Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3

Jochem Verrelst *, Jordi Muñoz, Luis Alonso, Jesús Delegido, Juan Pablo Rivera, Gustavo Camps-Valls, José Moreno

Image Processing Laboratory (IPL), University of Valencia, Spain

ARTICLE INFO

ABSTRACT

ESA's upcoming satellites Sentinel-2 (S2) and Sentinel-3 (S3) aim to ensure continuity for Landsat 5/7, SPOT-5, SPOT-Vegetation and Envisat MERIS observations by providing superspectral images of high spatial and temporal resolution. S2 and S3 will deliver near real-time operational products with a high accuracy for land monitoring. This unprecedented data availability leads to an urgent need for developing robust and accurate retrieval methods. Machine learning regression algorithms may be powerful candidates for the estimation of biophysical parameters from satellite reflectance measurements because of their ability to perform adaptive, nonlinear data fitting.

By using data from the ESA-led field campaign SPARC (Barrax, Spain) we have compared the utility of four state-of-the-art machine learning regression algorithms and four different S2 and S3 band settings to assess three important biophysical parameters: leaf chlorophyll content (*Chl*), leaf area index (LAI) and fractional vegetation cover (FVC). The tested Sentinel configurations were: S2-10 m (4 bands), S2-20 m (8 bands), S2-60 m (10 bands) and S3-300 m (19 bands), and the tested methods were: neural networks (NN), support vector regression (SVR), kernel ridge regression (KRR), and Gaussian processes regression (GPR).

GPR outperformed the other retrieval methods for the majority of tested configurations and was the only method that reached the 10% precision required by end users in the estimation of *Chl*. Also, although validated with an RMSE accuracy around 20%, GPR yielded optimal LAI and FVC estimates at highest S2 spatial resolution of 10 m with only four bands. In addition to high accuracy values, GPR also provided confidence intervals of the estimates and insight in relevant bands, which are key advantages over the other methods. Given all this, GPR proved to be a fast and accurate nonlinear retrieval algorithm that can be potentially implemented for operational monitoring applications.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

The lifetime of current Earth Observation (EO) sensors such as Landsat 5/7 and SPOT-5 is coming to an end. Any gap in data availability would affect on-going monitoring programs. To guarantee the availability of data to service providers and users, Global Monitoring for Environment and Security (GMES) took the initiative to a new generation of environmental Earth Observation missions, the so-called Sentinel missions (Martimort et al., 2007). GMES is a joint initiative of the European Commission and the European Space Agency (ESA), founded to establish a European capacity for the provision and use of operational monitoring information from space for environment and security applications (http://www.gmes.info).

Five different Sentinel concepts have been planned. In particular, Sentinel-2 (S2) and Sentinel-3 (S3) are designed to provide continuity to monitoring services over global terrestrial surfaces by relying on superspectral (more than ten and less than 50 bands, i.e. in-between multispectral and hyperspectral resolution) high resolution (S2) and medium resolution (S3) observations. At the same time, further improvements of existing monitoring services from space is needed to better understand the environment dynamics at local and global scales (Hilker et al., 2008). Therefore, along with these new missions, there is a demand of enhanced operational retrieval strategies of relevant biophysical parameters (Martimort et al., 2007).

Leaf chlorophyll content (*Chl*), green leaf area index (LAI) and fractional vegetation cover (FVC) are among the most important biophysical parameters retrievable from satellite reflectance data (Lichtenthaler, 1987; Whittaker & Marks, 1975). These parameters give insight in the phenological stage and health status (e.g., development, productivity, stress) of crops and forests. *Chl* can be considered a bio-indicator of the plant's actual health status (Lichtenthaler et al., 1996; Sampson et al., 2003),and of vegetation gross primary productivity (Gitelson et al., 2006). Besides, LAI and FVC characterize the structure and functioning of vegetation cover (Hirose et al., 1997).

* Corresponding author. Tel.: +34 963544067; fax: +34 96 354 32 61.
E-mail address: jochem.verrelst@uv.es (J. Verrelst).

Because of its role as the interface between ecosystem and atmosphere and involvement in many processes, *Chl*, FVC and LAI are crucial in aboveground biomass estimation, vegetative evapo-transpiration calculation and the energy-exchange evaluation of terrestrial vegetation (Baret & Guyot, 1991; Bicheron & Leroy, 1999; Buermann et al., 2001; Kakani et al., 2003).

When it comes to the implementation of candidate retrieval methods into operational Sentinel data processing chains, it is mandatory to invest in models that are both accurate and robust. The GMES objective for near real-time delivery requires that the dependency on ancillary data should be kept to the minimum (Martimort et al., 2007). Simple, yet widely accepted, empirical methods such as those based on vegetation indices, red-edge position, or spectral integral approaches are prone to perform unstable when applied to images which differ from those where the method was designed for (Verrelst et al., 2010). Conversely, despite being effective in local campaigns, physically-based methods such as inverting a radiative transfer (RT) model require site-specific information for proper model parameterization, which is not always instantly available. Hence, in view of monitoring applications across the globe, methods based on vegetation indices may be too simplistic while physical models may be too cumbersome for delivering near real-time and accurate products (Baret & Buis, 2008).

Alternatively, machine learning regression algorithms have the potential to generate adaptive, robust relationships and, once trained, they are very fast to apply (Hastie et al., 2009). 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 (e.g. Knudby et al. (2010)). They may therefore be more suitable candidates for operational applications. Effectively, algorithms such as neural networks (NNs) are already implemented in operational retrieval chains (e.g. Schiller & Doerffer (2005); Pozdnyakov et al. (2005); Verger et al. (2008, 2011)). It remains nevertheless to be questioned whether NNs offer the most flexible tools for parameter estimation, gaining insights 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 is why in the recent years, NNs are being replaced by more advanced, simpler to train regression methods. Specifically, during the last two decades, the family of *kernel methods* (Camps-Valls & Bruzzone, Dec, 2009) has emerged as an alternative to NNs in many scenarios. Kernel methods typically involve few and intuitive hyperparameters to be tuned, and can perform flexible input–output nonlinear mappings. In this work, we test the capabilities of four state-of-the-art machine learning regression algorithms given different S2 and S3 band settings: i) neural networks (Haykin, 1999), ii) support vector regression (Smola & Schölkopf, 2004), iii) kernel ridge regression (Shawe-Taylor & Cristianini, 2004), and iv) Gaussian processes regression (Rasmussen & Williams, 2006). The methods are compared in terms of accuracy, goodness of fit, robustness to low sample sized scenarios, and computational cost. The data used for the experiments came from the ESA-led field campaign SPARC, http://www.uv.es/leo/sparc/, which took place on the agricultural test site Barrax, Spain.

This setting brings us to the following main objective of the work: evaluation of the four machine learning regression algorithms on their ability to spectrally assess biophysical parameters given simulated S2 and S3 data. The potential use of the best algorithm will be further discussed for implementation in operational processing chains for monitoring services.

2. Sentinel

This section briefly describes the main characteristics of Sentinel-2 and −3, which form the basis for our simulated data.

2.1. Sentinel-2

The upcoming Sentinel-2 (S2) satellites capitalizes on the technology and the vast experience acquired with SPOT and Landsat over the past decades (Martimort et al., 2007). S2 is a polar-orbiting, superspectral high-resolution imaging mission. The mission is envisaged to fly a pair of satellites with the first planned to launch in 2013. Each S2 satellite carries a Multi-Spectral Imager (MSI) with a swath of 290 km. It provides a versatile set of 13 spectral bands spanning from the visible and near infrared (VNIR) to the shortwave infrared (SWIR), featuring four bands at 10 m, six bands at 20 m and three bands at 60 m spatial resolution. S2 incorporates three new bands in the red-edge region, which are centered at 705, 740 and 783 nm. Band settings are provided in Table 1. The pair of S2 satellites aims to deliver data taken over all land surfaces and coastal zones every five days under cloud-free conditions, and typically every 15–30 days considering the presence of clouds (Drusch et al., 2010). To serve the objectives of GMES, S2 satellites will provide imagery for the generation of high-level operational products (level 2b/3) such as land-cover and land-change detection maps and geophysical variables such as *Chl*, LAI and leaf water content maps (Drusch et al., 2010). To ensure that the final product can meet the user requirements, the GMES user committee defined an accuracy goal of 10% (Drusch et al., 2010).

2.2. Sentinel-3

The pair of Sentinel-3 (S3) satellites will provide global, frequent and near real-time ocean, ice and land monitoring. It continues ENVISAT's altimetry, the superspectral, medium-resolution VNIR ocean and land-surface observations of ERS, ENVISAT and SPOT-Vegetation, and includes enhancements to meet the operational revisit requirements and to facilitate new products and evolution of services. S3 will be equipped with the Ocean and Land Colour Instrument (OLCI), which will provide continuity of the existing MERIS mission. Six new bands have been added upon recommendations to improve the existing MERIS atmospheric and aerosol correction capabilities (Nieke et al., 2009). The OLCI ground resolution requirement depends on where the data are acquired: above open ocean, or over coastal zones and land. OLCI products require a spatial resolution at sub-satellite point of 1200 m over open ocean and sea ice, and 300 m over coastal zones, while land products require a resolution of 300 m globally. OLCI aimed to be optimized to measure the ocean color over the open ocean and coastal zones, but in such a way that it will not saturate over land targets. Its spectral bands are in the VNIR spectral range (from 403 to 1040 nm) with bandwidths ranging from 3.75 to 40 nm. Band settings are provided in Table 2.

Table 1
Sentinel-2 MSI band settings.

Band #	B1	B2	B3	B4	B5	B6	B7	B8	B8a	B9	B10	B11	B12
Band center (nm)	443	490	560	665	705	740	783	842	865	945	1375	1610	2190
Band width (nm)	20	65	35	30	15	15	20	115	20	20	30	90	180
Spatial resolution (m)	60	10	10	10	20	20	20	10	20	60	60	20	20

Table 2
Sentinel-3 OLCI band settings.

