

Evaluation of feature ranking and regression methods for oceanic chlorophyll-a estimation

Katalin Blix and Torbjørn Eltoft

Abstract—This paper evaluates two alternative regression techniques for oceanic Chlorophyll-a (Chl-a) content estimation. One of the investigated methodologies is the recently introduced Gaussian Process Regression (GPR) model. We explore two feature ranking methods derived for the GPR model, namely Sensitivity Analysis (SA) and Automatic Relevance Determination (ARD). We also investigate a second regression method, the Partial Least Squares Regression (PLSR) for oceanic Chl-a content estimation. Feature relevance in the PLSR model can be accessed through the Variable Importance in Projection (VIP) feature ranking algorithm. The paper thus analyses three feature ranking models, SA, ARD and VIP, which are all derived from different fundamental principles, and uses the ranked features as inputs to the GPR and PLSR to assess regression strengths. We compare the regression performances using some common performance measures, and show how the feature ranking methods can be used to find the lowest number of features to estimate oceanic Chl-a content by using the GPR and PLSR models, while still producing comparable performance to the state-of-art algorithms. We evaluate the models on a global MERIS matchup dataset. Our results show that the GPR model has the best regression performance for most of the input feature sets we used, and our conclusion is this model can favorably be used for Chl-a content retrieval, already with two features, ranked by the either the SA or ARD methods.

I. INTRODUCTION

Continuous monitoring of the occurrence and distribution of phytoplankton has high ecological [1] and economical importance (<http://oceancolor.gsfc.nasa.gov/>). Phytoplankton content can be indirectly estimated from the Chlorophyll-a (Chl-a) concentration. Similar to terrestrial plants, phytoplankton also use photosynthesis in order to live and grow. Chl-a is the key molecule for capturing light, which is the driving of photosynthesis [2]. Hence, Chl-a content is used as an indicator for several biophysical processes, which can be used for various applications.

Phytoplankton removes CO_2 from the atmosphere, through the photosynthetic process [3], and therefore the monitoring of phytoplankton via Chl-a has important relevance in climate studies [4], [5], [6].

Chl-a is also used to determine water-quality. Eutrophication of coastal waters and lakes has been increasing in the past

decades, leading to degraded water-quality [7], [8]. A symptom of degraded water quality is an increase of algae biomass, which may be measured by the concentration of Chl-a. Hence, estimates of aquatic Chl-a concentration may also be used to derive information about water-quality in coastal waters.

Monitoring can be achieved by optical sensors onboard satellites. It is often required to have high spatial resolution in order to monitor water-quality on a finer scale. However, optical remote sensing has its limitations with regard to spectral-spatial resolution [9], [10]. In order to achieve high spatial resolution, the number of spectral bands are limited. Therefore, it is critical to know the number, position and width of the bands required to retrieve Chl-a for the given aquatic condition, without losing accuracy in the estimation.

Satellite derived Chl-a concentration is usually based on globally tuned parametric bio-optical models, such as NASA's Ocean Color (OC) models [11], [12], [13], [14], [15]. In the remainder we refer to these models as the OC algorithms. The OC algorithms are polynomial regression models, which are trained by relating *in situ* Chl-a content to Remote Sensing Reflectance $R_{rs}(\lambda)$ (sr^{-1}), measured at predefined wavelengths through a so called band ratio, R . There is a variety of Chl-a content retrieval models based on band ratios [16]. In this work, we will confine ourselves to band ratios used in NASA's OC algorithms. This band ratio is calculated at the spectral position of the Chl-a absorption peak [17], and given by $R_{rs}(\lambda_{blue})/R_{rs}(\lambda_{green})$ [13]. Even though these algorithms are fast, simple and reflect the biophysical properties of aquatic Chl-a, they have certain weaknesses. This is due to the fact that the absorption spectrum varies with the amount of Chl-a concentration in the water, and it is also affected by the amount of other surfactant materials in the ocean waters near to the surface [17]. Furthermore, the coefficients of the polynomial in the OC regression models are determined by using a global training dataset. In order to allow a model to adapt to local variations, the coefficients need to be adjusted by extending the training data with measurements from the region of interest. Several studies have shown that algorithms based on band-ratios result in erroneous retrieval of Chl-a content [16] due to the regional variations of the optical properties of ocean waters. In order to overcome these difficulties in the retrieval of Chl-a content from remotely sensed data with OC models, it is important to use the correct combination of spectral bands in the computation of the band ratios.

Several studies have proposed alternative regression models for increasing the accuracy, reliability and effectiveness in the monitoring of oceanic Chl-a content from optical remote sensing data. (For further details we refer to [16] for a review

Manuscript received January 18, 2018;

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

TE is with the Centre for Integrated Remote Sensing and Forecasting for Arctic Operations (CIRFA) and the Department of Physics and Technology, University of Tromsø - The Arctic University of Norway. E-mail: torbjorn.eltoft@uit.no.

of these algorithms.) Machine learning regression methods are known to have strong regression capabilities, and several such algorithms have been studied for Chl-a content estimation. The investigated models include neural networks [18]–[20], support vector regression [21]–[23], relevance vector regression [24], and the lately introduced Gaussian Process Regression (GPR) algorithm [25].

The objectives of this work are as follows. Firstly, we study the relevance of features (i.e. spectral bands and/or band ratios), and regression performances of two regression methods, namely the Gaussian Process Regression (GPR) [26], [27] and the Partial Least Squares Regression (PLSR) [28] models, when applied for Chl-a content estimation from satellite-based optical measurements. Both these two models are known to have good regression performance, and both have model-tailored methods for assessing the relevance of input feature.

The GPR model uses a Bayesian approach to learn the non-linear functional relationship between the input feature vectors and the output Chl-a measurements, and feature ranking¹ can be conducted using the Automatic Relevance Determination (ARD) and Sensitivity Analysis (SA). PLSR is a well-known linear regression model, which uses a so-called latent variable space to relate the input features to the Chl-a measurement. In PLSR, feature relevance is analyzed using a ranking method called Variable Importance in Projection (VIP). Secondly, using a set of regression performance measures, we evaluated the regression strength of individual spectral features and sets of spectral features, and used the performance tests to propose a lowest number of spectral bands and/or features needed to estimate Chl-a content without any significant loss of accuracy compared to the state-of-the-art Ocean Colour (OC) algorithms. Finally, we include an assessment of the uncertainty level of some Chl-a estimates.

The GPR model differs from other machine learning and parametric methods in its underlying fundamental principles. Instead of proposing a function to relate Rrs to Chl-a, the GPR model learns the function by using a Bayesian approach, which has an analytic closed form solution.

The GPR model has been shown to perform better than other machine learning methods [29] and parametric models [30] in terms of accuracy and speed for the retrieval of biophysical parameters. In addition to the estimated Chl-a content, the GPR model is able to output the certainty level of the estimates.

The relative relevance of the features being used in the regression process is not directly accessible in GPR, since it is a non-linear kernel method. Feature relevance of Gaussian Processes (GPs) in land Chl-a content estimation was proposed computed by the so-called Automatic Relevance Determination (ARD) method in [30], [31]. Another method, the Sensitivity Analysis (SA) of GPs was introduced in [32] for oceanic Chl-a content estimation.

PLSR is an iterative statistical model, which has several advantageous properties. It can reduce co-linearity and noise

in the dataset, and it can provide multidimensional outputs. Feature relevance can be accessed through a measure denoted the Variable Importance in Projection (VIP). Lately, another method for band selection in PLSR (and Random Forest and Support Vector Machine regression) was proposed in [33], the so called ensemble approach. This study was conducted for leaf Chl-a content estimation. PLSR has been widely used in chemometrics [34], [28], and in several fields where there are a large amount of control variables with corresponding multi-dimensional outputs, for example in controlling and monitoring industrial processes [35]. The PLSR model has also been successfully applied for Chl-a content estimation in optically challenging oceanic waters [36].

In this paper, we first demonstrate feature ranking by the ARD, SA and VIP methods on two simulated datasets; a simple low-dimensional dataset, and a more complicated test example, with a very high-dimensional feature space. The purpose of these controlled experiments is to give the readers some confidence in the applied methods.

Then we use a global Chl-a validated SeaBASS dataset [37], [38] to train the regression models and to evaluate the feature ranking methods. We conduct a performance study of the regression models discussed above with respect to estimation of Chl-a based on a MERIS (MEdium Resolution Imaging Spectrometer) dataset, and we compare feature ranking by SA, ARD and VIP for GPR, and PLSR. Finally, we demonstrate how uncertainty can be accessed for the proposed models. Note, we have performed the same study for two additional global datasets for the SeaWiFS (Sea-Viewing Wide Field-of-View Sensor) and MODIS-Aqua (MODerate-resolution Imaging Spectroradiometer) sensors. These results are in correspondence with the results for the MERIS dataset, and presented in Appendix.

The remainder of this paper is organized as follows. Section II reviews the GPR and PLSR models and the associated feature ranking methods. Section III illustrates the concept of the feature ranking methods on two simulated examples. Section IV details the experimental setup of this study. Section V evaluates and compares the performance of the feature ranking methods and regression models. Finally, Sec. VII concludes this paper and outlines future work.

II. FEATURE RANKING METHODS FOR REGRESSION

A. Gaussian Process Regression (GPR)

Here, we apply regression in the context of estimating oceanic Chl-a contents (outputs) from Rrs values (inputs) by fitting a flexible GPR model to the training data. This training dataset consists of *in situ* Chl-a contents and corresponding Rrs values, measured in mgm^{-3} and sr^{-1} , respectively. Furthermore denote Chl-a by $\{y_n\}_{n=1}^N$ and Rrs by $\{\mathbf{x}_n \in \mathbb{R}^D\}_{n=1}^N$, where $n = 1, \dots, N$ is the number of measurements, and $d = 1, \dots, D$ is the number of spectral bands. The GPR model assumes that the observed Chl-a content is a function (also called a latent function) of the Rrs values, and the latent function values or outputs follow a multivariate joint Gaussian distribution, if $(f(\mathbf{x}_1), \dots, f(\mathbf{x}_N)) \sim \mathcal{N}(\mathbf{0}, \mathbf{K})$, with zero mean and covariance matrix \mathbf{K} . The observed outputs

¹Feature ranking methods have refer to methods, which assign relative relevance to the input features.

are usually contaminated by noise ε_n , thus $y_n = f(\mathbf{x}_n) + \varepsilon_n$ for $n = 1, \dots, N$. The noise terms are assumed to be additive, independently, identically Gaussian distributed with zero mean and constant variance, i. e. $\varepsilon_n \sim \mathcal{N}(0, \sigma^2)$.

