Journal of
Applied Remote Sensing

Water quality monitoring using Landsat Themate Mapper data with empirical algorithms in Chagan Lake, China

Kaishan Song
Zongming Wang
John Blackwell
Bai Zhang
Fang Li
Yuanzhi Zhang
Guangjia Jiang

Water quality monitoring using Landsat Themate Mapper data with empirical algorithms in Chagan Lake, China

Kaishan Song,[a] Zongming Wang,[a] John Blackwell,[b] Bai Zhang,[a] Fang Li,[a] Yuanzhi Zhang,[c] and Guangjia Jiang[a]

[a] Chinese Academy of Sciences, Northeast Institute of Geography and Agricultural Ecology, Changchun 130012, China

[b] Charles Sturt University, International Centre of Water for Food Security, Wagga Wagga NSW 2678, Australia

[c] The Chinese University of Hong Kong, Institute of Space and Earth Information Science, Esther Lee Building, Shatin, Hong Kong

songks@neigae.ac.cn or kaissong@iupui.edu

Abstract. Lake Chagan represents a complex situation of major optical constituents and emergent spectral signals for remote sensing analysis of water quality in the Songnen Plain. As such it provides a good test of the combined radiometric correction methods developed for optical remote sensing data to monitor water quality. Landsat thematic mapper (TM) data and *in situ* water samples collected concurrently with satellite overpass were used for the analysis, in which four important water quality parameters are considered: chlorophyll-a, turbidity, total dissolved organic matter, and total phosphorus in surface water. Both empirical regressions and neural networks were established to analyze the relationship between the concentrations of these four water parameters and the satellite radiance signals. It is found that the neural network model performed at better accuracy than empirical regressions with TM visible and near-infrared bands as spectral variables. The relative root mean square error (RMSE*)* for the neural network was $< 10\%$, while the RMSE for the regressions was less than 25% in general. Future work is needed on establishing the dynamic characteristic of Chagan Lake water quality with TM or other optical remote sensing data. The algorithms developed in this study need to be further tested and refined with multidate imagery data [DOI: 10.1117/1.3559497]

Keywords: water quality; remote sensing; Chagan Lake; BP neural network.

Paper 09136RR received Oct. 20, 2009; revised manuscript received Dec. 19, 2010; accepted for publication Dec. 29, 2010; published online Mar. 14, 2011.

1 Introduction

Lakes are valuable water resources for fishing, transport, agriculture, industry, recreation, and tourism. The water quality of a number of lakes around the world has deteriorated to such an extent that natural recovery is virtually impossible. The most common ecological problem of inland water bodies is anthropogenic eutrophication mostly as a result of economic development and indiscriminate discharge of untreated or poorly treated sewage. It has become the most widespread water quality problem in China and many other countries as well.[1,2] Water quality deterioration and eutrophication have received more and more attention from the public and government, and many studies have been carried out on lakes around the world to assess their water trophic state.[3,4]

Digital evaluation of optical satellite information at visible and near-infrared wavelengths has been used to estimate different parameters in surface waters.[5,6] These investigations show that

1931-3195/2011/$25.00 © 2011 SPIE

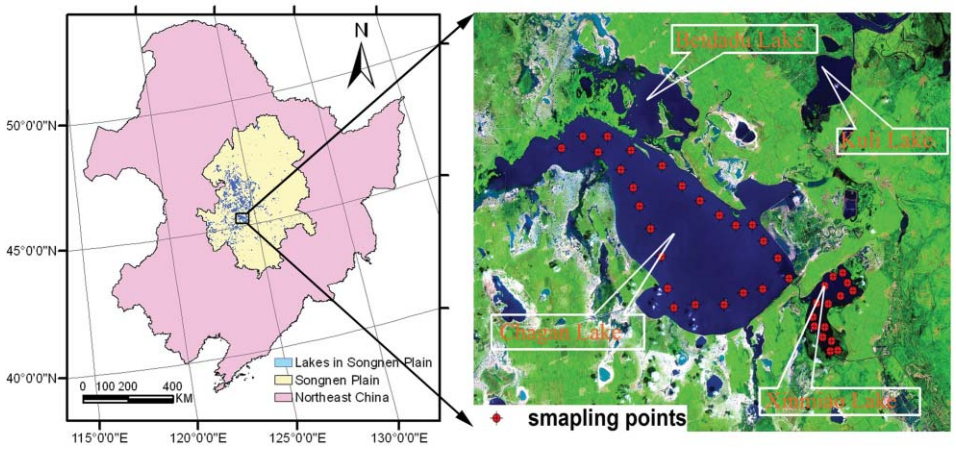

Fig. 1 Location map of the study area and sampling points in Chagan Lake.

Landsat thematic mapper (TM) data can provide relatively low-cost, simultaneous information on surface water conditions from numerous lakes and coastal areas situated within a large geographic area.[6,7] A bulk of studies indicated that Landsat-TM band 2 (525 to 605 nm) can be used to measure phytoplankton concentrations; band 3 (630 to 690 nm) can provide valuable information when large sediment loads occur and pigment as well; and band 4 (750 to 900 nm) is the spectral region where strong water absorption occurs, but the band reflectance is dominated by suspended matter.[2,6] Investigations also demonstrated that band ratios or band combinations are more valuable for water quality monitoring because atmospheric effects and illumination variation can be reduced somehow in band ratios process.[1–3]

There are thousands of lakes in the Songnen Plain, 6397 of them having an area greater than 60 ha totaling 432,600 ha (Fig. 1). All these lakes were formed in a similar geological and climatic environment.[8] As these lakes developed in saline environment, their waters tend to be alkaline (Table 1) and the surrounding land saline and alkaline (see white areas in Fig. 1). Average annual rainfall is about 430 mm, while the evaporation is about 1300 mm/yr. Evaporation is often greater than total water input for crop production and this can lead to secondary salinization of cropland which can then raise river and lake water salinity.[8] Chagan Lake is the largest lake in the Songnen Plain and can be regarded as representative of most lake waters featured in this semi-arid region.

Chagan Lake is optically dominated by scattering from suspended matter, chlorophyll, and yellow substance derived from dissolved and particulate organic matter. As the Chagan Lake is very shallow, the water can be very turbid even in calm weather conditions; the water quality is also highly affected by the input from the Songhua River. Traditional methods have been used to analyze the water quality over many years,[9] but optical properties of the water in the area have never been studied using remotely sensed data from both satellites and *in situ* collected spectral data. There is an urgent need to study Chagan Lake water quality using remotely sensed data for synoptic monitoring of the water quality parameters distributed characteristics. The approach developed will be applicable to other lakes in the Songnen Plain with a similar limnological

Table 1 General information about the three lakes in the study area (on average).

Lake name	Area (ha)	Perimeters (m)	Depth (m)	Volume (m^3)	Ph	Salinity (%)
Chagan	37,240	28,4745	1.52	5.42×10^8	9.18	0.66
Xinmiao	3,072	53,124	2.7	8.3×10^7	8.27	0.23
Kuli	2,877	43,016	2.3	6.6×10^7	8.15	0.21

environment, with only a slight modification. Furthermore, previous studies generally focused on water quality parameters, such as phytoplankton, total suspended matter, turbidity (Turb) and water clarity; few studies tried to estimate total phosphorus (TP) and total dissolved organic matter (TOC).[2] This study also tries to test the feasibility of estimating total phosphorus (TP) and total dissolved organic matter (TOC) with remotely sensed imagery data.

