space. For S2-20 m and CHRIS data, PCR seems to be hardly affected by a varying training/validation partition and noise injection. In fact, injection of some noise rather improved accuracies of CHRIS and HyMap. Hence, adding noise can lead to a closer match between training and validation data. However, results were never outstanding, and LCC prediction with HyMap data completely failed. Only for HyMap LAI results improved to r^2 up to 0.97 when 95% of the data were assigned to the training process. Therefore, on the whole, PCR is evaluated as suboptimal performing.

Second, the partial least square regression (PLSR) is an improved version of the PCR and widely used in EO applications. It systematically outperforms PCR in absence of noise. Improvements are particularly notable for LCC (r^2 up to 0.96 for CHRIS and HyMap). But PLSR is also more affected by the injection of noise. Low noise levels led to superior results, but above about 8% accuracies degraded rapidly.

Third, DTs yielded on the whole poorest results. Particularly, unacceptable poor results were obtained with low training data, and when many bands are involved. This suggests that DTs would not be a good choice for applying to hyperspectral data unless a large database is available. In fact, only good results were obtained [r^2 up to 0.94 (LCC) and 0.97 (LAI)] in case of "S2-20 m" (8 bands) when more than 80% was used for training and below 8% noise added.

Fourth, NNs are characterized by causing erratic patterns in each of the matrices. While being able to deliver very accurate results in some cases (e.g., for LAI using CHRIS and HyMap: r^2 up to 0.96 and 0.99) NN also showed to perform rather unstable, with large probability of delivering poor results. Particularly,

