Improving the accuracy of estimation of eutrophication state index using a remote sensing data-driven method: A case study of Chaohu Lake, China

Bo Xiang[1, 2]*, Jing-Wei Song[1, 2], Xin-Yuan Wang[1] and Jing Zhen[1, 2]
[1]*Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China*
[2]*Graduate School, Chinese Academy of Sciences, Beijing 100049, China*

ABSTRACT

Trophic Level Index (TLI) is often used to assess the general eutrophication state of inland lakes in water science, technology, and engineering. In this paper, a data-driven inland-lake eutrophication assessment method was proposed by using an artificial neural network (ANN) to build relationships from remote sensing data and in-situ TLI sampling. In order to train the net, Moderate Resolution Imaging Spectroradiometer (MODIS, which has a revisit cycle of 4 times per day) data were combined with in-situ observations. Results demonstrate that the TLI obtained directly from remote-sensing images using the data-driven method is more accurate than the TLI calculated from the water quality factors retrieved from remote-sensing images using a multivariate regression method. Spatially continuous and quasi-real time results were retrieved by using MODIS data. This method provides an efficient way to map the TLI spatial distribution in inland lakes, and provides a scheme for increased automation in TLI estimation.

Keywords: data driven, trophic level index, MODIS, artificial neural network, inland lake

INTRODUCTION

Some countries use Trophic State Index (TSI) to assess the trophic condition of a lake and associated changes (Carlson, 1977; Carlson and Simpson, 1996). Trophic Level Index (TLI) is another approach that was developed by the Ministry for the Environment of New Zealand (Burns and Bryers, 2000; Verburg et al., 2010). TLI is recommended by the China National Environmental Monitoring Centre for measuring inland lakes' eutrophication levels (CNEMC, 2009). Currently, TLI can either be calculated from water samples or from in-situ hyper-spectral data measurements (Li et al., 2006), and is usually calculated by interpolating all of these observation points. However, as demonstrated by Kuster (2004), nutrients in inland lakes have various shapes and complex spatial distributions. Some of the nutrient clusters even have a strip shape less than 30 m wide. Furthermore, the complicated non-linear relation is simplified as a linear relationship (Jiang et al., 2013). All this complexity of hydrology makes the retrieval of TLI by interpolation of only some ground observation points unreliable. On the other hand, to solve the laborious sampling and time-consuming chemical analysis problem, Yao used remote-sensing techniques that were fast, wide ranging, low-cost, and period dynamic (Yao et al., 2009). However, this method still cannot satisfy regional monitoring needs. TLI synthesizes many nutrition targets, and the synthetic eutrophication index is computed based on many nutrition targets retrieved from remote-sensing images. Some previous studies tried to obtain TLI by building multivariate regression from remote-sensing data to create a eutrophication index combined with in-situ measurements (He et al., 2009). All these studies provide a time-saving way of acquiring the spatial distribution of TLI. However, there are two deficiencies in such studies. Firstly, the revisit cycle of remote-sensing satellite data used in these studies is usually several days, which makes it almost impossible to collect the in-situ data at exactly the same time. Considering bad weather and other imaging conditions, the data were not suitable for daily monitoring. Secondly, the regression approach for retrieving TLI was a linear retrieval method, but it has been proven that a non-linear relationship exists (NSCEP, 1976) between satellite bands and the five water quality factors of interest. Studies (NSCEP, 1976; Yang et al., 2006) suggest that a non-linear model may be better adapted to retrieve results.

To overcome the accuracy and efficiency problem, we propose a rapid data-driven method for monitoring the TLI distribution of Chaohu Lake. A data-driven method (Darema, 2004; Darema et al., 2005) would be appropriate to rapidly retrieve TLI information directly from satellite images (Li, 2007; Ei Serafy et al., 2007; Song et al., 2014b). CBERS satellite data were substituted by Moderate Resolution Imaging Spectroradiometer (MODIS) because the MODIS revisit cycle is 4 times per day (Freeborn et al., 2011) and the spectrum covered by the image is wider. Multivariate regression (linear analysis) was replaced by an artificial neural network (non-linear analysis) regression to better retrieve TLI.

This paper introduces the data acquisition and processing procedures and reviews the feasibility of using ANN and MODIS to map the TLI distribution. The results of ANN regression and the time-series TLI distribution maps are discussed and presented, concluding with prospects for future improvement.

MATERIALS AND METHODS

Study area

Chaohu Lake is a typical large, shallow, subtropical lake. With a mean water depth of 2.69 m covering an area of 780 km², the lake is located between 30°58´ and 32°06´N and 116°24´ and 118°00´E. Its northwestern border is less than 10 km away from

---

the capital city of Anhui Province, Hefei, and its eastern border is very near Chaohu City (Fig. 1). Being one of the drinking water catchment areas for both cities, the water quality of Chaohu Lake is of the utmost importance. Some remote-sensing inversion work on water quality has been done in Chaohu (Mei et al., 2008b; Xie et al., 2010).

The lake is managed and monitored by two different administrations. Each administration takes responsibility for 6 observation points. These sample points are used to get observation data on surface water quality every other day from May to October and once a week for the rest of the year. The current 12 sampling points are the final choice of the Ministry of Environmental Protection and the local environmental protection department, based on rigorous analysis and verification synthesized from many different factors, such as lake area, lake basin form, condition of recharge, effluent and water intake, the location and scale of sewage disposal facilities, pollutant circulation, and migration and transformation of algae in water. The current 12 sampling points satisfactorily control water quality monitoring and verification in Chaohu Lake. At the same time, a more homogeneous distribution or more sample sites might not lead to better verification; it depends on the representativeness of the sample sites according to the characteristics of the water. Actually, in the early monitoring of Chaohu Lake in 2009, the sample sites were homogeneously distribution as shown in Fig. 2.

A nutrient-level distribution for Chaohu Lake was drawn based on the sample sites, and the area and distribution of chlorophyll, which is the main indicator of nutrient levels in lake water, was computed (Fig. 3). The central part and north central area are known to be consistently eutrophic, and therefore few sample sites were chosen in this region. More sample sites will be deployed in Chaohu Lake in the future to monitor changes in nutrient distribution and improve the accuracy of water quality data verification.