2 Study Area Description

Lake Chagan, located in the southwest of Songnen Plain, northeastern China (Fig. 1), is the 10th largest lake in China. It has a surface area of 372 km^2, with a mean depth of 1.52 m, and full storage capacity of 5.98×10^8 m^3. Xinmiao Lake is connected with Chagan Lake via a channel. Songhua River flows into Xinmiao Lake which is much clearer than Chagan Lake for constant flushing and more aquatic vegetation population. Kuli Lake is separate, but it used to be connected to Chagan Lake. There are no samples taken over the Kuli Lake due to a water vehicle not being available for the sampling survey. The descriptive information about these three lakes is listed in Table 1. Chagan Lake is also a eutrophic lake with high Chorophyll-a (Chl-a) concentration and low secchi disk depth (SDD).[9] The Songhua, Huolin, Tao'er, and Nenjiang Rivers all provide inputs to Chagan Lake. Natural precipitation and ground water are ancillary water suppliers for the lake. The lake's primary economic value is linked to fishery, but it is also important for agriculture and recreational purpose, and therefore a long-term monitoring program of the lake's conservation is under consideration. Xinmiao, Kuli, and Beidadu Lakes have clearer water than Chagan due to an abundance of aquatic vegetation, whereas Chagan Lake is quite turbid both in its shallowness and bottom re-suspension.

3 Material and Methodology

One scene Landsat 5-TM imagery was acquired (with path and row 119 and 29, respectively) and simultaneously, *in situ* spectral measurement, water sampling, and laboratory analysis were performed. Image process methods include radiometric calibration and georeference correction of TM image data. Multiple regression methods and back-propagation (BP) neural networks were used for the construction of empirical models.

3.1 *Data Collection*

3.1.1 *Water parameters testing*

Water samples at 40 sites were taken from the surface layer (0 to 0.5 m) over the lakes (see Fig. 1 for sample sites on the lakes), 25 from Chagan Lake, and 15 from Xinmiao Lake. The sampling sites were decided by visually assessing the water color differences with Landsat image data acquired previously (August 24, 2002). All the water samples were collected in 1000 ml acid-cleaned amber HDPE bottles and stored in a cooler with ice packs. 200 ml of water samples were filtered through 47 mm 0.45 μm pore size acetate filters for Chl-a testing. Prior to analysis, filters were dissolved in 10 ml of 90% buffered acetone and extracted in a dark freezer between 24 and 48 h, and finally Chl-a concentration was determined spectrophotometrically. For TOC determination, all samples were filtered through 0.7 μm pore size glass fiber filters, and then determined through Vario TOC (Elementar Anyalysysteme GmbH, Hanau, Germany). TP was analyzed using a molybdenum blue method after the samples were digested with potassium peroxydisulfate. Turbidity (NTU) was determined through Hydrolab multiparameter probes (Sound4a) positioned 25 cm below the water surface. Other parameters such as temperature (°C), specific conductance (mS), total dissolved solids (g/L), and *p*H were also collected with Sound4a. All the samples were analyzed in the laboratory within 48 h of water sampling. The statistical information about these data is presented in Table 2. It can be seen that water quality

Table 2 Statistical information on water samples in Chagan and Xinmiao Lakes.

	Chagan Lake				Xinmiao Lake				
	Min	**Max**	**Mean**	**Std.**	**Min**	**Max**	**Mean**	**Std.**	**Units**
Chl-a	13.54	30.68	16.92	8.08	6.11	15.74	8.28	3.56	μg/L
Turb	82.27	174.28	79.1850	56.84	11.47	58.34	34.98	17.52	NTUs
TOC	6.49	26.14	14.7949	7.431	6.49	16.67	9.26	3.22	mg/L
TP	0.060	0.190	0.092	0.052	0.03	0.082	0.053	0.027	mg/L

parameters are quite different between Chagan and Xinmiao Lakes, indicating a reasonable range for regression and the neural network model development with these *in situ* samples.

3.1.2 *Spectra data collection*

Simultaneously, field spectra were measured with a portable Fieldspec FSR VNIR spectrometer (Analytical Devices, Inc., Boulder, Colorado). The ASD radiometer has a spectral range between 350 and 1100 nm with radiometric resolution of about 3 nm, which is ideal for water quality applications. The instrument operates by comparing radiance from the surface of the lake to radiance from a near-Lambertian reference panel. The measurements were taken 1 m vertically above the lakes surface from a boat to best simulate a satellite reading. Each measured position was oriented to the boat side to minimize sun-glint and direct reflection from waves, but far away from the effect of the boats shadow.

An optically active component (OAC) difference in the Chagan and Xinmiao Lakes causes large variation in remote sensing reflectance, and the measured spectra over Chagan Lake are much higher than that collected over Xinmiao Lake (Fig. 2). The resultant spectrum reveals a low reflectance value at the wavelength range between 400 to 440 nm, which is due to absorption by both algal pigments (e.g., Chl-a) and dissolved organic matter.[10,11] Because of the influence from suspended matter, the absorption peak of the phytoplankton pigments at 440 nm is not clearly found in our datasets. Between 440 to 560 nm, all values of spectral reflectance show an almost linear increase. At the range of 550 to 560 nm, there is an obvious reflectance peak that results from low absorption by phytoplankton pigments, coupled with increasing backscattering as the suspended particle concentration increases. Chagan Lake exhibits more backscattering

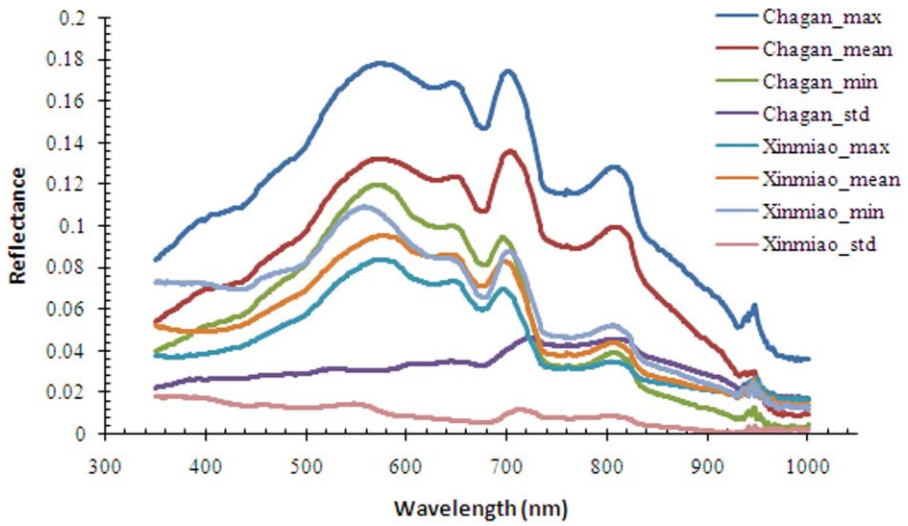

Fig. 2 Water sample reflectance over Chagan (25 spectra) and Xinmiao Lake (15 spectra).

than Xinmiao Lake due to higher chlorophyll and suspended sediment concentration (Table 2). The minimum reflectance occurs near 670 nm, which is attributed to maximum absorption by phytoplankton pigments.[10] The reflectance peak at 700 nm in all cases due to fluorescence by Chl-a pigment,[12] and this peak for Chagan Lake is higher than that for Xinmiao Lake due to the higher Chl-a concentration in Chagan Lake. All reflectance values drop quickly from 700 to 740 nm, and flatten off between 740 to 790 nm. Near 810 nm, the fourth small but visible reflectance peak occurs and then these reflectance values slowly decrease until they are flat again from 840 to 900 nm. All spectra shown in Fig. 2 were also used as a basis for the evaluation of spectral signature characteristics.