Consider now a new input Rrs data, \mathbf{x}_* , where the goal is to estimate the corresponding output Chl-a content, y_* . Then the Gaussian Process \mathcal{GP} defines a joint prior distribution of the available Chl-a observations $\mathbf{y} \equiv \{y_n\}_{n=1}^N$ and the unseen y_* . This can be written by

$$\begin{bmatrix} \mathbf{y} \\ y_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} + \sigma^2 \mathbf{I}_n & \mathbf{k}_* \\ \mathbf{k}_*^\top & k_{**} + \sigma^2 \end{bmatrix}\right), \quad (1)$$

where \mathbf{k}_* is the covariance between the training vector and the test point, k_{**} is the covariance between the test point with itself, and $\mathbf{K} + \sigma^2 \mathbf{I}_n$ is the noisy covariance matrix of the training inputs. Applying Bayesian inversion, it is possible to analytically compute the posterior distribution over the output y_* , given the new input, and the training dataset \mathcal{D}

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

$$\mu_{\text{GP}*} = \mathbf{k}_*^\top (\mathbf{K} + \sigma^2 \mathbf{I}_n)^{-1} \mathbf{y} = \mathbf{k}_*^\top \boldsymbol{\alpha} \quad (3)$$

$$\begin{aligned} \sigma_{\text{GP}*}^2 &= \sigma^2 + k_{**} - \mathbf{k}_*^\top (\mathbf{K} + \sigma^2 \mathbf{I}_n)^{-1} \mathbf{k}_* \\ &= \sigma^2 + k_{**} - \mathbf{k}_*^\top \mathbf{A} \mathbf{k}_*, \end{aligned} \quad (4)$$

where $\mathcal{D} = \{\mathbf{x}_n \in \mathbb{R}^D; y_n\}_{n=1}^N$ is the training data, $\boldsymbol{\alpha} = (\mathbf{K} + \sigma^2 \mathbf{I}_n)^{-1} \mathbf{y}$ is the weight vector of the GP mean, and $\mathbf{A} = (\mathbf{K} + \sigma^2 \mathbf{I}_n)^{-1}$ is the weight matrix of the GP variance.

Note that the predictive mean $\mu_{\text{GP}*}$ depends on the observations through the weight vector $\boldsymbol{\alpha}$, while the predictive variance $\sigma_{\text{GP}*}^2$ only depends on the inverse of the covariance function \mathbf{A} , and σ^2 is a regularization factor. Intuitively, the predicted Chl-a content in Eq. (3) is a linear combination of the observed Chl-a content values, while the certainty level, Eq. (4), only depends on the Rrs values, as seen from Eq. (5). In this work, we use the squared exponential kernel function to access similarity in the data by computing the elements of the covariance matrices. This can be written by

$$k(\mathbf{x}_m, \mathbf{x}_n) = \nu^2 \exp\left(-\frac{1}{2} \sum_{d=1}^D \left(\frac{x_m^d - x_n^d}{\lambda_d}\right)^2\right), \quad (5)$$

where the length-scale for band d , λ_d , and the positive scale factor, ν , are two hyperparameters of the kernel function. These hyperparameters, together with the noise variance σ^2 , are optimized by maximizing the marginal likelihood of the training data. For further details on the GPR model we refer to [26].

B. Feature ranking for GPR

1) *Automatic Relevance Determination (ARD)*: Relative relevance of the features can be accessed through optimizing the length-scale hyperparameters of the kernel function in Eq. (5) [30]. Since these hyperparameters controls the spread of the inputs on each spectral band, small values of λ_d indicate greater relevance. Therefore, the inverses of the optimized parameters allow the ranking of the spectral bands used in the GPR model. The length-scale hyperparameter is optimized through the maximization of the marginal likelihood function

with respect to the given parameter. The optimization is achieved by computing the partial derivatives with respect to λ_d of the negative log-marginal likelihood function. However, this method can result local maxima, which might lead to incorrect ranking of the spectral bands [30].

2) *Sensitivity Analysis (SA)*: We want to analyze the importance of spectral bands and features for a given function $\phi(\mathbf{x})$ by using a trained GPR model. To do so, let us define the sensitivity of spectral band (also called feature) j as

$$s_j = \int \left(\frac{\partial \phi(\mathbf{x})}{\partial x_j}\right)^2 p(\mathbf{x}) d\mathbf{x}, \quad (6)$$

where $p(\mathbf{x})$ is the probability density function over the D -dimensional input vector $\mathbf{x}_n = [x_n^1, \dots, x_n^D]^\top$. Intuitively, the objective of the sensitivity analysis is to evaluate changes 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 s_j will be positive for all bands and features. The empirical estimate of the sensitivity for the j th feature can be written as

$$s_j = \frac{1}{N} \sum_{n=1}^N \left(\frac{\partial \phi(\mathbf{x}_n)}{\partial x_n^j}\right)^2, \quad (7)$$

where N denotes the number of training samples.

In our study $\phi(\mathbf{x})$ represents the conditional mean function $\mu_{\text{GP}*}$. The resulting empirical estimate of the GP mean sensitivity is therefore:

$$\begin{aligned} s_{\mu_{\text{GP}*}}^j &= \frac{1}{N} \sum_{q=1}^N \left(\frac{\partial \phi(\mathbf{x}_q)}{\partial x_q^j}\right)^2 \\ &= \frac{1}{N} \sum_{q=1}^N \left(\frac{\partial \sum_{p=1}^N \alpha_p k(\mathbf{x}_p, \mathbf{x}_q)}{\partial x_q^j}\right)^2 \\ &= \frac{1}{N} \sum_{q=1}^N \left(\sum_{p=1}^N \frac{\alpha_p (x_p^j - x_q^j)}{\lambda_j^2} k(\mathbf{x}_p, \mathbf{x}_q)\right)^2. \end{aligned} \quad (8)$$

Note that the calculation of the empirical sensitivity is computed in closed-form using the training data points and the inferred $\boldsymbol{\alpha}$.

C. Partial Least Squares Regression (PLSR)

Assume once again the *in-situ* Chl-a (\mathbf{X}) and Rrs (\mathbf{y}) training dataset $D \equiv \{\mathbf{X}, \mathbf{y}\}$, where now the observations are collected in matrices, such that \mathbf{X} is an $N \times D$ input data-matrix consisting of $d = 1, \dots, D$ dimensions (spectral bands) and $n = 1, \dots, N$ observations, and let \mathbf{y} be the corresponding $N \times 1$ output-vector (Chl-a measurements), holding $n = 1, \dots, N$ observations.

The Partial Least Squares (PLS) model is based on introducing so-called latent variables, or X-scores, denoted by \mathbf{T} ($N \times H$). \mathbf{T} is relating \mathbf{X} and \mathbf{y} , and H is the number of latent variables (PLS components) [28], [39]. These latent variables are usually fewer than the number of features ($H < D$) and they are representing both \mathbf{X} and \mathbf{y} in the latent \mathbf{T} -space, such that the covariance between the projection of \mathbf{X} and \mathbf{y}

in the \mathbf{T} - space is maximized. Then the PLS model can be formally written by

$$\begin{aligned}\mathbf{X} &= \mathbf{TP}^T + \mathbf{E} \\ \mathbf{y} &= \mathbf{Tc} + \mathbf{f} \\ \mathbf{T} &= \mathbf{XW}^* \\ \mathbf{W}^* &= \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1},\end{aligned}\quad (9)$$

where \mathbf{P} ($D \times H$) is a matrix of the X-loadings and \mathbf{c} ($H \times 1$) is the y-loadings, and they are good representations (also referred to "summaries" in [28]) of \mathbf{X} and \mathbf{y} , respectively. The term \mathbf{W}^* ($D \times H$) holds the weights of \mathbf{X} , and defines the common latent variable space (X-scores). The error terms, \mathbf{E} ($N \times D$) and \mathbf{f} ($N \times 1$), are assumed to be iid. $\sim \mathcal{N}(0, \sigma^2)$. In order to impose orthogonal latent variables (\mathbf{T}), the weight matrix \mathbf{W} ($D \times H$) is introduced, and \mathbf{W} holds the eigenvectors of the variance-covariance matrix, $\mathbf{X}^T\mathbf{Y}\mathbf{Y}^T\mathbf{X}$. Thus the vectors of \mathbf{W} are orthonormal, and the row-vectors of \mathbf{T} are orthogonal to each other.

Then the PLS model can be used for regression by expressing \mathbf{y} as

$$\mathbf{y} = \mathbf{XW}^*\mathbf{c} + \mathbf{f} = \mathbf{Xb} + \mathbf{f}, \quad (10)$$

where $\mathbf{b} = \mathbf{W}^*\mathbf{c}$. This way \mathbf{y} can be estimated from \mathbf{X} , obtaining a meaningful relationship between \mathbf{X} and \mathbf{y} . The best fit is achieved by minimizing the error term \mathbf{f} in the PLSR model.

The X-scores, X- and y-loadings and the weights can be computed by using a PLS algorithm (An example of a PLS algorithm can be seen in Appendix VII). For further details on the PLS model and the various PLS algorithms we refer to [40], [41], [42], [43] and [44], [45].

The number of latent variables can be determined by using cross validation. However, in this work, the training-data is a multi-spectral dataset, where the maximum number of bands is 8 and $N \gg D$, and we keep $H = D$ in the training process.

D. Feature ranking for PLSR

Feature relevance in the PLSR model can be accessed directly from the regression coefficients \mathbf{b} ($D \times 1$) in Eq. (10). However, here we focus another way to assign relevance to the input features, called the Variable Importance in Projection (VIP) method.

1) *Variable Importance in Projection (VIP)*: The Variable Importance in Projection (VIP_j) measures the contribution to the total variance of the j^{th} input feature ($j = 1, \dots, D$), which is reflected by the weights (w_{hj}) from each component [46], [47]. It can be written by (Note, the dataset is centered and scaled.)

$$\text{VIP}_j = \sqrt{D \sum_{h=1}^H (c_h^2 t_h^T t_h) (w_{hj} / \|w_j\|^2) / \sum_{h=1}^H (c_h^2 t_h^T t_h)}. \quad (11)$$

VIP is a measure of the contribution of each feature through the variance explained by each latent variable. The term $(c_h^2 t_h^T t_h)$ is the variance of y explained by the h^{th} latent

variable. Thus, the VIP measure can also be expressed in term of Sum-of-Squares [48] by

$$\text{VIP}_j = \sqrt{D \sum_{h=1}^H SS_h (w_{hj} / \|w_j\|^2) / \sum_{h=1}^H SS_h}, \quad (12)$$

where SS_h is the percentage of y explained by the h^{th} latent variable. Intuitively, the VIP value is a sum of squares, weighted by the PLS weights w_j , which takes into account the explained variance in the PLSR model. The average of the $(\text{VIP}_j)^2$ is equal to one, therefore features with $\text{VIP}_j > 1$ are picked as the most relevant feature [39].

III. ILLUSTRATING THE CONCEPT OF THE FEATURE RANKING METHODS

In this section, we demonstrate the performance of the feature ranking methods and regression models on two controlled datasets. We simulate two cases; one simple low-dimensional and one complicated, very high-dimensional example. In both cases, the relationship between the input and output is known, and the output is constructed to be a function of both relevant and irrelevant input features. These experiments give us some insight into the performance of the methods, and provide potentially more confidence in the results obtained, when they are applied to real data, where no ground-truth information is available.

A. Description of the data

In the first experiment we try to predict the response variable y from input vectors $\mathbf{x}_n = [x_n^1, x_n^2]$, where $x_n^1 \sim \mathcal{N}(0, 0.1)$, $x_n^2 \sim \mathcal{U}(0, 1)$, and $y_n = 2x_n^2$, $n = 1, \dots, 1000$. The output (predicted mean) changes only in the second dimension, while it is fairly constant in the first. We expect that the feature ranking methods would identify the second dimension to be important in the prediction of y_n .

For the second experiment, we use a high-dimensional dataset $D = \{\mathbf{x}_n, y_n\}_{n=1}^N$, where $\mathbf{x}_n = [x_n^1, \dots, x_n^{1600}]$ is the input, and y_n is a scalar output, for $n = 1, \dots, 200$. Let R^{1600} define a 1600-dimensional feature space, and let $R_i^{\kappa_i}$, for $i = 1, 2, 3, 4$, be four 121-dimensional subspaces of R^{1600} , where κ_i denote the sets of feature indices of these subspaces. Let furthermore for each $\mathbf{x}'_n \in R^{1600}$ be independent random variables distributed according to a Gaussian distribution with zero mean and 0.5 variance, $N(0, 0.5)$. Let z_n^i for $i = 1, 2, 3, 4$ be four random variables, also distributed by $N(0, 0.5)$, and let $y_n = z_n^1 z_n^2$. Define

$$x_n^j = \begin{cases} iz_n^i & \text{for } j \in \kappa_i \text{ and } i = 1, 2, 3, 4, \\ x_n^j & \text{otherwise,} \end{cases} \quad (13)$$

for $j = 1, \dots, 1600$. The output y_n is hence only the product of z_n^1 and z_n^2 , and we expect that the feature ranking methods, (note, ARD for feature ranking could not be computed for the toy example because the optimization of the hyper-parameters (λ) failed due to the high-dimensional dataset), assign relevance only to the corresponding subspaces, R^{κ_1} and R^{κ_2} .

