We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



9

Monitoring of Water Quality Using Remote Sensing Data Mining

Xing-Ping Wen[1] and Xiao-Feng Yang[2]
[1]*Faculty of Land Resource Engineering, Kunming University of science and technology,*
[2]*Research Center for Analysis and Measurement, Kunming University of Science and Technology*
China

1. Introduction

Remote sensing techniques play increasingly important role over recent decades in both problems of global climate change and frequent deterioration of the status of aquatic ecology, driven by the ever-increasing needs of growing populations for drinking water, polluted by overland runoff from point and non-point sources, as well as fish and other seafood (Pozdnyakov et al., 2005). With the advent of new sensor technologies, it is possible to monitor land cover / land change in large area simultaneously and quickly (Wen and Yang, 2009a, Wen and Yang, 2009b). Remote sensing techniques have been widely used in water quality assessment (Alparslan et al., 2007, Brando and Dekker, 2003, Chen et al., 2007, Giardino et al., 2007, Hadjimitsis and Clayton, Kondratye et al., 1998, Koponen et al., 2002, Pozdnyakov et al., 2005, Ritchie et al., 2003, Seyhan and Dekker, 1986, Wang and Ma, 2001). Many documents describe water quality monitoring using different satellite sensor (Maillard and Pinheiro Santos, 2008, Giardino et al., 2007, He et al., 2008, Alparslan et al., 2007, Verma et al., 2008, Wang and Ma, 2001, Zhang et al., 2003, Martinez et al., 2007, Boken, 2007, Wang et al., 2004). The spectrum characteristics of water and pollutants are essential to water quality monitoring and assessment. The spectral characteristics of the signal received from water are a function of hydrological, biological and chemical characteristics of water, and other interference factor (Seyhan and Dekker, 1986). Suspended sediments increase the radiance emergent from surface waters in the visible and near infrared proportion of the electromagnetic spectrum(Ritchie et al., 1976), so it is promising and feasible to detect water pollutants using spectral signatures in the visible and near infrared band. Wang assessed the water quality of Taihu lake using Landsat TM imagery (Wang and Ma, 2001), and the result indicated that three visible bands of TM1, TM2 and TM3 were correlated with some water quality parameters from the lake. Alparslan (Alparslan et al., 2007) assessed water quality at Ömerli Dam using the first four bands of Landsat 7 ETM +satellite data. Hadjimitsis (Hadjimitsis and Clayton, 2009) assessed temporal variations of water quality in inland water bodies using atmospheric corrected satellite remotely sensed image data. It found that atmospheric correction was essential to water quality assessment using satellite remotely sensed imagery because it improved significantly the water reflectance. In this paper, atmospherically corrected Landsat ETM+ imagery is used to monitors water quality using stepwise multiple linear regression analysis, southwest China. Five over 30 square

kilometers lakes which are Dianchi, Fuxian, Yangzong, Qilu and Xingyun Lake near Kunming city in Yunnan province are investigated. It concludes that ETM+ 1, ETM+ 2 and ETM+ 3 are important bands to monitor water quality.

2. The study area

The study area is located at Kunming city, the capital of Yunnan provinces, in southwest China (fig. 1). Yunnan literally means "south of the clouds" in Chinese. Proud to be one of five lake-rich regions in China, Yunnan Province alone had nine lakes with areas of over 30 square kilometers. They include: Dianchi Lake, Fuxian Lake, Qilu Lake, Yangzong Lake, Xingyun Lake, Erhai Lake, Lugu Lake, Yilong Lake and Chenghai Lake. Dianchi Lake nicknamed "Sparkling Pearl Embedded in the Plateau" is a large inter-land freshwater fault lake located on the Yunnan-Guizhou Plateau close to Kunming. It covers about 300 km^2 with about 39 km long from north to south, and it is the sixth largest freshwater lake in China and the largest in Yunnan Province. However, until the first wastewater plant was built in 1990, 90 percent of Kunming's wastewater was pumped untreated into Dianchi Lake. Pollution was a major problem for the lake (Zhang et al., 1996, Liu and Zhang, 1996). According to recent reports, water quality across Yunnan's network of rivers and lakes had been deteriorating steadily over the last several years. Among them, Dianchi Lake was one of the most serious pollution lakes in Yunnan province. Although water from Lake Dianchi at one point made up 40% of the drinking water for Kunming City several years ago, nowadays the city had had to shift to other sources of water due to the lake's severe algal blooms. Dianchi Lake is rated grade V (the worst grade) which makes the water unfit for agricultural or industrial uses.