3.2 *Landsat TM Data*

When Landsat TM data are used to retrieve quantitative variables concerning the study area, a procedure to correct the measured radiance for the atmospheric condition is required.[13,14] The remaining amount of radiance that reaches the sensor (target radiance) can range from 25% at 450 nm (blue region of spectrum) to nearly 0% at 850 nm (near-infrared spectrum).[15,16] Using a hypothesis of homogeneous ground reflectance, the radiant energy received by a sensor contains several components, and could be expressed as:[6]

$$L_{\text{sat}} = T_r(L_{\text{atm}} + L_{\text{up}}), \tag{1}$$

and

$$L_{\text{up}} = L_{\text{refr}} + L_{\text{refl}} = L_w + L_{\text{refl}}, \tag{2}$$

where L_{sat} is the radiance gathered by the sensor, and T_r is the atmospheric transmittance due to absorption by atmospheric gases such as ozone, water vapor, and carbon dioxide. L_{atm} is the atmospheric radiance due to light scattering from gases and aerosols, and L_{up} is the radiance from the ground target. L_w is the radiance refracted from inside the water body, while L_{refl} is the radiance reflected from the water surface.

In this study, Landsat-5 TM data was acquired on October 14, 2004. The resolution of TM data is 30 m (except thermal band 6 with 120 m). Atmospheric correction was conducted using the 6S (second simulation of satellite signal in the solar spectrum) model. This model corrects at-sensor radiance images for the solar illuminance, radiation, the Rayleigh and aerosol scattering. The input to 6S specifies geometrical, spectral, atmospheric, and target conditions. The visibility was obtained from local weather stations (set to 25 km in this case). Consequently, we used standard models for other atmospheric parameters. These models were chosen as mid-latitude summer and continental aerosol models. The ground truth reflectance made it possible for the code to simulate the radiance at the TM sensor altitude.[16,17] This radiance simulation was used for the atmospheric correction of TM visible and near-infrared bands. The TM image was geo-referenced with ground control points collected with PROMARK-2 DGPS, and the land area in the image was masked off during the water quality modeling process. In order to extract the TM data from the ground truth points (water sampling stations), the mean and standard deviation of TM reflectance values were calculated using defined windows of 120 m $\times$ 120 m (4 $\times$ 4 pixel) for each ground truth point, and the mean value was used to establish the empirical models.

TM bands 1 to 4 are in the spectral range where light passes through the water, providing useful information about water quality. According to Ref. 3, TM1 (450 to 520 nm) can be used to measure the irradiance attenuation due to the absorption of aquatic humus and chlorophyll-a; TM2 (520 to 600 nm) can be used to measure phytoplankton concentrations; TM3 (630 to 690 nm) is more difficult to interpret due to multiprocesses affecting it, but it can provide valuable information when large sediment loads occur; and TM4 (760 to 900 nm) is the product of rapidly increasing water absorption, and suspended matter will greatly affect its irradiance value. TM6 (10,400 to 12,500 nm) can provide accurate information on water temperature.

Water temperature using thermal inversion was out of the scope of this research, so TM bands 1 to 4 were used to inverse water quality in this study.

3.3 *Retrieval Method*

There are two main methods to estimate surface water quality parameters. One is the use of empirical algorithms derived from remotely sensed reflectance data. They can provide site-specific predictions of water quality parameters with reasonable accuracy, but with no universal application.[18,19] The other approach is the use of analytic inversion models that require the solution of radiative transfer equations for deriving absorption and scattering coefficients.[12] The latter approach allows remote sensing measurements to be understood in terms of the inherent optical properties, and provides insight into the characteristics of the differences in algorithm coefficients for various regions.[11,19,20]

3.3.1 *Multivariate approach*

In this study, multivariate retrieval algorithms for water quality variables using TM data can be expressed as

$$D_{\mathrm{TM}} = A_0 + \sum_{i=1}^{k} A_i(\mathrm{TM}_i) \tag{3}$$

where TM_i is the reflectance values of TM bands, k is the TM band number from 1 to 4, and A_0, A_i are the empirical regression coefficients derived using the observations from the ground truth dataset.

3.3.2 *Neural network approach*

For case-II water, the presence of OACs creates an optically complex situation.[16,21] A neural network can meet the task of modeling the transfer function of OACs upwelling radiance. This follows because that (1) a neural network can model a large number of nonlinear behaviors without prior knowledge on the nature of nonlinearity and has the advantage over the standard linear regression with which nonlinearity cannot be modeled properly; (2) it is preferred over a nonlinear regression because the latter requires a priori knowledge of the nature of the non-linear behavior. A neural network is often suitable for situations where non-linearity or chaotic is in datasets to be analyzed, or when the signal is deeply hidden within noise or other signals.[22,23] For this study, a three layer neural network algorithm was used: an input layer, a hidden layer, and an output layer (Fig. 3). The first layer distributes the input parameters of extracted data at different TM bands or band combinations to the second layer. The second layer (hidden layer) has a varying number of neurons, where each input parameter is multiplied by its connection's weights and all the inputs to the neurons are summed and passed through a nonlinear sigmoid function (we choose tansig in this study). The third layer receives the output of the second layer in which it is processed through neurons again.[21]

For the input and hidden layers, the input to the neurons are multiplied by their related relative weights, summed and added to the bias. The weights control which inputs and connections in the network are more important than others. The bias controls the activation level of the neuron, when the resulting sum is passed through a nonlinear activation function:

$$z_j = g(y_j) = \tanh(y_j) = \tanh\left(\sum_{i=1}^{n} w_{ij}x_i + b_j\right), \tag{4}$$