In this example, we also performed regression to compare the performances of the GPR and PLSR models. The toy example used here was inspired by [49].

B. Experiment 1

Figure 1 shows the results obtained in the first example for the ARD, SA and VIP methods. The left panel shows that the output (predicted mean) changes only in the second dimension. All the three methods can identify the feature which contributes the most to the prediction of y_n (Fig. 1 right panel).

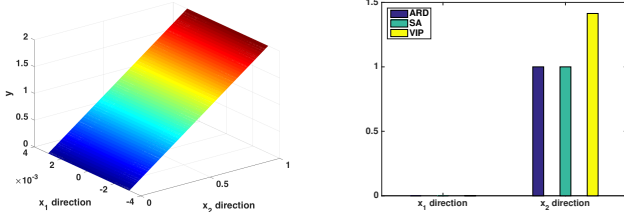


Fig. 1. The data for the first experiment (left) and the result of the feature ranking methods (right).

C. Experiment 2

First, applying the SA and the VIP to the second experimental dataset returns a sensitivity map s_d , where $d = 1600$. Furthermore, transforming s_d into a matrix (image) allows the visualization of the performance of the feature ranking methods. Figure 2 shows the sensitivity map for the VIP and the SA. The two important features, z_n^1 and z_n^2 , correspond to the squares in the left-top and right-bottom part in Fig. 2, respectively. It can be observed, that both feature ranking methods could successfully identify the relevant features. However, in the case of the VIP, all features above the value 1 count as important features [47]. The computed sensitivity map (Fig. 2 left) reveals that inputs at the top-right and bottom-left area, corresponding to z_n^3 and z_n^4 , respectively, which are not relevant in the prediction of the output, were also assigned to have a sensitivity above the value of 1. Overall, VIP seems to show higher sensitivity to the irrelevant inputs than the SA.

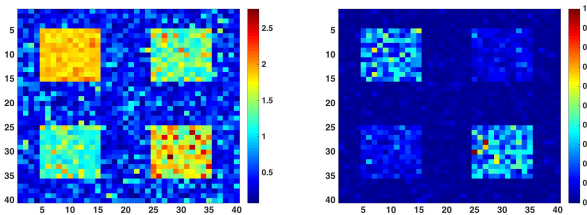


Fig. 2. Sensitivity map of the VIP (left) and the SA (right).

In addition, we performed regression on this toy dataset by using the PLSR and GPR models. Figure 3 shows the targets and the predicted values. In order to assess the strength of the regression, we computed several regression performance measures (Table I): Bias, accuracy by the Normalized-Root-Mean-Square Error (NRMSE) and goodness of fit as measured by squared Pearson's correlation coefficient (R^2). It can be observed in Table I that the GPR model has the lowest bias,

NRMSE values and the highest correlation, $R^2 = 1$. Hence the GPR model shows a better regression performance than the PLSR model. (The description of the computation of the regression performance measures can be seen in Sec. IV-C.2.)

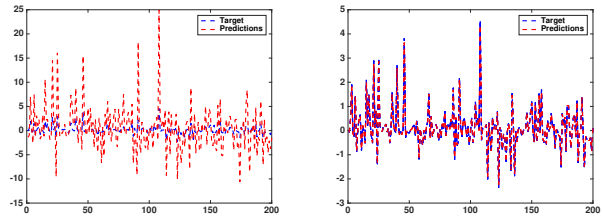


Fig. 3. Target values and predicted values for the PLSR model (left) and the GPR (right).

TABLE I
SUMMARY OF THE COMPUTED REGRESSION PERFORMANCE MEASURES
FOR THE PLSR AND GPR MODEL FOR THE TOY DATA.

| Method | Bias | NRMSE | R^2 |
|--------|--------|--------|--------|
| PLSR | 2.7023 | 0.5208 | 0.5836 |
| GPR | 0.0271 | 0.0062 | 1.0000 |

D. Concluding remarks

From these simulations we may draw the following conclusions.

The first example showed that all the three feature ranking methods were sensitive to the relevant feature in the case of the low-dimensional controlled dataset.

The second experiment revealed that both the SA and VIP methods could successfully identify the important features in a very high dimensional dataset. The GPR resulted in more accurate regression than the PLSR model for this example.

Based on these experiments, we find it reasonable to apply the presented ranking methodologies to multi-spectral data in order to find the most relevant spectral bands in Chl-a estimation.

IV. EXPERIMENTAL SETUP

Next, we describe the experimental setup and show the results of the three ranking algorithms, the SA, ARD and VIP, when applied to a Chl-a/Rrs matchup dataset, acquired by the ESA's MERIS sensor. (Note; we also performed the same analysis as presented below for a SeaWiFS (NASA) and a MODIS-Aqua (NASA) matchup dataset, which have different spectral resolutions, and therefore may give slightly different conclusions [17]. The results of these analyses can be found in the Appendix B.) The datasets can be obtained from the SeaBASS database (<http://seabass.gsfc.nasa.gov> and <https://oceancolor.gsfc.nasa.gov/>).

TABLE II
SUMMARY OF THE MERIS DATASET.

| MERIS | |
|----------------------------------|---------------------------------|
| Chl-a range (mgm ⁻³) | 0.017 - 40.23 |
| Nr. of samples | 567 |
| Bands (λ_c (nm)) | 413 443 490 510 560 620 665 681 |
| Band width | 10 nm and 7.5 nm |

A. Description of the dataset

Table II summarizes the MERIS dataset with respect to the center wavelength (λ_c), bandwidth, the range of the Chl-a contents, and the total number of samples.

The MERIS dataset consists of 567 measurements, measured between April 2002 and March 2012. It can be seen that the Chl-a content spans a wide range of concentration with values in the range between 0.017 and 40.23 mgm⁻³. The bandwidth is here 10 nm for bands 1- 7, and 7.5 nm for band 8. Figure 4 shows a few of the measured Rrs values for

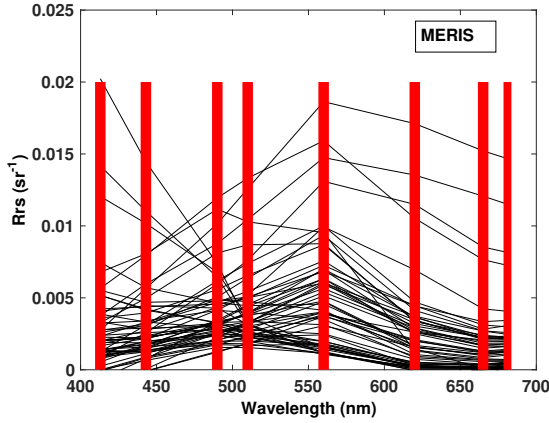


Fig. 4. The Rrs (sr⁻¹) spectrum of the MERIS dataset. The red bars indicate the location of the spectral bands.

the MERIS dataset. The red bars indicate the position of the bands, and the width of the bars illustrates the band widths. In the following, we will number the bands chronologically 1, 2,...,8, where 1 corresponds to the smallest and 8 to the longest wavelengths.

The Rrs-values show large variations across the dataset, corresponding to both Case 1 and Case 2 conditions [50], which is bound to cause randomness in the estimated Chl-a contents. By definition [51] Case 1 conditions refer to waters, dominated by phytoplankton, and phytoplankton associated products, while Case 2 conditions can contain other constituents, and usually correspond to optically complex waters.

B. Feature sets

1) *Set A: Spectral band feature set:* This feature set contains 8 features, the spectral bands of the MERIS dataset, ordered chronologically as noted above. Feature 1 is the band centered at 413 nm, and feature 8 corresponds to the spectral band at 681 nm.

2) *Set B: Extended spectral band feature set:* We extended the spectral band feature set by adding three additional features. These features are the band ratios from the OC2, OC3 and OC4 state-of-art models [11], [12], [13], [14] and [15]. These band ratios are the ratios of the measured Rrs in the blue and the green regions. The bands included in the band ratios are determined from the optical properties of the Chl-a absorption spectrum [13]. The three additional features in Set B are defined by:

$$R_{OC2} = \frac{Rrs(490nm)}{Rrs(560nm)} \quad (14)$$

$$R_{OC3} = \frac{\max(Rrs(443, 490nm))}{Rrs(560nm)} \quad (15)$$

$$R_{OC3} = \frac{\max(Rrs(443, 490, 510nm))}{Rrs(560nm)}. \quad (16)$$

Hence, Set B consists of 11 features, features 1 to 8 are the spectral bands chronologically ordered, and features 9, 10 and 11 are the band ratios, corresponding to R_{OC2} , R_{OC3} and R_{OC4} , respectively.

C. Test setup

The test setup consists of a feature ranking analysis, and three regression performance tests.

1) *Feature ranking:* First, we used Set B for ranking the relevance of the features using the SA, ARD and VIP methods.² We also performed feature ranking on Set A. This was to help determine which spectral bands, in the absence of the band ratio features, that are important for Chl-a retrieval, and to possibly add some insight to the physics of the problem.³

2) *Regression:* We carried out regression by splitting the dataset into 50% for training and 50% for testing. This was done by sorting the dataset based on the increasing Chl-a content. Then we split the dataset, with odd numbers forming the training set and even numbers forming the test set, respectively. This allowed us to have approximately similar statistical variations in the training and test datasets.

Regression strength was evaluated by computing the following regression performance measures: The Bias, the Normalized Root Mean Squared Errors (NRMSE) and the Squared Correlation Coefficient (R^2). These measures are expressed by

$$\text{Bias} = \frac{1}{N} \sum_{i=1}^N |(y_i - \hat{y}_i)| \quad (17)$$

$$\text{NRMSE} = \frac{1}{y_{\max} - y_{\min}} \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (18)$$

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (19)$$

²Note, in the SA method, we assume that all features can be treated as independent variables, although the band ratio features are functionally made up of other spectral bands in the feature set. Despite this fact, our results indicate that this has had no practical impact on the results.

³The GPR model has a computational load of $O(n^3)$. However, there are several techniques, which can increase computational efficiency based on dimensionality reduction [52], and feature ranking for GPR can be an important tool in this regard.

where N is the number of observations in the test set, y is the true Chl-a content, \hat{y} is the predicted Chl-a, y_{\max} is the maximum observed value, y_{\min} is the minimum observed value, and \bar{y} is the mean of the observed Chl-a contents in the test set.

We performed regression studies in three test setups.

Test 1: First, we used Set B to evaluate the GPR and PLSR models, when only one feature was used in the regression models. For each feature, we computed the regression performance measures, and the study would hence find which single feature would result in the strongest regression.

Test 2: In the next step, we used features from Set B and gradually extended the number of features input the regression models by sequentially adding one more feature at a time, following the order of importance determined by the SA, ARD and VIP methods, respectively. In each case we computed the Bias, NRMSE and R^2 values. This revealed how the number of features affected the regression performance, and how many and which features that would produce the best values for the regression performance measures. Furthermore, the three ranking methods could also be comparatively evaluated. Here, we assigned numbers to the ranked features from 1 to 11 according to the SA, ARD and VIP. Feature number 1 corresponds to the most important feature for the given ranking method, while the number 11 is the least relevant feature. Hence, since the ranking methods evaluate the importance of the features differently, the actual feature associated with a given number and ranking method needs to be looked up in Table IV.

