

# Gaussian Process Sensitivity Analysis for Oceanic Chlorophyll Estimation

Katalin Blix, Gustau Camps-Valls, *Senior Member, IEEE* and Robert Jenssen, *Member, IEEE*

**Abstract**—Gaussian Process Regression (GPR) have experienced tremendous success in biophysical parameter retrieval in the last years. The GPR provides a full posterior predictive distribution so one can derive mean and variance predictive estimates, i.e. point-wise predictions and associated confidence intervals. GPR typically uses translation invariant covariances that make the prediction function very flexible and nonlinear. This, however, makes the relative relevance of the input features hardly accessible, unlike in linear prediction models. In this paper, we introduce the Sensitivity Analysis (SA) of the GPR predictive mean and variance functions to derive feature rankings and spectral spacings, respectively. The methodology can be used to uncover knowledge in any kernel-based regression method, it is fast to compute, and it is expressed in closed-form. The methodology is evaluated on GPR for global ocean chlorophyll prediction, revealing the most important spectral bands and their spectral spacings. We illustrate the (successful) methodology in several datasets and sensors.

**Index Terms**—Kernel methods, Gaussian Process Regression (GPR), Sensitivity Analysis (SA), Oceanic Chlorophyll Prediction

## I. INTRODUCTION

Being able to monitor ocean chlorophyll content from remotely sensed data provides the possibility to monitor the health status of oceans through the photosynthetic activity [1]. Changes in the photosynthetic activity result in changes in the chlorophyll fluorescence [2], [3]. Therefore detecting chlorophyll fluorescence from space can reveal the distribution of the marine primary producers, the phytoplankton [4], [5], [6], [7]. This has deep ecological [8] and economical implications<sup>1</sup>. In addition, monitoring ocean chlorophyll content also provides a tool to achieve deeper understanding of the contribution of CO<sub>2</sub> to the climate [9], [10], [11].

In this scenario, ocean chlorophyll estimation from space requires accurate and fast mapping algorithms. It is standard practice to use parametric bio-optical models, such as the OC2 and OC4 models [12], [13]. They, however, assume explicit relationships between the input reflectance bands and the chlorophyll content. They are relatively simple (empirical) nonlinear mapping functions (most of the times simple band ratios and polynomial functions), and very importantly, the

model complexity must be controlled based on experience. In recent years, many alternative algorithms have appeared using statistical machine learning algorithms for chlorophyll content estimation from multi- and hyperspectral data. Many methods have been studied: neural networks [14], support vector regression [15]–[17], and the relevance vector machine (RVM) [18]. The recently introduced GPR model has been shown to outperform other oceanic chlorophyll content estimation methods [19]. For oceanic chlorophyll content estimation in remotely sensed data GPR framework has been successfully applied by [20], [21]. GPR differs from other machine learning methods not only in its predictive power, but also in its underlying fundamental principles [22], [23]. The other advantageous property of GPR is that it provides additional information about the prediction: the predictive variance. Thus the output of the regression is not only the estimated chlorophyll content, but also the estimated variance, which reveals the confidence of the prediction.

Although GPR has shown an excellent predictive performance, the information about the relative relevance of the features being used for regression is lost, since the model is a non-linear kernel method that defines an implicit (not accessible) feature mapping. In [24] we presented the *sensitivity analysis* (SA) of the mean function in the GPR model and applied it to three oceanic chlorophyll matchup datasets.

Our contribution in this work is the extension of our method to the predictive variance function of the GPR model. In addition, we further exploit the SA for both the predictive mean and variance function by evaluating their performances on both controlled examples and new updated global oceanic chlorophyll relevant datasets. We compare the methodology with state-of-art oceanic chlorophyll content estimation models for the Sea-Viewing Wide Field-of-View Sensor (SeaWiFS), the Medium Resolution Imaging Spectrometer (MERIS) and the NASA operational Moderate Resolution Imaging Spectroradiometer onboard AQUA (MODIS-Aqua). Furthermore, we present our results on sensitivity maps, which are aimed to show the possibility for applying our method for practical purposes.

SA reduces to study the variance (uncertainty) of the predictive function in terms of the uncertainties of the input features. The family of SA methods is vast and depends on the number of problem *constraints*, also known as *settings*. In this work, we focus on the particular field of *local methods*, which involve taking the partial derivative of model's output with respect to input features to assess its impact. Interestingly, in the case of the GPR model, such gradients can be computed in closed-form for most of the covariance functions.

Introducing the SA for the GPR mean function for determining feature relevance has the advantage that it's not limited to a specific kernel function, such as in the case of the

Manuscript received December 16, 2016;

KB is at the Department of Physics and Technology, University of Tromsø - The Arctic University of Norway. E-mail: katalin.blix@uit.no.

GCV is with the Image Processing Laboratory (IPL), Universitat de València. C/ Catedrático Escardino, Paterna (València) Spain. Web: <http://isp.uv.es>. E-mail: gustau.camps@uv.es.

RJ is with the Machine Learning @ UiT Lab at the Department of Physics and Technology, University of Tromsø - The Arctic University of Norway. E-mail: robert.jenssen@uit.no.

This paper has been partially supported by the Research Council of Norway, FRIPRO grant 239844 (RJ), and in part by the European Research Council under the ERC consolidator grant SEDAL-647423.

<sup>1</sup><http://oceancolor.gsfc.nasa.gov/>

Automatic Relevance Determination (ARD), where the length-scale parameters of the Squared Exponential (SE) kernel are optimized in order to assign feature relevance [25], [26]. Furthermore, the SA of the GPR variance function, to the best of our knowledge is the only existing method that can reveal the spectral spacing of the input space in the GPR model.

In order to gain an intuitive interpretation, we first present the SA on a controlled example, and then apply it to five global chlorophyll related datasets from different sensors: SeaBAM, SeaWiFS, MODIS-Aqua, and two MERIS complementary datasets. SA can efficiently reveal the most important spectral bands for chlorophyll content prediction, and the spectral spacing of the input space globally. In addition, we compare GPR using only the most relevant spectral bands for regression with spectral-band-ratio models. We also present sensitivity maps for both the GPR predictive mean and variance, which open the possibility to present the distribution of the most relevant wavelengths on a global scale, and also to access information about the (spatially resolved) distribution of the spectral sampling of the inputs. The sensitivity maps might indicate the detection of the distribution of chlorophyll fluorescence, thus opening the possibility to monitor ocean status through a fundamentally different, mathematically solid, approach.

Finally, we validate the results of the SA of the GPR mean function on a global scale by producing global chlorophyll content maps, allowing visual comparison with the actual measured chlorophyll content maps and predicted chlorophyll content maps computed with parametric models.

The objective of performing the SA of the GPR model on oceanic chlorophyll datasets was that GPR has been shown to have a strong regression capacity in the estimation of biophysical parameters, therefore the methodology could be efficiently used in practice for oceanic chlorophyll content estimation from remotely sensed data. However, the driving mechanisms of the GPR model hasn't been fully understood yet. This is in contrast to parametric models, where the estimated coefficients (weights) allow the direct interpretation of the relevance of the spectral bands. Applying the SA to the GPR mean function for oceanic matchup datasets, revealed the most important spectral bands in the regression model. This not only shows that the algorithm performs well and can be extended to a variety of kernel methods, but also provides a tool for having a deeper understanding in the optical properties of the oceans. Furthermore, using the SA of the GPR variance function for these datasets, results a unique interpretation of the spectral spacing of the input space. In addition, our aim by presenting sensitivity maps for the oceanic chlorophyll content datasets was to show that the SA can be used for mapping the most relevant bands and their spectral spacing which might be important when the biophysical and optical properties of the oceans are in focus. These sensitivity maps can be used for information retrieval purposes from remotely sensed data in the very important task of oceanic chlorophyll content estimation.

The remainder of the paper is organized as follows. Section II reviews the GPR model, presents the SA and an illustrative toy example. Section III details the data collection and experimental setup used in this paper. Section IV gives the experimental results for estimation of ocean chlorophyll content, comparison of the GPR (using only those bands which were ranked as

most relevant of the SA) with parametric state-of-art models for ocean chlorophyll content estimation, sensitivity maps and validation maps. Finally, Section V concludes the paper and outlines the further work.

## II. SENSITIVITY ANALYSIS IN GAUSSIAN PROCESSES

We first review the standard formulation of the GPR model briefly, then present the SA of the GPR predictive mean and variance, and illustrate its performance in a toy example.

### A. Regression with Gaussian Processes

Standard regression approximates 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  $\{\mathbf{x}_n \in \mathbb{R}^D\}_{n=1}^N$  plus constant power Gaussian noise, i.e.  $y_n = f(\mathbf{x}_n) + \varepsilon_n$ ,  $\varepsilon_n \sim \mathcal{N}(0, \sigma^2)$ . Instead of proposing a parametric form for  $f(\mathbf{x})$  and learning its parameters in order to fit observed data well, GPR proceeds in a Bayesian, non-parametric way [22], [23]. A zero mean<sup>2</sup> Gaussian Process ( $\mathcal{GP}$ ) prior is placed on the latent function  $f(\mathbf{x})$  and a Gaussian prior is used for each latent noise term  $\varepsilon_n$ ,  $f(\mathbf{x}) \sim \mathcal{GP}(\mathbf{0}, k_\theta(\mathbf{x}, \mathbf{x}'))$ , where  $k_\theta(\mathbf{x}, \mathbf{x}')$  is a covariance function parameterized by  $\theta$ , and  $\sigma^2$  is a hyperparameter that specifies the noise power. Essentially, a  $\mathcal{GP}$  is a stochastic process whose marginals are distributed as a multivariate Gaussian. In particular, given the priors  $\mathcal{GP}$ , 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 covariance matrix  $\mathbf{K}_{\mathbf{ff}}$  with  $[\mathbf{K}_{\mathbf{ff}}]_{ij} = k_\theta(\mathbf{x}_i, \mathbf{x}_j)$ .

If we consider a test location  $\mathbf{x}_*$  with corresponding output  $y_*$ , the  $\mathcal{GP}$  defines a joint prior distribution between the observations  $\mathbf{y} \equiv \{y_n\}_{n=1}^N$  and  $y_*$ .

