A Unified Model for Remotely Estimating Chlorophyll a in Lake Taihu, China, Based on SVM and In Situ Hyperspectral Data

Article *in* IEEE Transactions on Geoscience and Remote Sensing · September 2009

DOI: 10.1109/TGRS.2009.2014688 · Source: IEEE Xplore

CITATIONS
69

READS
220

3 authors, including:

Deyong Sun
Nanjing University of Information Science & Technology
79 PUBLICATIONS **1,518** CITATIONS

SEE PROFILE

Qiao Wang
Zhejiang University
180 PUBLICATIONS **2,563** CITATIONS

SEE PROFILE

A Unified Model for Remotely Estimating Chlorophyll a in Lake Taihu, China, Based on SVM and *In Situ* Hyperspectral Data

Deyong Sun, Yunmei Li, and Qiao Wang

***Abstract*—Accurate estimation of chlorophyll a (Chla) in inland turbid lake waters by means of remote sensing is a challenging task due to their optical complexity. In order to explore the best solution, we observed water quality parameters and measured water reflectance spectra in Lake Taihu for 14 days in November 2007. After initial wavelength analysis and iterative optimization, the best three-wavelength factor (TWF) was determined as $[R_{\mathrm{rs}}^{-1}(661) - R_{\mathrm{rs}}^{-1}(691)]R_{\mathrm{rs}}(727)$. Linear models and a support vector machine (SVM) model with TWFs as the inputs were established for retrieving Chla concentration level. It is found that linear models with a single TWF performed worse than the SVM model. The SVM model is highly accurate, whose R^2 and root-mean-square error are 0.8961 and 2.67 mg/m^3, respectively. Validation of the SVM model using data sets obtained at another sampling time reveals small errors. Thus, this model can be used to extract Chla concentration levels in Lake Taihu waters but is not restricted by the sampling time. These findings underline the rationale of the TWF model and demonstrate the robustness of the SVM algorithm for remotely estimating Chla in Lake Taihu waters.**

***Index Terms*—Chlorophyll a (Chla), hyperspectral data, Lake Taihu, support vector machine (SVM), three-wavelength factor (TWF).**

I. Introduction

In OPEN-OCEAN case-I waters, Chla concentration is typically derived from the blue and green spectral bands [1], [2]. Application of case-I-derived algorithms to case-II productive waters that contain widely variable and uncorrelated Chla, suspended solids, and dissolved matters often results in a poor prediction [3]–[6]. This is because absorption by colored dissolved organic matter (CDOM), nonalgal particulates, and phytoplankton pigments overlap in the blue spectrum. This overlap degrades the usefulness of this spectral range for Chla retrieval. In case-II waters, nonalgal particulates and CDOM sometimes dominate water column optics [7]. For example, spring phytoplankton accounts for 12.5%–26.4% of total absorption of water column (except pure water) in typical visible light bands in Lake Taihu, China [8]. Owing to the optical complexity of case-II waters, accurate retrieval of Chla concentration in turbid productive waters by means of remote sensing is challenging.

In such waters, Gordon *et al.* [35] proposed a model for depicting the relationship between spectral reflectance, $R_{\mathrm{rs}}(\lambda)$, and the inherent optical properties of water, namely, the total absorption (a) and the backscattering (b_b) coefficients. This model considerably broadens our knowledge about the effects of inherent optical properties of water components on water reflectance spectra. On the basis of this relationship, a three-wavelength-factor (TWF) model has been developed for Chla estimation with highly accurate results [5], [6], [9], [10]. It has the following form:

$$\mathrm{Chla} \propto \left[R_{\mathrm{rs}}^{-1}(\lambda_1) - R_{\mathrm{rs}}^{-1}(\lambda_2)\right] \times R_{\mathrm{rs}}(\lambda_3) \tag{1}$$

where $R_{\mathrm{rs}}(\lambda_i)$ stands for remotely sensed water reflectance at wavelength λ_i, usually in the visible light range of 400–800 nm.

For turbid productive waters, the total absorption coefficient is a sum of the absorption coefficients of phytoplankton pigment (mainly Chla), CDOM, nonalgal particles, and pure water. To retrieve Chla concentration, the absorption coefficient of phytoplankton pigment a_{ph} must be isolated. To accomplish this, the reciprocal reflectance at wavelength λ_1 in (1) should be maximally sensitive to a_{ph}. However, in turbid productive waters, $R_{\mathrm{rs}}^{-1}(\lambda_1)$ might also be strongly affected by absorption of CDOM, nonalgal particles, and pure water, and backscattering b_b by all particulate matters, in addition to absorption by Chla. The effect of $(a_{\mathrm{nap}} + a_{\mathrm{CDOM}})$ and b_b can be minimized using a second wavelength at which $R_{\mathrm{rs}}^{-1}(\lambda_2)$ is the least sensitive to the absorption by Chla $(a_{\mathrm{ph}}(\lambda_2) \ll a_{\mathrm{ph}}(\lambda_1))$, and the absorption coefficients of nonalgal particles and CDOM $(a_{\mathrm{nap}} + a_{\mathrm{CDOM}})$ at wavelength λ_2 are quite close to that at wavelength λ_1. The difference $R_{\mathrm{rs}}^{-1}(\lambda_1) - R_{\mathrm{rs}}^{-1}(\lambda_2)$ in (1), however, is still affected by b_b, so the estimated Chla value is subject to variations in backscattering among different samples. To account for the influence of this variability, a third wavelength λ_3 has been suggested. It minimizes the effect of constituent absorption, $(a_{\mathrm{ph}} + a_{\mathrm{nap}} + a_{\mathrm{CDOM}}) - 0$, and hence $a(\lambda_3) \sim a_w$. The near-infrared (NIR) range beyond 710 nm, where $a \gg b_b$ and $R_{\mathrm{NIR}} \propto b_b$, is an excellent candidate for λ_3 as it meets the requirement.

It is possible to obtain relatively high accuracies with the TWF model. For instance, Zimba and Gitelson [9] found

Manuscript received September 27, 2008; revised December 20, 2008 and January 18, 2009. First published March 31, 2009; current version published July 23, 2009. This work was supported in part by the National Science and Technology Support Project of China under Grant 2008BAC34B05, by the Foundation of Graduate Innovation Training Plan of Jiangsu Province under Grant CX08B_015Z, by the Excellent Thesis Cultivation Fund of Nanjing Normal University under Grant 181200000220, and by the National Natural Science Foundation under Grant 40571110.

The authors are with the Key Laboratory of Virtual Geographic Environment, Nanjing Normal University, Ministry of Education, Nanjing 210046, China (e-mail: sundeyong1984@163.com; liyunmei@njnu.edu.cn; wangqiao@sepa.gov.cn).

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

Digital Object Identifier 10.1109/TGRS.2009.2014688

a strong linear relationship between measured Chla concentrations and those inverted from the TWF $[R_{\rm rs}^{-1}(650) - R_{\rm rs}^{-1}(710)]R_{\rm rs}(740)$. Nevertheless, its performance does vary with the unique optical properties of a water body distinctly. Thus, Gitelson *et al.* [6] proposed a group of optimal wavelengths of 675, 695, and 730 nm in estimating Chla in the turbid productive Chesapeake Bay water. The optimal wavelength ranges of λ_1, λ_2, and λ_3 are considered, respectively, as 660–670, 700–730, and 740–760 nm after investigating many lakes and reservoirs in Nebraska and Iowa [10]. Xu *et al.* [11] used the optimal wavelengths of 690, 693, and 799 nm to estimate Chla concentrations in Lake Xinmiao, China. After studying the optimal positions of wavelengths λ_1, λ_2, and λ_3 in Lake Taihu, China, Zhou *et al.* [12] found their values to be 666, 688, and 725 nm, respectively. However, it remains unknown whether these wavelengths are still optimal in other seasons as they were derived from samples collected in a single season. This paper overcomes this limitation by identifying the best wavelengths from samples collected at one time and validating the results against samples collected in another time. In this way, the universality of the identified optimal wavelength is known. Such knowledge is essential in accurately estimating Chla in a given turbid water body.