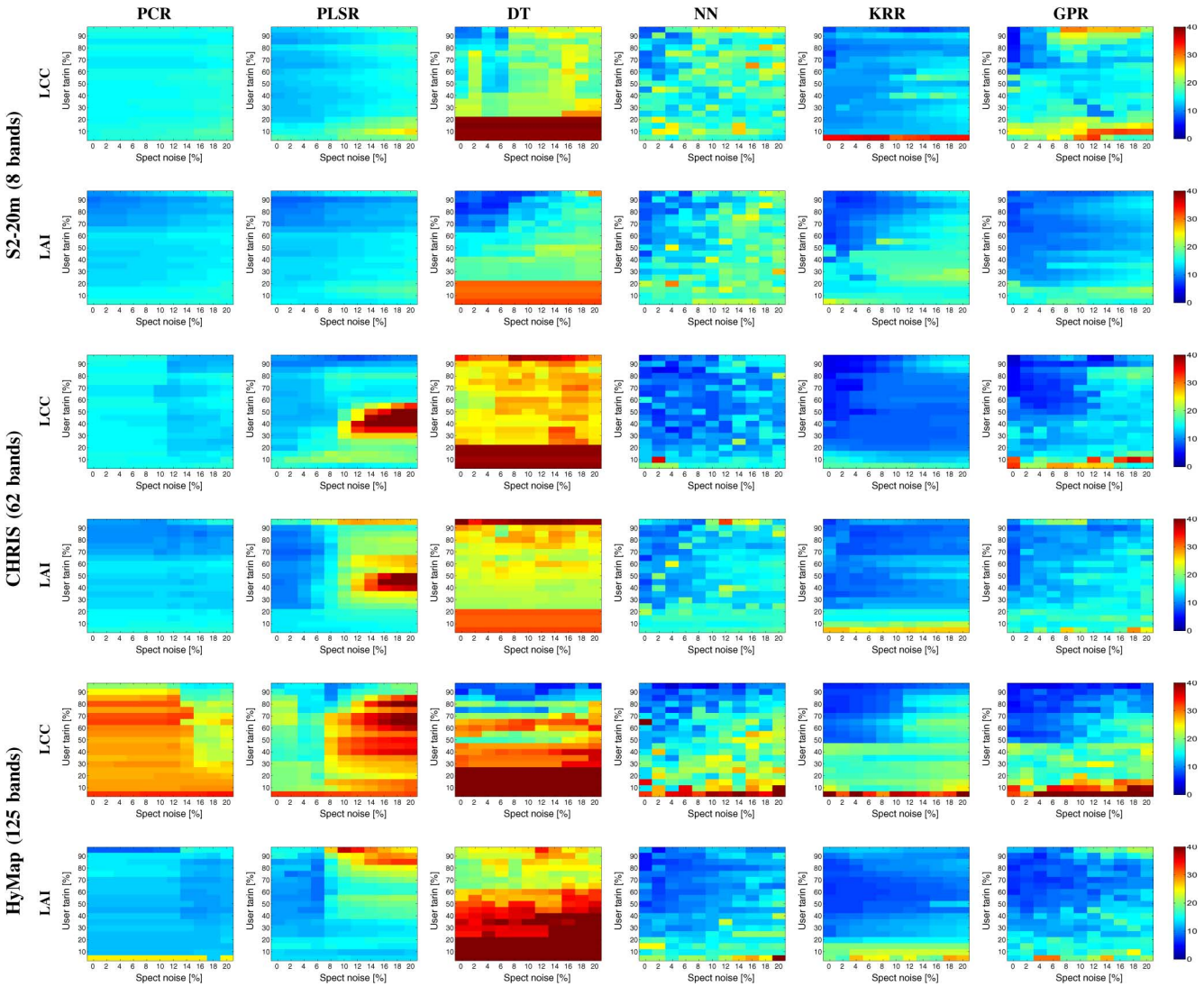

Fig. 6. NRMSE matrices of validation results for LCC and LAI retrieval using a regression algorithm displaying the impact of percentage noise (X-axis) against multiple solutions (Y-axis). The more bluish, the lower relative errors and thus the better the retrieval.

when more noise is involved and when less data are dedicated for training NN tends to perform more unstable. This erratic behavior can be explained by the complicated training phase whereby a highly specialized model is developed, but therefore easily faces the problem of overfitting. The lack of robustness to noise along with the complexity in training are therefore major drawbacks of NN.

Fifth, from all evaluated regression algorithms, KRR yielded most robust results. It led to excellent accuracies with r^2 maxima between 0.94 and 0.99 for all datasets, and more importantly also proved to perform very stable with increasing noise levels. Also, it should be noted that from all tested nonlinear MLRAs, this regression technique is fastest in developing their models (see [12] for a quantitative comparison).

Finally, GPR appears to be the most promising regression algorithm. It easily leads to excellent performances, with r^2 maxima between 0.94 and 0.99 for all datasets. Though, in comparison to KRR, GPR is somewhat more affected by noise injection. Note that the predictive mean equations for KRR and GPR are exactly the same so in principle the results should be exactly identical. Nevertheless, in GPR, we used a very flexible kernel to account for different lengthscales per feature. Although beneficial without noise, this turns to be a curse in these particular experiments because noise affected the marginal likelihood estimation of hyperparameters. For this reason, it can be concluded that GPR performs slightly less robust than KRR.

B. Biophysical Parameter Mapping

The developed models can be applied to any EO imagery given the same band settings as those presented during the training phase. From all considered regression techniques, GPR was evaluated as reaching highest accuracies for the majority of cases. Moreover, GPR has unique additional features: 1) it reveals most relevant bands when developing the model and 2) it provides uncertainty intervals (σ) associated with the mean predictions (μ). Therefore the consecutive approach was to apply GPR to a hyperspectral HyMap imagery for mapping LCC and LAI estimates along with associated uncertainty intervals (Fig. 7). In the obtained μ maps, the irrigated circular agricultural fields are clearly differentiated, including within-field variability.

