Retrieval of Biophysical Parameters With Heteroscedastic Gaussian Processes

Miguel Lázaro-Gredilla, *Member, IEEE*, Michalis K. Titsias,
Jochem Verrelst, and Gustavo Camps-Valls, *Senior Member, IEEE*

***Abstract*—An accurate estimation of biophysical variables is the key to monitor our Planet. Leaf chlorophyll content helps in interpreting the chlorophyll fluorescence signal from space, whereas oceanic chlorophyll concentration allows us to quantify the healthiness of the oceans. Recently, the family of Bayesian nonparametric methods has provided excellent results in these situations. A particularly useful method in this framework is the Gaussian process regression (GPR). However, standard GPR assumes that the variance of the noise process is independent of the signal, which does not hold in most of the problems. In this letter, we propose a nonstandard variational approximation that allows accurate inference in signal-dependent noise scenarios. We show that the so-called variational heteroscedastic GPR (VHGPR) is an excellent alternative to standard GPR in two relevant Earth observation examples, namely, Chl vegetation retrieval from hyperspectral images and oceanic Chl concentration estimation from *in situ* measured reflectances. The proposed VHGPR outperforms the tested empirical approaches, as well as statistical linear regression (both least squares and least absolute shrinkage and selection operator), neural nets, and kernel ridge regression, and the homoscedastic GPR, in terms of accuracy and bias, and proves more robust when a low number of examples is available.**

***Index Terms*—Biophysical parameter retrieval, Gaussian processes (GPs), heteroscedastic models, retrieval.**

I. Introduction

THE main goal of remote sensing is to monitor the Earth and its interaction with the atmosphere. The analysis can be done at local or global scales by looking at biogeochemical cycles, atmospheric situations, and vegetation dynamics [1], [2]. All these complex interactions can be studied through the definition of physical parameters either representing different properties for land (e.g., surface temperature, crop yield, defoliation, biomass, and leaf area coverage), water (e.g., yellow substance, ocean color, suspended matter, or chlorophyll concentration), or the atmosphere (e.g., temperature and moisture profiles at different altitudes). The field of physical parameter estimation constitutes an intermediate modeling step necessary to transform the measurements into useful estimates [3].

Manuscript received July 1, 2013; revised July 23, 2013; accepted July 24, 2013. This work was supported in part by the Spanish Ministry of Economy and Competitiveness (MINECO) under Project TIN2012-38102-C03-01.

M. Lázaro-Gredilla is with the Department of Signal Processing and Communications, Universidad Carlos III de Madrid, 28911 Leganés, Spain (e-mail: miguel@tsc.uc3m.es).

M. K. Titsias is with the Department of Informatics, Athens University of Economics and Business, 104 34 Athens, Greece (e-mail: mtitsias@aueb.gr).

J. Verrelst and G. Camps-Valls are with the Image Processing Laboratory (IPL), Universitat de València, 46980 València, Spain (e-mail: jochem.verrelst@uv.es; gustavo.camps@uv.es).

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

Digital Object Identifier 10.1109/LGRS.2013.2279695

The problem of physical parameter estimation has been approached with either statistical, physical, or hybrid methods [4]. In this letter, we will focus on the statistical approximation. Statistical models can be either parametric or nonparametric. Parametric models rely on physical knowledge of the problem and build explicit parametrized expressions that relate a few spectral channels with the biogeophysical parameter of interest. A plethora of vegetation indexes have been introduced, for instance, to study the vegetation status by estimating chlorophyll content and other leaf pigments [5]. Many of these indexes have been derived from high-resolution spectrometers, which include many (up to more than 200) hyperspectral bands. The simple calculation of these indexes has made possible deriving reasonable maps of vegetation properties in a quick and easy way. Nevertheless, the majority of the indexes only use up to five bands, thus under-exploiting the full potential of the hyperspectral datacube [6]. The same happens for oceanic or atmospheric parametric models.

Alternatively, nonparametric models are adjusted to predict a variable of interest using a training data set of input–output data pairs, which come from concurrent measurements of the parameter and the corresponding reflectance/radiance observation. Several nonparametric approaches have been introduced for physical parameter retrieval. Partial least squares regression has been, for instance, used for mapping canopy nitrogen [7], and a nonlinear extension has been introduced via kernels in [8] for chlorophyll content prediction. The support vector regression [9] has yielded good results in modeling oceanic chlorophyll [10].

Recently, the family of Bayesian nonparametric methods has provided excellent results. The relevance vector machine (RVM) [11] was used for the sparse approximation of oceanic chlorophyll [12]. To avoid problems of over-sparsification of the solution given by the RVM, a particularly useful method in this framework is Gaussian process regression (GPR) [13], which has been used in remote sensing problems [14]–[16]. GPR is simpler and generally more robust than other statistical regression tools. In addition to the good numerical performance and stability, GPR requires a relatively small training data set, it can adopt very flexible kernel functions, it identifies the relevant bands and observations in establishing relationships with a variable, and finally it provides confidence intervals for the predictions.