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

$$p(y_* | \mathbf{x}_*, \mathcal{D}) = \mathcal{N}(y_* | \mu_{\mathcal{GP}*}, \sigma_{\mathcal{GP}*}^2) \quad (1)$$

$$\mu_{\mathcal{GP}*} = \mathbf{k}_{\mathbf{f}*}^\top (\mathbf{K}_{\mathbf{ff}} + \sigma^2 \mathbf{I}_n)^{-1} \mathbf{y} = \mathbf{k}_{\mathbf{f}*}^\top \boldsymbol{\alpha} \quad (2)$$

$$\begin{aligned} \sigma_{\mathcal{GP}*}^2 &= \sigma^2 + k_{**} - \mathbf{k}_{\mathbf{f}*}^\top (\mathbf{K}_{\mathbf{ff}} + \sigma^2 \mathbf{I}_n)^{-1} \mathbf{k}_{\mathbf{f}*} \\ &= \sigma^2 + k_{**} - \mathbf{k}_{\mathbf{f}*}^\top \mathbf{A} \mathbf{k}_{\mathbf{f}*}, \end{aligned} \quad (3)$$

where  $\mathbf{k}_{\mathbf{f}*}$  is the covariance between the training vector and the test point,  $\boldsymbol{\alpha} = (\mathbf{K}_{\mathbf{ff}} + \sigma^2 \mathbf{I}_n)^{-1} \mathbf{y}$  is the weight vector of the GPR mean,  $k_{**}$  is the covariance between the test point with itself and  $\mathbf{A} = (\mathbf{K}_{\mathbf{ff}} + \sigma^2 \mathbf{I}_n)^{-1}$  is the weight matrix of the GPR variance.

Note that the predictive mean in Eq. (2)  $\mu_{\mathcal{GP}*}$  depends on the observations through the weight vector  $\boldsymbol{\alpha}$ , while the confidence intervals  $\sigma_{\mathcal{GP}*}^2$  (Eq. (3)) only depend on the inverse of the regularized covariance function  $\mathbf{A}$ .

### B. Sensitivity Analysis of features from Gaussian Processes

GPR offer some advantages over other regression methods. Since they yield a full posterior predictive distribution over  $y_*$  (Eq. (1)), it is possible to obtain not only mean predictions for test data,  $\mu_{\mathcal{GP}*}$  (Eq. (2)), but also the so-called "error-bars", assessing the uncertainty of the mean prediction,  $\sigma_{\mathcal{GP}*}^2$  (Eq.

<sup>2</sup>It is customary to subtract the sample mean to data  $\{y_n\}_{n=1}^N$ , and then to assume a zero mean model.

(3)). In this work, we focus on extracting knowledge from trained GPR model. To do so, let us define the sensitivity of feature  $j$  as

$$s_j = \int \left( \frac{\partial \phi(\mathbf{x})}{\partial \mathbf{x}_j} \right)^2 p(\mathbf{x}) d\mathbf{x}, \quad (4)$$

where  $p(\mathbf{x})$  is the probability density function over the  $D$ -dimensional input vector  $\mathbf{x}_n = [x_n^1, \dots, x_n^D]^\top$ , and  $\phi(\mathbf{x})$  represents either the predictive mean,  $\mu_{\text{GP}^*}$ , or variance,  $\sigma_{\text{GP}^*}^2$ . Intuitively, the objective of the sensitivity analysis of features is to measure the changes of the derivative of the function  $\phi(\mathbf{x})$  in the  $j$ th direction. In order to avoid the possibility of cancellation of the terms due to its signs, the derivatives are squared. Therefore, the resulting sensitivities will be positive  $s_j \geq 0$  for all bands. The empirical estimate of the sensitivity for the  $j$ th feature can be written as

$$s_j = \frac{1}{N} \sum_{n=1}^N \left( \frac{\partial \phi(\mathbf{x}_n)}{\partial x_n^j} \right)^2, \quad (5)$$

where  $N$  denotes the number of training samples. Before calculating the sensitivity, let us define the covariance prior that we used in this work, the standard isotropic scaled Gaussian kernel function:

$$k(\mathbf{x}_m, \mathbf{x}_n) = \nu^2 \exp \left( -\frac{1}{2} \sum_{d=1}^D \left( \frac{x_m^d - x_n^d}{\lambda_d} \right)^2 \right), \quad (6)$$

where  $\lambda_d$  is the length-scale for the dimension  $d$ , and  $\nu$  is a positive scale factor. The hyperparameters of this GP prior are collectively grouped in  $\theta = [\nu, \sigma, \lambda_1, \dots, \lambda_D]$ .

The resulting empirical estimate of the GPR mean sensitivity is therefore:

$$\begin{aligned} s_{\mu_{\text{GP}^*}}^j &= \frac{1}{N} \sum_{q=1}^N \left( \frac{\partial \phi(\mathbf{x}_q)}{\partial x_q^j} \right)^2 \\ &= \frac{1}{N} \sum_{q=1}^N \left( \frac{\partial \sum_{p=1}^N \alpha_p k(\mathbf{x}_p, \mathbf{x}_q)}{\partial x_q^j} \right)^2 \\ &= \frac{1}{N} \sum_{q=1}^N \left( \sum_{p=1}^N \frac{\alpha_p (x_p^j - x_q^j)}{\lambda_j^2} k(\mathbf{x}_p, \mathbf{x}_q) \right)^2, \end{aligned} \quad (7)$$

and for the GPR variance sensitivity is:

$$\begin{aligned} s_{\sigma_{\text{GP}^*}}^j &= -2N\nu^2 \times \\ &\sum_{q=1}^N \left( \sum_{p,q=1}^N A_{pq} (x_p^j - x_q^j) k(\mathbf{x}_p, \mathbf{x}_q)^2 / \lambda_j^2 \right)^2. \end{aligned} \quad (8)$$

Note that the calculation of the empirical sensitivity is computed in closed-form using only training data points and the inferred  $\alpha$  and  $\mathbf{A}$ . The SA derived here is inspired by [27] who however only regarded a support vector machine and brain research context, and who did not extend the analysis to variance.

### C. Proof of concept

We show the concept of the SA on a synthetic example. The goal of this experiment is to examine whether the SA of the GPR mean computed by Eq. (7) function can identify the relevant feature in the regression process. At the same time,

we compute the SA of the GPR variance function by using Eq. (8), so that the spacing,  $t$ , of the input features can be revealed.

Assume that the input consists of two features,  $\mathbf{x}_n = [x_n^1, x_n^2]$ , where  $x_n^1 = A \sin(2\pi t)$  is the relevant feature,  $x_n^2 \sim \mathcal{N}(0, \sigma^2)$  is irrelevant, and  $A \gg \sigma$ . The output is the sum of the two input features,  $y_n = x_n^1 + x_n^2$ . Time sampling is uniform for  $t \leq 0$  and logarithmically for  $t > 0$ . In order to trace the evolution of the sensitivities as  $t$  grows, we compute  $s_{\mu}^i$  and  $s_{\sigma}^i$  by using Eq. (7) and Eq. (8), respectively, for  $i = 1, 2$  through time  $t$ . Figure 1[top] shows the sensitivities for the GPR mean (Eq. (7)) for the relevant feature ( $s_{\mu}^1$ ) and for the noise  $s_{\mu}^2$ , respectively. It can be observed that the SA could consistently identify the relevant feature. The sensitivities of the GPR variance (computed by Eq. (8)) are shown in Fig. 1[bottom]. It can be seen how it correctly captures the change at  $t = 0$  related to the sampling rate. The SA of the GPR variance as expected, assigned greater values to the relevant feature. This example shows how the SA of the GPR can be used for determining the most relevant features, and to uncover the sampling rates of the input variables by using Eq. (7) and Eq. (8), respectively.

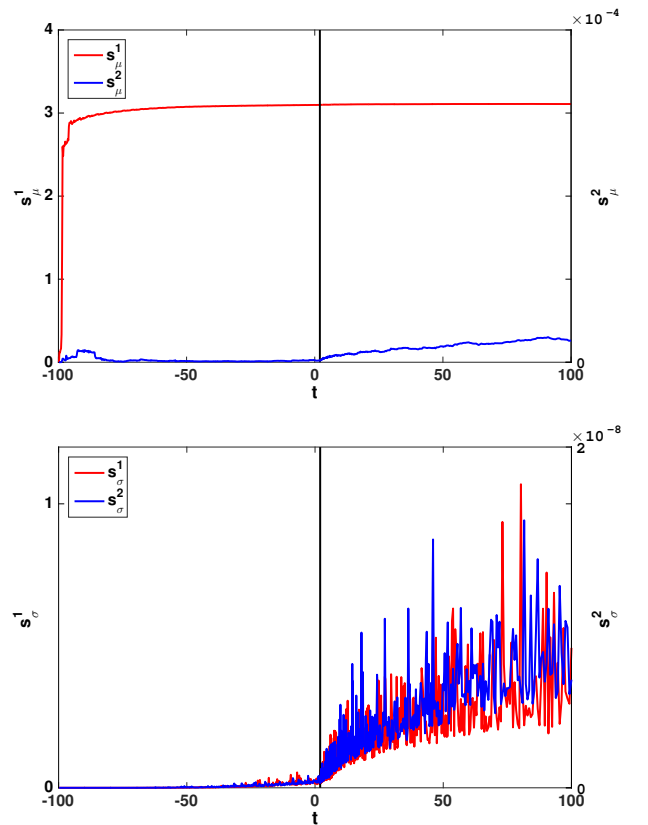


Fig. 1. Evolution of the sensitivities through time of the GPR mean (top) and variance (bottom) for the relevant (red) and irrelevant (blue) feature.

## III. DATA COLLECTION

In this work, we show results of the SA in five chlorophyll relevant datasets, acquired by different sensors and thus different spectral resolutions and complexity [28]: SeaBAM,

TABLE I  
DESCRIPTIVE STATISTICS OF THE FIVE DATASETS.