TABLE I
VALIDATION STATISTICS [NOISE LEVEL (%), TRAINING (%), RMSE, NRMSE, AND r^2] SORTED ACCORDING TO BEST EVALUATED NRMSE FOR LCC AND LAI RETRIEVAL

		Regression algorithm	Noise (%)	Training (%)	RMSE	NRMSE (%)	r^2
S2-20m (8 bands)	LCC	Principal component regression	4	65	7.11	13.79	0.87
		Partial least square regression	0	85	5.58	10.90	0.92
		Decision tree	6	90	4.56	8.95	0.94
		Neural network	2	95	3.94	7.72	0.96
		Kernel ridge regression	0	75	3.48	6.82	0.97
		Gaussian processes regression	**0**	**90**	**2.40**	**4.71**	**0.98**
	LAI	Principal component regression	0	90	0.50	9.30	0.93
		Partial least square regression	0	90	0.45	8.36	0.95
		Decision tree	6	85	0.33	6.09	0.97
		Neural network	0	90	0.39	7.17	0.96
		Kernel ridge regression	**2**	**90**	**0.36**	**6.03**	**0.98**
		Gaussian processes regression	0	80	0.38	7.03	0.95
CHRIS (62 bands)	LCC	Principal component regression	12	90	5.68	11.14	0.91
		Partial least square regression	12	95	4.07	7.99	0.96
		Decision tree	0	55	10.59	20.51	0.71
		Neural network	0	95	1.94	3.81	0.99
		Kernel ridge regression	0	85	1.95	3.81	0.99
		Gaussian processes regression	**0**	**95**	**1.06**	**3.01**	**0.99**
	LAI	Principal component regression	16	85	0.50	9.30	0.92
		Partial least square regression	4	70	0.46	8.20	0.93
		Decision tree	6	85	1.05	13.32	0.68
		Neural network	**0**	**65**	**0.37**	**6.50**	**0.96**
		Kernel ridge regression	0	95	0.25	6.70	0.97
		Gaussian processes regression	0	90	0.37	6.88	0.95
HyMap (125 bands)	LCC	Principal component regression	16	85	6.48	13.18	0.89
		Partial least square regression	8	85	3.84	7.82	0.96
		Decision tree	12	90	2.25	4.69	0.98
		Neural network	4	95	1.19	3.60	0.99
		Kernel ridge regression	0	75	2.28	4.64	0.98
		Gaussian processes regression	**0**	**80**	**1.73**	**3.51**	**0.99**
	LAI	Principal component regression	8	95	0.39	8.26	0.97
		Partial least square regression	6	70	0.55	8.83	0.88
		Decision tree	6	85	0.54	10.17	0.88
		Neural network	**2**	**95**	**0.23**	**4.74**	**0.99**
		Kernel ridge regression	0	60	0.42	6.78	0.94
		Gaussian processes regression	0	80	0.30	5.66	0.95

Best NRMSE result per sensor and biophysical parameter is bold typed.

In the uncertainty maps, the lower the σ (whiter color) indicate the more certain the retrieval as processed by the trained model. The delivery of uncertainty estimates allows us to provide insight on a pixelwise basis when applied to any image and so enables the interpretation at which land covers retrievals were associated with great certainty and land covers would benefit from additional sampling. It can be observed that, particularly over the circular agricultural parcels, LCC was processed with high certainty. This is less obvious for LAI retrievals; however, it should be kept in mind that σ is also related to the magnitude of the mean estimates (μ). For this reason, relative uncertainties (σ/μ) may provide a more meaningful interpretation. These maps can function as a spatial mask that enables displaying only pixels with great certainty. Moreover, uncertainty maps can also give information about the portability of the regression models when applied to images over areas other than the training site [41], [42].

VI. DISCUSSION

The hereby presented MLRA toolbox allows evaluating and applying a wide range of regression techniques in a semiautomatic and user-friendly way. As a case study, we applied the MLRA toolbox to compare six regression algorithms on their performance and robustness along ranges of varying training/validation distribution and noise variance. These algorithms can be categorized in either data dimensionality transformations (PCR, PLSR) and nonlinear algorithms (DT, NN, KRR, GPR). For all used datasets (S2-20 m, CHRIS and HyMap) pronounced differentiation in their best performances emerged. While for PCR, PLSR and DT best accuracies fell within a range of 4.7%–20.5% (r^2: 0.86–0.97) the MLRA algorithms NN, KPR and GPR yielded higher accuracies between 3.5% and 7.7% (r^2: 0.94–0.99). Hence, each of these MLRAs reached accuracies below 10%, which is typically demanded in operational products. These excellent performances can be explained by that MLRAs may find the nonlinear feature relations by building more flexible and adaptive models than those restricted to linear projections or regression. The excellent performance of the MLRAs becomes even more apparent when comparing against other classic retrievals methods. The same validation dataset reached r^2 on the order of 0.85 by using vegetation indices from CHRIS data [40], and r^2 up to 0.77 for the same S2-20 m bands by using LUT inversion of the PROSAIL radiative transfer (RT) model through cost functions [46]. Moreover, apart from the evaluated algorithms here, others can be added relatively easily. Meanwhile, a