However, standard GPR assumes that the variance of the noise process is independent of the signal, which does not hold in most of the problems (see Fig. 1). This strong assumption of *homoscedasticity* is generally broken in many biophysical retrieval problems because the acquisition process is typically affected by noise in different amounts depending on the measured range of the variable. In order to deal with

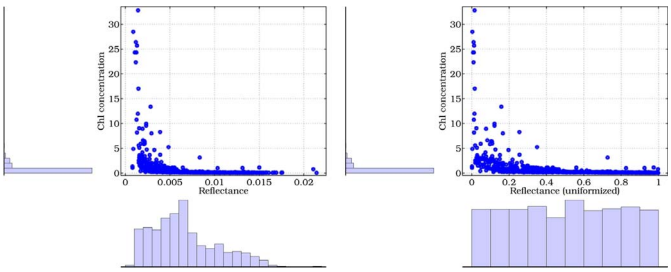

Fig. 1. Illustration of a *heteroscedastic* relation between the measured reflectance (at 412 nm) and the chlorophyll concentration in the SeaBAM data set used in the experiments. The left plot shows a clear heteroscedastic and nonlinear relationship between the independent (reflectance) and dependent (chlorophyll) variables. The right plot reveals that, even after marginal uniformization of the input (based on the cumulative density function of the marginal probability density function), the relation is still heteroscedastic.

input-dependent noise conditions, *heteroscedastic* models need to be designed; however, their tractability is not easy. In this letter, we propose a nonstandard *variational approximation* that allows accurate inference in signal-dependent noise scenarios. We show that the so-called variational heteroscedastic GPR (VHGPR) is an excellent alternative to standard GPR and generally outperforms other parametric and nonparametric approaches such as neural networks (NNs) and kernel regression.

The remainder of the letter is outlined as follows: Section II briefly reviews the theory of standard GPR. Section III introduces the proposed VHGPR for parameter retrieval. Section IV illustrates the performance of VHGPR in two relevant Earth observation examples: Chl vegetation retrieval from hyperspectral images, and oceanic Chl concentration estimation from *in situ* measured reflectances. We conclude the letter and outline future research lines in Section V.

II. GPR

Gaussian processes (GPs) are a well-established powerful nonparametric framework for nonlinear regression (see [13] for a thorough treatment). As it is common in most regression approaches, GPR models the observations (often referred to as *outputs*) $\{y_n\}_{n=1}^N$ as the sum of some unknown latent function $f(\mathbf{x})$ of the inputs $\{\mathbf{x}_n\mathbb{R}^D\}_{n=1}^N$ plus *constant power (homoscedastic)* Gaussian noise, i.e.,

$$y_n = f(\mathbf{x}_n) + \varepsilon_n, \qquad \varepsilon_n \sim \mathcal{N}(0, \sigma^2).$$

However, instead of proposing a parametric form for $f(\mathbf{x})$ and learning its parameters, in order to fit the observed data well, GPR proceeds in a Bayesian nonparametric form. A zero-mean[1] GP prior is placed on the latent function $f(\mathbf{x})$ and a Gaussian prior is used for each latent noise term ε_n

$$f(\mathbf{x}) \sim \mathcal{GP}\left(\mathbf{0}, k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}')\right); \quad \varepsilon_n \sim \mathcal{N}(0, \sigma^2) \tag{1}$$

where $k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}')$ is a covariance function parameterized by $\boldsymbol{\theta}$, and σ^2 is a hyperparameter that specifies the noise power.

A GP is a stochastic process whose marginals are distributed as a multivariate Gaussian. In particular, given priors (1), samples drawn from $f(\mathbf{x})$ at the set of locations $\{\mathbf{x}_n\}_{n=1}^N$ follow a joint multivariate Gaussian with zero mean and a covariance matrix $\mathbf{K_{ff}}$ with $[\mathbf{K_{ff}}]_{ij} = k_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j)$.

[1]It is customary to subtract the sample mean to data $\{y_n\}_{n=1}^N$ and then to assume a zero mean model.