	SeaBAM		SeaWIFS	
Chlorophyll range (mgm <sup>-3</sup> )	0.019 - 32.787		0.024 - 129.332	
Nr. of samples	919		1465	
Band ( $\lambda_c$ (nm))	$\mu$	$\sigma^2$	$\mu$	$\sigma^2$
412	0.0066	0.1374·10 <sup>-4</sup>	0.0036	0.1001·10 <sup>-4</sup>
443	0.0059	0.0969·10 <sup>-4</sup>	0.0038	0.0625·10 <sup>-4</sup>
490	0.0049	0.039·10 <sup>-4</sup>	0.0041	0.0468·10 <sup>-4</sup>
510	0.0032	0.0163·10 <sup>-4</sup>	0.0038	0.047·10 <sup>-4</sup>
555	0.002	0.0163·10 <sup>-4</sup>	0.0038	0.0910·10 <sup>-4</sup>
670			0.001	0.0178·10 <sup>-4</sup>
	MERIS (synthetic)		MERIS (real)	
Chlorophyll range (mgm <sup>-3</sup> )	0.021 - 53.4429		0.017 - 40.23	
Nr. of samples	5000		567	
Band ( $\lambda_c$ (nm))	$\mu$	$\sigma^2$	$\mu$	$\sigma^2$
413	0.0258	0.0006	-1.7594	1760.2
443	0.0323	0.008	0.0031	0.0597·10 <sup>-4</sup>
490	0.0476	0.0018	0.0042	0.0624·10 <sup>-4</sup>
510	0.0524	0.0022	0.0045	0.0793·10 <sup>-4</sup>
560	0.0606	0.0033	0.0057	0.1784·10 <sup>-4</sup>
620	0.0285	0.0012	0.003	0.134·10 <sup>-4</sup>
665	0.0222	0.0008	0.0022	0.095·10 <sup>-4</sup>
681	0.0234	0.0007	0.0022	0.0873·10 <sup>-4</sup>
	MODIS-Aqua			
Chlorophyll range (mgm <sup>-3</sup> )	0.0153 - 25.4985			
Nr. of samples	579			
Band ( $\lambda_c$ (nm))	$\mu$	$\sigma^2$		
412	0.0028	0.8138·10 <sup>-5</sup>		
443	0.0032	0.4778·10 <sup>-5</sup>		
488	0.0036	0.302·10 <sup>-5</sup>		
531	0.0037	0.4422·10 <sup>-5</sup>		
547	0.0037	0.5556·10 <sup>-5</sup>		
667	0.0009	0.1302·10 <sup>-5</sup>		
678	0.001	0.1186·10 <sup>-5</sup>		

SeaWIFS, MODIS-Aqua, and two complementary MERIS datasets. (For further details on the SeaBAM and MERIS (synthetic) datasets we refer to [13], [14], [17], [18] and [19]. The SeaWIFS, MODIS-Aqua and MERIS (real) datasets can be obtained from the SeaBASS database<sup>3</sup>.) Table I summarizes the main parameters of the descriptive statistics of these datasets, such as the center wavelengths ( $\lambda_c$ ), the mean  $\mu$  and variance  $\sigma$  of each channel, the range of the chlorophyll-a concentrations and the total number of samples. Note, that we used the reflectances measured in Remote sensing reflectance (Rrs) for chlorophyll content prediction purposes. The SeaBAM dataset gathers 919 ocean chlorophyll measurements around the United States and Europe. The matchup dataset consists of coincident *in situ* remote sensing reflectance on five channels, which correspond to some of the SeaWIFS channels and chlorophyll-a concentration measurements. The bandwidths of the channels are 20 nm, and they are situated in the range between 402 nm and 565 nm. The chlorophyll-a concentrations range between 0.019-32.787 mgm<sup>-3</sup>. In addition, we applied the SA of features to three global remote sensing ocean chlorophyll data, the SeaWIFS, the MODIS-Aqua and the MERIS dataset [29]. The SeaWIFS dataset covers the spectral region between 402 nm and 680 nm on six channels. We used 1465 chlorophyll-a measurements with coincident Rrs between September 1997 and November 2010. Chlorophyll-a concen-

trations span a quite wide range, between 0.024 and 129.332 mgm<sup>-3</sup>. The MODIS-Aqua dataset has seven channels ranging from 405 nm to 683 nm. The data we used here has 579 measurements between July 2002 and November 2012, where the chlorophyll-a molecule concentrations are between 0.0153 and 25.4985 mgm<sup>-3</sup>. Finally, the MERIS dataset has the same channels as the synthetic MERIS data, consisting of 567 measurements between April 2002 and March 2012, where the range of the chlorophyll-a concentration is between 0.017 and 40.23 mgm<sup>-3</sup>. We applied the SA to these global data and computed sensitivity maps for an extracted area, East-USA. An additional MERIS dataset is formed by synthetic data, where 5000 coincident chlorophyll-a concentrations and Rrs were simulated [17]. The chlorophyll-a concentrations range between 0.021 and 53.4429 mgm<sup>-3</sup>. The Rrs were simulated on eight channels. The channels are placed between 407.5 nm and 685 nm, with a bandwidth of 10 nm and 7.5 nm. The means and the variances of the channels show similar values for all the five datasets. Generally, the means are situated close to zero and the variances are small. Note, the MERIS (real) dataset's mean value of band 1 differs from the rest of the means. The corresponding variance is large. This might indicate fault measurement(s) in the dataset at this band.

#### IV. EXPERIMENTAL RESULTS

Here we present the experimental results on the previous five different datasets. We first describe the experimental setup, and then study the provided feature ranking from the GPR model. Furthermore, we compare GPR, using only those bands for regression, which were assigned to greatest relevance by the SA, with commonly applied parametric models. Finally, we provide spatially explicit SA maps for both the predictive mean and variance.

##### A. Experimental setup

We trained five different GPR models for the corresponding datasets. In all cases, we standardized the input features and removed the mean of the observed chlorophyll content. We split the available data randomly into a training set (50%) and a test (hold-out) set. The hyperparameters  $\theta$  were optimized by maximizing the marginal log-likelihood [22] using the training set. Results of the best models are shown for the test set in Table II. We show different quality measures for the models: bias (mean error, ME), accuracy (root-mean-square error, RMSE, and mean absolute error, MAE), and goodness of fit (Pearson's correlation coefficient,  $\rho$ ). It can be noted that in all cases, the GPR models are accurate and generally unbiased, so a sensitivity analysis is feasible.

TABLE II  
SUMMARY OF THE TEST RESULTS IN THE 5 DATASETS. (The numbers in the parantheses refer to the most relevant channels which were used as inputs in the GPR.)

Database	ME	RMSE	MAE	$\rho$
SeaBAM (2, 4 and 5)	+0.0037	0.1493	0.1104	0.9679
SeaWIFS (4, 5 and 6)	-0.0887	0.3149	0.2361	0.9236
MODIS-Aqua (4, 5 and 6)	+0.0229	0.2461	0.1866	0.9188
MERIS (synthetic) (5, 6, 7 and 8)	0.004	0.084	0.0232	0.9996
MERIS (real) (5, 6, 7 and 8)	<10 <sup>-7</sup>	0.21	0.1464	0.9261

<sup>3</sup><http://seabass.gsfc.nasa.gov/seabasscgi/search.cgi>



### B. Sensitivity Analysis of the five datasets

We perform the SA of the GPR mean and variance functions for all five datasets (results are given in Fig. 2). For the SeaWiFS dataset the SA of the GPR mean revealed that band 4 (510 nm) is the most sensitive (Fig. 2 [top-left]), which matches previous results [18], and the accurate bio-optical model OC4. The second most sensitive band corresponds to 670 nm center wavelength (band 6). This is in good correspondence with the three-band reflectance difference model [30], where it was shown that adding band 6 is advantageous and results improved chlorophyll content prediction, especially when chlorophyll content increases. The inclusion of band 6 is based on similar principles as using band 5 (555 nm). The SA assigned the third greatest relevance to band 5, which is commonly used in band-ratio models, such as the OC2/OC3/OC4 and the three-band reflectance difference model [31] and [30]. Fig. 2[bottom-left] shows the result of the SA of the variance, where the most stable spectral band was band 6. Similar results are obtained in the SeaBAM dataset, where the SA assigned the greatest importance to bands 5 (555 nm), 4 (510 nm) and 2 (443 nm), see Fig. 2 [top-left]. Again, these results matches Morel, CalCOFI-2 and Ocean Color (OC) parametric models. The SA of the variance resulted that band-5 (555 nm) has the most stable spectral variance (Fig. 2[bottom-left]).

For the MODIS-Aqua dataset, bands 4, 5 and 6 were found to have the highest sensitivities of the GPR mean (Fig. 2[top-middle]). These channels correspond to 531 nm, 547 nm and 667 nm, respectively. Band 5 also used in the OC2 and OC3 parametric models. The position of band 4 on MODIS-Aqua was selected to improve the detection of the accessory pigments [32], while band 6 is one of the MODIS-Aqua channels to detect chlorophyll fluorescence [33], [34]. Nevertheless, the SA of the variance, Fig. 2[bottom-middle], assigned the lowest sensitivity to channel 6.

For the MERIS dataset, we performed the SA on both synthetic and real datasets. For the synthetic dataset, the SA resulted that band 8 (681 nm) has the greatest importance (Fig. 2[top-right]) with relative low spectral variance (Fig. 2[bottom-right]). This might be the indication of chlorophyll fluorescence [35], [36], [37]. Channel 7 and 8 were included on MERIS for the detection of the chlorophyll fluorescence signal. Being able to detect chlorophyll fluorescence has special importance when chlorophyll content mapping in coastal waters is in focus, since the presence of gelbstoff and suspended matter might mask the water-leaving radiance from chlorophyll-a, when spectral-band ratios are applied [35].

For the real MERIS dataset, the SA of the GPR mean resulted in that band 5 (560 nm) is the most sensitive (Fig. 2 [top-right]), also with the highest sensitivity of the GPR variance (Fig. 2[bottom-right]). This result is in good correspondence with the OC2/OC3/OC4 parametric models. Note, that band 6 (620 nm), 7 (665 nm) and 8 (681 nm) were also found to have high sensitivities in comparison to the rest of the channels. Looking at the sensitivities of the variance for these four channels reveals that band 8 has the lowest sensitivity of the predictive variance.

Applying the SA of features for the GPR mean for these global datasets might reveal the most relevant spectral band for global oceanic chlorophyll prediction. Apart from the synthetic MERIS dataset, in all cases the most sensitive band fell into

the spectral region between 510 nm and 560 nm. The SA of the GPR predictive variance opens the possibility to access the spectral sampling of the channels. This additional information might help selecting channels for analysis in an automated way, since channels with high sensitivity of the GPR mean and low sensitivity of the GPR variance should be preferred.

### C. Comparison of methods

We compared the performance of the GPR (using only the most sensitive bands) with parametric bio-optical models ([13], [31] and [38]) in all the five datasets. These models can be written as it follows [18]: Morel-1 and CalCOFI 2-band linear are expressed by  $C = 10^{a_0+a_1R}$ , Morel-3 and CalCOFI 2-band cubic interpolators are  $C = 10^{a_0+a_1R+a_2R^2+a_3R^3}$  and models OC2/OC3/OC4 are described by  $a_0 + \sum_{i=1}^4 a_i \log_{10} R^i$ , where  $R$  indicates the logarithmic ratio between the blue and green wavelengths, and the  $a_i$  are the coefficients. Note that the coefficients and the wavelengths used for determining  $R$  are sensor-specific (and they can be found at NASA's ocean color web site <http://oceancolor.gsfc.nasa.gov/>). Model performances were evaluated by computing the same measures as in Sec. IV-A. The goal of this comparison-study was to evaluate the regression strength of the GPR by using only the most important spectral bands, and compare them with the commonly used state-of-art algorithms. Therefore, the measures were computed by using the available datasets for both training and testing. We used bootstrapping for accessing model performance. (Note in Table II the prediction strength of the method was in focus, therefore the available datasets were randomly divided into training- testing data, as described in Sec. IV-A.) The results of the models for the five datasets can be seen in Table III. The measures are the mean values of 100 bootstrap samples. The distribution of the bootstrap measures is presented in Fig. 4. The box-plots reveal that the computed model measures from the bootstrap samples for the GPR model (indicated by GP in Fig. 4) has a narrow range (except for the RMSE and MAE in the case of the SeaWiFS and MODIS-Aqua datasets), low bias and high accuracy.