To demonstrate the feasibility of our method, high temporal and spectral resolution MODIS data from April 2 to July 13, 2013, were acquired. MODIS Surface-Reflectance Products (MOD09) were downloaded from the National Aeronautics
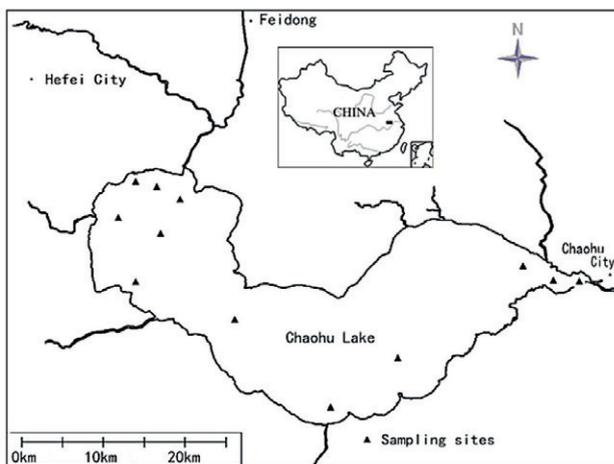
Figure 1
Location of sampling sites within Chaohu Lake
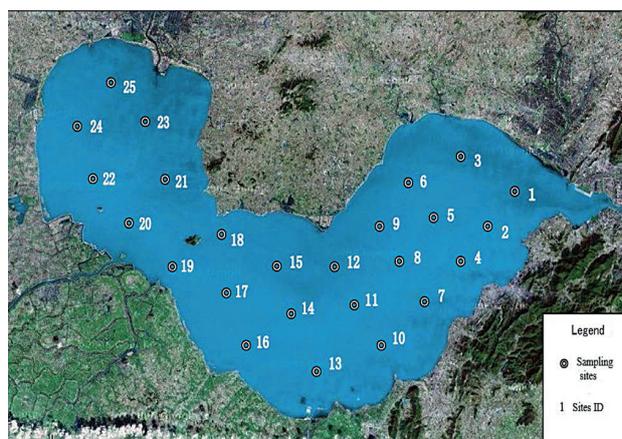
Figure 2
Satellite image of Chaohu Lake showing the location of sampling sites used in 2009
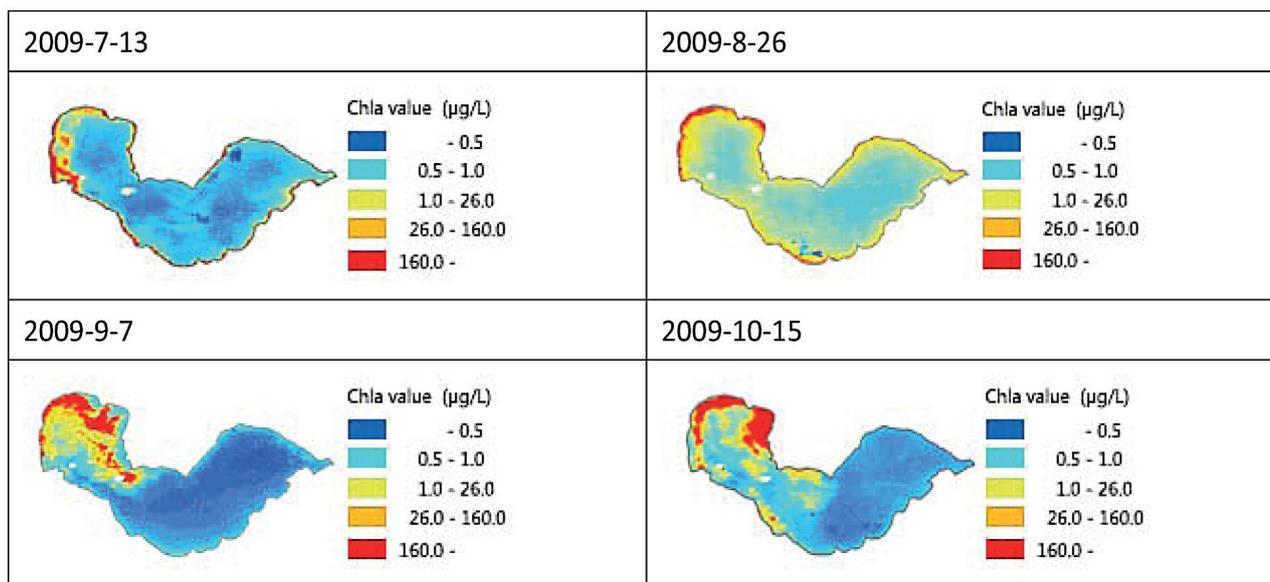
Figure 3
Distribution of chlorophyll a in Chaohu Lake in 2009

and Space Administration (NASA) (Vermote, 2013). MOD09 is the Level 2 product automatically generated from the MODIS Level 1B land bands intended to estimate the surface spectral reflectance. Atmospheric effects are almost removed in MOD09, though there are articles reporting on the deficiencies of MOD09 (Guang et al., 2013).

In-situ samples were obtained from 2 April to 13 July 2013 with a 7-day interval from April to May and 1-day interval from June to July. The sites were accessed by motor boat and water samples were taken from 0.5 m below the water surface using plastic samplers. A volume of 1.5 to 3 ℓ water was taken. Water quality was also assessed using an EXO2 multifunctional measuring instrument, which was lowered 1–2 cm into the water. The samples were placed in a box filled with ice and stored in the dark for a short period before laboratory analysis. All laboratory analyses were done at the local environmental monitoring centres. A spectrophotometric method was used to determine chlorophyll *a* according to Wei et al. (2002). A dichromate method was used to determine COD according to National Standard GB/T11914-89 (Yin, 1989). An ammonium molybdate spectrophotometry method was used to determine TP according to National Standard GB11893-89 (Yuan and Yao, 1989). An alkaline potassium persulfate digestion UV spectrophotometric method was used to determine TN according to National Standard HJ636-2012 (DLMEMC, 2012). The measured results were returned 2 days after the water samples were taken to the laboratory. The Environmental Protection Monitoring Station of the Chaohu Management Bureau has specialized departments and personnel responsible for daily equipment calibration and the manufacturer's engineers do regular equipment calibration. Three replicate samples were collected for each sampling point, and the average for each sampling site and time reported.