The objectives of this paper are as follows: 1) to find the optimal TWFs for turbid Lake Taihu waters; 2) to develop various models for the accurate estimation of Chla, including two linear models with a single TWF and a support vector machine (SVM) model with three TWFs; and 3) to compare the performance of the proposed models with that of linear models and the SVM model, and select the best one for Lake Taihu waters. These proposed models should be more accurate and have wide applicability that is not limited by sampling time.

II. Materials and Methods

A. Stations and Sampling

The study area is Lake Taihu, one of the largest freshwater lakes in China. It has an area of 2338.1 km^2. As a typical shallow inland lake, it has an average depth of 1.9 m [13]. Surface water samples were collected using the Niskin bottles at 74 stations distributed regularly over the entire lake from 8th to 21st of November 2007 (Fig. 1). Then, they were immediately deep-frozen for preservation. These water samples were brought back to laboratory for analysis within the same day. During sampling, climatic conditions were also recorded, including wind speed, wind direction, water temperature, etc.

B. In Situ Measurement of Hyperspectral Reflectance

Water reflectance spectra were concurrently measured using a portable ASD FieldSpec spectroradiometer. With a viewing field of 25°, this instrument has a sensitivity range from 350 to 1050 nm at an increment of 1.58 nm with 512 wavebands. The measurement of water surface reflectance spectra followed the method described by Mueller *et al.* [14] and Jiao *et al.* [15]. When the boat was anchored, the radiance spectra of the reference panel, water, and the sky were measured ten times

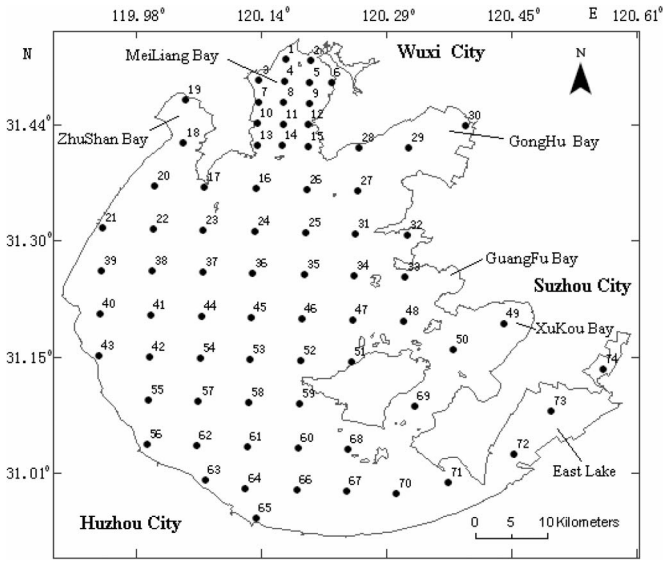


Fig. 1. Sketch map shows the distribution of sample stations.

each. The spectra at each sampling station were averaged to derive $R_{\rm rs}(\lambda)$. In the process of deriving $R_{\rm rs}(\lambda)$, the reflectance of skylight at the air–water surface (r) was taken as 2.2% for calm weather, 2.5% for wind speed of up to 5 m/s, and 2.6%–2.8% for wind speed of about 10 m/s, and the reflectance of the gray board (ρ_p) has been accurately calibrated to 30%.

C. Phytoplankton Pigment Absorption Chla Measurements

The measurements of phytoplankton pigment absorption coefficient $a_{\rm ph}(\lambda)$, nonalgal particle absorption $a_{\rm NAP}(\lambda)$, and CDOM absorption $a_{\rm CDOM}(\lambda)$ were carried out using the quantitative filter technique [16], which was described in detail by Sun *et al.* [8]. The Chla, total suspended matter (TSM), inorganic suspended matter (ISM), organic suspended matter (OSM) concentrations were measured according to the investigation criteria about lakes of China [17]. Chla was extracted with ethanol (90%) at 80° and analyzed spectrophotometrically at 665 and 750 nm, with correction for phaeopigments.

D. Data Selection and Analysis

At some stations, the reflectance spectra were very similar to that of vegetation but bore less resemblance to water's because of contamination by algal blooms. They were discarded. Lake Taihu is an optically shallow lake where bottom reflectance strongly affects water spectral reflectance in the shallower portion [18]. To ensure the highest quality of data sets, samples contaminated by bottom reflectance were eliminated, leaving 47 available for further analyses. These samples were randomly divided into two parts: a training data set of 32 samples and a validation data set of 15 samples. The linear regression and SVM (to be described in the next section) methods were used to develop Chla estimation models using the training data set. The mean absolute percentage error (MAPE) and root-mean-square error (rmse) were calculated for model validation

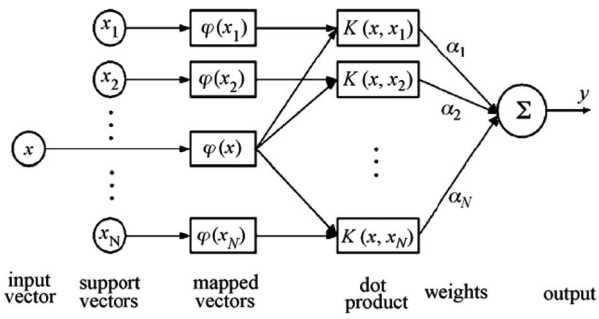


Fig. 2. Schematic structure of the SVM model.

using the second data set. These accuracy criteria are calculated as

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_i'}{y_i}\right| \tag{2}$$

$$rmse = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2} \tag{3}$$

where n is the number of samples, y_i is the measured value, and y_i' is the estimated value.

E. SVM

SVM is a universal learning method proposed by Vapnik [19], which has been widely applied to remote sensing research at present [20]–[22]. Its structure is shown in Fig. 2. The basic concept of SVM for regression is to map the data set x into a high-dimensional feature space via nonlinear mapping and to solve the linear regression problem in this feature space. Hence, given a set of data $T = (x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_l, y_l) \in (X \times Y)^l$, where x_i denotes the input vector, y_i denotes the corresponding output value, X and Y separately denote the input vector data set and the corresponding output value data set, and l denotes the total number of data patterns, the SVM regression function is expressed as

$$f(x) = w \cdot \varphi(x) + b \tag{4}$$

where $\varphi(x)$ denotes the nonlinear mapping function, w denotes the weight vector, and b denotes the bias term. The coefficients w and b can be estimated by minimizing the regularized risk function. For more detailed description of SVM for regression, please refer to the tutorial on support vector regression [23].

SVM is a kind of learning method based on kernel functions. Three kernel functions have been commonly used in SVM: radial basis function (RBF) $K(x, x_i) = \exp(-\|x - x_i\|^2/2\sigma^2)$, polynomial basis function $K(x, x_i) = ((x \cdot x_i) + b)^d$ $(b \geq 0$, d is the natural number), and sigmoid function $K(x, x_i) = \tanh(k(x \cdot x_i) + v)(k > 0, v < 0)$. In the RBF kernel, only one variable needs to be determined; thus, fewer free parameters in SVM need to be resolved, which is beneficial to SVM parameter optimization. SVM constructed from the RBF kernel has an excellent nonlinear forecasting performance. In addition, the RBF kernel has less numerical difficulties than other kernels [24]. By comparison, the polynomial basis function or sigmoid function parameter optimization is complex because two variables in these two functions need determination. Therefore, the RBF kernel was used in SVM in this paper.