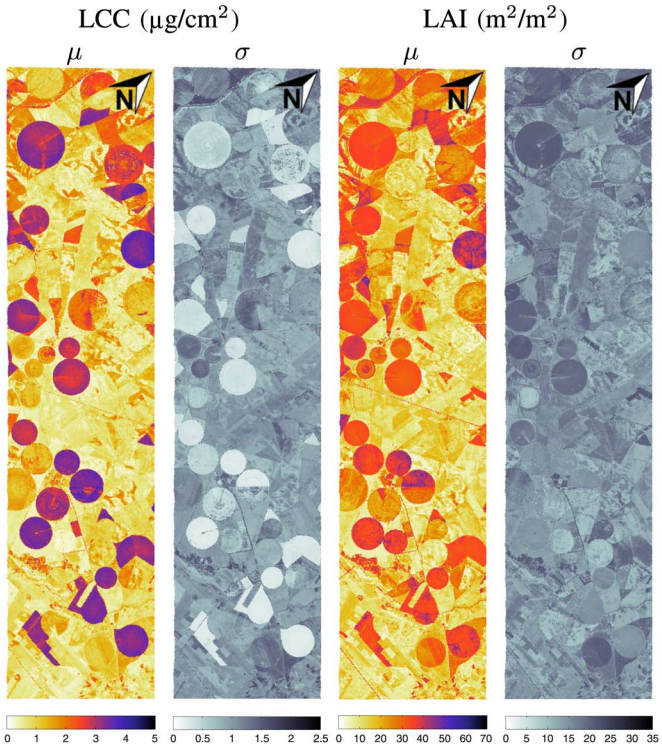

Fig. 7. LCC ($\mu g/cm^2$) and LAI (m^2/m^2) mean estimates (μ) and associated uncertainties (σ) for a HyMap flightline over Barrax agricultural area (Spain).

wide array of new MLRAs have already been implemented, among others: SVR, extreme learning machines (ELMs), and VHGPR [44], [55].

An urging open question is about evaluating how well these algorithms perform when being fed by large datasets as generated by canopy RT models. The advantage of RT models is that a broad range of land cover situations can be simulated (e.g., up to hundred thousands), leading to a dataset several times bigger than what can be collected during a field campaign. Operational processing chains typically rely on this hybrid approach [15], and similar strategies could be developed by the ARTMO toolbox. It remains, however, to be investigated how well kernel-based MLRAs perform with large datasets. This is not a trivial point. For instance, the computational load of the GPR increases exponentially with each added sample, making that this function faces difficulties when being trained by several thousand (in principle distinct) samples. Alternatively, dimensionality reduction techniques may largely overcome the burden of large datasets. Currently, a diversity of linear and nonlinear PCA techniques are being implemented (e.g., kernel PCA) in order to apply dimensionality reduction. On the other hand, redundancy also takes place along the simulated spectra, e.g., because not all RTM parameters lead to spectral variations, causing redundancy along the dataset. Therefore, the emerging field of redundant data reduction is expected to further reduce the dataset while preserving good performance [56], [57]. The field is also related to active learning approaches [58]. These dimensionality reduction techniques are foreseen to be implemented as well, which will eventually facilitate a smooth coupling between RTM-generated simulated spectra and powerful MLRAs for generic and operational retrieval applications.