In this paper, five over 30 square kilometers lakes near Kunming city in Yunnan province were investigated. There are Dianchi, Fuxian, Yangzong, Qilu and Xingyun Lake (fig. 2).



Fig. 1. China provincial boundaries and the study area.

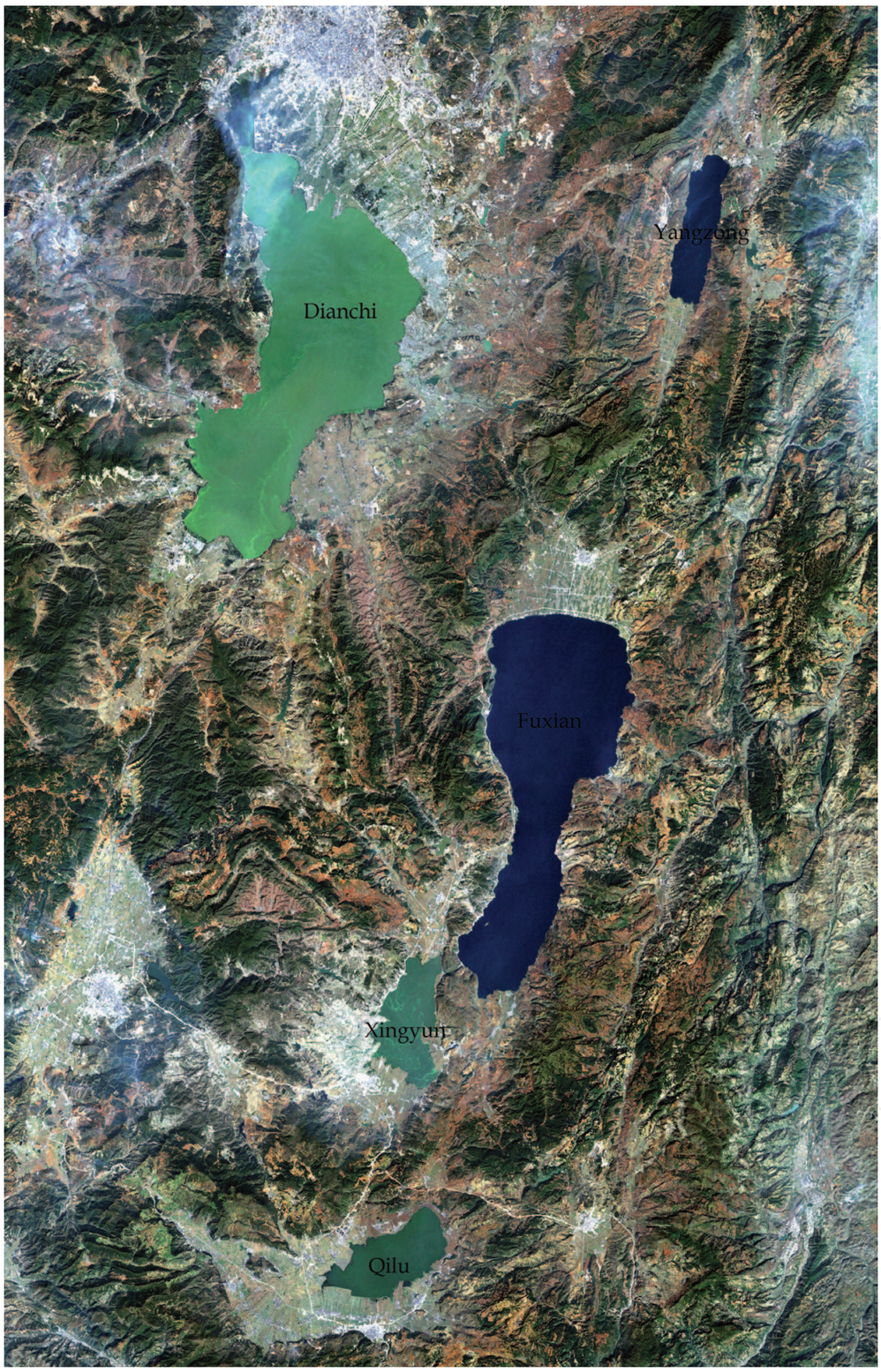


Fig. 2. The color composite ETM+ image of band 3 (Red), band 2 (Green) and band 1 (Blue) after atmospheric correction acquired in November 2, 2000.