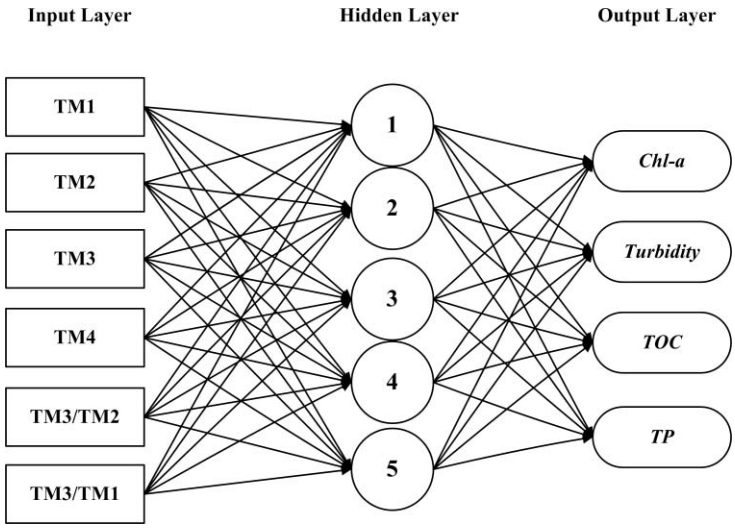

Fig. 3 A neural network diagram for this study.

where g is the sigmoid activation function and z_j is the output of the hidden layer after the nonlinear summation. The activation function is what gives the network its ability to model nonlinear problems, and has a promising application in quantitative remote sensing application.[16,21]

The nodes in the hidden layer depend on the complexity of the approximated function and sample numbers. Although a large number of nodes can more accurately simulate the function, they could have a risk of possible overfitting.[24,25] Based on experimentation, five hidden layer nodes were applied in the model of TM bands 1 to 4 and two band ratios to estimate water qualities. The neural network was trained using Matlab Neural Network Toolbox. The error minimum for each trial was calculated using the Leverberg–Marquardt method.[26] All the input TM bands 1 to 4 and band ratios data were preprocessed, and water quality parameters (Chl-a, Turb, TOC, TP) were scaled to (0, 1) by a factor a, which depends on the valid range of parameters, i.e., the difference between maximum and minimum values of a water quality parameter considered. In addition, the weights related to each connection were randomly generated before training the model. For each number of hidden nodes, the training process was run 2 to 20 times with random initial weights. The weight configuration returning the smallest error was retained for validation. TM image data was output as ASCII from ENVI to the Matlab Neural Network Toolbox for mapping these four water quality parameters after the neural network model had been trained.

4 Results and Discussions

4.1 *Regression Analysis*

In this study, the correlations among the four water quality parameters (Chl-a, Turb, TOC, and TP) were analyzed. The correlation between Chl-a and Turb was –0.49. This result differs slightly from some other studies.[6,16] However, the correlation between Chl-a and TOC was up to 0.86, showing that Chl-a was significantly related to TOC in the study area. Turb had a correlation of 0.53 with TOC, thus Turb is also related to TOC from associated dissolved organic matter in this study area. TP had a low correlation of 0.47 and 0.36 with Chl-a and TOC, respectively, but the correlation of TP and Turb was up to 0.78, TP is generally related to chlorophyll concentration, but with a delayed effect, so data collected at the same time may not have as high a correlation. As for the correlation between TP and Turb, further work needs to be carried out to explain the situation.

Table 3 Correlation (R^2) among Chl-a, Turb, TOC, and TP with TM bands 1 to 4 and band ratios.

	TM1	TM2	TM3	TM4	TM4/1	TM4/3	TM3/1	TM3/2
Chl-a	0.783	0.703	0.616	0.147	0.255	0.433	0.620	0.621
Turb	0.578	0.603	0.689	0.803	0.842	0.752	0.681	0.632
TOC	0.776	0.781	0.735	0.607	0.489	0.230	0.642	0.389
TP	0.622	0.593	0.554	0.514	0.349	0.542	0.381	0.419

The correlation between Chl-a, Turb, TOC, and TP with reflectance of Landsat TM bands is presented in Table 3. It can be seen that Chl-a had a high correlation with TM1, TM2, and TM3 (all in the visible spectral region), but the correlation was quite low with TM4 (near-infrared spectral region). All visible and near-infrared bands were significantly correlated to Turb, and the highest correlation of TM4 at the near-infrared region was 0.803. The same is true for TOC, but with a lower correlation compared with Turb. The highest correlation between TOC and Landsat spectra was in the visible bands. From Table 3, it can be seen that TM bands 1 to 4 all had a significant correlation with TP, as the correlation between TP and Turb was quite high. This possibly explains why TM visible and near-infrared bands all had a high correlation with TP in this study.

Correlation analysis was also undertaken with TM band ratios (Table 3). It can be seen that correlation between Chl-a, TOC, and TP and any band ratio was not higher than that for any single band. The highest correlation for Chl-a was with band ratio TM3/TM2. Band ratio algorithms can sometimes improve Turb inversion accuracy, but not significantly. This result is different from other result,[3,6,16] Generally speaking, band ratio can reduce some of random noises caused by illumination and background,[3,7,27] but not necessarily improve empirical model with imagery data.[28] All the other possible band combinations were also tested in this work, but most of the band combination does not show the superiority to single band. The reason will be further explored in our future study region.

The band ratio algorithm did not always improve water parameter inversion accuracy, and a single TM band often had higher correlation with water quality variables in this study. A stepwise linear regression model was established with TM visible, near-infrared bands, and band ratio wherever there was a high correlation. In this study, a data subgroup of 25 samples (15 from Chagan Lake and 10 from Xinmiao Lake) were used to establish the regression models. The Chl-a concentration, Turb, TOC, and TP were finally formulated as:

$$\text{Chl} - a = 1.13\text{TM1} - 0.43\text{TM3/TM2} - 41.91, \tag{5}$$

$$\text{Turb} = 11.31\text{TM4/TM1} - 2.03\text{TM3} - 16.42, \tag{6}$$

$$\text{TOC} = 0.26\text{TM1} + 0.85\text{TM2} - 0.16\text{TM3} - 20.06, \tag{7}$$

$$\text{TP} = 0.042\text{TM1} + 0.061\text{TM2} - 0.001\text{TM3} - 0.175. \tag{8}$$

The remaining part of the dataset was used to validate these empirical models. Finally, R^2 and root mean square error (RMSE) were calculated with all samples. The results are shown in Table 4, and are discussed in Sec. 4.7.

Table 4 Comparison of regression analysis and neural network inversion accuracy.

Parameters (Units)	Regression				Neural network			
	Calibration		Validation		Calibration		Validation	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Chl-a(ug/L)	0.84	2.87	0.78	3.42	0.98	0.803	0.962	0.822
Turb (NTU)	0.87	17.81	0.88	17.65	0.98	5.567	0.963	6.12
TOC(mg/L)	0.85	3.11	0.86	2.92	0.95	1.45	0.92	1.52
TP(mg/L)	0.74	0.024	0.63	0.031	0.99	0.03	0.94	0.04

4.2 *Empirical Neural Network Model*

Similarly, all the sample data sets ($n = 40$) were divided into two groups, one group of data set ($n = 25$) was used to train the network, and the rest of the data set was used to test the network. The output of the network was compared to the *in situ* measurements of surface water quality parameters, including Chl-a, Turb, TOC, and TP. Validation of the neural network was done with the other group of data subset ($n = 15$). The RMSE and the R^2 of the simulation results were calculated with all samples (see Table 4). Application of multiple linear regression models [Eqs. (5)–(8)] for retrieval of Chl-a, Turb, TOC and TP yielded a R^2 value ranging from 0.74 to 0.87 for the calibration samples, and the R^2 value for validation samples ranged from 0.63 to 0.88. Comparing the calibration and validation results, the regression models performed were stable except that for TP in which the R^2 value was relatively low. Compared to regression models, the neural network models had a pronounced improvement for both calibration and validation samples in which the R^2 value ranged from 0.92 to 0.99, correspondingly the RMSE value reduced significantly (see Table 4).