There are two commonly used SVMs for regression, namely, the ε-SVR algorithm and its extension, the v-SVR algorithm [25]. The v-SVR algorithm was chosen for its significant advantage over ε-SVR in automatically adjusting the width of the ε-tube around the function being approximated. The v-SVR implementation was provided by the LibSVM software library [26].

TABLE I
DESCRIPTIVE STATISTICS OF THE BIO-OPTICAL PARAMETERS INCLUDING CHLA (CHLOROPHYLL-A), TSM, OSM, ISM, $a_{\mathrm{CDOM}}(440)$, $a_{\mathrm{ph}}(440)$, AND $a_{\mathrm{NAP}}(440)$ (ABSORPTION COEFFICIENTS OF CDOM, PHYTOPLANKTON PIGMENTS, AND NONALGAL PARTICLES AT 440 nm). SD: STANDARD DEVIATION; CV (= SD/MEAN OF PARAMETER) COEFFICIENT OF VARIATION IN PERCENT

Statistics	Minimum	Maximum	Mean	SD	CV (%)
Chla (mg m^{-3})	0.98	35.29	12.06	7.84	65
TSM (mg l^{-1})	8.53	116.60	32.65	24.03	74
OSM (mg l^{-1})	3.20	19.53	8.77	3.04	35
ISM (mg l^{-1})	2.60	103.40	23.89	23.00	96
$a_{\mathrm{CDOM}}(440)$ (m^{-1})	0.05	1.42	0.47	0.31	66
$a_{\mathrm{ph}}(440)$ (m^{-1})	0.15	3.03	0.92	0.65	71
$a_{\mathrm{NAP}}(440)$ (m^{-1})	0.51	7.10	1.75	1.27	73

III. RESULTS

A. Bio-Optical Properties of Lake Taihu Waters

The bio-optical parameters of Lake Taihu waters exhibit wide variations (Table I), which indicates the presence of complex water column conditions. Lake Taihu waters are highly turbid judging from the TSM variation, which ranges from 8.53 to 116.60 mg/L with a mean of 32.65 mg/L ($SD = 24.03$ mg/L). Mean TSM concentration is much higher than that of other turbid water bodies, such as Chesapeake Bay [6], sandpit lakes and reservoirs in eastern Nebraska, U.S. [27], and some other U.S. lakes [10]. ISM concentration ranges from 2.60 to 103.40 mg/L with a mean of 23.89 mg/L, which is higher than that of OSM (8.77 mg/L). In addition, it has the largest variation among all water quality parameters ($CV = 96\%$). TSM has a correlation coefficient of 0.99 with ISM, R(TSM, ISM), but only 0.39 with OSM. This implies that the ISM dominates the TSM in Lake Taihu. Nonalgal particulates have a mean water absorption coefficient of 1.75 m^{-1}($SD = 1.27$ m^{-1}), the largest among all water components, while the mean values of phytoplankton pigments and CDOM are 0.92 and 0.47 m^{-1}, respectively (Table I). Chla ranges from 0.98 to 35.29 mg/m^3 with a mean value of 12.06 mg/m^3 and a standard deviation of 7.84 mg/m^3.

In Lake Taihu waters, spectral reflectance is highly variable over the visible and NIR spectrum (Fig. 3). The spectra were quite similar in magnitude to typical reflectance spectra of turbid productive waters [6], [28]–[30]. The reflectance is relatively low over 400–500 nm, owing to the phytoplankton pigment absorption. The peak reflectance over 550–580 nm is due to weak absorption by phytoplankton pigment and cell-induced scattering. The relatively low reflectance at near 670 nm is attributed to the absorption by Chla [31]. Around

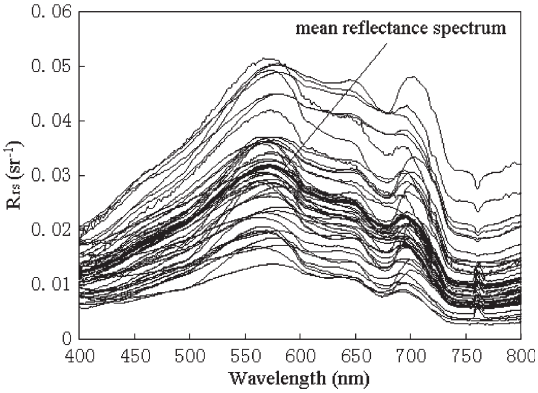


Fig. 3. Remote sensing reflectance spectra and mean spectrum in Lake Taihu waters.

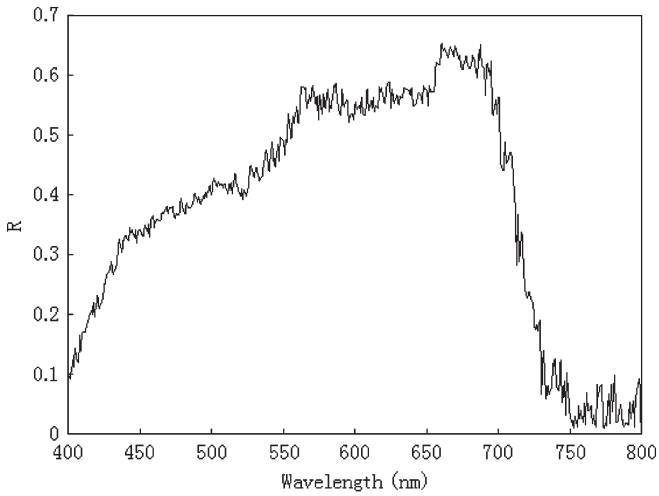


Fig. 4. Pearson correlation coefficients between $a_{\rm ph}(\lambda)$ and Chla concentrations.

700 nm is a noticeable reflectance peak, which indicates that the waters are rich in Chla, and hence, this peak is indicative of Chla abundance in water column [3]. As in other turbid productive waters [6], the position of this peak shifts from 688 nm toward about 706 nm as Chla increases from 9 mg/m^3 to above 70 mg/m^3. All these spectral characteristics are also obvious in the mean spectrum of remote sensing reflectance (thick line in Fig. 3).

B. TWF

For accurately estimating Chla concentration in turbid productive waters, it is imperative to minimize the influence of other components, mainly suspended particles and CDOM, on water reflectance spectra. The proper selection of wavelengths λ_1, λ_2, and λ_3 is critical to minimizing the influence. This selection requires initial wavelength determination and iterative analysis. λ_1, λ_2, and λ_3 all have their own characteristics. Relatively easier to determine than λ_2, the initial $\lambda_1$and λ_3 values were set first, followed by λ_2 via iterative optimization. λ_1 should be located around the wavelength where Chla concentration is the most sensitive to phytoplankton pigment absorption $a_{\rm ph}(\lambda)$. Pearson correlation analysis between Chla concentration and $a_{\rm ph}(\lambda)$ shows that the highest R of 0.6524 ($p < 0.001$) occurs at 661 nm (Fig. 4) because Chla's absorption peak in the red spectrum domain often appears in the vicinity of 670 nm. In addition, the peak position usually varies with the algal type in a water body. No highly noticeable correlation exists at another absorption peak, usually 440 nm in the blue. This absence is attributable to the effects of other pigment absorption [32]. Thus, 661 nm was selected as the initial wavelength for λ_1.