Band #	O1	O2	O3	O4	O5	O6	O7	O8	O22	O9	O10
Band center (nm)	400	412.5	443	490	510	560	620	665	673.75	681	709
Band width(nm)	15	10	10	10	10	10	10	10	7.5	7.5	10

Band #	O11	O12	O13	O23	O14/15	O16/17	O18	O19	O20	O21
Band center (nm)	754	761	764.375	767.5	779	865	885	900	940	1020
Band width(nm)	7.5	2.5	3.75	2.5	15	20	10	10	20	40

3. Machine learning regression algorithms

Machine learning approaches learn the relationship between the input (e.g. reflectances) and output (e.g. biophysical parameters) by fitting a flexible model directly from 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. We considered four state-of-the-art machine learning regression algorithms for the purpose of comparison, which were i) neural networks (NN), ii) support vector regression (SVR), iii) kernel ridge regression (KRR), and iv) Gaussian processes regression (GPR). All these regression methods are popular in various application domains thanks to its relatively fast training, good performance, and robustness to the overfitting problem.

3.1. Neural networks

The most common approach to develop nonparametric and nonlinear regression is based on artificial neural networks (NN) (Haykin, 1999). A NN is a (potentially fully) connected structure of neurons organized in layers. A neuron basically performs a linear regression 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 a NN with only one neuron, the results would be similar (or slightly better because of the nonlinearity) than those obtained with linear regression. In this work, we have used the standard multi-layer NN model (Fig. 1), whose neuron j in layer $l+1$ yields $x_j^{l+1} = f(\sum_i w_{ij}^l x_i^l + w_{bj}^l)$, where w_{ij}^l are the weights connecting neuron i in layer l to neuron j in layer $l+1$, w_{bj}^l are the bias term of neuron j in layer l, and f is a nonlinear activation function (in our case the hyperbolic tangent was used). The output (prediction) of the model for the sample $\mathbf{x}_i$ is denoted as $f(\mathbf{x}_i)$. Training a NN needs selecting a structure (number of hidden layers and nodes per layer), properly initialization of 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 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.
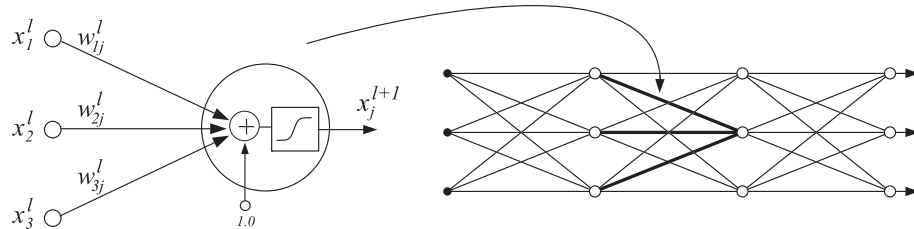
3.2. Support vector regression

A successful alternative to the use of NNs is the support vector regression (SVR) method (Schölkopf & Smola, 2002; Smola & Schölkopf, 2004). The method defines a linear prediction model over mapped samples to a much higher dimensional space, which is nonlinearly related to the original input space of reflectances (B bands). Given a set of training input–output pairs, $\{\mathbf{x}_i, y_i\}_{i=1}^n$, the SVR defines a mapping $\phi : \mathbf{x}_i \rightarrow \phi(\mathbf{x}_i) \in \mathbb{R}^H$, where $H \geq B$. The SVR prediction model is $y_i = \mathbf{w}^\top \phi(x_i) + b$. The standard SVR formulation uses Vapnik's ε-insensitive cost function, in which errors e_i up to ε are not penalized, and all further deviations will incur in a linear penalization. Briefly, SVR estimates weights $\mathbf{w}$ by minimizing the following regularized functional:

$$\frac{1}{2}\|\mathbf{w}\|^2 + C \sum_i \left(\xi_i + \xi_i^*\right) \tag{1}$$

with respect to $\mathbf{w}$ and $\{\xi_i^{(*)}\}_{i=1}^n$, constrained to:

$$y_i - \mathbf{w}^\top \phi(\mathbf{x}_i) - b \leq \varepsilon + \xi_i \quad \forall i = 1, \dots, n \tag{2}$$

$$\mathbf{w}^\top \phi(\mathbf{x}_i) + b - y_i \leq \varepsilon + \xi_i^* \quad \forall i = 1, \dots, n \tag{3}$$

$$\xi_i, \xi_i^* \geq 0 \qquad \forall i = 1, \dots, n \tag{4}$$

where $\xi_i^{(*)}$ are positive slack variables to deal with training samples with a prediction error larger than ε ($\varepsilon > 0$), and C is the penalization parameter applied to these ones. Note that C controls the trade-off between the minimization of errors and the regularization term, thus controlling the generalization capabilities. The usual procedure for solving SVRs introduces the linear restrictions (2)-(2) into(1) using Lagrange multipliers α_i, computes the Karush-Kuhn-Tucker conditions, and solves the dual problem using standard quadratic programming procedures, which yields the final solution:

$$\hat{y}_i = \sum_{j=1}^n \left(\alpha_j - \alpha_j^*\right) K\left(\mathbf{x}_i, \mathbf{x}_j\right) + b, \tag{5}$$

where K is a *kernel* function that is expressed as the dot product of mapped examples $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$. The kernel function plays a crucial role since it implements the notion of similarity between data points, i.e. between reflectances. Examples of typical kernel functions are the linear, $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^\top \mathbf{x}_j$, the polynomial $K(\mathbf{x}_i, \mathbf{x}_j) = \left(\mathbf{x}_i^\top \mathbf{x}_j + 1\right)^d$, or the one used in our implementation, the Gaussian

Fig. 1. Illustration of a feedforward neural network. A set of neurons (on the leftmost part of the figure) is included in a structure of connected neurons.

Function (RBF) kernel, $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|2/(2\sigma^2))$. Therefore, in SVR, three free parameters are tuned: the regularization parameter C, the tolerance value ε, and the kernel parameter σ. We used a cross-validation strategy for their optimization. The model is very fast to train and apply. Note that in (2), only the $m \leq n$ training examples $\mathbf{x}_j$ whose corresponding Lagrange multipliers α_j are non-zero count in the solution, which are called *support vectors* (SVs). Sparsity in the SVR is a direct consequence of the loss function; as the value of ε increases, the number of support vectors is reduced. This feature is crucial to avoid overfitting.

3.3. Kernel ridge regression

The previous SVR model uses a particularly robust cost function, and additional linear constraints. This gives rise to a quadratic programming (QP) problem to be solved. When Gaussian noise is assumed to be present in the observation, the squared loss is optimal from a maximum likelihood point of view. The Kernel Ridge Regression (KRR) minimizes the squared residuals in a higher dimensional feature space, and can be considered as the kernel version of the regularized least squares linear regression (Schölkopf & Smola, 2002; Shawe-Taylor & Cristianini, 2004). Notationally, we want to minimize the regularized squared loss function,

$$\frac{1}{2}\|w\|^2 + \frac{C}{2}\sum_i \xi_i^2, \tag{6}$$

which is subject to an equality constraint $\xi = y_i - (\mathbf{w}^\top \phi(\mathbf{x}_i) + b)$. The usual solution involves several steps. Replacing the constraint into Eq. (6), taking derivatives wrt model weights $\mathbf{w}$, and equating them to zero, leads to an equivalent problem depending on the unknown mapping function ϕ. 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 set $\mathbf{X}_*$ is obtained as a function of the dual weights $\alpha = [\alpha_1, ..., \alpha_n]^\top$ (one per sample), as follows:

$$f(\mathbf{X}_*) = K(\mathbf{X}_*, \mathbf{X})\left(K(\mathbf{X}, \mathbf{X}) + \frac{1}{C}\mathbf{I}\right)^{-1}\mathbf{y} = K(\mathbf{X}_*, \mathbf{X})\alpha, \tag{7}$$

which only involves computing similarities between training and testing samples via kernel functions. Also, note that for obtaining the model, only the inversion of the Gram (or kernel) matrix $\mathbf{K}$ of size $n \times n$ regularized by $1/C$ is needed. Therefore, in KRR only the regularization parameter C and the kernel parameters have to be tuned. For the latter, we again used the RBF kernel for the sake of fair comparison. Parameters were again optimized via cross-validation. Note that the prediction model is exactly the same as for the SVR. The main advantage of KRR is that the solution can be expressed in closed-form, so QP procedures are not needed. The main shortcoming is that the model is not sparse: all examples used for training have a weight α_i in the final solution.

3.4. Gaussian Processes Regression (GPR)

Gaussian processes regression (GPR) has been recently introduced as a powerful regression tool (Pasolli et al., 2010; Rasmussen & Williams, 2006). The model provides a *probabilistic* approach for learning generic regression problems with kernels. The GPR model establishes a relation between the input (B-bands spectra) $\mathbf{x} \in \mathbb{R}^B$ and the output variable (canopy parameter) $y \in \mathbb{R}$ in the same way as SVR or KRR (see Eq. 7). However, two main advantages of GPR must be noted.

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

$$\mathbb{V}[f(\mathbf{X}_*)] = K(\mathbf{X}_*, \mathbf{X}_*) - K(\mathbf{X}_*, \mathbf{X})\left(K(\mathbf{X}, \mathbf{X}) + \frac{1}{C}\mathbf{I}\right)^{-1}K(\mathbf{X}, \mathbf{X}_*) \tag{8}$$

Note that the mean prediction in Eq. (7) is a linear combination of observations $\mathbf{y}$, while the predictive variance in Eq. (8) 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 in the GPR model, because all hyperparameters can be learned efficiently by maximizing the marginal likelihood in the training set. See Rasmussen & Williams (2006); Camps-Valls et al. (2009); Verrelst et al. (in press) for further details. We used a scaled Gaussian kernel function,

$$K\left(\mathbf{x}_i, \mathbf{x}_j\right) = \nu \exp\left(-\sum_{b=1}^B \frac{\left(\mathbf{x}_i^{(b)} - \mathbf{x}_j^{(b)}\right)^2}{2\sigma_b^2}\right), \tag{9}$$

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 gives 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 interested reader is referred to the excellent book (Rasmussen & Williams, 2006) for more details on the theoretical aspects of GPR. A Matlab implementation of GPR is freely available at http://www.gaussianprocess.org/gpml/.

4. Methodology

4.1. SPARC database

The database used for training the models should be representative of actual plant canopies and conditions as observed by the Sentinel sensors. 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 and SPARC-2004 campaigns took place in Barrax, La Mancha, Spain (coordinates 303'N, 286'W, 700 m altitude). The test area has a rectangular form and an extent of 5 km × 10 km, and is characterized by a flat morphology and large, uniform land-use units. The region consists of approximately 65% dry land and 35% irrigated land. The annual rainfall average is about 400 mm.

