Article

Developing a New Machine-Learning Algorithm for Estimating Chlorophyll-a Concentration in Optically Complex Waters: A Case Study for High Northern Latitude Waters by Using Sentinel 3 OLCI

Katalin Blix [1,*,†], **Juan Li** [2], **Philippe Massicotte** [2] and **Atsushi Matsuoka** [2,3]

[1] Department of Physics and Technology, Faculty of Science and Technology, UiT the Arctic, University of Norway, 9019 Tromsø, Norway

[2] Takuvik Joint International Laboratory, Department de Biologie, Universite Laval, 1045 ave. de la Medecine, Quebec, QC G1V 0A6, Canada

[3] Takuvik Joint International Laboratory, CNRS, Universite Laval, 1045 ave. de la Medecine, Quebec, QC G1V 0A6, Canada

* Correspondence: katalin.blix@uit.no; Tel.: +47-483-49-399

† Current address: UiT the Arctic University of Norway, P.O. box 6050 Langnes, NO-9037 Tromsø, Norway.

check for updates

Abstract: The monitoring of Chlorophyll-a (Chl-a) concentration in high northern latitude waters has been receiving increased focus due to the rapid environmental changes in the sub-Arctic, Arctic. Spaceborne optical instruments allow the continuous monitoring of the occurrence, distribution, and amount of Chl-a. In recent years, the Ocean and Land Color Instruments (OLCI) onboard the Sentinel 3 (S3) A and B satellites were launched, which provide data about various aquatic environments on advantageous spatial, spectral, and temporal resolutions with high SNR. Although S3 OLCI could be favorable to monitor high northern latitude waters, there have been several challenges related to Chl-a concentration retrieval in these waters due to their unique optical properties coupled with challenging environments including high sun zenith angle, presence of sea ice, and frequent cloud covers. In this work, we aim to overcome these difficulties by developing a machine-learning (ML) approach designed to estimate Chl-a concentration from S3 OLCI data in high northern latitude optically complex waters. The ML model is optimized and requires only three S3 OLCI bands, reflecting the physical characteristic of Chl-a as input in the regression process to estimate Chl-a concentration with improved accuracy in terms of the bias (five times improvements.) The ML model was optimized on data from Arctic, coastal, and open waters, and showed promising performance. Finally, we present the performance of the optimized ML approach by computing Chl-a maps and corresponding certainty maps in highly complex sub-Arctic and Arctic waters. We show how these certainty maps can be used as a support to understand possible radiometric calibration issues in the retrieval of Level 2 reflectance over these waters. This can be a useful tool in identifying erroneous Level 2 Remote sensing reflectance due to possible failure of the atmospheric correction algorithm.

Keywords: ocean color monitoring; Chlorophyll-a; machine-learning; arctic; Sentinel 3 OLCI

1. Introduction

Arctic waters have been going through significant changes due to the rapid thinning and retreating sea ice over the past decade [1]. Larger areas of open waters, increasing nutrient supplies and favorable light conditions allowed phytoplankton blooms to occur, where they have not been observed before [2–4]. Phytoplankton are photosynthetic marine organisms [5], and their presence indicates

vulnerable aquatic ecosystems [6]. Monitoring the occurrence of phytoplankton in high northern latitude waters is of great importance in climate research [7–9], and industries, such as offshore [10,11], shipping [12], tourism [13], fishing [14], and aquaculture [15] (regarding Harmful Algae Blooms). Browning of high northern latitude lakes have been observed due to thawing permafrost [16], which impacts the ecosystems.

Phytoplankton can be monitored from space through the Chlorophyll-a (Chl-a) pigment. There are several sensors onboard satellites providing information about aquatic Chl-a concentration on various spatial, spectral, and temporal resolutions. These sensors include the MODerate-resolution Imaging Spectroradiometer onboard Aqua satellite (MODIS-Aqua) [17], Visible Infrared Imaging Radiometer Suit (VIIRS) sensor on Suomi National Polar-orbiting Partnership (Suomi NPP) and NOAA-20 weather satellites [18], Operational Land Imager (OLI) on Landsat 8 [19], and the Multi-Spectral Instrument (MSI) on the Sentinel 2 A and B satellites [20]. The Ocean and Land Color Instruments (OLCI) onboard the Sentinel 3 (S3) A and B satellites [21] have advantageous spatial, spectral, and temporal resolutions for monitoring high northern latitude waters. S3 OLCI provides two standard Chl-a products, the open ocean and the complex water Chl-a products. Open ocean Chl-a is retrieved by using the OC4Me algorithm [22,23], which is an empirical polynomial function, based on so-called band-ratios. The term band ratio refers to the maximum ratio between the spectral bands, where absorption takes place (442.5, 490 and 510 nm) and on the band centered at 560 nm, where there is little or no absorption. The OC4Me algorithm has been used frequently in open oceans, where it has been shown to have a robust and stable performance. However, the OC4Me algorithm is not designed for complex water monitoring, where the absorption of the Chl-a can be influenced by Colored Dissolved Organic Matter (CDOM) and the backscatter of Total Suspended Matter (TSM). The complex water Chl-a product is designed for complex aquatic environments. This is retrieved by using a Neural Network (NN) algorithm [24,25]. Although this might provide an alternative solution which is able to estimate Chl-a concentration in complex waters, validation experiments have shown that the approach is sensitive to TSM, and often results in erroneous Chl-a estimates [26].

Furthermore, it is often challenging to determine the water type in advance, especially in the rapidly changing sub-Arctic, Arctic waters. This causes difficulties in the choice of Chl-a product. Therefore, our objectives are as follows:

- Introduce a unified Chl-a retrieval algorithm to monitor Chl-a concentration in complex high northern latitude waters, such as inland, coastal waters, the Marginal Ice Zone (MIZ), and for open Arctic oceans.
- Design the model specifically to S3 OLCI so that the advantageous spectral, spatial and temporal resolutions of the instrument can be used for monitoring these complex high northern latitude waters.
- Provide a tool to assess uncertainties in the retrieval of the Level 2 (L2) radiometric data, i.e., the Remote sensing reflectance (Rrs), which is the input to many of the Ocean Color algorithms, including the model introduced here.

A combination of machine-learning (ML) approaches has a great potential to achieve these objectives. ML has been demonstrated to perform successfully in ocean color applications. For instance, in [27,28] NN methods have been introduced to improve retrieval of Rrs for both open and complex waters. NN has also been studied to estimate Chl-a from remotely sensed data [29,30]. In addition to NN, other ML methods have also been studied for bio-physical parameter retrieval, which includes Support Vector Regression [31–33], Relevance Vector Regression [34], and Gaussian Process Regression (GPR) [35]. Several studies have shown that the Machine-Learning Gaussian Process Regression (ML GPR) model performs outstandingly in comparison to other parametric and ML methods [36–39].

The ML GPR method is not only a flexible and analytic approach, but it also outputs the certainty level of the estimated Chl-a concentrations, computed according to an analytic scheme. These certainty estimates allow assessment of whether the input Rrs data differs from the Rrs data used in the training of the ML GPR model. This means that in case one has control over the quality of the training Rrs data,

possible errors in the *new* S3 OLCI Rrs data, which might arise due to calibration issues and/or failures of the atmospheric correction algorithm, can be revealed.

In the work [26], the Automatic Model Selection Algorithm (AMSA) was applied, which was introduced in [40] to a complexity-diverse aquatic environment. AMSA combines feature ranking and selection with ML regression methods to determine the number and combination of spectral bands and/or features needed to obtain the strongest performance for a given aquatic environment. Hence, applying the approach to an aquatic environment representing a wide range of optical complexities, might allow determination of a model, which can be generalized and optimized to complex high northern latitude waters.

The ML GPR model was established on Lake Balaton, and was shown to improve Chl-a retrieval from S3 OLCI Rrs data [26] in the lake. The ML GPR Balaton model takes as input only three Rrs values, measured on bands centered at 442.5, 665, and 681.25 nm as input to estimate Chl-a concentration. The regression power of the method decreases, when additional OLCI Rrs bands are added. Using all the available OLCI bands in the ML GPR Balaton model resulted in a Normalized Root Mean Squared Error (NRMSE) of 0.11, a bias of 2.25 and Pearson correlation coefficient of 0.79, whereas using only the three selected bands as inputs resulted in a NRMSE of 0.10, a bias of 2.12 and a Pearson correlation coefficient of 0.83) [26]. This suggests that using only these bands in complexity-diverse waters might provide sufficient spectral information to estimate Chl-a concentration. This ML GPR Balaton model was evaluated on a Hydrolight simulated data partitioned to correspond to various Arctic waters with promising performance [41].

Finally, in this work the ML GPR Balaton model was optimized, analyzed, evaluated and tested on a comprehensive in situ dataset, including a significant number of measurements of optically complex waters obtained from Arctic, coastal, and open oceans. Here, we show that the ML GPR Balaton model has reasonable performance in estimating Chl-a concentration from S3 OLCI L2 Rrs data, where other approaches fail. We present the predictive performance of the optimized ML GPR Balaton model on highly complex sub-Arctic and Arctic waters. We show how the certainty maps can provide information about the quality of the L2 input Rrs data. The detailed description and flowchart of the analysis are presented in Section 2.2.3.

2. Materials and Methods

2.1. Data

In this work, we focused on Chl-a concentration retrieval from Rrs measured on three OLCI bands, which are included in the ML GPR Balaton model. These bands are centered at 442.5, 665, and 681.25 nm. Hence both the in situ and satellite derived radiometric data include measurements only on these three bands. We used two datasets of optically complex waters, an Arctic and coastal data (COASTlOOC), for optimizing and training the ML GPR Balaton model. The location of the stations can be seen in Figure 1. The red circles show the positions of the measurements of the COASTlOOC dataset, and the black circles represent the locations of the Arctic samples. It can be seen that the measurements originate from a wide spatial extent and from various water conditions.

We chose three test sites for evaluating the predictive performance of the optimized ML GPR Balaton model. These three scenes include highly complex aquatic environments. The locations of the test sites are indicated with squares in Figure 1. It is challenging to estimate Chl-a concentration in these test sites from remotely sensed data due to the optical complexity of these waters. Therefore, they provide excellent case studies to examine the predictive strength of the proposed model. Furthermore, we use these scenes to present how uncertainties can be assessed by using the ML GPR Balaton model.
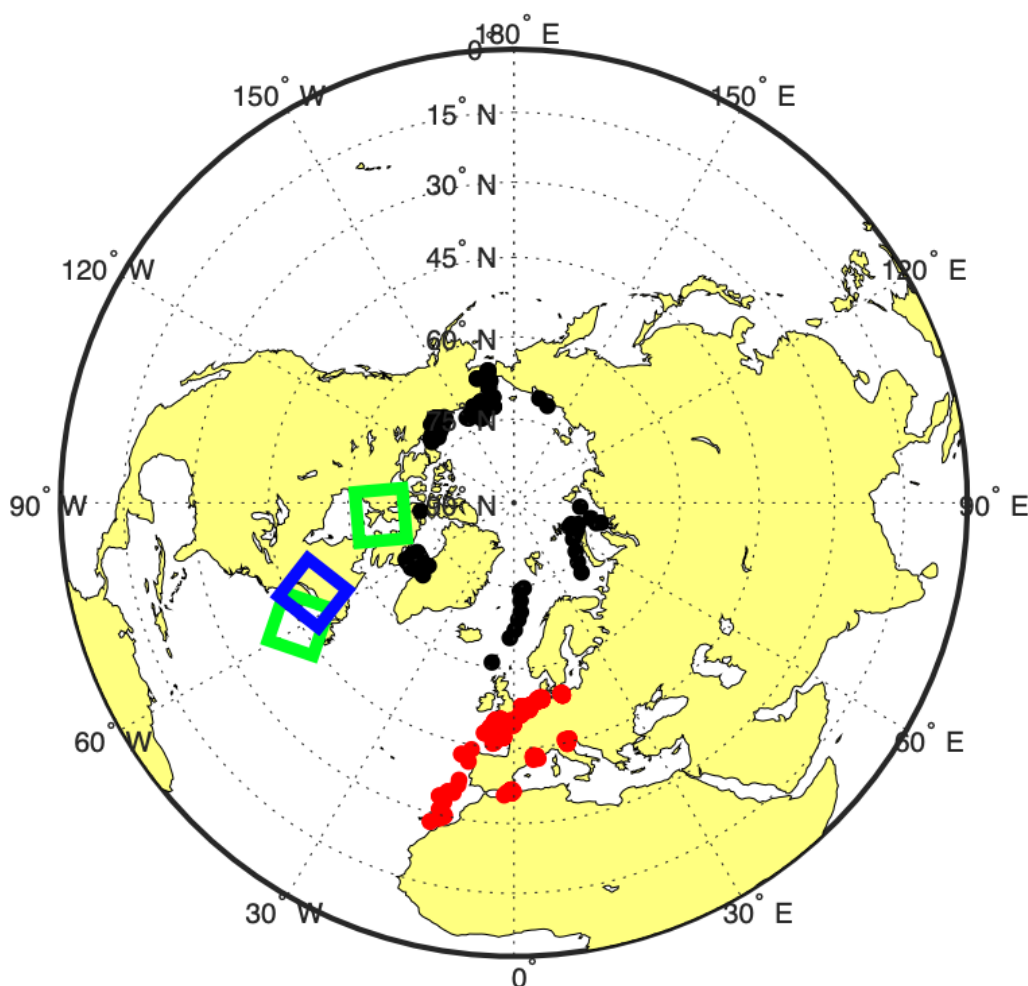
Figure 1. Location of the measurements. Red circles correspond to the COASTlOOC data and black circles to the Arctic dataset. The blue square shows the swath including the high northern latitude complex waters used for prediction (Section 2.1.2). The two green squares indicate the test sites used for predicting Chl-a in the Marginal Ice Zone, open and coastal waters (Appendix A).