VII. Conclusion

ARTMO's new "MLRA toolbox" enables applying and analyzing the predictive power of various MLRAs in a semi-automatic manner. Various regularization options have been implemented into the toolbox, e.g., training/validation data splitting, adding noise, and regression models can be developed and evaluated per land cover class. Data can either come from field campaigns or from simulations as generated by RTMs. The predictive power of multiple nonparametric regression algorithms was evaluated across gradients of varying training/validation distribution and increasing noise levels. By using the local SPARC dataset and multispectral simulated S2 and hyperspectral CHRIS and HyMap imagery over the Barrax (Spain) agricultural area, KRR and GPR emerged as most robust and best performing regression algorithms (r^2 up to 0.94–0.99 and NRMSE down to 7.0%–3.0%). Moreover, GPR provides additional uncertainty estimates on a pixelwise basis, which provides insight in the performance of the model. In all generality, the linear nonparametric algorithms such as the popular PLSR performed systematically poorer than the nonlinear, kernel-based regression algorithms (KRR, GPR).

The presented experimental results demonstrated the utility of the MLRA toolbox, which essentially has been developed to serve efficient and optimized surface properties mapping.

Acknowledgment

The authors would like to thank the three anonymous reviewers for their valuable comments.

References

[1] R. H. Whittaker and P. L. Marks, "Methods of assessing terrestrial productivity," in *Primary Productivity of the Biosphere*, New York, NY, USA: Springer, 1975, pp. 55–118.
[2] H. K. Lichtenthaler, "Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes," *Methods Enzymol.*, vol. 148, pp. 350–382, 1987.
[3] Z. Malenovský, H. Rott, J. Cihlar, M. Schaepman, G. García-Santos, R. Fernandes, and M. Berger, "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, 2012.
[4] H. K. Lichtenthaler, M. Lang, M. Sowinska, F. Heisel, and J. A. Miehe, "Detection of vegetation stress via a new high resolution fluorescence imaging system," *J. Plant Physiol.*, vol. 148, no. 5, pp. 599–612, 1996.
[5] W. A. Dorigo, R. Zurita-Milla, A. J. W. de Wit, J. Brazile, R. Singh, and M. E. Schaepman, "A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 9, no. 2, pp. 165–193, 2007.
[6] J. Delegido, J. Verrelst, L. Alonso, and J. Moreno, "Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content," *Sensors*, vol. 11, no. 7, pp. 7063–7081, 2011.
[7] J. Delegido, J. Verrelst, C. Meza, J. Rivera, L. Alonso, and J. Moreno, "A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems," *Eur. J. Agronomy*, vol. 46, pp. 42–52, 2013.
[8] G. Gianquinto, F. Orsini, M. Fecondini, M. Mezzetti, P. Sambo, and S. Bona, "A methodological approach for defining spectral indices for assessing tomato nitrogen status and yield," *Eur. J. Agronomy*, vol. 35, no. 3, pp. 135–143, 2011.
[9] F. Baret and S. Buis, "Estimating canopy characteristics from remote sensing observations. Review of methods and associated problems," in *Advances in Land Remote Sensing: System, Modeling, Inversion and Application*, New York, NY, USA: Springer, 2008, pp. 172–301.
[10] G. Camps-Valls and L. Bruzzone, Eds., *Kernel Methods for Remote Sensing Data Analysis*. New York, NY, USA: Wiley, 2009.
[11] G. Camps-Valls, D. Tuia, L. Gómez-Chova, S. Jiménez, and J. Malo, "Remote sensing image processing," in *Synthesis Lectures on Image, Video, and Multimedia Processing*, vol. 12, pp. 1–194, 2012.

[12] J. Verrelst, J. Muñoz, L. Alonso, J. Delegido, J. Rivera, G. Camps-Valls, and J. Moreno, "Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3," *Remote Sens. Environ.*, vol. 118, pp. 127–139, 2012.

[13] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer-Verlag, 2009.

[14] C. Bacour, F. Baret, D. Béal, M. Weiss, and K. Pavageau, "Neural network estimation of LAI, fAPAR, fCover and $\mathrm{LAI} \times \mathrm{Cab}$, from top of canopy MERIS reflectance data: Principles and validation," *Remote Sens. Environ.*, vol. 105, no. 4, pp. 313–325, 2006.