λ_3 serves to eliminate the influences of water light field distribution parameters Q, f, and total backscattering coefficient b_b. At this wavelength, the absorption coefficient of water components (except pure water) should be close to zero, namely, $a_t(\lambda_3) \approx a_w(\lambda_3)$ and $a_t(\lambda_3) \gg b_b(\lambda_3)$. Moreover, b_b should be approximately a constant at wavelengths of λ_1, λ_2, and λ_3. The 730–800 nm is identified as the best spectrum domain for λ_3 [29]. In this domain, pure water absorption coefficient at 750 nm is the highest, which can satisfy the condition of $a_t(\lambda_3) \gg b_b(\lambda_3)$ better. Therefore, the initial wavelength of λ_3 was set to 750 nm. The wavelength of λ_2 was determined via iterative optimization.

Pearson correlation coefficient (R) and the $rmse$ of Chla estimated from the linear regression model were used to evaluate the performance of all wavelengths. First, λ_1 and λ_3 were set, respectively, at 661 and 750 nm, and the optimal λ_2 was found to fall within 400–800 nm. Fig. 5(a) shows that the largest R is 0.8826 and the smallest $rmse$ is 3.73 mg/m^3, both at 691 nm. Therefore, λ_2 was set at 691 nm. Afterward, λ_3 was analyzed with λ_1 being 661 nm and λ_2 being 691 nm in the same manner. The optimal wavelength for λ_3 was found to be 727 nm. Optimization was undertaken iteratively until λ_1, λ_2, and λ_3 positions became stabilized. After four iterations, the optimal wavelengths of λ_1, λ_2, and λ_3 were set, respectively, at 661, 691, and 727 nm, and their corresponding R and $rmse$ are 0.8872 and 3.66 mg/m^3, respectively (Table II).

Initial wavelengths play an important role in TWF determination. Different initial wavelengths lead to different TWF results. Apart from the caution in setting the initial wavelengths, which was ignored in previous studies, we also examined the impact of different wavelengths. Review of initial wavelengths used in other studies identified two groups of wavelengths: 1) $\lambda_1 = 670$ nm and $\lambda_3 = 740$ nm, and 2) $\lambda_2 = 700$ nm and $\lambda_3 = 750$ nm. They were used to validate the proposed TWF model and to find optimal wavelengths for λ_1, λ_2, and λ_3. Use of the first group of values resulted in them being 664, 690, and 798 nm after 11 iterations of optimization. The corresponding R and $rmse$ were 0.8812 and 3.75 mg/m^3, both being worse than the respective figures above. The iteration took a long time to complete. The second group achieved the same accuracy, but the iteration took much longer time. This implies that these two groups of initial wavelengths are not the most appropriate for Lake Taihu waters.

C. Model Development

Two linear models $[R_{\rm rs}^{-1}(661) - R_{\rm rs}^{-1}(691)]R_{\rm rs}(727)$ and $[R_{\rm rs}^{-1}(664) - R_{\rm rs}^{-1}(690)]R_{\rm rs}(798)$ and an SVM model were developed for Chla estimation in Lake Taihu waters. The linear

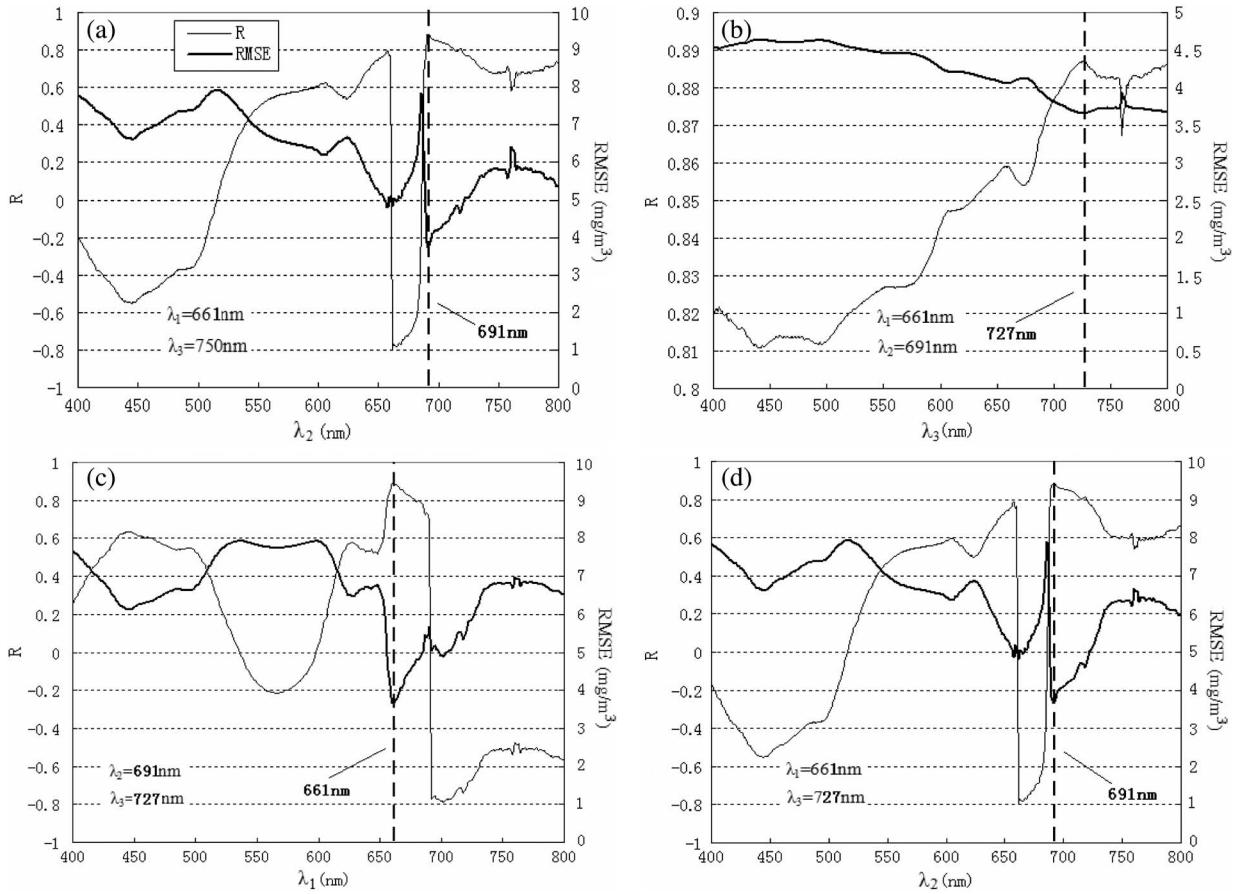


Fig. 5. R and $rmse$ variations in the course of iterative optimizations. (a) Finding optimal λ_2 with $\lambda_1 = 661$ nm and $\lambda_3 = 750$ nm. (b) Finding optimal λ_3 with $\lambda_1 = 661$ nm and $\lambda_2 = 691$ nm. (c) Finding optimal λ_1 with $\lambda_2 = 691$ nm and $\lambda_3 = 727$ nm. (d) Finding optimal λ_2 again with $\lambda_1 = 661$ nm and $\lambda_3 = 727$ nm.