2.1.1. Training Data for Model Optimization

The Arctic Data

The Arctic dataset contains measurements in the MIZ and from Arctic oceans, and it was composed by five cruises, the France-Canada-USA joint Arctic campaign MALINA [42], ICESCAPE2010 (Impacts of Climate on Ecosystems and Chemistry of the Arctic Pacific Environment), ICESCAPE2011 [43], Tara Oceans Polar Circle expedition [44], and GREEN EDGE [45].

Sampling of MALINA and GREEN EDGE was conducted onboard the Canadian Icebreaker CCGS Amundsen, and the two ICECAPEs were onboard the US Icebreaker USCGC Healy. The Tara expedition was conducted using a French schooner, Tara. Note that in some cases, during the same time periods of the CTD deployment on the ship, a barge and/or zodiac were also deployed to obtain surface water samples, and to avoid contamination due to the main ship including water quality as well as ship shadow. In MALINA, ICESCAPE2010, and ICESCAPE2011, samples were taken both from near sea ice and far from it. Details on the number of samples, sampling date and area are listed in Table 1.

Table 1. Summary of the Arctic dataset.

Cruise	Nr. of Samples	Date	Area
MALINA	35	30 July 2009–26 August 2009	Southern Beaufort Sea
ICESCAPE2010	18	15 June 2010–22 July 2010	Chukchi and Beaufort Sea
ICESCAPE2011	37	25 June 2011–29 July 2011	Chukchi and Beaufort Sea
TARA	28	24 May 2013–5 November 2013	Kara and Laptev Sea
GREEN EDGE	31	24 June 2016–10 July 2016	Baffin Bay

Rrs Measurements

All $Rrs(\lambda)$ of the five cruises used in this study were determined by the Compact-Optical Profiling System (C-OPS) [46] instrument, which is a radiometer system for determining apparent optical properties in aquatic systems. It consists of two 7 cm diameter radiometers: one measures in-water upwelling radiance $L_u(\lambda, z)$, and the other one measures downward irradiance $E_d(\lambda, z)$, pressure/depth, and dual axes tilts. Both radiometers are equipped with up to 19 optical-filter micro-radiometers. In this study, only three channels at 443, 665, 681 nm are used. The above-water global solar irradiance (E_s) was also measured to take into account the incident light field changes. Other corrections including dark currents, pressure tares, aperture offsets, tilt filtering (5 or less), and self-shading were also taken into account. Pressure tares were used to have radiometric data with vertical resolution of 1 cm [47]. Dark currents together with pressure tares were used to achieve low uncertainties for radiometric measurements (for most cases within a few percent). Data with tilt above 5 degree was not used. Self-shading was corrected in terms of diameter of the sensor and absorption values [48]. Finally, $Rrs(\lambda)$ was obtained following the protocols described in [46,47]. (Depth interval was determined so that $E_d(0-)$ was calculated from $E_d(0+)$ through air-water interface, and $E_d(0-)$ was measured by using the in-water sensor agreeing within a few percent. Within the depth interval, $\ln(L_u)$ data was fitted and extrapolated into the surface, and $L_u(0-)$ was obtained. $L_w(0+)$ was then calculated through air-water interface. Rrs was calculated using $L_w(0+)$ normalized by $E_d(0+)$ measured during the same time interval as $L_u(z)$.)

The COASTlOOC Data

The second dataset originates from the coastal surveillance through observation of ocean color (COASTlOOC) project [49], and it includes measurements from both coastal and open oceans around Europe. Water samples and coincident radiometric measurements were collected from complex aquatic environments, such as river plums and coastal waters, and open ocean. Irradiance was converted to Rrs by using a Q-factor 3.8 as described in [50]. Note, Chl-a measurements were obtained following the same procedure as in case of the Arctic data, whereas Rrs was derived by a different technique.

The Merged Global Data

The COASTlOOC and Arctic datasets were merged to build a comprehensive dataset, including Arctic, coastal, and some open waters. This dataset was used for optimizing, training, evaluating, and validating the ML GPR Balaton model. Hence, the total number of samples available for training and validation is 521, and overall Chl-a range is 0.02 and 29.07 mg m^{-3}. The Rrs data included only the three spectral bands used by the ML GPR Balaton model. The model using only the three selected spectral bands was evaluated and compared with a model, when all the available spectral bands were used for estimating Chl-a [26]. The 3 bands model resulted in improvements in all the computed statistical measures. Figure 2 shows the Rrs of the Arctic and COASTlOOC datasets for all the available bands corresponding to OLCI and for the three selected bands. Note that although there is a loss of information, when the number of bands are reduced to three, using only the selected three most relevant bands can improve Chl-a retrieval. This might be due to the fact that only these selected

three bands contain the information needed for Chl-a estimation in the ML GPR Balaton model for complexity-diverse waters.
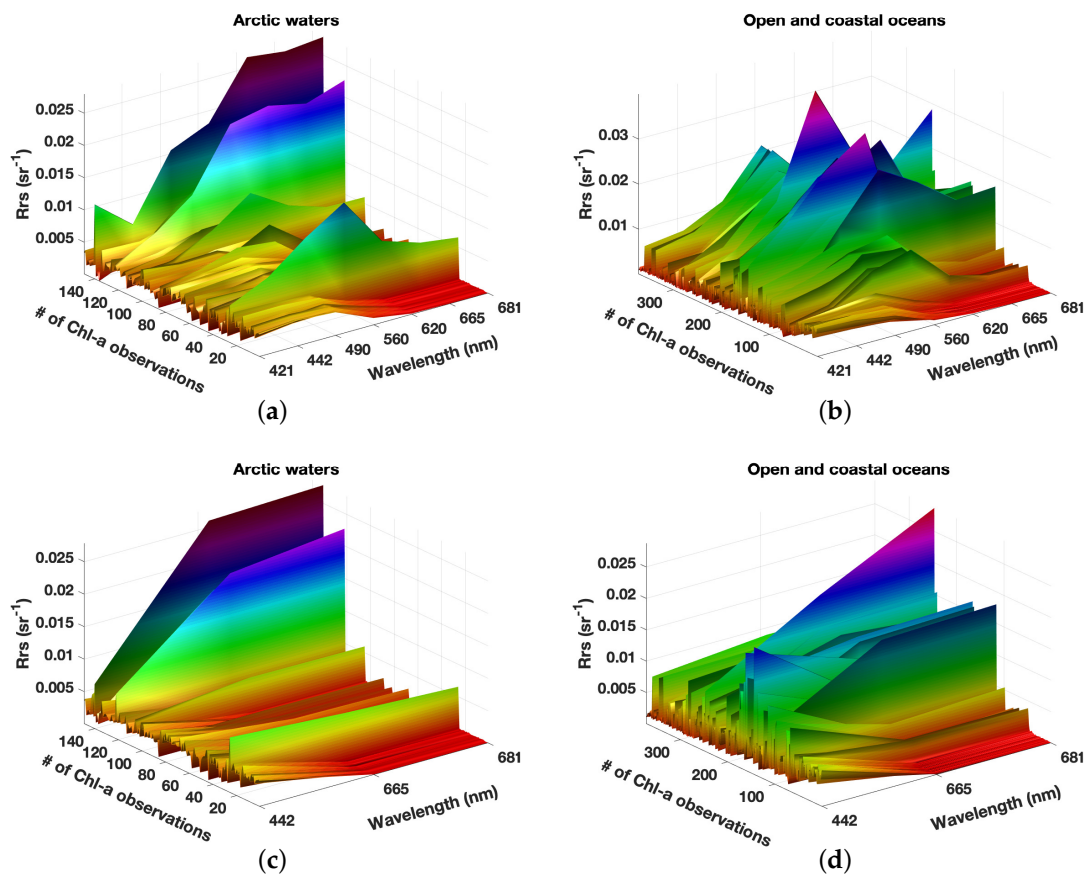
Figure 2. Top: The Rrs spectra of the Arctic (**a**) and COASTlOOC dataset (**b**) measured on all the available OLCI bands. **Bottom**: The Rrs spectra of the Arctic (**c**) and COASTlOOC dataset (**d**) measured on the three most important OLCI bands from the Balaton model. The term "# of Chl-a observations" refers to the observed Chl-a concentration, which is sorted by an increasing order. Hence, measurements corresponding to low numbers are representing low Chl-a concentrations, while observations labeled with larger numbers are high Chl-a concentrations.

Total Chl-a Measurements

High Performance Liquid Chromatography (HPLC) analysis was used in this study. First, 25 mm GF/F filters were used to filter phytoplankton pigments samples, then certain process were taken including filters extracted in 100 % methanol, disrupted by sonication and filtered (GF/F Whatman) before analyzing it by using HPLC on the same day. For the Arctic data, the detailed protocols of MALINA, TARA-Arctic, and GREEN EDGE samples followed the analytical procedure described in [51], and the method by [52] was applied to ICESCAPEs samples. (For further details on the dataset see for example [53].) HPLC-determined concentration of total Chl-a was determined as the sum of mono- and divinyl Chl-a concentration, Chlorophyllide-a and the allomeric and epimeric forms of Chl-a. The range of Chl-a was between 0.02 and 10.11 mg m^{-3} for the Arctic data, and it was between 0.05 and 29.07 mg m^{-3} for the COASTlOOC data, thus representative for various water conditions.

2.1.2. Data for Prediction with the Trained ML GPR Balaton Model

We evaluated the trained ML GPR Balaton model on a swath acquired by S3 OLCI including complex sub-Arctic and Arctic waters. The locations of the study sites can be seen in Figures 3 and 4.

