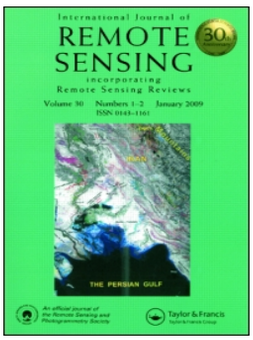

International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: http://www.tandfonline.com/loi/tres20

Application of a generalized additive model (GAM) for estimating chlorophyll-a concentration from MODIS data in the Bohai and Yellow Seas, China

Yueqi Wang, Dongyan Liu & DanLing Tang

To cite this article: Yueqi Wang, Dongyan Liu & DanLing Tang (2017) Application of a generalized additive model (GAM) for estimating chlorophyll-a concentration from MODIS data in the Bohai and Yellow Seas, China, International Journal of Remote Sensing, 38:3, 639-661

To link to this article: http://dx.doi.org/10.1080/01431161.2016.1268733

INTERNATIONAL JOURNAL OF REMOTE SENSING, 2016
VOL. 38, NO. 3, 639–661
http://dx.doi.org/10.1080/01431161.2016.1268733

Taylor & Francis
Taylor & Francis Group

Application of a generalized additive model (GAM) for estimating chlorophyll-*a* concentration from MODIS data in the Bohai and Yellow Seas, China

Yueqi Wang[a], Dongyan Liu[a,b] and DanLing Tang[c]

[a]Key Laboratory of Coastal Zone Environmental Processes and Ecological Remediation, Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai, Shandong, P.R. China; [b]State Key Laboratory of Estuarine and Coastal Research, East China Normal University, Shanghai, P.R. China; [c]Research Centre for Remote Sensing of Marine Ecology & Environment, State Key Laboratory of Tropical Oceanography, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, P.R. China

ABSTRACT

In optically complex waters, it is important to evaluate the accuracy of the standard satellite chlorophyll-*a* (chl-*a*) concentration algorithms, and to develop accurate algorithms for monitoring the dynamics of chl-*a* concentration. In this study, the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite remote-sensing reflectance and concurrent *in situ* measured chl-*a* (2010–2013) were used to evaluate the standard OC3M algorithm (ocean chlorophyll-a three-band algorithm for MODIS) and Graver–Siegel–Maritorena model version 1 (GSM01) algorithm for estimating chl-*a* concentration in the Bohai and Yellow Seas (BYS). The results showed that the chl-*a* algorithms of OC3M and GSM01 with global default parameters presented poor performance in the BYS (the mean absolute percentage difference (MAPD) and coefficient of determination (R^2) of OC3M are 222.27% and 0.25, respectively; the MAPD and R^2 of GSM01 are 118.08% and 0.07, respectively). A novel statistical algorithm based on the generalized additive model (GAM) was developed, with the aim of improving the satellite-derived chl-*a* accuracy. The GAM algorithm was established using the *in situ* measured chl-*a* concentration as the output variable, and the MODIS above water remote-sensing reflectance (visible bands at 412, 443, 469, 488, 531, 547, 555, 645, 667, and 678 nm) and bathymetry (water depth) as input variables. The MAPD and R^2 calculated between the GAM and the *in situ* chl-*a* concentration are 39.96% and 0.67, respectively. The results suggest that the GAM algorithm can yield a superior performance in deriving chl-*a* concentrations relative to the standard OC3M and GSM01 algorithms in the BYS.

ARTICLE HISTORY
Received 22 November 2015
Accepted 21 November 2016

1. Introduction

Satellite-derived chlorophyll-*a* (chl-*a*) concentrations have been an essential index for large spatial-scale and long temporal-scale oceanographic studies. However, both

CONTACT Dongyan Liu dyliu@yic.ac.cn Yantai Institute of Coastal Zone Research, CAS, 17th Chunhui Road, Laishan District, Yantai 264003, P.R. China

careful validation and sophisticated algorithm development are major requirements in order to retrieve accurate chl-*a* products from satellite imagery. In global oceans, the empirical algorithms and semi-analytical algorithms were commonly used to estimating chl-*a* concentration from ocean colour sensors. Empirical algorithms rely on a specific feature, such as the ratio between the blue and green band reflectance and combinations of more spectral bands, modelled to chl-*a* measurements using statistical regression (O'Reilly et al. 1998; Dall'Olmo et al. 2005; Moses et al. 2009; Hu, Lee, and Franz 2012). Currently, standard Moderate Resolution Imaging Spectroradiometer (MODIS) chl-*a* products provided by the US National Aeronautics and Space Administration (NASA) Ocean Biology Processing Group (OBPG) are derived using the empirical OC3M (ocean chl-*a* three-band algorithm for MODIS) algorithm. The OC3M chl-*a* has been widely applied for estimating phytoplankton biomass (Liu and Wang 2013; González Taboada and Anadón 2014), investigating harmful algal blooms (Tang et al. 2004; Hu et al. 2010; Park, Ruddick, and Lacroix 2010) and evaluating eutrophication (Kitsiou and Karydis 2011; Banks et al. 2012). However, chl-*a* in optically complex waters, derived using the standard OC3M algorithm, must be evaluated systematically and used carefully because the presence of complex constituents, such as detritus and coloured dissolved organic matter (CDOM), can significantly increase the uncertainty in chl-*a* retrievals (Moore, Campbell, and Dowell 2009).

The semi-analytical models constitute another algorithm for estimating chl-*a* concentration from ocean colour sensors. The semi-analytical algorithms rely on radiative transfer solutions from knowledge of the inherent optical properties (IOPs) and try to isolate the spectral influence of several optical variables (Carder et al. 1999; Maritorena, Siegel, and Peterson 2002; Wang, Boss, and Roesler 2005). The Graver–Siegel–Maritorena model version 1 (GSM01) is a widely used semi-analytical algorithm (Maritorena, Siegel, and Peterson 2002) and was well adopt for open waters because its parameters were simulated from a large global *in situ* dataset. A standard GSM01 chl-*a* product, provided by the European Node for Global Ocean Colour project (http://www.globcolour.info/), has been widely used in estimating phytoplankton biomass (González Taboada and Anadón 2014; Dave and Lozier 2015). The success of the semi-analytical approach depends on correct radiative transfer solutions and the accurate parameterization of the absorption and scattering coefficients of each individual constituent in the waters. Because these parameters are geographically specific and not generally known accurately in complex waters, retrievals are not always successful. For example, Shang et al. (2014) examined several chl-*a* algorithms in the northern South China Sea, and concluded that the GSM01 algorithm with default parameterization led to high errors in coastal waters. Komick, Costa, and Gower (2009) found that the standard GSM01 algorithm was ineffective at estimating the chl-*a* concentrations in turbid waters, but a modified version using different CDOM and phytoplankton absorption models exhibits significantly improved accuracy. Overall, the successful application of the GSM01 algorithm requires appropriate tuning of the default parameters in the model based on the local bio-optical dataset.

Because of the constraints associated with OC3M and GSM01 algorithms when estimating chl-*a* concentration in optically complex waters, some statistical inversion approaches have successfully been developed to retrieve chl-*a* concentrations from satellite images in complex waters, with the advantages that require no prior knowledge

of the optical features and easy to implement. For example, neural network based algorithms were successfully implemented for quantifying the chl-*a* concentration from satellite reflectance (Vilas, Spyrakos, and Torres Palenzuela 2011; Ioannou et al. 2013). The generalized additive model (GAM) is another promising statistical method, which has been applied for satellite ocean colour data analysis (Hastie and Tibshirani 1986, 1990). As neural network based algorithm, a GAM-based algorithm does not require any prior knowledge of the structural relationship between the response variable and the predictors, and is able to determine the relation from the input variables by itself. In addition with GAM method, it is possible to examine the specific importance and influence of each predictor in the estimation (Hastie and Tibshirani 1990). In recent years, the GAM method had been proposed for estimating trends of chl-*a* concentration in global oceans (Boyce, Lewis, and Worm 2010), quantifying the influence of the environmental regime on chl-*a* concentration variation (Irwin and Finkel 2008; Raitsos et al. 2012), estimating sea surface salinity using MODIS ocean colour reflectance (Urquhart et al. 2012), and mapping Secchi depth using MODIS-Aqua and auxiliary data (Stock 2015). In this context, GAM might be a good choice for chl-*a* concentration retrieval from satellite ocean colour reflectance.

The main goals of this study are (1) to evaluate the performance of the global standard chl-*a* algorithms (OC3M and GSM01) for estimation of chl-*a* concentration in the turbid coastal waters of the BYS compared to *in situ* chl-*a* concentration data and (2) to develop an accurate GAM algorithm based on a statistical analysis between the concurrent remote-sensing reflectance at available MODIS-visible wavelengths and the *in situ* chl-*a* measurements for estimating the chl-*a* concentration in the BYS.

2. Materials and methods

2.1 *Study area*

The Bohai and Yellow Seas (BYS), located in northern China, are connected by the Bohai Strait (Figure 1). The Bohai Sea is a shallow inland sea with an average water depth of 18 m. The Yellow Sea, located between mainland China and the Korean Peninsula, is a typical shallow epicontinental sea with an average depth of approximately 50 m (Liu and Wang 2013). The optical properties of BYS are significantly influenced by high suspended sediment from river discharges and bottom reflectance from the shallow sea floor (Wang, Tang, and Shi 2007; Shi and Wang 2010b), which can contribute to the large errors of the standard chl-*a* algorithms (Shi and Wang 2010a, 2010b). To improve the accuracy of the satellite chl-*a* products, Chen and Quan (2013) developed an improved empirical algorithm using a different combination of bands in the standard OC3M algorithm over the Yellow River Estuary, and Huang et al. (2013) modified and validated semi-analytical inversion models of the ocean colour in coastal Yellow Sea and East China Sea. In general, the development of empirical algorithms has been hindered by their need for prior information about the optical properties to support the selection of the bands (Siswanto et al. 2011; Chen and Quan 2013), while the semi-analytical algorithm was usually hindered by improper default parameter settings in the model (Huang et al. 2013; Shang et al. 2014). Therefore, it is necessary to perform a systematic evaluation of the satellite standard chl-*a* products and further develop more accurate