Applying only the most sensitive bands to the GPR can outperform other commonly used parametric models, which indicates the strength of the SA.

Note, that parametric models have been previously compared to machine learning methods, for example in [18], where no statistically significant difference was found. Furthermore GPR model using all available features has been shown to outperform other machine learning methods [25]. The goal of our comparison study is to show the strength of the SA in the important task of chlorophyll content estimation from remotely sensed data. In addition to find the most relevant features for chlorophyll content estimation, we also examine how the GPR model using only the most important spectral bands performs in comparison to the state-of-art algorithms.

We tested the statistical significant of model's difference by performing a one-way Analysis of Variance (ANOVA) on the estimates. We performed the statistical analysis of the bias and accuracy of the residuals by computing the F-value and p-value for each cases [39]. Table IV shows the results of the ANOVA analysis for the five datasets. Significant statistical differences can be observed for both the bias and accuracy for

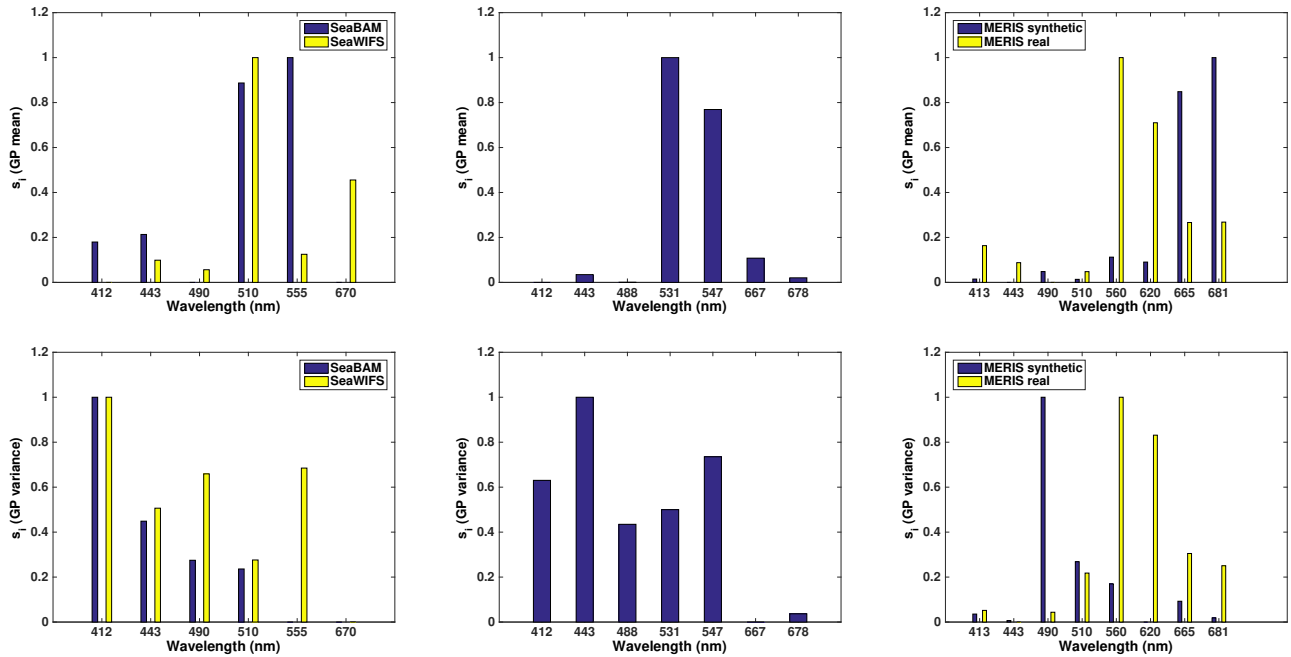


Fig. 2. SA of the GPR mean (top row) and variance (bottom row) for the SeaWIFS and SeaBAM dataset (left column), MODIS-Aqua dataset (middle column) and MERIS dataset (right column).

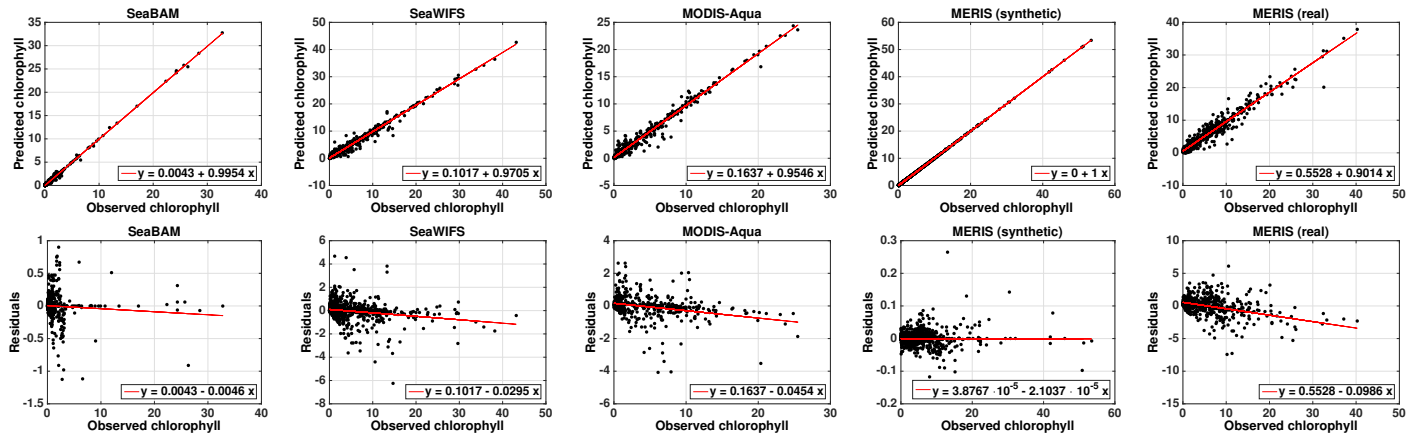


Fig. 3. Using only the most relevant bands in the GPR model for the five datasets. Observed versus predicted chlorophyll (top row) and observed chlorophyll versus residuals (bottom row).

the SeaBAM, MODIS-Aqua and MERIS (synthetic) dataset. In the case of the SeaWIFS and MERIS (real) datasets the statistical analysis couldn't reveal any difference in the bias between the GPR with the most relevant bands and the rest of the models. However, the accuracy shows a great deviation between the models for these datasets. Figure 3 presents the scatter plots of the observed versus predicted chlorophyll values (top row) and the observed chlorophyll versus residuals (bottom row) of the five datasets. Good linear agreement can be observed on the observed vs. predicted chlorophyll scatter plots. The observed chlorophyll vs. residuals scatter plots show a random scattering around zero with a relative small variance.

#### D. Sensitivity maps for the predictive mean and variance

We illustrate the performance of the SA by computing sensitivity maps for the SeaWIFS, MODIS-Aqua and MERIS dataset. The sensitivity maps were computed by extracting a coastal area of Eastern USA. Then the SA of the GPR mean and variance for the measurements in this area were computed. We chose the  $k$ -nearest neighbours for each data points, performed the SA on the group of these data points through an iteration process, and picked the most sensitive band for the GPR mean and the least sensitive for the GPR variance. Finally, we used these bands to produce sensitive maps by spatial interpolation. We used natural neighbour interpolation [40] for spatial illustration of the SA. Although other interpolation methods, such as kriging, are also commonly used, natural neighbor interpolation has been showed to have

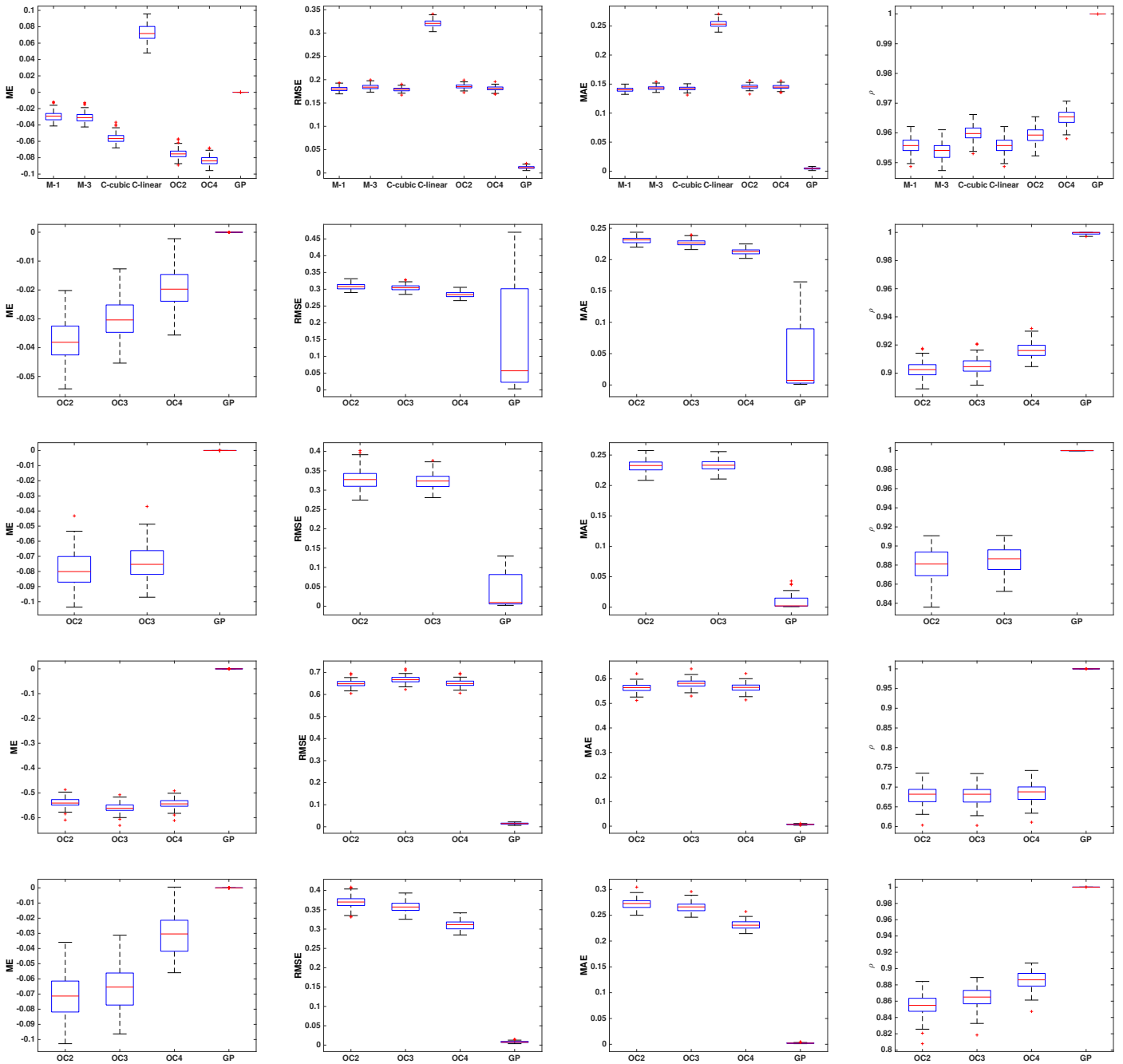


Fig. 4. Box-plots of the bootstrap model criteria for the SeaBAM (top row), SeaWIFS (second row), MODIS-Aqua (third row), MERIS synthetic (fourth row) and MERIS real (bottom row) datasets. The models for the SeaBAM dataset are indicated by M-1 and M-3 for Morel-1 and Morel-3, and C-linear and C-cubic for CalCOFI 2-band linear and CalCOFI 2-band cubic, respectively. The Ocean Color models are the OC2, OC3 and OC4 algorithms, and the GPR model using the most relevant bands is denoted by GP.

a good performance for this type of data as well [41], [42].