If we consider a test location $\mathbf{x}_*$ with corresponding output y_*, priors (1) induce the following joint prior distribution between the observations $\mathbf{y} \equiv \{y_n\}_{n=1}^N$ and y_*:

$$\begin{bmatrix} \mathbf{y} \\ y_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K_{ff}} + \sigma^2\mathbf{I}_n & \mathbf{k_{f*}} \\ \mathbf{k_{f*}}^\top & k_{**} + \sigma^2 \end{bmatrix}\right).$$

Collecting available data in $\mathcal{D} \equiv \{\mathbf{x}_n, y_n | n = 1, \ldots, N\}$, it is possible to analytically compute the posterior distribution over the unknown output y_*

$$\begin{aligned} p(y_*|\mathbf{x}_*, \mathcal{D}) &= \mathcal{N}\left(y_*|\mu_{\mathrm{GP}*}, \sigma^2_{\mathrm{GP}*}\right) \\ \mu_{\mathrm{GP}*} &= \mathbf{k_{f*}}^\top(\mathbf{K_{ff}} + \sigma^2\mathbf{I}_n)^{-1}\mathbf{y} \\ \sigma^2_{\mathrm{GP}*} &= \sigma^2 + k_{**} - \mathbf{k_{f*}}^\top(\mathbf{K_{ff}} + \sigma^2\mathbf{I}_n)^{-1}\mathbf{k_{f*}}. \end{aligned}$$

The corresponding hyperparameters $\{\boldsymbol{\theta}, \sigma\}$ are typically selected by type-II maximum likelihood, using the marginal likelihood (also called evidence) of the observations, which is also analytical (now explicitly conditioning on $\boldsymbol{\theta}$ and σ):

$$\log p(\mathbf{y}|\boldsymbol{\theta}, \sigma) = \log \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K_{ff}} + \sigma^2\mathbf{I}). \tag{2}$$

When the derivatives of (2) are also analytical, which is often the case, conjugated gradient ascent is typically used for optimization. The full training process takes $\mathcal{O}(N^3)$ time.

This illustrates a number of advantages of GPs: First, since they yield a full posterior predictive distribution over y_*, it is possible to obtain not only the mean predictions for test data but also the so-called "error-bars," assessing the uncertainty of the mean prediction. Second, the whole procedure only depends on a very small set of hyperparameters, which renders it virtually overfitting free. Apart from noise power σ^2, it is typical to only use two additional hyperparameters, defining signal power and smoothness, globally collected in θ. Third, the tuning of such hyperparameters can be performed using continuous optimization of the evidence, as opposed to the nonderivable grid search performed during cross-validation in other regression approaches.

III. Heteroscedastic GP

As we pointed out in the previous section, standard GPR is homoscedastic, i.e., assumes constant noise power σ^2 for all observations. This assumption can be too restrictive for some problems. Heteroscedastic GPs, on the other hand, let noise power vary smoothly throughout input space, by changing the prior over ε_n to

$$\varepsilon_n \sim \mathcal{N}(0, e^{g(\mathbf{x}_n)})$$

and placing a GP prior over $g(\mathbf{x}) \sim \mathcal{GP}(\mu_0\mathbf{1}, k_{\theta_g}(\mathbf{x}, \mathbf{x}'))$. Note that the exponential is needed[2] in order to describe the nonnegative variance. The hyperparameters of the covariance functions of both GPs are collected in $\boldsymbol{\theta}_f$ and $\boldsymbol{\theta}_g$, accounting for the signal and noise relations, respectively.

Relaxing the homoscedasticity assumption into heteroscedasticity yields a richer and more flexible model that contains the standard GP as a particular case corresponding to a constant $g(\mathbf{x})$. Unfortunately, this also hampers analytical tractability; hence, approximate methods must be used to obtain posterior distributions for $f(\mathbf{x})$ and $g(\mathbf{x})$, which are, in turn, required to compute the predictive distribution over y_*. Next, we summarize previous approaches to deal with the problem and the proposed variational alternative.

[2]Of course, other transformations are possible, just not as convenient.

A. Previous Approaches

The heteroscedastic GP (HGP) model was first described in [17], where an expensive Markov chain Monte Carlo (MCMC) procedure was used in order to implement full Bayesian inference. A faster but more limited method is presented in [18], in order to perform maximum *a posteriori* (MAP) estimation. These approaches have certain limitations: MCMC is hundreds of times slower, whereas MAP estimation does not integrate out all latent variables and is prone to overfitting.

B. Proposed Variational Heteroscedastic GP

Variational techniques allow to approximate intractable integrals arising in Bayesian inference and machine learning in general. They are typically used to 1) provide analytical approximations to the posterior probability of the unobserved variables and, hence, do statistical inference over these variables and 2) derive a lower bound for the marginal likelihood (or "evidence") of the observed data, which allows model selection because higher marginal likelihoods relate to greater probabilities of a model generating the data.

In order to overcome the aforementioned problems, a sophisticated variational approximation called *Marginalized Variational* (MV) approximation has been recently introduced in [19]. The MV approximation renders (approximate) Bayesian inference in the HGP model both fast and accurate. In [19], an analytical expression for the Kullback–Leibler divergence between a proposal distribution and the true posterior distribution of $f(\mathbf{x})$ and $g(\mathbf{x})$ (up to a constant) was provided. Minimizing this quantity, with regard to both the proposal distribution and the hyperparameters, yields an accurate estimation of the true posterior, while simultaneously performing model selection. Furthermore, the expression of the approximate mean and variance of the posterior of y_* (i.e., predictions) can be computed in closed form. We will refer to this variational approximation for heteroscedastic GPR as VHGPR.