3. Remote sensing data

3.1 Radiometric and atmospheric correction of Landsat ETM+ imagery

This paper uses Landsat ETM+ imagery acquired in November 2, 2000 to monitor water quality near Kunming city. The Landsat 7 satellite was successfully launched from Vandenburg Air Force Base on April 15, 1999. It is a near polar-orbiting, earth mapping orbit with a 16-day repeat cycle, sun-synchronous satellite at an altitude of 705 km above the Earth. Its payload is a single nadir-pointing instrument, the Enhanced Thematic Mapper Plus (ETM+). The ETM+ sensor provides for a nadir-viewing, eight-band multispectral scanning radiometer capable of providing high-resolution image information of the Earth's surface. It detects spectrally-filtered radiation in VNIR, SWIR, LWIR and panchromatic bands from the sunlit Earth in a 185 km wide swath. In order to retrieve the ground object spectra, the radiometric and atmospheric correction is the necessary step. The following equation referred to NASA website is used to convert digital number (DN) to radiance units:

$$L_\lambda = \text{Grescale} * \text{QCAL} + \text{Brescale} \quad (1)$$

where:

L_λ = Spectral radiance at the sensor's aperture in watts/ (meter squared * ster * μm)

Grescale = Rescaled gain (the data product "gain" contained in the Level 1 product header or ancillary data record) in watts/(meter squared * ster * μm)/DN

Brescale = Rescaled bias (the data product "offset" contained in the Level 1 product header or ancillary data record) in watts/(meter squared * ster * μm)

QCAL = the quantized calibrated pixel value in DN

The imagery is atmospherically corrected using FLAASH model (Cooley et al., 2002). Fig. 3 is the comparison of four different target curves before and after atmospheric correction. Spectral reflectance curves of four different targets are similar to spectral reflectance curves of corresponging standard targets after atmospheric correction. The atmospherically corrected image is shown in fig. 2

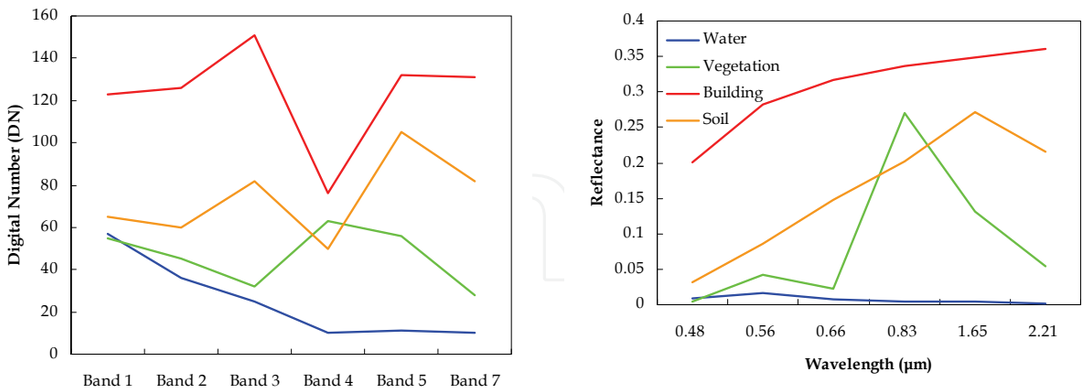


Fig. 3. Digital number curves of four different targets before atmospheric correction (left). Spectral reflectance curves of corresponding four different targets after atmospheric correction (right).

3.2 The classification of image and identification of five lakes

Five lakes are identified from the ETM+ image using classification method for avoiding effects of other pixels. Firstly, the Maximum Likelihood Classification (MLC) method

(Richards, 1999) is applied to extract water areas from the imagery. Then, the classification result is covert into vector file and polygon areas lower 30 square kilometers are removed using Geographic Information System (GIS) software. Finally, the ETM+ imagery is applied mask using the processed vector file, and only five lakes image was obtained (fig. 4).