[15] A. Verger, F. Baret, and M. Weiss, "Performances of neural networks for deriving LAI estimates from existing CYCLOPES and MODIS products," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 2789–2803, 2008.

[16] J. Verrelst, J. Rivera, L. Alonso, and J. Moreno, "ARTMO: An automated radiative transfer models Operator toolbox for automated retrieval of biophysical parameters through model inversion," in *Proc. EARSeL 7th SIG-Imag. Spectrosc. Workshop*, Edinburgh, U.K., 2011.

[17] J. Arenas-García, K. B. Petersen, G. Camps-Valls, and L. K. Hansen, "Kernel multivariate analysis framework for supervised subspace learning," *IEEE Signal Process. Mag.*, vol. 30, no. 4, pp. 16–29, 2013.

[18] I. T. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer-Verlag, 1986.

[19] H. Wold, "Partial least squares," *Encyclopedia Statist. Sci.*, vol. 6, pp. 581–591, 1985.

[20] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemom. Intell. Lab. Syst.*, vol. 2, no. 1–3, pp. 37–52, 1987.

[21] P. Geladi and B. Kowalski, "Partial least-squares regression: A tutorial," *Anal. Chim. Acta*, vol. 185, no. C, pp. 1–17, 1986.

[22] N. C. Coops, M.-L. Smith, M. Martin, and S. V. Ollinger, "Prediction of eucalypt foliage nitrogen content from satellite-derived hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1338–1346, Jun. 2003.

[23] D. Gianelle and F. Guastella, "Nadir and off-nadir hyperspectral field data: Strengths and limitations in estimating grassland biophysical characteristics," *Int. J. Remote Sens.*, vol. 28, no. 7, pp. 1547–1560, 2007.

[24] M. A. Cho, A. K. Skidmore, F. Corsi, S. E. van Wieren, and I. Sobhan, "Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 9, no. 4, pp. 414–424, 2007.

[25] X. Ye, K. Sakai, M. Manago, S.-I. Asada, and A. Sasao, "Prediction of citrus yield from airborne hyperspectral imagery," *Precis. Agric.*, vol. 8, no. 3, pp. 111–125, 2007.

[26] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Monterey, CA, USA: Wadsworth and Brooks, 1984.

[27] M. Weiss and F. Baret, "Evaluation of canopy biophysical variable retrieval performances from the accumulation of large swath satellite data," *Remote Sens. Environ.*, vol. 70, pp. 293–306, 1999.

[28] S. Liang, H. Fang, M. Kaul, T. G. Van Niel, T. R. Mcvicdr, J. Pearlman *et al.*, "Estimation of land surface broadband albedos and leaf area index from EO-1 ALI data and validation," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1260–1268, Jun. 2003.

[29] S. Haykin, *Neural Networks—A Comprehensive Foundation*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice Hall, Oct. 1999.

[30] J. Smith, "LAI inversion using backpropagation neural network trained with multiple scattering model," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 5, pp. 1102–1106, Sep. 1993.

[31] S. Gopal and C. Woodcock, "Remote sensing of forest change using artificial neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 2, pp. 398–404, Mar. 1996.

[32] D. S. Kimes, R. F. Nelson, M. T. Manry, and A. K. Fung, "Attributes of neural networks for extracting continuous vegetation variables from optical and radar measurements," *Int. J. Remote Sens.*, vol. 19, no. 14, pp. 2639–2662, 1998.

[33] A. Verger, F. Baret, and F. Camacho, "Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: Evaluation over an agricultural area with CHRIS/PROBA observations," *Remote Sens. Environ.*, vol. 115, no. 2, pp. 415–426, 2011.

[34] F. Baret, M. Weiss, R. Lacaze, F. Camacho, H. Makhmara, P. Pacholcyzk, and B. Smets, "GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 1: Principles of development and production," *Remote Sens. Environ.*, vol. 137, pp. 299–309, 2013.

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

[36] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[37] G. Camps-Valls, J. Muñoz-Marí, L. Gómez-Chova, L. Guanter, and X. Calbet, "Nonlinear statistical retrieval of atmospheric profiles from MetOp-IASI and MTG-IRS infrared sounding data," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1759–1769, May 2012.