The area indicated as number 1 includes a part of the St. Lawrence river, a smaller river, and a lake. The RGB image indicates various degrees of complexity for this area. Site 2 is the delta of the St. Lawrence river, which is known to have high turbidity and rapidly changing water complexity [54]. Site 3 and 4 are two lakes in Quebec: Lake Mistassini and Lac Saint-Pierre. These chosen sites provide the possibility to test the performance of the ML GPR Balaton model on optically highly complex inland waters. For instance, it can be seen in Figure 4, site 3 (Lake Mistassini) gray features appearing along the shore of the lakes. These features are not likely to be areas with high Chl-a concentration, they are possibly shallower and sediment-dominated patterns. Therefore, they can be used to evaluate the sensitivity of the ML GPR Balaton model to various degree of turbidity.
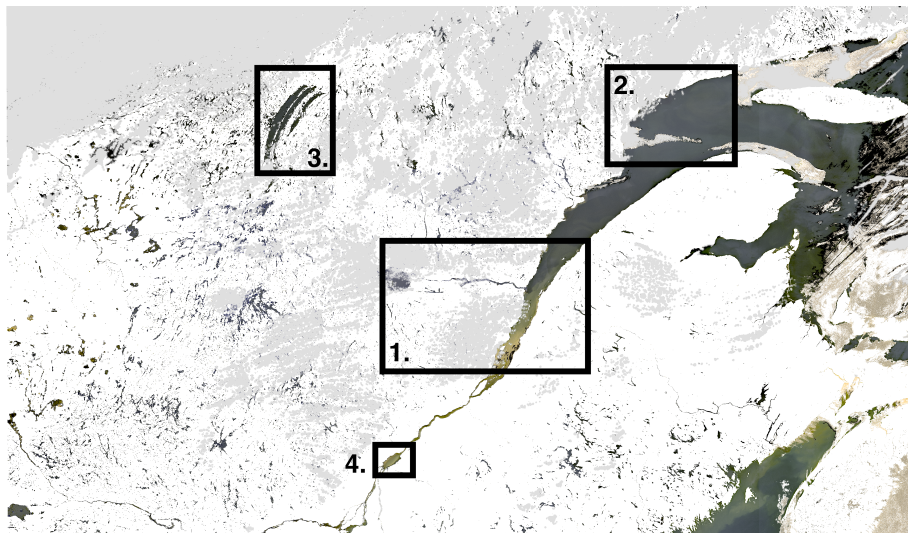
Figure 3. Inland waters in Quebec province. The RGB image was acquired at 8 June 2017 by Sentinel 3 OLCI. Cloud masks are indicated with gray. The four investigated areas are shown in the rectangles.
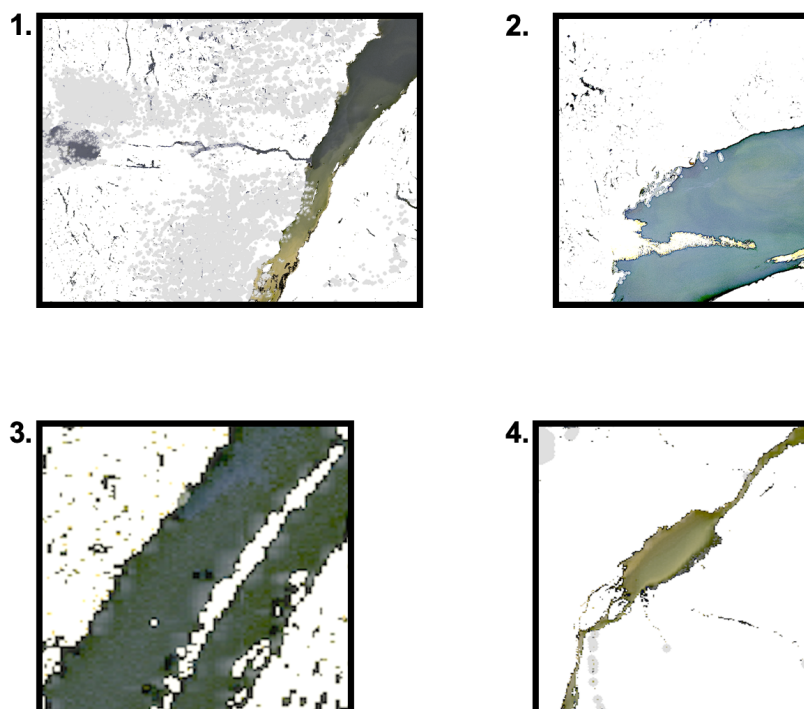
Figure 4. The location of the investigated waters.

These areas were used to test the performance of the unified ML GPR Balaton model for Chl-a estimation in highly complex waters by using the Rrs values on the three wavelengths (442.5, 665, and 681.25 nm). The Rrs data was atmospherically corrected Level 2 (L2) reflectances, on 300 m full resolution. The processing of the data was implemented in the Sentinel Application Platform (SNAP). The Level 2 Rrs data was produced by using the C2RCC atmospheric correction algorithm. The recommended flags were used to mask out invalid pixels. Then this data was used to predict Chl-a concentration with the trained ML GPR Balaton model for these complex waters. This prediction step was performed in the MATLAB toolbox.

2.2. Methodology

The ML GPR Balaton model was obtained on Lake Balaton by using the Automatic Model Selection Algorithm (AMSA) [40]. Lake Balaton is a complexity-diverse aquatic environment, including eutrophic, mesotrophic, oligotrophic, turbid, and clear water conditions. Following recent findings, the model has great potential to show similar performance, when applied to high northern latitude complex waters as well. (For further details on the establishing of the ML GPR Balaton model we refer to [26].) We optimized and evaluated the ML GPR Balaton model on optically complex waters of the Arctic and COASTlOOC datasets. The approach and the interpretation of the methodology are presented in Sections 2.2.1 and 2.2.2.

2.2.1. The ML GPR Balaton Model

Assume a training dataset, which in this case is the merged Arctic and COASTlOOC datasets, consisting of the output y_n Chl-a measurements and the corresponding input $\mathbf{x}_n$ Rrs, where $n = 1, \ldots N$ is the number of measurements, $\mathbf{x}_n$ is the three-dimensional input Rrs vector, and the dimensions are the wavelengths centered at 442.5, 665, and 681.25 nm. The Chl-a is assumed to be a function of the Rrs(λ_i), for $i = 1$, 2 and 3, and the noise ϵ, and can be written by

$$y_n = f(\mathbf{x}_n) + \epsilon_n, \quad \text{i.e.} \quad \epsilon_n \sim N(0, \sigma^2). \tag{1}$$

The noise follows a normal distribution (N) with zero mean and σ^2 variance. Then a zero mean Gaussian Process (*GP*) prior is placed over the function values $f(\mathbf{x}_n)$ and the noise, i.e., $f(\mathbf{x}) \sim GP(\mathbf{0}, k(\mathbf{x}, \mathbf{x}'))$. The term $k(\mathbf{x}, \mathbf{x}')$ is the covariance matrix, and the elements of the covariance matrix are computed by the kernel function. By definition, observations drawn from a *GP* at locations $\{\mathbf{x}_n\}_{n=1}^{N}$, also follow a joint Gaussian distribution [55]. Using Bayes theorem, the posterior distribution for a new output y_* can be analytically derived conditioned on the training data D and new input observations $\mathbf{x}_*$. This can be expressed by $p(y_* | \mathbf{x}_*, D) = N(y_* | \mu_{\text{GP}*}, \sigma_{\text{GP}*}^2)$, where $\mu_{\text{GP}*}$ is the predicted Chl-a and $\sigma_{\text{GP}*}^2$ is the certainty level of the estimate. The new input observations are the Rrs(λ_i) for either the training data (in the model evaluation) or the L2 OLCI Rrs over high northern latitude complex waters (in prediction). Then the estimated Chl-a can be expressed by

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

and the variance (certainty level of the estimated Chl-a concentration) is

$$\sigma_{\text{GP}*}^2 = \sigma^2 + k_{**} - \mathbf{k}_{\mathbf{f}*}^\top (\mathbf{K}_{\mathbf{ff}} + \sigma^2 \mathbf{I}_n)^{-1} \mathbf{k}_{\mathbf{f}*}, \tag{3}$$

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 *GP* mean, k_{**} is the covariance between the test point with itself, and $\mathbf{K}_{\mathbf{ff}} + \sigma^2 \mathbf{I}_n$ is the noise contaminated covariance matrix of the training data, where $\mathbf{I}_n$ is the identity matrix.

The elements of the covariance matrices are computed with the commonly used Squared Exponential (SE) kernel function. This can be expressed by $k(\mathbf{x}_p, \mathbf{x}_q) = \nu^2 \exp(-0.5 \sum_{d=1}^{D} ((x_p^d - x_q^d)/(\lambda_d))^2)$ for element p and q, where the hyper-parameters ν and λ are the scaling factor and

length-scales, respectively, and $d = 1, \ldots D$ is the dimension. In this case, $D = 3$, corresponding to the three wavelengths. The total number of hyper-parameters in this case is 5, which includes the scaling factor, length-scales and the noise variance. These hyper-parameters were optimized by maximizing the negative log-marginal likelihood function with respect to the hyper-parameters. In this work, we used Bayesian optimization.

2.2.2. Interpretation of the ML GPR Balaton Model

The ML GPR Balaton model assumes that Chl-a concentration is a function of the Rrs measured at the three wavelengths and the additional noise, i.e., Chl-a $= f(\text{Rrs}(\lambda_i)) + \epsilon$, for $\lambda_i = 442.5, 665$, and 681.25. The Rrs is furthermore a function of the water leaving radiance L_w, $\text{Rrs}(\lambda_i) = g(L_w(\lambda_i))$. This functional relationship (g) can be written by $\text{Rrs}(\lambda_i) = (L_w(\lambda_i)/F_0 f_s \cos(\theta_s) t_{ds}) f_b(\lambda_i) f_{(\lambda_i)}$, where F_0 is the extraterrestrial solar irradiance, f_s is a factor to adjust F_0 for the variation in Earth-Sun distance, t_{ds} is the total downward transmittance of the atmosphere, f_b is the bidirectional reflectance correction factor, and $f_{(\lambda_i)}$ is the correction for out-of-band response. L_w is retrieved from the measured total Top Of Atmosphere (TOA) radiance by using atmospheric correction algorithm [56–59]. In this case, the C2RCC algorithm was used.

Hence, Chl-a $= f(g(\mathbf{L}_w(\boldsymbol{\lambda}))) + \epsilon$. The water leaving radiance contains the spectral characteristics of the water bodies, and in our model three wavelengths are used to describe the Chl-a pigment. Furthermore, the function f that relates Chl-a to L_w through the Rrs, can be learned and analytically expressed. This can be interpreted by collecting all the available measured $\mathbf{Rrs}(\boldsymbol{\lambda})_n$ for $n = 1, \ldots, N$, and connecting them in the spectral space by a *GP*, so that the non-linear function of $\mathbf{Rrs}(\boldsymbol{\lambda})_n$ will be jointly Gaussian distributed. The spectral distance between the observed $\mathbf{Rrs}(\boldsymbol{\lambda})_n$ is controlled through the kernel function, more specifically, in case of the SE kernel, the length-scale hyper-parameter provides the measure for the similarity. Small values of the optimized length-scales indicate that the observations are close to each other in the spectral space, while large values show little spectral similarity. This length-scale hyper-parameter is used to parametrize the kernel function, which is used to compute the elements of the covariance matrices in the ML GPR Balaton model, and these covariance matrices are used in the final expression of the estimated Chl-a concentration (Equation (2)). Hence, the approach can be tracked, analytically expressed, and interpreted.

The additional automatic output of the ML GPR model, is the variance (certainty level) of the estimates (Equation (3)). This is a highly advantageous property, since it is independent of the Chl-a measurements, and it reveals whether the new observed Rrs is similar to the input data used in the training process.

Note, the L2 Rrs, which is the function of $\mathbf{L}_w(\boldsymbol{\lambda})_n$ is obtained by using atmospheric correction, hence the approach relies on the atmospheric correction algorithm.

2.2.3. Description of the Analysis

First, we evaluated the ML GPR Balaton model in a statistical analysis. The merged dataset was used for studying the learning strength of the approach and for cross-validation. We computed statistical measures commonly used in remote sensing [60,61]. These measures are the bias, which is the sum of the absolute mean errors, the Normalized Root Mean Squared Errors, NRMSE $= 1/(y_{\max} - y_{\min})\sqrt{(1/N)\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$, Pearson correlation coefficient, $R^2 = (\sum_{i=1}^{N}(\hat{y}_i - \overline{y})^2)/(\sum_{i=1}^{N}(y_i - \overline{y}_i)^2)$, and the *p*-value to assess the significance. N denotes the number of observations, y is the measured Chl-a concentration, $\hat{y}$ is the predicted Chl-a, $y_{\max}$ is the maximum observed value, $y_{\min}$ is the minimum observed value, and $\overline{y}$ is the mean of the observed Chl-a concentrations. Note, the NRMSE and R^2 are unitless, and the unit of the bias is mg m^{-3}.

Then an image acquired by S3 OLCI was chosen, when cloud coverage was relatively low, for assessing the predictive strength of the approach in application. The swath includes high northern latitude complex waters: lakes, rivers, and the estuary of the St. Lawrence river. Hence, it provides an excellent site to test the method on various complex high northern latitude waters. In addition to

estimating Chl-a concentration by the ML GPR Balaton model, we also present the certainty levels of the estimates, and the state-of-the-art L2 Chl-a products computed by using NNs and OC4 approaches. Note, although the state-of-the-art complex water product is obtained by the NN, we included the estimates of the OC4 algorithm as well, due to the general robustness of the approach and uncertainties in the composition of water constituents of the underlying water bodies.

