Journal of Applied Sciences

ISSN 1812-5654

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 13 (10): 1767-1773, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.1767.1773

Online Evaluation of Water Quality by Mid-infrared Spectroscopy in River Network Areas of Suzhou City

[1,2,3]Ligang Fang, [1,3]Jinxiang Li, [3]Zhaobin Liu, [1]Changbo Tang, [4]Zhu Liang
[1]Jiangsu Province Support Software Engineering Center for Modern Information Technology Application in Enterprise, Suzhou, China
[2]Guangdong Open Laboratory of Geospatial Information Technology and Application, Guangzhou, China
[3]Department of computer engineering, Suzhou vocational university, Suzhou, China
[4]Suzhou Environmental Monitor Centre, Suzhou Environmental Protection Bureau, Suzhou, China

Abstract: Fifty typical stations were selected to measure mid-infrared spectroscopy in the river network areas of Suzhou city, on November 23, 2009 and five water quality parameters were analyzed for every measured station, including Total Phosphorous (TP), Total Nitrogen (TN), K, Mg and Cu. The correlations between the water quality parameters and spectral reflectance were analyzed. The single band, multiple linear regression and synergy interval partial least squares (SiPLS) models were built based on reflectance. The results show that, for TN and TP, the SiPLS models have the smallest RMSE and TN and TP can be detected efficiently by in situ mid-infrared spectral analysis ($r>0.76$). For the prediction of K, Mg, the results indicate that the optimal method is with the single band models of the reflectance and the concentration derived from the single band models are strongly correlated with the measured concentration ($r>0.79$). The study indicated that the water quality parameters can be monitored by mid-infrared spectroscopy technology, which would provide the water supply and conservancy authorities with referenced spatial information to manage water resources.

Key words: Mid-infrared spectra, water quality parameters, online evaluation, Suzhou

INTRODUCTION

In recent years, a large bloom of blue-green algae in Taihu Lake caused water quality to deteriorate severely and dissolved oxygen in water to decline, which eventually resulted in disaster of plankton and aquatic biological (Qian and He, 2009). A large bloom of blue-green algae also can produced all kinds of natural poison, such as microcystin, which can intrude into human body by drinking water or aquatic product contaminated with serious threat to human health.

When blue-green algae bloom, nutritive cell will transform fixed N_2 to protein for photosynthesis. Blue-green algae will also absorb P for bloom of cell. Therefore, a large bloom of blue-green algae is not only closely related to ecological environment of water (such as type and quantity of microorganisms, sunshine and temperature) but also organic elements (such as N_2, NH_3 and P) (Watanabe, 1994). Large amounts of N and P fertilizer in land eco-systems and phosphoric food additives in Industry were discharged out at Taihu Lake and adjoining river. Eutrophication of lake was gradually aggravated year after year and the content of N, P and C increased enormously in the lake deposit, which resulted to bloom of blue-green algae.

Research indicates that TP and TN concentrations have high positive correlation with blue-green algae. So, the contents of TP and TN are important parameters for predication of blue-green algae. Besides, some organic elements and minor elements also affect bloom of blue-green algae. However, measures for water indexes monitoring and prewarning were based on field detection in Suzhou and Wuxi city, which are high cost, low efficiency and precision.

At present, mid-infrared spectroscopy has been used in some studies about chemical composition. For example, Kang J. etc. developed a new simultaneous analytical method for the quantitative determination of total phenols, titratable acidity, pH, alcohol of wine using Attenuated Total Reflection (ATR) FTIR spectroscopy combined with the partial least square (PLS) algorithm (Kang *et al.*, 2010). Lin F. etc. suggested that FTIR spectroscopy may be considered as a diagnosis technology for leaf nitrogen content in rice (Lin *et al.*,

Corresponding Author: Ligang Fang, Jiangsu Province Support Software Engineering Center for Modern Information Technology Application in Enterprise, Suzhou, China
Tel: 86-0512-66507429 Fax: 86-0512-66503531

2009). In addition, Mid-infrared spectroscopy combined with chemometrics was used to determine total sugar content in Chinese rice wine (Shen *et al.*, 2009) and mid-infrared spectroscopy to assess the authenticity of tea varieties (Feng *et al.*, 2007). However, there is less study about water quality indexes based on mid-infrared spectroscopy.