In the 2003 campaign (12–14 July) biophysical parameters were measured within a total of 113 Elementary Sampling Units (ESU) among different crops. ESU refers to a plot size of about 20^2 m. The same field data were collected in the 2004 campaign (15–16 July) within a total of 18 ESUs among different crops. For both years, within each ESU the averaged leaf *Chl* was derived by measuring about 50 samples with a calibrated CCM-200 Chlorophyll Content Meter (Gandía et al., 2004). Green LAI was derived from canopy measurements made with a LiCor LAI-2000 digital analyzer. Each ESU was assigned to a LAI value, which was obtained by the average of 24 measures (8 data readings × 3 replications) (Fernández et al., 2005). FVC were estimated from ground measurements using hemispherical photographs taken with a digital camera with a fish-eye lens (Jiménez-Muñoz et al., 2005). The final FVC estimate for each ESU was calculated as the average of twelve measurements. In total 9

crop types (garlic, alfalfa, onion, sunflower, corn, potato, sugar beet, vineyard and wheat) were sampled, with field-measured values of LAI that vary between 0.4 and 6.3, *Chl* between 2 and 55 $\mu g/cm^2$ and FVC between 0 and 1. Further details on the measurements can be found in the data acquisition report (ESA, 2004). Additionally, 30 random bare soil spectra with a biophysical (*Chl*, LAI, FVC) value of zero were added to broaden the dataset to non-vegetated samples.

4.2. Simulation of Sentinel-2 and -3 configurations

Sentinel MSI and OLCI band settings were simulated on the basis of Compact High Resolution Imaging Spectrometry (CHRIS) data. 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, site of the covered area and spatial resolution because of on-board memory storage reasons (Barnsley et al., 2004). The radiometric resolution of CHRIS is 12 bits, which is the same as S2's MSI. We made use of nominal nadir CHRIS observations in Mode 1 (62 bands, maximal spectral information) for the four SPARC campaign days, where field measurements of surface properties were measured in conjunction with satellite overpasses. CHRIS Mode 1 has a spatial resolution of 34 m at nadir. The spectral resolution provides a bandwidth from 5.6 to 33 nm depending on the wavelength. The images were geometrically corrected (Alonso & Moreno, 2005), followed by atmospheric correction according to the method proposed in Guanter et al. (2005). The nadir image from 12 July 2003 was used for spectral and spatial resampling to the settings of S2 and S3. Because featuring bands with different pixel sizes (10, 20 and 60 m), it is of special interest simulating S2 configurations as a function of pixel size. Nearest neighbor was used for the spatial resampling and a Gaussian model with FWHM spacings was used for spectral resampling. Constrained by the spectral range of CHRIS, the following Sentinel configurations were generated:

- S2 MSI: spatial resolution of 10 m, 4 bands. Includes the bands with pixel size 10 m, being bands B2, B3, B4 and B8 (490–665 and 842 nm). Hereafter referred as S2-10 m.
- S2 MSI: spatial resolution of 20 m, 8 bands. Includes the bands with pixel size 20 m, and bands with pixel size 10 m that were coarse-grained to 20 m, being bands B2 to B8a (490–865 nm). Hereafter referred as S2-20 m.
- S2 MSI: spatial resolution of 60 m, 10 bands. Includes the bands with pixel size 60 m, and bands with pixel size 10 m and 20 m that were coarse-grained to 60 m, being bands B1 to B9 (443–945 nm). Hereafter referred as S2-60 m.
- S3 OLCI: spatial resolution of 300 m, 19 bands. Includes all bands that fall within the CHRIS range, being bands O1 to O20 (413–940 nm) and were coarse-grained to 300 m. Hereafter referred as S3-300 m.

4.3. Experimental setup

The SPARC field dataset, along with the ensemble of canopy reflectance spectra according to the four Sentinel configurations, were divided into training and validation datasets. The performance of machine learning algorithms depends on the quality of the training data. In order to avoid skewed results, each method was run 50 times with different random realizations of training and testing partitions. The methods were compared against the standard least squares linear regression (LR) as a baseline reference. Model's performance was evaluated with the absolute root-mean-squared error (RMSE) and the relative RMSE (%) to assess accuracy, and the coefficient of determination (r^2) to account for the goodness-of-fit. The relative RMSE was used to compare the performances across the different methods and parameters.

The models were trained with varying training/testing partitions of increasing difficulty; from 50/50% up to 90/10% with increments of 10%. Averaged relative RMSE revealed that the machine learning regression methods performed slightly better when increasing training samples at the expense of testing samples, with best performance at a partition of 90% testing samples. When choosing an adequate sampling distribution, it is however important to strive for accurate retrievals over a wide range of land cover types. This requires that a wide range of training samples should be included in the generation of the model while sufficient testing samples should be kept aside for ensuring a meaningful validation. Henceforth, a partition of 80% (129 points) training and 20% (32 points) testing data was chosen.

5. Results

5.1. Method evaluation

Relative RMSE results of the regression methods for *Chl*, LAI and FVC given the S2-10 m (4 bands), S2-20 m (8 bands), S2-60 m (10 bands) and S3-300 m (19 OLCI bands) configurations are presented in Fig. 2. The error bars provide an idea of model robustness with respect to the input data. Pooling the samples 50 times hardly impacted the training performance of a model but, once validated, gave sign of some instability. Particularly, NN performed unstably once applied to validation data (e.g. see S2-20 m and S2-60 m). Hence, of importance for further analysis is how accurately a trained model performs when validated against ground reference measurements rather than the training performance. Given the validation results, GPR provided best accuracies for the majority of Sentinel configurations. GPR performed better than the other regression methods for all Sentinel configurations in estimating LAI, for S2-20 m and S2-60 m in estimating *Chl*, and for S2-10 m in estimating FVC. On the whole, SVR provided the second best accuracies.

Furthermore, all kernel methods (GPR, SVR and KRR) rendered robust results when having relatively few bands available. For LAI and FVC the performance of these methods hardly changed across the different S2 configurations (4, 8 and 10 bands). The relative RMSE stabilized around 20%. This suggests that these parameters can be optimally mapped at high spatial resolution of 10 m with only four bands. These bands, inherited from Landsat, SPOT and MERIS, are well known for their sensitivity to vegetation properties; B2: vegetation senescing and soil background, B3: total chlorophyll in vegetation, B4: maximal chlorophyll absorption, and importantly, B8: to LAI (Drusch et al., 2010).

For *Chl*, although obtaining a similar accuracy of about 20%, the S2-10 m configuration lacked bands in the red-edge for reaching optimal retrievals. GPR yielded more accurate *Chl* estimations with the S2-40 m and S2-60 m configurations, which include bands in the red-edge (B5: 705 nm, B6: 740 nm). From all tested cases, the best results were obtained by GPR for the estimation of *Chl* with the S2-60 m configuration. The model yielded a mean relative RMSE of 10.5% with a standard deviation (SD) of 3.2%, and was on the order of the GMES goal accuracy of ≤10%. The overall good performance throughout all Sentinel configurations proved that GPR acts as a powerful regressor, not only for superspectral data, but also for multispectral data.

In fact, despite having 19 bands available, S3-300 m did not lead to superior results. Accuracies were not optimal, probably due to the coarse spatial resolution of S3-300 m compared to the field sampling of SPARC. The majority of field measurements were employed within agricultural fields (and not in surrounding fallow land), which may lead to inconsistencies with the composite estimates at S3-300 m resolution. Further, a similar study demonstrated that GPR captures the majority of relevant spectral variation by less than four bands, while additional bands hardly improves the fittings (Verrelst et al., in press). This may be another reason why the performance of S3-300 m is similar to those of the S2 configurations.
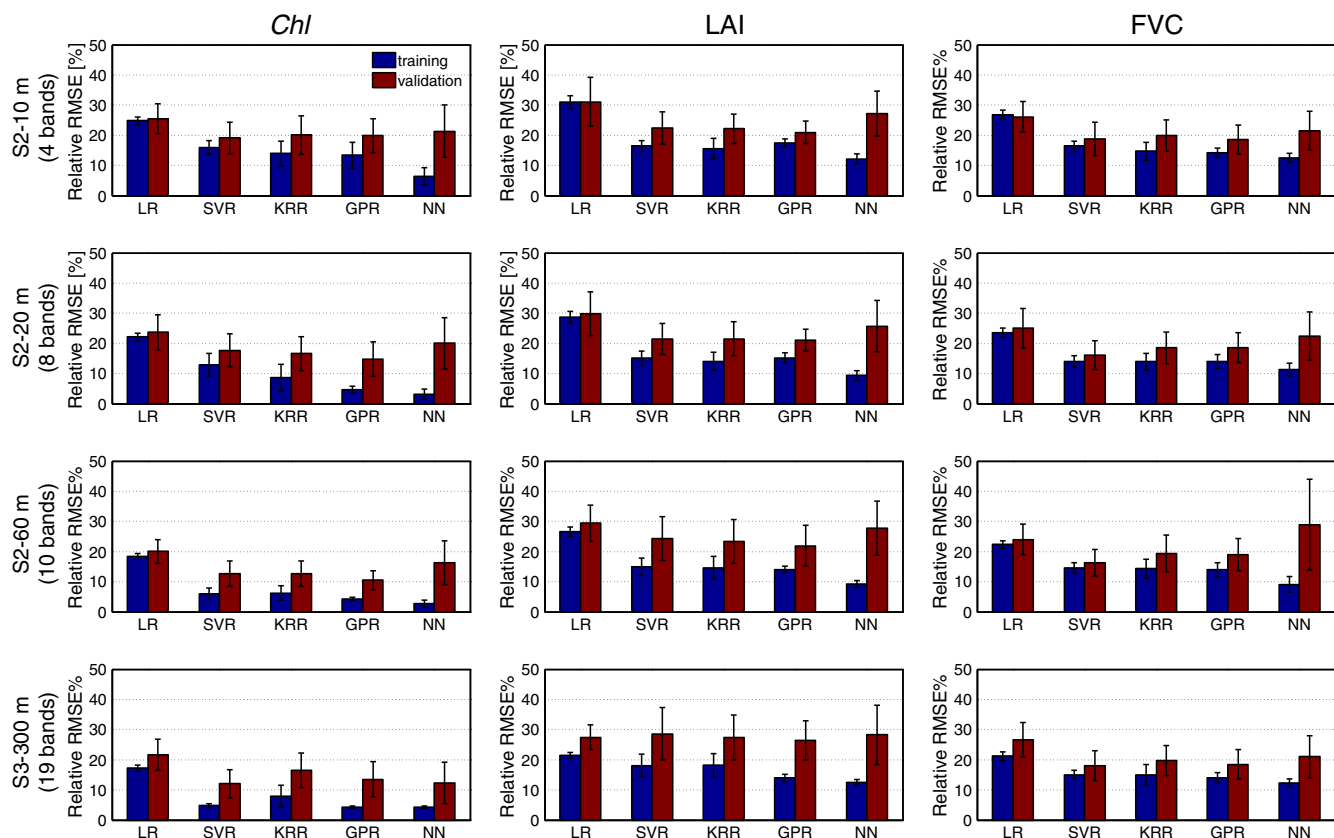
Fig. 2. Relative RMSE (%) results (mean and ±1 SD) for *Chl*, LAI and FCV retrievals using LR, KRR, GPR and NN for the configurations of S2-10 m, S2-20 m, S2-60 m and S3-300 m.