Test 3: Finally, we used Set A to perform the same sequential procedure as in Test 2. Using only the ranked spectral bands for regression, allows to determine which bands and the minimum number needed, without having a significantly decrease in regression strength. Here the number 1 is assigned to the most relevant spectral band according to the three ranking methods, while the number 8 represents the least important band. The actual feature associated with a given number and ranking method needs to be looked up in Table V.

By comparing the results of Tests 2 and 3 with the results of Test 1 we can assess the increase in regression strength, when an increasing number of the ranked features are used in the regression. Hence, Test 1 can be seen as a reference performance level.

3) *Comparison to the OC models:* The OC models are empirical fourth-order models. They use band ratios. The estimated Chl-a content can be expressed by

$$\text{Chl-a} = 10^{a_0 + \sum_{i=1}^4 a_i \left(\log_{10} \left(\frac{\text{Rrs}(\lambda_{\text{blue}})}{\text{Rrs}(\lambda_{\text{green}})} \right) \right)^i}, \quad (20)$$

where a_0 and a_i are the (polynomial) coefficients, $\text{Rrs}(\lambda_{\text{blue}})$ is the maximum of the measured Rrs values in the blue region, and $\text{Rrs}(\lambda_{\text{green}})$ is the measured Rrs on the green band. The sensor specific coefficients and bands used for the band ratios are listed in Table III. For further details on the OC models we refer to NASA's ocean color website⁴.

⁴oceancolor.gsfc.nasa.gov

4) *Uncertainty of the estimates:* Last, but not least we illustrate the uncertainty of the estimates of the GPR model by choosing the strongest model, and comparing the uncertainty of the estimates with the *best* (lowest number of features and still strong regression performance) model. This shows how the uncertainty level changes, when we reduce the number of bands in the GPR model.

V. RESULTS

A. Feature ranking

Fig. 5 and Table IV summaries the results of the SA, ARD and VIP feature ranking on the MERIS dataset, when all the features were used (Set B). As can be seen, both the SA and the VIP methods assigned the highest relevance to the three band ratios, but they ranked their individual relevance differently. SA gave highest relevance to R_{OC3} , R_{OC2} and R_{OC4} , in that order, and compared to these features, the relevance of the eight band features is more or less negligible. In the VIP method, only features with score above 1 are considered important. As seen in the left panel of Fig. 5, all band ratio features are scored above one, whereas the band features are below, and hence, less important.

The ARD method, (middle panel of Fig. 5), ranked the features differently. It gave highest relevance to the band centred at 665 nm, and high relevance to the bands at 443 nm and at 620 nm, in addition to the band ratio features. However, except for the bands 560 nm and 413 nm, which both have very low scoring, the relative differences in importance for ARD are not as pronounced as for the other two ranking methods.

Fig. 6 and Table V show the results of the ranking methods, when only the spectral bands are used (Set A). As can be seen, all ranking methods assigned high relevance to the band positioned at 560 nm. This band is the denominator in all band ratio features, since this is a reference band because there is little or no absorption by Chl-a in this region [53], and the results reconfirm its importance. SA gave high importance also to the bands at 413 nm and 620 nm, whereas VIP, in addition to 560 nm, only gave the band at 413 nm a score above 1. The other bands are scored slightly below 1. ARD also puts high relevance to the bands at 490 nm and 510 nm. Both these bands are included in the R_{OC4} band ratio. In summary, these results suggest that the bands used in the band ratios (560 nm and 490 nm) are important. The high relevance of the band at 413 nm, as suggested by both SA and VIP, may be explained by the fact that the dataset also includes samples from eutrophic waters.

B. Regression experiments

Test 1: Fig. 7 shows the three regression measures, Bias, NRMSE, and R^2 , for the single feature regression setup for the GPR model (upper panel), and for the PLSR model (lower panel). The numbers ticked on the x -axis are denoting band features in increasing order of wave length, and the solid red line horizontally across each figure is inserted as a reference to ease the visual comparisons. As noted, using only one feature at a time, both the GPR and PLSR models resulted in the best performance for the three band ratio features, and according to

TABLE III
COEFFICIENTS IN THE OC MODELS FOR THE MERIS DATASET

| MERIS | | | | | | | |
|-------|-----------------|------------|----------------|----------------|----------------|----------------|----------------|
| Model | blue band | green band | a ₀ | a ₁ | a ₂ | a ₃ | a ₄ |
| OC2 | 490 | 560 | 0.2389 | -1.9369 | 1.7627 | -3.0777 | -0.1054 |
| OC3 | 443 > 490 | 560 | 0.2521 | -2.2146 | 1.5193 | -0.7702 | -0.4291 |
| OC4 | 443 > 490 > 510 | 560 | 0.3255 | -2.7677 | 2.4409 | -1.1288 | -0.4990 |

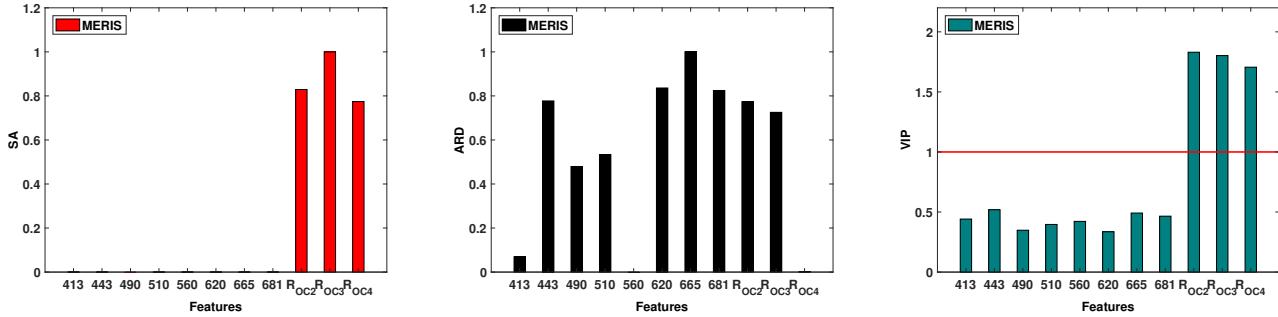


Fig. 5. SA of the GP mean (left), ARD (middle) and VIP (right) for Set B. For the VIP method features above the red line are important in the estimation of Chl-a, while bands below are not likely to contribute to the prediction.

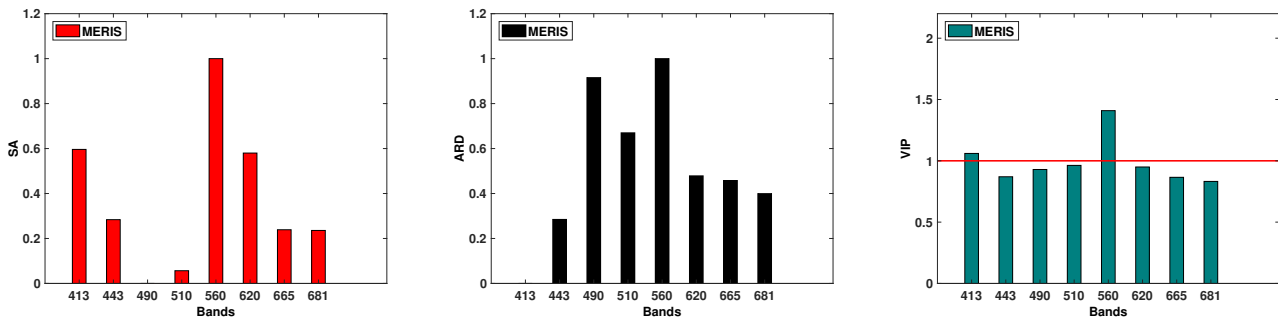


Fig. 6. SA of the GP mean (right), ARD (middle) and VIP (left) for Set A. For the VIP method features above the red line are important in the estimation of Chl-a, while bands below are not likely to contribute to the prediction.

TABLE IV
RANKED FEATURES FOR SET B.

| MERIS | | | |
|-----------------|------------------|------------------|------------------|
| Ranked features | SA | ARD | VIP |
| 1 | R _{OC3} | 665 nm | R _{OC2} |
| 2 | R _{OC2} | 620 nm | R _{OC3} |
| 3 | R _{OC4} | 681 nm | R _{OC4} |
| 4 | 560 nm | 443 nm | 443 nm |
| 5 | 620 nm | R _{OC2} | 665 nm |
| 6 | 413 nm | R _{OC3} | 681 nm |
| 7 | 665 nm | 510 nm | 413 nm |
| 8 | 681 nm | 490 nm | 560 nm |
| 9 | 443 nm | 413 nm | 510 nm |
| 10 | 510 nm | R _{OC4} | 490 nm |
| 11 | 490 nm | 560 nm | 620 nm |

TABLE V
RANKED FEATURES FOR SET A.

| MERIS | | | |
|--------------|-----|-----|-----|
| Ranked bands | SA | ARD | VIP |
| 1 | 560 | 560 | 560 |
| 2 | 413 | 490 | 413 |
| 3 | 620 | 510 | 510 |
| 4 | 443 | 620 | 620 |
| 5 | 665 | 665 | 490 |
| 6 | 681 | 681 | 443 |
| 7 | 510 | 443 | 665 |
| 8 | 490 | 413 | 681 |

the measures, all three have approximately similar regression strength, (i.e. lowest Bias and NRMSE, and the highest R^2). We also note that the four bands with the longest wavelengths, i.e. those centred at 560, 620, 665 and 681 nm, showed

good regression performance, especially for the GPR model. For this regression model, the four band features with the shortest wavelengths showed a significantly weaker individual regression strength. On the other hand, for the PLSR model, this difference in performance between the short and the long wavelengths is less pronounced.

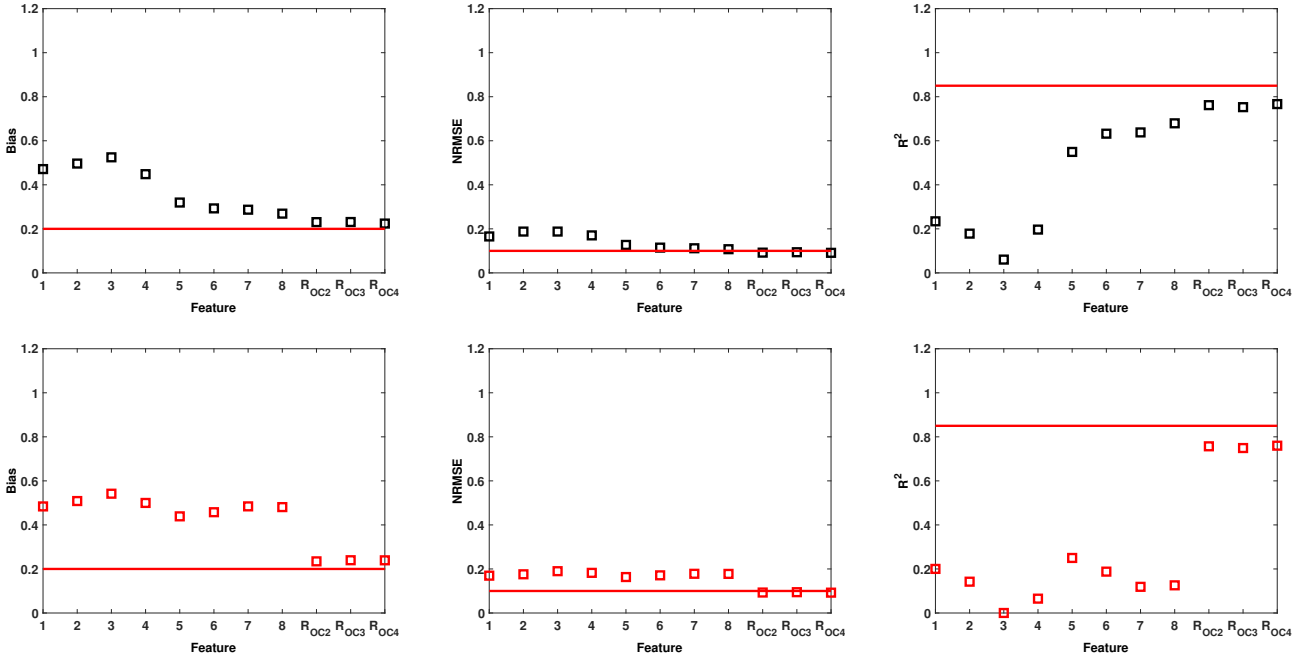


Fig. 7. Computed measures for the MERIS dataset for the GPR (top-row) and PLSR (bottom-row) model. Regression performance measures were computed by performing regression with only one feature at the time for Set B. (The red line is a reference line, allowing an easier comparison of the performance of the GPR and PLSR models.)