Figure 5 shows the flowchart of the approach. The 3 bands ML GPR Balaton model, which was established on data from Lake Balaton, was tuned and optimized on the Arctic and COOSlOOC datasets. We evaluated the 3 bands ML GPR Balaton model by using bootstrapping in 1000 iterations. The final model was optimized on all the available data, which was the merged Arctic and COOSTlOOC datasets. Finally, the optimized model was studied for Chl-a estimation and uncertainty assessment on data acquired by S3 OLCI over high northern latitude Arctic waters.
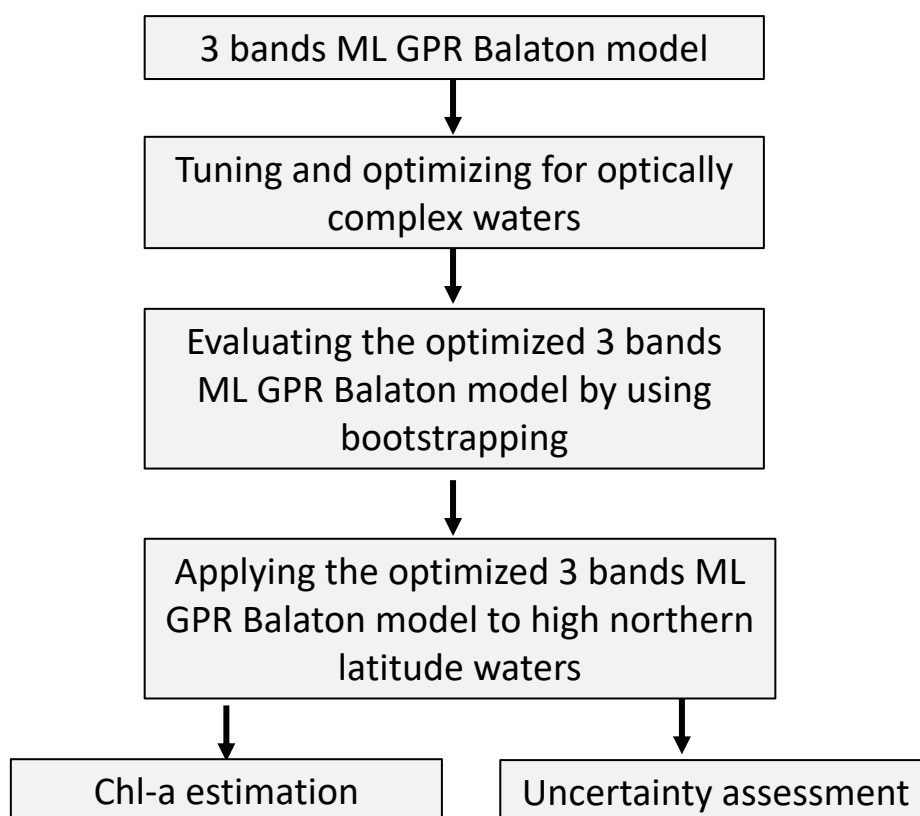
```
        ┌─────────────────────────────────┐
        │   3 bands ML GPR Balaton model   │
        └─────────────────────────────────┘
                        │
                        ▼
        ┌─────────────────────────────────┐
        │  Tuning and optimizing for       │
        │  optically complex waters        │
        └─────────────────────────────────┘
                        │
                        ▼
        ┌─────────────────────────────────┐
        │  Evaluating the optimized 3      │
        │  bands ML GPR Balaton model by   │
        │  using bootstrapping             │
        └─────────────────────────────────┘
                        │
                        ▼
        ┌─────────────────────────────────┐
        │  Applying the optimized 3 bands  │
        │  ML GPR Balaton model to high    │
        │  northern latitude waters        │
        └─────────────────────────────────┘
            │                        │
            ▼                        ▼
  ┌──────────────────┐   ┌──────────────────────┐
  │  Chl-a estimation │   │ Uncertainty assessment│
  └──────────────────┘   └──────────────────────┘
```

Figure 5. Flowchart of the approach.

3. Results

3.1. Statistical Analysis

The trained ML GPR Balaton model was evaluated by performing cross-validation (bootstrapping). The dataset was randomly divided into 90 % for training and 10 % for testing. The statistical measures were computed for the test set for 1000 iterations. At each iteration step, a new model was optimized for the randomly chosen training data, and then tested on the rest of the observation. This assured randomness, hence revealing weaknesses and strengths of the approach. The distribution of the computed statistical measures for the 1000 cross-validations can be seen in Figure 6.

The histograms show the distribution of the computed statistical measures for the test set for the 1000 iterations. The distribution of the NRMSE is skewed towards lower values (Figure 6a). Most of the computed values are distributed around 0.125. The bias Figure 6b seems to follow a normal distribution with a mean value around 1.3 mg m^{-3} for optically complex waters. The histogram

of R^2 Figure 6c also shows skewed distribution, and most of the computed values are close to 0.9. The p-value Figure 6d indicates high significance. The histograms show the performance of the 3 bands ML GPR Balaton model on randomized test data. The aim of this study is to understand how the proposed model performs on various test data, for instance, when the test data differs from the training data. This way the bootstrapping procedure provides information about the statistics of the regression performance measures, whereas dividing the data into one training and one test set gives us only one value of these measures. The heights (*y*-axis) of the bars of the histograms show the occurrence frequencies of the values of the statistical measures, indicated on the *x*-axis.
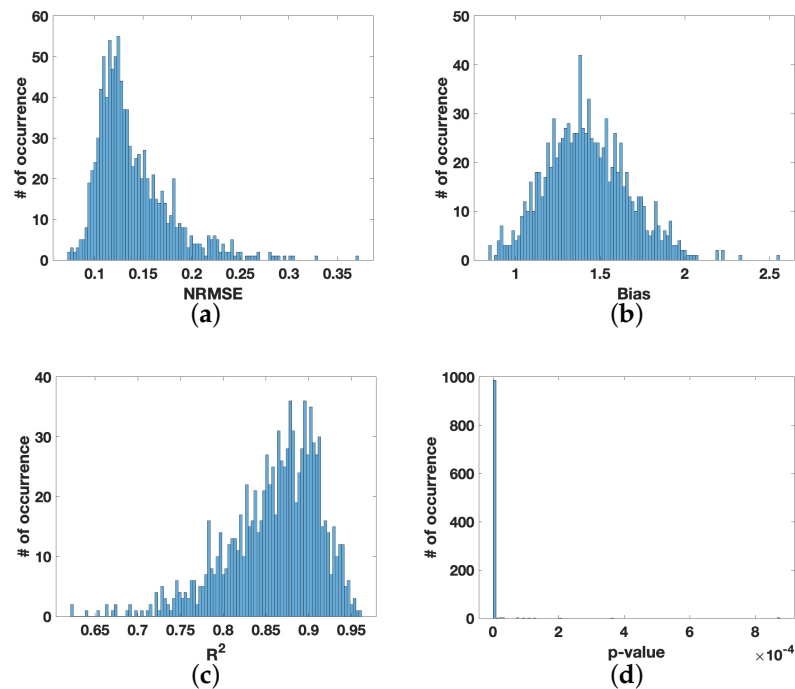
Figure 6. Results of the cross validation for the computed statistical measures for 1000 iterations of the merged Arctic and COASTlOOC datasets: the NRMSE (**a**), Bias (**b**), R^2 (**c**) and p-value (**d**).

Figure 7 shows the boxplot and medians of the computed measures of the 1000 bootstrap samples. The edges of the boxes indicate the 25th (bottom edge) and 75th (top edge) percentiles. The red line in the boxes is the median. The whiskers show the minimum (bottom whisker) and maximum (top whisker) values.
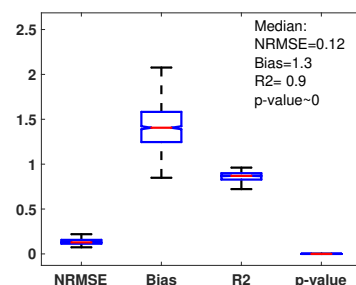
Figure 7. Boxplot of the computed statistical measures of the cross-validation for the computed statistical measures for 1000 iterations of the merged Arctic and COASTlOOC datasets.

3.2. Chl-a Maps in High Northern Latitude Complex Waters

We evaluated the ML GPR Balaton model on the four regions as defined in Section 2.1.2 (Figures 3 and 4). We computed the Chl-a estimates and the corresponding certainty maps pixelwise. Blue-green pixels in the certainty maps correspond to relatively high confidence, and yellow-red pixels show relatively lower certainty. These low certainty pixels reveal areas, where the new observed input Rrs data differs from the data used for obtaining the trained ML GPR Balaton model. At the same time, higher certainty shows regions, where the input Rrs data is similar to the one used for training. Hence, these areas will likely have similar accuracy as the computed measures in the cross-validation study.

Figure 8 shows histograms of the in situ Rrs (left column) used for training the ML GPR Balaton model and the L2 OLCI Rrs (middle and right column) used for predicting Chl-a. The L2 OLCI Rrs seems to differ from the in situ measurements, and it includes negative values. Areas in the predicted maps, corresponding to these deviations are expected to show low certainties. However, the range of the in situ Rrs in the L2 OLCI Rrs (Figure 8 right column) reveals that most of the pixels are distributed in the range of the in situ values, and these histograms show similar distribution to the histograms of the in situ Rrs.
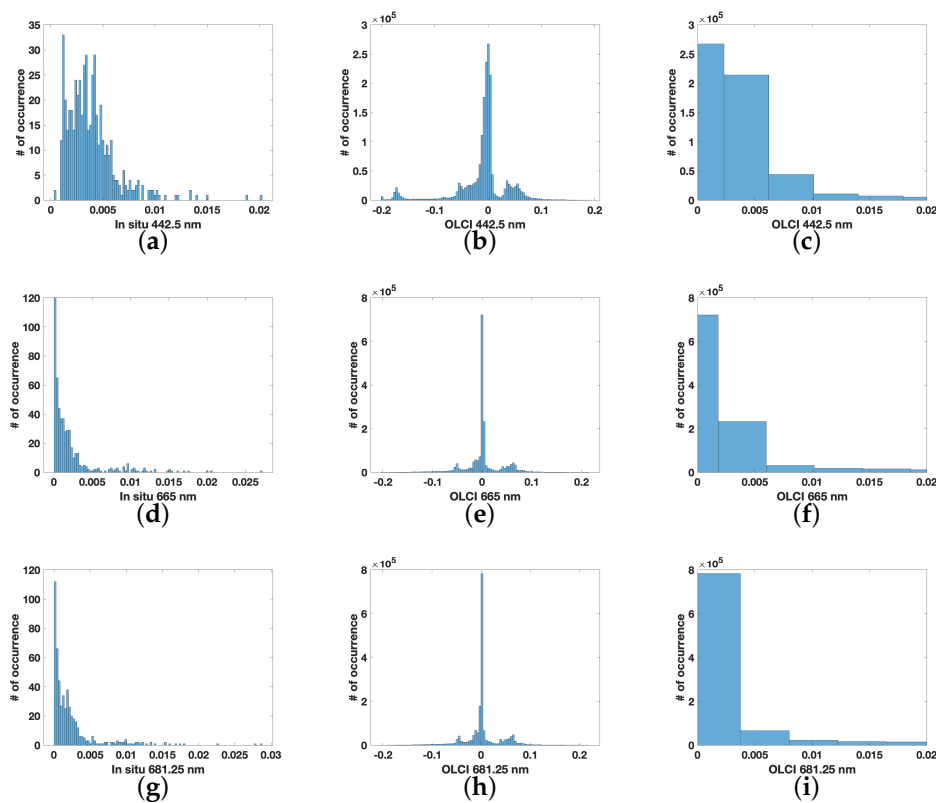
Figure 8. The histograms of the in situ Rrs used for training the ML GPR Balaton model (**a**,**d**,**g**), L2 OLCI Rrs (**b**,**e**,**h**) and L2 OLCI Rrs presented only the range corresponding to the in situ values (**c**,**f**,**i**).

Figure 9 shows the histograms of the in situ Chl-a used for training the ML GPR Balaton model (a), the estimated Chl-a by using the ML GPR Balaton model (b), the L2 OLCI NN complex waters product (c) and the L2 OLCI open water product (d). Most of the values are situated in the 0–10 mg m^{-3} range. The range of the estimates varies for the three methods, the NN does not provide Chl-a estimates above 25 mg m^{-3}, and the OC4 range is between 0 and 30 mg m^{-3}. Although most of the estimated values are also between 0 and 30 mg m^{-3} for the ML GPR Balaton model, it has some values around 40–50 mg m^{-3}, exceeding the upper limit of the values used for training. This suggests that input Rrs data differed from data used in the training process (as seen in Figure 8), and the assigned higher

values to the Chl-a concentration are based on the learned functional relationship. It is likely to have low certainty for pixels representing this higher Chl-a range.