Training data selection was based on the following principles:

- Sampling points close to the shore were excluded from the dataset. Since the resolutions of MOD09 are 250 m (Band 1 and Band 2) and 500 m (Band 3 to Band 7), off-shore water will be mixed with coastal water, which will cause spectral distortion.
- Cloud-contaminated pixels are excluded in the MODIS cloud mask product (MOD35).

Pixel values for each sampling point were extracted. Only Bands 1 to 5 were used in this study in order to avoid redundancy. Bands 6 and 7 have rarely been reported as being suitable for monitoring water. A total of 63 samples of in-situ data concurrent with MOD09 were obtained and the dataset was randomly divided into 3 groups: the training set (45 points), the validation set (9 points) and the testing set (9 points).

Feasibility of mapping TLI spatial distribution from MODIS

Chaohu Lake is a phosphorus-controlled lake, and TLI is used by the managers of the lake as an indicator for assessing the trophic level (Zhang et al., 2013; Wang et al., 2007). TLI is a chlorophyll-based index aiming to characterize, in both a qualitative and quantitative manner, the level of eutrophication of lakes. It is a weighted sum based on chlorophyll *a* and several other substances. Currently, the TLI of Chaohu is calculated using the national standard and the following equations (Wang et al., 2002; Jin et al., 1990):

$$\text{TLI}(\Sigma) = \sum_{j=1}^{m} W_j \times \text{TLI}(j) \tag{1}$$

$$W_j = r_{ij}^2 / \sum_{j=1}^{m} r_{ij}^2 \tag{2}$$

TLI_j is the *j*th composite indicator with the corresponding weight W_j. The r_{ij} value given in Table 1 gives the correlation coefficient for the relationship between the reference chlorophyll concentration and each indicator. The TLI of all the observation points was calculated from 5 components including chemical oxygen demand (COD), total phosphorus (TP), total nitrogen (TN), chlorophyll *a* (Chl–*a*), and Secchi depth (SD). Formulas for the TLI of each component are given below:

$$\text{TLI(Chl–}a) = 10(2.5 + 1.086 \ln(\text{Chl–}a)) \tag{3}$$

$$\text{TLI(TP)} = 10(9.436 + 1.624 \ln(\text{TP})) \tag{4}$$

$$\text{TLI(TN)} = 10(5.453 + 1.694 \ln(\text{TN})) \tag{5}$$

$$\text{TLI(SD)} = 10(5.118 + 1.94 \ln(\text{SD})) \tag{6}$$

$$\text{TLI(COD)} = 10(0.109 + 2.661 \ln(\text{COD})) \tag{7}$$

Equations 3 to 7 are empirical regression equations based on a survey of eutrophication levels of more than 20 lakes in China (Jin et al., 1990). According to *Specifications for Lake Eutrophication Survey (2nd Edition)* (Jin et al., 1990), these 20 lakes are representative of a range of depths, sizes and meteorological conditions. For example, the sample includes Aydingkol Lake with a depth of less than 1 m, and Nam-Co Lake with a depth of 125 m; as well as Tianchi Lake with a surface area of less than 10 km², and Qinghai Lake with a surface area of 4 500 km². Climatic conditions represented range from those of plain to highland and mountainous regions. Most of these lakes were highly eutrophic. Therefore, the method presented is reliable in lakes with a moderate to high eutrophication status. The empirical coefficients can be used in other lakes only after determining the nutrition composition of a typical lake in the region of interest.

The units for each component are given in Table 1. The score for TN and TP was adjusted in the range from 0 to 100 according to the international standard (Jin et al., 1990) for lake eutrophication levels and was analysed for correlations between the index of the eutrophication and water quality parameter. Also taken into consideration was that TN and TP is generally higher in Chinese lakes relative to those in developed countries. Trophic status is categorised using the ΣTLI as follows:

- Oligotrophic: ΣTLI < 30
- Mesotrophic: 30 ≤ ΣTLI ≤ 50
- Light eutrophic: 50 < ΣTLI ≤ 60
- Mid eutrophic: 60 < ΣTLI ≤ 70
- Hypereutrophic: ΣTLI > 70

From the definition of ΣTLI, it has a linear relationship with the natural logarithm value of the five indicator values. Both band ratio combination methods (Gons et al., 2002; Song et al., 2013),

TABLE 1
Correlation between Chl-*a* and other substances influencing TLI

	Chl-*a* (mg/m3)	TP (mg/ℓ)	TN (mg/ℓ)	SD (m)	COD (mg/ℓ)
r_{ij}	1	0.84	0.82	−0.83	0.83
r_{ij}^2	1	0.7056	0.6724	0.6889	0.6889

namely the bio-optic method (Ma et al., 2006; D›Alimonte et al., 2012) and non-linear algorithms (Schwarz et al., 2002; Wu et al., 2009; Keiner, 1999; Zhang et al., 2002), were applied in retrieving the five indicators from satellite images. Therefore, it was theoretically possible to use TLI with satellite images and in-situ observations mapping the spatial distribution because there was a close relationship between the combination of satellite image bands and TLI values. However, there were two reasons why the TLI retrieval application was hindered using the white box method: bio-optic and multivariate regression. Firstly, current standard atmospheric correction procedures, especially developed for inland waters, were unable to remove the atmospheric effects on the data. Atmospheric effects will cause anomalies in bands in lakes, and this will lead to systematic errors in retrieving indicators. Secondly, errors caused by interactions will increase when the turbidity of the water rises. For example, the Chl-*a* estimation algorithm, based on the response at 469 nm, is more likely to be affected by concentrations of suspended sediments and coloured dissolved organic matter (CDOM), which can be more prevalent in shallow waters like Chaohu Lake (Darecki and Stramski, 2004; Shutler, 2007).

Therefore linear regression and the white box method are not suitable for retrieving the five components of TLI because systematic errors will be introduced into the final result. Machine learning is an effective way to solve non-linear problems in general. As a practical theory and algorithm of machine leaning, artificial neural network algorithms are more suitable in fitting non-linear relationships of attribute-value pair problems (Wu et al., 2005; Mitchell, 1997).