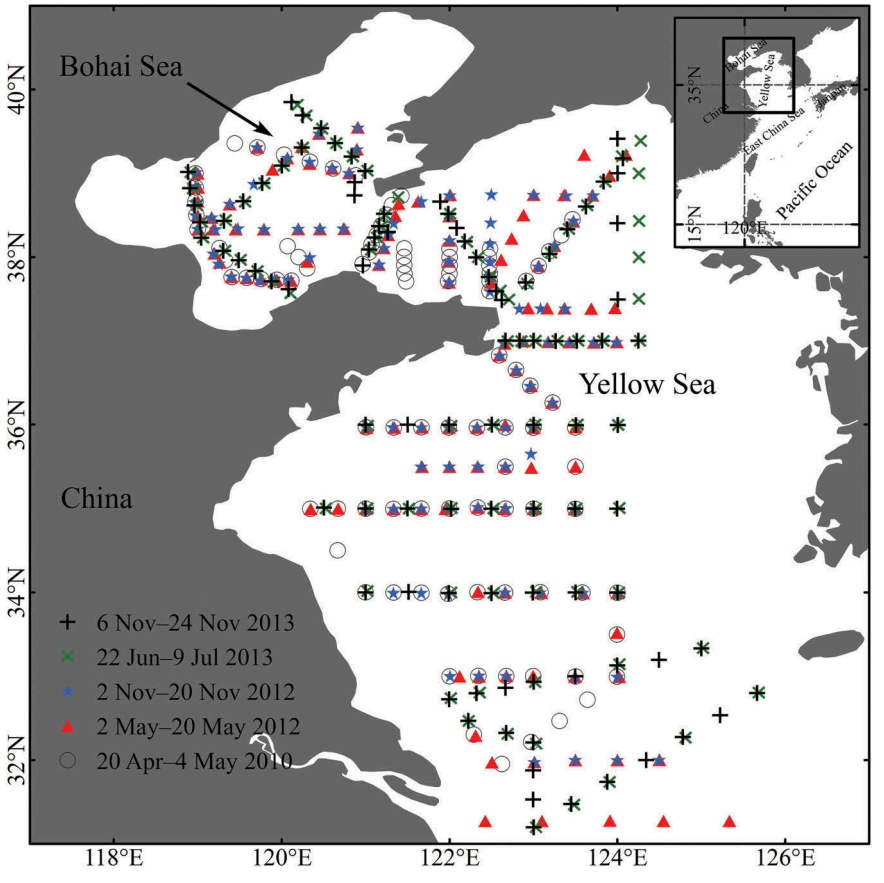

Figure 1. Map showing the study area (the Bohai and Yellow Seas) and the sites of the *in situ* chl-*a* measurements.

and operational algorithms that can be properly applied in the BYS. In this sense, statistical inversion approach based on GAM was used for the first time to derive the chl-*a* concentration from ocean colour reflectance in the BYS.

2.2 In situ *measurements of chl*-a

The *in situ* measured sea surface chl-*a* data was used for chl-*a* algorithm accuracy assessment and new algorithm development. These chl-*a* data were collected from five cruises conducted during the years 2010–2013 (Figure 1). The five cruises used almost the same sampling method for the chl-*a* measurements, and each site had three duplicate samples for quality control. Samples of 1000 ml of water were collected from the surface layer with a measurement depth of less than 5 m and then filtered through 47 mm Whatman GF/F filters under low vacuum. The filters were preserved in the dark at a low freezing temperature (–20°C) before laboratory analysis. In the laboratory, the chl-*a* concentration was determined spectrophotometrically (TU-1800, Persee, China) after being extracted with 15 ml of 90% acetone in the dark for 24 h at 4°C (Lorenzen 1967). The longitude, latitude, and water depth of each site was also recorded as auxiliary data.

2.3 *MODIS imagery and data extraction*

The daily MODIS-aqua standard local area coverage (LAC) remote-sensing reflectance (R_{rs}) images for 10 available visible bands (wavelength at 412, 443, 469, 488, 531, 547, 555, 645, 667, and 678 nm) and standard chl-*a* products were downloaded from the NASA Goddard Space Flight Center (GSFC, http://oceancolor.gsfc.nasa.gov/cms/). The R_{rs} data sets were derived using the most update-to-date standard calibration and atmospheric correction algorithms (Gordon 1997; Wang and Shi 2007), with a spatial resolution of approximately 1 km ×1 km.

For validation of standard algorithms and new algorithms development, the standard criterion for extracting concurrent data from satellite images, and *in situ* measurements was commonly as follows: (1) a time difference of ±3 h was set to acquire matching pairs; (2) the median value from a 5 × 5 window centred at each *in situ* site was used to define the coincident values in satellite images; (3) to minimize the sensor and/or algorithm noise and avoid strong non-homogeneous windows, the satellite data were only used when the number of valid pixels in the 5 × 5 window exceeded 15 and the coefficient of variation (CV) was <0.15. As a result, a dataset with only 23 matching pairs was produced by the match-up using this standard criterion. Because of the 23 samples is far less than the needs for GAM model training with 13 independent variables, a relaxed criterion for match-up analysis was adopted in this study. The relax criterion has the same procedures as the standard criterion except for the time window is ±24 h and the CV is less than 0.5. In total, 180 matching pairs between the satellite and *in situ* measurements were extracted for statistical analyses.

In this study, the concurrent dataset produced from the relaxed match-up criterion (with 180 samples) was applied for the GAM development and the accuracy comparison between different chl-*a* algorithms. A strict accuracy comparison was also presented in Section 4.1, using the concurrent dataset produced by the standard match-up criterion (with 23 samples), in order to investigate the effect of the relaxed criterion on the results of match-up analyses.

2.4 *Standard satellite chl-a algorithms*

The standard satellite chl-*a* products were calculated using the following algorithms, which are freely available to end users: (1) OC3M, a fourth-order empirical algorithm based on two MODIS R_{rs} band ratios ($R_{rs,443}/R_{rs,547}$ and $R_{rs,488}/R_{rs,547}$) (O'Reilly et al. 2000); (2) GSM01, an optimized semi-analytical algorithm that simultaneously retrieves chl-*a* concentration, absorption coefficient for dissolved and detrital materials and the particulate backscatter coefficient at the reference band, from spectral measurements of remote-sensing reflectance (Lee, Carder, and Arnone 2002; Maritorena, Siegel, and Peterson 2002; Franz and Werdell 2010). The default parameters for the GSM01 model were estimated by Maritorena, Siegel, and Peterson (2002) using a simulated annealing approach from a global *in situ* data set, and selecting 443 nm as the reference wavelength. A Levenberg–Marquardt nonlinear least-squared method is commonly used to solve the unknown parameters in the model.

2.5 *Application of GAM algorithm*

In this study, the GAM method was implemented for the retrieval of chl-*a* concentrations from MODIS R_{rs} images. A GAM is a flexible statistical model that extends the traditional linear model to account for nonlinear relationships between the response and predictor variables (Hastie and Tibshirani 1986, 1990). The development of a GAM includes three phases: (1) the original design of the GAM based on the property of the response variable, i.e. deciding the smoothing function, link and distribution functions; (2) consideration of the predictor parameters, determining the potential parameters to test as input independent variables in the GAM; (3) model optimization based on a forward and backward stepwise model fitting approach, to select the related independent variables and decide the final formula of the model.

2.5.1 *Structure of the GAM*

The GAM constructs non-parametric smoothing functions of the predictor variables to replace the generalized linear functions in linear models with the form (Hastie and Tibshirani 1986, 1990; Wood 2006):

$$y = B_0 + \sum_{j=1} f_j(x_j) + \varepsilon, \tag{1}$$

where y is the response variable, B_0 is the overall mean of the response, x_j are the predictor variables, ε is the residual, and f_j are smoothing functions such as splines, kernels, and linear functions, which are obtained using a scatterplot smoothing algorithm with a back-fitting procedure that allows for choosing the appropriate function for each of the x_j.

A cubic regression spline smoothing function with a potential knot point at each independent variable was used in this study. A log link was used to establish the relationship between the mean value of the response variable y and the smooth function of the parameter x together with a Gamma distribution because the chl-*a* data present a highly positive skewed distribution of non-negative values (Campbell 1995).

2.5.2 *Independent variables consideration*

The critical step in the GAM algorithm development is the determination of independent variables that may account for the significant influence of the response variable. In this study, *in situ* measured chl-*a* data were the dependent variable, and the independent variables were chosen from two geographic parameters (coordinate and water depth) and the MODIS-derived optical parameters (remote-sensing reflectance in visible bands). The geographic variables were considered, because the chl-*a* concentration presents significant spatial heterogeneity (Shi and Wang 2012; Yamaguchi et al. 2012; Chen and Liu 2015) and the bathymetry is the prominent factor controlling the chl-*a* distribution in the BYS (Liu and Wang 2013). Specifically, the distance to the land increases with the water depth, which can further influence the land-source nutrients discharge from land into the sea (Wang, Wang, and Zhan 2003); the bathymetry also imposes significant impact on the current system and front structure, which can further influence the horizontal and vertical transport of nutrients (Chen 2009; Liu and Wang 2013). Therefore, the bathymetry was considered as

one of the potential factors affecting the variability of chl-*a* concentrations through controlling the distribution of nutrients in the BYS.

2.5.3 *GAM optimization and cross-validation*

To optimize the GAM formula, a forward and backward stepwise model fitting approach was used based on Akaike's information criterion (AIC) statistic (Bozdogan 1987; Hastie and Tibshirani 1990; Wood 2006). All the independent variables were included in the original model as smoothed terms. By using the AIC, the significance of each term in the model could be assessed, and the stepwise approach was enabled the removal of insignificant variables (predictors) from the final model. Hence, the final model gives the combined effect of each significant independent variable (predictor) on the dependent variable (response). A detailed description of the GAM can be found in the studies of Wood (2004) and Hastie and Tibshirani (1990), while examples of using in phytoplankton research are provided by Boyce, Lewis, and Worm (2010) and Raitsos et al. (2012).

Model validation was performed using cross-validation techniques. For all cases, 80% of the original match-up data points were randomly selected as a training dataset for the GAM, and the resultant model was validated on the remaining 20% of the data. This cross-validation procedure was repeated five times to ensure that the results are not biased by random selection of the data points.

2.5.4 *Analysis of the GAM results*

A final GAM model was established based on the response variable and its significant influential predictors. The specific importance and effect of each predictor imposed on the response can be examined from the GAM results. The relative significance of each predictor can be quantified and compared by the significance (*p*-value) associated with each smoothed term in the GAM. The effect of each predictor on the response can be described by the effective degree of freedom (EDF) and the function plot of each smoothed term.