1) MV Approximation: The standard variational approximation defines a lower bound on the evidence of the model

$$F\left(q(\mathbf{f}),q(\mathbf{g})\right)=\log p(\mathbf{y})-\mathbf{KL}\left(q(\mathbf{f})q(\mathbf{g})\|p(\mathbf{f},\mathbf{g}|\mathbf{y})\right) \quad (3)$$

by subtracting the Kullback–Leibler divergence (which is always nonnegative) from it. Since the value of $\log p(\mathbf{y})$ is independent of the variational densities $q(\mathbf{f})$ and $q(\mathbf{g})$, selecting them to maximize this bound is equivalent to minimizing $\mathbf{KL}(q(\mathbf{f})q(\mathbf{g})\|p(\mathbf{f},\mathbf{g}|\mathbf{y}))$, i.e., obtaining the best possible factorized approximation to the posterior in the KL sense.

As it stands, F depends on the two N-dimensional variational distributions $q(\mathbf{f})$ and $q(\mathbf{g})$. It is possible to obtain a simpler and tighter bound, by optimally removing its dependence w.r.t. $q(\mathbf{f})$. This new bound is known as the MV bound and can be expressed as

$$F\left(q(\mathbf{g})\right)=\log Z\left(q(\mathbf{g})\right)-\mathbf{KL}\left(q(\mathbf{g})\|p(\mathbf{g})\right) \quad (4)$$

with $Z(q(\mathbf{g}))=\int e^{\int q(\mathbf{g})\log p(\mathbf{y}|\mathbf{f},\mathbf{g})\mathbf{dg}}p(\mathbf{f})\mathbf{df}$ (see [19]).

The MV bound (4) upper bounds the standard variational bound (3) and (since it is the particular case of it) also lower bounds the evidence. Hence, $\log p(\mathbf{y})\geq F(q(\mathbf{g}))=F(q^*(\mathbf{f}),q(\mathbf{g}))\geq F(q(\mathbf{f}),q(\mathbf{g}))$.

2) MV Bound for the HGP Model: For the HGP likelihood and priors, the MV bound can be computed in closed form if we restrict $q(\mathbf{g})=\mathcal{N}(\mathbf{g}|\boldsymbol{\mu},\boldsymbol{\Sigma})$, i.e., to be a multivariate normal distribution. Note that we do not need to impose any constraint on $q(\mathbf{f})$ because the MV bound does not depend on it. Using $\mathbf{K_{ff}}$ and $\mathbf{K_{gg}}$ to name the covariance matrices resulting from evaluating covariance functions of $f(\mathbf{x})$ and $g(\mathbf{x})$ at the inputs respectively, (4) becomes

$$\begin{aligned}F(\boldsymbol{\mu},\boldsymbol{\Sigma})&=\log\int e^{\int\mathcal{N}(\mathbf{g}|\boldsymbol{\mu},\boldsymbol{\Sigma})\log p(\mathbf{y}|\mathbf{f},\mathbf{g})\mathbf{dg}}\mathcal{N}(\mathbf{f}|\mathbf{0},\mathbf{K_{ff}})\mathbf{df}\\&\quad-\mathbf{KL}\left(\mathcal{N}(\mathbf{g}|\boldsymbol{\mu},\boldsymbol{\Sigma})\|\mathcal{N}(\mathbf{g}|\boldsymbol{\mu}_0\mathbf{1},\mathbf{K_{gg}})\right)\\&=\log\mathcal{N}(\mathbf{y}|\mathbf{0},\mathbf{K_{ff}}+\mathbf{R})-\frac{1}{4}\mathrm{trace}(\boldsymbol{\Sigma})\\&\quad-\mathbf{KL}\left(\mathcal{N}(\mathbf{g}|\boldsymbol{\mu},\boldsymbol{\Sigma})\|\mathcal{N}(\mathbf{g}|\mu_0\mathbf{1},\mathbf{K_{gg}})\right)\end{aligned} \quad (5)$$

where $\mathbf{R}$ is a diagonal matrix with elements $[\mathbf{R}]_{ii}=e^{[\boldsymbol{\mu}]_i-[\boldsymbol{\Sigma}]_{ii}/2}$. Observe the correspondence between the first term of the MV bound in (5) and the evidence of a standard homoscedastic GP in (2). Optimizing this bound w.r.t. $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and hyperparameters $\boldsymbol{\theta}_f$ and $\boldsymbol{\theta}_g$, we simultaneously perform model selection and find an approximate posterior distribution for $\mathbf{g}$, i.e., $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$. With this information, finding the posterior distributions for $\mathbf{f}$ and y_* is immediate (see details in [19]). The full training process takes $\mathcal{O}(N^3)$ time, just as a standard GP (although the multiplicative constant is larger), thus enabling fast and accurate inference for heteroscedastic models.