Results are shown in Figures 5-7. The left maps show the positions of the in-situ chlorophyll measurements (red dots), the second figures illustrate the interpolated measured chlorophyll values, while the third and right figures show the sensitivity maps for the GPR predictive mean and variance, respectively.

The sensitivity maps show that the SA of the GPR mean

assigns higher wavelengths to areas where the chlorophyll is present. Interestingly, it can be observed that there are areas with low chlorophyll content (second column) and corresponding higher wavelengths (third column) (in Fig. 6 middle part). This might indicate the presence of suspended particulate materials, which tend to result higher values in the reflectance spectra with increasing concentration [28]. Therefore, computing sensitivity maps in addition to estimated

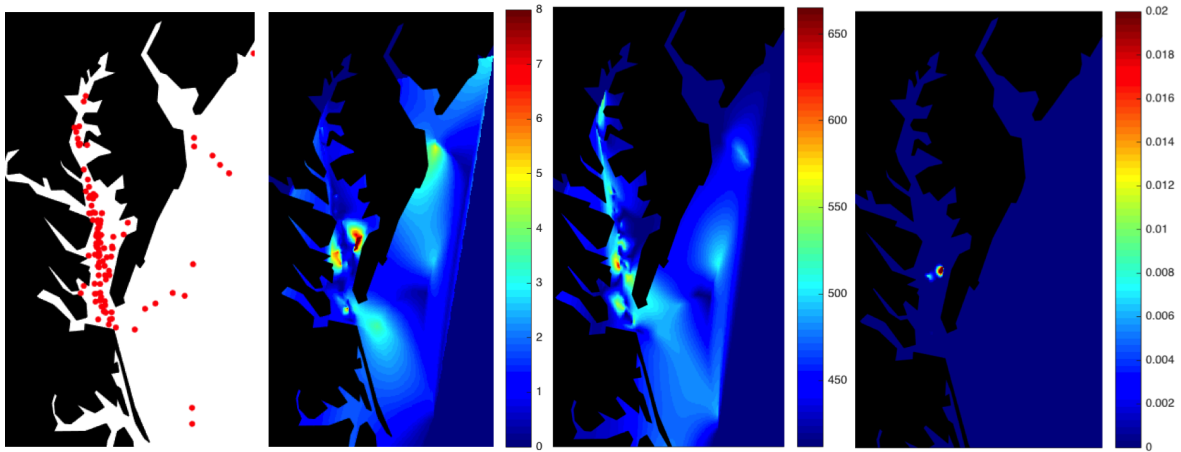


Fig. 5. Position of the in-situ chlorophyll measurements marked by red points (left), chlorophyll content map in  $\text{mgm}^{-3}$  (second), sensitivity map of the GPR mean (third) and variance (right) for the SeaWiFS dataset.

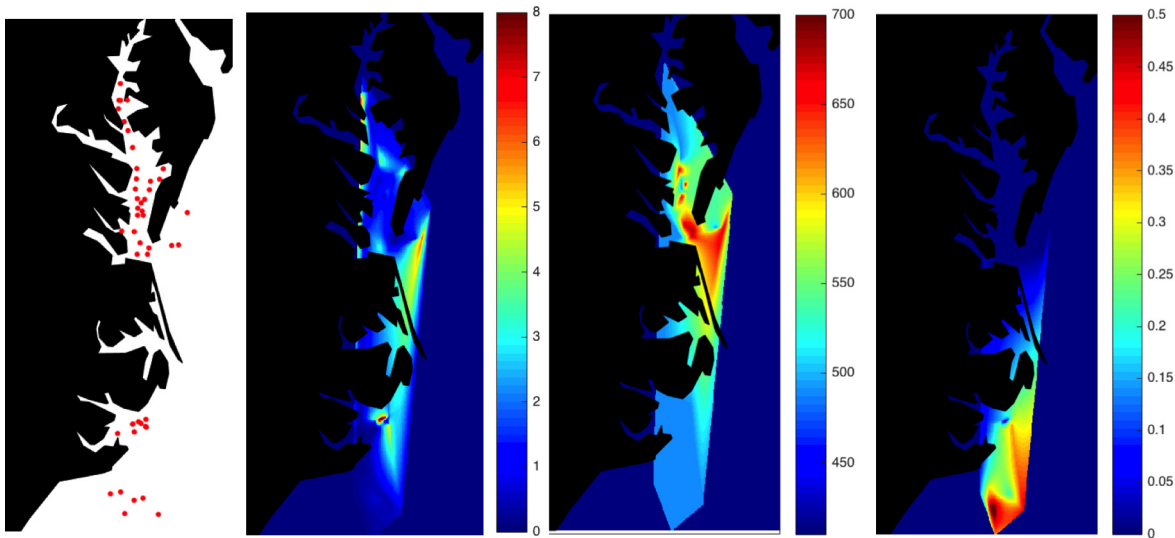


Fig. 6. Position of the in-situ chlorophyll measurements marked by red points (left), chlorophyll content map in  $\text{mgm}^{-3}$  (second), sensitivity map of the GPR mean (third) and variance (right) for the MODIS-Aqua dataset.

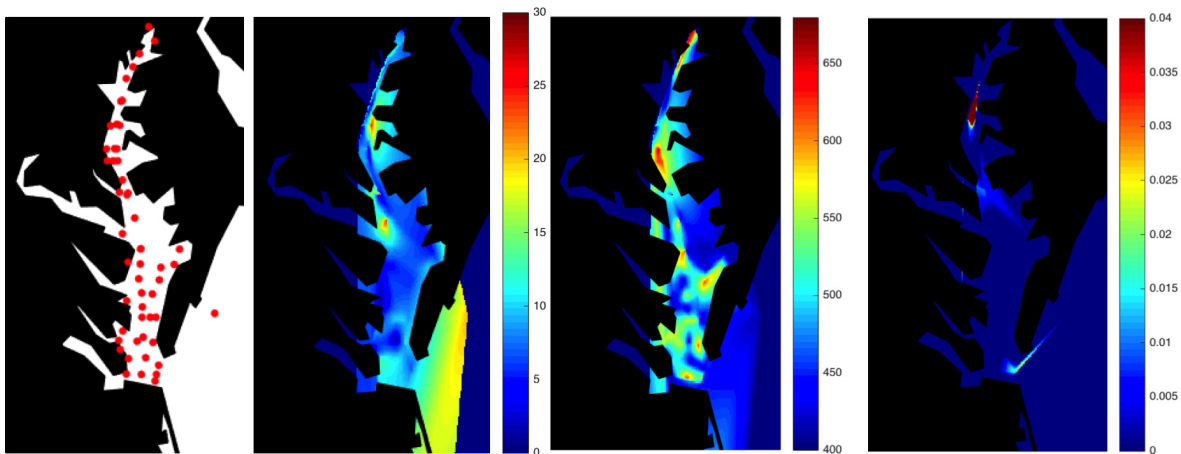


Fig. 7. Position of the in-situ chlorophyll measurements marked by red points (left), chlorophyll content map in  $\text{mgm}^{-3}$  (second), sensitivity map of the GPR mean (third) and variance (right) for the MERIS dataset.



TABLE III

MODEL COMPARISON OF THE TEST RESULTS FOR BIO-OPTICAL MODELS AND GPR FOR ALL DATASETS. THE COMPUTED MODEL MEASURES ARE THE MEAN VALUES OF 100 BOOTSTRAP-SAMPLES.

SeaBAM				
Model	ME	RMSE	MAE	$\rho$
Morel-1	-0.0289	0.18	0.1404	0.9558
Morel-3	-0.0309	0.1844	0.1432	0.954
CalCOFI 2-band cubic	-0.056	0.1791	0.1424	0.9598
CalCOFI 2-band linear	+0.0729	0.3209	0.2539	0.9558
Ocean chlorophyll 2, OC2	-0.075	0.1856	0.1456	0.9593
Ocean chlorophyll 4, OC4	-0.0835	0.1811	0.1451	0.9652
GPR (2, 4 and 5)	$<10^{-16}$	0.0117	0.0047	1.0000
SeaWIFS				
Ocean chlorophyll 2, OC2	-0.376	0.308	0.2312	0.9025
Ocean chlorophyll 3, OC3	-0.0297	0.3046	0.2269	0.9048
Ocean chlorophyll 4, OC4	-0.0194	0.2839	0.2129	0.9165
GPR (4, 5 and 6)	$<10^{-14}$	0.149	0.035	0.9994
MODIS-Aqua				
Ocean chlorophyll 2, OC2	-0.0788	0.3283	0.2319	0.8802
Ocean chlorophyll 3, OC3	-0.0742	0.3236	0.2328	0.885
GPR (4, 5 and 6)	$<10^{-15}$	0.0345	0.0078	0.9999
MERIS (synthetic)				
Ocean chlorophyll 2, OC2	-0.5397	0.6489	0.5634	0.6799
Ocean chlorophyll 3, OC3	-0.5606	0.667	0.5813	0.6795
Ocean chlorophyll 4, OC4	-0.5439	0.6506	0.5653	0.6862
GPR (5, 6, 7 and 8)	$<10^{-11}$	0.0144	0.0073	1.0000
MERIS (real)				
Ocean chlorophyll 2, OC2	-0.0719	0.3699	0.2715	0.8549
Ocean chlorophyll 3, OC3	-0.0668	0.3571	0.2654	0.8641
Ocean chlorophyll 4, OC4	-0.0315	0.3100	0.2311	0.8853
GPR (5, 6, 7 and 8)	$<10^{-15}$	0.0081	0.0022	1.0000

TABLE IV

SUMMARY OF THE STATISTICAL ANALYSIS (ONE-WAY-ANOVA) IN THE 5 DATASETS.