2.6 *Algorithm evaluation*

To evaluate the results of the chl-*a* algorithms in relation to *in situ* chl-*a* values, the parameters of mean bias (MB), mean absolute percentage difference (MAPD) and root mean square log error (RMSLE) associated with the linear regression were calculated. These statistical parameters are expressed as

$$\mathrm{MB} = \frac{1}{N}\sum_{i=1}^{N}(C_{\mathrm{sa}} - C_{\mathrm{in}}), \tag{2}$$

$$\mathrm{MAPD} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{C_{\mathrm{sa}} - C_{\mathrm{in}}}{C_{\mathrm{in}}}\right| \times 100\%, \tag{3}$$

$$\mathrm{RMSLE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\log_{10}(C_{\mathrm{sa}}) - \log_{10}(C_{\mathrm{in}}))^2}, \tag{4}$$

where C_{sa}, C_{in}, and N are the satellite derived chl-*a*, the *in situ* measured chl-*a*, and the number of match-ups, respectively. MAPD gives the accuracy or uncertainty of the difference, while MB gives the systematic error or direction of bias (overestimation or underestimation) with respect to the *in situ* values; RMSLE is a robust parameter for comparison of the different chl-*a* algorithms via consideration of the log-normal distribution of chl-*a* values (Campbell 1995). Scatter plots of *in situ* measured chl-*a* values *versus* satellite-derived chl-*a* values for each algorithm were produced, and the slope and intercept of the linear fit and its coefficient of determination (R^2) were also obtained.

3. Results

3.1 *Characteristics of the in situ chl-a concentration*

The *in situ* measured chl-*a* concentrations (n = 503) in the study area span three orders of magnitude and range between 0.08 and 26.81 mg m^{-3} (Figure 2). The mean value (1.59 mg m^{-3}) is much higher than the global average (about 0.24 mg m^{-3}) (Gregg and Conkright 2002) and also higher than the average (about 0.61 mg m^{-3}) over the northern temperate oceans (30–60° N) (Dasgupta, Singh, and Kafatos 2009), because of the proximity to the coast of most measurement sites. However, compared with the most turbid estuary regions in the world, such as Chesapeake Bay (the average of *in situ* chl-*a* concentration is about 7.7 mg m^{-3}) (Werdell et al. 2009), the *in situ* chl-*a* concentration in the BYS is much lower. Based on the oceanic provinces defined by Antoine, André,

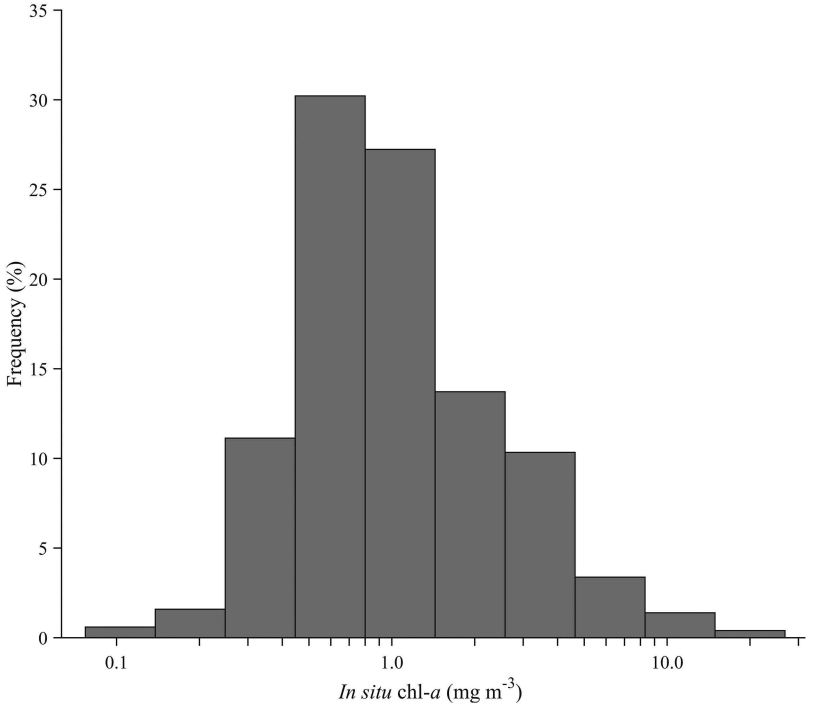

Figure 2. Frequency distribution of *in situ* measured chl-*a* concentrations (mg m^{-3}) for the whole data set (n = 503).

Table 1. Statistics of the *in situ* measured sea surface chl-*a* concentration for the five cruises.

Year	Month	Number	Average ± SD	Max	Min
2010	April, May	96	2.75 ± 3.05	16.28	0.20
2012	May	118	1.21 ± 1.06	9.43	0.10
2012	November	101	0.66 ± 0.31	2.50	0.25
2013	June, July	83	2.43 ± 3.27	26.81	0.08
2013	November	105	1.18 ± 0.82	4.35	0.11
Whole dataset		503	1.59 ± 2.14	26.81	0.10

SD, Max, and Min represent the standard deviation, minimum, and maximum, respectively. (Unit: mg m^{-3}.)

and Morel (1996), only 0.4% of the sampling sites (i.e. two) are oligotrophic with chl-*a* concentrations less than 0.1 mg m^{-3}, 54.1% are mesotrophic with concentrations between 0.1 and 1.0 mg m^{-3} and 45.5% are eutrophic with concentrations greater than 1.0 mg m^{-3}.

The statistics of *in situ* measured chl-*a* concentration for the five cruises were illustrated in Table 1. The seasonal variability of chl-*a* concentration was produced by taking an average value of each cruise, and the standard deviation of chl-*a* concentration in each cruise quantified its spatial heterogeneity. The average of chl-*a* concentration shows an obvious seasonal variation with high values in spring (April and May) and summer (June and July), with low values in autumn (November). The standard deviation of chl-*a* concentration also presents seasonal variation with high values in summer and spring, while low values in autumn (Table 1). The high value of standard deviation combined with the extreme high maximum (more than 10 mg m^{-3}) in spring and summer may be related to the occurrence of algae blooms during these two seasons (Liu et al. 2009; He et al. 2013; Liu and Wang 2013). Note that we cannot describe a complete annual cycle of chl-*a* concentration in this study due to the lack of *in situ* samples in winter.

3.2 *OC3M algorithm evaluation*

The accuracy of the OC3M chl-*a* algorithm was evaluated by comparison between the satellite-derived and *in situ* measured chl-*a* concentrations. A scatter plot of the *in situ versus* OC3M chl-*a* values and the calculated parameters are shown in Figure 3. High values of MAPD (222.27%) and RMSLE (0.50) indicate poor performance of the OC3M algorithm in the BYS. A value of MB = 2.04 indicated a distinct overestimation in chl-*a* values produced by OC3M. The distribution of coloured-coded points on the scatter plot exhibits a distinct division between shallow (coastal) and deep (offshore) waters (Figure 3). The deviation of colour-coded points from 1:1 line is larger in coastal waters than offshore, indicating that the OC3M chl-*a* algorithm results in larger errors in the coastal waters than in the offshore waters over the BYS. Although the absolute error was large in OC3M chl-*a*, the relative pattern might be reasonable, according to the significant correlation ($R^2 = 0.25$, $p < 0.01$) and positive slope (0.49).

3.3 *GSM01 algorithm evaluation*

The accuracy of the GSM01 chl-*a* algorithm can be addressed by a comparison between the satellite-derived and *in situ* measured chl-*a* values, as shown in Figure 4. Compared to OC3M, the GSM01 presented a lower error with MAPD = 118.08% and RMSLE = 0.45,

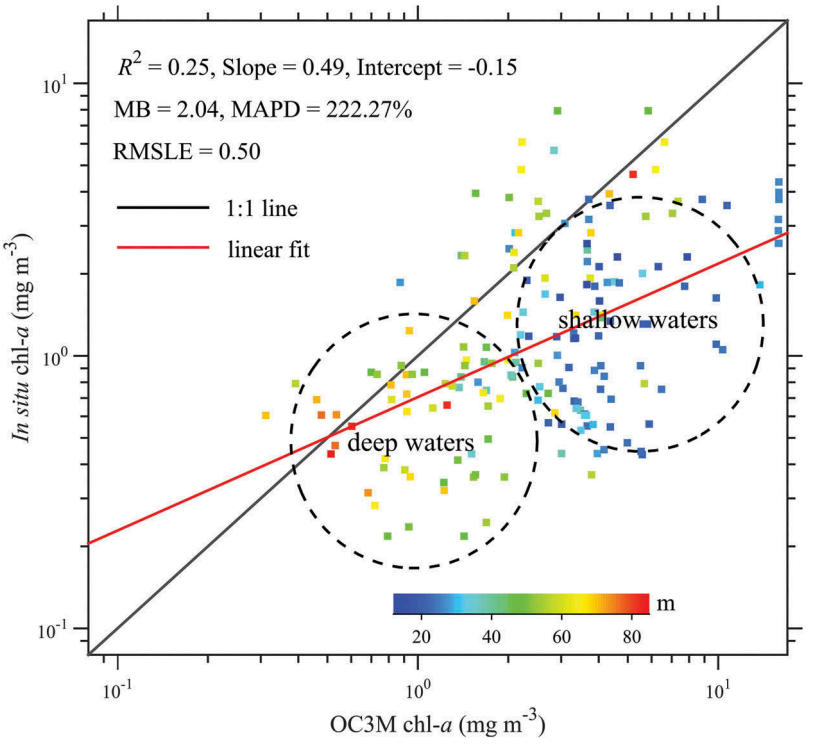

Figure 3. Scatter plot of the *in situ* measured chl-*a* values *versus* the satellite-derived chl-*a* calculated using the OC3M algorithm in the BYS. The colour scale in the figure indicates water depth of the site.

and overall slighter overestimation (MB = 0.15). However, the R^2 = 0.07 indicated no significant correlation between the GSM01 and the *in situ* chl-*a* values. The distribution of the colour-coded points and the slope of the linear regression (0.2) indicated that GSM01 overestimates the chl-*a* concentration in coastal waters but underestimates the concentration in the offshore waters over the BYS.

