ISPRS Journal of Photogrammetry and Remote Sensing 86 (2013) 157–167

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Gaussian processes uncertainty estimates in experimental Sentinel-2 LAI and leaf chlorophyll content retrieval

Jochem Verrelst *, Juan Pablo Rivera, José Moreno, Gustavo Camps-Valls

Image Processing Laboratory (IPL), Parc Científic, Universitat de València, 46980 Paterna, València, Spain

ARTICLE INFO

Article history:
Received 15 March 2013
Received in revised form 24 September 2013
Accepted 27 September 2013

Keywords:
Uncertainty estimates
Gaussian processes regression
Biophysical parameters
Sentinel-2
Chlorophyll content
Leaf area index

ABSTRACT

ESA's upcoming Sentinel-2 (S2) Multispectral Instrument (MSI) foresees to provide continuity to land monitoring services by relying on optical payload with visible, near infrared and shortwave infrared sensors with high spectral, spatial and temporal resolution. This unprecedented data availability leads to an urgent need for developing robust and accurate retrieval methods, which ideally should provide uncertainty intervals for the predictions. Statistical learning regression algorithms are powerful candidats for the estimation of biophysical parameters from satellite reflectance measurements because of their ability to perform adaptive, nonlinear data fitting. In this paper, we focus on a new emerging technique in the field of Bayesian nonparametric modeling. We exploit Gaussian process regression (GPR) for retrieval, which is an accurate method that also provides uncertainty intervals along with the mean estimates. This distinct feature is not shared by other machine learning approaches. In view of implementing the regressor into operational monitoring applications, here the portability of locally trained GPR models was evaluated. Experimental data came from the ESA-led field campaign SPARC (Barrax, Spain). For various simulated S2 configurations (S2-10m, S2-20m and S2-60m) two important biophysical parameters were estimated: leaf chlorophyll content (LCC) and leaf area index (LAI). Local evaluation of an extended training dataset with more variation over bare soil sites led to improved LCC and LAI mapping with reduced uncertainties. GPR reached the 10% precision required by end users, with for LCC a NRMSE of 3.5–9.2% (r^2: 0.95–0.99) and for LAI a NRMSE of 6.5–7.3% (r^2: 0.95–0.96). The developed GPR models were subsequently applied to simulated Sentinel images over various sites. The associated uncertainty maps proved to be a good indicator for evaluating the robustness of the retrieval performance. The generally low uncertainty intervals over vegetated surfaces suggest that the locally trained GPR models are portable to other sites and conditions.

1. Introduction

Biophysical parameter products such as leaf chlorophyll content (LCC) and leaf area index (LAI) have become standard products by space agencies and research institutions. At the same time, proper usage of such products requires that associated uncertainty information needs to be provided (Buermann et al., 2001; Morisette et al., 2006; Fang et al., 2012). Two categories of product uncertainty information are generally available in the literature, either physical or theoretical (Fang et al., 2012). Physical uncertainties indicate the departure of product values from hypothetical true values and are obtained through the collection of ground-based validation data. Theoretical uncertainties are caused by uncertainties in the input data and model imperfections and are usually estimated during the retrieval process (Baret et al., 2007; Knyazikhin et al., 1998; Pinty et al., 2011). Several operationally delivered products are nowadays accompanied with theoretical uncertainties in the form of a quantitative quality indicator (Knyazikhin et al., 1998; Pinty et al., 2011). An alternative method was recently proposed in (Fang et al., 2012) where authors apply an independent uncertainty calculation through a triple collocation method. As such, independently the uncertainties of MODIS, CYCLOPES, and GLOBCARBON LAI products were evaluated. From these products, it was concluded that particularly CYCLOPES generally meet the quality requirements (±0.5) proposed by the Global Climate Observing System (GCOS) (GCOS, 2011). The CYCLOPES algorithm is based on a neural network (NN) trained from the 1D SAIL radiative transfer model (RTM) (Baret et al., 2007).

While NNs have proven robust in various operational processing chains, the retrieval schemes are far from being perfect and various limitations have been identified. For instance, LAI estimates from CYCLOPES are less accurate at higher values due to the saturation effect in the radiative transfer simulation and the

* Corresponding author. Tel.: +34 963544067.
E-mail address: jochem.verrelst@uv.es (J. Verrelst).

0924-2716/$ - see front matter © 2013 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS) Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.isprsjprs.2013.09.012

NN inversion algorithm (Bacour et al., 2006; Weiss et al., 2007). Furthermore, NNs not only behave as a black box model, but they are also relatively unpredictable when used with input spectra that deviate (even slightly) from what has been presented during the training stage (Atzberger, 2004; Baret and Buis, 2008). It is therefore 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. For these reasons, alternatives that overcome these limitations are needed. In part, this is why in the recent years NNs are being replaced by other more advanced, simpler to train, machine learning regression algorithms (MLRAs). Actually, during the last two decades, the family of *kernel methods* (Camps-Valls and Bruzzone, 2009) has emerged as an alternative to NNs in many applications. Kernel methods typically involve few and intuitive hyperparameters to be tuned, and can perform flexible input–output nonlinear mappings. They are able to cope with the strong nonlinearity of the functional dependence between the biophysical parameter and the observed reflected radiance. Intimately related to the field of kernel methods, here we find the new emerging field of Bayesian nonparametric modeling. The framework of Bayesian nonparametrics gives a Bayesian treatment of statistical inference. This field has given rise to particularly powerful methods, such as relevance vector machines and Gaussian processes. These methods are able to provide high accuracies and at the same time provide uncertainty intervals for the predictions (e.g. Camps-Valls et al., 2006; Verrelst et al., 2012b). They may therefore be more suitable candidates for operational applications, especially now that Earth observation is reaching a mature state.

In 2014, the European Space Agency's (ESA) forthcoming Sentinel-2 (S2) mission will start delivering high-resolution optical images on a global scale. This unprecedented data availability leads to an urgent need for developing robust and accurate retrieval methods. Recently, Verrelst et al. (2012a) have tested the capabilities of four state-of-the-art MLRAs given different Sentinel-2 and Sentinel-3 band settings. Selected MLRAs were NNs, support vector regression, kernel ridge regression, and Gaussian processes regression (GPR). The methods were compared in terms of accuracy, goodness of fit, robustness to low sample sized scenarios, and computational cost. Training and validation data came from the ESA-led field campaign SPARC, which took place on the agricultural test site Barrax, Spain. The main conclusion was that, in general, GPR outperformed the other regressors in terms of speed and computational costs. At the same time, GPR yielded superior accuracies for the majority of tested cases. Moreover, in contrary to NNs, GPR provide directly theoretical uncertainty estimates through Gaussian probability (cf. Section 3). These uncertainty estimates opened a new source of information. For instance, they make possible to assess the robustness of the retrievals at various spatial scales. In Verrelst et al. (2013), retrievals from hyperspectral airborne and spaceborne data over the Barrax area were compared. In this way, the uncertainty measure provided information about the upscaling quality, i.e., if the uncertainties are kept constant then the upscaling can be considered stable. Even though retrievals proved to be robust over vegetated areas, high uncertainties appeared over non-vegetated surfaces, which suggested that the training dataset was not representative enough for those land cover types. Furthermore, since statistical approaches are often criticized because of limited generalization and transferability (e.g. Colombo et al., 2003; Meroni et al., 2004), it remains to be questioned how robust the locally-trained GPR models function when applied to other sites and conditions. In this respect, the delivery of additional uncertainty estimates may enable to evaluate the portability of the regression model. Specifically, when uncertainty intervals as produced by a locally trained GPR model over an arbitrary site are on the same order as those produced over the successfully validated reference site, then it can be reasonably assumed that the parameter retrievals are also of the same quality as the retrievals of the reference site. Thus, when successfully validated over a reference imagery then the uncertainty estimates can work as a quality indicator. This concept has been evaluated here, and brings us to the following specific objectives of the present paper: (1) to evaluate the impact of experimental training data on the development of GPR models, particularly in view of improved retrievals over non-vegetated surfaces, and (2) to evaluate the portability of a locally trained GPR model to other sites and conditions by making use of its associated uncertainty intervals.