5.2. Processing time

As shown in the previous section, GPR proved to act as a robust regressor. Another criterion for being a successful candidate into an operational processing chain is the processing speed of both developing (training) and testing the regression model. For each scenario the time to fulfill a series the time to fulfill a series of 50 replications was recorded for evaluation purposes. All calculations were done in a Matlab environment on a Windows XP Intel(R) Core(TM)2 Quad CPU, 2.4 GHz, 3.00 GB RAM processor. An overview of the averaged elapsed time for the different Sentinel settings is shown in Table 3. Standard linear regression (LR) hardly took computational time. Also KRR proved to be a fast regressor, whereas SVR needed considerably more time to develop a model, particularly when having only few bands available (up to 11 seconds). Though, these models run relatively fast compared to NN. For instance, NN computed a S3-300 m model in about 55 seconds. The complex optimization during the learning process of the network occurred at the expense of a long computational time. But once calibrated, its application on actual data is almost instantaneous. In comparison, GPR computed a S3-300 m model in about 2 seconds, and a S2-10 m model in even less than a second. Such fast running time suggests that GPR is computationally efficient.

Table 3
Mean elapsed time (in seconds) and standard deviation (SD) for developing a model per regression method, parameter and Sentinel configuration. SD were not provided when they were in order of milliseconds.

Sentinel	LR	SVR	KRR	GPR	NN
S2-10 m	0.012	11.0(3.1)	0.6	0.7	6.4(1.2)
S2-20 m	0.012	9.7(2.0)	0.6	1.0	16.2(4.7)
S2-60 m	0.012	9.2(1.4)	0.6	1.2	22.7(7.3)
S3-300 m	0.012	5.11(0.6)	0.6	2.0	55.4(45.8)

Summarizing, from the tested regression methods, GPR provided the most accurate estimations within a relatively fast running time. Therefore, GPR's predictive features across the different Sentinel configurations will be further analyzed in the next sections.

5.3. GPR band analysis

An interesting feature of GPR is that it immediately provides insight into the relevance of contributing samples and bands, hence overcoming the 'black box' concern of previous methods such as NN. Specifically, band ranking may reveal the bands that contribute most to the development of a GPR model. At the same time, because of being a data-driven method it should be kept in mind that band ranking results are restricted to the used training dataset. Nevertheless, by sorting the bands on their relevance and counting the band rankings over the 50 repetitions some general trends can be derived. In a similar work where the importance of single bands were evaluated using the same dataset, it was demonstrated that the four most relevant bands contributed about 96% as opposed to having all bands included in developing a GPR model (Verrelst et al., in press). Thereby, the top ranked band alone contributed already about 81%, and is thus the band that matters mostly. Tracking the four most relevant bands was also pursued here, as so their positions can be compared across all Sentinel scenarios (Fig. 3).

Starting from S2-10 m, all bands reached top rankings given the three parameters. This can be partly explained by the low number bands, which increases the chance that eventually every band is granted with a top ranking, though large differences among the bands are notable. The band centered at 490 nm was least top ranked, while the band centered at 842 nm was most frequently top ranked. The latter NIR band was evaluated as most important in establishing a relationship with FVC for all 50 repetitions, and also for *Chl* this
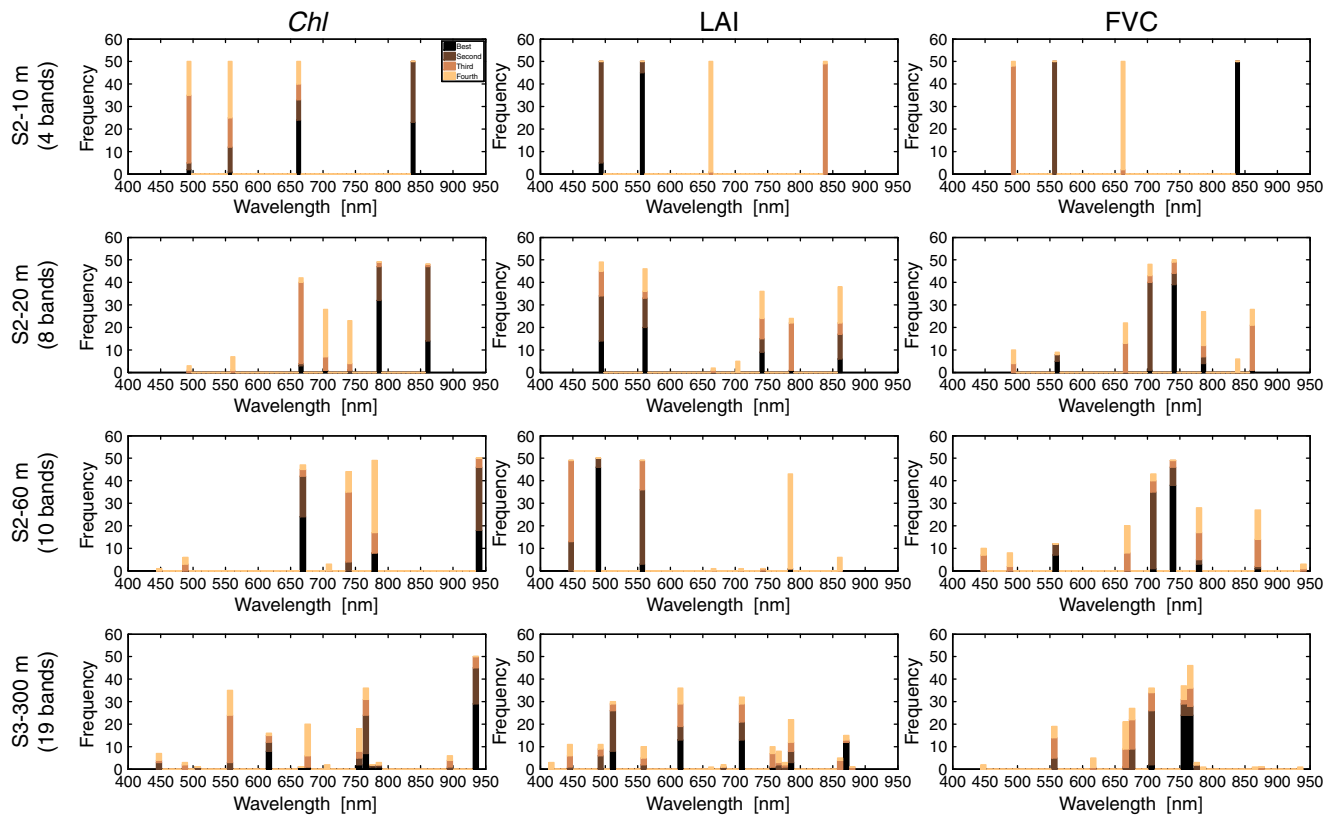
Fig. 3. GPR four most relevant bands for *Chl*, LAI and FCV and the configurations of S2-10 m, S2-20 m, S2-60 m and S3-300 m.

band was half of the times evaluated as most relevant band. For LAI, the most relevant band appeared to be centered at 550 nm. When having more bands added, such as in S2-20 m configuration, the aforementioned dominance diversified into more specific spectral regions. For *Chl*, the relevant spectral region broadened between 670 nm and 865 nm. New bands in the red-edge centered at 705 nm and 740 nm gained importance. For LAI, the 490–550 nm region still dominated, but now also new bands in the NIR 740–865 nm region played a major role. For FVC, the dominant spectral region tended to converge towards the red-edge, with the most relevant band centered at 740 nm. With atmospheric bands (B1 and B9) added in S2-60 m the configuration, for *Chl* most relevant bands were at 665 nm and also at 945 nm, which is an atmospheric band. For LAI, the relevant spectral region moved towards shorter wavelengths with 490 nm as dominant band, but now B1 also played a role. The new bands hardly delivered a significant contribution for FVC; the red-edge 740 nm band still dominated. The S3-300 m situation was slightly different as twice more bands were available with somewhat different band positions. For *Chl*, the most relevant band emerged in the NIR centered at 940 nm while the bands centered around 620 nm and 761 nm were of second importance. For LAI, almost the whole VNIR was evaluated as important, with relevant bands centered at 510 nm, 620 nm, 709 nm and 885 nm. For FVC, the same trend was observed as in the other Sentinel scenarios, with dominant bands positioned between 754 nm and 761 nm.

5.4. GPR pixel-wise retrievals of biophysical parameters

GPR was subsequently used to pixel-wise estimate *Chl*, LAI and FVC. For each parameter and Sentinel configuration a single GPR model was trained and validated with a random 80/20% dataset partition and then run over the simulated Sentinel image. The retrieval process was completely automated and image-based. Training and validation RMSE accuracies of each model are provided in Table 4.

Excellent validation accuracies were obtained with the estimation of *Chl* for the configurations of S2-20 m, S2-60 m and S3-300 m. The prediction of validation data was almost perfect ($r^2 = .99$) and RMSE stayed well below the by GMES defined threshold of 10% (7.1, 8.3 and 6.6%, respectively). Also the LAI and FVC validation data were characterized by high prediction accuracies throughout all Sentinel scenarios (r^2 between 0.89 and 0.95).