3.4 *Results of GAM chl-a algorithm construction*

A final GAM model for the best set of variables obtained from the AIC evaluation can be written as

$$\text{chl-}a = B_0 + \sum s(R_{rs}(\lambda)) + s(\text{depth}) + \varepsilon$$
$$\lambda = 412, 443, 469, 488, 531, 547, 555, 645, 667, 678. \quad (5)$$

In the final GAM model, the depth of sites and all the 10 visible R_{rs} bands were included as predictors. The final GAM function explained 84.8% of the variance of chl-*a*. The scatter plot of *in situ* measured *versus* GAM predicted chl-*a* is shown in Figure 5. Compared with OC3M (Figure 3) and GSM01 (Figure 4), the GAM model displays better agreement with the *in situ* observations. The GAM algorithm shows better statistical agreement than the OC3M and GSM01 algorithms with R^2 = 0.67, MAPD = 39.96%, and RMSLE = 0.20. The distribution of colour-coded points (Figure 5) presents a

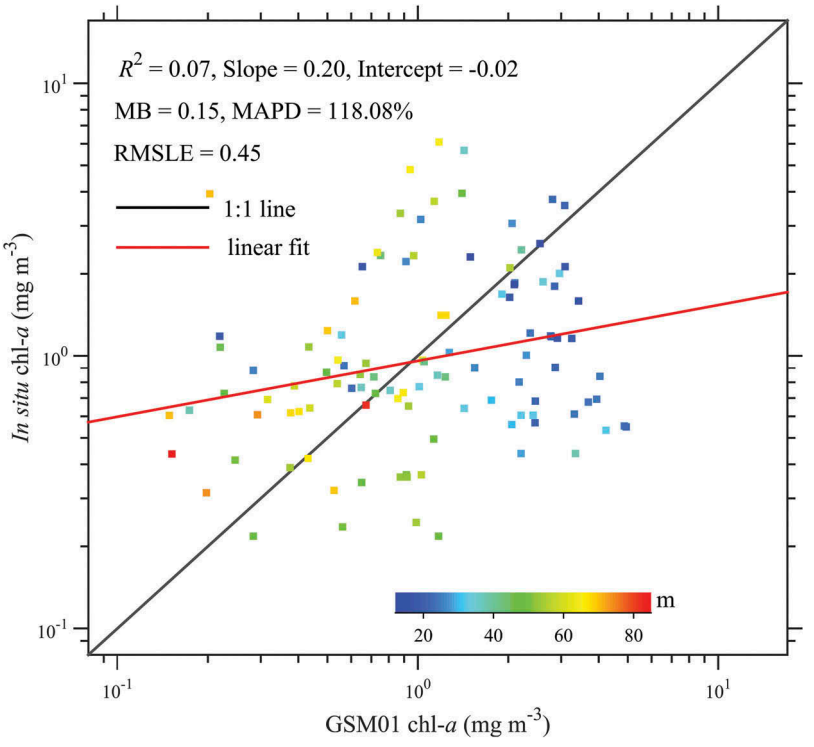

Figure 4. Scatter plot of the *in situ* measured chl-*a* values *versus* the satellite-derived chl-*a* calculated using the GSM01 algorithm in the BYS. The colour scale in the figure indicates the water depth of the site.

homogeneous distribution with water depth, indicating that the GAM algorithm performs as well in shallow waters as in offshore waters. The GAM method has been shown to be the more effective algorithm for quantifying the chl-*a* concentration in the BYS.

For the GAM method, it is possible to examine the specific importance and effect of each of predictors (R_{rs} and depth Equation (5)) in the estimation of chl-*a* concentration. Table 2 lists the effective degree of freedom and approximate significance associated with each smoothed term in the final GAM model. The result shows that all the 11 variables included in the GAM are statistically significant (Table 2, *p*-value <0.05) and thus important in the chl-*a* retrieval. Of the 11 predictors included in the GAM, R_{rs} at 412, 443, and 555 nm present the highest influential behaviour on chl-*a* estimation (smallest values of *p*-value in Table 2). The effect of each predictor on the response (chl-*a*) can be examined by the smooth function plots in Figures 6(*a*)–(*k*), in which the positive slope of smoothed line indicates positive effect of the predictor imposed on the chl-*a* estimation, and *vice versa*. Most of the predictor variables exhibit a significant nonlinear effect on the chl-*a* estimation (Table 2 and Figures 6 (*a*)–(*k*), EDF >1), except for $R_{rs,448}$, $R_{rs,547}$, and $R_{rs,678}$ (Table 2 and Figures 6(*a*)–(*f*), EDF = 1). The chl-*a* decreases with depth at 0–20 m and 60–80 m, roughly invariable with depth from 20 to 60 m and increases with depth above 80 m (Figure 6(*a*)). The $R_{rs,412}$, $R_{rs,448}$, $R_{rs,547}$, and $R_{rs,667}$ have negative effect on the chl-*a* estimation, whereas $R_{rs,443}$, $R_{rs,469}$, $R_{rs,555}$, $R_{rs,645}$, and $R_{rs,678}$ have positive effect on chl-*a* estimation (Figures 6(*b*)–(*k*)). The $R_{rs,531}$ has no-monotonic effect

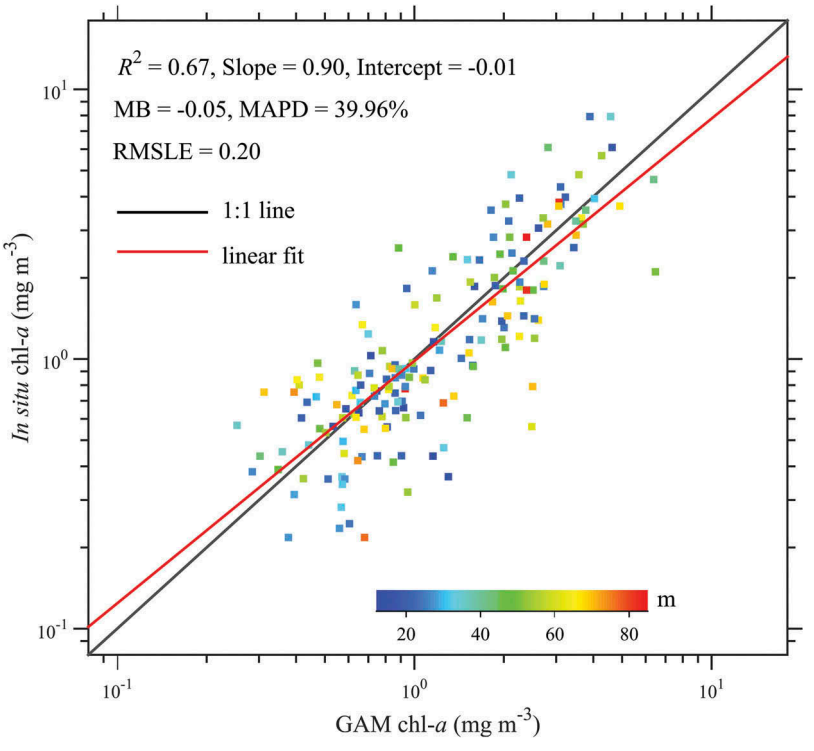

Figure 5. Scatter plot of the *in situ* measured chl-*a* values *versus* the satellite-derived chl-*a* calculated using GAM algorithm in the BYS. The colour scale in the figure indicates the water depth of the site.

Table 2. Effective degree of freedom (EDF) and approximate significance (*p*-value) of each of GAM smoothed terms in Equation (5).

Smoothed term	EDF	*p*-Value
Depth	7.95	1.42×10^{-2}
$R_{rs,412}$	5.50	4.15×10^{-13}
$R_{rs,443}$	3.37	7.44×10^{-12}
$R_{rs,469}$	2.04	4.36×10^{-3}
$R_{rs,488}$	1.00	7.57×10^{-7}
$R_{rs,531}$	8.56	1.11×10^{-3}
$R_{rs,547}$	1.00	2.11×10^{-3}
$R_{rs,555}$	5.52	2.36×10^{-11}
$R_{rs,645}$	2.30	7.12×10^{-4}
$R_{rs,667}$	1.91	8.35×10^{-5}
$R_{rs,678}$	1.01	1.76×10^{-3}

on chl-*a* estimation, with chl-*a* slightly decrease with $R_{rs,531}$ when $R_{rs,531}$ <0.005 and increase with $R_{rs,531}$ when $R_{rs,531}$ >0.005 (Figure 6(*f*)).

3.5 *Map comparisons between different algorithms*

To provide spatial comparisons between the chl-*a* products calculated from different algorithms, Figures 7(*a*)–(*e*) present the MODIS composite chl-*a* maps associated with

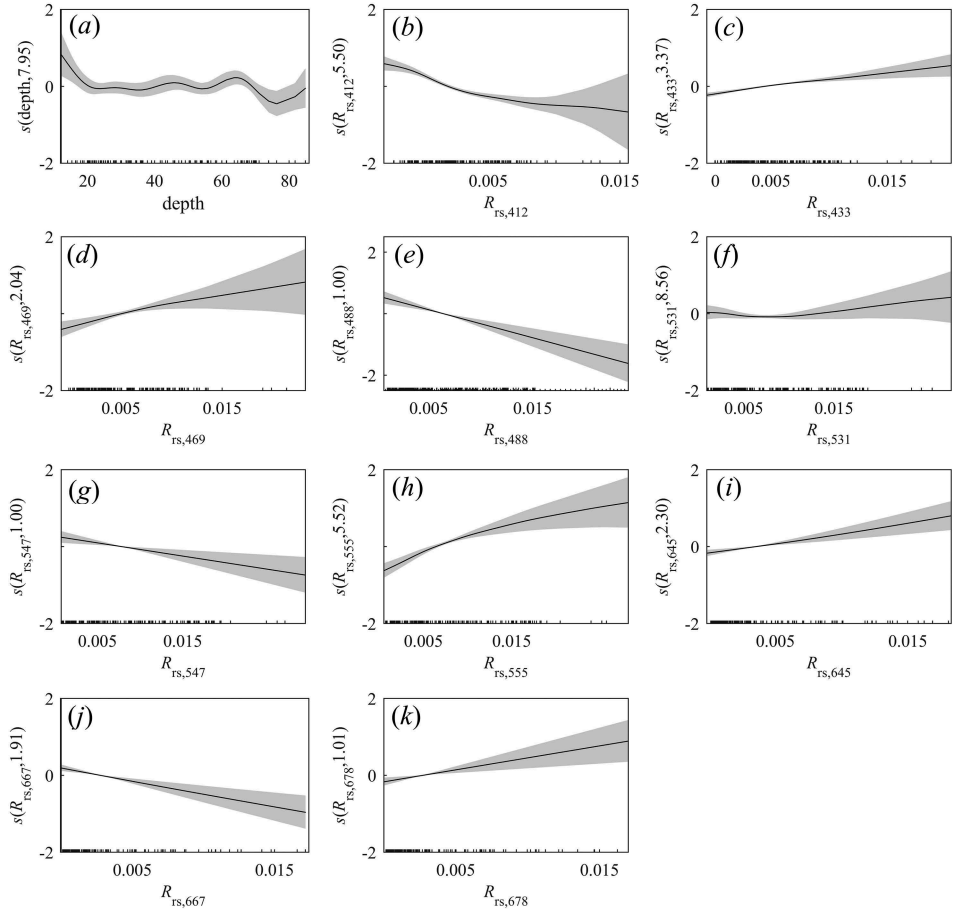

Figure 6. (*a*)–(*k*) Smooth function plots for predictor variables in the GAM. Estimated smooth functions (solid lines) with 95% confidence intervals (shaded area) are shown for each predictor. The numbers in the labels of the *y*-axis denote the effective degrees of freedom. The comb on the *x*-axis shows where the value of predictor data points lie.