The remainder of the paper is organized as follows. Section 2 briefly describes the Sentinel-2 concept, while Section 3 revises the Bayesian nonparametric field in general and the Gaussian process regression algorithm in particular. In the methodology (Section 4) the used training dataset, experimental Sentinel-2 images, and experimental setup are described. Results focus first on a local evaluation of the GPR models (Section 5.2) and then moves to the evaluation of algorithm's performance on other sites (Section 5.3). Section 6 discusses main findings and Section 7 concludes the paper.

2. Sentinel-2

ESA's S2 satellites capitalize on the technology and the vast experience acquired with SPOT and Landsat over the past decades (Drusch et al., 2012). 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 2014. 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 (Table 1). S2 incorporates three new bands in the red-edge region, which are centered at 705, 740 and 783 nm. 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. To serve the objectives of Copernicus (The European Earth Observation Programme), S2 satellites will provide data 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 LCC, LAI and leaf water content maps. To ensure that the final products can meet user requirements, the Copernicus user committee defined an accuracy goal of 10% (Drusch et al., 2012).

3. Bayesian nonparametrics and Gaussian processes

Finding a functional relation between input (e.g. reflectances) and output (e.g. physical parameter) variables is the main goal of statistical learning. The problem is complex and elusive because possibly an infinite number of functions can be found to fit the data. This problem is known as the excess of capacity of the class of functions implementing the model, and has been referred to the problem of overfitting. The key is to constrain model's capacity, in either one of the following two ways: imposing strong prior knowledge or via regularization schemes that promote simpler solutions. In the last decades, statistical learning (inference) has witnessed an overwhelming interest in kernel methods because they implement nonlinear models and still rely on linear algebra operations (Camps-Valls and Bruzzone, 2009). Kernel methods are very appealing for physical parameter retrieval, mainly because they deal efficiently with low-sized datasets of potentially high dimensionality, which are the situations we find in parameter

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

retrieval from multispectral or hyperspectral imagery using models trained by field campaign data. Also, since kernel methods do not assume an explicit prior data distribution but are inherently nonparametric models, they cope well with remote sensing data specificities and complexities (Camps-Valls and Bruzzone, 2009).

In the context of statistical inference and for remote sensing products, one is not solely interested in high accuracies of the algorithm but also in producing uncertainty intervals for the predictions. This calls for a Bayesian treatment of the inference problem. While recently kernel methods have advanced the field of remote sensing data analysis (Camps-Valls and Bruzzone, 2009), the problem of uncertainty estimation with nonparametric models has been elusive in most of the approaches. In this context, the emerging field of *nonparametric Bayesian modeling* constitutes a proper theoretical framework to tackle the problem of physical parameter retrieval (O'Hagan, 1994; Rasmussen and Williams, 2006; Orbanz and Teh, 2010).[1] Essentially, a nonparametric Bayesian model is a Bayesian model on an infinite-dimensional parameter space, which corresponds to the set of possible patterns, e.g. the class of smooth functions for regression (retrieval). The field has been very active in the last decade, and has delivered successful model instantiations. Some of them has been actually introduced in the field of remote sensing data analysis. For example, the relevance vector machine (RVM) introduced in Tipping (2001) is a nonparametric Bayesian model that assumes a Gaussian prior over the weights in order to enforce sparsity, and uses expectation-maximization to infer the parameters. In Camps-Valls et al. (2006) the RVM model was used for oceanic chlorophyll content estimation. The model, however, may incur in too sparse solutions that do not fit well in parts of the space not considered in the retained relevant vectors. Lately, Bayesian nonparametric modeling with Gaussian Processes (Rasmussen and Williams, 2006) have received much attention in the field of machine learning, and has been also introduced in the remote sensing application field (Verrelst et al., 2012a,b, 2013). This paper will focus on GPR.

GPR provides a probabilistic (Bayesian) approach for learning generic regression problems with kernels (Rasmussen and Williams, 2006). The GPR model establishes a relation between the input (B-bands spectra) $\mathbf{x} \in \mathbb{R}^B$ and the output variable (canopy parameter) $y \in \mathbb{R}$ of the form:

$$\hat{y} = f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i K(\mathbf{x}_i, \mathbf{x}), \tag{1}$$

where $\{\mathbf{x}_i\}_{i=1}^N$ are the spectra used in the training phase, $\alpha_i \in \mathbb{R}$ is the weight assigned to each one of them, and we intentionally dropped the bias term of the regression assuming centered observations, and K is a function evaluating the similarity between the test spectrum $\mathbf{x}$ and all N training spectra, $\mathbf{x}_i$, $i = 1, \ldots, N$. We used a scaled Gaussian kernel function,

$$K(\mathbf{x}_i, \mathbf{x}_j) = v \exp\left(-\sum_{b=1}^{B} \frac{(\mathbf{x}_i^{(b)} - \mathbf{x}_j^{(b)})^2}{2\sigma_b^2} \right) + \delta_{ij} \cdot \sigma_n^2, \tag{2}$$

where v is a scaling factor, B is the number of bands, σ_b is a dedicated parameter controlling the spread of the relations for each particular spectral band b, σ_n is the noise standard deviation and δ_{ij} is the Kronecker's symbol. The kernel is thus parametrized by signal (v, σ_b) and noise (σ_n) hyperparameters, collectively denoted as $\boldsymbol{\theta} = \{v, \sigma_b, \sigma_n\}$.

For training purposes, we assume that the observed variable is formed by noisy observations of the true underlying function $y = f(\mathbf{x}) + \epsilon$. Moreover we assume the noise to be additive independently identically Gaussian distributed with zero mean and variance σ_n. Let us define the stacked output values $\mathbf{y} = (y_1, \ldots, y_n)^\top$, the covariance terms of the test point $\mathbf{k}_* = (k(\mathbf{x}_*, \mathbf{x}_1), \ldots, k(\mathbf{x}_*, \mathbf{x}_n))^\top$, and $k_{**} = k(\mathbf{x}_*, \mathbf{x}_*)$. From the previous model assumption, the output values are distributed according to:

$$\begin{pmatrix} \mathbf{y} \\ f(x_*) \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \mathbf{K} + \sigma_n^2 \mathbf{I} & \mathbf{k}_* \\ \mathbf{k}_*^\top & k_{**} \end{pmatrix} \right) \tag{3}$$

For prediction purposes, the GPR is obtained by computing the posterior distribution over the unknown output $\mathbf{y}_*$, $p(\mathbf{y}_*|\mathbf{x}_*, \mathcal{D})$, where $\mathcal{D} \equiv \{\mathbf{x}_n, y_n | n = 1, \ldots, N\}$ is the training dataset. Interestingly, this posterior can be shown to be a Gaussian distribution, $p(y_*|\mathbf{x}_*, \mathcal{D}) = \mathcal{N}(y_*|\mu_{\mathrm{GP}*}, \sigma^2_{\mathrm{GP}*})$, for which one can estimate the *predictive mean* (point-wise predictions):

$$\mu_{\mathrm{GP}*} = \mathbf{k}_*^\top (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y}, \tag{4}$$

and the *predictive variance* (confidence intervals):

$$\sigma^2_{\mathrm{GP}*} = k_{**} - \mathbf{k}_*^\top (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{k}_*. \tag{5}$$

The corresponding hyperparameters $\boldsymbol{\theta}$ are typically selected by Type-II Maximum Likelihood, using the marginal likelihood (also called *evidence*) of the observations, which is also analytical. When the derivatives of the log-evidence are also analytical, which is often the case, conjugated gradient ascent is typically used for optimization (see (Rasmussen and Williams, 2006; Camps-Valls et al., 2009) for further details).

Three important properties of the method are worth stressing here. First, the obtained weights α_i after optimization gives the relevance of each spectrum $\mathbf{x}_i$. The predictive mean is essentially a weighted average of the canopy parameter values associated to the training samples closest to the test sample. 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. These features have been extensively studied in (Verrelst et al., 2012a,b) and proved to be valuable for gaining insight in relevant bands. Finally, a GPR model provides not only a pixelwise prediction for each spectrum but also an uncertainty (or confidence) level for the prediction. Hence in contrary to other approaches (e.g. NN) uncertainty intervals are directly delivered along with mean estimates. The interested reader is referred to the book by Rasmussen and 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/.

We illustrate the uncertainty intervals estimated by GPR and through standard boostrapping (Wu, 1986) of the nonlinear regression solution in Fig. 1. Note that GPR basically accounts for a reduction of uncertainty based on the relative local density of the input data points, not the outputs. This is obvious by looking at the

[1] Excellent online lectures on Bayesian nonparametrics are available at: http://videolectures.net/mlss09uk_teh_nbm/ and http://videolectures.net/mlss09uk_orbanz_fnbm/.