Neural networks

Neural networks are a kind of machine learning algorithm that imitates brain processes and were originally developed to solve non-linear problems like fitting, pattern recognition, clustering, and time-series prediction (Mitchell, 1997). They have been successfully applied in environmental sciences (Sattari et al., 2012; Krasnopolsky and Chevallier, 2003; Krasnopolsky and Schiller, 2003; Song et al., 2014a). Figure 4 (a) shows that traditional TLI retrieval requires all water quality factors to be calculated from remote-sensing data by means of neural networks. In this study, a neural network was used as a transferring function to link MODIS bands and in-situ TLI values. The neural network used in this study comprised of 3 layers (input layer, hidden layer, and output layer) and was a feed-forward type including back-propagation of errors. The flowchart of the satellite bands to the TLI neural network is presented in Fig. 4 (b). Traditional TLI retrieval requires an intermediate process to retrieve all water quality factors, but the remote sensing data-driven method can omit this intermediate process. The red dashed box in Fig. 4 (a) shows the difference between traditional TLI retrieval and the remote-sensing data-driven retrieval method.

In this study, the input layers used the first five bands of the surface reflectance product from MODIS (MOD09). For each observation point, the corresponding reflectance values were extracted from MOD09 as well as the ancillary data. Band 3, Band 4 and Band 5 (500-m resolution) were interpolated to the same pixel size as Band 1 and Band 2 (250-m resolution) by using the nearest neighbour algorithm. The output of the
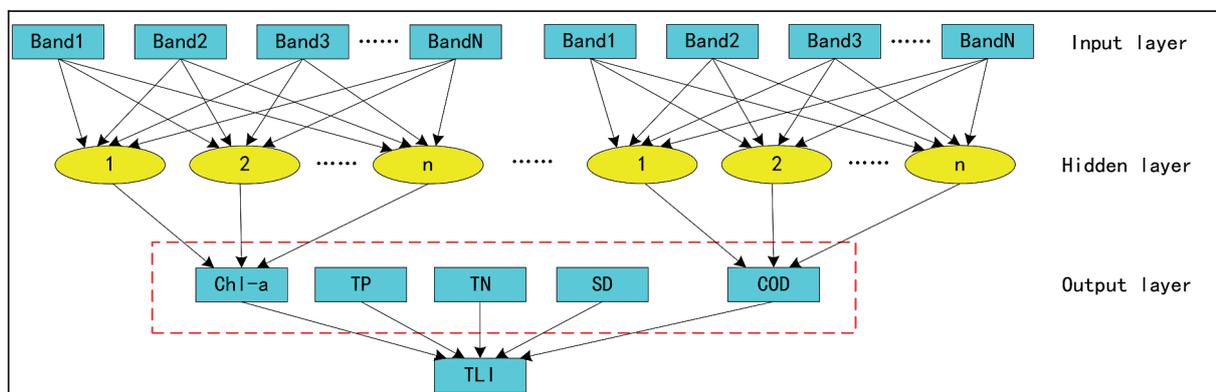
(a) Traditional TLI inversion process
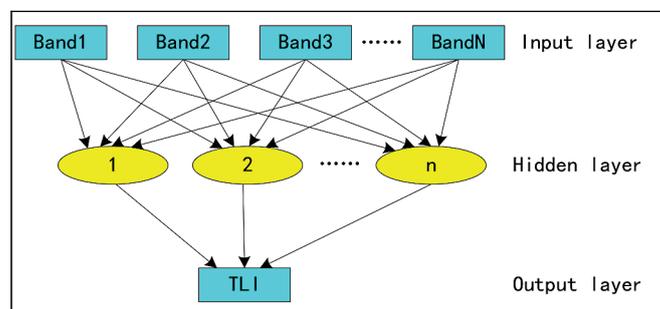
(b) Remote sensing-driven TLI inversion process

Figure 4
Architecture of satellite bands to TLI neural network

network was the TLI calculated by in-situ measurement of the corresponding pixel. In the hidden layer:

$$\text{Net}_j = f_{\text{sig}}\left(\sum_{i=0}^{N} V_{ij}\,\text{Band}_i\right) \tag{8}$$

where: Net_j represents the jth node in the hidden layer and v_{ij} represents weight for unit j corresponding to the ith input. f represents the activation function of a node. In this study we used a sigmoid function as follows:

$$f_{\text{sig}}(x) = \frac{1}{1 + \exp(-x)} \tag{9}$$

In the output layer, ΣTLI is calculated through the function as follows:

$$\Sigma\text{TLI} = f_0\left(\sum_{j=0}^{n} \omega_j\,\text{Net}_j\right) \tag{10}$$

where: ω_j is the corresponding weight for each hidden node. f_0 is the activation function. In this study, we used the commonly-used linear function. Both ω_j and v_{ij} were assigned with random values initially, and then modified by the delta rule according to the learning samples.

Out of a total of 63 points, 45 were training data, 9 were validation data, and 9 were testing data. The Levenberg-Marquardt approximation method was employed to minimize errors in each trial. In order to avoid over-training, hidden layer nodes between 2 and 5 were tested and results showed that 4 nodes was the optimal number. We also implemented a multivariate regression (MR) method as a comparison experiment. Correlations of the five bands, MR results, and neural network (Nodes 2 to 5) to TLI are given in Table 2. Two commonly-used indexes for measuring accuracies were utilized for testing the performances of the method, i.e., coefficient of determination (R^2) and mean square error (MSE):

$$\text{MSE} = \frac{\sum_{t=1}^{n}(R_t - F_t)^2}{n} \tag{11}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(R_t - F_t)^2}{\sum_{t=1}^{n}(R_t - \overline{R})^2} \tag{12}$$

The number of nodes in the hidden layer depends on the complexity of the relationship between input and output. Using

TABLE 2
The coefficient of determination (R^2) between TLI and MODIS data, multivariate regression (MR) results, and neural network results with different node numbers.

Values used to calculate correlations to TLI	R^2
MODIS Band 1	−0.1746
MODIS Band 2	0.1257
MODIS Band 3	−0.3280
MODIS Band 4	−0.3258
MODIS Band 5	0.0631
Multivariate regression (MR)	0.5677
Neural network (2 nodes)	0.7918
Neural network (3 nodes)	0.8482
Neural network (4 nodes)	0.8937
Neural network (5 nodes)	0.8576

enough nodes ensures the nonlinearity of the neural network in fitting the data. However, excessive nodes will lead to over-fitting when the network not only learns the real model but also takes in noise. In order to find the optimal node number, the coefficient of determination (R^2) and MSE were calculated to find the optimal value. R^2 indicates how well data fit a statistical model, and ranges from 0 to 1. If R^2 is closer to the numerical value 1, it indicates that this model fits the data better. Additionally, negative values of R^2 may occur when fitting non-linear functions to data. We trained each neural network with node numbers from 2 to 5 and found that the network with 4 nodes had the minimum MSE and highest correlation. Besides this, we used the thresholds of each level of TLI to classify the results and compare with the in-situ value. The comparison of the neural network and multivariate regression is presented in Table 3.