Database	Bias		Accuracy	
	F-value	p-value	F-value	p-value
SeaBAM	67.76	$<0.001$	156.45	$<0.001$
SeaWIFS	2.38	$<0.1$	18.52	$<0.001$
MODIS-Aqua	5.12	$<0.01$	31.94	$<0.001$
MERIS (synthetic)	3243.61	0	4441.51	0
MERIS (real)	0.99	$<0.4$	202.93	$<0.001$

chlorophyll content maps might open the possibility to retrieve further information about the constituents of the oceans through their optical properties. Looking at the chlorophyll content maps together with the SA of the GPR mean maps might give an intuition about the connection between the amount of chlorophyll and the most important wavelengths. The SA of the GPR mean maps represent the geographical distribution of the most important wavelengths, while the SA of the GPR variance maps show how the distribution of the spectral spacing varies in the same area.

#### E. Verifying the results of the SA of the GPR mean on the SeaWIFS dataset

In order to validate the results on a global scale we present global chlorophyll content maps (Fig. 8) for the SeaWIFS dataset. The global validation maps for the MODIS-Aqua and MERIS (real) datasets and the local validation maps for the SeaWIFS, MODIS-Aqua and MERIS (real) datasets can be found under Appendix. We use the same procedure for spatial interpolation as in Section IV-D. Figure 8 shows

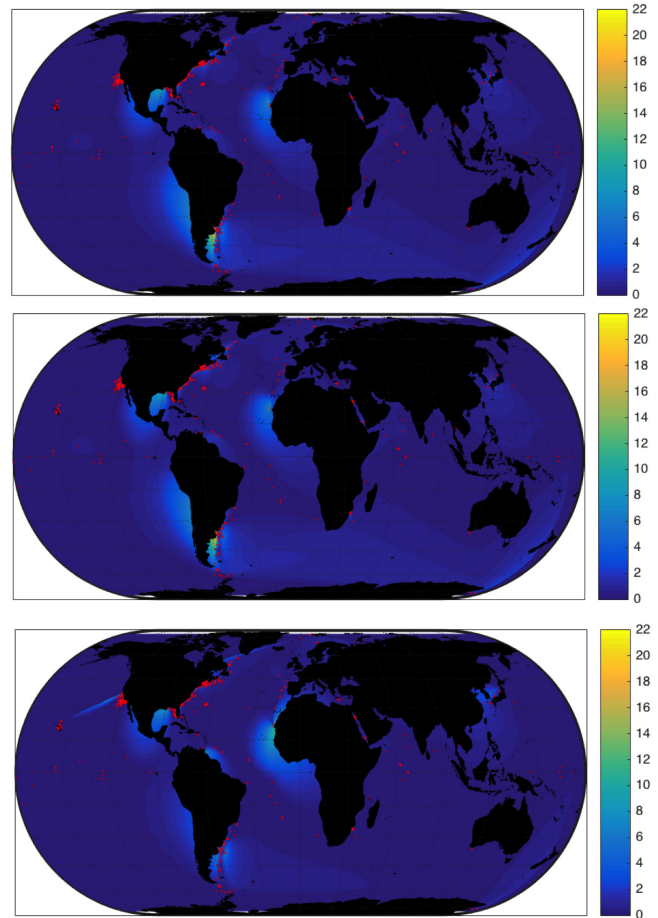


Fig. 8. Global in-situ chlorophyll content map (top), predicted chlorophyll content map for the GPR with the most relevant bands (middle) and predicted chlorophyll content maps for the OC4 parametric model (bottom) for the SeaWIFS dataset. The total number of samples is 1465 and the unit of the chlorophyll content is  $\text{mgm}^{-3}$ . The red dots indicate the position of the measurements (interpolation points).

the results of the in-situ chlorophyll content map (top), the predicted chlorophyll content map using GPR with the most relevant bands (middle) and predicted chlorophyll content map applying a parametric model (bottom). We chose the parametric model with the lowest RMSE-value (Table III). It can be observed that the chlorophyll content map of the GPR with the most relevant bands looks almost identical as the true chlorophyll content map, while the parametric model seems to overestimate the predicted values. Thus, the SA of GPR can be used to determine feature relevance and selection.

Note, the aim of presenting validation maps is to visualize the strength of the SA rather than to produce accurate global chlorophyll content maps, which would have been challenging for these datasets due to the number of samples and the wide time frame the chlorophyll samples were taken at. Our focus was to illustrate that using the SA of the GPR mean function for identifying the most important spectral bands in the regression process and using only these bands as inputs for the GPR for chlorophyll content estimation, can compete with the frequently applied parametric models. Therefore, the

methodology opens the possibility for practical application purposes.

## V. CONCLUSIONS AND FURTHER WORK

We derived empirical estimates for the sensitivity of the GPR predictive mean and variance functions. After applying the SA to a controlled example, we illustrated the performance of the method on five global datasets. We found that the SA of the GPR mean assigned the highest sensitivity to bands in the range between 510 nm and 560 nm. This is in good correspondence with the reflectance spectra of the chlorophyll. Bands positioned on higher wavelengths also got ranked as relevant bands for chlorophyll content prediction. This might indicate preference for bands associated to chlorophyll fluorescence. Being able to monitor chlorophyll fluorescence allows the possibility to detect changes in photosynthesis, thus to monitor the health status of oceans. In addition, the detection of chlorophyll fluorescence might be a useful tool, when other substances beside chlorophyll are also present. This might be specially the case for coastal waters. Besides the sensitivity of the GPR mean we also derived the SA of the GPR variance for the five global datasets, and uncovered the relevance of the (spectral) sampling of the bands. Knowing the spectral distribution of the inputs might allow the deeper understanding of the underlying biophysics, as well as the design of further sensors. Furthermore, we compared the performance of the GPR using only the most sensitive bands for regression with parametric models. The computed measures revealed the SA could identify the most important features, and thus using only these features as inputs to the GPR could outperform other models.

Finally, we presented the SA of the GPR on sensitivity maps for a given region. These spatially-explicit maps highlight the usefulness of the GPR sensitivity analysis to study the distribution of the most relevant wavelengths and to reveal the optimality of the spectral sampling density. In addition, we compared the SA of the GPR mean function for chlorophyll content prediction on a global scale by computing global chlorophyll content maps for the actual chlorophyll content, the GPR model with the most sensitive bands and with the parametric model of the lowest computed RMSE value. These global maps confirmed that using the bands which were assigned to have the greatest relevance to perform GPR shows good correspondence and spatial comparability. For future work, we plan to produce sensitivity maps on a time scale as well, with the aim of detecting changes in oceanic chlorophyll fluorescence. It does not escape our notice that the methodology can be used for global sensitivity analysis of radiative transfer models, as well as to further evaluate current GPR emulators.

## APPENDIX

### GLOBAL AND LOCAL VALIDATION MAPS

Figure 9 shows the global validation maps for the MODIS-Aqua and MERIS (real) datasets. In the case of the MODIS-Aqua dataset (left column), it seems that both the GPR and the parametric model results overestimates along the Western coast of Europe and Africa and underestimates around the Northern part the Indian ocean. The predicted chlorophyll contents show good correspondence with the true values along

the coasts of America for both models. Comparing the GPR and the parametric model with the in-situ chlorophyll content map for the MERIS real dataset (right column) reveals an overall overestimation and underestimation in the predicted chlorophyll contents, respectively. However, the distribution of the chlorophyll seems to follow the same pattern as the true chlorophyll content map for both cases. In general, it can be concluded that the predicted chlorophyll contents are in good correspondence with the true values. Even though there might occur over- and underestimates in the predicted values, using the most sensitive bands to perform GPR for chlorophyll content prediction on a global scale shows just as good performance as the parametric model (with the lowest RMSE value). Therefore, the SA of GPR can be used to determine feature relevance and selection.

The validation maps were also implemented for the same area as in Sec. IV-D. Figure 10 shows the results.