Fig. 4. Five lakes in the study area.

4. Methodology

4.1 Optimum index factor of lakes image

Statistics is the science of making effective use of numerical data. In this paper, mean, standard deviation and correlation coefficient of reflectance values from lakes are calculated (table 1, table 2). Due to the influence of algas in lakes, the mean value of band 4 is higher. The standard deviation is widely used to measure the variability or dispersion. A low standard deviation indicates that the data value tend to be very close to the mean, whereas the high standard deviation indicates that the data is spread out over a large range of values. The band with the higher standard deviation contains the higher amount of

'information' than other bands. The correlation coefficient of different bands represents duplication among band pairs. The Optimum Index Factor (OIF) developed by (Chavez et al., 1982) is a statistic value that can be used to select the optimum combination of three bands in a satellite image with which a color composite image is created. The optimum combination is the one with the highest sum of standard deviations and the least amount of duplication (lowest correlation coefficient).

Statistic value	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
Mean	0.02238	0.05448	0.03518	0.05247	0.01207	0.00649
Standard deviation	0.01360	0.03064	0.02159	0.09680	0.02020	0.00845

Table 1. Mean and standard deviation of pixel values from lakes in different bands

Correlation	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
Band 1	1.000000	0.727497	0.738473	0.682501	0.627298	0.496868
Band 2	0.727497	1.000000	0.956101	0.540135	0.448750	0.368707
Band 3	0.738473	0.956101	1.000000	0.523605	0.454983	0.392675
Band 4	0.682501	0.540135	0.523605	1.000000	0.933209	0.689372
Band 5	0.627298	0.448750	0.454983	0.933209	1.000000	0.793760
Band 7	0.496868	0.368707	0.392675	0.689372	0.793760	1.000000

Table 2. Correlation coefficient of different bands

OIF is calculated using the formula as follows:

$$OIF = \frac{Std_i + Std_j + Std_k}{\left|Cor_{i,j}\right| + \left|Cor_{i,k}\right| + \left|Cor_{j,k}\right|} \quad (2)$$

OIF = The optimum combination of three bands.
Std_i = The standard deviation of band i.
$|Cor_{i,j}|$ = The absolute value of the correlation coefficient of band i and band j.
Firstly, the number of possible combinations of three bands is determined, for 6 bands there are 20 combinations. Then, for each combination of three bands, the OIF is calculated. Finally, the OIF values are ranked and band 7, band 4 and band 3 with the highest OIF is the optimum combination. Band 4 and band 3 usually used to calculate Normalized Difference Vegetation Index (NDVI) indicates live green plants in lakes. The color composite image of band 7, band4 and band 3 is shown in fig. 5. Comparing with the fig. 2, it provides a contrast to different water quality.

4.2 Spectra of different water quality

In order to compare spectra of different water quality, typical water pixels are selected from five lakes respectively using regions of interest tools (ROIs) and the average spectra of different water quality from five lakes are calculate and illustrated (fig. 6). As shown from fig. 6, the reflectance value of band 2, band 3 and band 1 change dramaticlly according to different water quality, which are relate to the water pollutant of different lakes.



Fig. 5. The color composite image of band 7 (Red), band 4 (Green) and band 3 (Blue).

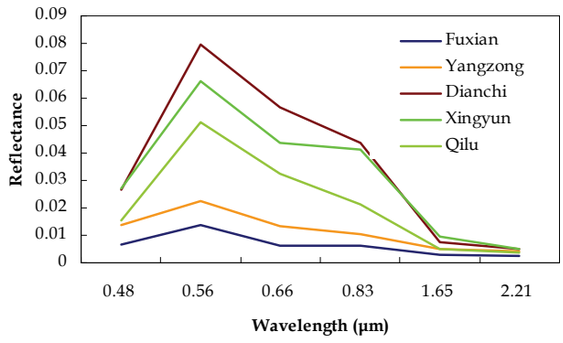


Fig. 6. Average spectra of different water quality from five lakes.