Test 2: In this experiment, we gradually extend the number of features input to the regression models by sequentially adding one more feature at a time, following the order of importance determined by the SA, ARD and VIP methods. The resulting regression performance measures as function of the number of ranked input features are summarised in Fig. 8 for GPR (upper panel) and PLSR (lower panel). The figures show that for both regression models, the regression performance improves as more and more input features are used. The best regression performance is achieved when as many as 10 features, ranked by the SA method, were applied to the GPR model. This only excludes the band positioned at 490 nm, but this band is already contributing to the regression, as it is included in the band ratios. We also note that the *improvement curves*, i.e. the reduction in Bias and NRSME and increase in R^2 from left to right, vary with ranking method and also with the regression model. We note that the curves associated with the SA ranking in general provides the best regression performance for both GPR and PLSR. For the SA ranking we observe that the GPR-curves have a clear step-wise trend, with big improvement steps at 3 and 5 feature inputs. There is basically no improvement of adding adding R_{OC2} to R_{OC3} , but significant improvements when also R_{OC4} is used. Similar, there is little change in the measures by adding band feature 560 nm, but big improvement when band 620 nm is added. The curves for the PLSR model also have step-wise appearance, but the steps are at different numbers of feature inputs, and the curves seem to achieve the optimal performance with 8 number of features following the SA ranking.

The GPR and PLSR models showed similar trends in

performance when extending the feature sets. Also the values for the performance measures were quite similar. The most noticeable difference occurred in the NRMSE value, where the GPR model showed a slightly lower value.

Test 3: Adding sequentially the ranked band features to the GPR and PLSR models revealed improvements, already when the second most important band was added (Fig. 9). We observe that in general the band features ranked by the ARD method showed the best regression performance measures as we extended the input sets, both for the GPR and PLSR models. Again we note that the ARD ranked the bands at 560 and 490 nm as the most relevant bands, and these bands correspond to the bands used in the R_{OC2} band ratio. However, the GPR model converged to a higher R^2 value, when many features were used, and the overall best performance was achieved with the GPR model using all features.

C. Comparison to the OC models

Finally, we compared the regression performance of the GPR and PLSR models with the OC2, OC3 and OC4 models. These comparisons are summarised in Table VI in terms of resulting performance measures associated with some selected input features. The 9 first rows display the performance measures associated with the three OC-models and the GPR and the PLSR models using R_{OC2} , R_{OC3} and R_{OC4} as inputs, respectively. Note that both the GPR and PLSR perform better than all the OC models. The best result is obtained with GPR using R_{OC4} as input feature. In the next rows, we present the numerical results for the GPR and PLSR using 1, 2, 8, and 10 input features, as described below.

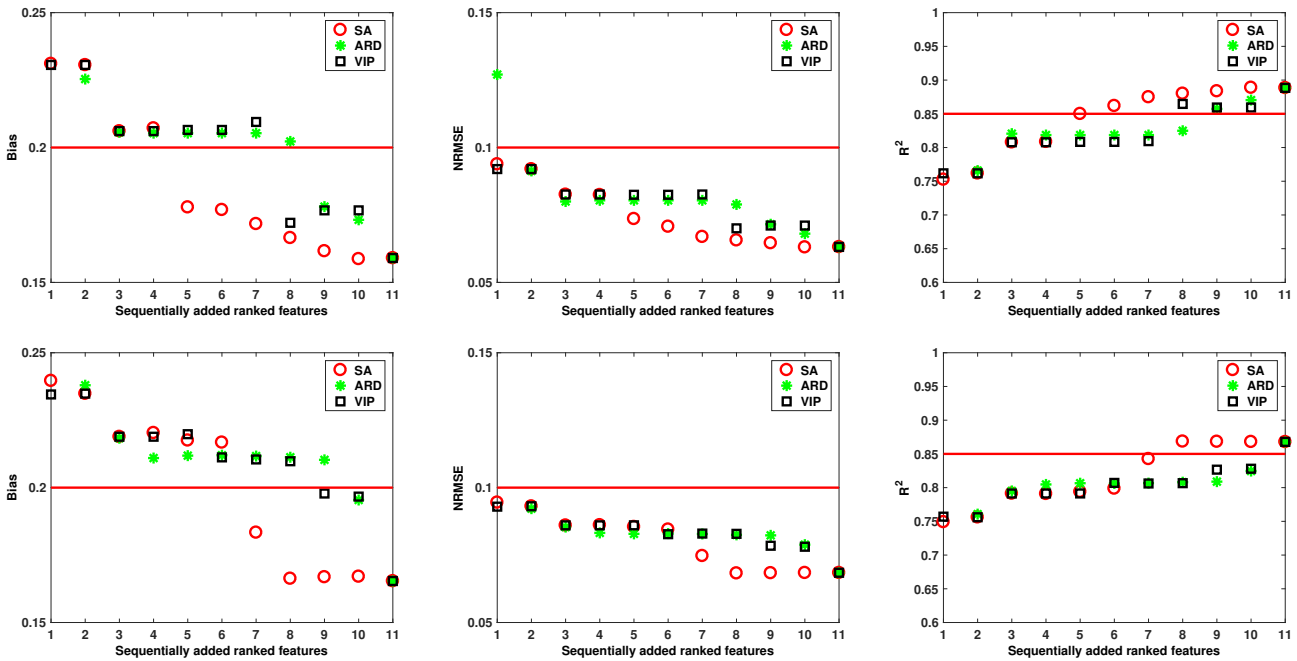


Fig. 8. Computed measures for Set B for the GPR (top-row) and PLSR (bottom-row) model. Here the numbers represent the ranked features of the SA (red circle), ARD (green star) and VIP (black square) methods. The ranked features were added sequentially as inputs to the GPR and PLSR models.(The red line is a reference line, allowing an easier comparison of the performance of the GPR and PLSR models.)

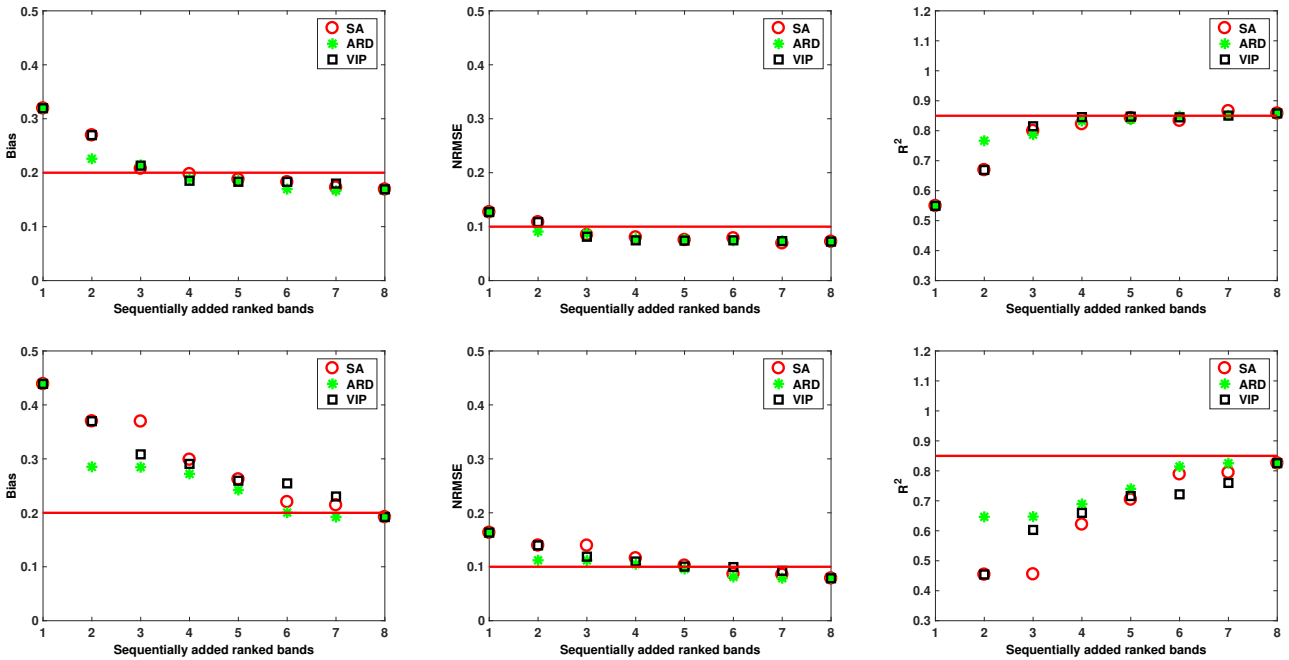


Fig. 9. Computed measures for Set A for the GPR (top-row) and PLSR (bottom-row) model. Here the numbers represent the ranked features of the SA (red circle), ARD (green star) and VIP (black square) methods. The ranked features were added sequentially as inputs to the GPR and PLSR models. (The red line is a reference line, allowing an easier comparison of the performance of the GPR and PLSR models.)

- i) *Single band feature:* We display the regression performance of band features 7 (665 nm) and 8 (681 nm) when used as single feature inputs. These correspond to the band features with best performance, in the single feature experiment (Fig. 7).
- ii) *Two band features:* Here we display the regression performance when two bands defines the input vector. We have chosen the bands at 560 and 490

TABLE VI
COMPARISON OF THE OC MODELS WITH GPR AND PLSR MODELS FOR THE MERIS DATASET.

| Regression model | Bias | NRMSE | R ² |
|--|---------------|---------------|----------------|
| OC2 | 0.2715 | 0.1114 | 0.7101 |
| GPR with R _{OC2} | 0.2306 | 0.0920 | 0.7618 |
| PLSR with R _{OC2} | 0.2345 | 0.0929 | 0.7570 |
| OC3 | 0.2676 | 0.1090 | 0.7241 |
| GPR with R _{OC3} | 0.2311 | 0.0938 | 0.7526 |
| PLSR with R _{OC3} | 0.2391 | 0.0945 | 0.7491 |
| OC4 | 0.2347 | 0.0949 | 0.7671 |
| GPR with R _{OC4} | 0.2241 | 0.0912 | 0.7666 |
| PLSR with R _{OC4} | 0.2392 | 0.0924 | 0.7598 |
| GPR with band centered at 665 nm | 0.2861 | 0.1143 | 0.6521 |
| PLSR with band centered at 665 nm | 0.4842 | 0.1780 | 0.1189 |
| GPR with band centered at 681 nm | 0.2691 | 0.1076 | 0.6793 |
| PLSR with band centered at 681 nm | 0.4809 | 0.1776 | 0.1257 |
| GPR with bands centered at 490 and 560 nm | 0.2259 | 0.0909 | 0.7666 |
| PLSR with bands centered at 490 and 560 nm | 0.2853 | 0.1120 | 0.6465 |
| GPR with all bands | 0.1692 | 0.0723 | 0.8578 |
| PLSR with all bands | 0.1922 | 0.0787 | 0.8257 |
| GPR with all features, except band centered at 490 nm | 0.1587 | 0.0631 | 0.8889 |
| PLSR with all features, except band centered at 490 nm | 0.1670 | 0.0684 | 0.8680 |

nm, which according to Fig. 9 would give the best performance (ARD-ranking).