TABLE II
CORRELATION COEFFICIENT (R) AND $rmse$ BETWEEN TWFS AND CHLA CONCENTRATIONS, AND THE CORRESPONDING WAVELENGTHS AFTER EVERY ITERATION OF OPTIMIZATION, ~: WAVELENGTH OPTIMIZED (IN NANOMETERS)

Optimization times	λ_1	λ_2	λ_3	R		RMSE (mg m^{-3})	
				Maximum	Corresponding wavelength	Minimum	Corresponding wavelength
1	661	~	750	0.8826	691	3.73	691
2	661	691	~	0.8872	727	3.66	727
3	~	691	727	0.8872	661	3.66	661
4	661	~	727	0.8872	691	3.66	691

models were used to retrieve Chla concentrations separately (Fig. 6) using (5) and (6). Their R^2 coefficients are 0.8583 and 0.8550, respectively, while the $rmse$'s of Chla estimation are 3.12 and 3.16 mg/m^3. The first model has a slightly better performance than the second model, which is consistent with the analysis in Section III-B

$$\text{Chla} = 212.1\left[R_{\text{rs}}^{-1}(661) - R_{\text{rs}}^{-1}(691)\right]R_{\text{rs}}(727) + 14.2 \quad (R^2 = 0.8583, N = 32, rmse = 3.12, P \le 0.001) \tag{5}$$

$$\text{Chla} = 212.4\left[R_{\text{rs}}^{-1}(664) - R_{\text{rs}}^{-1}(690)\right]R_{\text{rs}}(798) + 11.0 \quad (R^2 = 0.8550, N = 32, rmse = 3.16, P \le 0.001). \tag{6}$$

TWFs are all related closely to Chla concentrations ($R \ge 0.8812, rmse \le 3.75$ mg/m^3) (Table III). $[R_{\text{rs}}^{-1}(661) - R_{\text{rs}}^{-1}(691)]R_{\text{rs}}(727)$ and $[R_{\text{rs}}^{-1}(661) - R_{\text{rs}}^{-1}(691)]R_{\text{rs}}(750)$ were both obtained in the course of iterative optimizations with initial wavelengths of $\lambda_2 = 700$ nm and $\lambda_3 = 750$ nm and $[R_{\text{rs}}^{-1}(664) - R_{\text{rs}}^{-1}(690)]R_{\text{rs}}(798)$. Therefore, these TWFs were fed to the SVM model, and Chla concentration was treated as the only output. In this way, Chla estimation was changed into a regression problem of SVM with three inputs and one output. It is expressed as

$$X = \left(\left[R_{\text{rs}}^{-1}(661) - R_{\text{rs}}^{-1}(691)\right]R_{\text{rs}}(727), \left[R_{\text{rs}}^{-1}(661) - R_{\text{rs}}^{-1}(691)\right]R_{\text{rs}}(750), \left[R_{\text{rs}}^{-1}(664) - R_{\text{rs}}^{-1}(690)\right]R_{\text{rs}}(798)\right)$$
$$Y = C_{\text{Chla}}. \tag{7}$$

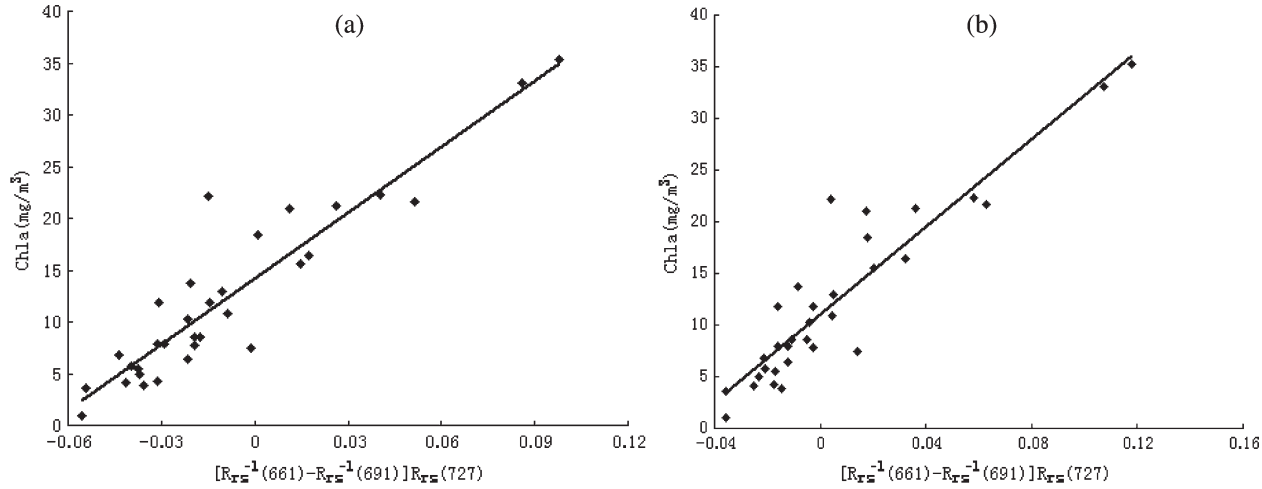


Fig. 6. Linear regression analysis between TWFs and Chla concentrations. (a) $[R_{\rm rs}^{-1}(661) - R_{\rm rs}^{-1}(691)]R_{\rm rs}(727)$. (b) $[R_{\rm rs}^{-1}(664) - R_{\rm rs}^{-1}(690)]R_{\rm rs}(798)$.

TABLE III
CORRELATION COEFFICIENT (R) AND $rmse$ BETWEEN TWFS AND CHLA CONCENTRATIONS

TWFs	R	RMSE (mg m^{-3})
$[R_{rs}^{-1}(661)-R_{rs}^{-1}(691)]R_{rs}(727)$	0.8872	3.66
$[R_{rs}^{-1}(661)-R_{rs}^{-1}(691)]R_{rs}(750)$	0.8826	3.73
$[R_{rs}^{-1}(664)-R_{rs}^{-1}(690)]R_{rs}(798)$	0.8812	3.75

The performance of the SVM model depends largely on the selection of the model parameters. There are two parameters in the RBF kernel: penalty of estimation error C and kernel σ. The goal is to find the best value for the (C, σ) pair so that the SVM model can be the most accurate. For this purpose, the data set was trained repeatedly with different combinations of (C, σ) values. Just as with the evaluation of the linear models, $R^2$and $rmse$ were used to judge the performance. The (C, σ) pair, which yielded the highest accuracy, was ultimately set to (10 000, 0.15), and the corresponding R^2 and $rmse$ were, respectively, 0.8961 and 2.67 mg/m^3. This setting led to a better performance than the linear models.

D. Model Validation

Previous scholars have proposed several Chla estimation models based on TWFs, including $[R_{\rm rs}^{-1}(671) - R_{\rm rs}^{-1}(710)]R_{\rm rs}(740)$ [29], $[R_{\rm rs}^{-1}(675) - R_{\rm rs}^{-1}(695)]R_{\rm rs}(730)$ [6], $[R_{\rm rs}^{-1}(666) - R_{\rm rs}^{-1}(688)]R_{\rm rs}(725)$ [12], and $[R_{\rm rs}^{-1}(690) - R_{\rm rs}^{-1}(693)]R_{\rm rs}(799)$ [11]. Parameter values in these empirical models were derived from different turbid waters. Therefore, they had to be determined from the training data set. Plugging of these new values into the proposed TWFs resulted in several new models such as the "Dall'Olmo's" model, Gitelson's model, Zhou's model, and Xu's model. The predictive errors of these models were analyzed using the validation data set (Table IV). The $rmse$'s of the TWF models proposed by others range from 4.70 to 5.26 mg/m^3, while the $MAPE$'s are within 45.59% and 52.01%. In general, these models have larger predictive errors than the linear models and the SVM model developed in this paper. Thus, these models are inapplicable to Lake Taihu waters.