RESULTS

The coefficient of determination (R^2) and RMSE of the neural network were much better than for the MR results. This is due to the non-linear nature of the neural network in transferring the satellite bands into the five components of TLI and the final TLI value. Table 3 shows that the accuracy of classification using the MR method was not satisfactory. The difference between the retrieved result (59.5) and the actual result (60.5) was judged to indicate a classification error. Errors in classification occur because classification is based on using thresholds to directly separate results. Classification is done according to the score interval for different eutrophication level statuses. Accuracy of classification is evaluated using the actual observed values for the 12 water quality sample sites. The closer the retrieved value from the remote sensing imagery is to the observed value, the higher the accuracy of the classification result. Nevertheless, the large number of uncertainties in the water body environment of lakes leads to fuzziness in the categorisation and standardization of the different indicators, which may lead to the possibility of classification errors. Even though the retrieved value from the remote sensing imagery may be very close to the observed value, because the observed value led to a different classification, this will be judged as a classification error. Consequently, an improved evaluation method based on fuzzy mathematics should be used to evaluate the water quality level. Figure 5 shows a comparison of neural network results (4 nodes), MR results, and the TLI calculated from sampled data. The MSE of the 4-node hidden layer neural network appears to be much better than that of the MR results.

The results show that the application of neural networks in TLI retrieved from MODIS images outperforms the multivariate retrogression both in its regression accuracy and in the retrieval stability. Thus, the neural network based TLI mapping method can be used as a complement to the conventional data sampling method.

These three groups of images were obtained on 2 April 2013, 3 June 2013, and 11 July 2013, respectively.

TABLE 3
Comparison of regression results for neural network and multivariate regression (MR) methods

	MR	NN(2)	NN(3)	NN(4)	NN(5)
MSE	10.4689	8.0696	6.2538	5.3452	5.6777
Accuracy of classification	0.6508	0.7302	0.7777	0.8571	0.7619
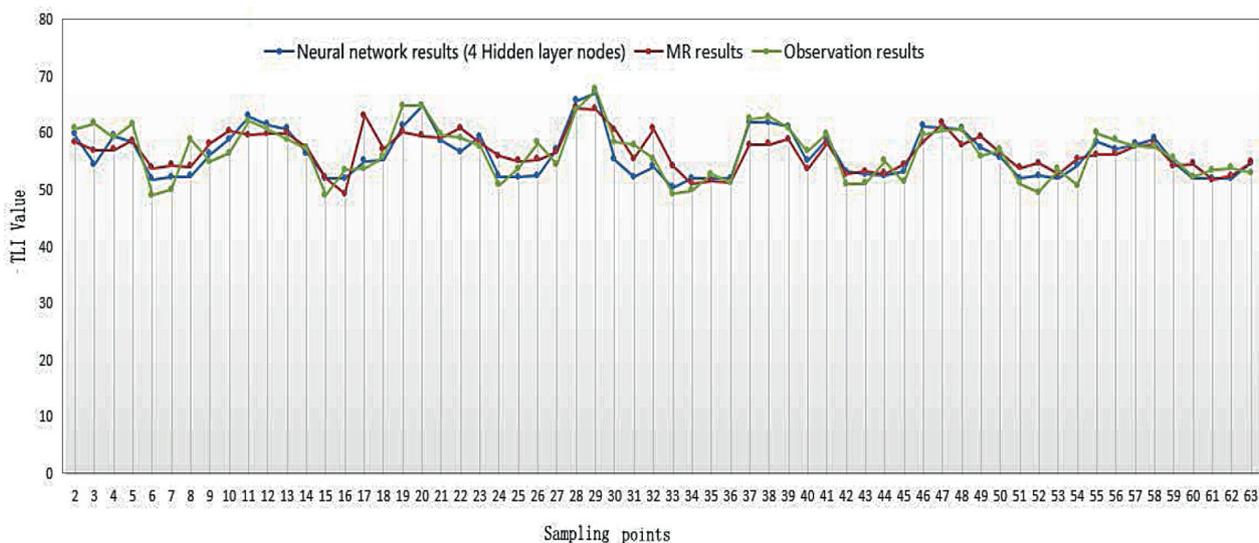
Figure 5

Comparison of neural network results (4 nodes), MR results, and TLI calculated from observation data

DISCUSSION

As Fig. 6 indicates, the neural network–retrieved spatial distribution of TLI outperformed the MR approach in two ways:

- The range of the TLI retrieved was set between the minimum and maximum of the training data. In this way, extreme values were avoided. Even though higher TLI values could be retrieved, this conservative algorithm maintained stability and controllability. By contrast, the range of the MR results extended both the minimum and maximum of the training data over a larger scale.
- Results calculated from the neural network data-driven method show great spatial heterogeneity, which means that more eutrophication state information was mined from the satellite images. On the contrary, the results of MR are spatially smooth. The linear MR method involves moderate- and low-resolution satellite images sharing the characteristic of having mixed pixels, which leads to similarity of the neighbouring pixels and a spatially smooth result. This means that the neural network results are indeed pixel-based while MR results suffer from the mixed-pixel effect.

One obstacle that affects the performance of both neural networks and MR is the atmosphere. MOD09 is generated by the Second Simulation of a Satellite Signal in the Solar Spectrum, Vector (6S) model with several atmospheric parameters taken either from the National Centres for Environmental Prediction (NCEP) (ozone, pressure) or from the MODIS data (aerosol, water vapour) (Vermote and Vermeulen, 1999). Future work might focus on atmospheric correction, and recently a better algorithm has been reported (Guang et al., 2013), which suggests that future improvement in the retrogression accuracy might depend on the new method.

Another factor that influences the performance of a neural network's data retrieval capability is the spatial resolution of MODIS imagery and disturbance of water when sampling. The low resolution of MODIS imagery means the pixel value is a mixture of the sampling area and surrounding area. As Kuster points out, the spatial distribution of water surface algae is at less than 30 m depth (Kuster, 2004). This may also apply to TLI distribution. The groundtruth is the point value

while the corresponding pixel is the mixture. Besides this, the water sampling method will disturb the surface water and push the scum and subsurface aggregations away from the ship. The joint force of the two factors brings uncertainty to the final results.