Figures 10–13 show the results for the four study sites. Region 1 includes a small lake, a segment of the St. Lawrence river, and a flow connecting the lake with the St. Lawrence river. In general, the three approaches assign different Chl-a values for these regions. For instance, the ML GPR Balaton model shows lower estimates for regions, where the NNs estimates high Chl-a (see the lake and the lower part of the St. Lawrence river in the image). The highest certainty of the ML GPR model is observed in the lake and the upper middle part of the St. Lawrence river in the image. The estimates of the OC4 model show large contrasts in Chl-a concentrations.

The estimated Chl-a maps for region 2 can be seen in Figure 11, which is the mouth of the St. Lawrence river, and known to have a wide variety of optical properties. In this case, all three estimates show similarities in the pattern of the Chl-a concentration. The NNs estimates seem to have sensitivity to cloud shadows, which appears as areas with high Chl-a concentration (c). The ML GPR Balaton and OC4 model seem to assign similar numerical values to the Chl-a concentrations. The certainty map shows that areas by the edges of the clouds differ from the observed training input data, which is indicated by low certainty. This might suggest that the applied masks should have removed these pixels.
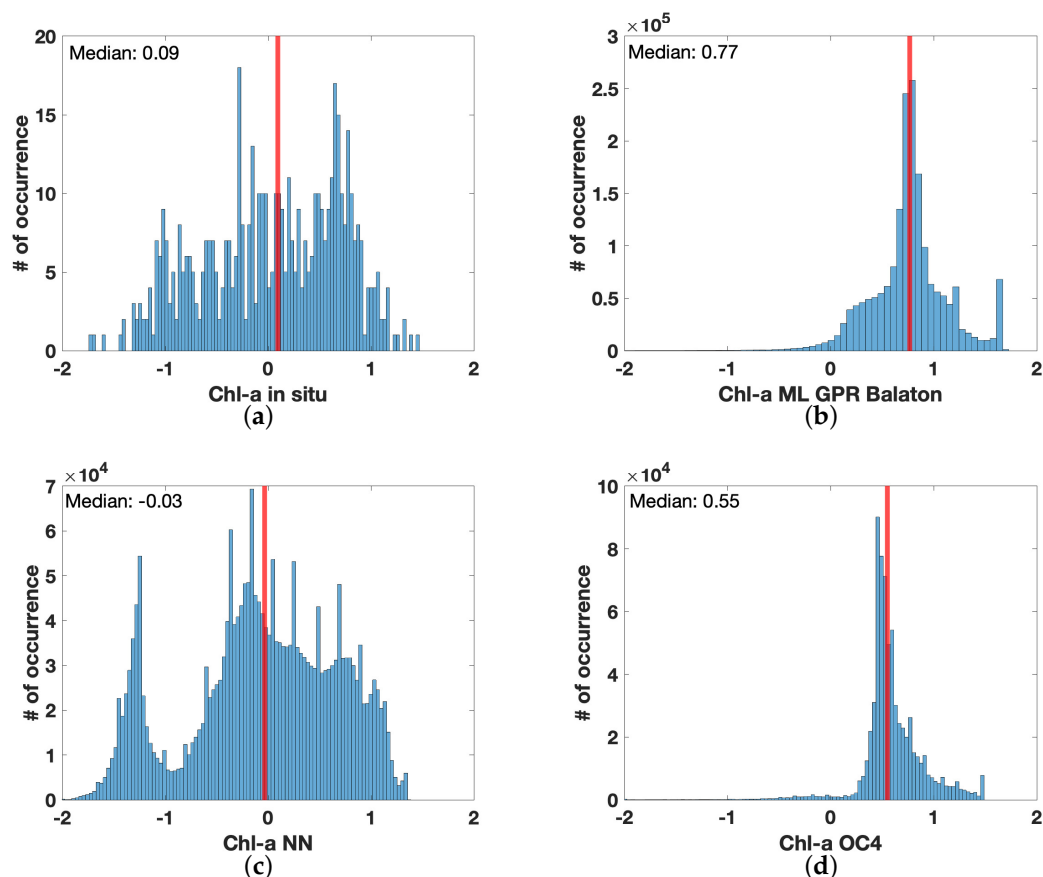
Figure 9. The histograms of the *log* transformed in situ Chl-a used for training the ML GPR Balaton model (**a**), estimated Chl-a by using the ML GPR Balaton model (**b**), NN complex water estimates (**c**) and OC4 open water estimates (**d**). The red lines indicate the median of the estimates.
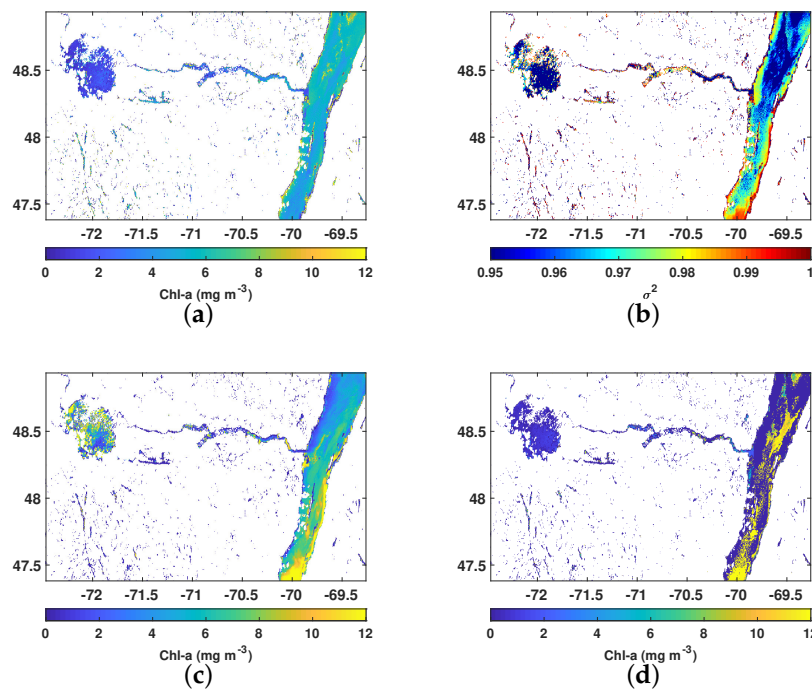
Figure 10. Region 1: Estimated Chl-a concentration (**a**) and the corresponding certainty map (**b**). Chl-a concentration map by the NNs (**c**) and OC4 (**d**). The unit of Chl-a is mg m^{-3}, and the certainty (variance) is unitless.
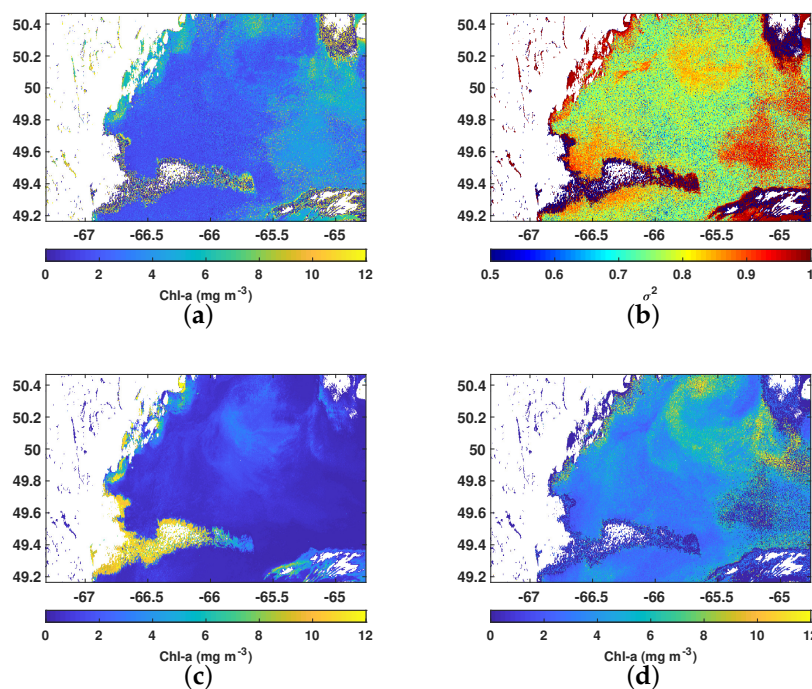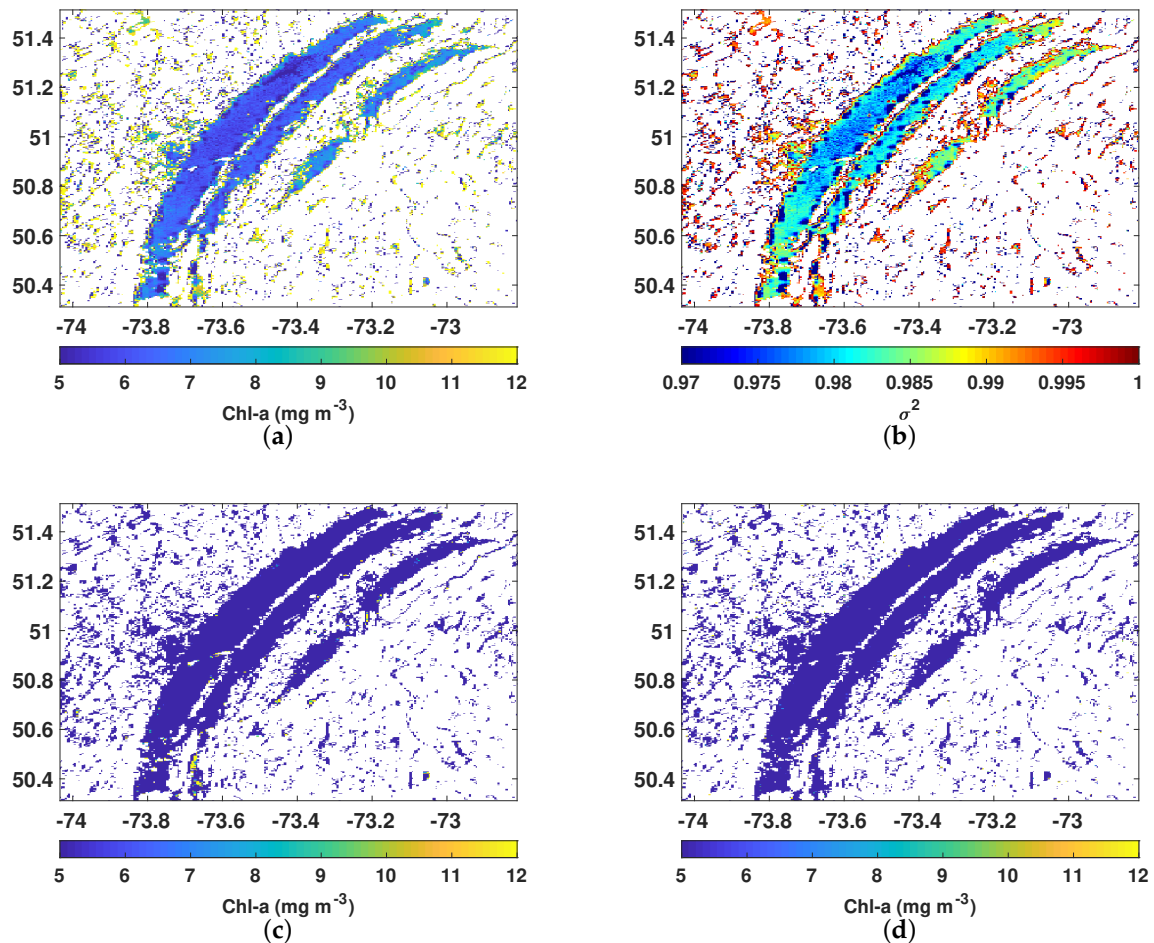

Figure 11. Region 2: Estimated Chl-a concentration (**a**) and the corresponding certainty map (**b**). Chl-a concentration map by the NNs (**c**) and OC4 (**d**). The unit of Chl-a is mg m^{-3}, and the certainty (variance) is unitless.