the *in situ* measured chl-*a* values during the five cruises. The chl-*a* distributions in the first three columns were derived using the OC3M, GSM01, and GAM algorithms, respectively (Figures 7(*a*)–(*e*)). The fourth column shows the spatial distributions of *in situ* measured chl-*a* concentration (colour of scatter indicates chl-*a* values) (Figures 7(*a*)–(*e*)). Overall, all satellite chl-*a* images showed similar relative spatial patterns with higher chl-*a* in coastal waters and lower values offshore. During spring 2010 (Figure 7 (*a*)(i–iv)), both the OC3M and GAM images presented a prominent chl-*a* bloom in the centre of the Yellow Sea (Figure 7(*a*)(i, iii))), as indicated by the *in situ* chl-*a* plots (Figure 7(*a*(iv))), whereas this is not captured by the GSM01 algorithm. In spring 2012 (Figure 7(*b*(i–iv))), all chl-*a* products showed consistent patterns with chl-*a* increased from deep to shallow waters, but the OC3M chl-*a* (Figure 7(*b*(i))) displays high chl-*a* patches inconsistent with the *in situ* chl-*a* (Figure 7(*b*(iv))). During autumn 2012 and autumn 2013 (Figure 7(*c*(i–iv),*e*(i–iv)), the spatial distribution produced by GAM and *in situ* chl-*a* were relatively homogenous (Figure 7(*c*(iii,iv)) (2012); Figure 7(*e*(iii,iv))

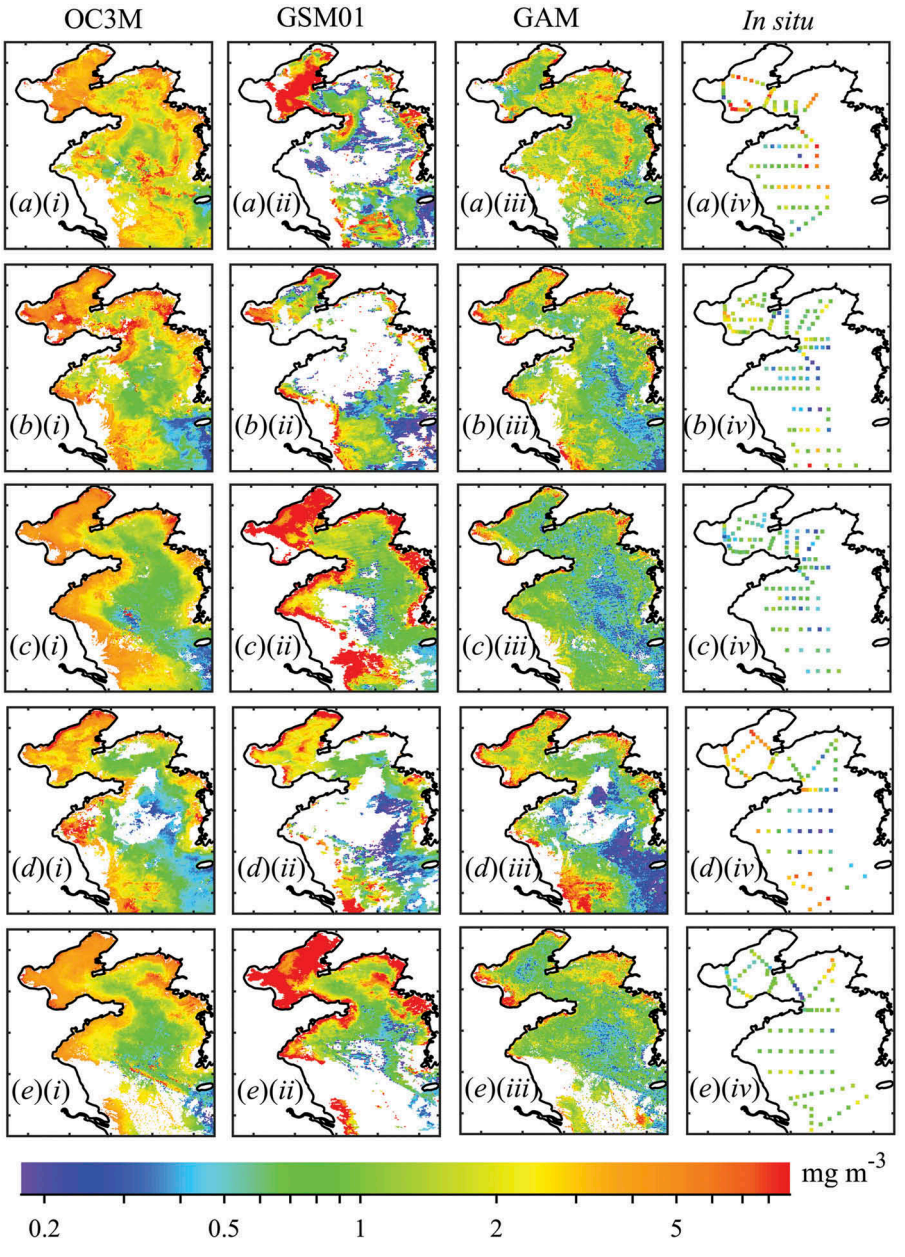

Figure 7. Composite chl-*a* maps created using the three algorithms (OC3M, GSM01, and GAM) and the *in situ* measured chl-*a* plots for the five cruise periods. Each row represents the period of (*a*) 20 April–04 May 2010, (*b*) 02–20 May 2012, (*c*) 02–20 November 2012, (*d*) 22 June–09 July 2013, and (*e*) 06–24 November 2013, respectively. Each column presents the chl-*a* distributions derived using the algorithms of OC3M, GSM01, and GAM, respectively.

(2013)), whereas OC3M and GSM01 images were characterized by a high gradient between the offshore and the coastal waters (Figure 7(*c*(i,ii)) (2012); Figure 7(*e*(i,ii)) (2013)). In summer 2013 (Figure 7(*d*(i–iv))), all the three satellite chl-*a* products show relatively consistent spatial patterns with *in situ* chl-*a*. Overall, the results indicate that

the chl-*a* maps produced using the GAM algorithm have a closer correspondence with the *in situ* chl-*a* distribution than the other two algorithms. In particular, the erroneously high chl-*a* overestimations obtained using the global standard chl-*a* algorithm (OC3M and GSM01) were reduced in the GAM chl-*a* maps. It is also clear that the GSM01 algorithm (Figures 7(*a*(ii), *b*(ii), *c*(ii),*d*(ii), e(ii))) missed significant features seen in the *in situ* chl-*a* measurements, the missing features resulting from the masking out of negative values arising from the inversion failing to find a realistic solution or because of negative R_{rs} values caused by atmospheric correction errors (Tilstone et al. 2013).

4. Discussion

4.1 *Uncertainties and limitations of the results*

Because of many compromised schemes were applied in the match-up analyses and model development processes, we must admit that somewhat uncertainties and certain limitations still existed in the results of this study.

First, the uncertainties of the validation can result from the relaxed match-up criterion. The threshold values of temporal window (±24 h) and CV (0.5) may be too large for the BYS with significant terrestrial and tidal influences. For example, the hourly ocean colour products from the Korean Geostationary Ocean Color Imager (GOCI) demonstrated significant diurnal variations in ocean optical, biological, and biogeochemical properties in the Bohai Sea, Yellow Sea, and East China Sea (Wang et al. 2013). To address the uncertainties from the relaxed match-up analysis, a strict validation was given using the standard match-up criterion in which the temporal window and CV was set to ±3 h and 0.15, respectively. Figure 8 and Table 3 show the results of the strict match-up analyses between *in situ* measured and satellite derived chl-*a* products. In spite of slightly deterioration in accuracy of GAM algorithm (R^2 = 0.57), the result still confirmed that the GAM algorithm (MAPD = 38.41%, R^2 = 0.57) receive better performance than standard OC3M (MAPD = 140.24%, R^2 = 0.12) and GSM01 algorithms (MAPD = 92.82%, R^2 = 0.17), as well as the result from the relaxed match-up analysis.

Second, it is common to use *in situ* R_{rs} and *in situ* chl-*a* products for algorithm validation and new model development, rather than directly use the satellite R_{rs} and *in situ* chl-*a* products, as satellite R_{rs} itself may has inherent uncertainties. For example, the satellite retrievals of R_{rs} at visible bands from MODIS present significant biases with a median of APD (absolute percentage of difference) of 18–28% along the yellow sea coast (Cui et al. 2014). These uncertainties in satellite R_{rs} may propagate into the established GAM model in our study.

Moreover, due to the fact that the GAM model established in this study is based on the relationship between the apparent optical properties (AOPs) and chl-*a* concentrations, the GAM algorithm actually belongs to the category of empirical models. This model is simple and easy to implement. However, it lacks a physical foundation and has a limited operational range defined by the optical characteristics of the waters for which they are trained.