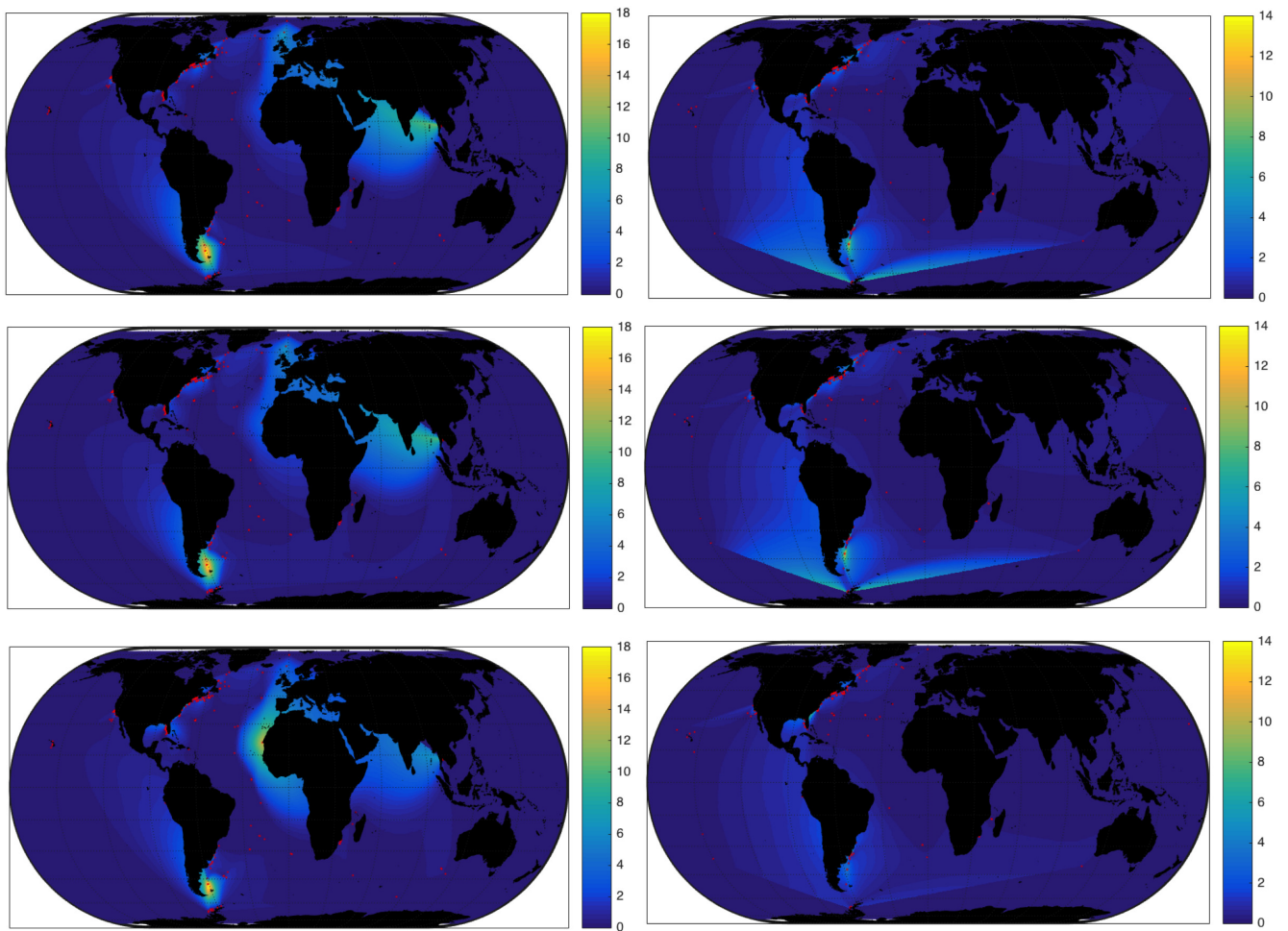


Fig. 9. Global in-situ chlorophyll content maps (top row), predicted chlorophyll content map for the GPR with the most relevant bands (middle row) and predicted chlorophyll content maps for the OC3 (MODIS-Aqua) and OC4 (MERIS real) parametric model (bottom row) for the MODIS-Aqua (left column) and MERIS real (right column) datasets. The total number of samples is 579 for the MODIS-Aqua dataset and 567 for MERIS real dataset. The unit of the chlorophyll content is  $\text{mgm}^{-3}$ . The red dots indicate the position of the measurements (interpolation points).

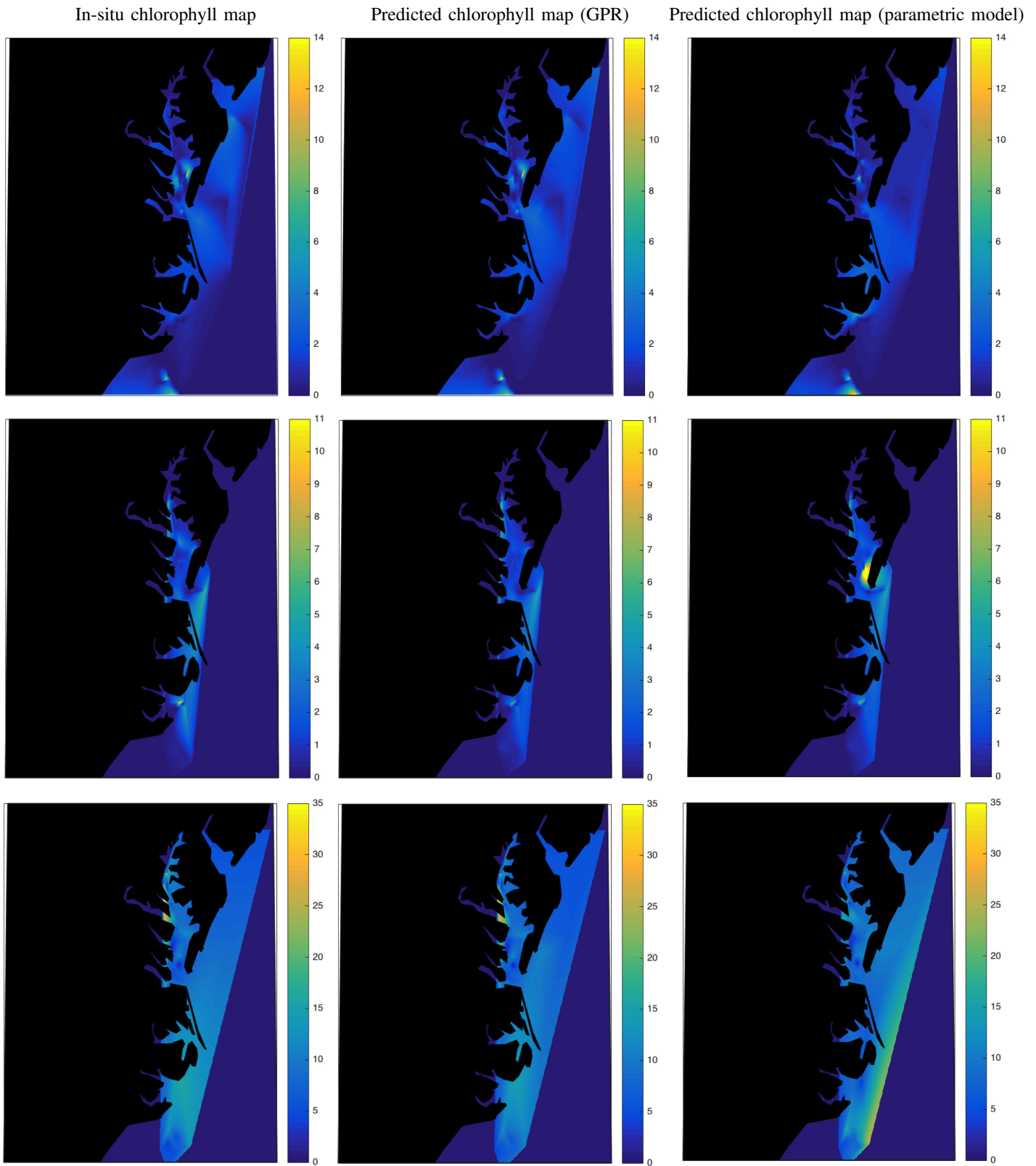


Fig. 10. Local in-situ chlorophyll content maps (left column), predicted chlorophyll content map for the GPR with the most relevant bands (middle column) and predicted chlorophyll content maps for the best parametric model (right column) for the SeaWiFS (top row), MODIS-Aqua (middle row) and MERIS real (bottom row) datasets. The unit of the chlorophyll content is  $\text{mgm}^{-3}$ .



## ACKNOWLEDGMENT

We would like to thank Dr. Georg Elvebakk at the Department of Mathematics and Statistics (UiT the Arctic University of Norway) and Dr. A. Ruescas-Orient at the Image Processing Laboratory (Universitat de València) for their useful comments and discussion on the results.