STUDY METHOD

Sample and spectra measurement: The field measurements and collections of water samples were carried out from 0930 to 1430 on December 23, 2009 in river network areas of Suzhou city. Spectra, Total Phosphorous (TP), Total Nitrogen (TN) sampled in stations including 50 locations were measured and analyzed and K, Mg and Cu were sampled in 10 of all locations. The accurate positions of all the sampling stations were recorded by the global positioning system, as shown in Fig. 1.

Online sample and spectral measurement equipment consists of water sample collection and measurement pool, as shown in Fig. 2. Water sample collection device is composed of hose Connection, wriggle pump and filter membrane. The work procedure of the equipment is as follows: water sample is extracted from water supply under test by wriggle pump (type: BT100FJ, maximum flow: 380 mL min^{-1}) and filtered by filter membrane of 0.45 μm which can remove the interference of suspended particle and aquatic residues in spectral analysis, finally, the water sample treated is transferred to the measurement pool and measured by infrared spectrometer. Spectral measurements of water samples are carried out using InfraSpec VFA-IR Spectrometers manufactured by Carl Zeiss Company in German. The instrument operates in the bands from 4850 to 10570 nm and there are 128 bands within the range. 10 spectral curves are continuously measured at a time and averaged. The samples are collected in the bottle for water quality analysis laboratory after spectral measurement. In the equipment, automatic cleaning device is used to clean measurement pool at regular intervals, which can eliminate impurity attached to the measurement pool and improve the accuracy of spectral measurement.

Analyses of water samples: We collected water samples after spectra measurement using Infra Spec VFA-IR, all

Fig. 1: Map of sampling position of the river network areas in the SuZhou city, china. Spectra, Total Phosphorous (TP), Total Nitrogen (TN) sampled in stations including 50 locations were measured and analyzed and K, Mg and Cu were sampled in 10 of all locations on December 23 in 2009

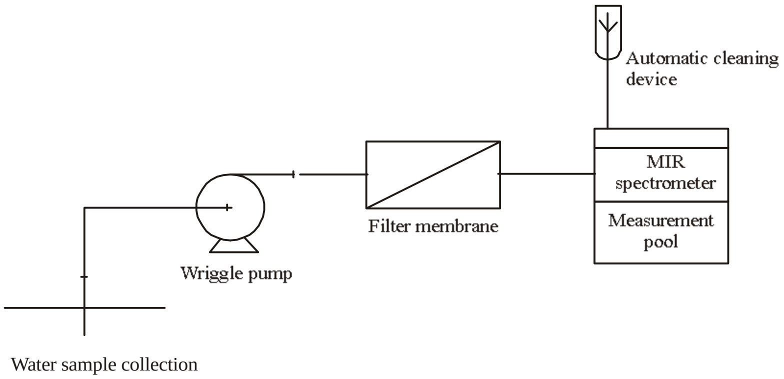

Fig. 2: Online sample and spectral measurement equipment

the samples were kept in an icebox and transported to water quality analysis laboratory in Suzhou as soon as possible in polypropylene sampling bottles. The entire procedure, including sample collection, storing and measurement, was performed strictly according to the Technical Specifications Requirements for Monitoring of Surface Water and Waste Water of China. Total Phosphorous (TP), Total Nitrogen (TN), K, Mg and Cu sampled in stations including 50 locations were measured and analyzed on December 23 in 2009.