The RGB image for region 3, Lake Mistassini (in Figures 3 and 4) reveals some features along the shores of the lake. These features appear as grayish patterns in the RGB image, and not likely to be areas with high primary productivity. Interestingly, the ML GPR Balaton model could not only capture

these features with relatively high certainty, but it showed no sensitivity to possible suspended matter, and hence assigned low Chl-a concentration to these pixels (Figure 12). This indicates that the ML GPR Balaton model can discriminate between waters of various degrees of turbidity, and identify the signal from Chl-a in different water conditions. Note that none of the state-of-the-art L2 Chl-a concentration products could estimate Chl-a concentration in this lake.

Figure 12. Region 3, Mistassini Lake: Estimated Chl-a concentration (**a**) and the corresponding certainty map (**b**). Chl-a concentration map by the NNs (**c**) and OC4 (**d**). The unit of Chl-a is mg m^{-3}, and the certainty (variance) is unitless.

The last study site, region 4 was the Lac Saint-Pierre. Figure 13 shows the estimated Chl-a concentration maps. It can be seen that both the ML GPR Balaton model and the NNs estimated higher Chl-a values in the same areas; however, the ML GPR model assigned lower values than the NNs. Using NNs to estimate Chl-a concentration in turbid waters has been previously shown to overestimate Chl-a. Furthermore, the certainty map shows low certainty in areas, where the estimated Chl-a concentration is higher. The OC4 model seems to output a *binary-like* image for this area, which is less likely to be correct.
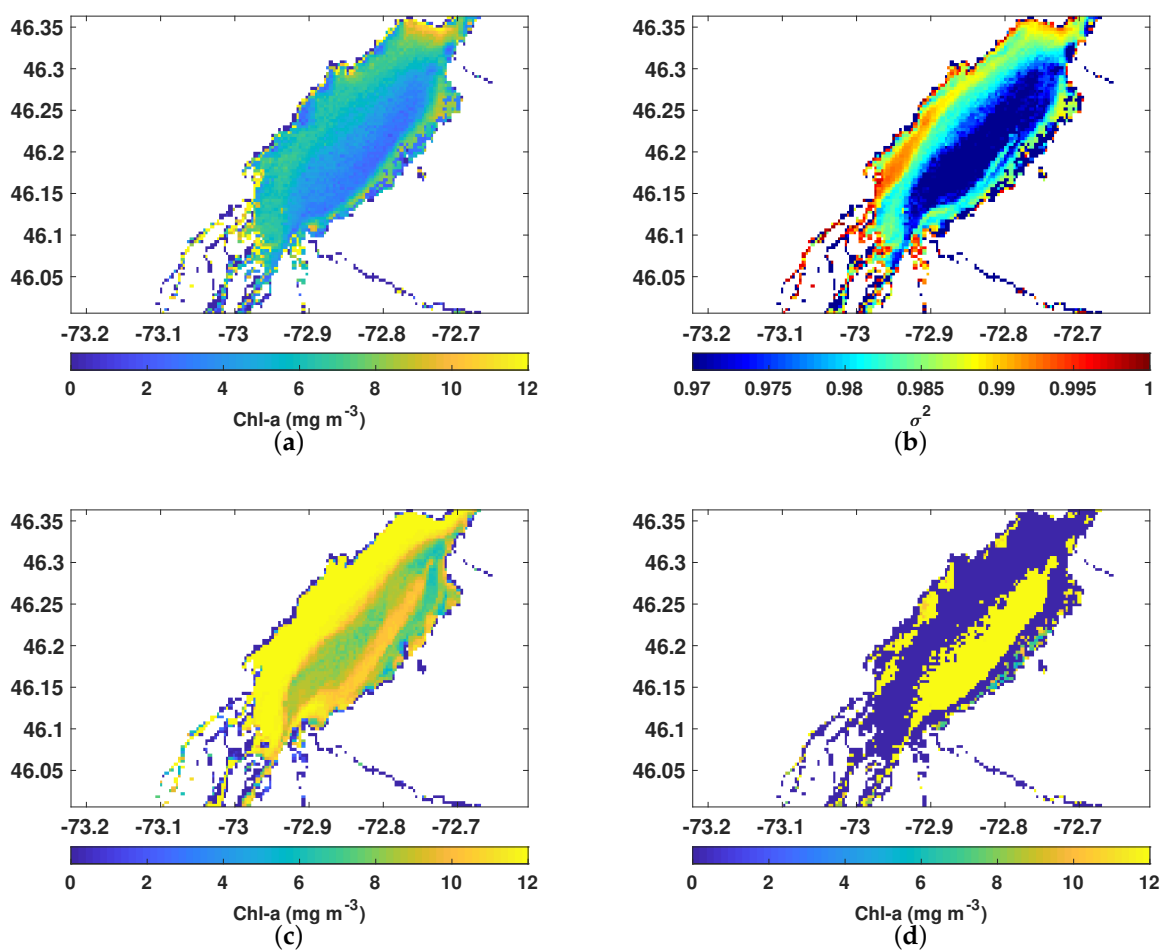
Figure 13. Region 4, Lac Saint-Pierre: Estimated Chl-a concentration (**a**) and the corresponding certainty map (**b**). Chl-a concentration map by the NNs (**c**) and OC4 (**d**). The unit of Chl-a is mg m^{-3}, and the certainty (variance) is unitless.

We have also applied the optimized ML GPR Balaton model to data acquired by S3 OLCI over the Marginal Ice Zone, coastal, and open Arctic waters. These Chl-a maps can be seen in Appendix, and they show that the approach also applies to various Arctic waters.

4. Discussions

This work aimed to evaluate the hypothesis that the Automatic Model Selection Algorithm (AMSA) can be used to establish a Chl-a concentration retrieval algorithm tailored for S3 OLCI for complexity-diverse waters. In addition, this algorithm can be generalized to perform well in high northern latitude complex waters, where the state-of-the-art approaches fail.

We optimized the unified ML GPR Balaton model on in situ radiometric measurements and corresponding Chl-a observations collected from Arctic, coastal (including river plumes) and open waters. Using in situ data for model optimization has several advantages in this case. It is independent on a certain calibration and atmospheric correction algorithm, which might change during a mission and by region. Furthermore, it can reveal uncertainties in the retrieval of satellite products. Hence, the model can contribute to improve the quality of the satellite derived water quality estimates.

We conducted a statistical analysis, which showed that ML GPR Balaton model performs well on data representing a wide variety of optical properties. Our findings (Figure 7) were consistent with previous results obtained on Lake Balaton. This suggests that the results obtained in this work are

comparable with the results in case of Lake Balaton. Most importantly, the ML GPR Balaton model seems to be robust, stable, and reliable.

Note, the ML GPR Balaton model uses three features to estimate Chl-a concentration. These features are spectral bands positioned at 442.5, 665, and 681.25 nm center wavelength. This is in contrast with the commonly used machine-learning approaches, where all spectral information (all bands) is used [24,25]. The bias significantly decreased, when only these three spectral bands were used in the ML GPR Balaton model, instead of using all the OLCI bands in the visible part of the spectral region. (For further details see [26].) These three bands not only improve Chl-a retrieval in complexity-diverse waters from S3 OLCI Rrs, but also reflect the physical properties of the water bodies. The blue band, positioned at 442.5 nm is probably the dominant band for clear, oligotrophic waters. At the same time, red bands are commonly used in turbid waters. The bands centered at 665 and 681.25 nm correspond to the second absorption peak of Chl-a and to the fluorescence band, respectively, and they are probably favored for turbid and/or eutrophic waters. Red bands are commonly used in parametric and semi-analytic Chl-a retrieval algorithms for complex waters [22,61,62].

The predicted Chl-a maps showed that the ML GPR Balaton model can estimate Chl-a concentration in complex high northern latitude waters, and capture patterns which are usually challenging for other Chl-a retrieval algorithms. The presented maps were focusing on Canadian inland waters and the St. Lawrence river estuary due to the difficulties in satellite derived Chl-a estimation in these regions. The optical complexity and the ongoing rapid transitions in water conditions due to environmental changes result in challenges in the retrieval of Level 2 water quality products. The proposed approach proceeds by using Bayesian inversion, which allows the model to reveal uncertainties in the input Level 2 products. We showed that the approach provides the pixelwise certainty level of the estimated Chl-a concentration, which is a highly advantageous property, since it can reveal areas where the acquired Rrs data differs from the data used for training the model. This means that low certainty can be due to deviations in the optical properties of the waterbodies. However, if the training data is representative, which is likely to be the case in this work, then low certainty is most likely due to erroneous Rrs retrieval of the applied atmospheric correction algorithm. Hence, the certainty maps provide a support to understand the challenges in ocean color monitoring by using S3 OLCI. These certainty maps can be used as a mask, to disregard areas with relative high uncertainties, and keep the estimates, where the computed statistical measures are valid.

Even though in this work, we presented the performance of the unified ML GPR Balaton model on high northern latitude complex inland waters, we also computed prediction maps for S3 OLCI Rrs data acquired over the Marginal Ice Zone (MIZ) and sub-Arctic/ Arctic coastal and open oceans. These results can be seen in the Appendix, and they also support that the model can be applied to various Arctic waters, with different degrees of optical complexity. It is advantageous to have a unified Chl-a retrieval algorithm for S3 OLCI, because it is often challenging to assess information about the type of water to be monitored in advance. Furthermore, rapid changes in Arctic waters can also change the type of water with time, causing difficulties to choose the correct type of S3 OLCI L2 Chl-a product. This work shows a possible solution to overcome these difficulties.

5. Conclusions

It can be concluded that the ML GPR Balaton model performs well and has the potential to be used for operational purposes for complex sub-Arctic/ Arctic waters. The cross-validation (bootstrapping) study resulted in a median of the normalized root mean squared error of 0.13, a median of the bias of 1.3 mg m^{-3}, a median of the Pearson correlation coefficient of 0.91, and a low *p*-value, indicating high significance. With regard to the wide variety of the data used for optimizing the ML GPR Balaton model, these computed measures would be valid for a new dataset, under the assumption that the atmospheric correction algorithm is properly calibrated. Using in situ data for model tuning and optimization instead of matchups allows the approach to be independent of the atmospheric correction algorithm used to produce the satellite Level 2 data.

We showed how the ML GPR Balaton approach can be a tool to locate and assist in solving calibration issues, with the use of the certainty maps. Low certainty can help to locate areas, where Rrs values differ from the representative training data. Hence, this work not only presents a candidate unified machine-learning model for monitoring a wide range of water conditions, but also shows how the approach can be used to locate waters, where calibration issues might need to be addressed.

For future work, we will automate the approach to detect uncertain areas. This will be done by defining a threshold from uncertainty estimates, and disregard values above the threshold. This would yield an uncertainty mask based on an analytical approach. Further studies include introducing the approach for other water quality parameters.

Appendix A. Chl-a Maps in the Marginal Ice Zone, Coastal, and Open Oceans

We applied the ML GPR Balaton model to two additional swaths acquired over the MIZ and coastal and open oceans. The RGB images of the two areas are presented in Figure A1.


(**a**)   (**b**)
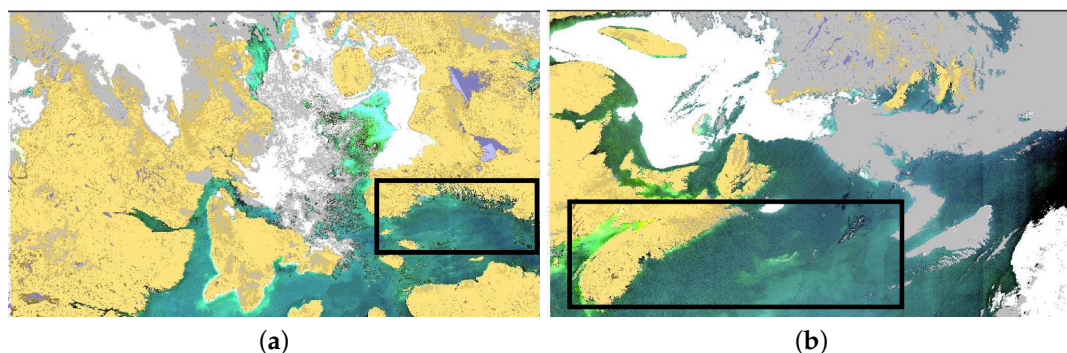
Figure A1. (**a**) RGB image of the Marginal Ice Zone, acquired on the 8th of July 2017 by Sentinel 3 OLCI. (**b**) RGB image of coastal and open oceans acquired at 10 June 2017 by Sentinel 3 OLCI. Masks: land (orange), cloud (gray) and snow-ice (white). The black squares show the selected areas for the enlarged images.