There is another very important question about vertical stratification of water. Vertical water column structure may lead to changes in concentration. Chaohu Lake is, however, an expansive shallow lake, and so the results obtained from sampling will be unaffected by vertical stratification. For other types of lakes, we are developing a three-dimensional model based on hydrodynamics, which will consider the influence of water column structure on water quality. In the future, the remote-sensing data will be assimilated with the three-dimensional hydrodynamic model and can be used to monitor and analyse the results of model simulation.

Finally, because the number of sample points is not sufficient, the accuracy of the model is bound to be affected, especially the accuracy of the results for the non-sampled area. Based on the accuracy error recorded between the actual measured value and the inversion value of the neural network model, the deployment of more sampling points in the areas of Chaohu Lake recording poor accuracy will improve the accuracy of the model while ensuring that a suitable number of points are sampled.

CONCLUSIONS

We implemented a rapid data-driven method for monitoring the TLI distribution of Chaohu Lake from MODIS satellite remote-sensing data on the basis of an artificial neural network. Advantages and potential future improvements of this innovative method are listed below:

- Results demonstrate that water quality distribution can be predicted with information retrieved from satellite remote-sensing images. From the perspective of information theory, reducing the intermediate steps may improve accuracy because of less information loss in the transformation process.
- The TLI distribution mapping interval can be improved. If weather permits, the mapping of TLI can be done twice a day. Compared with sampling method requiring access by

Acquisition time	Neural network method	MR method
2013-4-2		
2013-6-3		
2013-7-11		
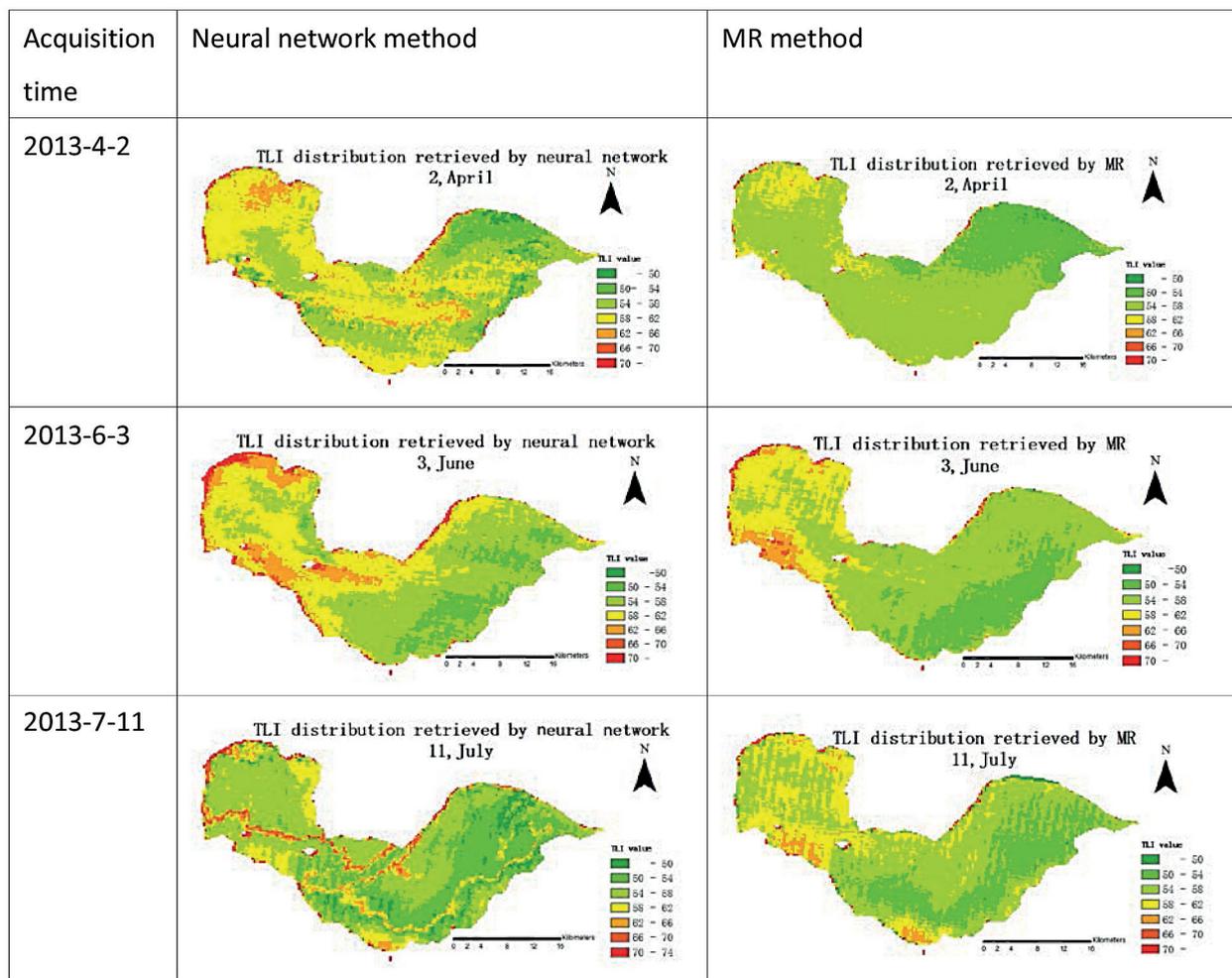
Figure 6
The retrieval results contrasting the neural network and MR

boat, this method saves both time and effort.

- Inaccuracy caused by surface water disturbance can be avoided to some extent. However, this depends on the accuracy of the training data acquired by the boat sampling method. This problem can be avoided if automatic water sampling stations are created in Chaohu Lake.
- Higher spatial resolution images with an appropriate revisit cycle may be used to improve the mapping result. To address the limitations of weather and data acquisition, high temporal resolution MODIS data were used in this study. Better spatial resolution Landsat imagery will provide better details for determining TLI. More accurate pixels concurrent to the observation point will be extracted from the image, which will lead to much better regression performance and more detailed information for areas with diffuse blue-green algae.
- This method demonstrates that ANN is much more suited to TLI determination than the MR method. The key reason for the improved performance lies in the non-linearity of neural networks. Various machine-learning algorithms have been extensively studied and successfully applied in fitting problems. Future work may focus on substituting a back-propagation neural network. For example, genetic programming has been applied for better performance in chlorophyll *a* concentration retrieval (Chang et al., 2013).