- iii) *All band features*: For comparison, we also include the results when using all band features as input.
- iv) *Ten input features*: We finally combine the 3 band ratio features with 7 bands according to the SA ranking. This resulted in the overall best regression performance (see Fig. 8).

The GPR model with ten input features showed the strongest regression strength, and it actually outperforms all the OC models. The second strongest model was the PLSR with the same features. Both GPR and PLSR performed well in comparison to the OC models, also with few input bands. Hence, the results of Table VI suggest that Chl-a content retrieval can be improved in comparison to the OC models by using the GPR model with only two bands.

D. Uncertainty level of the GPR model

Based on our results in Sec. V-C, we illustrate the advantageous property of the GPR model, i.e. its ability to assign an uncertainty level to the estimates. Figure 10 (top) shows the estimated Chl-a values by using the obtained strongest GPR model (Table VI), the actual measured Chl-a values and the uncertainty level of the estimates for the MERIS dataset. Figure 10 (bottom) shows, how the uncertainty level changes, when the GPR model uses only two spectral bands, 490 and 560 nm.

For some of the estimated values, the uncertainty level increases slightly when fewer features are used. This is in good correspondence with the computed regression performance measures.

The interesting observation to note is that the uncertainty level does not reveal a significant increase when only the two most important spectral bands are used in model, compared to the strongest GPR model, with 10 input features.

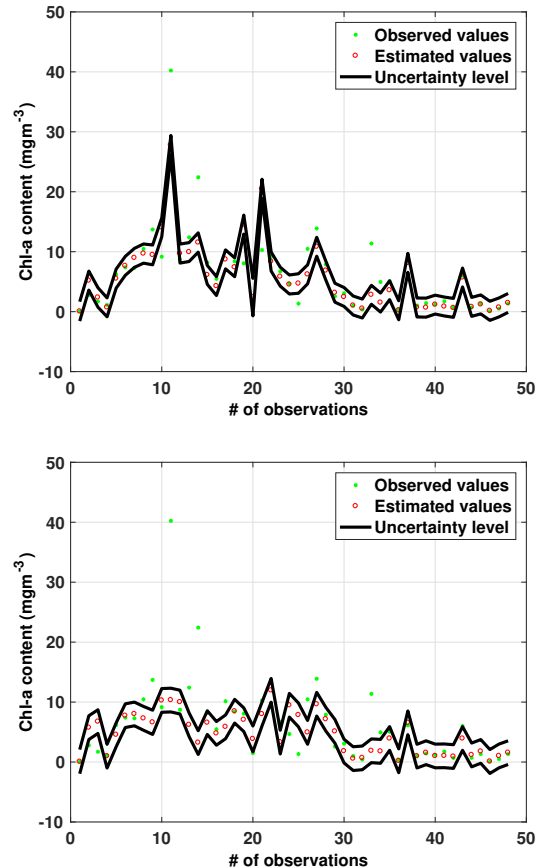


Fig. 10. Observed and estimated Chl-a contents by using the GPR model for the MERIS dataset. The black solid lines indicate the uncertainty level of the estimates. The top figure shows some of the observations by using the GPR model with the features that resulted the best values for the regression performance measures. The bottom panel illustrates the GPR model by using only two bands for regression.

VI. ILLUSTRATIVE EXAMPLE

We illustrate the effect of using different algorithms and bands for Chl-a estimation on a test image acquired in July in 2015 by MODIS-Aqua over high-latitude Arctic oceans (N = 89.9931°, S = 65.7186°, W = -174.2612°, E = 5.236°). Sea ice concentration was estimated to 33.7383 % and the cloud-coverage was 51.908 %. The quasi true color image can be seen in Fig. 11.

Figure 12 shows the Chl-a content maps estimated by the OC3 algorithm (top), by the GPR model with all the spectral bands (middle), and by the GPR model with bands centered at 488 and 678 nm (bottom). We observe that the estimated Chl-a maps have some differences. The GPR maps show lower concentration values, and reveal more details than the map of the OC3 algorithm. Both GPR estimates illustrate how the model captures internal structures, presumably associated to ocean current eddies. It is expected that phytoplankton blooms follow the pattern of the current eddies.

The corresponding computed regression performance measures for these cases were (see Table XII, Appendix B): bias = 0.2272, NRMSE = 0.1057, R² = 0.7868 (OC3); bias =

0.1628, NRMSE = 0.0702, $R^2 = 0.8844$ (GPR, all bands) bias = 0.1774, NRMSE = 0.0746, $R^2 = 0.8684$ (GPR, 488 and 678 nm). Based on these measures, the best results should be expected to be achieved by using the GPR model with all bands, followed by the GPR model, and with bands centered at 488 and 678 nm. The most pronounced difference between the maps computed by the GPR model with all bands (middle-panel) and GPR with bands centered at 488 and 678 nm (bottom-panel) is in the amount of the assigned Chl-a content. Which one is most correct, cannot be concluded without in-situ information. This example, together with the computed statistics from the training set, shows that both the regression model and the input feature vector are important in ocean color applications, and that more research is needed to select the most reliable methodology.

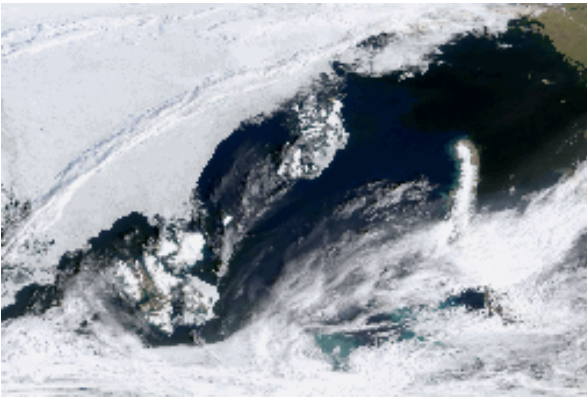


Fig. 11. Quasi true color image of the test site.

VII. CONCLUSION AND FUTURE WORK

In this work, we studied feature ranking and regression performances of two regression methods, namely the Gaussian Process Regression (GPR) and the Partial Least Squares Regression (PLSR) models, when applied for Chl-a content estimation based on a global MERIS dataset. In the GPR model, we use a Bayesian approach to learn the non-linear functional relationship between the input feature vectors and the output Chl-a measurements, and the feature ranking was conducted using the Automatic Relevance Determination (ARD) and Sensitivity Analysis (SA). The PLSR is a well-known linear regression model, which uses a so-called latent variable space to relate the input features to the Chl-a measurement. In PLSR, feature relevance was analysed using a ranking method called Variable Importance in Projection (VIP).

From the eight spectral bands of a MERIS matchup data set, we created two input feature sets. One (Set A) consisting of all the spectral bands, and another extended feature set (Set B), which in addition to all bands, also consisted of the three band ratio features, denoted R_{OC2} , R_{OC3} and R_{OC4} , which are the inputs used in the state-of-art OC2, OC3, and OC4 regression models. The relevance of features were analysed by all the ranking methods, and subsequently input to the two regression models in a test setup consisting of three tests. Using three measures, the Bias, NRSME, and R^2 , the individual regression

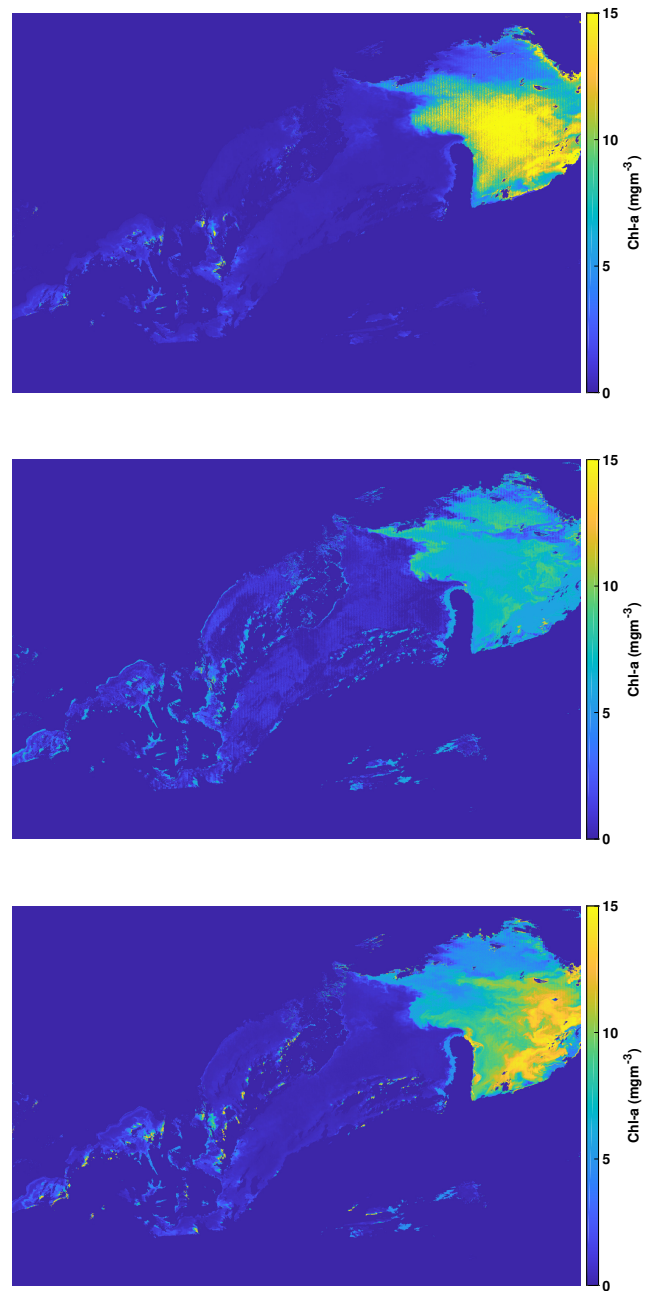


Fig. 12. Chl-a content estimates by using the OC3 algorithm (top), GPR model with all bands (middle) and the GPR model with bands centered at 488 and 678 nm (bottom).

strength of each feature as single input was computed. Next, we evaluated the regression strength of sets of features by gradually extending the number of features, adding one more feature at a time, following the order of importance determined by the SA, ARD and VIP methods, respectively. We did this analysis first for Set B, and then for Set A.

Our results show that the all feature ranking methods can successfully assign sensible relevance to the features. Since the methods operate according to different ranking criteria,

it is expected that they might assess the features differently. Both SA and VIP assigned the highest relevance to the three band ratio features, whereas ARD gave highest scores to the spectral bands centred at 665 nm, 443 nm, and at 620 nm. ARD also found that the band ratio features to be important. When applied to the features in Set A, i.e. only the band features, all three methods agreed to give highest relevance to the band at 560 nm, but the order of the next relevant bands were somewhat different.