Total Phosphorous and Total Nitrogen were measured by the continuous flow analyzer (SAN^{++}) which is manufactured by SKALAR Company in Netherlands. TN and TP were measured according to the technical specifications requirements (Kroon, 1993; ISO/DIS 15681, 2003).

Statistical analysis: Multiple Linear Regression (MLR) is a commonly used calibration algorithm which is simple and easy to interpret. However it fails when variables are more than samples and is easily affected by the collinearity between the variables (Naes and Mevik, 2001). In this study the variable number of full visible and near infrared spectra was larger than samples. Therefore it was not possible to run MLR directly and the effective variable selection was necessary before MLR model establishment. Selected variables with less collinearity would be helpful to improve the MLR model. The optimal band combination was determined by the lowest value of predicted residual error sum of squares. The performance of the calibration model was evaluated in terms of the correlation coefficients (r) and root mean square error (RMSE). The best model should be the one with the smallest RMSE and the largest r.

Partial Least Squares (PLS) analysis is widely used for calibration in current chemometric analysis. PLS is performed to establish a regression model to predict physiological concentrations. PLS finds the fundamental relations between the variable matrix Y (the properties of interest) and the variable matrix X (the spectra). PLS is particularly suited when the number of variables is greater than the samples and when there is multicollinearity among X values. Synergy interval PLS (SiPLS) splits the data set into a number of intervals (variable-wise) and calculates all possible PLS model combinations of two, three or four intervals, which is used to establish the optimal models(Zhang *et al*., 2005; Chen *et al*., 2008; Norgaard and Leardi 2005). RMSE and r are used to evaluate precision of the models:

$$\mathrm{RMSE} = \frac{\sum_{i=1}^{n}\left(x_i - y_i\right)^2}{n} \tag{1}$$

where, x_i: actual value (mg L^{-1}); y_i: predicted value (mg L^{-1}); n: number of samples in the calibration set.

DATA ANALYSIS AND DISCUSSION

Correlation analysis between the water quality parameters and reflectance: The water quality contents, TN, TP, Cu, Mg and K, in the water samples of Suzhou river were correlated with spectral data. The wavelength that was most strongly correlated with the selected water quality indexes was selected as the optimal one.

In-situ TN and TP of all locations were highly variable due to serious pollution of river network areas in Suzhou city, as is shown in Table 1. TN concentrations varied between 4.72 and 7.75 mg L^{-1} and TP between 0.09 and 0.45 mg L^{-1}.

The descriptive statistic and analysis of variance between reflectance and water constituents were performed in SPSS14.0 software. From the Table 1, there were higher correlation between TN and other water quality indexes (r>0.5) and high correlation between TP and Mg with the correlation coefficient of -0.851. In addition, K and Cu was significantly correlation with the correlation coefficient of 0.962.

Figure 3 showed the correlation of TN as a function of the wavelength. Maximum of correlation coefficient was

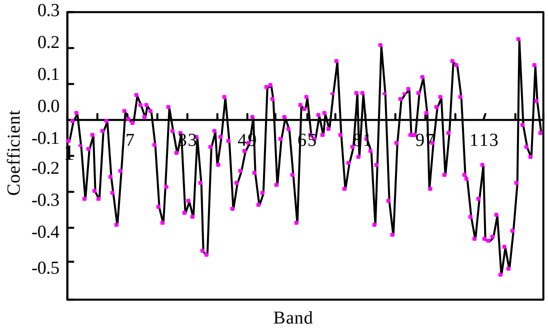

Fig. 3: The correlation between TN and the single band reflectance

Table 1: The descriptive statistics and correlation between K, Mg, Cu, TN and TP

	Descriptive statistics			Correlation efficient				
Parameter	Mean	Max	Min	K (mg L^{-1})	Mg (mg L^{-1})	Cu (mg L^{-1})	TN (mg L^{-1})	TP (mg L^{-1})
K (mg L^{-1})	22.02	24.3	20.1	1	0.013	0.962	0.573	0.437
Mg (mg L^{-1})	8.85	10.1	8.39		1	0.257	0.769	-0.851
Cu (mg L^{-1})	0.003	0.016	0.001			1	0.7	0.242
TN (mg L^{-1})	5.53	7.75	4.72				1	-0.473
TP (mg L^{-1})	0.13	0.45	0.09					1