Fig. 4 provides maps of the retrieved parameters for the different Sentinel scenarios. To start with the most detailed S2-10 m map, within-field variations are clearly perceivable by all three parameters. Particularly the pronounced spatial variation of *Chl* clearly marks the irrigated circular fields with green biomass. These irrigated fields are characterized by a *Chl* above 40 μg/cm², a LAI above 3 and a FVC

Table 4

Training (train) and validation (test) statistics (r^2, absolute RMSE and relative RMSE (%)) for the Sentinel configurations and the parameters using GPR.

| Parameter	r^2		abs. RMSE		rel. RMSE (%)	
	train	test	train	test	train	test
S2-10 m:						
Chl	0.943	0.931	4.83	5.36	17.1	19.7
LAI	0.942	0.910	0.43	0.51	16.6	23.7
FVC	0.964	0.954	0.07	0.08	14.8	16.5
S2-20 m:						
Chl	0.994	0.991	1.62	1.92	5.8	7.1
LAI	0.944	0.916	0.42	0.49	17.2	23.0
FVC	0.979	0.950	0.07	0.08	15.7	17.4
S2-60 m:						
Chl	0.996	0.987	1.24	2.31	4.4	8.3
LAI	0.962	0.922	0.35	0.47	14.3	22.3
FVC	0.960	0.951	0.08	0.08	15.7	17.4
S3-300 m:						
Chl	0.996	0.992	1.26	1.81	4.4	6.6
LAI	0.962	0.894	0.35	0.55	14.3	25.5
FVC	0.971	0.949	0.06	0.08	13.2	17.7
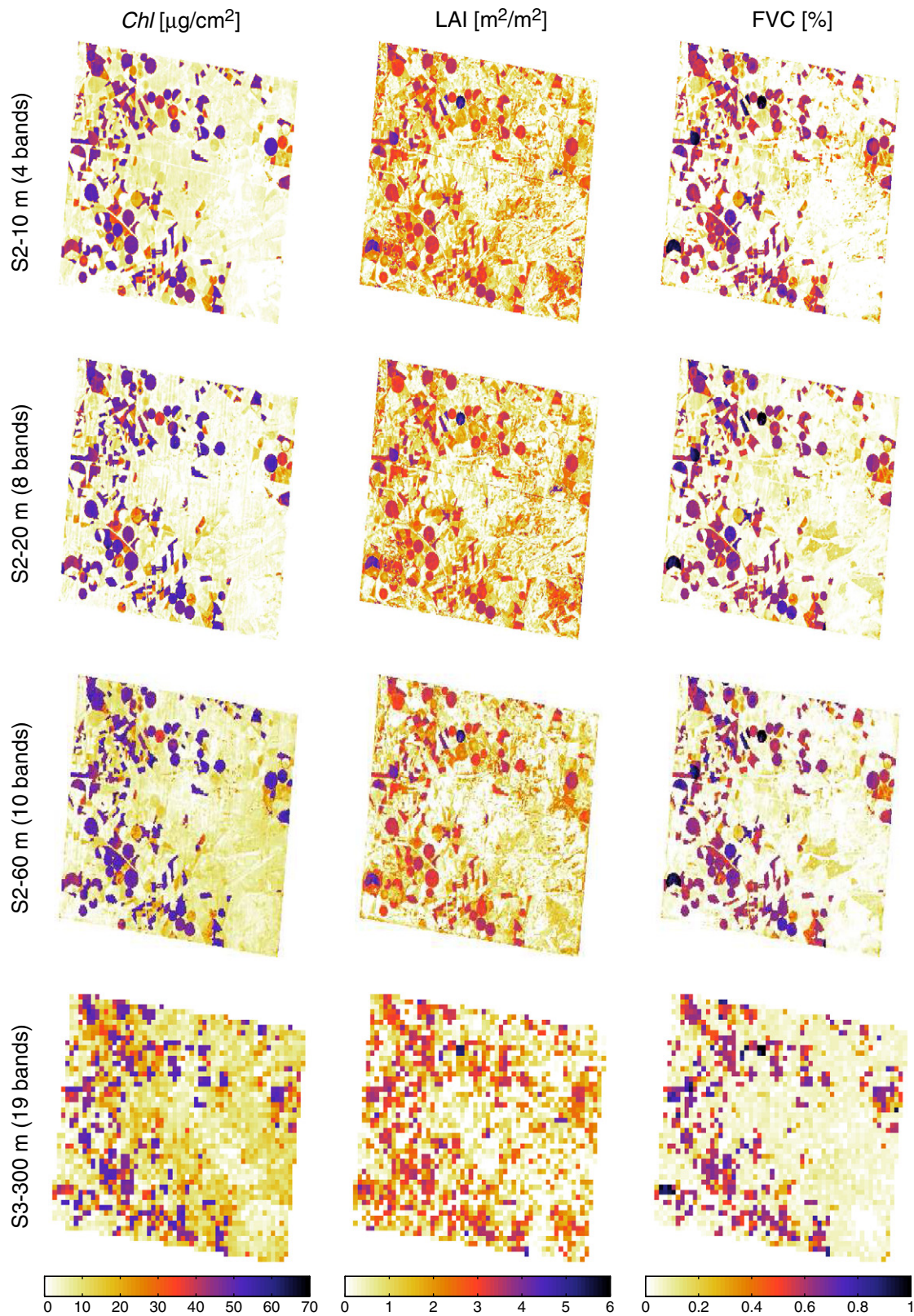
Fig. 4. GPR estimation maps for *Chl*, LAI and FCV, for the configurations of S2-10 m, S2-20 m, S2-60 m and S3-300 m.

above 60%. Areas with low *Chl*, LAI and FVC (the whitish parts) are mainly bare soils, fallow lands or rainfed senescing or harvested cereal fields (wheat, barley). The coarser S2-20 m and S2-60 m maps show essentially the same spatial patterns with about the same magnitudes of estimates. For instance, the circular fields and within-field

variability are still clearly observable. This implies that bands can be added to the benefit of improved retrievals at the expense of relatively small pixel coarsening. Overall, the different S2 scenarios demonstrated that S2s unique band setting, which allows flexibility in spatial and spectral configurations, is beneficial for crop monitoring

applications. Considerably coarser maps were obtained with the S3-300 m configuration. Although within-field spatial information has been mostly lost, spatial patterns of the irrigated fields were still observable. Nevertheless, small differences in retrievals as a consequence of the coarsening were notable. *Chl* retrievals no longer exceeded 60 μg/cm^2, most likely because of the merging of green irrigated areas with dry non-irrigated areas, leading to composite pixels. Such composite pixels are expected to occur frequently at the medium spatial resolution of S3, thereby smoothing out field-level variations.

5.5. GPR corresponding pixel-wise confidence values

A unique feature of GPR is that it provides associated confidence intervals along with the estimations. A confidence value here expresses the standard deviation (SD) that accompanies the mean estimation of a parameter. Intuitively, these confidence values are comparable to uncertainties encountered by field instruments. A degree of measurement uncertainty is inevitable, e.g. due to detector and external noise or due to a limited sensitivity margin of the detector. For instance, LAI instruments are typically unable to provide accurate values below a margin of 0.5. A LAI retrieval with a SD below 0.5 can therefore be considered as accurately retrieved. In turn, when a confidence value falls above 0.5 then it becomes more alerting. This does not necessarily mean that the estimate is wrong. Rather, it informs that the input spectrum deviates from what has been presented during the training stage, thereby imposing uncertainties to the retrieval.

Fig. 5 shows the confidence maps corresponding to the estimation maps. Since the same data were used, the confidence maps can be used to compare the retrieval performance across the different Sentinel configurations. When comparing the confidence maps it can be noted that the estimation of *Chl* led to poorer confidences (overall higher SDs) using S2-10 m data compared to configurations with more bands included. This is probably because of the aforementioned suboptimal band setting of S2-10 m for this parameter compared to the other Sentinel configurations (see also Fig. 2 and Table 4). With more bands available, the confidences greatly improved for the majority of the image, suggesting the relevance of the added bands for *Chl* retrieval. At the same time, it can be observed from the estimation maps that the differences in estimates are minimal. Although slightly poorer confidences were obtained with LAI and FVC, they stayed stable across the different Sentinel configurations. This suggests that adding bands hardly affected the retrieval performance of these parameters.

Within these maps, areas with reliable retrievals are clearly distinguished from areas with unreliable retrievals. Reliable retrievals (low SDs) were found on irrigated areas and harvested fields. Unreliable retrievals (high SDs) were found on areas with remarkably different spectra, such as bright, whitish calcareous soils (center, right), or harvested rainfed barley fields with remaining bright straws covering the surface (center). Particularly LAI suffered from poor confidences over the whitish calcareous soils in the right part of the image. FVC had only little problem with these soils when having only four bands available. Hence, a practical implication of the confidence maps is that they detect those areas that may benefit from a denser sampling regime.

6. Discussion

The upcoming S2 and S3 missions open opportunities to implement novel retrieval algorithms in operational processing chains. Of big interest are retrieval algorithms that are accurate, fast, robust, and sufficiently flexible to make full use of the new S2 MSI and S3 OLCI bands. Machine learning regression algorithms are able to cope with most of these objectives. Space agencies and monitoring services already make use of machine learning methods such as NN for

operational products (Pozdnyakov et al., 2005; Schiller and Doerffer, 2005; Verger et al., 2008, 2011). It remains however to be questioned whether these are the most adequate algorithms to fulfill all these requirements. In this paper the utility of SVR, KRR, GPR, NN were compared given various Sentinel configurations. NN faced two known disadvantages; it needed a relatively long processing time for training a model and it behaved relatively unpredictable when used with input spectra that deviate from what has been presented during the training stage (Atzberger, 2004; Baret & Buis, 2008; Kimes et al., 1998). Alternatively, GPR was evaluated as a more promising regressor in terms of processing speed and accuracy. Moreover, GPR is transparent in terms of model development (Pasolli et al., 2010; Rasmussen & Williams, 2006).

In the following sections a comparison of GPR results against similar studies is presented. A discussion on band preferences, encountered limitations and recommendations for further research closes this section.

6.1. Comparison with previous works

In a similar study using the SPARC dataset and S2 band settings, (Richter et al., 2009) retrieved LAI values by inverting PROSAIL (PROSPECT + SAIL) for the crops sugar beet and maize. While sugar beet showed accurate estimations of 8–9%, maize only reached an accuracy of 16–22%. The poorer estimation accuracies of maize were explained by the inability of the turbid SAIL model to simulate leaf clumping induced by the row structure of the crop. GPR also reached the latter accuracy, taken into account that field data spanned several crop types including row crops. When broadening to other datasets, literature revealed that local (hyperspectral) studies yielded RMSE accuracies between 0.41 and 1.4 for similar crops (Richter et al., 2009). In comparison, GPR here achieved RMSE precisions between 0.35 and 0.85.