4.3 Stepwise multiple linear regression analysis of water quality level

Water quality is physical, chemical and biological characteristics of water. Water quality standards vary significantly due to different environmental conditions, ecosystems, and intended human uses. Due to the complexity of water quality, there are many types of measurements of water quality indicators. In this paper, according to spectra of different water quality, the water quality is classfied into six level. Typical sample training data are selected based on their spectra from five lakes using regions of interest tools (ROIs). Average spectra of training data in different level are shown in fig. 7. The sixth level is the

average spectrum of alga in lakes. As shown from fig. 7, with the increase of pollutant, the reflectance of bands increases. The average spectrum of level 1 is similar to the spectrum of distilled water.

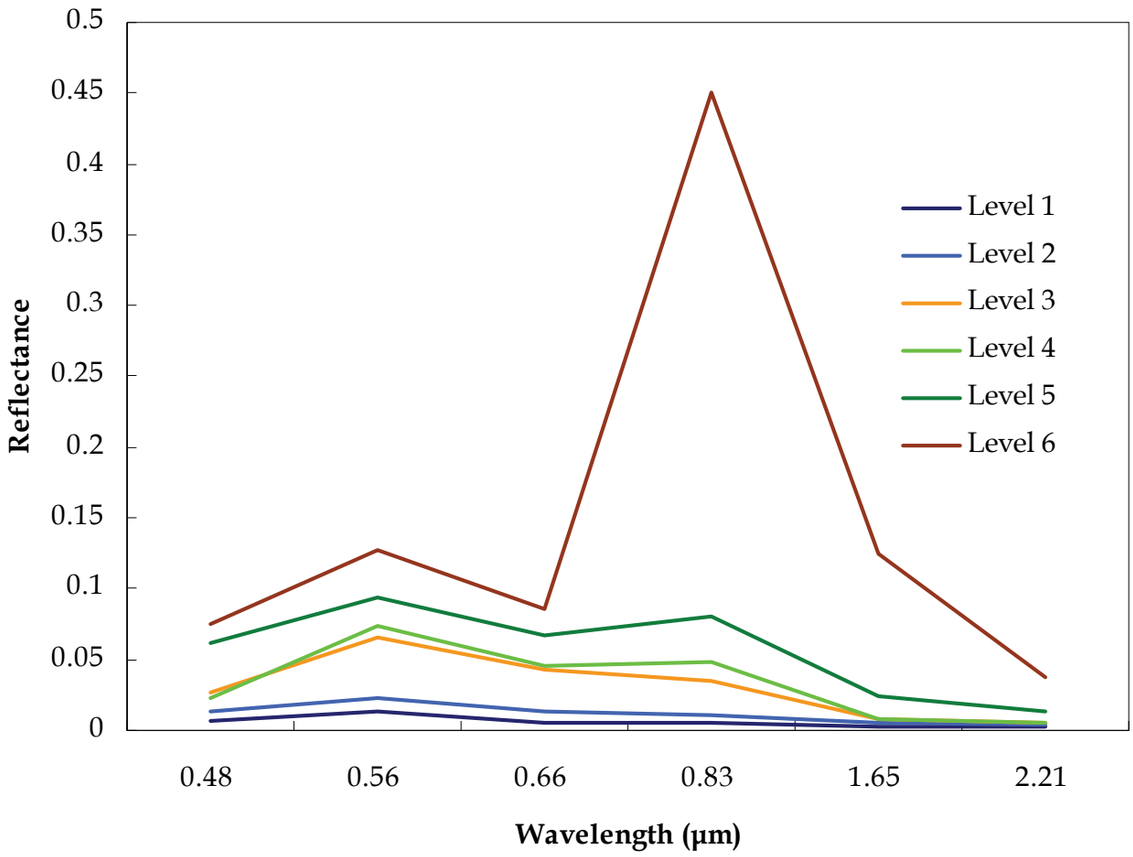


Fig. 7. Average spectra of sample training data in different water quality level.

The relationship between water quality and the reflectance is analyzed using stepwise multiple linear regression analysis (Hocking, 1976). Water quality level fitting formula using all sample training data is as follows:

$$\text{Water quality level} = \rho_1 * 7.26 + \rho_2 * 25.75 + \rho_3 * 17.14 + 0.80 \quad (3)$$

Where:

ρ_i = The reflectance of ETM+ band *i*.