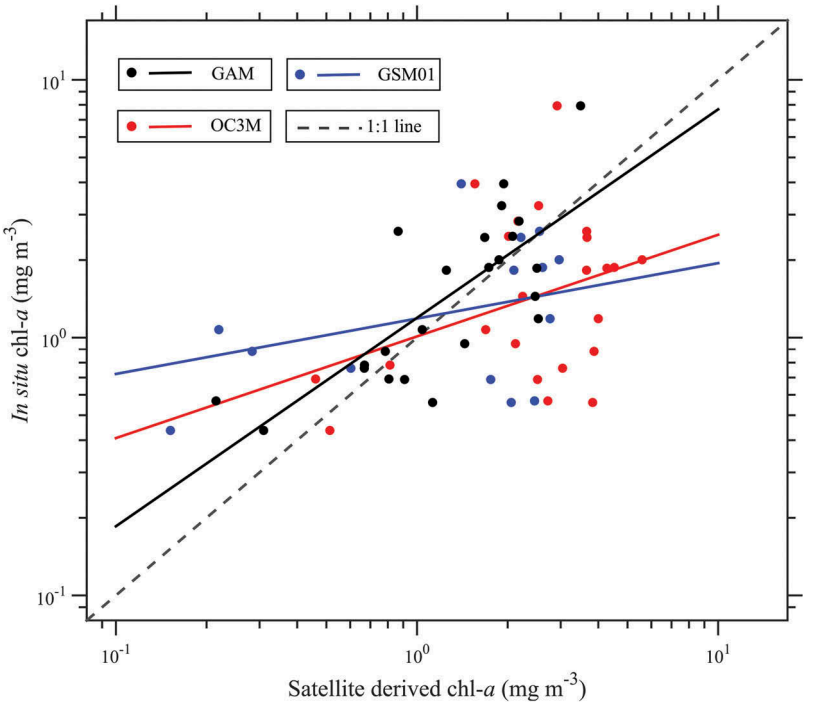

Figure 8. Scatter plot of the *in situ* measured chl-*a* values *versus* the satellite-derived chl-*a* calculated using OC3M, GSM01 and GAM algorithm, respectively. The strict match-up criterion was used to produce the concurrent dataset (n = 23). The solid line indicates fitted line corresponding to each algorithm.

Table 3. Statistics results for the match-up analysis between *in situ* measured and satellite derived chl-*a* products calculated using OC3M, GSM01, and GAM algorithm, respectively.

Algorithm	R^2	Slope	Intercept	RMSLE	MAPD (%)
OC3M	0.12	0.40	0.004	0.40	140.24
GSM01	0.17	0.21	0.074	0.49	92.82
GAM	0.57	0.81	0.077	0.22	38.41

The strict match-up criterion was used to produce the concurrent dataset (n = 23).

4.2 *Performance of the standard algorithms*

The statistical and spatial comparisons showed that the OC3M algorithm presented poor performance in magnitude (MAPD = 222.27%) but somewhat reasonable performance in relative patterns (R^2 = 0.25, p < 0.05). OC3M has an obvious overestimation (MB = 2.04), notably in coastal waters. This overestimation is mainly attributed to errors in atmospheric correction and model development. In the current standard MODIS data, the atmospheric correction is basically assumed to be black at two near-infrared (NIR) bands (748 and 869 nm) for the open ocean. However, this NIR black assumption is invalid for turbid waters, leading to significant errors in MODIS ocean colour data (Wang 2007). For example, Son, Wang, and Shon (2011) found unreasonable high chl-*a* patches in the central Yellow Sea due to the large errors arising from the standard-NIR atmospheric correction algorithm. The OC3M chl-*a* algorithm uses the blue-green band ratio of R_{rs},

with a low ratio corresponding to high chl-*a* values (O'Reilly et al. 1998; O'Reilly et al. 2000). However, additional detritus, coloured dissolved organic matter and suspended sediment can significantly increase the absorption coefficient of the water and decrease the R_{rs} in the blue band, while it enhances the particulate backscattering coefficient and increases the R_{rs} in the green band. Hence, the chl-*a* is overestimated by the lower than normal blue-green band ratios in coastal waters (Dierssen 2010). There are a few methods that can help to improve the accuracy of OC3M chl-*a* empirical algorithms. First, the coefficients in OC3M fourth-order function can be tuned using new regressions based on local bio-optical samples (Gregg et al. 2009; Siswanto et al. 2011). Second, the algorithm using other band combinations might obtain more accurate chl-*a* products, e.g. the fluorescence line height algorithm (Gower, Doerffer, and Borstad 1999), the maximum chlorophyll index (Gower et al. 2005) and algorithms using red and NIR bands (Moses et al. 2009; Le et al. 2013).

The standard GSM01 chl-*a* algorithm in our study presented a low accuracy in magnitude (MAPD = 118.08%) and relative pattern (R^2 = 0.07). The poor performance of this algorithm was reported in previous studies in high absorption and backscattering waters (Komick, Costa, and Gower 2009; Tilstone et al. 2013; Shang et al. 2014). The large errors of the GSM01 algorithm are probably due to imperfect atmospheric correction and the R_{rs}-based model parameterization. As discussed above, the failure of the standard NIR-based atmospheric correction in coastal waters can cause the use of inaccurate R_{rs} in model development. In the standard GSM01 parameterization, the default values of g_i (transition from the quasi-single backscatter albedo remote-sensing reflectance below the air-water interface), S_{cdom} (spectral decay constant for non-algal detritus and dissolved organic matter absorption), η (power-law exponent for the particulate backscattering coefficient), and a_{ph}* (chl-*a* specific absorption coefficient) in semi-analytical function may not reflect the truth of the local area, and these errors will propagate into the retrieved chl-*a* values. To improve the GSM accuracy, a necessary local parameterization based on *in situ* measurements is suggested (Franz and Werdell 2010) and implemented (Komick, Costa, and Gower 2009; Tilstone et al. 2013; Shang et al. 2014). For example, Komick, Costa, and Gower (2009) modified the standard GSM model by using a different CDOM absorption model and a new estimated phytoplankton absorption; the modified GSM (RMSLE = 0.446) presented better accuracy than the standard GSM01 (RMSLE = 0.772) over the western Canada coastal waters. Shang et al. (2014) tuned the parameters of S_{cdom}, η, and a_{ph}* in the GSM model and slightly reduced the derived chl-*a* errors (MAPD reduced from 147% to 80%) over the northern South China Sea.

Overall, without consideration of the errors in MODIS R_{rs} data, the success of standard chl-*a* algorithms (OC3M and GSM01) is highly dependent on the prior knowledge of the optical response to chl-*a* variability. For example, in the OC3M algorithm, the blue and green band ratio was selected as the input variable based on an optical feature analysis between different levels of chl-*a* concentration (O'Reilly, Maritorena, and Siegel 2000). In GSM01 algorithm, the format and default parameters of radiative transfer solutions were determined from many bio-optical experiments (Lee, Carder, and Arnone 2002; Maritorena, Siegel, and Peterson 2002). If the prior knowledge was not clearly captured by the model, the standard chl-*a* product is likely to produce inaccurate retrievals.

4.3 *Performance of GAM algorithm*

Compared to the standard OC3M and GSM01 algorithms, the GAM algorithm produced more accurate chl-*a* retrievals over the BYS, and it obtained similar accuracy in the coastal (shallow) and offshore (deep) waters. In fact, the GAM belongs to the category of empirical algorithms because it was based on the statistical relationship between the *in situ* measured chl-*a* concentration and the AOP of the water. Compared with the OC3M empirical algorithm, the GAM method does not require any prior knowledge about the bio-optical response of water. Compared with other statistical algorithms such as the neural network based method (Vilas, Spyrakos, and Torres Palenzuela 2011; Ioannou et al. 2013), the GAM algorithm can quantify the significant effect of each predictor on the chl-*a* estimation, and can depict the relationship between chl-*a* and each of the predictors used in the model. Overall, the advantages of the GAM chl-*a* algorithm can be summarized as: (1) no requirement of any prior knowledge about optical response to the chl-*a* concentration; (2) able to account for nonlinear effects of predictors on the chl-*a* estimation; (3) it is possible to examine the significance and influence of each of the predictors in the chl-*a* retrievals. In this study, an accurate GAM chl-*a* algorithm was established using the predictors of 10 visible bands of R_{rs} and water depth over BYS, and all the predictors had a significant effect on the chl-*a* estimation in the model (Table 2). The effects of each R_{rs} band and water depth on the chl-*a* estimation were clearly described by smooth function plots (Figure 6(*a*)–(*k*)), but a more detailed examination is out of the scope of this study. It is interesting that the $R_{rs,412}$ and $R_{rs,443}$ presented the most significant effects on the chl-*a* estimation, although the short wavelength bands usually have poor accuracy due to improper atmospheric correction in the BYS (Cui et al. 2014). This phenomena indicates that the GAM algorithm may accommodate some errors originate from the improper atmospheric correction in R_{rs}, especially in the blue wavelength band. As a good indicator of suspended sediments concentration in the BYS (Bi et al. 2011; Wang et al. 2014), the $R_{rs,555}$ also presents strong effect on chl-*a* estimation in the GAM model, which demonstrates that the GAM algorithm may decrease the influence of suspended sediment imposing on chl-*a* estimation in turbid waters. In summary, the improved accuracy in the GAM chl-*a* products might be attributed to the following aspects: (1) the GAM used more optical bands (which may capture more detailed features of the complex coastal waters in the model) and also considered some geographical features (represented by water depth); (2) the statistically based GAM algorithm might be less susceptible to the R_{rs} errors than the optical response based OC3M and GSM01 algorithms; (3) the GAM can eliminate some of biases originate from the improper atmospheric correction in R_{rs}; (4) the GAM algorithm can better accommodate the effect of suspended sediments on chl-*a* estimation in turbid coastal waters.

In future studies, to further improve the performance of GAM algorithm, the following schemes could be considered. First, the standard MODIS R_{rs} data used in the GAM may have errors resulting from the failure of the standard-NIR atmospheric correction in the BYS (Siegel et al. 2000; Wang 2007; Wang, Tang, and Shi 2007). Hence, an accurate R_{rs} with shortwave infrared atmospheric correction may improve the accuracy of the GAM algorithm. Second, the GAM in this study only used visible band R_{rs} in the model development to make a comparison between standard algorithms (OC3M and GSM)

and the GAM algorithm. However, some of the previous studies suggested that the band ratios and/or other band combinations in the model can improved the retrieved accuracy (O'Reilly et al. 2000; Siswanto et al. 2011; Hu, Lee, and Franz 2012); thus, it is proposed that some new selections and combinations of bands might be used to improve the performance of the GAM algorithm.