TABLE IV
COMPARISON OF THE PREDICTIVE ERROR FROM DIFFERENT MODELS

Estimation models	MAPE (%)	RMSE (mg m^{-3})
Dall'Olmo's model (2005)	51.92	4.77
Gitelson's model (2007)	45.59	4.96
Zhou's model (2008)	46.16	4.70
Xu's model (2008)	52.01	5.26
Linear model 1	49.55	4.42
Linear model 2	44.95	4.58
SVM model	22.88	2.56

The $MAPE$ of the SVM model was 22.88%, which was much lower than that of the two linear models in (5) and (6) (Table IV). However, the two linear models have $rmse$'s of 4.42 and 4.58 mg/m^3, respectively, or approximately 1.7 and 1.8 times than that of the SVM model (Table IV). According to the scatterplots of observed and modeled Chla concentrations for the two linear models and the SVM model (Figs. 7 and 8), the points of the SVM model converge toward the 1 : 1 line, but that of the linear models scattered. Thus, it is concluded that the SVM model suits Lake Taihu waters best.

E. Model Sensitivity and Applicability Analysis

The model sensitivity reflects the dependence degree of the model to samples, which should be analyzed for evaluating the model performance. The developed SVM model in this paper needs three TWFs as input parameters. Thus, we should separately analyze the sensitivity of the SVM model to each TWF input. Here, we used the following function to quantify the sensitive degree of the SVM model to TWFs, which has been widely used in relative studies [33], [34]:

$$\Delta Chla = |SVM(x + \Delta x) - SVM(x)| \tag{8}$$

where $\Delta Chla$ represents the sensitive degree of the SVM model to input parameters (in milligrams per cubic meter), SVM represents the developed SVM model in this paper, x represents the input parameter, and Δx represents the uncertainty of the parameter x. Note that x here includes three input parameters, that is, TWF1, TWF2, and TWF3. When computing the sensitive degree of the SVM model to one TWF, the other two TWFs should stay the same. As to the Δx, we

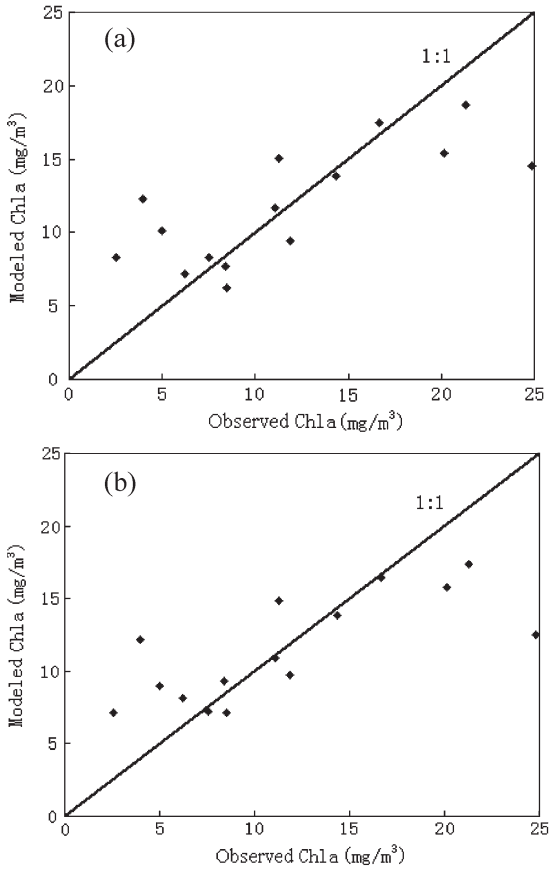


Fig. 7. Scatterplots of observed and modeled Chla concentrations using two linear models. (a) Linear model 1 and (b) linear model 2.

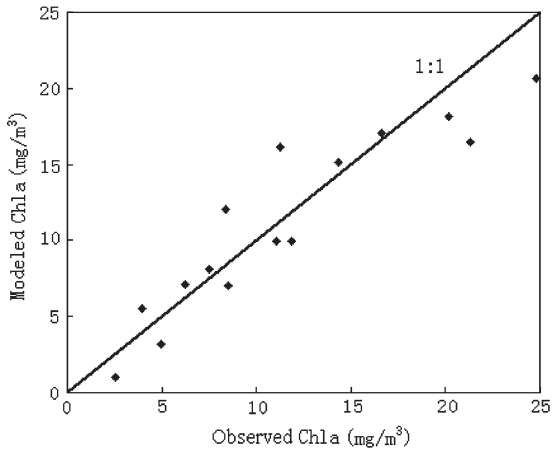


Fig. 8. Scatterplot of observed and modeled Chla concentrations estimated using the SVM model.

set six values for each TWF according to the variation range of TWFs (Table V).

The sensitive degree results have been presented in Table V. When $\Delta \times 1$ reaches 0.0766 (half of the length of span of TWF1 variation), $\Delta Chla$ is 41.90 mg $\cdot$ m^{-3}, which shows the lowest sensitivity to the SVM model compared to TWF2 and TWF3 (86.93 and 238.78 mg $\cdot$ m^{-3}, respectively). When Δx's are the smallest values we set (the length of span of TWF variations divided by 100), the $\Delta Chla$'s for TWF1, TWF2, and TWF3 are 1.80, 0.95, and 1.45 mg $\cdot$ m^{-3}, respectively. Note that the $\Delta Chla$'s for TWF2 all display the smallest errors (except the last) compared to TWF1 and TWF3.

The SVM model featured in this paper is based on the *in situ* data set obtained in November 2007. It is preferable to verify its performance in other seasons when inherent optical properties of water components may vary from those in November 2007. Previous experiments on Lake Taihu waters have accumulated some *in situ* data sets obtained in July and August 2005, October 2006, and March 2007. The ranges of Chla concentrations varied among the data sets. The largest variation from 7.12 to 155.05 mg/m^3 in Chla concentration took place in summer, while the slightest variation from 2.03 to 7.27 mg/m^3 occurred in spring (Table VI). The $MAPE$ ranged from 31.37% in summer to 54.48% in spring. It is worthy to note that the $rmse$ in summer was 11.20 mg/m^3, much higher than in spring and autumn. Obviously, summer Chla averaging (38.88 ± 37.50)mg/m^3 was higher than in other seasons, owing to increased phytoplankton quantity. However, the Chla used in establishing the SVM model was low and ranged from 0.98 to 35.29 mg/m^3 with an average of (12.06 ± 7.84)mg/m^3. Thus, the relatively large error in summer is attributed mainly to the large variations in water optical properties in two different environments, namely, water bodies of model development and summer test. Comparatively speaking, the $rmse$'s of the SVM model in spring and autumn are relatively low, which are 3.20 and 5.41 mg/m^3, respectively.

IV. Discussion and Conclusion

The determination of TWFs is based on the relationship between reflectance $R_{\rm rs}(\lambda)$ and two inherent optical properties (total absorption coefficient a and backscattering coefficient b_b) [35]. The combination of three wavelengths λ_1, λ_2, and λ_3 can minimize the effects of suspended particles and CDOM and increase the sensitivity of water reflectance to Chla in waters. Therefore, the TWF models have a substantial theoretical backing, which will considerably promote the use of universal Chla estimation models in turbid productive waters. Moreover, the models do not require data on inherent optical properties of in-water constituents, which can be variable and difficult to obtain. In addition, the low cost of hand-held radiometers and their ease of use offer widespread applicability in monitoring the diverse conditions of turbid productive waters [6], [10].