IV. Experimental Results

We evaluate the performance of the proposed VHGPR recently presented in [19] and compare it with classical parametric models based on vegetation indexes and statistical regression methods in machine learning, such as least squares linear regression (LR), least absolute shrinkage and selection operator (LASSO), regression trees (TREE), kernel ridge regression (KRR), and GPs. Code for running all the statistical methods is provided at http://www.uv.es/gcamps/code/simpleR.html.

The method's performance is assessed in terms of accuracy [root-mean-square error (RMSE), mean absolute error (MAE)], bias (mean error, ME), and goodness-of-fit (Pearson's correlation coefficient, R). We did a one-way analysis of variance (ANOVA) to compare the means and squared errors of the models. In all the experiments, the available data were randomly split into two half-sets: 50% samples for training and 50% samples for testing performance. This experiment was repeated 100 times, and the average results are provided. We optimized the parameters through a threefold cross-validation in the training set and then show the performance for the test set. The minimum RMSE was taken as the selection criterion for all methods, except for GPR and VHGPR where the maximum marginal was the chosen optimizing criterion. In order to meet the positivity constraint of the dependent variable, before training, we transformed the data logarithmically, as proposed elsewhere.[3]

[3]While the interpretation of the model parameters cannot be made in terms of the original response, after transforming back the predictions, one can perfectly compute the accuracy and bias of the model's prediction. Alternatives to such convenient procedure may consider *learning* the warping transformation of the output variable [20] or including a multivariate Gaussianization transformation in the formulation [21].

TABLE I
ABSOLUTE |ME|, RMSE, MAE, AND CORRELATION COEFFICIENT BETWEEN THE ACTUAL AND ESTIMATED CHL CONCENTRATION (R) OF MODELS IN THE TEST SET. BEST RESULTS ARE HIGHLIGHTED IN BOLD AND THE SECOND BEST IN ITALICS

Method	|ME|	RMSE	MAE	R
Morel-1†,‡	0.023	0.178	0.139	0.956
Morel-3	0.025	0.182	0.143	0.954
CalCOFI 2-band cubic	0.051	0.177	0.142	0.960
CalCOFI 2-band linear	0.079	0.325	0.256	0.956
Ocean Chlorophyll 2, OC2	0.031	0.169	0.133	0.960
Ocean Chlorophyll 4, OC4	0.039	0.162	0.129	0.966
LR	0.240	0.295	0.256	0.957
LASSO	*0.002*	0.170	0.130	0.958
TREE	**0.001**	0.197	0.140	0.943
NN	0.013	*0.145*	*0.108*	*0.970*
KRR	0.018	0.155	0.112	0.967
GPR	0.011	0.150	0.105	0.968
VHGPR	0.008	**0.143**	**0.104**	**0.971**

† Morel-1 and CalCOFI 2-band linear are described by $C = 10^{a_o + a_1 R}$, Morel-3 and CalCOFI 2-band cubic are cubic interpolators described by $C = 10^{a_o + a_1 R + a_2 R^2 + a_3 R^3}$, and models OC2/OC4 are described by $C = a_o + 10^{a_1 + a_2 R + a_3 R^2 + a_4 R^3}$, where for Morel models $R = \log(Rrs443/Rrs555)$, for CalCOFI and OC2 models $R = \log(Rrs490/Rrs555)$, and for OC4 model $R = \log(\max\{Rrs443, Rrs490, Rrs510\}/Rrs555)$.

‡ The results provided in this table for Morel, CalCOFI and OC2/OC4 models slightly differ from the ones given in [22] since they are computed only for the test samples. In addition, models in [22] used all available data to fit the models and hence no validation procedure was followed.

A. Experiment 1: Oceanic Chlorophyll Concentration

We used the SeaBAM data set [22], which gathers 919 *in situ* measurements of chlorophyll concentration around the United States and Europe. The data set contains coincident *in situ* chlorophyll concentration and remote sensing reflectance measurements $R_{rs}(\lambda)[sr^{-1}]$ at some wavelengths (412, 443, 490, 510, and 555 nm) that are present in the SeaWiFS ocean color satellite sensor. The chlorophyll concentration values range from 0.019 to 32.79 mg/m and shows nonlinear and heteroscedastic relations (cf. Fig. 1). Smart ways to "linearize" the problem consist of computing (*ad hoc*) spectral ratios. However, in this letter, we aim to work directly with the original (untransformed) data. More information about the data can be obtained at http://seabass.gsfc.nasa.gov/seabam/seabam.html.