5. Conclusion

In this study, a novel statistical algorithm for satellite-derived chl-*a* was established based on the GAM using *in situ* measured chl-*a* and standard remote-sensing reflectance at available MODIS visible bands. The accuracy of the GAM and the global standard chl-*a* algorithms (OC3M and GSM01) were evaluated and compared using the *in situ* measured chl-*a* collected from five independent cruises over the BYS. The OC3M and GSM01 algorithms present poor performance, as indicated by the large biases in absolute values (MAPD is 222.27% and 118.08%, for OC3M and GSM01, respectively) and the low correlation in relative patterns (R^2 is 0.25 and 0.07, for OC3M and GSM01, respectively). Hence, the standard chl-*a* products produced by global standard chl-*a* algorithms must be used carefully when investigating chl-*a* features in the BYS. The GAM algorithm produced a superior performance in estimating the chl-*a* concentrations in the BYS when compared to the standard OC3M and GSM01 algorithms. The MAPD and R^2 calculated from the GAM derived and *in situ* measured chl-*a* are 39.96% and 0.67, respectively. We also indicate how, in a future study, the GAM algorithm, which is an effective method for deriving accurate chl-*a* concentration from the MODIS satellite reflectance in the BYS can be further developed to improve its performance.

Acknowledgement

The authors would like to thank the Ocean Biology Processing Group of NASA for providing the MODIS data set (http://oceancolor.gsfc.nasa.gov/).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was supported by the National Natural Science Foundation of China [41376121]; the Strategic Priority Research Program of the Chinese Academy of Sciences [XDA11020405]; and the Natural Science Foundation of Shandong Province [JQ201414].

References

Antoine, D., J.-M. André, and A. Morel. 1996. "Oceanic Primary Production: 2. Estimation at Global Scale from Satellite (Coastal Zone Color Scanner) Chlorophyll." *Global Biogeochemical Cycles* 10 (1): 57–69. doi:10.1029/95GB02832.

Banks, A. C., P. Prunet, J. Chimot, P. Pina, J. Donnadille, E. Jeansou, M. L. Lux, et al. 2012. "A Satellite Ocean Color Observation Operator System for Eutrophication Assessment in Coastal Waters." *Journal of Marine Systems* 94 (S): S2–S15. doi:10.1016/j.jmarsys.2011.11.001.

Bi, N., Z. Yang, H. Wang, D. Fan, X. Sun, and K. Lei. 2011. "Seasonal Variation of Suspended-Sediment Transport through the Southern Bohai Strait." *Estuarine, Coastal and Shelf Science* 93 (3): 239–247. doi:10.1016/j.ecss.2011.03.007.

Boyce, D. G., M. R. Lewis, and B. Worm. 2010. "Global Phytoplankton Decline over the past Century." *Nature* 466 (7306): 591–596. doi:10.1038/nature09268.

Bozdogan, H. 1987. "Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions." *Psychometrika* 52 (3): 345–370. doi:10.1007/BF02294361.

Campbell, J. W. 1995. "The Lognormal Distribution as a Model for Bio-Optical Variability in the Sea." *Journal of Geophysical Research* 100 (C7): 13237–13254. doi:10.1029/95jc00458.

Carder, K. L., F. R. Chen, Z. P. Lee, S. K. Hawes, and D. Kamykowski. 1999. "Semianalytic Moderate-Resolution Imaging Spectrometer Algorithms for Chlorophyll a and Absorption with Bio-Optical Domains Based on Nitrate-Depletion Temperatures." *Journal of Geophysical Research: Oceans* 104 (C3): 5403–5421. doi:10.1029/1998jc900082.

Chen, C.-T. A. 2009. "Chemical and Physical Fronts in the Bohai, Yellow and East China Seas." *Journal of Marine Systems* 78 (3): 394–410. doi:10.1016/j.jmarsys.2008.11.016.

Chen, J., and J. Liu. 2015. "The Spatial and Temporal Changes of Chlorophyll-A and Suspended Matter in the Eastern Coastal Zones of China during 1997–2013." *Continental Shelf Research* 95: 89–98. doi:10.1016/j.csr.2015.01.004.

Chen, J., and W. Quan. 2013. "An Improved Algorithm for Retrieving Chlorophyll-A from the Yellow River Estuary Using MODIS Imagery." *Environmental Monitoring and Assessment* 185 (3): 2243–2255. doi:10.1007/s10661-012-2705-y.

Cui, T., J. Zhang, J. Tang, S. Sathyendranath, S. Groom, Y. Ma, W. Zhao, and Q. Song. 2014. "Assessment of Satellite Ocean Color Products of MERIS, MODIS and Seawifs along the East China Coast (In the Yellow Sea and East China Sea)." *ISPRS Journal of Photogrammetry and Remote Sensing* 87: 137–151. doi:10.1016/j.isprsjprs.2013.10.013.

Dall'Olmo, G., A. A. Gitelson, D. C. Rundquist, B. Leavitt, T. Barrow, and J. C. Holz. 2005. "Assessing the Potential of Seawifs and MODIS for Estimating Chlorophyll Concentration in Turbid Productive Waters Using Red and Near-Infrared Bands." *Remote Sensing of Environment* 96 (2): 176–187. doi:10.1016/j.rse.2005.02.007.

Dasgupta, S., R. P. Singh, and M. Kafatos. 2009. "Comparison of Global Chlorophyll Concentrations Using MODIS Data." *Advances in Space Research* 43 (7): 1090–1100. doi:10.1016/j.asr.2008.11.009.

Dave, A. C., and M. S. Lozier. 2015. "The Impact of Advection on Stratification and Chlorophyll Variability in the Equatorial Pacific." *Geophysical Research Letters* 42 (11): 4523–4531. doi:10.1002/2015GL063290.

Dierssen, H. M. 2010. "Perspectives on Empirical Approaches for Ocean Color Remote Sensing of Chlorophyll in a Changing Climate." *Proceedings of the National Academy of Sciences* 107 (40): 17073–17078. doi:10.1073/pnas.0913800107.

Franz, B. A., and P. J. Werdell. 2010. "A Generalized Framework for Modeling of Inherent Optical Properties in Ocean Remote Sensing Applications." *Proceedings of Ocean Optics, Anchorage, Alaska*, September 27–October 1.

González Taboada, F., and R. Anadón. 2014. "Seasonality of North Atlantic Phytoplankton from Space: Impact of Environmental Forcing on a Changing Phenology (1998–2012)." *Global Change Biology* 20 (3): 698–712. doi:10.1111/gcb.12352.

Gordon, H. R. 1997. "Atmospheric Correction of Ocean Color Imagery in the Earth Observing System Era." *Journal of Geophysical Research: Atmospheres* 102 (D14): 17081–17106. doi:10.1029/96jd02443.

Gower, J., R. Doerffer, and G. Borstad. 1999. "Interpretation of the 685nm Peak in Water-Leaving Radiance Spectra in Terms of Fluorescence, Absorption and Scattering, and Its Observation by MERIS." *International Journal of Remote Sensing* 20 (9): 1771–1786. doi:10.1080/014311699212470.

Gower, J., S. King, G. Borstad, and L. Brown. 2005. "Detection of Intense Plankton Blooms Using the 709 Nm Band of the MERIS Imaging Spectrometer." *International Journal of Remote Sensing* 26 (9): 2005–2012. doi:10.1080/01431160500075857.

Gregg, W. W., N. W. Casey, J. E. O'Reilly, and W. E. Esaias. 2009. "An Empirical Approach to Ocean Color Data: Reducing Bias and the Need for Post-Launch Radiometric Re-Calibration." *Remote Sensing of Environment* 113 (8): 1598–1612. doi:10.1016/j.rse.2009.03.005.

Gregg, W. W., and M. E. Conkright. 2002. "Decadal Changes in Global Ocean Chlorophyll." *Geophysical Research Letters* 29 (15): 20–24. doi:10.1029/2002GL014689.

Hastie, T., and R. Tibshirani. 1986. "Generalized Additive Models." *Statistical Science* 1: 297–310. doi:10.1214/ss/1177013604.

Hastie, T. J., and R. J. Tibshirani. 1990. *Generalized Additive Models*. Vol. 43. London: Chapman and Hall/CRC.

He, X. Q., Y. Bai, D. Pan, C. Chen, D. F. Wang, and F. Gong. 2013. "Satellite Views of Seasonal and Inter-Annual Variability of Phytoplankton Blooms in the Eastern China Seas over the past 14 Yr (1998-2011)." *Biogeosciences* 10: 4721–4739. doi:10.5194/bg-10-4721-2013.

Hu, C., Z. Lee, and B. Franz. 2012. "Chlorophyll Aalgorithms for Oligotrophic Oceans: A Novel Approach Based on Three-Band Reflectance Difference." *Journal of Geophysical Research: Oceans* 117 (C1): C01011. doi:10.1029/2011jc007395.

Hu, C., D. Li, C. Chen, J. Ge, F. E. Muller-Karger, J. Liu, F. Yu, and M. He. 2010. "On the Recurrent Ulva Prolifera Blooms in the Yellow Sea and East China Sea." *Journal of Geophysical Research* 115 (C5): C05017. doi:10.1029/2009jc005561.

Huang, J., L. Chen, X. Chen, and Q. Song. 2013. "Validation of Semi-Analytical Inversion Models for Inherent Optical Properties from Ocean Color in Coastal Yellow Sea and East China Sea." *Journal of Oceanography* 69 (6): 713–725. doi:10.1007/s10872-013-0202-8.

Ioannou, I., A. Gilerson, B. Gross, F. Moshary, and S. Ahmed. 2013. "Deriving Ocean Color Products Using Neural Networks." *Remote Sensing of Environment* 134: 78–91. doi:10.1016/j.rse.2013.02.015.

Irwin, A. J., and Z. V. Finkel. 2008. "Mining a Sea of Data: Deducing the Environmental Controls of Ocean Chlorophyll." *Plos ONE* 3 (11): e3836. doi:10.1371/journal.pone.0003836.

Kitsiou, D., and M. Karydis. 2011. "Coastal Marine Eutrophication Assessment: A Review on Data Analysis." *Environment International* 37 (4): 778–801. doi:10.1016/j.envint.2011.02.004.

Komick, N. M., M. P. F. Costa, and J. Gower. 2009. "Bio-Optical Algorithm Evaluation for MODIS for Western Canada Coastal Waters: an Exploratory Approach Using in Situ Reflectance." *Remote Sensing of Environment* 113 (4): 794–804. doi:10.1016/j.rse.2008.12.005.