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

[39] 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.

[40] 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, pt. 2, pp. 1832–1843, May 2012.

[41] J. Verrelst, L. Alonso, J. Rivera Caicedo, J. Moreno, and G. Camps-Valls, "Gaussian process retrieval of chlorophyll content from imaging spectroscopy data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 867–874, Apr. 2013.

[42] J. Verrelst, J. Rivera, J. Moreno, and G. Camps-Valls, "Gaussian processes uncertainty estimates in experimental Sentinel-2 LAI and leaf chlorophyll content retrieval," *ISPRS J. Photogramm. Remote Sens.*, vol. 86, pp. 157–167, 2013.

[43] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, and J. Calpe-Maravilla, "Biophysical parameter estimation with adaptive Gaussian Processes," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.* (IGARSS'2009), Capetown, South Africa, 2009.

[44] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, M. Lázaro-Gredilla, and J. Verrelst. (Jun. 2013). "simpleR: A simple educational MATLAB toolbox for statistical regression, v 2.1" [Online]. Available: http://www.uv.es/gcamps/code/simpleR.html.

[45] 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.

[46] J. Verrelst, J. Rivera, G. Leonenko, L. Alonso, and J. Moreno, "Optimizing LUT-based RTM inversion for semiautomatic mapping of crop biophysical parameters from Sentinel-2 and -3 data: Role of cost functions," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 257–269, Jan. 2014.

[47] J. Rivera, J. Verrelst, J. Delegido, F. Veroustraete, and J. Moreno, "On the semi-automatic retrieval of biophysical parameters based on spectral index optimization," *Remote Sens.,* 2014 (in review).

[48] 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.

[49] S. Gandía, G. Fernández, J. C. Garcia, and J. Moreno, "Retrieval of vegetation biophysical variables from CHRIS/PROBA data in the SPARC campaign," in *Proc. 2nd CHRIS/PROBAWorkshop*, Frascati, Italy: ESA/ESRIN, 2004.

[50] G. Fernández, J. Moreno, S. Gandía, B. Martínez, F. Vuolo, and F. Morales, "Statistical variability of field measurements of biophysical parameters in SPARC-2003 and SPARC-2004 campaigns," in *Proc. SPARC Workshop*, Frascati, Italy: ESA/ESRIN, 2005.

[51] M. J. Barnsley, J. J. Settle, M. A. Cutter, D. R. Lobb, and F. Teston, "The PROBA/CHRIS mission: A low-cost smallsat for hyperspectral multiangle observations of the earth surface and atmosphere," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1512–1520, Jul. 2004.

[52] L. Alonso and J. Moreno, "Advances and limitations in a parametric geometric correction of CHRIS/PROBA data," in *Proc. 3rd CHRIS/Proba Workshop*, Frascati, Italy: ESA/ESRIN, 2005.

[53] L. Guanter, L. Alonso, and J. Moreno, "A method for the surface reflectance retrieval from PROBA/CHRIS data over land: Application to ESA SPARC campaigns," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 12, pp. 2908–2917, Dec. 2005.

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

[55] M. Lazaro-Gredilla, M. 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, Apr. 2014.

[56] X. Shen, H. Wu, and Q. Zhu, "Training support vector machine through redundant data reduction," in *ACM Int. Conf. Proc. Ser.,* 2012 pp. 25–28.

[57] S. Wang, Z. Li, and X. Zhang, "Bootstrap sampling based data cleaning and maximum entropy SVMs for large datasets," in *Proc. Int. Conf. Tools Artif. Intell.* (ICTAI), 2012, vol. 1, pp. 1151–1156.

[58] 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, pp. 606–617, 2011.

Juan Pablo Rivera Caicedo received the B.Sc. degree in agricultural engineering from the National University of Colombia, Bogotá, Colombia and University of Valle, Cali, Colombia, in 2001, the Master degree in irrigation engineering from the Centro de Estudios y Experimentación de Obras Publicas (CEDEX), Spain, in 2003, and the M.Sc. degree in remote sensing from the University of Valencia, Valencia, Spain.