ACKNOWLEDGEMENTS

REFERENCES

BURNS NM and BRYERS G (2000) Protocols for monitoring trophic levels of New Zealand lakes and reservoirs. URL: https://www.mfe.govt.nz/withyou/funding/docs/5090_nzlm_protocol_complete_text.pdf (Accessed 15 October 2013).
CARLSON RE (1977) A trophic state index for lakes. *Limnol. Oceanogr.* **22** (2) 361–369.

CARLSON RE and SIMPSON J (1996) *A Coordinator's Guide to Volunteer Lake Monitoring Methods*. North American Lake Management Society, Madison. 96 pp.

CHANG NB, XUAN ZM and YANG YJ (2013) Exploring spatiotemporal patterns of phosphorus concentrations in a coastal bay with MODIS images and machine learning models. *Remote Sens. Environ.* **134** (7) 100–110.

CHANG NB, YANG YJ, DARANPOB A, JIN KR and JAMES T (2012) Spatiotemporal pattern validation of chlorophyll-*a* concentrations in Lake Okeechobee, Florida, using a comparative MODIS image mining approach. *Int. J. Remote Sens.* **33** (7) 2233–2260.

CNEMC (CHINA NATIONAL ENVIRONMENTAL MONITORING CENTER) (2009) Technical regulations for classification and evaluation method of lake (reservoir) eutrophication. URL: http://www.cnemc.cn/publish/100/news/news_8849.html (Accessed 15 October 2013).

DLMEMC (DALIAN MUNICIPAL ENVIRONMENTAL MONITORING CENTER) (2012) URL: http://kjs.mep.gov.cn/hjb-hbz/bzwb/shjbh/sjcgfffbz/201203/W020120410336075623072.pdf (Accessed 31 March 2015).

D'ALIMONTE D, ZIBORDI G, BERTHON JF, CANUTI E and KAJIYAMA T (2012) Performance and applicability of biooptical algorithms in different European seas. *Remote Sens. Environ.* **124** 402–412.

DARECKI M and STRAMSKI D (2004) An evaluation of MODIS and SeaWiFS bio-optical algorithms in the Baltic Sea. *Remote Sens. Environ.* **89** 326–350.

DAREMA F (2004) Dynamic data driven applications systems: A new paradigm for application simulations and measurements. In: *Computational Science-ICCS 2004: 4th International Conference,* 6–9 June 2004, Krak'ow.

DAREMA F, ROTEA M, GOLDBERG M, NEWLON DH, CHERNIAVSKY JC, FIGUEROA JE, HUDSON JE, FRIEDMAN C, LYSTER P and BOHN R (2005) DDDAS: Dynamic Data Driven Application Systems. URL: http://www.nsf.gov/pubs/2005/nsf05570/nsf05570.htm (Accessed 20 September 2013).

El SERAFY GY, BLAAS M, ELEVELD MA and VAN DER WOERD HJ (2007) Data assimilation of satellite data of suspended particulate matter in Delft3D-WAQ for the North Sea. In: *Proceedings of the Joint EUMETSAT/AMS Conference,* 24–28 September 2007, Darmstadt.

FREEBORN PH, WOOSTER MJ, and ROBERTS G (2011) Addressing the spatiotemporal sampling design of MODIS to provide estimates of the fire radiative energy emitted from Africa. *Remote Sens. Environ.* **115** 475–489.

GONS HJ, RIJKEBOER M and RUDDICK KG (2002) A chlorophyll-retrieval algorithm for satellite imagery (Medium Resolution Imaging Spectrometer) of inland and coastal waters. *J. Plankton Res.* **24** (9) 947–951.

GUANG J, XUE Y, YANG LK, MEI LL and HE XW (2013) A method for retrieving land surface reflectance using MODIS data. *IEEE J. Selected Topics Appl. Earth Obs. Remote Sens.* **6** (3) 1564–1570.

HE BY, LIANG SW, DING C, YANG XQ and HU K (2009) Estimating trophic level indexes of lakes in Wuhan by using CBERS-2 IMAGE. *Changjiang Liuyu Ziyuan Yu Huanjing* **18** (12) 1181–1186.

JIANG SY, XIONG QX and ZHU JQ (2013) Evaluation of lake eutrophication based on the HJ-1 satellite multispectral data. In: *International Forum on Computer and Information Technology,* 24–25 December 2013, Shenzhen.

JIN XC , TU QY, ZHANG ZS, JIANG XC, WANG Y, ZHU X, SHU JH, XU NN, HUANG CZ, XU RX and co-authors (1990) *Specifications for Lake Eutrophication Survey* (2nd Edition). China Environment Science Press, Beijing. 291–295.

KEINER LE (1999) Estimating oceanic chlorophyll concentrations with neural networks. *Int. J. Remote Sens.* **20** (1) 189–194.

KRASNOPOLSKY VM and CHEVALLIER F (2003) Some neural network applications in environmental sciences. Part II: advancing computational efficiency of environmental numerical models. *Neural Netw.* **2003** (16) 335–348.

KRASNOPOLSKY VM and SCHILLER H (2003) Some neural network applications in environmental sciences. Part I: Forward and inverse problems in geophysical remote measurements. *Neural Netw.* **2003** (16) 321–334.

KUSTER T (2004) Quantitative detection of chlorophyll in cyanobacterial blooms by satellite remote sensing. *Limnol. Oceanogr.* **49** (6) 2179–2189.

LI Y (2007) An integrated water quality modeling system with dynamic remote sensing feedback. PhD thesis, Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology.

LI YM, HUANG JZ, WEI YC, LU WN and SHI JZ (2006) Evaluating eutrophic state of Taihu Lake by in situ hyperspectra. *Huan jing ke xue* **27** (9) 1770–1775.

MA RH, TANG JW and DAI JF (2006) Bio-optical model with optimal parameter suitable for Taihu Lake in water colour remote sensing. *Int. J. Remote Sens.* **27** (19) 4305–4328.

MEI CQ, WANG XY and LI WD (2008a) Application of BP network model in the evaluation of eutrophication in Chaohu. *Energ. Environ.* **01** 9–11.