Table I reports the obtained results for a set of empirical (top) and statistical learning methods (bottom). The proposed VHGPR outperforms the rest, in accuracy (RMSE, MAE) and goodness-of-fit (R), and closely follows the decision tree in bias (ME). VHGPR performs better than GPR in all quality measures, particularly for bias (+27% in ME), but also for the accuracy (+4.7% in RMSE), and slightly for correlation (+0.3% in R). From the ANOVA, statistical differences between VHGPR and the rest of the models were observed for both bias ($F = 30.0, p < 0.01$) and accuracy ($F = 24.7, p < 0.001$). However, in particular, VHGPR, GPR, and NN were not statistically different in accuracy ($F = 0.09, p = 0.688$), but they showed differences in bias ($F = 10.0, p < 0.01$).

As an alternative robustness test to assess differences among models, we show in Fig. 2 (left) the evolution of RMSE for the most accurate models as a function of the number of predictions used. The curves are the result of averaging 100 realizations, and for each realization, we computed the RMSE with a fixed number of predictions chosen at random. In the limit, one obtains the results in Table I. Interestingly, the VHGPR model is consistently better than the rest of the methods and reveals better convergence rates to a lower error bound.

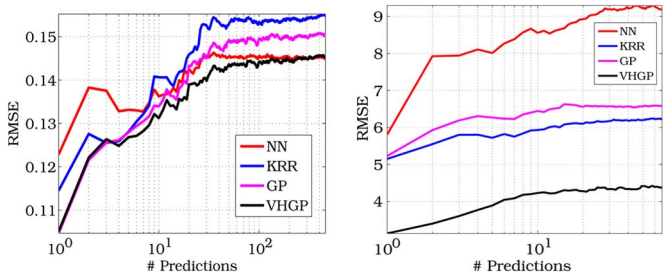

Fig. 2. Performance (RMSE) in the test set (averaged over 100 random realizations of the training-test data splitting) for different nonlinear statistical algorithms as a function of number of used predictions for (left) the SeaBAM and (right) the Barrax site.

B. Experiment 2: Vegetation Chlorophyll Concentration

Here, we aim to retrieve vegetation parameters using *in situ* measurements and hyperspectral images. The data were obtained in the SPARC-2003 (SPectra bARrax Campaign) and SPARC-2004 campaigns in Barrax, La Mancha, Spain. The region consists of approximately 65% dry land and 35% irrigated land. The methodology applied to obtain the *in situ* leaf-level Chl data consisted of measuring samples with a calibrated CCM-200 chlorophyll content meter in the field. Concurrently, we used Compact High Resolution Imaging Spectrometer (CHRIS) images Mode 1 (62 spectral bands, 34-m spatial resolution at nadir). The images were preprocessed, i.e., geometrically and atmospherically corrected [15]. Summarizing, a total of $n = 135$ datapoints in a 62-D space and the measured chlorophyll concentration constitute the database. Some very noisy bands were removed to improve the results.

Averaged results for the 100 realizations are shown for the test set in Table II. It is clearly observed that nonparametric methods show the best results, with both GPR and VHGPR performing best of the tested methods. In particular, Table II shows noticeable gains of +93.5% in ME, +33.5% in RMSE, +46.44% in MAE, and +2% in correlation. The ANOVA confirmed the differences between VHGPR and the other models for both bias ($F = 7.3, p < 0.01$) and accuracy ($F = 52.1$, $p < 0.001$). These results are confirmed by looking at the evolution of the RMSE as a function of the considered predictions in Fig. 2 (right), which clearly shows how VHGPR outperforms the other models.

The best VHGPR model was used for prediction on the whole CHRIS image to generate a pixel-by-pixel map of Chl and its confidence map (see Fig. 3). The maps clearly show the irrigated crops (the circles in orange-red), the seminatural areas (light blue), and the bare soil areas (dark blue). Although these maps cannot be used as a validation *per se*, the confidence maps allow us to draw conclusions on the performance of the retrievals. For example, the high confidences (western part of the image) were the fields sampled the most, whereas low confidence predictions (center of the image) correspond to areas particularly underrepresented in the training data, such as dry barley, harvested barley, and bright bare soils. This product

TABLE II
RESULTS OF NARROWBAND/BROADBAND INDEXES PROPOSED IN RELEVANT LITERATURE TESTED IN THE PRESENT STUDY ALONG WITH RECENT NONPARAMETRIC MODELS. BEST RESULTS ARE HIGHLIGHTED IN BOLD AND THE SECOND BEST IN ITALICS