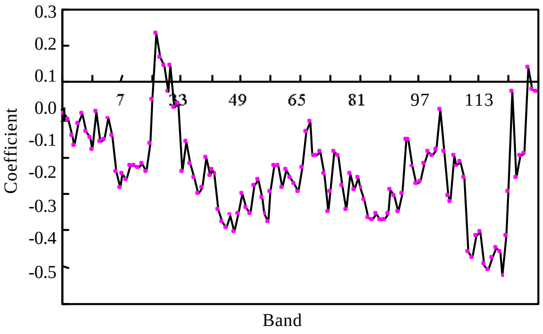

Fig. 4: The correlation between TP and the single band reflectance

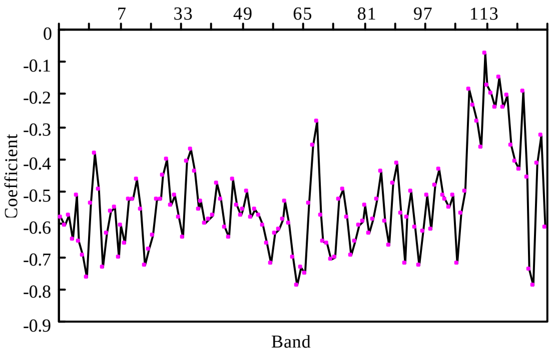

Fig. 5: The correlation between K and the single band reflectance

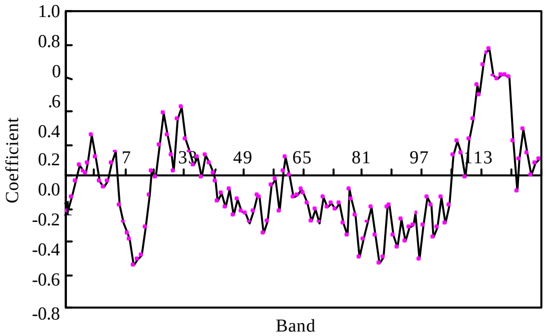

Fig. 6: The correlation between Mg and the single band reflectance

in the 114th band (r = 0.54). Moreover, TN was also better correlated to the reflectance in the second, 6th, 16th, 24th, 39th, 61st and 112th to 118th band (r>0.4).

From Fig. 4, negative correlations were found between the spectra reflectance and TP concentration. There were better correlation between TP and the single band reflectance in the 45th, 47th and 110th to 120th band (r>-0.4) and maximum coefficient, -0.53, emerged in the 119th band.

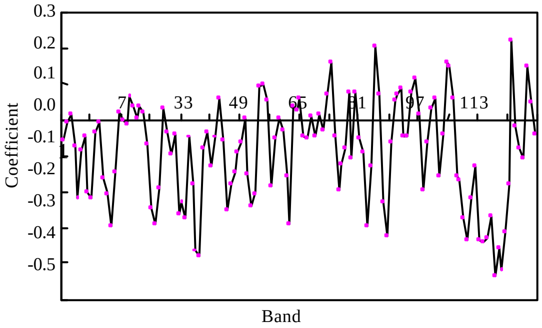

Fig. 7: The correlation between Cu and the single band reflectance

Significantly negative correlations were found between the spectral reflectance and K concentration. Fig. 5 showed correlation coefficient between K and most bands were high than -0.5 and maximum coefficient, -0.79, emerged in the 63rd and 125th band. There was also better correlation between Mg and the single band reflectance (Fig. 6). Positive maximum of correlation coefficient was in the 115th band (r = 0.77) and negative maximum was in the 19th (r = -0.54). However, Cu index had the bad correlation coefficient of less than 0.43 between the water quality parameter and the spectra reflectance.