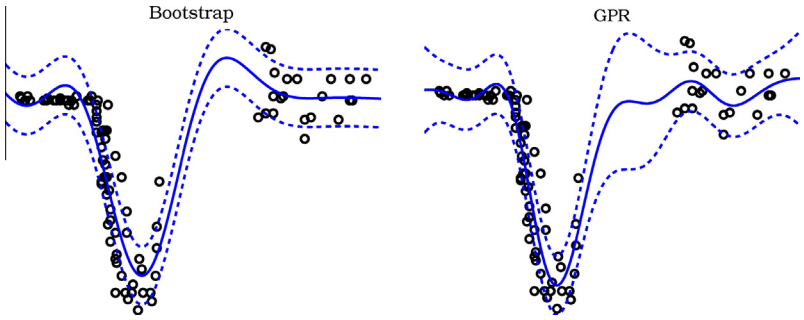

Fig. 1. Toy regression example in 1-D: we predict the *y*-values from the *x*-values in a synthetic signal generated by a combination of Gaussian bumps. We show the predictive variance or confidence interval estimates via bootstrap variance estimates (left), and Gaussian process regression (right).

predictive variance equation above. On the other hand, the variance of the bootstrap variance estimate quickly vanishes if one moves away from data points in the set. This behavior correctly reflects the fact that the predictions will be practically zero far away from points in the training sets. Gaussian processes indicate that the uncertainty is high because no data has been observed in that area.

4. Methodology

4.1. SPARC database

GPR has been trained with a local ground dataset coming from SPARC (SPectra bARrax Campaign). The SPARC-2003 and SPARC-2004 campaigns took place in the Barrax agricultural site in Central Spain (coordinates 39°N, −2°1′E, 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, mainly by center pivot irrigation systems. It leads to a patchy landscape with large circular fields. 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 LCC was derived by measuring about 50 samples with a calibrated CCM-200 Chlorophyll Content Meter. 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). 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 and LCC between 2 and 55 μg/cm^2. Further details on the measurements can be found in the data acquisition report (Moreno and Participants of the SPARC campaigns, 2004). Additionally, 30 random bare soil spectra with a biophysical (LCC, LAI) value of zero were added to broaden the dataset to non-vegetated samples. This 'original' dataset used in Verrelst et al. (2012a,b, 2013) in training GPR models and hereafter refers to 'training original' or '*TrOr*'.

In the latter studies, the GPR uncertainty maps showed that mean estimates over vegetated areas were associated to low uncertainties. At the same time, considerably higher uncertainties were encountered over areas of fallow land and bare soils. These higher uncertainties can be attributed to the relatively poor contribution of non-vegetated land cover types in the training dataset. For this reason, we have extended the SPARC training dataset with 60 new spectra that cover all kinds of non-vegetated surfaces, i.e. spectra with an LCC and LAI of zero. Most of the spectra were taken over bare soil surfaces, but also man-made surfaces (e.g., build-up areas, roads) and water bodies have been included. This 'extended' training dataset hereafter refers to 'training extended' or '*TrEx*'.

4.2. Reference and other simulated Sentinel-2 images

Because actual S2 data is not available yet, we opted for simulating it 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. 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 corrected for atmospheric effects according to the method proposed in Guanter et al. (2005). Since most of ground truth data were collected during the 2003 campaign, the nadir image from 14 July 2003 was used as reference image for spectral and spatial resampling to the settings of S2. The image is shown in Fig. 2. The majority of ESUs are located on the circular green fields, while non-vegetated samples came from the yellowish-white[2] surfaces. Because configured with different pixel sizes (10, 20 and 60 m), it is of special interest to simulate S2 bands as a function of pixel's size. A nearest neighbor strategy was used for the spatial resampling and a Gaussian model with full-width-half-maximum spacings was used for spectral resampling. Constrained by the spectral range of CHRIS, experimental data according to the following three Sentinel settings were generated, 'S2-10m': four bands at 10 m, 'S2-20m': eight bands at 20 m (4 bands at 20 m plus the S2-10m bands coarse-grained at 20 m), and 'S2-60m': ten bands at 60 m (2 bands at 60 m plus the earlier bands coarse-grained at 60 m).

Since the objective of this work was to evaluate the ability of transferring Barrax-trained GPR models for the various Sentinel settings to other images, multiple Mode 1 CHRIS images over terrestrial surfaces were arbitrarily collected. The only requirement was that they are predominantly cloud-free and acquired during spring or summertime. The CHRIS images include: multi-temporal images over Barrax (Spain); July 2004, and June 2009. Multitemporal images over Demmin (Germany), May and July 2006, and an image over Los Monegros (Spain), August 2006, Las Tablas (Spain), July 2006 and Sudbury (Canada), August 2007. The sites are

[2] For interpretation of color in Fig. 2, the reader is referred to the web version of this article.

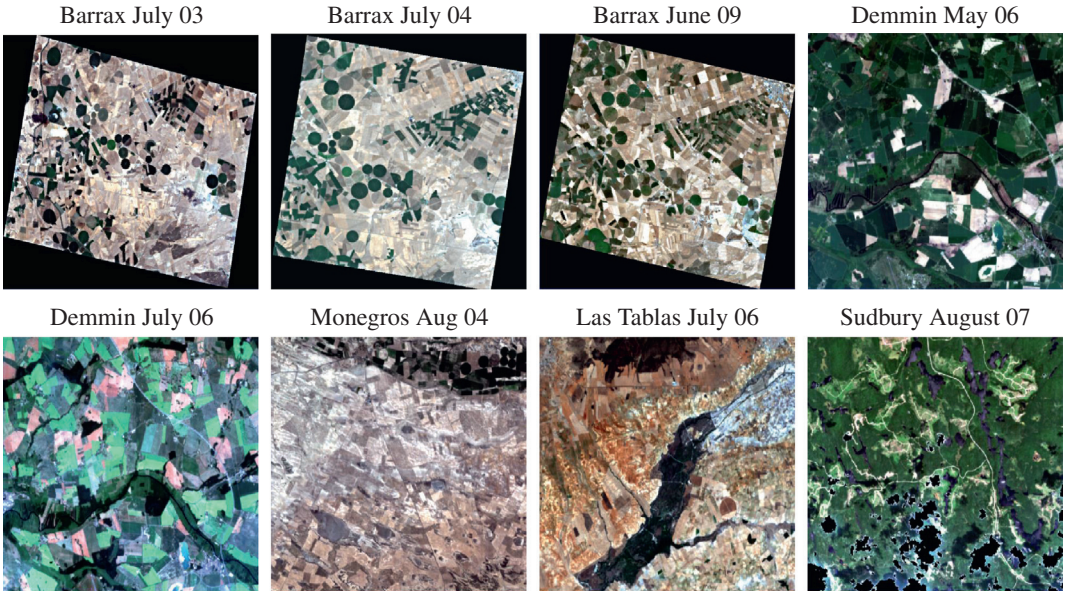

Fig. 2. RGB compositions of CHRIS images used for evaluating the performance of the locally-trained GPR models. Barrax July 03 is the reference image.

Table 2
Description of test sites.