MEI CQ, WANG XY and PENG P (2008b) Application of MODIS data for monitoring blue-green algal bloom in Chaohu Lake. *Remote Sens. Technol. Appl.* **23** (3) 328–332.

NIWA (NATIONAL INSTITUTE OF WATER AND ATMOSPHERIC RESEARCH, MINISTRY FOR THE ENVIRONMENT) (2010) URL: http://www.mfe.govt.nz/environmental-reporting/fresh-water/lake-water-quality-indicator/index.html (Accessed 15 October 2013).

MITCHELL TM (1997) *Machine Learning*. McGraw Hill, New York. 60–63.

NSCEP (NATIONAL SERVICE CENTER FOR ENVIRONMENTAL PUBLICATIONS, CORVALLIS ENVIRONMENTAL RESEARCH LABORATORY, OFFICE OF RESEARCH AND DEVELOPMENT, U.S. ENVIRONMENTAL PROTECTION AGENCY) (1976) Trophic Classification of Lakes Using LANDSAT-1 (ERTS-1) Multispectral Scanner Data. URL: http://nepis.epa.gov/Exe/ZyPDF.cgi/9100T1G3.PDF?Dockey=9100T1G3.PDF (Accessed 21 March 2015) 108–111.

SATTARI MT, APAYDIN H and OZTURK F (2012) Flow estimations for the Sohu Stream using artificial neural networks. *Environ. Earth Sci.* **2012** (66) 2031–2045.

SCHWARZ JN, KOWALCZUK P, KACZMAREK S, COTA GF, MITCHELL BG, KAHRU M, CHAVEZ FP, CUNNINGHAM A, MCKEE D, GEGE P and co-authors (2002) Two models for absorption by coloured dissolved organic matter (CDOM). *Oceanologia* **44** (2) 209–241.

SHUTLER JD, LAND PE, SMYTH TJ and GROOM SB (2007) Extending the MODIS 1 km ocean colour atmospheric correction to the MODIS 500 m bands and 500 m chlorophyll *a* estimation towards coastal and estuarine monitoring. *Remote Sens. Environ.* **107** (4) 521–532.

SONG JW, WANG XY, LIAO Y, ZHEN J, ISHWARAN N, GUO HD, YANG RX, LIU CS, CHANG C and ZONG X (2014a) An improved neural network for regional giant panda habitat suitability mapping: A case study in Ya'an Prefecture. *Sustainability* **6** (7) 4059–4076.

SONG JW, XIANG B, WANG XY, WU L and CHANG C (2014b) Application of dynamic data driven application system in environmental science. *Environ. Rev.* **22** (999) 287–297.

SONG K, LI L, Li Z, TEDESCO L, HALL B and SHI K (2013) Remote detection of cyanobacteria through phycocyanin for water supply source using three-band model. *Ecol. Inf.* **2013** (15) 22–33.

VERBURG P, HAMILL K, UNWIN M and ABELL J (2010) Lake water quality in New Zealand 2010: Status and trends. URL: http://www.mfe.govt.nz/publications/ser/lake-water-quality-in-nz-2010/ (Accessed 15 October 2013).

VERMOTE E (2013) MOD 09 - Surface Reflectance; Atmospheric Correction Algorithm Products. URL: http://modis.gsfc.nasa.gov/data/dataprod/dataproducts.php?MOD_NUMBER=09 (Accessed 15 October 2013).

VERMOTE EF and VERMEULEN A (1999) Atmospheric correction algorithm: spectral reflectances (MOD09). ATBD version, 4. URL: http://mod09val.ltdri.org/publications/atbd_mod09.pdf (Accessed 15 October 2013).

WANG MC, LIU XQ and ZHANG JH (2002) Evaluate method and classification standard on lake eutrophication. *Zhongguo Huanjing Jiance* **18** (5) 47–49.

WANG T, HUANG WJ, LIU LY, JIA JH and SHEN T (2007) Hyperspectral monitoring model of eutrophication in Poyang Lake. *Sci. Surv. Mapp.* **32** (4) 18.

WEI FS, BI T, QI WQ, SUN ZG, XI JQ, ZHANG YH, WANG ZG, CHEN G, YU XH, ZENG SN and co-authors (2002) *Monitoring and Analysis Method of Water and Waste Water* (4th edn). China Environment Science Press, Beijing. 670–671.

WU JS, ZHOU YJ and JIN L (2005) The neural network and its research development. *J. Guangxi Teachers Educ. Univ. (Nat. Sci. Ed.)* **22** (1) 92–97.

WU M, ZHANG W, WANG XJ and LUO DG (2009) Application of MODIS satellite data in monitoring water quality parameters of Chaohu Lake in China. *Environ. Monit. Assess.* **148** (1–4) 255–264.

XIE J, WANG XY, ZHANG J, LI W (2010) Analysing developing trend of chlorophyll-*a* concentration in Chaohu Lake based on TM/ETM+ image. *China Environ. Sci.* **30** (5) 677–682.

YANG DT, PAN DL, ZHANG XY, ZHANG XF, HE XQ and LI SJ (2006) Retrieval of chlorophyll a and suspended solid concentrations by hyperspectral remote sensing in Taihu Lake, China. *Chin. J. Oceanol. Limnol.* **24** (4) 428–434.

YAO YJ, ZHU L and SUN MZ (2009) Interior water body water quality remote sensing monitoring method based on the HJ-1 satellite multi-spectral data. In: *2009 International Symposium of HAIHE Basin Integrated Water and Environment Management*, 16 October 2009, Beijing.

YIN W (1989) Water quality – determination of the chemical oxygen demand – dichromate method. URL: http://www.sepa.gov.cn/image20010518/3632.pdf (Accessed 31 March 2015).

YUAN YL and YAO Y (1989) Water quality – determination of total phosphorus ammonium molybdate – Spectrophotometric method. URL: http://www.sepa.gov.cn/image20010518/3655.pdf (Accessed 31 March 2015).

ZHANG R, GAO LM, XI BD, SU J, HUO SL, JI DF, Lü NQ, and ZHU JC (2013) Improved TLI index method and its application in nutritional states evaluation in Chaohu Lake. *Huanjing Gongcheng Xuebao* **7** (6) 2127–2133.

ZHANG Y, PULLIAINEN J, KOPONEN S and HALLIKAINEN M (2002) 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** (2) 327–336.