In other words, the correlation coefficients between the single band reflectance and water constituents were higher and their maximum is more than 0.53 except Cu index.

Multiple linear regression analysis: The effective variable selection was necessary before MLR model establishment. Selected variables with less collinearity would be helpful to improve the MLR model. The sensitive wavelengths reflecting the characteristics of spectra for quality parameters were obtained based on correlation coefficients. After correlation coefficients were sorted between the quality parameters and reflectance (Fig. 3-6), the band with optimal correlation coefficient was selected first and suboptimum band was selected in turn in sensitive wavelengths analysis. In addition, the only peak band was selected in a interval with strong correlation, which was used to overcome the problem of collinearity encountered with linear multivariate least squares regression models. According to existing experience rules, the ratio of the number of samples and variables is greater than or equal to 5 in MLR analysis (Shao *et al*., 2011). So, for TN and TP the number of variables is not beyond 10 in modeling set of 50 samples and for K, Mg and Cu the number of variables is not beyond 2. The selection results of sensitive wavelengths are summarized in Table 2.

From Table 2, Cu and TN have same sensitive band in the 117th band and Cu and TP have also same sensitive band in the 119th band. Mg and TP has same sensitive band in the 115th band and similar sensitive band near 118th band, which shows that there is a relationship between Mg and TP.

The study makes full use of the advantages of MLR which is simple and easily interpreted and avoids the disadvantages of MLR. The selected variables were set as the inputs of MLR. The results are shown in Table 3.

Synergy interval partial least squares analysis: Use SiPLS to calculate all PLS models on all combinations of 2, 3 or 4 intervals. List the results by siplstable function. Next, remake PLS models on selected interval combinations with plsmodel function (you can use this function for any combination of intervals, i.e. also more than four). Study the results by plsrmse and plspvsm functions. The results are presented in the tables.

From Table 4, when the wavelength is divided into 26 intervals, RMSE of the model is smallest (0.3326), which was built by the combination of 12th, 20th and 23th intervals and the correlation coefficient of measured and predicted values (r) was 0.82.

Table 2: The selection results of sensitive bands based on correlation coefficients

Quality indexes	Sensitive bands
TN	2, 6, 16, 24, 39, 61, 114, 117...
TP	45, 47, 56, 84, 87, 111, 115, 119...
K	63, 125, 8, 65...
Mg	115, 118, 111, 19...
Cu	117, 119, 38, 114...

Table 3: The calibration results of TN, TP, K, Mg and Cu based on MLR with selected variables

Quality indexes	LVs	r	RMSE (mg L^{-1})
TN	5	0.73	0.349
TP	2	0.58	0.069
K	2	0.82	0.943
Mg	2	0.91	0.242
Cu	1	0.43	0.004

Lvs: The number of latent variables

Table 4: Part of the results of TN based on SiPLS

The No. of intervals	The No. of principal component	Intervals selected	RMSE (mg L^{-1})
10	2	(5 10)	0.3856
16	2	(8 15)	0.3862
26	5	(12 20 23)	0.3326
64	3	(31 45 57)	0.3548

Table 5: Part of the results of TP based on SiPLS

The No. of intervals	The No. of principal component	Intervals selected	RMSE (mg L^{-1})
10	3	(2 9)	0.0653
16	4	(3 15)	0.06476
26	4	(4 23 25)	0.06203
64	6	(8 56 63)	0.06047

From Table 5, RMSE of the model is smallest (0.06047), which was built by the combination of 8th, 56th and 63th intervals and the correlation coefficient of measured and predicted values (r) was 0.76, when the wavelength is divided into 64 intervals.