ARD ranked the spectral band at 490 nm as second most important. We note that the bands at 490 and 560 nm are included in all the band ratio features. The spectral region at 490 nm corresponds to the shifted Chl-a absorption peak, and has been used to avoid contribution from CDOM [54]. Even though this wavelength mostly represents accessory pigments, due to the correlation of these accessory pigments with Chl-a, the spectral band at 490 nm can successfully be used to derive Chl-a concentration. The band centred at 560 nm is a reference wavelength, since phytoplankton absorption is at the minimum around this green band [54].

In regression Test 1 (single feature input), we found that the three band ratio features achieved far the best regression performance in both regression models. This is not surprising, given the fact that these features are composed of two spectral bands, carefully selected. We also found that spectral bands with the longest wavelengths, i.e. 560, 620, 665, and 681nm, were performing significantly better than the shorter wavelengths, especially for the GPR model. It has previously been shown that bands in the red part of the visible region of the electromagnetic spectrum can successfully be applied for Chl-a retrieval due to the second absorption maximum of the Chl-a molecule [13]. Our results support this finding.

Not surprisingly, Test 2 showed that for both regression models, the regression performance improved as more and more input features are used as input. The best regression performance is achieved when 10 features, ranked by the SA method, were applied to the GPR model. We note that the curves associated with the SA ranking in general provided the best regression performance for both GPR and PLSR. We also found that the improvement-curves of GPR associated with the SA ranking had a clear step-wise trend, with big improvement steps at 3 and 5 feature inputs. These jumps would intuitively give clues to which features to select first, if dimensionality reduction were to be applied to the input feature space.

Test 3 also showed that the regression performance gradually improved when applying more and more input spectral bands, and the overall best performance was obtained with all 8 spectral bands. However, we note that the gain in performance for the GPR model by increasing the number of input spectral bands from 4 to 8, is only minor (see Fig.9).

Our comparisons between the GPR and PLSR regression models and the three OC-models clearly demonstrated that both GPR and PLSR performed better than the state-of-the-art models for several different sets of input features. Based on the performance measures we have used, we also find that the GPR model in all cases has the strongest regression performance.

Note that we also performed the same study for two additional global matchup datasets, the SeaWiFS and MODIS-

Aqua datasets. Due to lack of space, we did not include a detailed description of these studies here, but some important results have been tabulated in Appendix B. We found similar results for the SeaWiFS and MODIS-Aqua datasets as those reported in this paper for the MERIS dataset.

Based on the current studies, we conclude that there is a big potential for improvements in Chl-a retrieval from satellite-based observations by selecting the most appropriate regression model in combination with an optimal set of input features.

For future work, we plan to perform extensive validation studies of the GPR model, and compare its performance with state-of-art Chl-a retrieval algorithms, and other algorithms e.g. Neural Networks, on optically complex aquatic environments, such as coastal and Arctic waters, and mid-latitude shallow lakes.

ACKNOWLEDGMENTS

This research is partly funded by CIRFA partners and the Research Council of Norway (grant number 237906).

REFERENCES

- [1] C. S. Reynolds, *The Ecology of Phytoplankton*. Cambridge University Press, 2006.
- [2] Govindjee, *Bioenergetics of Photosynthesis*. Academic Press, 1975.
- [3] T. Volk and M. I. Hoffert, *Ocean Carbon Pumps: Analysis of Relative Strengths and Efficiencies in Ocean-Driven Atmospheric CO₂ Changes*. American Geophysical Union, 2013. [Online]. Available: <http://dx.doi.org/10.1029/GM032p0099>
- [4] 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 CO₂ in the Southern Ocean," *Science*, vol. 283, no. 5400, pp. 365–367, 1999. [Online]. Available: <http://science.sciencemag.org/content/283/5400/365>
- [5] M. Hein and K. Sand-Jensen, "CO₂ increases oceanic primary production," *Nature*, vol. 388, pp. 526–527, 1997.
- [6] M. Hofmann, B. Worm, S. Rahmstorf, and H. J. Schellnhuber, "Declining ocean chlorophyll under unabated anthropogenic CO₂ 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>
- [7] N. T. T. Ha, K. Koike, and M. T. Nhuan, "Improved accuracy of chlorophyll-a concentration estimates from modis imagery using a two-band ratio algorithm and geostatistics: As applied to the monitoring of eutrophication processes over tien yen bay (northern vietnam)," *Remote Sensing*, vol. 6, no. 1, pp. 421–442, 2014. [Online]. Available: <http://www.mdpi.com/2072-4292/6/1/421>
- [8] X.-e. Yang, X. Wu, H.-l. Hao, and Z.-l. He, "Mechanisms and assessment of water eutrophication," *Journal of Zhejiang University SCIENCE B*, vol. 9, no. 3, pp. 197–209, Mar 2008. [Online]. Available: <https://doi.org/10.1631/jzus.B0710626>
- [9] F. A. Al-Wassai and N. V. Kalyankar, "Major limitations of satellite images," *CoRR*, vol. abs/1307.2434, 2013. [Online]. Available: <http://arxiv.org/abs/1307.2434>
- [10] J. Hill, C. Diemer, O. Ståhl, and T. Udelhoven, "A local correlation approach for the fusion of remote sensing data with different spatial resolutions in forestry applications," in *In: IAPRS, Vol. 32, Part 7-4-3 W6*, pp. 3–4, 1999.
- [11] 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.
- [12] A. Morel and S. Maritorena, "Bio-optical properties of oceanic waters: A reappraisal," *Journal of Geophysical Research: Oceans*, vol. 106, no. C4, pp. 7163–7180, 2001. [Online]. Available: <http://dx.doi.org/10.1029/2000JC000319>
- [13] J. E. O'Reilly, S. Maritorena, 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. 24 937–24 953, 1998.
- [14] J. E. O'Reilly, S. Maritorena, 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. F. Cota, S. B. Hooker, C. R. McClain, K. L. Carder, F. Müller-Karger, L. Harding, A. Magnuson, D. Phinney, G. F. Moore, J. Aiken, K. R. Arrigo, R. Letelier, and M. Culver, "SeaWiFS postlaunch calibration and validation analyses, part 3," *Nasa Tech. Memo. 2000-206892*, vol. 11, 2000.
- [15] 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.
- [16] 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>
- [17] I. S. Robinson, *Measuring the Oceans from Space: The principles and methods of satellite oceanography*. Praxis Publishing Ltd, 2004.
- [18] 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.
- [19] R. Doerffer and H. Schiller, "The meris case 2 water algorithm," *International Journal of Remote Sensing*, vol. 28, no. 3-4, pp. 517–535, 2007. [Online]. Available: <https://doi.org/10.1080/01431160600821127>
- [20] M. Hieronymi, D. Müller, and R. Doerffer, "The olci neural network swarm (onns): A bio-geo-optical algorithm for open ocean and coastal waters," *Frontiers in Marine Science*, vol. 4, p. 140, 2017. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fmars.2017.00140>
- [21] 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.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] C. E. Rasmussen and C. K. I. Williams, *Gaussian Process for Machine Learning*. Cambridge MA: MIT Press, 2006.
- [27] 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.
- [28] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, pp. 109 – 130, 2001.
- [29] 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.
- [30] 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>
- [31] J. Verrelst, J. P. Rivera, A. Gitelson, J. Delegido, J. Moreno, and G. Camps-Valls, "Spectral band selection for vegetation properties retrieval using gaussian processes regression," *International Journal of Applied Earth Observation and Geoinformation*, vol. 52, pp. 554 – 567, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0303243416301234>
- [32] K. Blix, G. Camps-Valls, and R. Jenssen, "Gaussian process sensitivity analysis for oceanic chlorophyll estimation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 4, pp. 1265–1277, April 2017.
- [33] H. Feilhauer, G. P. Asner, and R. E. Martin, "Multi-method ensemble selection of spectral bands related to leaf biochemistry," *Remote Sensing of Environment*, vol. 164, pp. 57 – 65, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0034425715001339>
- [34] I. E. Frank and J. H. Friedman, "A statistical view of some chemometrics regression tools," *Technometrics*, vol. 35, no. 2, pp. 109–135, 1993. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/00401706.1993.10485033>
- [35] X. Wang, U. Kruger, and B. Lennox, "Recursive partial least squares algorithms for monitoring complex industrial processes," *Control Engineering Practice*, vol. 11, pp. 613–632, 2003.
- [36] K. Ryan and K. Ali, "Application of a partial least-squares regression model to retrieve chlorophyll-a concentrations in coastal waters using hyper-spectral data," *Ocean Science Journal*, vol. 51, no. 2, pp. 209–221, 2016. [Online]. Available: <http://dx.doi.org/10.1007/s12601-016-0018-8>
- [37] 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.
- [38] P. J. Werdell, S. W. Bailey, G. S. Fargion, C. Pietras, K. D. Knobelspiess, G. C. Feldman, and C. R. McClain, "Unique data repository facilitates ocean color satellite validation," *EOS Trans. AGU*, vol. 84, no. 38, p. 387, 2003.
- [39] R. Gosselin, D. Rodrigue, and C. Duchesne, "A Bootstrap-VIP approach for selecting wavelength intervals in spectral imaging applications," *Chemometrics and Intelligent Laboratory Systems*, vol. 100, pp. 12–21, 2010.

- [40] N. L. Afanador, "Important Variable Selection in Partial Least Squares for Industrial Process Understanding and Control," Ph.D. dissertation, Radboud University Nijmegen, 2014.
- [41] H. Abdi, "Partial least squares regression and projection on latent structure regression (PLS Regression)," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 1, pp. 97 – 106, 2010. [Online]. Available: <http://dx.doi.org/10.1002/wics.51>
- [42] S. Rännar, F. Lindgren, P. Geladi, and S. Wold, "A PLS kernel algorithm for data sets with many variables and fewer objects. part 1: Theory and algorithm," *Journal of Chemometrics*, vol. 8, pp. 111 – 125, 1994. [Online]. Available: <http://dx.doi.org/10.1002/cem.1180080204>
- [43] S. de Jong, "SIMPLS: An alternative approach to partial least squares regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 18, no. 3, pp. 251 – 263, 1993. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/016974399385002X>
- [44] B. S. Dayal and J. F. MacGregor, "Improved PLS algorithms," *Journal of Chemometrics*, vol. 11, no. 1, pp. 73 – 85, 1997. [Online]. Available: [http://dx.doi.org/10.1002/\(SICI\)1099-128X\(199701\)11:1<73::AID-CEM435>3.0.CO;2-](http://dx.doi.org/10.1002/(SICI)1099-128X(199701)11:1<73::AID-CEM435>3.0.CO;2-)
- [45] K. Song, D. Lu, L. Li, S. Li, Z. Wang, and J. Du, "Remote sensing of chlorophyll-a concentration for drinking water source using genetic algorithms (GA)-partial least square (PLS) modeling," *Ecological Informatics*, vol. 10, pp. 25 – 36, 2012, ecological Informatics and Ecosystem Conservation. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1574954111000756>
- [46] L. Eriksson, E. Johansson, N. Kettaneh-Wold, and S. Wold, "Multi- and Megavariate Data Analysis. principles and applications," *Journal of Chemometrics*, vol. 16, no. 5, pp. 261 – 262, 2001.
- [47] P. Jonsson, "Surface Status Classification, Utilizing Image Sensor Technology and Computer Models," Ph.D. dissertation, Mid Sweden University, 2015.
- [48] T. Mehmood, K. H. Liland, L. Snipen, and S. Sæbø, "A review of variable selection methods on Partial Least Squares Regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 118, pp. 62 – 69, 2012.
- [49] 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>
- [50] L. Gross-Colzy, S. Colzy, R. Frouin, and P. Henry, "A general ocean color atmospheric correction scheme based on principal components analysis: Part 1. performance on case 1 and case 2 waters," in *In: SPIE, Vol. 6680, pp. 6680 - 12*, 2007. [Online]. Available: <http://dx.doi.org/10.1117/12.738508>
- [51] C. Mobley, D. Stramski, W. Bissett, and E. Boss, "Optical modeling of ocean waters: Is the case 1- case 2 classification still useful?" *Oceanography*, 2004.
- [52] E. L. Snelson, "Flexible and efficient gaussian process models for machine learning," Ph.D. dissertation, Gatsby Computational Neuroscience Unit University College London, 2007.
- [53] H. R. Gordon, O. B. Brown, R. H. Evans, J. W. Brown, R. C. Smith, K. S. Baker, and D. K. Clark, "A Semianalytic Radiance Model of Ocean Color," *Journal of Geophysical Research*, vol. 93, pp. 10 909–10 924, 1988.
- [54] J. P. Cannizzaro and K. L. Carder, "Estimating chlorophyll a concentrations from remote-sensing reflectance in optically shallow waters," *Remote Sensing of Environment*, vol. 101, no. 1, pp. 13 – 24, 2006. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0034425705004025>
- [55] S. Wold, A. Ruhe, H. Wold, and W. J. D. III, "The partial least squares approach to generalized inverses," *SIAM J. Sci. Stat. Comput.*, vol. 5, pp. 735 – 743, 1984.