In the automatic procedure of stepwise multiple linear regression analysis, the important variables are band 2, band 3, band 1, band 7, band 5 and band 4 in sequence. The variance contributions of band 7, band 5 and band4 are less, so only band 1, band 2 and band3 are preserved in the formula. The correlation coefficient between sample data outcomes and predictive values calculated by formula (3) is 0.98. Therefore, the formula (3) is used to assess water quality. Firstly, the reflectance of bands is calculated using band math algorithm and the grey sum image is obtained. Secondly, the density slice method is applied to the processed grey imagery and the water quality level image is outputed. Finally, the result image is projected and outputed (fig. 8).

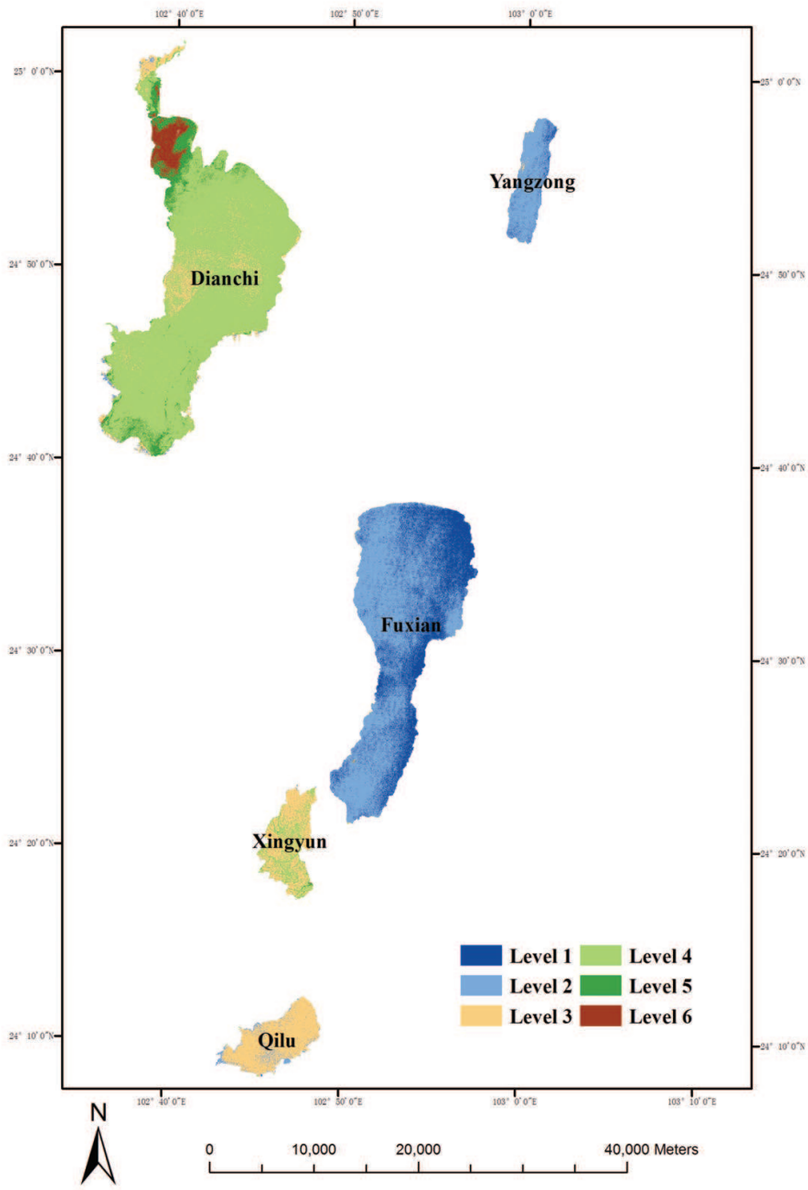


Fig. 8. The water quality monitoring image of lakes.

5. Results and discussions

In this paper, Landsat ETM+ imagery was used to monitors water quality of lakes near Kunming city in the southwest China. The water quality is distinguished with six levels from no pollution to contamination seriously. According to the stepwise multiple linear regression analysis formula, ETM+ band 1, band 2 and band 3 are close correlated with water quality. As shown from fig. 8, the water quality of Fuxian and Yangzong lakes are the best, and Dianchi and Xingyun lake are the worst, Qilu lake is in the middle. They are consistent with average spectra of five lakes water. The water quality level of Dianchi and Xingyun lakes are greater than 5, which has become highly polluted and eutrophic with

serious algae problems. The pollution of Qilu lake is to some degree. Suitable water pollution controls and provention for lakes will be needed.