Site	Acquisition time	Description
Barrax, Spain (39°N, −2°E)	14 July 03	The Barrax agricultural 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, mainly by center pivot irrigation systems. It leads to a patchy landscape with large circular fields. The annual rainfall average is about 400 mm
Barrax, Spain (39°N, −2°E)	16 July 04	
Barrax, Spain (39°N, −2°E)	19 June 09	
Demmin, Germany (53.5°N, −13.1°E)	08 May 06	The Demmin agricultural area is located in Northeast Germany and is based on a group of farms covering app. 25,000 ha. The surface is flat at 50 m a.s.l. with some slopes along the Tollense River. The field sizes are large in this area for German standards, about 80–250 ha. The main crops grown are wheat, barley, rape, maize and sugar. The annual rainfall ranges from 500 to 650 nm
Demmin, Germany (53.5°N, −13.1°E)	07 July 06	
Monegros, Spain (41.2°N, −0.1°1′E)	20 August 05	Los Monegros area is a semi-arid region, sparse vegetation and shallow as well as poorly developed soils. Agricultural activities are poor and may trigger land degradation processes. The area is characterized by various small playa lakes. These lakes form in small karstic depression by the dissolution of evaporitic subsurface layers, mainly gypsum and limestone. The playa lakes are usually dry in summer. The annual rainfall average is low, about 350 mm
Las Tablas, Spain (39.1°N, −3.4°E)	03 July 06	Las Tablas de Daimiel is a Complex of shallow pools and associated marshland, which lies in the great plain of La Mancha. The surface is flat at 600–620 m a.s.l. The site receives floodwater from the permanent freshwater Río Guadiana and the seasonal brackish Río Gigüela, and groundwater from an underground water basin. The surrounding area is used for rain-fed agriculture, mainly olives and wine yards. The annual average rainfall is about 450 mm
Sudbury, Canada (47.1°N, −81.4°E)	10 August 07	Sudbury is a boreal forested region located in the northern part of Ontario. It is a flat area 350 m a.s.l. The area is predominantly covered by black spruce and aspen stands with an understory of shrubs and herbs. The annual rainfall average is about 820 mm

described in Table 2. All these CHRIS images were corrected for atmospheric effects according to Guanter et al. (2005). A cloud masking over the Sudbury image was applied to remove the pixels with cloud contamination. Although not really necessary for this exercise, the Barrax images were also geometrically corrected according to Alonso and Moreno (2005). The images were subsequently resampled according to above-described Sentinel settings (S2-10m, S2-20m and S2-60m). The images are displayed in Fig. 2.

4.3. Experimental setup

The local *TrOr* and *TrEx* experimental datasets were divided into two subsets: 80% for training and the remaining 20% for validation. The subsequent undertaken approach was straightforward; for each biophysical parameter (LCC, LAI) and each S2 configuration (S2-10m, S2-20m, S2-60m) a *TrOr* and *TrEx* model were trained. The predictive power of the developed models was subsequently validated against the 20% validation data by using the absolute root-mean-squared error (RMSE) and the normalized RMSE (NRMSE [%] = RMSE/range of the parameters as measured in the field ∗100) to assess accuracy, and the coefficient of determination (r^2) to account for the goodness-of-fit. The NRMSE was used to compare the performances across the different methods and parameters. Once successfully validated, the *TrOr*- and *TrEx*-developed models were applied to the other experimental S2 images and the mean estimates and associated uncertainties were compared. To preserve a physical meaning, negative LCC or LAI estimates were automatically converted to near-zero values (0.0001) during the retrieval process.

5. Results

5.1. GPR performance with original and extended training dataset

The evaluation of the *TrOr*- and *TrEx*-developed GPR models starts with inspection of the validation results (Table 3). Excellent

prediction accuracies were obtained for all scenarios with a r^2 between 0.92 and 0.99 and NRMSE between 3.5% and 10.5%. Noteworthy hereby is that excellent LAI accuracies were already achieved with only 4 bands (S2-10m). This is encouraging for LAI mapping at high spatial resolution. The inclusion of more bands only improved accuracies marginally. Conversely, LCC clearly gained from the inclusion of extra bands, with a S2-20m configuration (8 bands: B2-B8a) that managed to reach NRMSE down to 3.5% (r^2: 0.99). It is well known that the inclusion of red-edge bands (B5 and B6) can considerably improve biophysical parameter estimation (e.g. Verrelst et al., 2012a; Delegido et al., 2011, 2013). However, the addition of 2 more bands in the blue (B1) and NIR (B9) at a coarser resolution of 60 m (S2-60m) slightly degraded accuracies.

Of more relevance here is the comparison of validation results as achieved by using models developed using *TrOr* and *TrEx*. It can be noticed that *TrEx* yielded slightly improved results. This is especially the case for LAI, where for all S2 scenarios NRMSE dropped with about 2%. Hence, the GPR model is able to incorporate more samples without losing accuracy. In fact the contrary occurred; for all S2 configurations, the NRMSE kept below the threshold of 10%.

Even though successfully validated, it is well understood that a validation dataset is rather limited as compared to the total variability observed in satellite images. Therefore, it is expected that inspection of the mean estimate and associated uncertainty maps on complete scenes will allow us to better understand the performances of both GPR models.

5.2. Local mapping over Barrax, Spain

Although LCC and LAI maps were generated for all S2 settings from the 2003 Barrax reference site, for the sake of brevity we only display generated maps at a high spatial resolution of 10 m. In turn the maps will allows us to appreciate the strength of GPR. Biophysical parameter retrieval was thus achieved with 4 bands only: 490, 560, 665 and 842 nm. The retrieval process was completely automated and image-based; the generation of a map was completed almost instantaneously.

Fig. 3[top] provides mean estimate (μ) maps and associated uncertainties ($\pm\sigma$) over the Barrax test site as generated by the original training dataset (*TrOr*). The mean estimate maps are first briefly interpreted. Within-field variations are clearly detected in both maps. Particularly, the pronounced spatial variation of LCC marks prominently the irrigated circular fields with green biomass. These irrigated fields are characterized by an LCC above 40, and an LAI above 3. Areas with low LCC and LAI (the whitish parts) are mainly bare soils, fallow lands or rain-fed senescent or harvested cereal fields (wheat, barley). The same maps are provided below, but then generated by the extended training dataset (*TrEx*). When comparing *TrOr*- and *TrEx*-generated LCC maps it can be observed that they are very similar. That similarity is also reflected when correlating both maps in a scatter plot, as displayed below. The large majority of the pixels fall right on the 1:1 line. In turn, more differences between *TrOr*- and *TrEx*-generated LAI maps occurred over non-green vegetated areas (e.g. fallow land and bare soils). The addition of bare soil spectra in the training dataset led thus to more meaningful LAI retrievals for a considerably amount of pixels, as was also observed by the validation dataset in Table 3. The scatter plot shows that the *TrEx*-developed LAI model causes that a substantial part of pixels are interpreted towards lower LAI estimates. While this may imply an improved accuracy, the associated uncertainty intervals will manifest the quality of the estimates.

Within these uncertainty maps, areas with reliable retrievals can be clearly distinguished from areas with unreliable retrievals. These differences are more obviously observed in the *TrOr*-generated maps. Reliable retrievals (low $\pm\sigma$) were found on irrigated areas and harvested fields. This is not surprising since the majority of training and validation samples came from these fields. Unreliable retrievals (high $\pm\sigma$) were found on areas with remarkably different spectra, such as bright, whitish calcareous soils (center, right), or harvested rain-fed barley fields with remaining bright straws covering the surface (center). Hence, as earlier noticed in Verrelst et al. (2012a,b), a practical implication of the uncertainty maps is the detection of areas that may benefit from a denser ground truth sampling regime. That was also the rationale for the collection of an extended training dataset over non-vegetated targets (i.e. *TrEx*). For both biophysical parameters, *TrEx* resulted in reduced uncertainties across the whole map (see also statistics in Table 4), and regions with large uncertainties have been considerably reduced. That trend is again visualized by the scatter plots displayed down Fig. 3; the majority of pixels fall below the 1:1 line. Consequently, the extended training dataset demonstrated its superiority; it was validated with high accuracy, yielded realistic LCC and LAI maps, and these maps were delivered with lower uncertainties.

Table 3
Validation statistics (r^2, absolute RMSE and NRMSE (%)) for the Sentinel configurations and LCC and LAI as generated by the *TrOr*- and *TrEx*-developed GPR models.

Parameter	r^2		abs. RMSE		NRMSE (%)	
	TrOr	*TrEx*	*TrOr*	*TrEx*	*TrOr*	*TrEx*
S2-10m:						
LCC	0.931	0.949	5.36	4.70	10.50	9.21
LAI	0.910	0.948	0.51	0.39	9.37	7.28
S2-20m:						
LCC	0.991	0.993	1.92	1.77	3.76	3.48
LAI	0.916	0.958	0.49	0.35	9.09	6.51
S2-60m:						
LCC	0.977	0.993	2.37	1.81	6.49	3.55
LAI	0.934	0.959	0.29	0.36	8.42	6.63

5.3. Evaluating portability GPR models to other experimental S2 images by inspecting scatterplots

Even though promising results have been obtained over the reference Barrax site, a key requirement for operational applications is to ensure that GPR models are portable to other regions with similar accuracy. Henceforth, the *TrOr*- and *TrEx*-developed GPR models were applied to simulated S2 images over various sites in Spain, Germany and Canada. Mean estimate and uncertainty *TrOr* and *TrEx* maps were generated and again correlated. Scatter plots are displayed in Fig. 4. Some interesting observations can be made from these scatter plots. For all tested images, LCC mean estimates fell right on the 1:1 line. It confirms the earlier observed trend that the extended training dataset did not lead to erratic LCC predictions. But also the scatter plots of the uncertainty estimates confirms the earlier observed systematic trend; for all images the *TrEx*-developed models yielded a systematic decrease in uncertainty, e.g. LCC $\pm\sigma$ hardly reached above 10 $\mu g/cm^2$.