FVC was somewhat more accurately predicted than LAI, though the expected GMES goal accuracy of ≤10% was not reached. FVC was validated consistently across all Sentinel configurations with relative errors of 16-18% (0.08 RMSE). Such accuracies are in line with earlier studies. For instance, French et al. (2010) obtained a RMSE accuracy of 0.13 in irrigated cotton fields. By feeding simulated training data into a generic NN, Bacour et al. (2006) achieved an accuracy of 0.09. Evaluation of their trained model on a MERIS image over southwest France, however, revealed that FVC values underestimated actual values. Interstingly, Verger et al. (2011) also reached an accuracy of 0.08 over Barrax with an NN approach, while Bannari et al. (2006) obtained a similar accuracy (0.09) by using a linear spectral mixture approach on hyperspectral data to map crop FVC. Bands in the SWIR were found especially important because of their sensitivity to soil and crop cover. It must be noted that compared to LAI considerably less FVC retrieval studies over crops were encountered. A reason for this is that canopy RT models used for inversion are typically turbid (e.g. SAIL) with LAI as the only driver of plant density. Alternatively, data-based techniques such as machine learning algorithms can relate spectral observations with any measured parameter (e.g. FVC).

GPR yielded best results in estimating *Chl*. The models that generated *Chl* maps were validated with excellent accuracies that ranged between 1.81 μg/cm^2 (S3-300 m) and 5.36 μg/cm^2 (S2-10 m). A vast amount of local *Chl* retrieval studies were found in literature, e.g. based on advanced RT modeling approaches, but hardly any with more successful accuracies. For instance, Jacquemoud et al. (2000) reported RMSE accuracies between 13.1 and 34.5 μg/cm^2 when four canopy RT models were inverted with airborne hyperspectral observations over maize and soybean with 2–5 free variables. For sugar beet, RMSE accuracies on the order of 10 μg/cm^2 were found when using hyperspectral and multispectral observations (Jacquemoud et al., 1995). More recently, Botha et al. (2007) reported overall *Chl* RMSE between 7.97 and 14.12 μg/cm^2 for potato cultivars by inverting PROSAIL with field measured
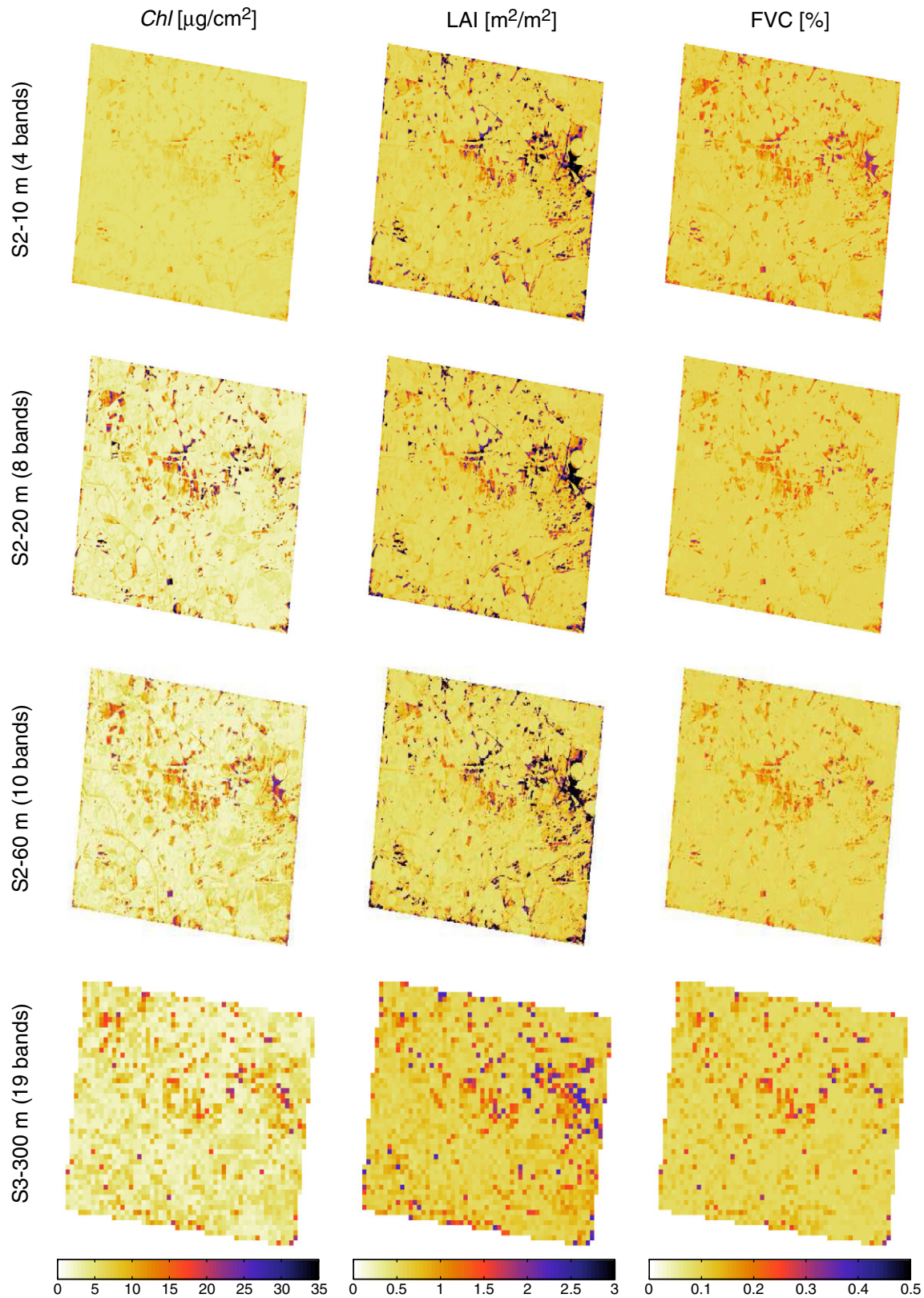
Fig. 5. GPR confidence maps for *Chl*, LAI and FCV, for the configurations of S2-10 m, S2-20 m, S2-60 m and S3-300 m. To facilitate visual comparison, the maps were color scaled to a maximum that is half of the maximum of their corresponding estimation maps.

hyperspectral data, and Kempeneers et al. (2008) reported overall RMSE of 4.7 $\mu g/cm^2$ by inverting PROFLIGHT (PROSPECT + FLIGHT) over orchard crops with airborne hyperspectral observations. Based on hyperspectral CHRIS observations and the same SPARC dataset covering various crop types as used in this study, Delegido et al. (2010) developed a *Chl*-sensitive spectral index, NAOC, which was linearly calibrated with the *Chl* data. An accuracy of 4.2 $\mu g/cm^2$ was reached. Of specific interest is the study in Houborg & Boegh (2008) that used four

SPOT bands at a resolution of 10 m, similar to the here presented S2-10 m case. By applying an extensive generic modeling inversion approach over agricultural crops, an accuracy of 5.0 $\mu g/cm^2$ was reached. This result is slightly better than the here achieved accuracy of 5.36 $\mu g/cm^2$ by S2-10 m, though the inversion approach was later adjusted by Houborg et al. (2009), reaching accuracies of 7.1 $\mu g/cm^2$.

In summary, GPR reached accuracies with S2 and S3 band settings comparable (LAI, FVC) or superior (*Chl*) to similar retrieval studies. These results are promising for operational applications, taken into account that a broad range of crop types with different plant geometries and growing stages were included in the training dataset.

6.2. GPR band analysis

Within the 2 seconds processing time GPR provided insight in relevant samples (not shown here) and relevant bands in establishing a relationship with the parameter of interest. Although each randomly chosen training dataset influenced the relevance of bands, by ranking the four best bands over 50 replications some interesting trends prevailed. For instance, the whole spectral range (440 to 940 nm) yielded bands in the top four in establishing relationships with the three biophysical parameters. This is encouraging for the planned S2 and S3 missions. However, some bands never reached the top four when having many bands available. The S2 B8 band (centered at 842 nm) was only a few times evaluated as 4th most important band in the 8-band S2-20 m configuration, and was no longer top four ranked when having 10 bands included. This band is inherited from SPOT and is characterized by a broad bandwidth (115 nm) that partially overlaps the B7 band (centered at 783 nm) and fully overlaps the B8a band (centered at 865 nm). Hence, it is likely that the spectral variation captured by the finer B7 and B8a bands makes the broadband redundant. For S3, although having 19 bands included, only the very first O1 band centered at 413 nm never reached the top four. Significant residual atmospheric effects were observed in this band and is therefore of little use for biophysical parameter retrieval. Optimal band sets were identified per parameter, which does not necessarily mean that bands should be removed for optimizing the retrieval process. This has been suggested sometimes, e.g. when using NNs (Verger et al., 2011), or when comparing simulated reflectance data against measured reflectance data with a cost function (Darvishzadeh et al., 2008a; Richter et al., 2009). In GPR, unlike in NNs, the relative relevance of the bands in developing a model are learned directly from the data so one could in principle use all of them (Rasmussen & Williams, 2006).

The observed varying bands preference throughout the different Sentinel scenarios contradicts somewhat with the established idea that each parameter is linked to specific spectral regions (le Maire et al., 2008; Thenkabail et al., 2004). Several authors concluded that maximum sensitivity to *Chl* appears in the region from 520 to 630 nm and also around 700 nm (Carter, 1994; Datt, 1998; Gitelson & Merzlyak, 1997; Gitelson et al., 1996, 2003; le Maire et al., 2008; Sims & Gamon, 2002). Consequently, multiple *Chl*-sensitive indices have been proposed using wavelengths in these regions, in combination with a NIR reference band, because *Chl* solely absorbs reflectance in the visible part of the spectrum and not in the NIR (Gamon & Surfus, 1999; Gitelson & Merzlyak, 1994; Zarco-Tejada et al., 2001). Also here GPR identified the bands centered at 620, 665, 705 and 740 nm as important players in *Chl* determination, but the band centered at 842 nm was evaluated as most important for the 4-band S2-10 m configuration (Fig. 3). With more bands included in following Sentinel configurations, this dominance diverged towards the beginning of the NIR shoulder (750–780 nm) and further into the NIR towards 940 nm. These spectral regions are closely associated with water absorption (Schepers et al., 1996; Thenkabail et al., 2000). Consequently, the *Chl* sensitivity in the NIR may be driven by *Chl*-water content covariance, as also observed by (Filella & Penuelas, 1994; Stagakis et al., 2010). At the same time, *Chl* is often correlated with

structural parameters such as LAI and FVC, which imposes additional spectral variations in the NIR (Darvishzadeh et al., 2008b; Filella & Penuelas, 1994; Verrelst et al., 2010). Because the SPARC dataset consist of irrigated fields with crops at various growing stages, it is very likely that such secondary (nonlinear) relationships were exploited by GPR, hence causing shifted band preferences.