Method	Formulation	ME	RMSE	MAE	R
GI	R_{672}/R_{550}	12.97	28.77	26.58	0.74
GVI	$(R_{682}-R_{553})/(R_{682}+R_{553})$	30.30	33.56	30.37	0.70
MCARI2	$1.2[2.5(R_{800}-R_{670})-1.3(R_{800}-R_{550})]$	5.99	14.89	12.02	0.72
mNDVI	$(R_{800}-R_{680})/(R_{800}+R_{680}-2R_{445})$	1.20	9.27	7.06	0.79
mNDVI_{705}	$(R_{750}-R_{705})/(R_{750}+R_{705}-2R_{445})$	1.22	9.13	6.30	0.80
mSR_{705}	$(R_{750}-R_{445})/(R_{705}+R_{445})$	2.52	10.94	7.99	0.76
mTVI	$1.2[1.2(R_{800}-R_{550})-2.5(R_{670}-R_{550})]$	5.99	14.89	12.02	0.72
NDVI	$(R_{800}-R_{670})/(R_{800}+R_{670})$	1.72	9.85	7.34	0.78
NDVI2	$(R_{750}-R_{705})/(R_{750}+R_{705})$	1.81	9.56	6.79	0.80
NPCI	$(R_{680}-R_{430})/(R_{680}+R_{430})$	16.53	31.05	27.13	0.72
OSAVI	$1.16(R_{800}-R_{670})/(R_{800}+R_{670}+0.16)$	1.72	9.85	7.34	0.78
PRI	$(R_{531}-R_{570})/(R_{531}+R_{570})$	25.58	35.96	32.14	0.77
PRI2	$(R_{570}-R_{539})/(R_{570}+R_{539})$	37.84	39.19	37.84	0.76
PSRI	$(R_{680}-R_{500})/R_{750}$	28.07	37.10	34.18	0.80
RDVI	$(R_{800}-R_{670})/\sqrt{(R_{800}+R_{670})}$	2.12	10.67	8.21	0.76
SIPI	$(R_{800}-R_{445})/(R_{800}-R_{680})$	17.18	31.54	28.62	0.76
SPVI	$0.4[3.7(R_{800}-R_{670})-1.2(R_{530}-R_{670})]$	1.79	11.09	8.70	0.71
SR1	R_{750}/R_{700}	3.16	11.76	8.49	0.75
SR3	R_{750}/R_{550}	0.87	9.78	7.51	0.75
SR4	R_{672}/R_{550}	12.97	28.77	26.58	0.74
SRPI	R_{430}/R_{680}	3.47	13.76	9.18	0.63
TVI	$0.5[120R_{750}-R_{550})-200(R_{670}-R_{550})]$	5.31	14.49	11.74	0.70
VOG	R_{740}/R_{720}	0.61	9.68	7.44	0.76
LR	ℓ_2 least squares	4.56	11.52	8.94	0.77
LASSO	ℓ_1 least squares	3.46	12.39	9.56	0.73
TREE	Pruning, min. split = 30	*0.14*	6.98	4.59	0.86
NN	Sigmoid links, one hidden layer	0.93	9.19	6.49	0.77
KRR	RBF kernel	0.73	6.22	5.24	0.89
GPR [15]	Anisotropic RBF kernel	1.69	6.57	5.19	*0.95*
VHGPR [19]	Anisotropic RBF signal/noise kernels	**0.10**	**4.37**	**2.78**	**0.97**

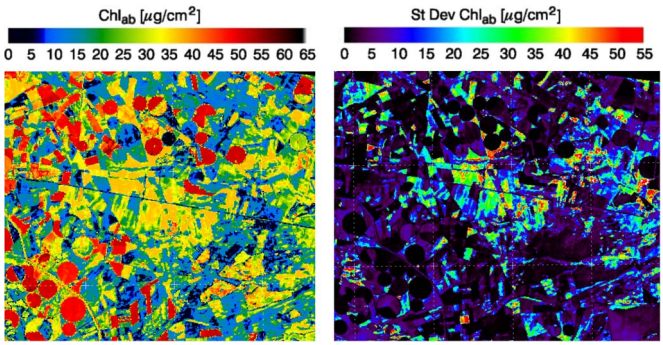

Fig. 3. (Left) Chlorophyll concentration prediction map and (right) predictive standard deviation obtained with VHGPR on the CHRIS 12-07-2003 nadir image.

may be used to set sensitivity margins of field instruments quite intuitively. These results confirm the very good properties of nonparametric models, in general, and VHGPR, in particular, for the estimation of vegetation properties from space.

V. CONCLUSION

Most of the real problems in remote sensing biophysical parameters exhibit nonlinear and heteroscedastic relations. Even with modern and sophisticated kernel methods, signal and noise relations are commonly disregarded. We have proposed the VHGPR for biophysical parameter retrieval. The model revealed as an excellent alternative to standard GPR in two relevant Earth observation estimation problems, in terms of accuracy and bias. Future work will consider the issue of the computational cost of the GP models via sparse-promoting priors, ensembles of models, and the analysis of the learned relations encoded in the kernel. In a more application side, we plan to exploit statistical regression in combination with radiative transfer models.