Also with respect to LAI estimates, we observed the same pattern encountered in the reference image, i.e. the majority of *TrEx*-processed pixels largely follows the 1:1 line. However, a considerable amount of pixels tended to deviate towards lower estimates. While this trend appeared only slightly over Barrax and Demmin, the down-estimating occurred strongly over the scarcely vegetated areas of Los Monegros and Las Tablas. Also over Sudbury (Canada), LAI values were underestimated despite being dominated by vegetated surfaces. But most important is that, for all images, a significant and systematic decrease in uncertainties took

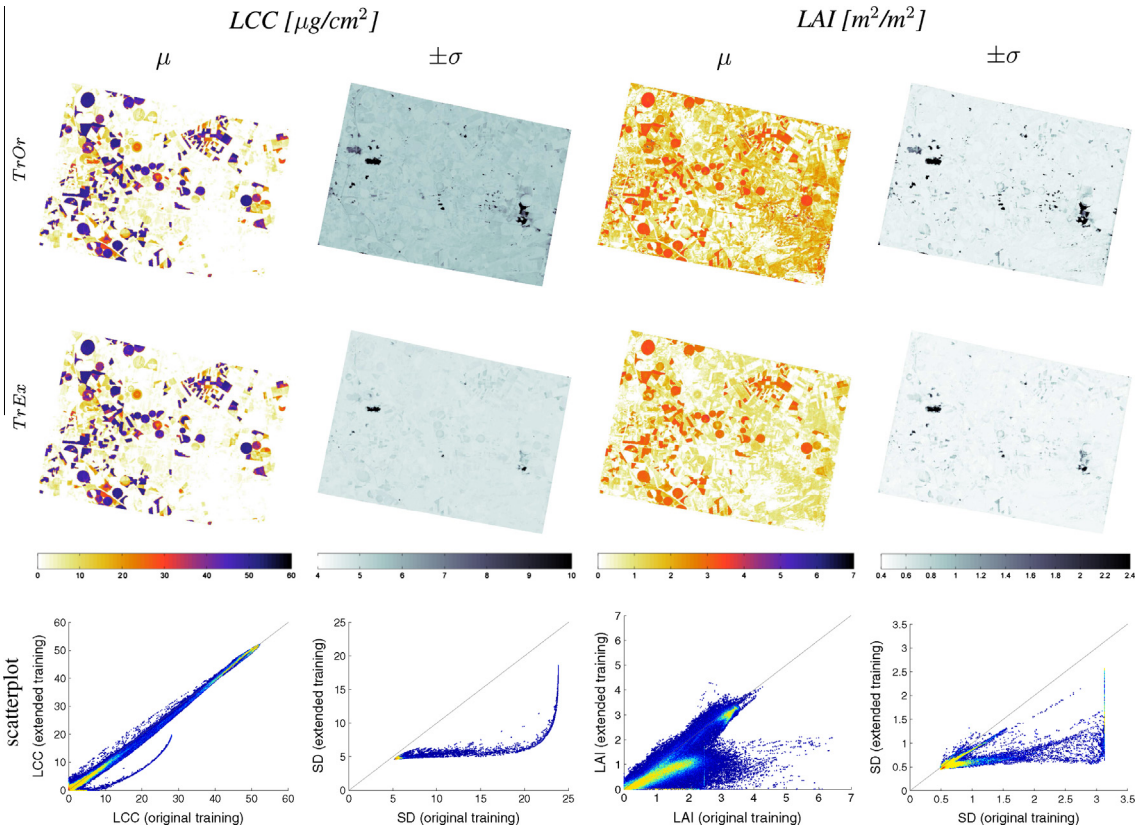

Fig. 3. Mean estimate (μ) and associated uncertainty ($\pm\sigma$) maps for LCC and LAI as generated by the S2-10m original training dataset (*TrOr*) [top], and extended training dataset (*TrEx*) [middle]. Scatterplots of *TrOr* vs. *TrEx* are shown below.

Table 4
Overview statistics (mean ($\bar{x}$) and standard deviation (SD)) for LCC and LAI μ (mean prediction) and $\pm\sigma$ (associated uncertainty) S2-10m maps as retrieved by *TrOr*- and *TrEx*-developed GPR models. 'Barrax July 03' is the reference image.

Image	μ LCC		$\pm\sigma$ LCC		μ LAI		$\pm\sigma$ LAI	
	TrOr $\bar{x}$ (SD)	*TrEx* $\bar{x}$ (SD)	*TrOr* $\bar{x}$ (SD)	*TrEx* $\bar{x}$ (SD)	*TrOr* $\bar{x}$ (SD)	*TrEx* $\bar{x}$ (SD)	*TrOr* $\bar{x}$ (SD)	*TrEx* $\bar{x}$ (SD)
Barrax July 03	7.71 (14.89)	7.54 (14.76)	5.53 (1.04)	4.75 (0.41)	1.21 (1.00)	0.82 (0.89)	0.58 (0.26)	0.51 (0.13)
Barrax July 04	9.48 (14.75)	9.22 (14.54)	5.50 (1.03)	4.75 (0.39)	1.37 (1.02)	1.13 (0.91)	0.60 (0.32)	0.51 (0.14)
Barrax June 09	12.44 (11.34)	10.66 (11.49)	6.10 (1.39)	4.98 (0.51)	0.91 (0.99)	0.73 (1.07)	0.73 (0.32)	0.73 (0.32)
Demmin May 06	25.38 (18.35)	24.31 (18.61)	5.75 (0.50)	5.00 (0.25)	2.37 (0.96)	1.70 (1.02)	0.66 (0.19)	0.59 (0.09)
Demmin July 06	26.60 (11.08)	25.69 (11.78)	5.53 (0.23)	4.93 (0.14)	1.85 (0.65)	1.37 (0.58)	0.55 (0.08)	0.55 (0.06)
Monegros Aug 05	2.16 (7.24)	2.42 (6.95)	7.22 (2.05)	4.96 (0.27)	1.81 (1.24)	0.24 (0.46)	1.44 (0.88)	0.67 (0.21)
Las Tablas July 06	7.99 (7.61)	7.42 (7.12)	5.67 (0.88)	4.78 (0.19	0.66 (0.83)	0.24 (0.38)	0.65 (0.35)	0.55 (0.11))
Sudbury Aug 07	6.92 (4.67)	3.65 (4.71)	8.30 (1.40)	5.97 (0.47)	2.32 (0.27)	0.63 (0.44)	1.80 (0.52)	0.96 (0.23)

place. Particularly those areas that earlier suffered from large uncertainties were now predicted with more certainty.

Overview statistics (mean and standard deviation) for prediction and uncertainty maps are provided in Table 4. With regard to portability evaluation, special attention goes to the mean uncertainty statistics along the different images. For both GPR models and LCC and LAI maps uncertainty statistics turned to be almost as good as the reference image. It can thus be concluded that the locally trained GPR model is generally applicable to other sites, thereby reaching uncertainties on the same order of the Barrax 2003 reference site. Only the Sudbury data was processed with considerably less certainty. Moreover, for all images, mean uncertainties have been considerably reduced when comparing *TrEx* with *TrOr*. Summarizing, the portability of the extended training dataset (*TrEx*) model has been successfully evaluated as mean uncertainties stabilized to about the same level as the reference Barrax image.

5.4. S2-10m biophysical parameter mapping

In this section we pay attention to the *TrEx*-generated maps, see Fig. 5. A first observation across the different sites is that LCC and LAI estimations fall within expected ranges. Variations in land cover are clearly observable and non-vegetated surfaces can be easily distinguished from vegetated surfaces. A second observation is that for most of the sites uncertainty maps show rather low values in general. Nevertheless, noticeable within-image variations are still occurring, particularly for LAI over the Barrax 2009 image. Only for Sudbury (Canada) systematically greater uncertainties appeared. Here a suboptimal atmospheric correction may explain the poorer performance.