4.3 *Retrieval accuracy comparison*

On the basis of both the determination coefficient (R^2) and RMSE, the estimation accuracy was greatly improved when the neural network model was applied (Table 4). The relative RMSEs (RMSE over measured average water quality parameters) of the estimation resulting from using the neural network algorithm were 6.2% for Chl-a, 7.4% for Turb, 9.7% for TOC, and 4.4% for TP, which were significantly lower than those resulting from the regression analysis with the corresponding value of 18.7%, 16.4%, 21.9%, and 24.2%, respectively. A graphical comparison between the result for the regression algorithm and that for the neural network models for Chl-a, Turb, TOC, and TP is shown in Figs. 4–7. In comparing the estimated and measured Chl-a concentration, the result derived from neural network model was much convergently distributed along the 1:1 line while there were four samples far from the 1:1 line for the result derived from the multiple linear regression model.

A similar result can be found for the estimated and measured Turb, the result from the neural network model well distributed along the 1:1 line, while that from the multiple linear regression model tended to underestimate turbidity (Fig. 5). Chagan Lake is a highly turbid water body, which can cause saturation for satellite received upwelling signals as found from other research as well,[3,19,29]. The high remote sensing reflectance in Fig. 2 also affirmed the higher suspended matter in the water. According to estimated and measured TOC, both the multiple linear regression model and the neural network model had a stable performance for TOC estimation (Fig. 6). It also was noteworthy that both models underestimated TOC when its concentration was above 25 mg/L in this case study. A multiple linear regression model tended to underestimate TP (Fig. 7), while the neural network model showed a much better result for its effective compensation of a nonlinearity problem.[2,24] It can be found that all results clearly indicate that the performances of the neural network model are superior to that of the regression model.

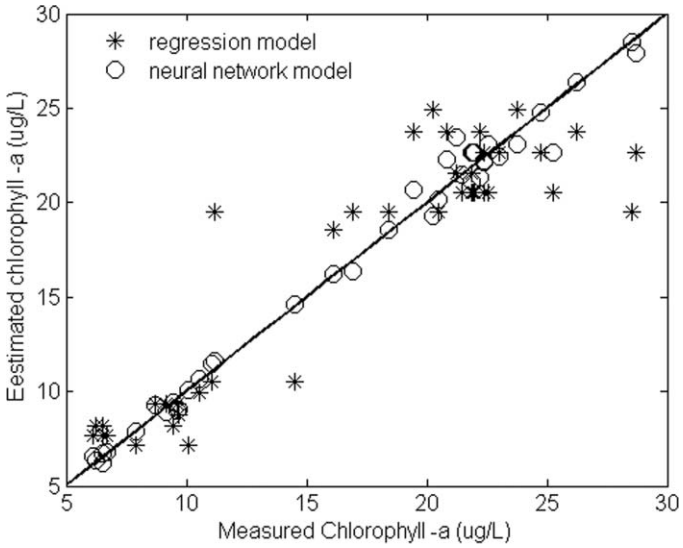

Fig. 4 Comparison of estimation results from regression analysis and neural network for Chl-a.

4.4 *Water Quality Parameter Mapping*

Both regression algorithm and neural network models were applied to all the lakes in the study area. Density-sliced maps of Chl-a, Turb, TOC, and TP concentration are shown in Figs. 8–11. The left graph in Figs. 8(a)–11(a) is produced by the regression model, while the right one Figs. 8(b)–11(a) is produced by the neural network throughout.

The lowest Chl-a concentration values were observed in the southern part of Xinmiao Lake and almost the whole area of Kuli Lake, close to the inlet of the Songhua River and Nenjiang River, respectively, which provide relatively clear and fresh water resources for these two lakes. Chl-a concentration between 8 to 16 ug/L was observed in the middle and north part of Xinmiao Lake which is connected with Chagan Lake via the narrow channel surrounded by aquatic grass, Kuli Lake, and the north part of Chagan Lake. Most of the near shore area of Chagan Lake had a Chl-a concentration between 16 to 24 ug/L, and the middle part had the highest

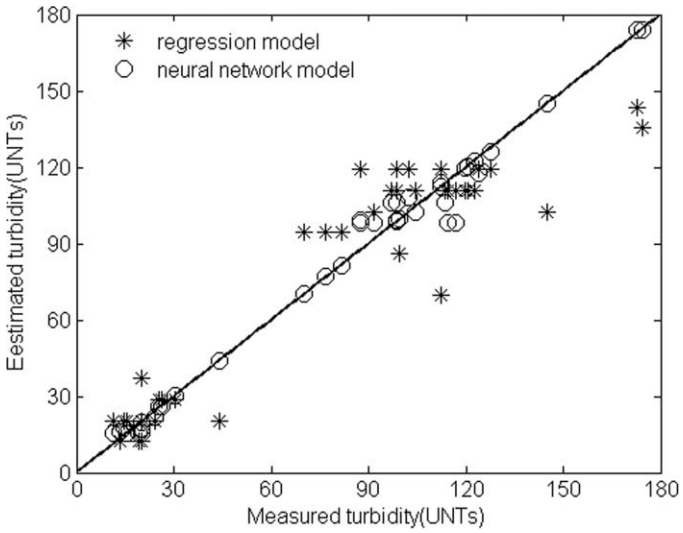

Fig. 5 Comparison of estimation results from regression analysis and neural network for turbidity.

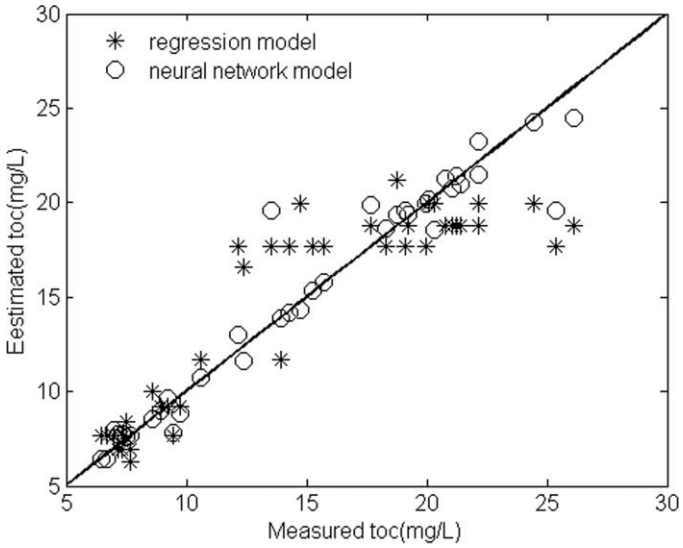

Fig. 6 Comparison of estimation results from regression analysis and neural network for TOC.