The predicted Chl-a maps by using the ML GPR Balaton model (a), the NN (c) and OC 4 (d) algorithms for the MIZ can be seen in Figure A2. It can be seen that the pattern of the estimated Chl-a concentration is similar for all the three algorithms in this case. However, the assigned Chl-a values differ. The certainty map (b) shows the relative certainty for the ML GPR Balaton model estimates.

Figure A3 shows the corresponding enlarged areas. This also shows that in this case, all three algorithms could capture the pattern of Chl-a. The difference here also manifests itself in the amount of the assigned Chl-a concentration. (The range of the units is the same as in Figure A2.) The certainty map (b) reveals high certainty in areas of the appearance of the bloom (green pixels), indicating reliable Chl-a estimates.
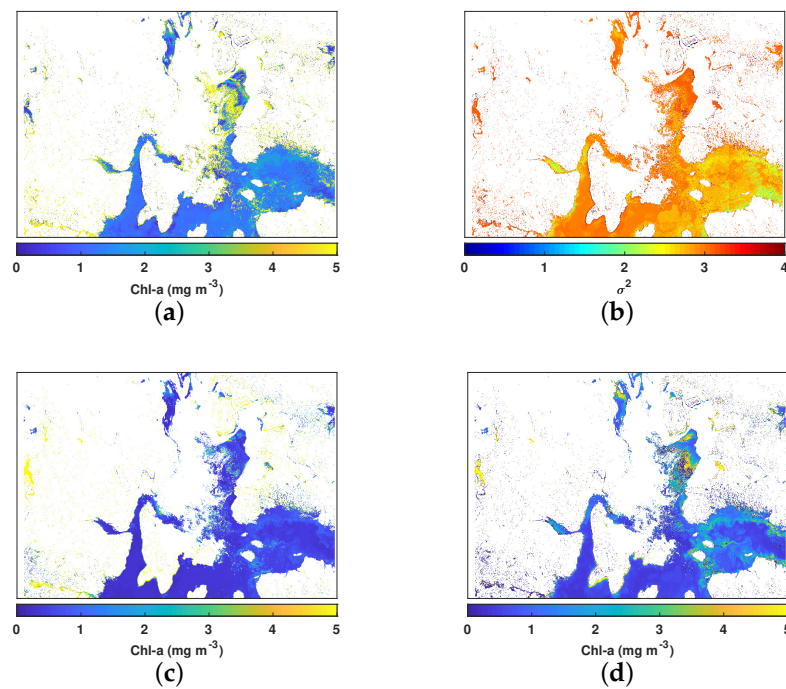
Figure A2. Estimated Chl-a in the Marginal Ice Zone: ML GPR Balaton model (**a**) and corresponding certainty map (**b**), NN (**c**) and OC4 (**d**). The unit of Chl-a is mg m^{-3}.
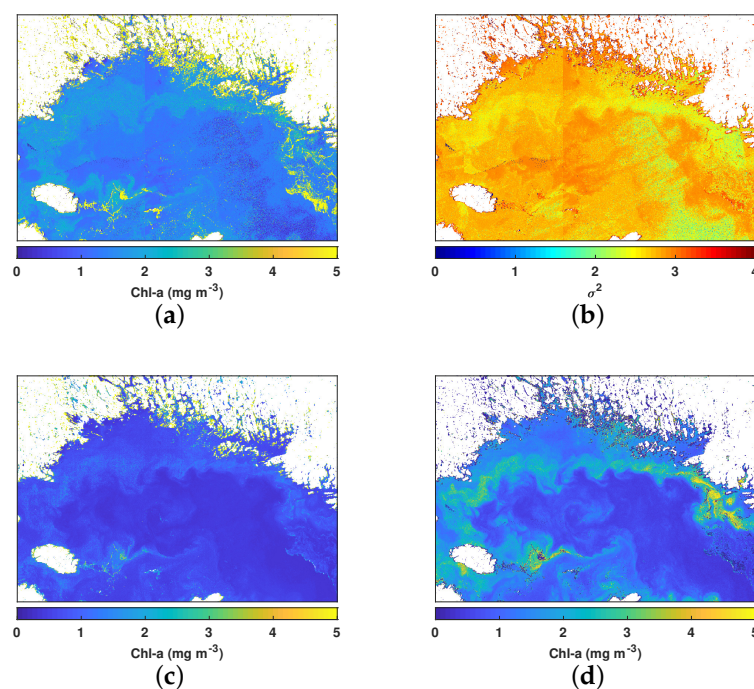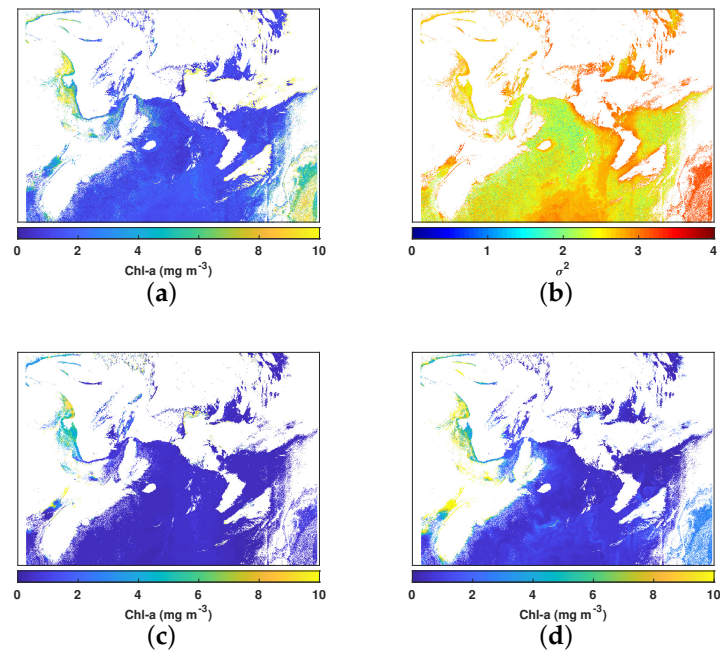

Figure A3. Estimated Chl-a in the Marginal Ice Zone for the enlarged area: ML GPR Balaton model (**a**) and corresponding certainty map (**b**), NN (**c**) and OC4 (**d**). The unit of Chl-a is mg m^{-3}.

Figures A4 and A5 show estimated Chl-a concentration for both open and coastal waters for the three algorithms in the East - Canadian sub-Arctic/ Arctic waters. For open oceans, the ML GPR Balaton model seem to estimate similar Chl-a concentration as the OC4 models. This was expected, since the OC4 model is designed for open ocean. The estimated Chl-a concentration in the coastal areas (for example bottom-left part of the images) shows larger variations between the three algorithms.

In general, the ML GPR Balaton model assigns lower Chl-a values, which is probably due to its insensitivity to TSM. The relative certainty is high, showing that input Rrs data is similar to the training data. Surprisingly, certainty decreases in the open ocean area, where there is a possible bloom. This might mean that the training data is lacking information from a similar bloom event that occurred when the image was taken.

Figure A4. Estimated Chl-a in coastal and open waters for the enlarged area: ML GPR Balaton model (**a**) and corresponding certainty map (**b**), NN (**c**) and OC4 (**d**). The unit of Chl-a is mg m^{-3}.
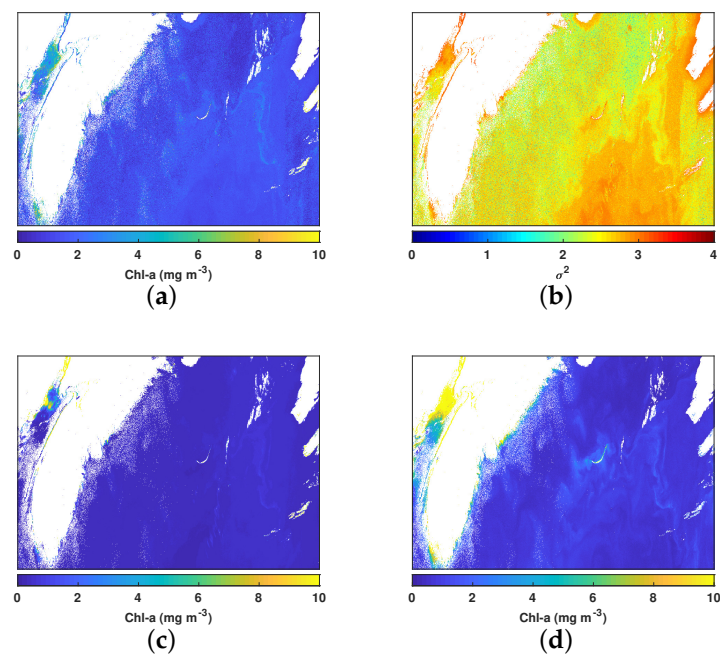
Figure A5. Estimated Chl-a in coastal and open Arctic waters: ML GPR Balaton model (**a**) and corresponding certainty map (**b**), NN (**c**) and OC4 (**d**). The unit of Chl-a is mg m^{-3}.

References

1. Meier, W.N.; Hovelsrud, G.K.; van Oort, B.E.; Key, J.R.; Kovacs, K.M.; Michel, C.; Haas, C.; Granskog, M.A.; Gerland, S.; Perovich, D.K.; et al. Arctic sea ice in transformation: A review of recent observed changes and impacts on biology and human activity. *Rev. Geophys.* **2014**, *52*, 185–217. [CrossRef]
2. Renaut, S.; Devred, E.; Babin, M. Northward Expansion and Intensification of Phytoplankton Growth During the Early Ice-Free Season in Arctic. *Geophys. Res. Lett.* **2018**, *45*, 10590–10598. [CrossRef]
3. Ardyna, M.; Babin, M.; Gosselin, M.; Devred, E.; Rainville, L.; Tremblay, J.E. Recent Arctic Ocean sea ice loss triggers novel fall phytoplankton blooms. *Geophys. Res. Lett.* **2014**, *41*, 6207–6212. [CrossRef]
4. Engelsen, O.; Hegseth, E.N.; Hop, H.; Hansen, E.; Falk-Pedersen, S. Spatial variability of chlorophyll-a on the Marginal Ice Zone of the Barents Sea, with relsations to sea ice and oceanographic conditions. *J. Mar. Syst.* **2002**, *35*, 79–97. [CrossRef]
5. Volk, T.; Hoffert, M.I. *Ocean Carbon Pumps: Analysis of Relative Strengths and Efficiencies in Ocean-Driven Atmospheric CO_2 Changes*; American Geophysical Union: Washington, DC, USA, 2013; pp. 99–110.
6. Johannessen, O.M.; Miles, M.W. Critical vulnerabilities of marine and sea ice–based ecosystems in the high Arctic. *Reg. Environ. Chang.* **2011**, *11*, 239–248. [CrossRef]
7. Arrigo, K.R.; Robinson, D.H.; Worthen, D.L.; Dunbar, R.B.; DiTullio, G.R.; VanWoert, M.; Lizotte, M.P. Phytoplankton Community Structure and the Drawdown of Nutrients and CO_2 in the Southern Ocean. *Science* **1999**, *283*, 365–367.
8. Hein, M.; Sand-Jensen, K. CO_2 increases oceanic primary production. *Nature* **1997**, *388*, 526–527. [CrossRef]
9. Hofmann, M.; Worm, B.; Rahmstorf, S.; Schellnhuber, H.J. Declining ocean chlorophyll under unabated anthropogenic CO_2 emissions. *Environ. Res. Lett.* **2011**, *6*, 034–035. [CrossRef]
10. Bird, K.J.; Charpentier, R.R.; Gautier, D.L.; Houseknecht, D.W.; Klett, T.R.; Pitman, J.K.; Moore, T.E.; Schenk, C.J.; Tennyson, M.E.; Wandrey, C.J.; et al. *Circum-Arctic Resource Appraisal: Estimates of Undiscovered Oil and Gas North of the Arctic Circle*; U.S. Geological Survey Fact Sheet 2008; U.S. Geological Survey: Denver, CO, USA, 2008; pp. 1175–1179.
11. Jacobsen, S.R.; Gudmestad, O.T. Evacuation From Petroleum Facilities Operating in the Barents Sea. In Proceedings of the ASME 2012 31st International Conference on Ocean, Offshore and Arctic Engineering, Rio de Janeiro, Brazil, 1–6 July 2012; Volume 6, pp. 457–466.
12. Melia, N.; Haines, K.; Hawkins, E.; Day, J.J. Towards seasonal Arctic shipping route predictions. *Environ. Res. Lett.* **2017**, *12*, 084005. [CrossRef]
13. Dawson, J.; Johnston, M.; Stewart, E. Governance of Arctic expedition cruise ships in a time of rapid environmental and economic change. *Ocean. Coast. Manag.* **2014**, *89*, 88–99. [CrossRef]
14. Choudhury, S.B.; Jena, B.; Rao, M.V.; Rao, K.H.; Somvanshi, V.S.; Gulati, D.K.; Sahu, S.K. Validation of integrated potential fishing zone (IPFZ) forecast using satellite based chlorophyll and sea surface temperature along the east coast of India. *Int. J. Remote. Sens.* **2007**, *28*, 2683–2693. [CrossRef]
15. Hommedal, S.; Lorentzen, E.A. What We Know about the So-Called Killer Alga in Northern Norway. 2019. Available online: http://www.imr.no/en/hi/news/2019/may/what-we-know-about-the-so-called-killer-alga-in-northern-norway (accessed on 3 September 2019).
16. Wauthy, M.; Rautio, M.; Christoffersen, K.S.; Forsström, L.; Laurion, I.; Mariash, H.L.; Peura, S.; Vincent, W.F. Increasing dominance of terrigenous organic matter in circumpolar freshwaters due to permafrost thaw. *Limnol. Oceanogr. Lett.* **2018**, *3*, 186–198. [CrossRef]
17. MODIS-Aqua. Available online: https://modis.gsfc.nasa.gov/ (accessed on 25 June 2019).
18. VIIRS. Available online: https://jointmission.gsfc.nasa.gov/ (accessed on 3 September 2019).
19. Landsat-8 OLI. Available online: https://landsat.gsfc.nasa.gov/operational-land-imager-oli/ (accessed on 3 September 2019).
20. Sentinel 2 MSI. Available online: https://sentinel.esa.int/web/sentinel/missions/sentinel-2 (accessed on 3 September 2019).
21. Sentinel 3 OLCI. Available online: https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-3 (accessed on 3 September 2019).
22. O'Reilly, J.E.; Maritirena, S.; Mitchell, B.G.; Siegel, D.A.; Carder, K.L.; Garver, S.A.; Kahru, M.; McClain, C. Ocean color chlorophyll algorithms for SeaWiFS. *J. Geophys. Res.* **1998**, *103*, 24937–24953. [CrossRef]