MODELS AND DISCUSSION

The optimal models for water quality evaluation: Single band regression analyses using the reflectance indicate that the wavelength at 114th band is the most suitable for estimating TN concentration in river network areas of Suzhou city. MLR model analyse indicates 5 LVs are best selection for estimating TN concentration and the combination of 12th, 20th and 23th intervals is optimal for SiPLS model when the wavelength is divided into 26 intervals. Compared with the other models (Table 6), the SiPLS model has the smallest RMSE (0.33) and the TN concentration derived from this model is strongly correlated with the measured TN concentration (r = 0.82).

For the prediction of TP, from Table 7, the SiPLS model has the smallest RMSE (0.060) and the TP concentration derived from this model is strongly correlated with the measured TP concentration (r = 0.76).

From Table 8, the band with the maximum correlation is 63th, with RMSE = 0.849. So, the optimal bands were determined from the absolute maximal RMSE. The results indicate that the optimal method of prediction for K is with the single band models of the reflectance. The results show that it is possible to predict K concentration of the surface water from in situ spectral measurements.

Magnesium was predicted using reflectance of 115th band with a quadratic polynomial model (Table 9), which has the smallest RMSE (0.229) and the Mg concentration

Table 6: Summary of the three regression models between TN and reflectance

Model	r	RMSE (mg L^{-1})
Single band		
Linear	0.54	0.400
Exponetial	0.54	0.400
Logarithmic	0.55	0.400
Quadratic polynomial	0.59	0.400
MLR	0.73	0.349
SiPLS	0.82	0.332

Table 7: Summary of the three regression models between TP and reflectance

Model	r	RMSE (mg L^{-1})
Single band		
Linear	0.53	0.069
Exponetial	0.53	0.069
Logarithmic	0.58	0.069
Quadratic polynomial	0.60	0.069
MLR	0.58	0.069
SiPLS	0.76	0.060

Table 8: Summary of the three regression models between K and reflectance

Model	r	RMSE (mg L^{-1})
Single band		
Linear	0.79	0.849
Exponetial	0.79	0.849
Logarithmic	0.79	0.849
Quadratic polynomial	0.79	0.849
MLR	0.82	0.943

Table 9: Summary of the three regression models between Mg and reflectance

Model	r	RMSE (mg L^{-1})
Single band		
Linear	0.77	0.317
Exponetial	0.78	0.317
Logarithmic	0.77	0.317
Quadratic polynomial	0.90	0.229
MLR	0.91	0.242

Table 10: Summary of the three regression models between Cu and reflectance

Model	r	RMSE (mg L^{-1})
Single band		
Linear	0.43	0.004
Exponetial	0.43	0.004
Logarithmic	0.43	0.004
Quadratic polynomial	0.43	0.004
MLR	0.43	0.004

derived from this model is strongly correlated with the measured Mg concentration (r = 0.90).

There are small correlation coefficients between the spectral reflectance and Cu concentration for all 128 bands of mid-infrared spectroscopy. The single band with the maximum correlation coefficients are 117th band (r = 0.43). The results show that it is difficult to predict Cu concentration of the surface water from in situ spectral measurements. From Table 1, obviously, lower concentration Cu samples from the surface water of river network areas of Suzhou city also limit correlation analysis and modeling.

CONCLUSIONS

This study was conducted to correlative analysis between mid-infrared reflectance spectra and water quality indexes of the river network areas of Suzhou city, sampled on November 23, in 2009. The descriptive statistic and analysis of variance between reflectance and water constituents were studied. There were higher correlation between TN and other water quality indexes (r>0.5) and high correlation between TP and Mg with the correlation coefficient of -0.851. In addition, single band regression analysis, MLR analysis and SiPLS analysis are used for the correlation analysis between reflectance and water constituents. The results show that, for TN, the SiPLS model has the smallest RMSE (0.33) and the TN concentration derived from this model is strongly correlated with the measured TN concentration (r = 0.82), for TP, the SiPLS model has the smallest RMSE (0.060) and the TP concentration derived from this model is strongly correlated with the measured TP concentration (r = 0.76). For the prediction of K, Mg, the results indicate that the optimal method is with the single band models of the reflectance and the concentration derived from the single band models are strongly correlated with the measured concentration. The study suggested that mid-infrared spectroscopy may be considered as a diagnosis technology for TN, TP, K and Mg content in water.