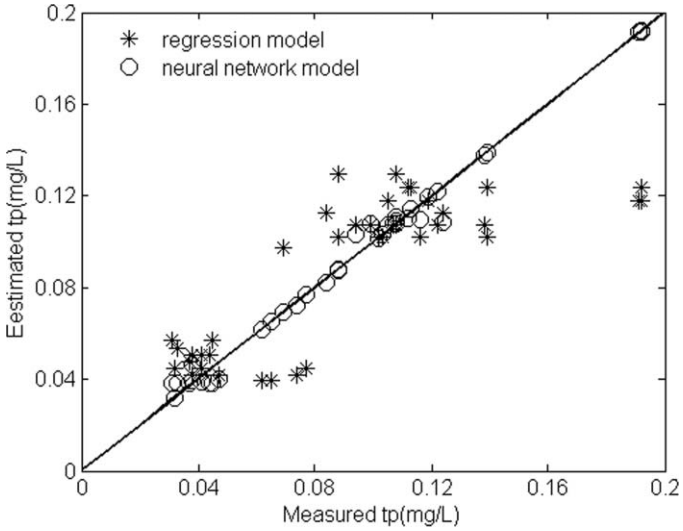

Fig. 7 Comparison of estimation results from regression analysis and neural network for TP.

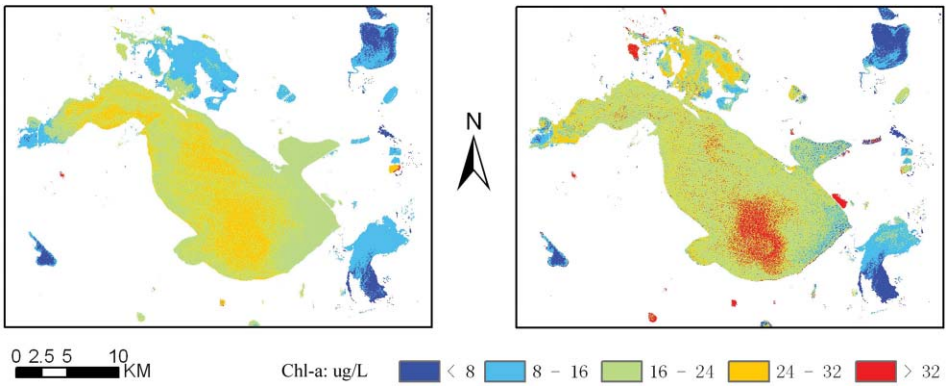

Fig. 8 Variation map of Chl-a retrieved from (a) regression and (b) neural network estimation.

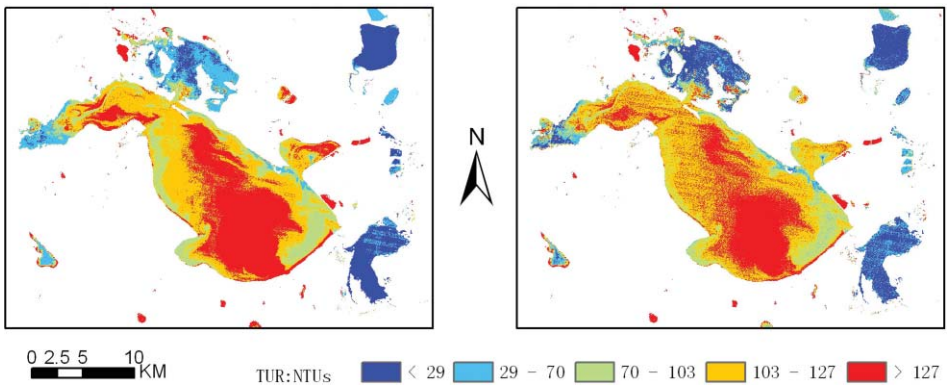

Fig. 9 Variation map of Turb retrieved from (a) regression and (b) neural network estimation.

Chl-a concentration. From Fig. 4, it can be seen that the empirical regression model tends to underestimate the Chl-a concentration in the study area; however, the neural network model was more accurate in estimating all water quality parameter values. Comparing Fig. 8(a) and Fig. 8(b), it can be seen that the Chl-a map produced with the neural network had more distinct patterns than that with the regression model.

Figure 9 shows that the lower Turb values (29 to 70 NTUs) were distributed in most of Xinmiao and Kuli Lake, in addition to some parts of Beidadu Lake. Some near shore part of the Chagan Lake had a Turb concentration value between 70 to 103 UNTs, and a large area toward the lake center had a Turb concentration of 103 to 127 NTUs with the highest concentration at the lake center. It can be seen from Fig. 5 that it does not perform as well as the neural network model does. However, both the regression model and the neural network generated a similar turbidity distribution over Chagan Lake, which indirectly proved that both models could map turbidity concentration relatively accurately.

Figure 10 shows the lowest TOC values distributed in part of Xinmiao and Kuli Lakes. Most of Beidadu Lake and the water inlet and outlet of Chagan Lake had TOC concentrations between 8 to 11mg/L. For the regression analysis, TOC values ranged 11 to 16 mg/L on the surface water near Chagan lakeshore. The peak value was found in the middle of Chagan Lake ranging from 16 to 19 mg/L. Figure 6 shows that the regression model underestimated TOC values with the corresponding remote sensing reflectance on the TM image. The neural network modeling result shows TOC values greater than 19 mg/L over most of Chagan Lake, implying that this approach outperformed the regression analysis shown in Fig. 6.

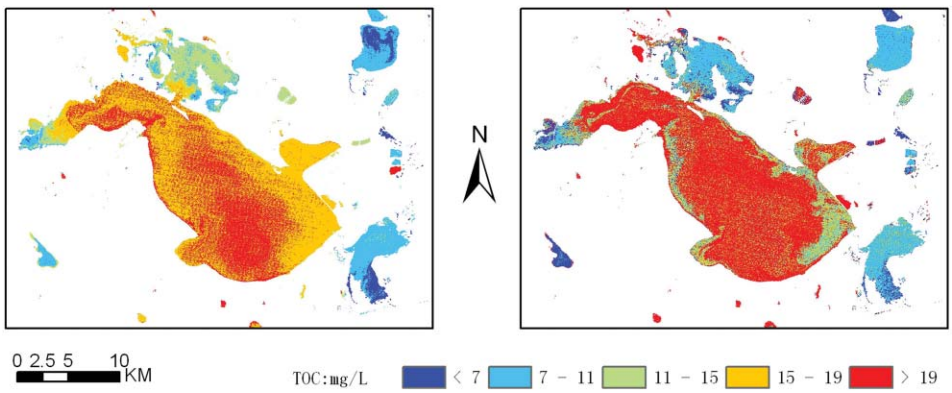

Fig. 10 Variation map of TOC retrieved from (a) regression and (b) neural network estimation.