While the uncertainty maps provide some information about the robustness of the retrievals, one has nevertheless to be careful with its interpretation. Note that $\pm\sigma$ represents the uncertainty interval around the mean predictions, meaning that they need to

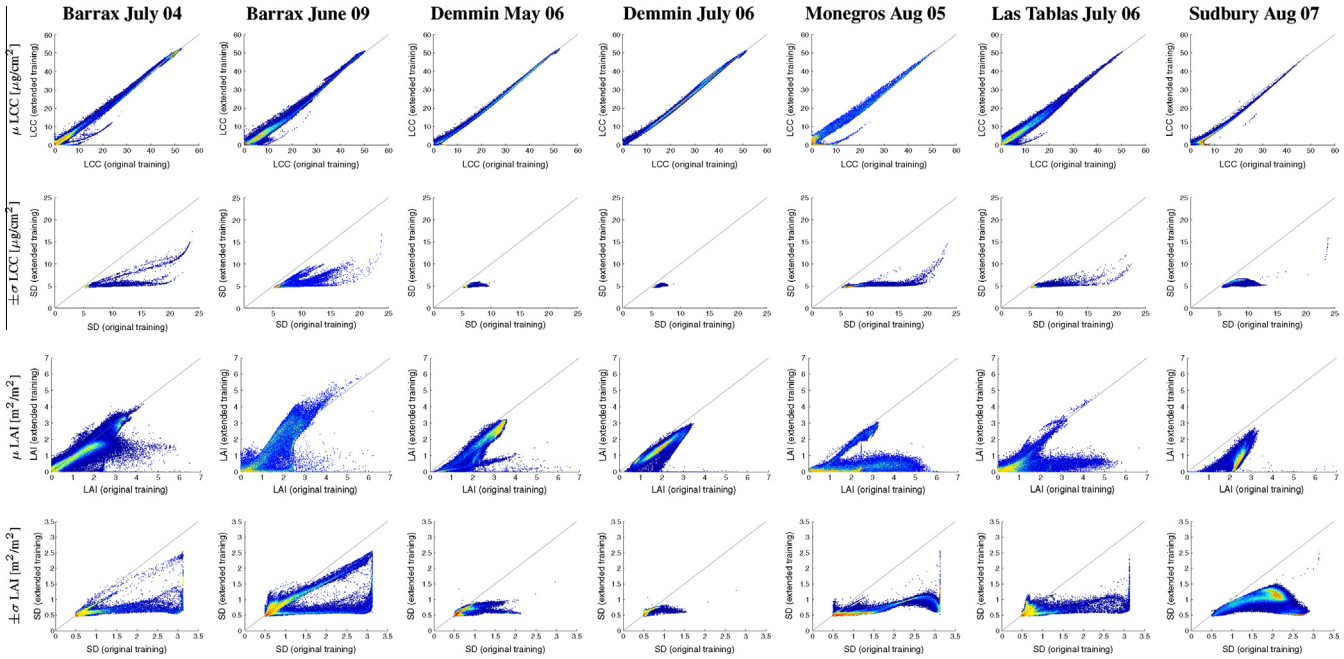

Fig. 4. Scatterplots of maps generated by the S2-10m original training datasets (*TrOr*) vs. extended training dataset (*TrEx*) for μ LCC [top], $\pm\sigma$ LCC [below], μ LCC [below] and $\pm\sigma$ LAI [bottom].

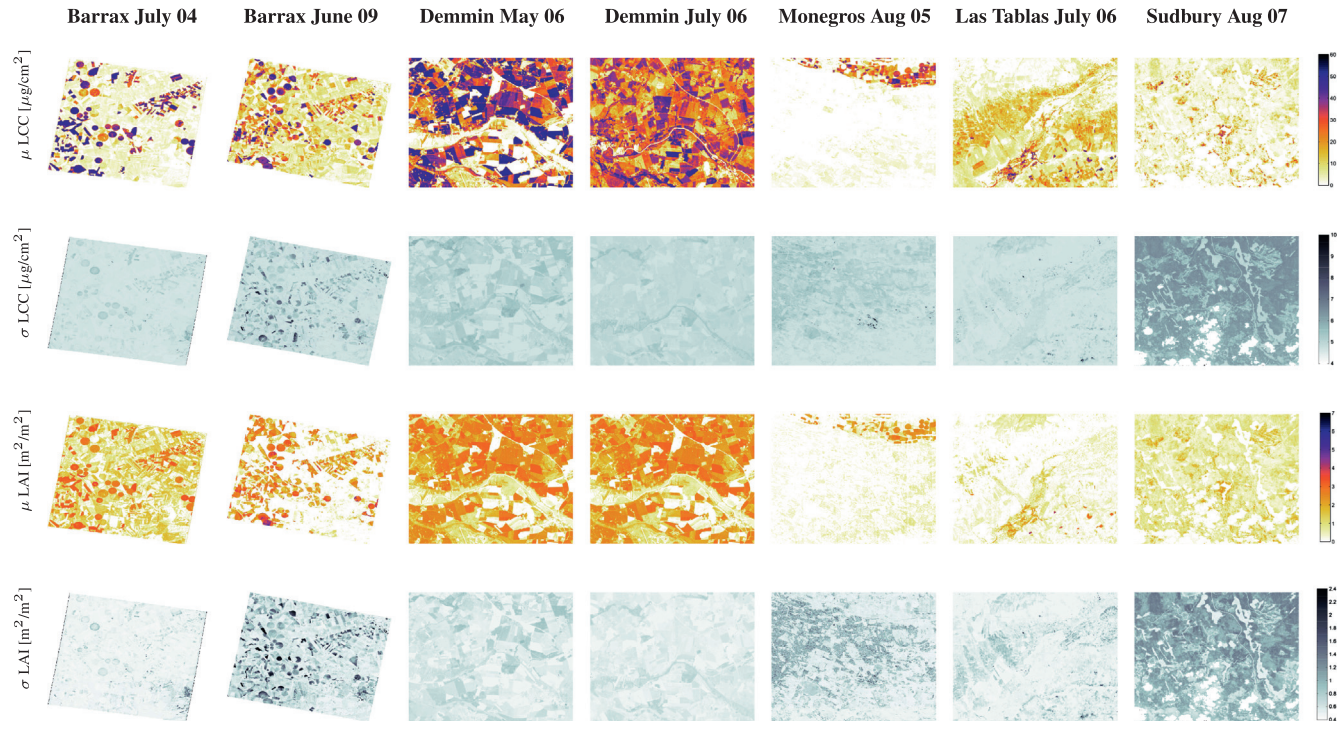

Fig. 5. Mean estimation (μ) and uncertainty ($\pm\sigma$) S2-10m maps for LCC and LAI as generated by the *TrEx*-developed GPR model.

be interpreted in relation to the estimates. For instance, an LCC uncertainty interval of about 5 would be more problematic for a mean estimate of 5 μg/cm^2 than of 50 μg/cm^2. Therefore, to evaluate the robustness of the GPR models it requires calculating the coefficient of variation:

$$CV = \frac{\sigma}{\mu} * 100. \quad (6)$$

CV maps provide relative uncertainties and are displayed in Fig. 6. These maps can then be evaluated against an uncertainty threshold, e.g. as proposed by GCOS, i.e. 20% (GCOS, 2011). Considering the reference Barrax image, dark areas represent retrievals with high uncertainties. These are typically bare soil areas which are characterized by low estimates (close to zero) and a relative high $\pm\sigma$. Conversely, it can be observed that most reliable retrievals occurred on the irrigated agricultural sites. However, only 9.0% of the pixels fell below the GCOS's 20% threshold. This low number seems surprising because the map was earlier excellently validated. It underlines the limited meaning of a (sparse) validation dataset when inspecting a heterogeneous map as a whole.