## REFERENCES

- [1] D. Blondeau-Patissier, J. F. Gower, A. G. Dekker, S. R. Phinn, and V. E. Brando, "A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans," *Progress in Oceanography*, vol. 123, pp. 123 – 144, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0079661114000020>
- [2] Govindjee, *Bioenergetics of Photosynthesis*. Academic Press, 1975.
- [3] P. Falkowski and D. A. Kiefer, "Chlorophyll a fluorescence in phytoplankton: relationship to photosynthesis and biomass," *Journal of Plankton Research*, vol. 7, pp. 715–731, 1985.
- [4] S. M. McKibben, P. G. Strutton, D. G. Foley, T. D. Peterson, and A. E. White, "Satellite-based detection and monitoring of phytoplankton blooms along the Oregon coast," *Journal of Geophysical Research*, vol. 117, 2012.
- [5] A. Gitelson, M. Mayo, Y. Z. Yacobi, A. Parparov, and T. Berman, "The use of high-spectral-resolution radiometer data for detection of low chlorophyll concentrations in Lake Kinneret," *Journal of Plankton Research*, vol. 16, pp. 993–1002, 1994.
- [6] J. Fischer and U. Kronfeld, "Sun-stimulated chlorophyll fluorescence I: Influence of oceanic properties," *International Journal of Remote Sensing*, vol. 11, no. 12, pp. 2125–2147, 1990. [Online]. Available: <http://dx.doi.org/10.1080/01431169008955166>
- [7] M. J. Behrenfeld, T. K. Westberry, E. Boss, R. T. O'Malley, and et al., "Satellite-Detected Fluorescence Reveals Global Physiology of Ocean Phytoplankton," *Biogeosciences*, vol. 6 (5), pp. 779–794, 2009.
- [8] C. S. Reynolds, *The Ecology of Phytoplankton*. Cambridge University Press, 2006.
- [9] K. R. Arrigo, D. H. Robinson, D. L. Worthen, R. B. Dunbar, G. R. DiTullio, M. VanWoert, and M. P. Lizotte, "Phytoplankton Community Structure and the Drawdown of Nutrients and CO<sub>2</sub> in the Southern Ocean," *Science*, vol. 283, no. 5400, pp. 365–367, 1999. [Online]. Available: <http://science.sciencemag.org/content/283/5400/365>
- [10] M. Hein and K. Sand-Jensen, "CO<sub>2</sub> increases oceanic primary production," *Nature*, vol. 388, pp. 526–527, 1997.
- [11] M. Hofmann, B. Worm, S. Rahmstorf, and H. J. Schellnhuber, "Declining ocean chlorophyll under unabated anthropogenic CO<sub>2</sub> emissions," *Environmental Research Letters*, vol. 6, no. 3, pp. 034–035, 2011. [Online]. Available: <http://stacks.iop.org/1748-9326/6/i=3/a=034035>
- [12] H. R. Gordon, O. B. Brown, R. H. Evans, J. W. Brown, and R. C. Smith, "A semianalytic radiance model of ocean color," *Journal of Geophysical Research*, vol. 93, pp. 10909–10924, 2011.
- [13] S. B. Hooker, E. R. Firestone, J. E. O'Reilly, M. C. O'Brien, D. A. Siegel, D. Toole, D. Menzies, R. C. Smith, J. L. Mueller, B. G. Mitchell, M. Kahru, F. P. Chavez, P. Strutton, G. E. Cota, S. B. Hooker, C. R. McClain, K. L. Carder, L. Harding, A. Magnuson, D. Phinney, G. E. Moore, J. Aiken, K. R. Arrigo, R. Letelier, and M. Culver, "Volume 11, SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3," 2000.
- [14] P. Cipollini, G. Corsini, M. Diani, and R. Grass, "Retrieval of sea water optically active parameters from hyperspectral data by means of generalized radial basis function neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, pp. 1508–1524, 2001.
- [15] H. Zhan, P. Shi, and C. Chen, "Retrieval of oceanic chlorophyll concentration using support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41 (12), pp. 2947–2951, 2003.
- [16] E. J. Kwiatkowska and G. S. Fargion, "Application of machine-learning techniques toward the creation of a consistent and calibrated global chlorophyll concentration baseline dataset using remotely sensed ocean color data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41 (12), pp. 2844 – 2860, December 2003.
- [17] G. Camps-Valls, J. Muñoz-Marí, K. R. L. Gómez-Chova, and J. Calpe-Maravilla, "Biophysical parameter estimation with a semisupervised support vector machine," *IEEE Geoscience and Remote Sensing Letters*, vol. 6 (2), pp. 248 – 252, 2009.
- [18] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, J. Vila-Francés, J. Amorós-López, and J. Calpe-Maravilla, "Retrieval of oceanic chlorophyll concentration with relevance vector machines," *Remote Sensing of Environment*, vol. 105(1), pp. 23 – 33, 2006.
- [19] L. Pasolli, F. Melgani, and E. Blanzieri, "Gaussian process regression for estimating chlorophyll concentration in subsurface waters from remote sensing data," *IEEE Geoscience and Remote Sensing Letters*, vol. 7(3), pp. 464 – 468, July 2010.
- [20] Y. Bazi, N. Alajlan, and F. Melgani, "Improved Estimation of Water Chlorophyll Concentration With Semisupervised Gaussian Process Regression," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 7, pp. 2733–2743, 2012.
- [21] Y. Bazi, N. Alajlan, F. Melgani, H. AlHichri, and R. R. Yager, "Robust Estimation of Water Chlorophyll Concentrations With Gaussian Process Regression and IOWA Aggregation Operators," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 7, pp. 3019–3028, 2014.
- [22] C. E. Rasmussen and C. K. I. Williams, *Gaussian Process for Machine Learning*. Cambridge MA: MIT Press, 2006.
- [23] C. K. I. Williams and C. E. Rasmussen, "Gaussian processes for regression," in *Advances in Neural Information Processing Systems*, vol. 8, pp. 514 – 520, 1996.
- [24] K. Blix, G. Camps-Valls, and R. Jenssen, "Sensitivity analysis of gaussian processes for oceanic chlorophyll prediction," in *2015 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2015, Milan, Italy, July 26-31, 2015*, 2015, pp. 996–999. [Online]. Available: <http://dx.doi.org/10.1109/IGARSS.2015.7325936>
- [25] J. Verrelst, J. Muñoz, L. Alonso, J. P. Rivera, G. Camps-Valls, and J. Moreno, "Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for sentinel-2 and -3," *Remote Sensing of Environment*, 2012.
- [26] J. Verrelst, L. Alonso, G. Camps-Valls, J. Delegido, and J. Moreno, "Retrieval of Vegetation Biophysical Parameters Using Gaussian Process Techniques," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 5 PART 2, pp. 1832–1843, 2012. [Online]. Available: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84860332507&partnerID=40&md5=8f89c24c1927827bf249a795b35098fc>
- [27] P. M. Rasmussen, K. H. Madsen, T. E. Lund, and L. K. Hansen, "Visualization of nonlinear kernel models in neuroimaging by sensitivity maps," *NeuroImage*, vol. 55, no. 3, pp. 1120 – 1131, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1053811910016198>
- [28] I. S. Robinson, *Measuring the Oceans from Space: The principles and methods of satellite oceanography*. Praxis Publishing Ltd, 2004.
- [29] P. J. Werdell and S. W. Bailey, "The seawifs bio-optical archive and storage system (seabass): Current architecture and implementation," *NASA Tech. Memo. 2002-211617*, G. S. Fargion and C. R. McClain, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, p. 45, 2002.
- [30] C. Hu, Z. Lee, and B. Franz, "Chlorophyll a algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference," *Journal of Geophysical Research*, vol. 117, p. C01011, 2012.
- [31] J. E. O'Reilly, S. Maritirena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. McClain, "Ocean color chlorophyll algorithms for SeaWiFS," *Journal of Geophysical Research*, vol. 103, pp. 24937–24953, 1998.
- [32] W. E. Esaiás, M. R. Abbott, I. Barton, O. B. Brown, J. W. Campbell, K. L. Carder, D. K. Clark, R. H. Evans, F. E. Hoge, H. R. Gordon, W. M. Balch, R. Letelier, and P. J. Minnett, "An overview of MODIS capabilities for ocean science observations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, pp. 1250–1265, 1998.
- [33] J. F. R. Gower, L. Brown, and G. A. Borstad, "Observation of chlorophyll fluorescence in west coast waters of Canada using MODIS satellite sensor," *Can. J. Remote Sensing*, vol. 30, no. 1, pp. 17–25, 2004.
- [34] C. Hu, F. E. Muller-Karger, C. J. Taylor, K. L. Carder, C. Kelble, E. Johns, and C. A. Heil, "Red tide detection and tracing using MODIS fluorescence data: A regional example in SW Florida coastal waters," *Remote Sensing of Environment*, vol. 97, pp. 311–321, 2005.
- [35] J. F. R. Gower, R. Doerffer, and G. A. Borstad, "Interpretation of the 685nm peak in water-leaving radiance spectra in terms of fluorescence, absorption and scattering, and its observation by MERIS," *International Journal of Remote Sensing*, vol. 20, no. 9, pp. 1771–1786, 1999. [Online]. Available: <http://dx.doi.org/10.1080/014311699212470>
- [36] J. Gower, S. King, W. Y. G. Borstad, and L. Brown, "Use of 709 nm band of MERIS to detect intense plankton blooms and other conditions in coastal waters," *ESA, MERIS User Workshop*, 2003.
- [37] X.-G. Xing, D.-Z. Zhao, Y.-G. Liu, J.-H. Yang, P. Xiu, and L. Wang, "An Overview of Remote Sensing of Chlorophyll Fluorescence," *Ocean Science Journal*, vol. 42, no. 1, pp. 19–59, 2007.
- [38] P. J. Werdell and S. W. Bailey, "An improved bio-optical data set for ocean color algorithm development and satellite data product validation," *Remote Sensing of Environment*, vol. 98, pp. 122–140, 2005.
- [39] S. Salcedo-Sanz, C. Casanova-Mateo, J. Muñoz-Marí, and G. Camps-Valls, "Prediction of daily global solar irradiation using temporal gaussian processes," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 11, pp. 1936–1940, Nov 2014.
- [40] J.-D. Boissonnat and F. Cazals, "Natural neighbor coordinates of points on a surface," *Computational Geometry Theory and Applications*, vol. 19, pp. 155 – 173, 2001.
- [41] A. K. Mishra and N. Garg, "Analysis of Trophic State Index of Nainital Lake from Landsat-7 ETM Data," *Journal of the Indian Society of Remote Sensing*, vol. 39, no. 4, pp. 463–471, 2011. [Online]. Available: <http://dx.doi.org/10.1007/s12524-011-0105-3>

- [42] M. Rowe, E. Anderson, J. Wang, and H. Vanderploeg, "Modeling the effect of invasive quagga mussels on the spring phytoplankton bloom in Lake Michigan," *Journal of Great Lakes Research*, vol. 41, Supplement 3, pp. 49 – 65, 2015, complex interactions in Lake Michigan's rapidly changing ecosystem. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0380133015000040>



**Katalin Blix** received a B.S. degree in Geosciences (2010) from the Sogn og Fjordane University College and a Civil engineer/ M.S. degree in Technology with specialization to applied physics and mathematics (2014) from UiT the Arctic University of Norway. She is currently pursuing the Ph.D. degree in remote sensing at UiT the Arctic University of Norway.



**Gustau Camps-Valls** received a B.Sc. degree in Physics (1996), in Electronics Engineering (1998), and a Ph.D. degree in Physics (2002) all from the Universitat de València. He is currently an Associate Professor (hab. Full professor) in the Department of Electronics Engineering. His is a research coordinator in the Image and Signal Processing (ISP) group. He has been Visiting Researcher at the Remote Sensing Laboratory (Univ. Trento, Italy) in 2002, the Max Planck Institute for Biological Cybernetics (Tübingen, Germany) in 2009, and as Invited Pro-

fessor at the Laboratory of Geographic Information Systems of the École Polytechnique Fédérale de Lausanne (Lausanne, Switzerland) in 2013. He is interested in the development of machine learning algorithms for geoscience and remote sensing data analysis. He is an author of 120 journal papers, more than 150 conference papers, 20 international book chapters, and editor of the books "Kernel methods in bioengineering, signal and image processing" (IGI, 2007), "Kernel methods for remote sensing data analysis" (Wiley & Sons, 2009), and "Remote Sensing Image Processing" (MC, 2011). He's a co-editor of the forthcoming book "Digital Signal Processing with Kernel Methods" (Wiley & sons, 2015). He holds a Hirsch's index  $h=42$  (see Google Scholar page), entered the ISI list of Highly Cited Researchers in 2011, and Thomson Reuters ScienceWatch identified one of his papers on kernel-based analysis of hyperspectral images as a Fast Moving Front research. In 2015, he obtained the prestigious European Research Council (ERC) consolidator grant on Statistical learning for Earth observation data analysis. He is a referee and Program Committee member of many international journals and conferences. Since 2007 he is member of the Data Fusion technical committee of the IEEE GRSS, and since 2009 of the Machine Learning for Signal Processing Technical Committee of the IEEE SPS. He is member of the MTG-IRS Science Team (MIST) of EUMETSAT. He is Associate Editor of the IEEE Transactions on Signal Processing, IEEE Signal Processing Letters, IEEE Geoscience and Remote Sensing Letters, and invited guest editor for IEEE Journal of Selected Topics in Signal Processing (2012) and IEEE Geoscience and Remote Sensing Magazine (2015).



**Robert Jenssen** is an Associate Professor with the Department of Physics and Technology, at UiT the Arctic University of Norway. He directs the Machine Learning @ UiT research group: <http://site.uit.no/ml>. The group's focus is on innovating information theoretic learning, kernel machines, and deep learning research, an activity that is funded e.g. over the Norwegian Research Council's prestigious FRIPRO program, grant no. 239844 (Next Generation Learning Machines). Jenssen received the Honorable Mention for the 2003 Pattern Recognition Journal Best Paper

Award, the 2005 IEEE ICASSP Outstanding Student Paper Award, and the 2007 UiT Young Investigator Award. His paper, "Kernel Entropy Component Analysis", was the featured paper of the May 2010 issue of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. In addition, his paper, "Kernel Entropy Component Analysis for Remote Sensing Image Clustering" with L. Gomez-Chova and G. Camps-Valls received the IEEE Geoscience and Remote Sensing Society Letters Prize Paper Award in 2013. Jenssen is a research professor with the Norwegian Computing Center in Oslo, and a senior researcher at the Norwegian Center on E-Health Research. He was Guest Researcher with the Technical University of Denmark, Kongens Lyngby, Denmark, from 2012 to 2013, Technical University of Berlin, Berlin, Germany, from 2008 to 2009, and University of Florida, Gainesville, FL, USA, from 2002 to 2003.