REFERENCES

[1] C.D. Rodgers, *Inverse Methods for Atmospheric Sounding: Theory and Practice*. Singapore: World Scientific, 2000.
[2] S. Liang, *Quantitative Remote Sensing of Land Surfaces*. New York, NY, USA: Wiley, 2004.
[3] F. Baret and S. Buis, "Estimating canopy characteristics from remote sensing observations: Review of methods and associated problems," in *Advances in Land Remote Sensing: System, Modeling, Inversion and Applications*. Berlin, Germany: Springer Verlag, 2008.
[4] G. Camps-Valls, D. Tuia, L. Gómez-Chova, S. Jiménez, and J. Malo, Eds., "Remote sensing image processing," in *Collection "Synthesis Lectures Image, Video, Multimedia Processing"*. LaPorte, CO, USA: Morgan & Claypool, Sep. 2011.
[5] J. Berni, P. Zarco-Tejada, L. Suárez, and E. Fereres, "Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 722–738, Mar. 2009.
[6] M. Schaepman, S. Ustin, A. Plaza, T. Painter, J. Verrelst, and S. Liang, "Earth system science related imaging spectroscopy—An assessment," *Remote Sens. Environ.*, vol. 113, no. 1, pp. S123–S137, Sep. 2009.
[7] P. Townsend, J. Foster, R. J. Chastain, and W. Currie, "Application of imaging spectroscopy to mapping canopy nitrogen in the forests of the central Appalachian Mountains using Hyperion and AVIRIS," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1347–1354, Jun. 2003.
[8] J. Arenas-García and G. Camps-Valls, "Efficient kernel orthonormalized PLS for remote sensing applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2872–2881, Oct. 2008.
[9] A.J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.
[10] G. Camps-Valls, L. Bruzzone, J. L. Rojo-Álvarez, and F. Melgani, "Robust support vector regression for biophysical variable estimation from remotely sensed images," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 3, pp. 339–343, Jul. 2006.
[11] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learning Res.*, vol. 1, pp. 211–244, Sep. 2001.
[12] G. Camps-Valls, L. Gómez-Chova, J. Vila-Francés, J. Amorós-López, J. Muñoz-Marí, and J. Calpe-Maravilla, "Retrieval of oceanic chlorophyll concentration with relevance vector machines," *Remote Sens. Environ.*, vol. 105, no. 1, pp. 23–33, Nov. 2006.
[13] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. New York, NY, USA: MIT Press, 2006.
[14] L. Pasolli, F. Melgani, and E. Blanzieri, "Gaussian process regression for estimating chlorophyll concentration in subsurface waters from remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 3, pp. 464–468, Jul. 2010.
[15] J. Verrelst, L. Alonso, G. Camps-Valls, J. Delegido, and J. Moreno, "Retrieval of vegetation biophysical parameters using Gaussian process techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1832–1843, May 2012.
[16] J. Verrelst, L. Alonso, J. P. Rivera, J. Moreno, and G. Camps-Valls, "Gaussian process retrieval of chlorophyll content from imaging spectroscopy data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 867–874, Apr. 2013.
[17] P. Goldberg, C. Williams, and C. Bishop, "Regression with input-dependent noise: A Gaussian process treatment," in *Proc. Advances NIPS*, 1998, pp. 493–499.
[18] K. Kersting, C. Plagemann, P. Pfaff, and W. Burgard, "Most likely heteroscedastic Gaussian processes regression," in *Proc. ICML*, 2007, pp. 393–400.
[19] M. Lázaro-Gredilla and M. K. Titsias, "Variational heteroscedastic Gaussian process regression," in *Proc. Int. Conf. Machine Learning, ICML 2011*, Bellevue, WA, USA, 2011, pp. 841–848.
[20] M. Lázaro-Gredilla, "Bayesian warped Gaussian processes," in *Proc. Adv. Neural Inf. Process. Syst.*, P. Bartlett, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., 2012, vol. 25, pp. 1628–1636. [Online]. Available: http://books.nips.cc/papers/files/nips25/NIPS2012_0765.pdf
[21] V. Laparra, G. Camps-Valls, and J. Malo, "Iterative Gaussianization: From ICA to random rotations," *IEEE Trans. Neural Netw.*, vol. 22, no. 4, pp. 537–549, Apr. 2011.
[22] J. E. O'Reilly, S. Maritorena, B. G. Mitchell, D. A. Siegel, K. Carder, S. A. Garver, M. Kahru, and C. McClain, "Ocean color chlorophyll algorithms for SeaWiFS," *J. Geophys. Res.*, vol. 103, no. C11, pp. 24 937–24 953, Oct. 1998.