The study provides space data systems for forecasting of water environment and improving water quality plan. And then the study provides water quality parameters and biotechnologies for environmental monitoring based on satellite remote sensing technology, which will improve capacity of water environmental monitoring in Taihu Lake.

ACKNOWLEDGMENT

This study was supported by the National Science Foundation of China grant (41201338), Natural Science Fund grant of JiangSu Province(BK2012164), Science and Technology Project of SuZhou city(SYN201105, SYG201257), "Qinlan Project" of JiangSu Province University, Pre-research Fund of Suzhou Vocational University (2012SZDYY08) and Opening Project of JiangSu Province Support Software Engineering R and D Center for Modern Information Technology Application in Enterprise(SX201201). The authors would also like to express appreciations to colleagues in our laboratory and workers of Suzhou Environmental Monitor Centre for their valuable comments and other helps.

REFERENCES

Chen, Q., J. Zhao, M. Liu, J. Cai and J. Liu, 2008. Determination of total polyphenols content in green tea using FT-NIR spectroscopy and different PLS algorithms. J. Pharm. Biomed. Anal., 46: 568-573.

Feng, Y., X. Gu, J. Tang and S. Cheng, 2007. Discrimination of tea varieties by mid-infrared spectroscopy combined with pattern recognition. J. Food Sci. Biotechnol., 26: 7-11.

ISO/DIS 15681, 2003. Water quality: Determination of orthophosphate and total phosphorus contents by flow analysis (FIA and CFA)-Part 2: Method by Continuous Flow Analysis (CFA). http://www.iso.org/iso/catalogue_detail.htm?csnumber=35051.

Kang, J., X. Gu, J. Tang and L. Wang, 2010. Simultaneous quantified analysis of wine by mid-infrared reflectance spectrometry and partial least square. Chinese J. Spectroscopy Lab., 27: 789-797.

Kroon, H., 1993. Determination of nitrogen in water: Comparison of a continuous-flow method with on-line UV digestion with the original Kjeldahl method. Anal. Chim. Acta, 276: 287-293.

Lin, F., Z. Chen, J. Deng and K. Wang, 2009. Determination of leaf nitrogen in rice using FTIR spectroscopy. Plant Nutr. Fertilizer Sci., 15: 750-755.

Naes, T. and B.H. Mevik, 2001. Understanding the collinearity problem in regression and discriminant analysis. J. Chemometrics, 15: 413-426.

Norgaard, L. and R. Leardi, 2005. iToolbox manual: Important notes on the iToolbox for MATLAB. MATLAB, April 2005, USA.

Qian, Y. and P. He, 2009. Analysis of water environment variation in the Taihu Lake Basin. Yangtze River, 40: 40-43.

Shao, Y., Y. Bao and Y. He, 2011. Visible/near-infrared spectra for linear and nonlinear calibrations: A case to predict soluble solids contents and pH value in peach. Food Bioprocess Technol., 4: 1376-1383.

Shen, F., Y. Ying, Q. Sun and J. Hu, 2009. Determination of total sugar content in chinese rice wine by mid-infrared spectroscopy. Trans. Chinese Soc. Agric. Mach., 40: 143-145.

Watanabe, M., 1994. Production and Poison of Blue-Green Algae. University of Tokyo Press, Tokyo.

Zhang, L., L.M. Zhang, Y. Li, B.P. Liu, X.F. Wang and J.D. Wang, 2005. Application and improvement of partial-least-squares in fourier transform infrared spectroscopy. Spectroscopy Spectral Anal., 25: 1610-1613.