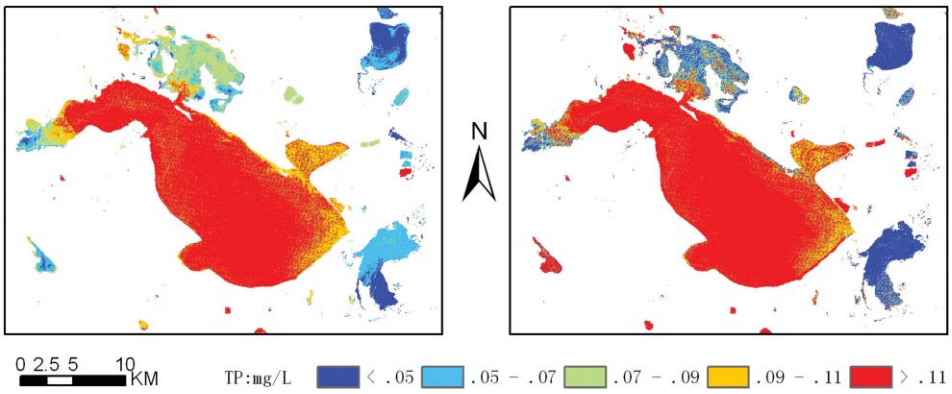

Fig. 11 Variation map of TP retrieved from (a) regression analysis (b) and neural network estimation.

By comparison, the neural network model performed much better for estimating TP than the other three water parameters. From Fig. 7, it can be seen that the regression model underestimated TP in Chagan Lake. TP distribution from the regression model differs from that of neural network estimation in some parts of the study area, though the mapping result for Chagan Lake itself were quite similar with both models (Fig. 11). This study shows that TP estimation with Landsat TM data is relatively accurate though the theoretical basis for TP spectral response is still not clear.[30,31] Future work and experiments in this area should be carried out to study how the water quality parameters, such as phosphorus, nitrogen, and biochemical oxygen demand, interact with water leaving radiance. However, statistical analyses of some water quality parameters, which have weak interaction with water leaving irradiance, still need to be carried out to improve our theoretical understanding.

4.5 *Future Research*

Several aspects in this study have highlighted the need for further work. First, future studies should have more *in situ* samples for training the neural network. Second, the neural network method should be expanded to evaluate multidate imagery data for operational applications. As previously mentioned, there are many lakes in the Songnen Plain, which have developed through similar historical environments, having similar surface water conditions. The method can therefore be applied to these other lake bodies where remote sensing of water quality parameters can play an important and time-saving role.

The small number of *in situ* samples available concurrent with TM overpass handicapped this study. The neural network model worked well, but more training samples will make it perform better. In any future study, more representative samples covering the entire range of water parameters in the study area should be obtained for the training purpose. For this study samples were not taken from the lake center, where high concentrations of Chl-a, Turb, and TOC were apparent from the satellite results. Another issue that must be dealt with in determining water quality from multidate imagery is the possibility that different phytoplankton species and different suspended sediment types could be present in the water through different seasons. These variations in phytoplankton species, inducing the variation in water spectral signatures, may change the inversion model parameters; even in some of the sensitive spectral bands.[11,21] Different phytoplankton species have their own phonological characteristics, and it is not possible that all these species will have the same spectral characteristics. Considering the high potential that the neural network method has for spectrally estimating water quality parameters, it is essential to obtain sufficient *in situ* water samples to train the network for water quality parameter estimation under all circumstances.

5 Conclusions

In this study, both a regression model and a neural network model were established to analyze Chagan Lake and its associated lakes for surface water quality parameters evaluation. It shows that the neural network is more useful in modeling the transfer function between the water quality parameters Chl-a, Turb, TOC, TP, and Landsat TM received radiances. Maps of the concentrations of these parameters were produced with both regression analysis and a neural network model. The neural network is able to model the nonlinear transfer function better than traditional regression analysis, though regression analysis is an easier method to use for linear transfer functions. The neural network with Landsat TM visible and near-infrared bands as inputs should prove useful in evaluating the water quality parameters in other lakes of the Songnen Plain with similar characteristics, where values of constituents such as chlorophyll, dissolved organic matter are unknown. In the future a new generation of satellite-borne sensors with higher spatial and spectral resolution will be launched. Imagery data acquisition with these sensors, coupled with the use of neural networks, will play a much greater and more accurate role in lake water parameter monitoring in the Songnen Plain.

Acknowledgments

The authors would like to thank the Key Knowledge Innovation Project of the Chinese Academy of Sciences (KZCX2-YW-340 and KZCX2-YW-QN305) for financial support. Special thanks are also due to Dr. Chen Ming and Dr. Wang Zhiqiang for their help with the field work for water samples collection and to Zhang Yuxia for her helpful work in analyzing these water quality parameters. Many thanks are also due to Professor Lin Li from Indiana University–Purdue University at Indianapolis for going through the final revised manuscript and giving valuable comments. The authors also thank the anonymous reviewers for their valuable comments and suggestions.

References

1. C. Giardino, M. Pepe, P. A. Brivio, P. Ghezzi, and E. Zilioli, "Detecting chlorophyll, Secchi disk depth and surface temperature in a sub-alpine lake using Landsat imagery," *Sci. Total Environ*. **268**, 19–29 (2001).
2. Y. P. Wang, H. Xia, J. Fu, and G. Y. Sheng, "Water quality change in reservoirs of Shenzhen, China: Detection using LANDSAT/TM data," *Sci. Total Environ*. **328**, 195–206 (2004).
3. A. G. Dekker and S. W. Peters, "The use of the Thematic Mapper for the analysis of eutrophic lakes: A case study in the Netherlands," *Int. J. Remote. Sens*. **14**(5), 799–821 (1993).
4. K. X. Xing, H. C. Guo, Y. F. Sun, and Y. T. Huang, "Assessment of the spatial-temporal eutrophic character in the Lake Dianchi," *J. Geogr. Sci.*. **15**(1), 37–43 (2005).
5. V. Klemas, D. Barlett, W. Philpo, and R. Roger, "Coasta and estuarine studies with ERTS-1 and Shklab," *Remote Sens. Environ*. **3**, 153–177 (1974).
6. Y. Zhang, J. Pulliainen, S. Koponen, and M. Hallikainen, "Application of an empirical neural network to surface water quality estimation in the Gulf of Finland using combined optical data and microwave data," *Remote Sens. Environ*. **81**, 327–336 (2002).
7. T. Lindell, D. Pierson, G. Premazzi, and E. Zilioli, *Manual for Monitoring European Lakes Using Remote Sensing Techniques*, EUR 18665 EN, Official Publications of the European Communities, p. 161, Luxembourg (1999).
8. X. T. Liu, *Management on Degraded Land and Agricultural Development in the Songnen Plain*, pp. 25–27, Science Press, Beingjing (2001) (in Chinese).
9. G. Wang, Y. Zhang, and F. Gao, "Important wetland of west Jilin and its eco-environmental function," *J. Soil & Water Conserve*. **15**(6), 121–124 (2001) (in Chinese).