In this paper, the SVM method took the following points into consideration: 1) Simple regression analysis between TWF and Chla concentrations can yield a linear model for Chla estimation, but its parameters are usually variable among different experiments, even for the same water body at different times; 2) although a TWF with the highest accuracy can be found through iterative optimization, some other TWFs also achieved high accuracies. This finding is consistent with the conclusion of "spectral ranges" proposed for three wavelengths [10], which indicates that errors of Chla estimation are minimal over specific spectral ranges. Thus, several TWFs were selected as the inputs to the SVM model to achieve a high level of performance; and 3) SVM is advantageous in that it is capable of learning from known examples and applying that knowledge to unknown data with higher accuracy on the training data.

TABLE V
SENSITIVE DEGREE OF THE SVM MODEL TO INPUT PARAMETERS, TWF1, TWF2, AND TWF3. THE $\Delta \times 1$, $\Delta \times 2$, AND $\Delta \times 3$ ARE THE UNCERTAINTY OF THE INPUT PARAMETERS CORRESPONDING TO THREE TWFS, RESPECTIVELY. IN THIS PAPER, SIX Δx's HAVE BEEN SET FOR EACH TWF, AND THEIR VALUES WERE OBTAINED BY DIVIDING THE VARIATION RANGE OF TWF BY 2, 5, 10, 20, 50, AND 100. THE VALUES IN BRACKETS ARE THE LENGTHS OF SPAN OF TWF VARIATIONS

TWF1	-0.0554 ~ 0.0978 (0.1532)					
Δx1	0.0015	0.0031	0.0077	0.0153	0.0306	0.0766
ΔChla (mg m^{-3})	1.80	3.61	9.06	18.02	34.19	41.90
TWF2	-0.0412 ~ 0.0726 (0.1138)					
Δx2	0.0011	0.0023	0.0060	0.0114	0.0228	0.0569
ΔChla (mg m^{-3})	0.95	1.93	5.12	11.19	25.97	86.93
TWF3	-0.0357 ~ 0.1179 (0.1536)					
Δx3	0.0015	0.0031	0.0077	0.0154	0.0307	0.0768
ΔChla (mg m^{-3})	1.45	3.11	9.37	23.74	65.00	238.78

TABLE VI
APPLICABILITY ANALYSIS OF THE SVM MODEL IN DIFFERENT SEASONS (N = the number of samples)

Sampling time (season)	N	Chla concentration (mg m^{-3})	MAPE (%)	RMSE (mg m^{-3})
July, August 2005 (Summer)	24	7.12~155.05 (38.88±37.50)	31.37	11.20
October 2006 (Autumn)	17	4.69~19.69 (12.50±5.30)	36.03	5.41
March 2007 (Spring)	10	2.03 ~ 7.27 (4.83±1.80)	54.48	3.20

However, one significant problem must be resolved satisfactorily before the universal model for Chla estimation can be developed, i.e., the determination of typical TWFs for turbid productive waters. Currently, the best TWF varies with different turbid waters, even with different samples from the same water body. This justifies the selection of several TWFs with the highest accuracy as the model inputs in this paper. However, for other turbid productive waters, the applicability of the TWFs proposed in this paper should be verified through more analysis.

In summary, the SVM model with three TWFs established in this paper performs better than the previous linear model proposed for Lake Taihu waters, with R^2 of 0.8961 and $rmse$ of 2.67 mg/m^3. Validation of its applicability in different seasons indicates that the SVM model can become a universal model for Chla estimation in Lake Taihu waters, and no significant difference in performance for measurements made in spring, summer, and autumn. Nevertheless, given that the performance of the SVM model varies with seasonality to some extent, more research is needed to further improve its performance by refining the wavelengths proposed in this paper in the future when more *in situ* samples become available.

ACKNOWLEDGMENT

The authors would like to thank C. Le and C. Huang for their help and useful suggestions and also L. Wang, X. Wang, K. Shi, R. Xia, Y. Yang, X. Jin, and Y. Wang for their participation in the field experiment. The authors would also like to thank J. Gao for his valuable comments on this paper and to two anonymous reviewers for their useful comments and suggestions.

REFERENCES

[1] A. Morel and L. Prieur, "Analysis of variations in ocean color," *Limnol. Oceanogr.*, vol. 22, no. 4, pp. 709–722, Jul. 1977.

[2] H. R. Gordon and A. Morel, *Remote Assessment of Ocean Color for Interpretation of Satellite Visible Imagery, A Review*, vol. 4. New York: Springer-Verlag, 1983.

[3] A. A. Gitelson, "The peak near 700 nm on reflectance spectra of algae and water: Relationships of its magnitude and position with chlorophyll concentration," *Int. J. Remote Sens.*, vol. 13, no. 17, pp. 3367–3373, 1992.

[4] H. J. Gons, "Optical teledetection of chlorophyll a in turbid inland waters," *Environ. Sci. Technol.*, vol. 33, no. 7, pp. 1127–1132, Apr. 1999.

[5] G. Dall'Olmo, A. A. Gitelson, and D. C. Rundquist, "Towards a unified approach for remote estimation of chlorophyll-a in both terrestrial vegetation and turbid productive waters," *Geophys. Res. Lett.*, vol. 30, no. 18, p. 1938, 2003. DOI:10.1029/2003GL018065.

[6] A. A. Gitelson, J. F. Schalles, and C. M. Hladik, "Remote chlorophyll-a retrieval in turbid, productive estuaries: Cheapeake Bay case study," *Remote Sens. Environ.*, vol. 109, no. 4, pp. 464–472, Aug. 2007.

[7] D. D'Alimonte, G. Zibordi, and J.-F. Berthon, "A statistical index of bio-optical seawater types," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 8, pp. 2644–2651, Aug. 2007.

[8] D. Y. Sun, Y. M. Li, Q. Wang, C. F. Le, C. C. Huang, and L. Z. Wang, "Parameterization of water component absorption in inland entrophic lake and its seasonal variability, a case study in Lake Taihu," *Int. J. Remote Sens.* to be published.

[9] P. V. Zimba and A. A. Gitelson, "Remote estimation of chlorophyll concentration in hyper-eutrophic aquatic systems: Model tuning and accuracy optimization," *Aquaculture*, vol. 256, no. 1–4, pp. 272–286, 2006.

[10] A. A. Gitelson, G. Dall'Olmo, W. Moses, D. C. Rundquist, T. Barrow, T. R. Fisher, D. Gurlin, and J. Holz, "A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation," *Remote Sens. Environ.*, vol. 112, no. 9, pp. 3582–3593, Sep. 2008.

[11] J. P. Xu, B. Zhang, K. S. Song, Z. M. Wang, D. W. Liu, and H. T. Duan, "Estimation of chlorophyll-a concentration in Lake Xinmiao based on a semi-analytical model," *J. Infrared Millim. Waves*, vol. 27, no. 3, pp. 197–201, 2008.

[12] G. H. Zhou, Q. H. Liu, R. H. Ma, and G. L. Tang, "Inversion of chlorophyll-a concentration in turbid water of Lake Taihu based on optimized multi-spectral combination," *J. Lake Sci.*, vol. 20, no. 2, pp. 153–159, 2008.

[13] B. Q. Qin, W. P. Hu, and W. M. Chen, *The Process and Mechanism of Water Environment Evolvement in Taihu Lake*. Beijing, China: Sci. Press, 2004, pp. 1–2.