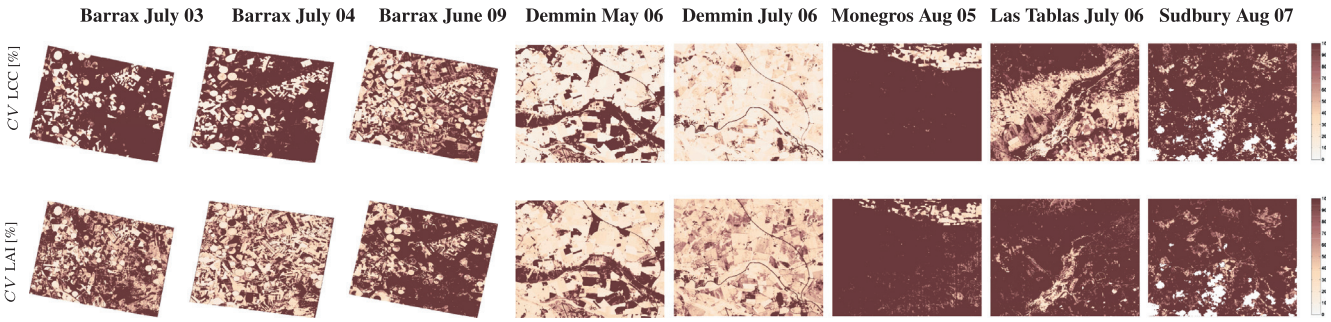

Fig. 6. Coefficient of variation (*CV*) S2-10m maps for LCC [top] and LAI [bottom] as generated by the *TrEx*-developed GPR model.

When comparing the *CV* maps across the different sites it can be observed that particularly over the agricultural areas meaningful estimates were obtained. For instance, the July 2006 Demmin LCC map is processed with low uncertainty over the whole image; 53.5% of pixels fell below the 20% threshold, and 75.6% below the 30% threshold. Also here agricultural parcels were processed with low uncertainty. Only non-vegetated surfaces such as water bodies are flagged with a high relative uncertainty. Spurious relative uncertainties are also observed across the Los Monegros and Las Tablas sites where bare soil dominate. With *TrEx* it was attempted to account for these bare soils, but due to near-zero mean estimates (0.0001) and, although $\pm\sigma$ is reduced, they still fall above near-zero threshold. *CV* turned therefore above 100%. Note that these relative uncertainty maps suggest that, at 10 m resolution, there is a greater problem of portability within an image than to other images. Solely the Sudbury maps show systematically higher relative uncertainties. This is probably due to the more difficult atmospheric correction as a consequence of cloud cover and lower sunlight intensity. The Sudbury case underlines the importance of an accurate and consistent atmospheric correction processing chain, which is expected to be provided by ESA (Level 2A Product). From a practical perspective, the *CV* map serves as a useful quality layer that allows masking out biophysical parameters estimates within an acceptable uncertainty (e.g. $CV < 20\%$), and at the same time can mask out non-vegetated surfaces (e.g. $CV > 100\%$).

6. Discussion

The forthcoming S2 mission opens opportunities to implement novel retrieval algorithms in operational processing chains. The interest is put on retrieval algorithms that are accurate, fast, robust, and sufficiently flexible to make fully use of the new S2 MSI bands. Machine learning regression algorithms (MLRAs) are able to cope with most of these objectives. In an earlier work, GPR was evaluated as a very promising regressor in terms of processing speed and accuracy when using a local training dataset (Verrelst et al., 2012a). At the same time, GPR is transparent in terms of model development (Rasmussen and Williams, 2006); it may provide a ranking of features (bands) and samples (spectra), thus alleviating the black-box problem. A discussion on its performance in comparison to alternative state-of-the-art retrieval approaches presented literature was provided in Verrelst et al. (2012a,b). In short, in a local setting GPR reached accuracies with S2 MSI band settings comparable (LAI) or superior (LCC) to competitive approaches. Nevertheless, the portability of this statistical approach to other sites remained questionable.

The here presented extended experimental training dataset (*TrEx*) not only further improved performances but also allowed a decrease in theoretical uncertainties. This proof of concept underlines the importance of a broad and diverse training dataset. More importantly, the GPR models were successfully applied to simulated S2 images covering various sites; associated uncertainties were on the same order as those generated by the reference image. The S2-10m examples demonstrated that excellent retrievals can be achieved already with 4 bands at a high spatial resolution of 10 m. Specifically, over fully vegetated surfaces relative uncertainties fell below the 20% requirements proposed by GCOS. This is encouraging for processing data from broadband sensors with a limited number of bands such as SPOT and Landsat or high resolution image such as Ikonos and Quickbird. On the other hand, S2 MSI encompasses additional bands at a coarser spatial resolution of 20 and 60 m. Note hereby that MSI's SWIR bands B11 and B12 have not been considered in this study because of falling outside the CHRIS' spectral range. These SWIR bands are known to be sensitive to vegetation structure (Brown et al., 2000) and can better distinguish between dried-out fallow and non-vegetated lands. It is expected that inclusion of the SWIR bands will further improve the retrieval quality and reduce uncertainties. Moreover, further improvements can be achieved, which may lead to further reduction of uncertainties over non-vegetated surfaces. For instance, one could develop a set of GPRs each working with different portions of the data or features. The uncertainty intervals could be optimally combined to improve the accuracy and reduce prediction bias. Alternatively, more sophisticated kernel functions that exploit signal-to-noise relations could be eventually considered (Gómez-Chova and Camps-Valls, 2012; Lazaro-Gredilla et al., 2013).

At the same time, it should be noted that in operational biophysical parameter retrievals, pixels over non-vegetated targets are flagged as non-valid and are not considered in uncertainty calculation (see Fang et al., 2012). Thereby, at coarser pixel's size of MODIS and SPOT/VEGETATION, non-vegetated areas are well-defined by land cover maps, such as desert, inland water body, urban surfaces. In contrast, at 10 m resolution non-vegetated pixels can virtually appear anywhere. Since those kinds of pixels can be easily identified (e.g., mean estimates of ≈ 0, or $CV > 100\%$), they could actually just be discarded when interpreting uncertainty estimates. Another remark to bear in mind is that the derived theoretical uncertainties are directly related to what has been presented during the training phase. Theoretical uncertainties are thus not intended to replace the true physical uncertainties of the biophysical parameter products but instead to provide additional complementary information. Physical uncertainties are mandatory to be provided and should be obtained through comprehensive validation datasets collected on various sites, such as that coordinated by the Committee on Earth Observation Satellites (CEOS) Land Product Validation (LPV) community (Morisette et al., 2006).

Finally, it does not escape our attention that only a limited set of images acquired during May-August have been evaluated. In an attempt to make the training data more representative at temporal and global scales, it should be able to cope with the majority of global land cover types along latitudinal gradients and over time.

Ideally, ground truth data (biophysical parameters plus associated radiometric data) should be collected over a broad variety of terrestrial surfaces and vegetation types at multiple phenological stages. This, however, is a tedious and expensive job. Alternatively, an urging open question in this respect is to evaluate how robust GPR performs when trained by artificial spectra, e.g. as generated by a radiative transfer model (RTM). The advantage of RTMs 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 such a hybrid approach (e.g. Baret et al., 2007; Bacour et al., 2006). Initial efforts in this direction have been undertaken by implementing a MRLA module into the ARTMO (Automated Radiative Transfer Models Operator) toolbox, which is a suite of RTMs into one GUI toolbox (Verrelst et al., 2012c). As such, GPR and other MLRA models can be automatically developed through lookup tables of simulated spectra and corresponding input parameters (e.g. LCC, LAI). Nevertheless, it should hereby be noted that, while GPR works successful for small training datasets (e.g. <2000 samples), its heavy computational load impedes the use of very large datasets. This limitation has to be resolved when aiming to develop generic models for global applications. For instance, alternative (sparse) versions of GPR have been proposed that can handle large scale datasets, see e.g. http://www.gaussianprocess.org/. Greedy algorithms in active learning settings can be also an alternative: GPRs trained on different data portions can provide a ranked list of the most informative spectra which could be used then to generate a final model. Forthcoming research will move in these directions, ultimately to reach a robust and generic retrieval processing scheme.