Le, C., C. Hu, J. Cannizzaro, D. English, F. Muller-Karger, and Z. Lee. 2013. "Evaluation of Chlorophyll-A Remote Sensing Algorithms for an Optically Complex Estuary." *Remote Sensing of Environment* 129: 75–89. doi:10.1016/j.rse.2012.11.001.

Lee, Z., K. L. Carder, and R. A. Arnone. 2002. "Deriving Inherent Optical Properties from Water Color: A Multiband Quasi-Analytical Algorithm for Optically Deep Waters." *Applied Optics* 41 (27): 5755–5772. doi:10.1364/AO.41.005755.

Liu, D., J. K. Keesing, Q. Xing, and P. Shi. 2009. "World's Largest Macroalgal Bloom Caused by Expansion of Seaweed Aquaculture in China." *Marine Pollution Bulletin* 58 (6): 888–895. doi:10.1016/j.marpolbul.2009.01.013.

Liu, D., and Y. Wang. 2013. "Trends of Satellite Derived Chlorophyll-A (1997-2011) in the Bohai and Yellow Seas, China: Effects of Bathymetry on Seasonal and Inter-Annual Patterns." *Progress In Oceanography* 116: 154–166. doi:10.1016/j.pocean.2013.07.003.

Lorenzen, C. J. 1967. "Determination of Chlorophyll and Pheo-Pigments: Spectrophotometric Equations1." *Limnology and Oceanography* 12 (2): 343–346. doi:10.4319/lo.1967.12.2.0343.

Maritorena, S., D. A. Siegel, and A. R. Peterson. 2002. "Optimization of a Semianalytical Ocean Color Model for Global-Scale Applications." *Applied Optics* 41 (15): 2705–2714. doi:10.1364/AO.41.002705.

Moore, T. S., J. W. Campbell, and M. D. Dowell. 2009. "A Class-Based Approach to Characterizing and Mapping the Uncertainty of the MODIS Ocean Chlorophyll Product." *Remote Sensing of Environment* 113 (11): 2424–2430. doi:10.1016/j.rse.2009.07.016.

Moses, W. J., A. A. Gitelson, S. Berdnikov, and V. Povazhnyy. 2009. "Satellite Estimation of Chlorophyll-a Concentration Using the Red and NIR Bands of MERIS—The Azov Sea Case Study." *IEEE Geoscience and Remote Sensing Letters* 6 (4): 845–849. doi:10.1109/lgrs.2009.2026657.

O'Reilly, J. E., S. Maritorena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. McClain. 1998. "Ocean Color Chlorophyll Algorithms for Seawifs." *Journal of Geophysical Research: Oceans* 103 (C11): 24937–24953. doi:10.1029/98JC02160.

O'Reilly, J. E., S. Maritorena, D. A. Siegel, M. C. O'Brien, D. Toole, B. G. Mitchell, M. Kahru, et al. 2000. "Ocean Color Chlorophyll a Algorithms for Seawifs, OC2, and OC4: Version 4." *Seawifs Postlaunch Calibration and Validation Analyses, Part* 3: 9–23.

Park, Y.-J., K. Ruddick, and G. Lacroix. 2010. "Detection of Algal Blooms in European Waters Based on Satellite Chlorophyll Data from MERIS and MODIS." *International Journal of Remote Sensing* 31 (24): 6567–6583. doi:10.1080/01431161003801369.

Raitsos, D. E., G. Korres, G. Triantafyllou, G. Petihakis, M. Pantazi, K. Tsiaras, and A. Pollani. 2012. "Assessing Chlorophyll Variability in Relation to the Environmental Regime in Pagasitikos Gulf, Greece." *Journal of Marine Systems* 94 (S): S16–S22. doi:10.1016/j.jmarsys.2011.11.003.

Shang, S. L., Q. Dong, C. M. Hu, G. Lin, Y. H. Li, and S. P. Shang. 2014. "On the Consistency of MODIS Chlorophyll a Products in the Northern South China Sea." *Biogeosciences* 11 (2): 269–280. doi:10.5194/bg-11-269-2014.

Shi, W., and M. Wang. 2010a. "Characterization of Global Ocean Turbidity from Moderate Resolution Imaging Spectroradiometer Ocean Color Observations." *Journal of Geophysical Research* 115 (C11): C11022. doi:10.1029/2010jc006160.

Shi, W., and M. Wang. 2010b. "Satellite Observations of the Seasonal Sediment Plume in Central East China Sea." *Journal of Marine Systems* 82 (4): 280–285. doi:10.1016/j.jmarsys.2010.06.002.

Shi, W., and M. Wang. 2012. "Satellite Views of the Bohai Sea, Yellow Sea, and East China Sea." *Progress In Oceanography* 104: 30–45. doi:10.1016/j.pocean.2012.05.001.

Siegel, D. A., M. Wang, S. Maritorena, and W. Robinson. 2000. "Atmospheric Correction of Satellite Ocean Color Imagery: the Black Pixel Assumption." *Applied Optics* 39 (21): 3582–3591. doi:10.1364/AO.39.003582.

Siswanto, E., J. Tang, H. Yamaguchi, Y. Ahn, J. Ishizaka, S. Yoo, S.-W. Kim, et al. 2011. "Empirical Ocean-Color Algorithms to Retrieve Chlorophyll-A, Total Suspended Matter, and Colored Dissolved Organic Matter Absorption Coefficient in the Yellow and East China Seas." *Journal of Oceanography* 67 (5): 627–650. doi:10.1007/s10872-011-0062-z.

Son, S., M. Wang, and J.-K. Shon. 2011. "Satellite Observations of Optical and Biological Properties in the Korean Dump Site of the Yellow Sea." *Remote Sensing of Environment* 115 (2): 562–572. doi:10.1016/j.rse.2010.10.002.

Stock, A. 2015. "Satellite Mapping of Baltic Sea Secchi Depth with Multiple Regression Models." *International Journal of Applied Earth Observation and Geoinformation* 40: 55–64. doi:10.1016/j.jag.2015.04.002.

Tang, D., H. Kawamura, H. Doan-Nhu, and W. Takahashi. 2004. "Remote Sensing Oceanography of a Harmful Algal Bloom off the Coast of Southeastern Vietnam." *Journal of Geophysical Research: Oceans* 109 (C3): C03014. doi:10.1029/2003jc002045.

Tilstone, G. H., A. A. Lotliker, P. I. Miller, P. M. Ashraf, T. S. Kumar, T. Suresh, B. R. Ragavan, and H. B. Menon. 2013. "Assessment of MODIS-Aqua Chlorophyll-A Algorithms in Coastal and Shelf Waters of the Eastern Arabian Sea." *Continental Shelf Research* 65: 14–26. doi:10.1016/j.csr.2013.06.003.

Urquhart, E. A., B. F. Zaitchik, M. J. Hoffman, S. D. Guikema, and E. F. Geiger. 2012. "Remotely Sensed Estimates of Surface Salinity in the Chesapeake Bay: A Statistical Approach." *Remote Sensing of Environment* 123: 522–531. doi:10.1016/j.rse.2012.04.008.

Vilas, L. G., E. Spyrakos, and J. M. Torres Palenzuela. 2011. "Neural Network Estimation of Chlorophyll a from MERIS Full Resolution Data for the Coastal Waters of Galician Rias (NW Spain)." *Remote Sensing of Environment* 115 (2): 524–535. doi:10.1016/j.rse.2010.09.021.

Wang, B.-D., X.-L. Wang, and R. Zhan. 2003. "Nutrient Conditions in the Yellow Sea and the East China Sea." *Estuarine, Coastal and Shelf Science* 58 (1): 127–136. doi:10.1016/s0272-7714(03)00067-2.

Wang, H., A. Wang, N. Bi, X. Zeng, and H. Xiao. 2014. "Seasonal Distribution of Suspended Sediment in the Bohai Sea, China." *Continental Shelf Research* 90: 17–32. doi:10.1016/j.csr.2014.03.006.

Wang, M. 2007. "Remote Sensing of the Ocean Contributions from Ultraviolet to Near-Infrared Using the Shortwave Infrared Bands: Simulations." *Applied Optics* 46 (9): 1535–1547. doi:10.1364/AO.46.001535.

Wang, M., J. Ahn, L. Jiang, W. Shi, S. Son, Y. Park, and J. Ryu. 2013. "Ocean Color Products from the Korean Geostationary Ocean Color Imager (GOCI)." *Optics Express* 21 (3): 3835–3849. doi:10.1364/OE.21.003835.

Wang, M., and W. Shi. 2007. "The NIR-SWIR Combined Atmospheric Correction Approach for MODIS Ocean Color Data Processing." *Optics Express* 15 (24): 15722–15733. doi:10.1364/OE.15.015722.

Wang, M., J. Tang, and W. Shi. 2007. "MODIS-Derived Ocean Color Products along the China East Coastal Region." *Geophysical Research Letters* 34 (6): L06611. doi:10.1029/2006gl028599.

Wang, P., E. S. Boss, and C. Roesler. 2005. "Uncertainties of Inherent Optical Properties Obtained from Semianalytical Inversions of Ocean Color." *Applied Optics* 44 (19): 4074–4085. doi:10.1364/AO.44.004074.

Werdell, P. J., S. W. Bailey, B. A. Franz, L. W. Harding Jr. Jr, G. C. Feldman, and C. R. McClain. 2009. "Regional and Seasonal Variability of Chlorophyll-A in Chesapeake Bay as Observed by Seawifs and MODIS-Aqua." *Remote Sensing of Environment* 113 (6): 1319–1330. doi:10.1016/j.rse.2009.02.012.

Wood, S. N. 2004. "Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models." *Journal of the American Statistical Association* 99 (467): 673–686. doi:10.1198/016214504000000980.

Wood, S. N. 2006. *Generalized Additive Models: an Introduction with R*. Boca Raton, FL: Chapman and Hall/CRC.

Yamaguchi, H., H.-C. Kim, Y. B. Son, S. W. Kim, K. Okamura, Y. Kiyomoto, and J. Ishizaka. 2012. "Seasonal and Summer Interannual Variations of Seawifs Chlorophyll a in the Yellow Sea and East China Sea." *Progress In Oceanography* 105: 22–29. doi:10.1016/j.pocean.2012.04.004.