10. A. A. Gitelson, G. Garbuzov, F. Szilagyi, K. H. Mittenzway, A. Karnieli, and A. Kaiser, "Quantitative remote sensing methods for real-time monitoring of inland waters quality," *Int. J. Remote. Sens*. **14**(7), 1269–1295 (1993).
11. C. E. Binding, H. J. John, R. P. Bukata, and G. B. William, "Spectral absorption properties of dissolved and particulate matter in Lake Erie," *Remote Sens. Environ*. **112**, 1702–1711 (2008).
12. H. Gordon and A. Morel, *Remote Assessment of Ocean Color for Interpretation of Satellite Visible Imagery: A Review*, pp. 3–51, Springer-Verlag, New York (1983).
13. C. Pattiaratchi, P. Avery, and A. Wyllie, "Estimates of water quality in coastal waters using multidate Landsat Thematic Mapper data," *Int. J. Remote Sens*. **15**, 1571–1584 (1994).
14. C. Song, C. E. Woodcock, K. C. Seto, M. P. Lenney, and S. A. Macomber, "Classification and change detection using Landsat TM data: When and how to correct atmospheric effects?," *Remote Sens. Environ*. **75**, 230–224 (2001).
15. H. Ouaidrari and E. F. Vermote, "Operational atmospheric correction of Landsat TM data," *Remote Sens. Environ.* **70**, 4–15 (1999).
16. Y. Zhang, J. Pulliainen, S. Koponen, and M. Hallikainen, "Water quality retrievals from combined Landsat TM data and ERS-2 SAR data in the Gulf of Finland," *IEEE Trans. Geosci Remote. Sens.* **41**(3), 622–629 (2000).
17. L. A. Milovich, Frulla, and D. A. Gagliardin, "Environment contribution to the atmospheric correction for Landsat-MSS images," *Int. J. Remote Sens*. **16**, 2515–2537 (1995).
18. T. Sabine and H. Kaufmann, "Lake water quality monitoring using hyperspectral airborne data—a semi-empirical multisensor and multitemporal approach for the Mechlenburg Kake District, Germany," *Remote Sens. Environ*. **81**, 228–237 (2002).
19. D. L. Woodruff, R. P. Stumpf, J. A. Scope, and H. W. Paerl, "Remote estimation of water clarity in optically complex estuarine waters," *Remote Sens. Environ*. **68**, 41–52 (1999).
20. Z. P. Lee and L. K. Kendall, "Absorption spectrum of phytoplankton pigments derived from hyperspectral remote-sensing reflectance," *Remote Sens. Environ.* **89**, 361–368 (2004).
21. L. E. Keiner and X. H. Yan, "Neural network model for estimating sea surface chlorophyll and sediments from thematic mapper imagery," *Remote Sens. Environ.* **66**, 153–165 (1998).
22. V. M. Krasnopolsky, "Neural network emulations for complex multidimensional geophysical mappings: Applications of neural network techniques to atmospheric and oceanic satellite retrievals and numerical modeling," *Rev. Geophys*. **45**, RG3009 (2007).
23. V. M. Krasnopolsky, L. Breaker, and W. Gemmill, "A neural network as a nonlinear transfer function model for retrieving surface wind speeds from the special sensor microwave imager," *J. Geophys. Res*. **100**(C6), 11033–11045 (1995).
24. V. M. Krasnopolsky, W. Gemmill, and L. A. Breaker, "Neural network multiparameter algorithm for SSM/I ocean retrievals: Comparisons and validations," *Remote Sens. Environ*. **73**, 133–142 (2000).
25. V. M. Krasnopolsky, M. S. Fox-Rabinovitz, and D. V. Chalikov, "New approach to calculation of atmospheric model physics: Accurate and fast neural network emulation of longwave radiation in a climate model," *Mon. Weather Rev*. **133**, 1370–1383 (2005).
26. R. Beale and T. Jackson, *Neural Computing: An Introduction*, Adam Hilger, Bristol (1990).
27. A. Gitelson, G. Dall'Olmo, M. Moses, D. 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*. **112**, 3582–3593 (2008).
28. R. K. Vincent, X. Qin, R. M. McKay, J. Miner, K. Czajkowski, and J. Savino, "Phycocyanin detection from Landsat TM data for mapping cyanobacterial blooms in Lake Erie," *Remote Sens. Environ*. **89**, 381–392 (2004).
29. G. Neukermans, K, Ruddick, B, Emilien, D, Ramon, B, Nechad, and P.-Y. Deschamps, "Mapping total suspended matter from geostationary satellites: A feasibility study with SEVIRI in the Southern North Sea," *Opt. Express*, **17**, 14029–14052 (2009).

30. C. F. Wu, J. P. Wu, J. G. Qi, L. S. Zhang, H. Q. Huang, L. P. Lou, and Y. X. Chen, "Empirical estimation of total phosphorus concentration in the mainstream of the Qiantang River in China using Landsat TM data," *Int. J. Remote. Sens.* **31**(9), 2309–2324 (2010).
31. T. Kutser, H. Arst, T. Miller, L. Krmann, and A. Milius, "Telespectrometrical estimation of water transparency, chlorophyll-a and total phosphorus concentration of Lake Peipsi," *Int. J. Remote. Sens.* **16**(16), 3069–3085 (1995).

Kaishan Song is an associate professor at Northeast Institute of Geography and Agroecology (NEIGAE), Chinese Academy of Sciences (CAS) and is currently working as a research scientist at Indiana University–Purdue University at Indianapolis. He received his MS degree in remote sensing and cartography from the Northeast Normal University, China in 2002, and his PhD degree in remote sensing application from NEIGAE, CAS in 2005. His current research interests include water quality monitoring with remotely sensed data, hyperspectral modeling for LAI and chlorophyll concentration estimation, vegetation bio-physical parameters inversion with MODIS and MERIS data coupled with empirical and radiative transfer model (PROSAIL). He is the author of more than 50 journal papers of which 7 are English peer-reviewed journal paper, and has written 1 book.

Zongming Wang is a professor at Northeast Institute of Geography and Agroecology (NEIGAE), Chinese Academy of Sciences (CAS). He received his MS degree in applied ecology from the Northwest Agricultural-forest University of China in 2002, and his PhD degree in remote sensing application from the NEIGAE, CAS in 2005. His main study field is application of RS and GIS in land resource and global change. He published more than 70 articles, 13 of which are English peer-reviewed journal paper.

Bai Zhang is a professor at Northeast Institute of Geography and Agroecology (NEIGAE), Chinese Academy of Sciences (CAS). He received his BS degree in physical geography from Peking University, China in 1983, and his PhD degree in remote sensing application from NEIGAE, CAS in 2005. His main study field is application of RS and GIS in land resource and landscape dynamics study. He published more than 180 articles, 7 of which are peer-reviewed English journal paper.

Yuanzhi Zhang is a professor at Institute of Space and Earth Information Science (ISEIS), the Chinese University of Hong Kong (CUHK). He received his MS degree in remote sensing and cartography from Institute of Remote Sensing Application (IRSA), CAS, China in 1991, and his PhD degree in remote sensing Science and GeoInformation (GIS) in Helsinki University of Technology (HUT), Finland in 2005. His main study interest includes: coastal zone monitoring and management; natural resources, land use/cover, and environmental processes; microwave remote sensing of land surface (water, soil, rock, forest, and snow/ice); inherent optical properties of lakes, rivers, coastal and marine waters. He acted as associate editor of African Journal of Agricultural Research since 2006. He published more than 60 articles, more than 30 of which are peer-reviewed English journal papers.

Biographies of other authors not available.