[14] J. L. Mueller, A. Morel, R. Frouin, C. Davis, R. Arnone, K. Carder, Z. P. Lee, R. G. Steward, S. Hooker, C. D. Mobley, S. McLean, B. Holben, M. Miller, C. Pietras, K. D. Knobelspiesse, G. S. Fargion, J. Porter, and K. Voss, *Ocean Optics Protocols for Satellite Ocean Color Sensor Validation*, vol. III. Greenbelt, MD: NASA Goddard Space Flight Center, 2003. Revision 4.

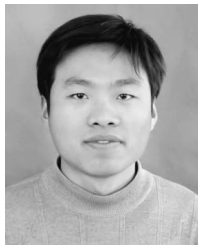
[15] H. B. Jiao, Y. Zha, J. Gao, Y. M. Li, Y. C. Wei, and J. Z. Huang, "Estimation of chlorophyll-a concentration in Lake Tai, China using *in situ* hyperspectral data," *Int. J. Remote Sens.*, vol. 27, no. 19, pp. 4267–4276, Oct. 2006.

[16] B. G. Mitchell, "Algorithms for determining the absorption coefficient of aquatic particulates using the quantitative filter technique (QFT)," *Proc. SPIE*, vol. 1302, pp. 137–148, 1990.

[17] X. Huang, *Eco-Investigation, Observation and Analysis of Lakes*. Beijing, China: Standard Press China, 1999, pp. 77–99.

[18] R. H. Ma, J. W. Tang, and J. F. Dai, "Bio-optical model with optimal parameter suitable for Lake Taihu in water color remote sensing," *Int. J. Remote Sens.*, vol. 27, pp. 4303–4326, 2006.

[19] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
[20] M. M. Chi and L. Bruzzone, "Semisupervised classification of hyperspectral images by SVMs optimized in the primal," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1870–1880, Jun. 2007.
[21] J. Muñoz-Marí, L. Bruzzone, and G. Camps-Valls, "A support vector domain description approach to supervised classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 8, pp. 2683–2692, Aug. 2007.
[22] H. G. Zhan, P. Shi, and C. Q. Chen, "Retrieval of oceanic chlorophyll concentration using support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2947–2951, Dec. 2003.
[23] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.
[24] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," Dept. Comput. Sci. Inf. Eng., Nat. Taiwan Univ., Taipei, Taiwan, 2003. Tech. Rep.
[25] B. Schölkopf, P. Bartlett, A. Smola, and R. Williamson, "Support vector regression with automatic accuracy control," in *Proc. Int. Conf. Artif. Neural Netw., Perspectives Neural Comput.*, L. Niklasson *et al.*, Eds., 1998, pp. 111–116.
[26] C. C. Chang and C. J. Lin, "Training nu-support vector regression: Theory and algorithms," *Neural Comput.*, vol. 14, no. 8, pp. 1959–1977, Aug. 2002.
[27] G. Dall'Olmo, A. A. Gitelson, D. C. Rundquist, B. Leavitt, T. Barrow, and J. C. Holz, "Assessing the potential of SeaWiFS and MODIS for estimating chlorophyll concentration in turbid productive waters using red and near-infred bands," *Remote Sens. Environ.*, vol. 96, no. 2, pp. 176–187, May 2005.
[28] Z. P. Lee, K. L. Carder, S. K. Hawes, R. G. Steward, T. G. Peacock, and C. O. Davis, "Model for the interpretation of hyperspectral remote-sensing reflectance," *Appl. Opt.*, vol. 33, no. 24, pp. 5721–5732, Aug. 1994.
[29] G. Dall'Olmo and A. A. Gitelson, "Effect of bio-optical parameter variability on the remote estimation of chlorophyll a concentration in turbid productive waters: Experimental results," *Appl. Opt.*, vol. 44, no. 3, pp. 412–422, Jan. 2005.
[30] J. F. Schalles, "Optical remote sensing techniques to estimate phytoplankton chlorophyll a concentrations in coastal waters with varying suspended matter and CDOM concentrations," in *Remote Sensing of Aquatic Coastal Ecosystem Processes: Science and Management Applications*, L. Richardson and E. Ledrew, Eds. Dordrecht, The Netherlands: Springer-Verlag, 2006, pp. 27–79.
[31] G. Oron and A. Gitelson, "Real-time quality monitoring by remote sensing of contaminated water-bodies: Waste stabilization pond effluent," *Water Res.*, vol. 30, no. 12, pp. 3106–3114, Dec. 1996.
[32] A. Bricaud, M. Babin, A. Morel, and H. Claustre, "Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: Analysis and parameterization," *J. Geophys. Res.*, vol. 100, no. C7, pp. 13 321–13 332, Jul. 1995.
[33] J. C. Jiménez-Muñoz and J. A. Sobrino, "A generalized single-channel method for retrieving land surface temperature from remote sensing data," *J. Geophys. Res.*, vol. 108, no. D22, pp. 4688–4695, 2003.
[34] L. Zhu, X. F. Gu, L. F. Chen, T. Yu, and Z. F. Wang, "Comparison of LST retrieval precision between single-channel and split-windows for high-resolution infrared camera," *J. Infrared Millim. Waves*, vol. 27, no. 5, pp. 346–353, 2008.
[35] H. R. Gordon, O. B. Brown, R. H. Evans, J. W. Brown, R. C. Smith, K. S. Baker, and D. K. Clark, "A semianalytic radiance model of ocean color," *J. Geophys. Res.*, vol. 93, no. D9, pp. 10 909–10 924, 1988.



Deyong Sun received the B.S. degree in geographical information systems from the Nanjing University of Technology, Nanjing, China, in 2005. He has been working toward the M.S. and Ph.D. degrees in the field of environment remote sensing at the Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing, since 2005.

He has published ten scientific papers, including journals and conference proceedings. His research interests include water color remote sensing and optical properties in optically complex water bodies.

Yunmei Li received the B.S. degree in mathematics from Yunnan University, Kunming, China, in 1987, and the Ph.D. degree in remote sensing technology and application from Zhejiang University, Hangzhou, China, in 2001.

Since 2001, she has been with the Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing, China, where she was a Postdoctoral Researcher until 2003, an Associate Professor from 2003 to 2007, and has been a Full Professor of remote sensing technology and application since 2007, teaching hyperspectral remote sensing, geographical information systems, and water color remote sensing. She is also currently the Head of the Water Color Remote Sensing Laboratory, Department of Geographical Science, Nanjing Normal University. She has been appointed as an Evaluator of project proposals. She has authored or coauthored more than 60 scientific publications, including journals, book chapters, and conference proceedings. Her current research interests are in the area of water color remote sensing (water component concentration retrieval, inherent optical properties of water bodies, and bio-optical modeling for optically complex waters), particularly for inland turbid lakes. She conducts and supervises research on these topics within the frameworks of several national projects.

Qiao Wang received the B.S. degree in mathematics from Nankai University, Tianjin, China, in 1982, and the M.S. degree in map science and the Ph.D. degree in geographical information systems from Wuhan University, Wuhan, China, in 1992 and 1996, respectively.

From 1992 to 1996, he was an Associate Professor with Wuhan University. From 1996 to 1998, he was a Postdoctoral Researcher with the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China. Since 1998, he has been a Researcher with the National Environment Protection Headquarters. Since 1999, he has been appointed as a Specific Professor with the Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing, China. He has conducted more than 20 projects about research on environment remote sensing, geographical information systems, etc., within national project frameworks. He has also been appointed as an Evaluator of project proposals. He has authored or coauthored more than 80 scientific publications, including journals, books, and conference proceedings.