6. Acknowledgment

This study was jointly supported by the NSFC of Yunnan province, China (KKSA200921019), Scientific Research Foundation of Kunming University of Science and Technology (KKZ3200821048), China Postdoctoral Science Foundation (20100471687) and the innovation team of ore-forming dynamics and prediction of concealed deposits, Kunming University of Science and Technology, Kunming, China.

7. References

Alparslan, E., Aydöner, C., Tufekci, V. & Tüfekci, H. (2007). Water quality assessment at Ömerli Dam using remote sensing techniques. *Environmental Monitoring and Assessment,* Vol. 135, No. 1, 391-398.

Boken, V. K. (2007). Linking landuse and groundwater quality in the Mississippi delta Using MODIS satellite data. *IEEE International Geoscience and Remote Sensing Symposium*, 5025-5027,

Brando, V. E. & Dekker, A. G. (2003). Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality. *Geoscience and Remote Sensing, IEEE Transactions on,* Vol. 41, No. 6, 1378-1387.

Chavez, P. S.; Berlin, G. L. & Sowers, L. B. (1982). Statistical method for selecting Landsat MSS ratios. *Journal of Applied Photographic Engineering,* Vol. 8, No. 1, 22-30.

Chen, Q.; Zhang, Y. & Hallikainen, M. (2007). Water quality monitoring using remote sensing in support of the EU water framework directive (WFD): A case study in the Gulf of Finland. *Environmental Monitoring and Assessment,* Vol. 124, No. 1, 157-166.

Cooley, T., Anderson, G. P., Felde, G. W., Hoke, M. L., Ratkowski, A. J., Chetwynd, J. H., Gardner, J. A., Adler-Golden, S. M., Matthew, M. W., Berk, A., Bernstein, L. S., Acharya, P. K., Miller, D. & Lewis, P. (2002). FLAASH, a MODTRAN4-based atmospheric correction algorithm, its application and validation. *Geoscience and Remote Sensing Symposium*, 1414-1418,

Giardino, C., Brando, V. E., Dekker, A. G., Strömbeck, N. & Candiani, G. (2007). Assessment of water quality in Lake Garda (Italy) using Hyperion. *Remote Sensing of Environment,* Vol. 109, No. 2, 183-195.

Hadjimitsis, D. & Clayton, C. (2009). Assessment of temporal variations of water quality in inland water bodies using atmospheric corrected satellite remotely sensed image data. *Environmental Monitoring and Assessment,* Vol. 159, No. 1-4, 281-292.

He, W., Chen, S., Liu, X. & Chen, J. (2008). Water quality monitoring in a slightly-polluted inland water body through remote sensing — Case study of the Guanting Reservoir in Beijing, China. *Frontiers of Environmental Science & Engineering in China,* Vol. 2, No. 2, 163-171.

Hocking, R. R. (1976). The Analysis and Selection of Variables in Linear Regression. *Biometrics,* Vol. 32, No. 1, 1-49.

Kondratye, K. Y.; Pozdnyakov, D. V. & Pettersson, L. H. (1998). Water quality remote sensing in the visible spectrum. *International Journal of Remote Sensing,* Vol. 19, No. 5, 957-979.

Koponen, S., Pulliainen, J., Kallio, K. & Hallikainen, M. (2002). Lake water quality classification with airborne hyperspectral spectrometer and simulated MERIS data. *Remote Sensing of Environment,* Vol. 79, No. 1, 51-59.

Liu, J. Q. & Zhang, Y. X. (1996). Distribution and degradation of Alkyl benzene sodium sulfonate, and its harmful effect on carp in Dianchi Lake. *GeoJournal,* Vol. 40, No. 1, 219-227.

Maillard, P. & Pinheiro Santos, N. A. (2008). A spatial-statistical approach for modeling the effect of non-point source pollution on different water quality parameters in the Velhas river watershed - Brazil. *Journal of Environmental Management,* Vol. 86, No. 1, 158-170.