LAI is probably one of the easiest parameters to detect because of affecting a wide spectral range (Thenkabail et al., 2004). Band combinations with green or red and bands in NIR are often found as the most successful ones (Fava et al., 2009; Hansen & Schjoerring, 2003), while (Stagakis et al., 2010) found a little higher performance when using red and green bands. Also GPR identified most relevant bands in the green, yellow and in the NIR for the S2 configurations. When having more bands available, as in S3, both bands in the visible, centered at 510, 620 nm, red-edge (709, 779 nm) and NIR (865 nm) were evaluated as most sensitive to LAI. The latter are assumed to be the effects of covariation between leaf biomass and canopy water content. This again highlights the important role that parameter covariance plays in adaptive fitting methods. The sensitive spectral regions towards FVC seems to be more straightforward, as it is directly related to the bands where largest variation occurs due to an increase in vegetation cover as opposed to a spectrally distinct background (e.g. bare soil).

Overall, the varying band preferences suggest that indirect reflectance responses to biophysical parameters play a significant role at the canopy level. Simple 2-band algorithms may therefore not always be the best choice for biophysical parameter estimation. Actually, Verrelst et al. (in press) concluded that running GPR through single bands outperformed the usage of vegetation indices because GPR can implement more flexible nonlinear fitting functions. Kernel regression approaches, in which no prior assumptions about band relevance are made, seem better suited to capture the available spectral variation related to the parameter of interest than conventional linear approaches.

6.3. Limitations and recommendations for future research

Despite the promising performances of GPR for S2 and S3 band settings, it does not escape to our knowledge the limitations of the used Sentinel data, which is simulated from CHRIS data. As a consequence of the narrower spectral range of CHRIS, the S2 MSI bands B10-B12 and S3 OLCI bands O1 and O21 were not included in the analysis. Sentinel bands are also configured with different radiometric accuracies than the CHRIS sensor. This is especially true for the S2 atmospheric bands at 60 m that are intended for atmospheric applications, such as aerosols correction (B1), water vapor correction and cirrus detection correction (B9 and B10) (Drusch et al., 2010). These bands are typically discarded in biophysical parameter retrieval approaches that rely on inversion of canopy RT models, because these models are unable to simulate atmospheric bands (Richter et al., 2009). Conversely, machine learning regression methods make full use of the available variation embedded in each single band for establishing a relationship without making assumptions. For instance, bands in the NIR and SWIR (B10-B12, O21) are known to be sensitive to soil and vegetation structure (Bannari et al., 2006) and can thus potentially improve the retrieval quality.

Although GPR was evaluated as the best performing algorithm, other promising nonparametric retrieval methods deserve to be tested as well. Lately *linear* multivariate feature extraction methods have been successfully introduced for biophysical parameter retrieval (Atzberger et al., 2010; Hansen & Schjoerring, 2003). Recently, *nonlinear* versions of these methods have been presented by (Arenas-García & Camps-Valls, 2008), which derive a lower number of highly informative features. Regardless of the performance outcome, however, we are not aware that any of these methods possess the interesting feature to directly deliver additional confidence maps. Moreover, from a

practical standpoint, confidence maps allow evaluating the robustness of estimations when having a trained model applied to other images in space and time.

Finally, in an attempt to make the training data more globally applicable, it would be desirable to add data from other land cover types. An urging research question in this respect is to evaluate how well GPR performs when fed by artificial spectra, e.g. as generated by a RT model. 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 approach (Verger et al., 2008, 2011). Future research will consider the evaluation of machine learning models when coupled with a RT model.

7. Summary and conclusions

ESA's forthcoming Sentinel-2 (S2) and Sentinel-3 (S3) satellites aim to improve the old generation of satellite sensors by providing superspectral imagery of high spatial and temporal resolution. S2 will be configured with band settings according to Landsat and SPOT, while S3 will include OLCI, which is a sensor similar to MERIS, in both cases with enhanced capabilities in terms of spatial and temporal resolution. Moreover, both Sentinels will be equipped with additional new bands for improved retrieval capacities and compensation of atmospheric effects. Along with these enhanced sensor configurations, there is also a need for improved parameter retrieval methods implementable into operational processing chains. In this respect, we have compared the utility of four state-of-the-art machine learning regression algorithms in estimating three key biophysical parameters *Chl*, LAI and FVC from satellite reflectance measurements in the range of 400–1000 nm. Field data came from the Barrax test site, covering a broad range of crops and growing stages. Hyperspectral CHRIS data was used for the simulation of the following S2 and S3 configurations: S2-10 m (4 bands), S2-20 m (8 bands), S2-60 m (10 bands), and S3-300 m (19 bands). GPR not only yielded better accuracies for the majority of cases, but also proved to be a fast and robust regressor. Despite that with S2 more bands can be included at lower spatial resolution, it was found that LAI and FVC can be accurately mapped with four bands at highest spatial resolution of 10 m. *Chl* benefited from extra bands in the red-edge for reaching best estimates. Specifically, *Chl* reached RMSE accuracies on the order of the by GMES requested precision threshold of 10%. In addition to high accuracies, GPR provide a confidence value (predictive variance) along with a mean predictive estimation, and insight in relevant bands thanks to the optimization of tailored kernel function. The latter revealed that parameter covariance contains an important source of information. GPR is concluded as a promising candidate for implementation into operational Sentinel processing chains.

Acknowledgments

References

Alonso, L., & Moreno, J. (2005). Advances and limitations in a parametric geometric correction of CHRIS/PROBA data. *Proceedings of the 3rd CHRIS/Proba Workshop*.

Arenas-García, J., & Camps-Valls, G. (Oct. 2008). Efficient kernel orthonormalized PLS for remote sensing applications. *IEEE Transactions on Geoscience and Remote Sensing*, *46*(10), 2872–2881.

Atzberger, C. (2004). Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. *Remote Sensing of Environment*, *93*(1–2), 53–67.

Atzberger, C., Guérif, M., Baret, F., & Werner, W. (2010). Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. *Computers and Electronics in Agriculture*, *73*(2), 165–173.

Bacour, C., Baret, F., Béal, D., Weiss, M., & Pavageau, K. (2006). Neural network estimation of LAI, fAPAR, fCover and LAI××Cab, from top of canopy MERIS reflectance data: Principles and validation. *Remote Sensing of Environment*, *105*(4), 313–325.

Bannari, A., Pacheco, A., Staenz, K., McNairn, H., & Omari, K. (2006). Estimating and mapping crop residues cover on agricultural lands using hyperspectral and IKONOS data. *Remote Sensing of Environment*, *104*(4), 447–459.

Baret, F., & Buis, S. (2008). Estimating canopy characteristics from remote sensing observations. Review of methods and associated problems. *Advances in Land Remote Sensing: System, Modeling, Inversion and Application*, 172–301.

Baret, F., & Guyot, G. (1991). Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment*, *35*(2–3), 161–173.

Barnsley, M. J., Settle, J. J., Cutter, M. A., Lobb, D. R., & Teston, F. (2004). The PROBA/-CHRIS mission: A low-cost smallsat for hyperspectral multiangle observations of the earth surface and atmosphere. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(7), 1512–1520.

Bicheron, P., & Leroy, M. (1999). A method of biophysical parameter retrieval at global scale by inversion of a vegetation reflectance model. *Remote Sensing of Environment*, *67*(3), 251–266.

Botha, E. J., Leblon, B., Zebarth, B., & Watmough, J. (2007). Non-destructive estimation of potato leaf chlorophyll from canopy hyperspectral reflectance using the inverted PROSAIL model. *International Journal of Applied Earth Observation and Geoinformation*, *9*(4), 360–374.

Buermann, W., Dong, J., Zeng, X., Myneni, R. B., & Dickinson, R. E. (2001). Evaluation of the utility of satellite-based vegetation leaf area index data for climate simulations. *Journal of Climate*, *14*(17), 3536–3550.

Camps-Valls, G., & Bruzzone, L. (Eds.). (Dec 2009). *Kernel methods for Remote Sensing Data Analysis*. UK: Wiley & Sons.

Camps-Valls, G., Gómez-Chova, L., Muñoz-Marí, J., & Calpe-Maravilla, J. (2009). Biophysical parameter estimation with adaptive Gaussian Processes. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS'2009. Capetown, South Africa*.

Carter, G. A. (1994). Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. *International Journal of Remote Sensing*, *15*(3).

Darvishzadeh, R., Skidmore, A., Schlerf, M., & Atzberger, C. (2008). Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sensing of Environment*, *112*(5), 2592–2604.

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., Corsi, F., & Cho, M. (2008). LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS Journal of Photogrammetry and Remote Sensing*, *63*(4), 409–426.

Datt, B. (1998). Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a+b, and total carotenoid content in eucalyptus leaves. *Remote Sensing of Environment*, *66*(2), 111–121.

Delegido, J., Alonso, L., González, G., & Moreno, J. (2010). Estimating chlorophyll content of crops from hyperspectral data using a normalized area over reflectance curve (NAOC). *International Journal of Applied Earth Observation and Geoinformation*, *12*(3), 165–174.

Drusch, M., Gascon, F., & Berger, M. (2010). Sentinel-2 mission requirements document. http://esamultimedia.esa.int/docs/GMES/GMES_Sentinel2_MRD_issue_2.0_update.pdf.

ESA (2004). SPARC data acquisition report. contract no: 18307/04/nl/ff, university valencia. http://earth.esa.int/campaigns/DOC/SPARC2004_Data_Acquisition_Report.pdf.