APPENDIX A

PLSR algorithm

Here we present the so called NIPALS (Non-linear Iterative Partial Least Squares) algorithm introduced by [55]. The NIPALS algorithm can be written by

for $h = 1, \dots, H$ (21)

$$\begin{aligned} \mathbf{w} &= \mathbf{X}^T \mathbf{y} \\ \mathbf{t} &= \mathbf{X} \mathbf{w} \\ c &= \mathbf{y}^T \mathbf{t} / \mathbf{t}^T \mathbf{t} \\ \mathbf{p} &= \mathbf{X}^T \mathbf{t} / \mathbf{t}^T \mathbf{t} \\ \mathbf{X} &= \mathbf{X} - \mathbf{t} \mathbf{p}^T \\ \mathbf{y} &= \mathbf{y} - \mathbf{t} c \end{aligned}$$

end for

$$\begin{aligned} \mathbf{W} &= \mathbf{w}_1, \dots, \mathbf{w}_H \\ \mathbf{T} &= \mathbf{t}_1, \dots, \mathbf{t}_H \\ \mathbf{P} &= \mathbf{p}_1, \dots, \mathbf{p}_H \\ \mathbf{c} &= c_1, \dots, c_H \end{aligned}$$

APPENDIX B

We performed the same experiments as in Sec. IV and V for two additional datasets, the SeaWiFS and MODIS-Aqua datasets. We found that the results were in good correspondence with our findings for the MERIS dataset, namely a satisfactory regression can be already achieved by using the spectral bands centered at 490 and 555 nm for the SeaWiFS dataset, and 488 and 678 nm for the MODIS-Aqua dataset.

Below we present the description of the datasets and features, and the results of the feature ranking methods, regression models and comparisons for these additional datasets.

The SeaWiFS and MODIS-Aqua datasets

The SeaWiFS and MODIS-Aqua datasets are summarized in Table VII. These datasets represent both Case 1 and 2 conditions.

TABLE VII
SUMMARY OF THE SEAWIFS AND MODIS-AQUA DATASETS.

| SeaWiFS | |
|----------------------------------|-----------------------------|
| Chl-a range (mgm ⁻³) | 0.024 - 129.332 |
| Nr. of samples | 1465 |
| Bands (λ_c (nm)) | 421 443 490 510 555 670 |
| Band width | 20 nm |
| MODIS-Aqua | |
| Chl-a range (mgm ⁻³) | 0.0153 - 25.4985 |
| Nr. of samples | 579 |
| Bands (λ_c (nm)) | 412 443 488 531 547 667 678 |
| Band width | 10 nm, 15 nm |

The band ratio features for the SeaWiFS dataset can be

TABLE VIII
COEFFICIENTS IN THE OC MODELS FOR THE SEAWIFS AND
MODIS-AQUA DATASETS

| Model | blue band | green band | a ₀ | a ₁ | a ₂ | a ₃ | a ₄ |
|------------|-----------------|------------|----------------|----------------|----------------|----------------|----------------|
| SeaWiFS | | | | | | | |
| OC2 | 490 | 555 | 0.2511 | -2.0853 | 1.5035 | -3.1747 | 0.3383 |
| OC3 | 443 > 490 | 555 | 0.2515 | -2.3798 | 1.5823 | -0.6372 | -0.5692 |
| OC4 | 443 > 490 > 510 | 555 | 0.3272 | -2.9940 | 2.7218 | -1.2259 | -0.5683 |
| MODIS-Aqua | | | | | | | |
| OC2 | 488 | 547 | 0.2500 | -2.4752 | 1.4061 | -2.8233 | 0.5405 |
| OC3 | 443 > 488 | 547 | 0.2424 | -2.7423 | 1.8017 | 0.0015 | -1.2280 |

written by:

$$R_{OC2} = \frac{Rrs(490nm)}{Rrs(555nm)} \tag{22}$$

$$R_{OC3} = \frac{\max(Rrs(443, 490nm))}{Rrs(555nm)} \tag{23}$$

$$R_{OC3} = \frac{\max(Rrs(443, 490, 510nm))}{Rrs(555nm)}, \tag{24}$$

and for the MODIS-Aqua dataset

$$R_{OC2} = \frac{Rrs(488nm)}{Rrs(547nm)} \tag{25}$$

$$R_{OC3} = \frac{\max(Rrs(443, 488nm))}{Rrs(547nm)}. \tag{26}$$

Feature ranking for the SeaWiFS and MODIS-Aqua datasets

The ranked features can be seen in Table IX, and the ranked spectral band features are presented in Table X.

TABLE IX
RANKED FEATURES FOR THE SEAWIFS AND MODIS-AQUA DATASETS.

| SeaWiFS | | | |
|-----------------|------------------|------------------|------------------|
| Ranked features | SA | ARD | VIP |
| 1 | R _{OC4} | 490 nm | R _{OC4} |
| 2 | R _{OC2} | R _{OC2} | R _{OC3} |
| 3 | R _{OC3} | R _{OC4} | R _{OC2} |
| 4 | 490 nm | 443 nm | 412 nm |
| 5 | 510 nm | 412 nm | 555 nm |
| 6 | 555 nm | 670 nm | 490 nm |
| 7 | 412 nm | 510 nm | 510 nm |
| 8 | 443 nm | R _{OC3} | 443 nm |
| 9 | 670 nm | 555 nm | 670 nm |
| MODIS-Aqua | | | |
| Ranked features | SA | ARD | VIP |
| 1 | R _{OC3} | 678 nm | R _{OC3} |
| 2 | R _{OC2} | 531 nm | R _{OC2} |
| 3 | 412 nm | R _{OC2} | 488 nm |
| 4 | 443 nm | 412 nm | 547 nm |
| 5 | 547 nm | 547 nm | 531 nm |
| 6 | 488 nm | 443 nm | 678 nm |
| 7 | 678 nm | 667 nm | 443 nm |
| 8 | 667 nm | 488 nm | 412 nm |
| 9 | 488 nm | R _{OC3} | 667 nm |

TABLE X
RANKED SPECTRAL BANDS FOR THE SEAWIFS AND MODIS-AQUA
DATASETS.

| SeaWiFS | | | |
|--------------|-----|-----|-----|
| Ranked bands | SA | ARD | VIP |
| 1 | 412 | 555 | 555 |
| 2 | 555 | 490 | 412 |
| 3 | 443 | 443 | 670 |
| 4 | 670 | 412 | 443 |
| 5 | 490 | 670 | 510 |
| 6 | 510 | 510 | 490 |
| MODIS-Aqua | | | |
| Ranked bands | SA | ARD | VIP |
| 1 | 488 | 488 | 547 |
| 2 | 678 | 678 | 412 |
| 3 | 547 | 412 | 531 |
| 4 | 667 | 531 | 443 |
| 5 | 412 | 547 | 488 |
| 6 | 443 | 667 | 678 |
| 7 | 531 | 443 | 667 |

Regression

Table XI and XII show the computed regression performance measures for the SeaWiFS and MODIS-Aqua datasets, respectively.

TABLE XI
COMPARISON OF THE OC MODELS WITH GPR AND PLSR MODELS FOR THE SEAWIFS DATASET.

| Regression model | Bias | NRMSE | R ² |
|---|---------------|---------------|----------------|
| OC2 | 0.2319 | 0.0990 | 0.8128 |
| GPR with R _{OC2} | 0.2142 | 0.0908 | 0.8403 |
| PLSR with R _{OC2} | 0.2159 | 0.0915 | 0.8373 |
| OC3 | 0.2275 | 0.0977 | 0.8180 |
| GPR with R _{OC3} | 0.2129 | 0.0914 | 0.8380 |
| PLSR with R _{OC3} | 0.2225 | 0.0939 | 0.8289 |
| OC4 | 0.2123 | 0.0907 | 0.8406 |
| GPR with R _{OC4} | 0.2079 | 0.0890 | 0.8464 |
| PLSR with R _{OC4} | 0.2302 | 0.0952 | 0.8243 |
| GPR with bands centered at 490 and 555 nm | 0.2101 | 0.0894 | 0.8450 |
| PLSR with bands centered at 490 and 555 nm | 0.2592 | 0.1082 | 0.7724 |
| GPR with all bands | 0.1792 | 0.0780 | 0.8820 |
| PLSR with all bands | 0.2394 | 0.1019 | 0.7980 |
| GPR with bands centered at 412, 443, 490, 670 nm, and features R _{OC2} and R _{OC4} | 0.1805 | 0.0780 | 0.8820 |
| PLSR with bands centered at 412, 443, 490, 670 nm, and features R _{OC2} and R _{OC4} | 0.2021 | 0.0854 | 0.8583 |

TABLE XII
COMPARISON OF THE OC MODELS WITH GPR AND PLSR MODELS FOR THE MODIS-AQUA DATASET.

| Regression model | Bias | NRMSE | R ² |
|--|---------------|---------------|----------------|
| OC2 | 0.2270 | 0.1098 | 0.7697 |
| GPR with R _{OC2} | 0.2072 | 0.0894 | 0.8115 |
| PLSR with R _{OC2} | 0.2123 | 0.0910 | 0.8053 |
| OC3 | 0.2272 | 0.1057 | 0.7868 |
| GPR with R _{OC3} | 0.2062 | 0.0880 | 0.8173 |
| PLSR with R _{OC3} | 0.2304 | 0.0958 | 0.7847 |
| GPR with band centered at 678 nm | 0.2713 | 0.1119 | 0.7042 |
| PLSR with band centered at 678 nm | 0.3760 | 0.1559 | 0.4403 |
| GPR with bands centered at 488 and 678 nm | 0.1774 | 0.0746 | 0.8684 |
| PLSR with bands centered at 488 and 678 nm | 0.2648 | 0.1064 | 0.7379 |
| GPR with all bands | 0.1628 | 0.0702 | 0.8844 |
| PLSR with all bands | 0.2049 | 0.0856 | 0.8271 |
| GPR with bands centered at 412, 531 and 678 nm, and features R _{OC2} | 0.1697 | 0.0725 | 0.8771 |
| PLSR with bands centered at 412, 531 and 678 nm, and features R _{OC2} | 0.1891 | 0.0837 | 0.8350 |