7. Conclusions

The delivery of uncertainty information is a prerequisite for the operational use of remote sensing products. Currently only few retrieval approaches provide such estimates. Gaussian Processes Regression (GPR), a machine learning regression algorithm (MLRA) based on Bayesian nonparametric modeling, is one of them. Such retrieval algorithm may be of interest in view of ESA's forthcoming Sentinel-2 (S2) mission. In this work these uncertainty estimates were used to evaluate the robustness and portability of locally-trained GPR models. Hyperspectral CHRIS data was used for the simulation of experimental S2 images, being: S2-10m (4 bands), S2-20m (8 bands) and S2-60m (10 bands). While providing accurate estimates when trained with a local dataset called SPARC (*TrOr*; Barrax, Spain), it was also observed that non-vegetated pixels were processed with great uncertainty. In an attempt to make the LCC and LAI GPR models more robust, an extended training dataset was introduced (*TrEx*), i.e. the original SPARC dataset plus 60 spectra over all kinds of non-vegetated surfaces (e.g., soils, man-made surfaces, water bodies). For the Barrax 2003 reference image, *TrEx*-developed GPR model delivered validation errors below 10% (NRMSE: 3.5–9.2%; r^2: 0.95–0.99), robust LCC and improved LAI estimates, and above all reduced associated uncertainty estimates. These locally-trained GPR models were subsequently applied to other experimental S2 images over various sites across Spain, Germany, Canada. The uncertainty estimates provided insight in the success of the models' performance. Overall, GPR models are portable to other images and uncertainty estimates can thereby function as quality layer to filter out unreliable retrievals. Uncertainty intervals were on the same order as the Barrax 2003 reference image and relative uncertainties over vegetated surfaces were below the 20% requirements proposed by GCOS. However, typically large uncertainty variation within an image was observed due to surface heterogeneity. GPR is concluded as a powerful regressor for remote sensing applications; not only it delivers accurate predictions, it is currently the only MLRA that provides associated uncertainty intervals.

Acknowledgements

This paper has been partially supported by the Spanish Ministry for Science and Innovation under Projects AYA2010-21432-C02-01 and TIN2012-38102-C03-01. Dr. Luis Guanter at the Free University of Berlin is thanked for providing the atmospherically-corrected CHRIS images. Two anonymous reviewers are thanked for providing comments that helped to improve the quality of the original manuscript.

References

Alonso, L., Moreno, J., 2005. Advances and limitations in a parametric geometric correction of CHRIS/PROBA data. In: Proceedings of the 3rd CHRIS/Proba Workshop, ESA/ESRIN, Frascati, Italy. <http://earth.esa.int/workshops/chris_proba_05/papers/06_alonso.pdf>.

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.

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.

Baret, F., Buis, S., 2008. Estimating canopy characteristics from remote sensing observations. Review of methods and associated problems. In: Liang, S. (Ed.), Advances in Land Remote Sensing: System, Modeling, Inversion and Application. Springer, pp. 171–200.

Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Niño, F., Weiss, M., Samain, O., Roujean, J., Leroy, M., 2007. LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. Part 1: principles of the algorithm. Remote Sensing of Environment 110 (3), 275–286.

Brown, L., Chen, J., Leblanc, S., Cihlar, J., 2000. A shortwave infrared modification to the simple ratio for LAI retrieval in boreal forests: an image and model analysis. Remote Sensing of Environment 71 (1), 16–25.

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.), 2009. Kernel Methods for Remote Sensing Data Analysis. Wiley & Sons, UK.

Camps-Valls, G., Gómez-Chova, L., Vila-Francés, J., Amorós-López, J., Muñoz-Marí, J., Calpe-Maravilla, J., 2006. Retrieval of oceanic chlorophyll concentration with relevance vector machines. Remote Sensing of Environment 105 (1), 23–33.

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

Colombo, R., Bellingeri, D., Fasolini, D., Marino, C., 2003. Retrieval of leaf area index in different vegetation types using high resolution satellite data. Remote Sensing of Environment 86 (1), 120–131.

Delegido, J., Verrelst, J., Alonso, L., Moreno, J., 2011. Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. Sensors 11 (7), 7063–7081.

Delegido, J., Verrelst, J., Meza, C., Rivera, J., Alonso, L., Moreno, J., 2013. A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems. European Journal of Agronomy 46, 42–52.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sensing of Environment 120, 25–36.

Fang, H., Wei, S., Jiang, C., Scipal, K., 2012. Theoretical uncertainty analysis of global MODIS, CYCLOPES, and GLOBCARBON LAI products using a triple collocation method. Remote Sensing of Environment 124, 610–621.

GCOS, 2011. Systematic Observation Requirements for Satellite-Based Products for Climate, 2011 Update, Supplemental Details to the Satellite-Based Component of the Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (2010 update, GCOS-154), pp. 138. <http://www.wmo.int/pages/prog/gcos/Publications/gcos-154.pdf>.

Gómez-Chova, L., Camps-Valls, G., September 2012. Learning with the kernel signal-to-noise ratio. In: IEEE Workshop on Machine Learning for Signal Processing, MLSP'12. Santander, Spain.

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.

Knyazikhin, Y., Martonchik, J., Myneni, R., Diner, D., Running, S., 1998. Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data. Journal of Geophysical Research D: Atmospheres 103 (D24), 32257–32275.

Lazaro-Gredilla, M., Titsias, M., Verrelst, J., Camps-Valls, G., 2013. Retrieval of biophysical parameters with heteroscedastic Gaussian processes. Geoscience and Remote Sensing Letters, IEEE PP (99), 1–5.

Meroni, M., Colombo, R., Panigada, C., 2004. Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar plantations. Remote Sensing of Environment 92 (2), 195–206.

Moreno, J., Participants of the SPARC campaigns, 2004. SPARC Data Acquisition Report. Contract No: 18307/04/NL/FF. University Valencia. <https://earth.esa.int/c/document_library/get_file?folderId=21020&name=DLFEs-386.pdf>.

Morisette, J., Baret, F., Privette, J., Myneni, R., Nickeson, J., Garrigues, S., Shabanov, N., Weiss, M., Fernandes, R., Leblanc, S., Kalacska, M., Sánchez-Azofeifa, G., Chubey, M., Rivard, B., Stenberg, P., Rautiainen, M., Voipio, P., Manninen, T., Pilant, A., Lewis, T., Iiames, J., Colombo, R., Meroni, M., Busetto, L., Cohen, W., Turner, D., Warner, E., Petersen, G., Seufert, G., Cook, R., 2006. Validation of global moderate-resolution LAI products: a framework proposed within the CEOS land product validation subgroup. IEEE Transactions on Geoscience and Remote Sensing 44 (7), 1804–1814.

O'Hagan, A., 1994. Bayesian Inference. Kendall's Advanced Theory of Statistics, vol. 2B. Arnold, London, UK.

Orbanz, P., Teh, Y.-W., 2010. Bayesian Nonparametric Models. Encyclopedia of Machine Learning. Springer.

Pinty, B., Andredakis, I., Clerici, M., Kaminski, T., Taberner, M., Verstraete, M., Gobron, N., Plummer, S., Widlowski, J.-L., 2011. Exploiting the MODIS albedos with the two-stream inversion package (JRC-TIP): 1. Effective leaf area index, vegetation, and soil properties. Journal of Geophysical Research D: Atmospheres 116 (9).

Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. The MIT Press, New York, US.

Tipping, M.E., 2001. The relevance vector machine. Journal of Machine Learning Research 1, 211–244.

Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J., Camps-Valls, G., Moreno, J., 2012a. Machine learning regression algorithms for biophysical parameter retrieval: opportunities for Sentinel-2 and -3. Remote Sensing of Environment 118, 127–139.

Verrelst, J., Alonso, L., Camps-Valls, G., Delegido, J., Moreno, J., 2012b. Retrieval of vegetation biophysical parameters using Gaussian process techniques. IEEE Transactions on Geoscience and Remote Sensing 50 (5 PART 2), 1832–1843.

Verrelst, J., Romijn, E., Kooistra, L., 2012c. Mapping vegetation density in a heterogeneous river floodplain ecosystem using pointable CHRIS/PROBA data. Remote Sensing 4 (9), 2866–2889.

Verrelst, J., Alonso, L., Rivera Caicedo, J., Moreno, J., Camps-Valls, G., 2013. Gaussian process retrieval of chlorophyll content from imaging spectroscopy data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (submitted for publication).

Weiss, M., Baret, F., Garrigues, S., Lacaze, R., 2007. LAI and fAPAR CYCLOPES global products derived from VEGETATION. Part 2: validation and comparison with MODIS collection 4 products. Remote Sensing of Environment 110 (3), 317–331.

Wu, C.F.J., 1986. Jackknife, bootstrap and other resampling methods in regression analysis. Annals of Statistics 14 (4), 1261–1295.