23. Morel, A.; Claustre, H.; Antoine, D.; Gentili, B. Natural variability of bio-optical properties in Case 1 waters: attenuation and reflectance within the visible and near-UV spectral domains, as observed in South Pacific and Mediterranean waters. *Biogeosciences* **2007**, *4*, 913–925. [CrossRef]

24. Doerffer, R.; Schiller, H. The MERIS Case 2 water algorithm. *Int. J. Remote. Sens.* **2007**, *28* , 517–535. [CrossRef]

25. Brockmann, C.; Doerffer, R.; Peters, M.; Kerstin, S.; Embacher, S.; Ruescas, A. Evolution of the C2RCC Neural Network for Sentinel 2 and 3 for the Retrieval of Ocean Colour Products in Normal and Extreme Optically Complex Waters. In Proceedings of the Living Planet Symposium, Prague, Czech Republic, 9–13 May 2016; ESA Special Publication: Noordwijk, The Netherlands, 2016; Volume 740, p. 54.

26. Blix, K.; Pálffy, K.; Tóth, V.R.; Eltoft, T. Remote Sensing of Water Quality Parameters over Lake Balaton by Using Sentinel-3 OLCI. *Water* **2018**, *10*, 1428. [CrossRef]

27. Fan, Y.; Li, W.; Voss, K.J.; Gatebe, C.K.; Stamnes, K. Neural network method to correct bidirectional effects in water-leaving radiance. *Appl. Opt.* **2016**, *55*, 10–21. [CrossRef]

28. Fan, Y.; Li, W.; Gatebe, C.K.; Jamet, C.; Zibordi, G.; Schroeder, T.; Stamnes, K. Atmospheric correction over coastal waters using multilayer neural networks. *Remote. Sens. Environ.* **2017**, *199*, 218 – 240. [CrossRef]

29. Cipollini, P.; Corsini, G.; Diani, M.; Grass, R. Retrieval of sea water optically active parameters from hyperspectral data by means of generalized radial basis function neural networks. *IEEE Trans. Geosci. Remote. Sens.* **2001**, *39*, 1508–1524. [CrossRef]

30. Hieronymi, M.; Müller, D.; Doerffer, R. The OLCI Neural Network Swarm (ONNS): A Bio-Geo-Optical Algorithm for Open Ocean and Coastal Waters. *Front. Mar. Sci.* **2017**, *4*, 140. [CrossRef]

31. Zhan, H.; Shi, P.; Chen, C. Retrieval of Oceanic Chlorophyll Concentration Using Support Vector Machines. *IEEE Trans. Geosci. Remote. Sens.* **2003**, *41*, 2947–2951. [CrossRef]

32. Kwiatkowska, E.J.; Fargion, G.S. 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 Trans. Geosci. Remote. Sens.* **2003**, *41*, 2844–2860. [CrossRef]

33. Camps-Valls, G.; Muñoz-Marí, J.; L. Gómez-Chova, K.R.; Calpe-Maravilla, J. Biophysical Parameter Estimation With a Semisupervised Support Vector Machine. *IEEE Geosci. Remote. Sens. Lett.* **2009**, *6*, 248–252. [CrossRef]

34. Camps-Valls, G.; Gómez-Chova, L.; Muñoz-Marí, J.; Vila-Francés, J.; Amorós-López, J.; Calpe-Maravilla, J. Retrieval of oceanic chlorophyll concentration with relevance vector machines. *Remote. Sens. Environ.* **2006**, *105*, 23–33. [CrossRef]

35. Pasolli, L.; Melgani, F.; Blanzieri, E. Gaussian Process Regression for Estimating Chlorophyll Concentration in Subsurface Waters From Remote Sensing Data. *IEEE Geosci. Remote. Sens. Lett.* **2010**, *7* , 464–468. [CrossRef]

36. Verrelst, J.; Muñoz, J.; Alonso, L.; Rivera, J.P.; Camps-Valls, G.; Moreno, J. Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3. *Remote. Sens. Environ.* **2012**, *118*, 127–139. [CrossRef]

37. Verrelst, J.; Alonso, L.; Camps-Valls, G.; Delegido, J.; Moreno, J. Retrieval of Vegetation Biophysical Parameters Using Gaussian Process Techniques. *IEEE Trans. Geosci. Remote. Sens.* **2012**, *50*, 1832–1843. [CrossRef]

38. Blix, K.; Camps-Valls, G.; Jenssen, R. Gaussian Process Sensitivity Analysis for Oceanic Chlorophyll Estimation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **2017**, *10*, 1265–1277. [CrossRef]

39. Blix, K.; Eltoft, T. Evaluation of feature ranking and regression methods for oceanic chlorophyll-a estimation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **2018**, *11*, 1403–1418. [CrossRef]

40. Blix, K.; Eltoft, T. Machine Learning Automatic Model Selection Algorithm for Oceanic Chlorophyll-a Content Retrieval. *Remote. Sens.* **2018**, *10*, 775. [CrossRef]

41. Blix, K.; Eltoft, T. A Generalized Chlorophyll-a Estimation Model for Complexity-Diverse Arctic Waters. In Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2019), Yokohama, Japan, 28 July–2 August 2019.

42. MALINA Data. Available online: http://www.obs-vlfr.fr/proof/php/malina/ (accessed on 5 March 2019).

43. ICESCAPE Data. Available online: https://seabass.gsfc.nasa.gov/ (accessed on 3 September 2019).

44. TARA Data. Available online: https://oceans.taraexpeditions.org/ (accessed on 3 September 2019).

45. GREEN EDGE Data. Available online: http://www.obs-vlfr.fr/proof/php/GREENEDGE/ (accessed on 3 September 2019).

46. Morrow, J.H.; Hooker, S.B.; Booth, C.R.; Bernhard, G.; Lind, R.N.; Brown, J. Advances in Measuring the Apparent Optical Properties (AOPs) of Optically Complex Waters. *NASA Tech. Memo* **2010**, *215856*, 42–50.

47. Hooker, S.B.; Morrow, J.H.; Matsuoka, A. Apparent optical properties of the Canadian Beaufort Sea—Part 2: 1 % and 1 cm perspective in deriving and validating AOP data products. *Biogeosciences* **2013**, *10*, 4511–4527. [CrossRef]

48. Gordon, H.R.; Ding, K. Self-shading of in-water optical instruments. *Limnol. Oceanogr.* **1992**, *37*, 491–500. [CrossRef]

49. Babin, M.; Morel, A.; Fournier-Sicre, V.; Fell, F.; Stramski, D. Light scattering properties of marine particles in coastal and open ocean waters as related to the particle mass concentration. *Limnol. Oceanogr.* **2003**, *48*, l843–859. [CrossRef]

50. Bélanger, S.; Babin, M.; Larouche, P. An empirical ocean color algorithm for estimating the contribution of chromophoric dissolved organic matter to total light absorption in optically complex waters. *J. Geophys. Res. Ocean.* **2008**, *113*. [CrossRef]

51. Ras, J.; Claustre, H.; Uitz, J. Spatial variability of phytoplankton pigment distributions in the Subtropical South Pacific Ocean: comparison between in situ and predicted data. *Biogeosciences* **2008**, *5*, 353–369. [CrossRef]

52. Heukelem, L.V.; Thomas, C.S. Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments. *Chromatogr. A* **2001**, *910*, 31–49. [CrossRef]

53. Matsuoka, A.; Boss, E.; Babin, M.; Karp-Boss, L.; Hafez, M.; Chekalyuk, A.; Proctor, C.W.; Werdell, P.J.; Bricaud, A. Pan-Arctic optical characteristics of colored dissolved organic matter: Tracing dissolved organic carbon in changing Arctic waters using satellite ocean color data. *Remote. Sens. Environ.* **2017**, *200*, 89–101. [CrossRef]

54. Massicotte, P.; Gratton, D.; Frenette, J.J.; Assani, A.A. Spatial and temporal evolution of the St. Lawrence River spectral profile: A 25-year case study using Landsat 5 and 7 imagery. *Remote. Sens. Environ.* **2013**, *136*, 433 – 441. [CrossRef]

55. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; The MIT Press: Cambridge, MA, USA, 2005.

56. Ahmad, Z.; Franz, B.A.; McClain, C.R.; Kwiatkowska, E.J.; Werdell, J.; Shettle, E.P.; Holben, B.N. New aerosol models for the retrieval of aerosol optical thickness and normalized water-leaving radiances from the SeaWiFS and MODIS sensors over coastal regions and open oceans. *Appl. Opt.* **2010**, *49*, 5545–5560. [CrossRef]

57. Bailey, S.W.; Werdell, P.J. A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote. Sens. Environ.* **2006**, *102*, 12–23. [CrossRef]

58. Gordon, H.R.; Wang, M. Influence of oceanic whitecaps on atmospheric correction of ocean-color sensors. *Appl. Opt.* **1994**, *33*, 7754–7763. [CrossRef] [PubMed]

59. Zibordi, G.; Mélin, F.; Berthon, J.F. Comparison of SeaWiFS, MODIS and MERIS radiometric products at a coastal site. *Geophys. Res. Lett.* **2006**, *33*. [CrossRef]

60. Liang, L.; Qin, Z.; Zhao, S.; Di, L.; Zhang, C.; Deng, M.; Lin, H.; Zhang, L.; Wang, L.; Liu, Z. Estimating crop chlorophyll content with hyperspectral vegetation indices and the hybrid inversion method. *Int. J. Remote. Sens.* **2016**, *37*, 2923–2949. [CrossRef]

61. Watanabe, F.; Alcântara, E.; Imai, N.; Rodrigues, T.; Bernardo, N. Estimation of Chlorophyll-a Concentration from Optimizing a Semi-Analytical Algorithm in Productive Inland Waters. *Remote. Sens.* **2018**, *10*, 227. [CrossRef]

62. Lins, R.C.; Martinez, J.M.; Motta Marques, D.D.; Cirilo, J.A.; Fragoso, C.R. Assessment of Chlorophyll-a Remote Sensing Algorithms in a Productive Tropical Estuarine-Lagoon System. *Remote. Sens.* **2017**, *9*, 516. [CrossRef]