Currently, he is involved in preparatory activities of the Fluorescence Explorer (FLEX). His research interests include retrieval of vegetation properties using airborne and satellite data, leaf and canopy radiative transfer modeling, and hyperspectral data analysis. His research activities led to the scientific toolbox ARTMO: http://ipl.uv.es/artmo.

Since January 2011, he is a member of the Laboratory for Earth Observation (LEO), the Image Processing Laboratory, University of Valencia, as a Ph.D. candidate.

Jochem Verrelst received the M.Sc. degree in tropical land use and geo-information science, and the Ph.D. degree in remote sensing from Wageningen University, Wageningen, Netherlands, in 2005 and 2010, respectively. The subject of his dissertation was on the space-borne spectrodirectional estimation of forest properties.

From 2010 to 2012, he has been a Marie Curie Postdoctoral Fellow at the Laboratory for Earth Observation (LEO), the Image Processing Laboratory, University of Valencia, Valencia, Spain. Currently, he is involved in preparatory activities of the Fluorescence Explorer (FLEX). His research interests include retrieval of vegetation properties using airborne and satellite data, canopy radiative transfer modeling, multi-angular and hyperspectral data analysis. His research activities led to the scientific toolbox ARTMO: http://ipl.uv.es/artmo.

Jordi Muñoz-Marí (M'11) received the B.Sc. degree in physics, the M.Sc. degree in electronics engineering, and the Ph.D. degree in electronics engineering (2003) from the Universitat de València, València, Spain, in 1993, 1996, and 2003, respectively.

Currently, he is an Associate Professor with the Electronics Engineering Department, Universitat de València, where he teaches electronic circuits and, programmable logical devices, digital electronic systems, and microprocessor electronic systems. His research interests are mainly related to the development of machine learning algorithms for signal and image processing. Visit http://www.uv.es/jordi/ for more information.

José Moreno (A'89–M'09) is currently with the Department of Earth Physics and Thermodynamics, Faculty of Physics of the University of Valencia, Valencia, Spain, as a Professor of earth physics, teaching and working on different projects related to remote sensing and space research as responsible for the Laboratory for Earth Observation. His main work is related to the modeling and monitoring of land surface processes by using remote sensing techniques. He has been involved in many international projects and research networks, including the preparatory activities and exploitation programmes of several satellite missions (ENVISAT, CHRIS/PROBA, GMES/Sentinels, and SEOSAT) and the Fluorescence Explorer (FLEX), and as a Candidate ESA Earth Explorer Mission. He is a Director of the Laboratory for Earth Observation (LEO) at the Image Processing Laboratory/Scientific Park.

Dr. Moreno has served as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (1994–2000) and has been a Member of the ESA Earth Sciences Advisory Committee (1998–2002), the Space Station Users Panel, and other international advisory committees.

Gustavo Camps-Valls (M'04–SM'07) received the Ph.D. degree in physics from the Universitat de València, Valencia, Spain, in 2002.

Currently, he is an Associate Professor with the Department of Electronics Engineering and a Leading Researcher with the Image Processing Laboratory (IPL), Universitat de València. Recently included in the ISI lists as a highly cited researcher. He co-edited the books "Kernel methods in bioengineering, signal and image processing" (IGI, 2007) and "Kernel methods for remote sensing data analysis" (Wiley & Sons, 2009).

Dr. Camps-Valls serves on the Program Committees of International Society for Optical Engineers (SPIE) Europe, International Geoscience and Remote Sensing Symposium (IGARSS), Machine Learning for Signal Processing (MLSP), and International Conference on Image Processing (ICIP). Since 2007, he is a Member of the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society and since 2009, he is a Member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society. He is involved in the EUMETSAT MTG-IRS Science Team. He is an Associate Editor of the "IEEE TRANSACTIONS ON SIGNAL PROCESSING," "IEEE SIGNAL PROCESSING LETTERS," and "IEEE GEOSCIENCE AND REMOTE SENSING LETTERS." Visit http://www.uv.es/gcamps for more information.