Fava, F., Colombo, R., Bocchi, S., Meroni, M., Sitzia, M., Fois, N., et al. (2009). Identification of hyperspectral vegetation indices for Mediterranean pasture characterization. *International Journal of Applied Earth Observation and Geoinformation*, *11*(4), 233–243.

Fernández, G., Moreno, J., Gandía, S., Martínez, B., Vuolo, F., & Morales, F. (2005). Statistical variability of field measurements of biophysical parameters in SPARC-2003 and SPARC-2004 campaigns. *Proceedings of the SPARC Workshop*.

Filella, I., & Penuelas, J. (1994). The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *International Journal of Remote Sensing*, *15*(7), 1459–1470.

French, A. N., Hunsaker, D. J., Clarke, T. R., Fitzgerald, G. J., & Pinter, P. J. (2010). Combining remotely sensed data and ground-based radiometers to estimate crop cover and surface temperatures at daily time steps. *Journal of Irrigation and Drainage Engineering*, *136*(4), 232–239.

Gamon, J. A., & Surfus, J. S. (1999). Assessing leaf pigment content and activity with a reflectometer. *The New Phytologist*, *143*(1), 105–117.

Gandía, S., Fernández, G., Garcia, J. C., & Moreno, J. (2004). Retrieval of vegetation biophysical variables from CHRIS/PROBA data in the SPARC campaign. *Proceedings of the 2nd CHRIS/PROBAWorkshop*.

Gitelson, A. A., Gritz, Y., & Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, *160*(3), 271–282.

Gitelson, A., & Merzlyak, M. N. (1994). Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. *Journal of Photochemistry and Photobiology. B, Biology*, *22*(3), 247–252.

Gitelson, A. A., & Merzlyak, M. N. (1997). Remote estimation of chlorophyll content in higher plant leaves. *International Journal of Remote Sensing*, *18*(12), 2691–2697.

Gitelson, A. A., Merzlyak, M. N., & Lichtenthaler, H. K. (1996). Detection of red edge position and chlorophyll content by reflectance measurements near 700 nm. *Journal of Plant Physiology*, *148*(3–4), 501–508.

Gitelson, A. A., Vina, A., Verma, S. B., Rundquist, D. C., Arkebauer, T. J., Keydan, G., et al. (2006). Relationship between gross primary production and chlorophyll content in crops: Implications for the synoptic monitoring of vegetation productivity. *Journal of Geophysical Research D: Atmospheres*, *111*(8).

Guanter, L., Alonso, L., & Moreno, J. (2005). A method for the surface reflectance retrieval from PROBA/CHRIS data over land: Application to ESA SPARC campaigns. *IEEE Transactions on Geoscience and Remote Sensing*, *43*(12), 2908–2917.

Hansen, P. M., & Schjoerring, J. K. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*, *86*(4), 542–553.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (2nd ed.). New York: Springer-Verlag.

Haykin, S. (Oct. 1999). *Neural Networks – A Comprehensive Foundation* (2nd ed.). : Prentice Hall.

Hilker, T., Coops, N. C., Wulder, M. A., Black, T. A., & Guy, R. D. (2008). The use of remote sensing in light use efficiency based models of gross primary production: A review of current status and future requirements. *The Science of the Total Environment*, *404*(2–3), 411–423.

Hirose, T., Ackerly, D. D., Traw, M. B., Ramseier, D., & Bazzaz, F. A. (1997). CO2 elevation, canopy photosynthesis, and optimal leaf area index. *Ecology*, *78*(8), 2339–2350.

Houborg, R., Anderson, M., & Daughtry, C. (2009). Utility of an image-based canopy reflectance modeling tool for remote estimation of lai and leaf chlorophyll content at the field scale. *Remote Sensing of Environment*, *113*(1), 259–274.

Houborg, R., & Boegh, E. (2008). Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and *SPOT* reflectance data. *Remote Sensing of Environment*, *112*(1), 186–202.

Jacquemoud, S., Bacour, C., Poilvé, H., & Frangi, J. -P. (2000). Comparison of four radiative transfer models to simulate plant canopies reflectance: Direct and inverse mode. *Remote Sensing of Environment*, *74*(3), 471–481.

Jacquemoud, S., Baret, F., Andrieu, B., Danson, F. M., & Jaggard, K. (1995). Extraction of vegetation biophysical parameters by inversion of the PROSPECT+SAIL models on sugar beet canopy reflectance data. Application to TM and AVIRIS sensors. *Remote Sensing of Environment*, *52*(3), 163–172.

Jiménez-Muñoz, J. C., Sobrino, J. A., Guanter, L., Moreno, J., Plaza, A., & Martínez, P. (2005). Fractional vegetation cover estimation from proba/CHRIS data: methods, analysis of angular effects and application to the land surface emissivity retrieval. *Proceedings of the 3rd CHRIS/Proba Workshop*.

Kakani, V. G., Reddy, K. R., Zhao, D., & Sailaja, K. (2003). Field crop responses to ultraviolet-B radiation: A review. *Agricultural and Forest Meteorology*, *120*(1–4), 191–218.

Kempeneers, P., Zarco-Tejada, P. J., North, P. R. J., de Backer, S., Delalieux, S., Sepulcre-Canto, G., et al. (2008). Model inversion for chlorophyll estimation in open canopies from hyperspectral imagery. *International Journal of Remote Sensing*, *29*(17–18), 5093–5111.

Kimes, D. S., Nelson, R. F., Manry, M. T., & Fung, A. K. (1998). Attributes of neural networks for extracting continuous vegetation variables from optical and radar measurements. *International Journal of Remote Sensing*, *19*(14), 2639–2662.

Knudby, A., LeDrew, E., & Brenning, A. (2010). Predictive mapping of reef fish species richness, diversity and biomass in zanzibar using IKONOS imagery and machine-learning techniques. *Remote Sensing of Environment*, *114*(6), 1230–1241.

le Maire, G., François, C., Soudani, K., Berveiller, D., Pontailler, J. Y., Bréda, N., et al. (2008). Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. *Remote Sensing of Environment*, *112*(10), 3846–3864.

Lichtenthaler, H. K. (1987). Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. *Methods in Enzymology*, *148*, 350–382.

Lichtenthaler, H. K., Lang, M., Sowinska, M., Heisel, F., & Miehe, J. A. (1996). Detection of vegetation stress via a new high resolution fluorescence imaging system. *Journal of Plant Physiology*, *148*(5), 599–612.

Martimort, P., Berger, M., Carnicero, B., Bello, U. D., Fernandez, V., Gascon, F., et al. (2007). The optical high-resolution mission for GMES operational services. *European Space Agency Bulletin*, 18–23.

Nieke, J., Frerick, J., Stroede, J., Mavrocordatos, C., & Berruti, B. (2009). *Status of the optical payload and processor development of ESA's sentinel 3 mission, vol. 7452*, .

Pasolli, L., Melgani, F., & Blanzieri, E. (2010). Gaussian process regression for estimating chlorophyll concentration in subsurface waters from remote sensing data. *IEEE Geoscience and Remote Sensing Society*, *99*, 464–468.

Pozdnyakov, D., Shuchman, R., Korosov, A., & Hatt, C. (2005). Operational algorithm for the retrieval of water quality in the Great Lakes. *Remote Sensing of Environment*, *97*(3), 352–370.

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning.* New York: The MIT Press.

Richter, K., Atzberger, C., Vuolo, F., Weihs, P., & D'Urso, G. (2009). Experimental assessment of the sentinel-2 band setting for RTM-based LAI retrieval of sugar beet and maize. *Canadian Journal of Remote Sensing*, *35*(3), 230–247.

Sampson, P. H., Zarco-Tejada, P. J., Mohammed, G. H., Miller, J. R., & Noland, T. L. (2003). Hyperspectral remote sensing of forest condition: Estimating chlorophyll content in tolerant hardwoods. *Forest Science*, *49*(3), 381–391.

Schepers, J. S., Blackmer, T. M., Wilhelm, W. W., & Resende, M. (1996). Transmittance and reflectance measurements of corn leaves from plants with different nitrogen and water supply. *Journal of Plant Physiology*, *148*(5), 523–529.

Schiller, H., & Doerffer, R. (2005). Improved determination of coastal water constituent concentrations from MERIS data. *IEEE Transactions on Geoscience and Remote Sensing*, *43*(7), 1585–1591.

Schölkopf, B., & Smola, A. (2002). *Learning with Kernels – Support Vector Machines, Regularization, Optimization and Beyond.* : MIT Press Series.

Shawe-Taylor, J., & Cristianini, N. (2004). *Kernel Methods for Pattern Analysis.* : Cambridge University Press.

Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, *81*(2–3), 337–354.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*, 199–222.

Stagakis, S., Markos, N., Sykioti, O., & Kyparissis, A. (2010). Monitoring canopy biophysical and biochemical parameters in ecosystem scale using satellite hyperspectral imagery: An application on a Phlomis fruticosa Mediterranean ecosystem using multiangular CHRIS/PROBA observations. *Remote Sensing of Environment*, *114*(5), 977–994.

Thenkabail, P. S., Enclona, E. A., Ashton, M. S., & Meer, B. V. D. (2004). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment*, *91*(3–4), 354–376.

Thenkabail, P. S., Smith, R. B., & Pauw, E. D. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment*, *71*(2), 158–182.

Verger, A., Baret, F., & Camacho, F. (2011). Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: Evaluation over an agricultural area with CHRIS/PROBA observations. *Remote Sensing of Environment*, *115*(2), 415–426.

Verger, A., Baret, F., & Weiss, M. (2008). Performances of neural networks for deriving LAI estimates from existing CYCLOPES and MODIS products. *Remote Sensing of Environment*, *112*(6), 2789–2803.

Verrelst, J., Schaepman, M. E., Malenovský, Z., & Clevers, J. G. P. W. (2010). Effects of woody elements on simulated canopy reflectance: Implications for forest chlorophyll content retrieval. *Remote Sensing of Environment*, *114*(3), 647–656.

Verrelst, J., Alonso, L., Camps-Valls, G., Delegido, J., Moreno, J., in press. Retrieval of vegetation biophysical parameters using gausian processes techniques. *IEEE Transactions on Geoscience and Remote Sensing*.

Whittaker, R. H., & Marks, P. L. (1975). Methods of assessing terrestrial productivity. *Primary Productivity of the Biosphere* (pp. 55–118).

Zarco-Tejada, P. J., Miller, J. R., Noland, T. L., Mohammed, G. H., & Sampson, P. H. (2001). Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(7), 1491–1507.