Martinez, J. M., Guyot, J. L., Cochonneau, G. & Seyler, F. (2007). Surface water quality monitoring in large rivers with MODIS data application to the amazon basin. *IEEE International Geoscience and Remote Sensing Symposium*, 4566-4569,

Pozdnyakov, D., Shuchman, R., Korosov, A. & Hatt, C. (2005). Operational algorithm for the retrieval of water quality in the Great Lakes. *Remote Sensing of Environment,* Vol. 97, No. 3, 352-370.

Richards, J. A. (1999). *Remote Sensing Digital Image Analysis,* Springer-Verlag. Berlin

Ritchie, J.; Zimba, P. & Everitt, J. (2003). Remote sensing techniques to assess water quality. *Photogrammetric Engineering and Remote Sensing,* Vol. 69, No. 6, 695-704.

Ritchie, J. C.; Schiebe, F. R. & McHenry, J. R. (1976). Remote sensing of suspended sediment in surface water. *Photogrammetric Engineering & Remote Sensing,* Vol. 42, No. 1539-1545.

Seyhan, E. & Dekker, A. (1986). Application of remote sensing techniques for water quality monitoring. *Aquatic Ecology,* Vol. 20, No. 1, 41-50.

Verma, V., Setia, R., Sharma, P. & Singh, H. (2008). Geoinformatics as a tool for the assessment of the impact of ground water quality for irrigation on soil health. *Journal of the Indian Society of Remote Sensing,* Vol. 36, No. 3, 273-281.

Wang, X. J. & Ma, T. (2001). Application of Remote Sensing Techniques in Monitoring and Assessing the Water Quality of Taihu Lake. *Bulletin of Environmental Contamination and Toxicology,* Vol. 67, No. 6, 863-870.

Wang, Y., Xia, H., Fu, J. & Sheng, G. (2004). Water quality change in reservoirs of Shenzhen, China: detection using LANDSAT/TM data. *Science of the Total Environment,* Vol. 328, No. 1, 195-206.

Wen, X. & Yang, X. (2009a). Change detection from remote sensing imageries using spectral change vector analysis. *Asia-Pacific Conference on Information Processing (APCIP 2009)*, 189-192, IEEE Computer Society, Shenzhen, China

Wen, X. & Yang, X. (2009b). A new change detection method for two remote sensing images based on spectral matching. *International Conference on Industrial Mechatronics and Automation (ICIMA 2009)*, 89-92, IEEE Computer Society, Chengdu, China

Zhang, X., Zhang, S., Ying, W., Ren, T., Xu, C., Zhong, Z. & Zhang, S. (1996). Heavy metals pollution on the sediments in lakes Dianchi, Erhai and Poyanghu and historical records. *GeoJournal,* Vol. 40, No. 1, 201-208.

Zhang, Y. Z., Pulliainen, J. T., Koponen, S. S. & Hallikainen, M. T. (2003). Water quality retrievals from combined Landsat TM data and ERS-2 SAR data, in the Gulf of Finland. *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 41, No. 3, 622-629.

IntechOpen

IntechOpen



Knowledge-Oriented Applications in Data Mining

Edited by Prof. Kimito Funatsu

ISBN 978-953-307-154-1
Hard cover, 442 pages
Publisher InTech
Published online 21, January, 2011
Published in print edition January, 2011

The progress of data mining technology and large public popularity establish a need for a comprehensive text on the subject. The series of books entitled by 'Data Mining' address the need by presenting in-depth description of novel mining algorithms and many useful applications. In addition to understanding each section deeply, the two books present useful hints and strategies to solving problems in the following chapters. The contributing authors have highlighted many future research directions that will foster multi-disciplinary collaborations and hence will lead to significant development in the field of data mining.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Xing-Ping Wen and Xiao-Feng Yang (2011). Monitoring of Water Quality Using Remote Sensing Data Mining, Knowledge-Oriented Applications in Data Mining, Prof. Kimito Funatsu (Ed.), ISBN: 978-953-307-154-1, InTech, Available from: http://www.intechopen.com/books/knowledge-oriented-applications-in-data-mining/monitoring-of-water-quality-using-remote-sensing-data-mining

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen