第 11卷 第 5期
2007年 9月

Vol. 11, No. 5
Sep., 2007

Article ID: 1007-4619 (2007) 05-0756-06

The Hyperspectral Data Monitoring Model of Chlorophyll-a of Summer in Taihu Lake, China

WEI Yu-chun, HUANG Jia-zhu, LI Yun-mei, GUANG Jie
(Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing 210046, China)

Abstract: In this paper, a new hyperspectral data model that estimates chlorophyll-a concentration (Chla) in Taihu Lake of summer is proposed. The model was developed based on measurement in situ in July 2004, and was validated by hyperspectral data in August 2004. Water samples were collected by Wuxi Taihu Lake Environment Monitoring Station and covered the typical water areas. At each site, hyperspectral data were measured ten times by field spectroradiometer ASD FieldSpec, and were converted into remote sensing reflectance. Different band combinations were calculated and compared, and the normalized band index was selected because it is more explicable. The model built by data in July 2004 is Chla(μg/L) = EXP(2.478 + 16.378 * N66), where N66 is (R696 − R661)/(R696 + R661). Goodness-of-fit statistics of the model R^2 is 0.9051, and $P < 0.0001$. Compared with other models, this one is more stable and is of less absolute error when used to estimate Chla in August 2004. The works in the paper also showed that hyperspectral data model can be used to estimate Chla by month in the summer of Taihu Lake.

Key words: chlorophyll-a; hyperspectral data; Taihu Lake; water environment
CLC number: TP79   Document code: A

1 INTRODUCTION

The chlorophyll-a concentration (Chla) is a main parameter that reflects the eutrophication degree in water body[1]. Using remote sensing data to estimate chlorophyll-a concentration in the water body has already become an important method, and the hyperspectral data are used as a main data source because of its higher spectrum resolution. Many authors have studied the relation between Chla and the hyperspectral data, and indicated that hyperspectral data can be used to estimate Chla[2—9].

Taihu Lake is one of the five biggest fresh water lakes in China, and is an important fresh water resource in the Yangtze River delta. Since 80th of the twentieth century with the fast development in industry, eutrophication in Taihu Lake has become more serious, to a certain degree, has affected the local economy growth.

Some researchers have obtained some conclusions about relation between hyperspectral data and Chla in Taihu Lake[5,8,9]. However, data used in these works mainly concentrate in the one period or the partial water area, and the model did not been tested by the extrapolation confirmation.

The purpose of this paper is to establish a model that is robust and can be used to make predication of Chla based on the hyperspectral data in Taihu Lake, summer.

2 REGION AND DATA

Taihu Lake is a large shallow water lake, which water

Received date: 2006-08-10; Accepted date: 2007-03-15

Foundation item: This work is supported by NSFC (40571110) and 863 Program of China under grant No. 2003AA131060. We would like to express our gratitude to Li Xuwen, Liu Ke and Jiao Hongbo for their works in carrying out this study.

Biography: WEI Yu-chun, Ph. D., works in Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, focus on environmental remote sensing and geo-model analysis. E-mail: weiyuchun@njnu.edu.cn

Address: WEI Yu-chun, HUANG Jia-zhu, LI Yun-mei, No. 1, Wenyuan Road, Nanjing City, Jiangsu province, 210046. E-mail: weiyuchun@njnu.edu.cn

area is 2338 km^2, and the mean depth range is from two to four meters. The region climate is the subtropical monsoon. The minimum precipitation appears in January and February, and the maximum appears in June and July. Take Taihu Lake as the center, the west is upstream where distributes hills and mountains, the east is the downriver where distributes the plain.

The Wuxi Taihu Lake Environment Monitoring Station has set up 21 samples for routine water quality monitoring in typical area from 1996 (Fig. 1). Water samples are collected and processed in the first 10 days of each month according to the national standard. Statistic result of data shows that the average Chla is 27 μg/L in the lake, the standard deviation is 56 μg/L. The average Chla in each month has a obvious seasonal variation (Fig. 2).

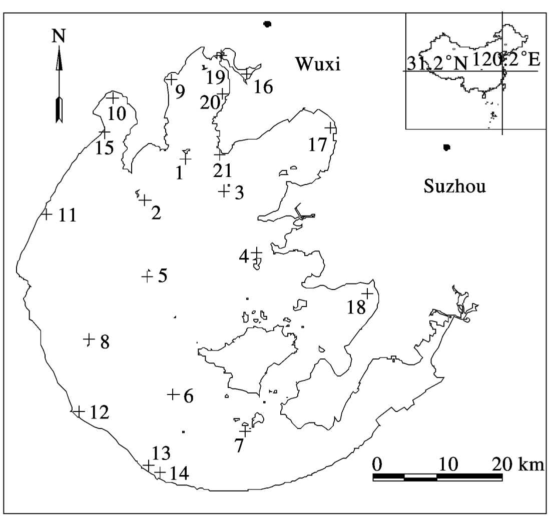

Fig. 1 Location of samples in Taihu Lake

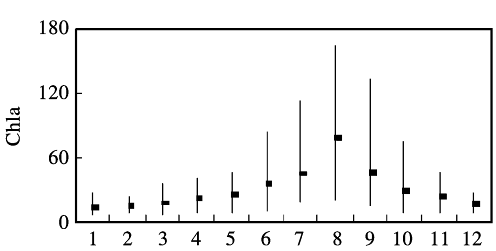

Fig. 2 Average and range of Chla monthly in Taihu Lake

In summer, the biggest variance of Chla appeared in August and September, June and July are the next. Average Chla increases from June obviously, and reaches the highest in August, then reduces gradually. Besides the summer, variation of average Chla in each month is not high, and the smallest variance appeared in winter. Therefore, the summer is the best season to establish remote sensing monitor model of Chla.

In the summer of 2004, water samples were collected and analyzed according to the national standard. In the same time, water surface spectrums were measured by filed spectroradiometer of ASD FieldSpec. The instrument's spectral bandwidth is 1.58 nm and spectrum resolution is 3 nm. A standard grey plate with reflectance 30% is used as the reference. Measurement procedure followed that described by Tang[10]. The instrument away from the water surface approximately one meter, measures the grey plate, water surface and the sky, repeating ten times for each object. Water surface status, wind direction and wind speed were recorded at the same time.

Data in this paper are collected in July and August of 2004, the corresponding date is 1—7 th in July and 2—6 th in August, respectively.

3 DATA PROCESSING

Spectrum measurement data were processed according to Tang[10]. Measure value was transformed into the remote sensing reflectance and denoted by Rn, where n is wavelength value. Wind speed parameter is 2.5%. The median value of ten measurements was taken as sample value. Because of the noise in measurement, reflectance with wavelength range from 420 nm to 900 nm was used in the paper.

In some sample sites, spectrum measurement did not carry on or was useless because of the influence from the weather condition and water surface status. Therefore, total number of samples used is 17 (not including sample 6, 12, 14 and 18) in July, and 20 in August (not including sample 12).

There are many methods of analyzing the relation between spectrum reflectance and Chla[11]. The general workflow includes three steps: a) data transform; b) determine the sensitive band combination, the wavelength's position and other characteristic parameter of spectrum data; c) build model. Considering the feasibility and usability, the correlation analysis and regression analysis method were used in this paper.

Linear relation between reflectance and Chla in July was checked by the scatter chart at the first, and logarithm relation was found. Therefore, Chla is transformed by the natural logarithm, and is denoted by lnChla.

Correlation analysis was used to get the band index. For a single band, the Pearson correlation coefficient between lnChla and original reflectance or spectrum derivation reflectance with order 1, 2 and 3 were calculated. For any two different bands, difference and ratio of reflectance were used to make band index, correlation between the band index and lnChla was calculated. The band or band index with the max correlation coefficient will be selected, and this is called the max correlation coefficient rule in the paper.

Regression analysis was used to build relationship model between lnChla and band index from data in July. Data in August were used to test the model. The model with the lowest RMSE (Root Mean Squared Error) was accepted in the last.

4 RESULTS AND DISCUSSION

4.1 Basic statistics characteristics

Average Chla in July and August of 2004 was 42μg/L and 143μg/L respectively (Table 1), while it was 39μg/L and 59μg/L in history in the same month. Chla in July 2004 was near the average in history, but that in August was higher. In spatial distribution, the lower Chla appeared in the middle part and the eastern part of the lake, the higher appeared in the west part. Obvious high value appeared in the north part Meiliang lake (sample No. 19, 20) and river enters of the lake (sample No. 10, 12, 15) in the western.

Table 1 Statistics characteristics of Chla (μg/L) in Taihu Lake

Month	Min	Max	Average	Stdev
July	7	192	42	43
August	5	1140	143	255

In generally, when algae's density is low in water, the chlorophyll-a in the algae matter can results in absorption, and a valley in spectrum curve can be found from wavelength 420 nm to 500 nm and at 675 nm. The most remarkable spectrum characteristic in water contained the algae is the peak nearby 700 nm. In addition, because of the influence from algae blue protein, valley or shoulder curve shape can appear in 642 nm[5].

Spectrum curves in July 2004 are shown in Fig. 3.

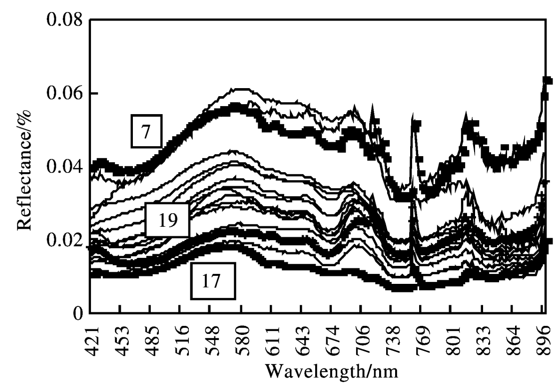

Fig. 3 Spectral reflectance curve of samples in July 2004, Taihu Lake

Characteristics of curve in the data are consistent with previous conclusion. In Fig. 3, Spectrum curve of three typical samples were pointed alone. No. 7 is Zeshan, where Chla is 11μg/L; No. 17 is Shadungang, Chla is 7μg/L, the min value of all samples; No. 19 is Dushankuo, Chla is 192μg/L, the max value of all samples. The spectrum reflectance does not change obviously with the increase of Chla, so it is difficult to estimate Chla by single band reflectance.

4.2 Correlation analysis between single band reflectance and lnChla

Correlation analysis results show in Fig. 4 and Table 2. Negative correlation exists between Chla and original reflectance. Under $P < 0.008$, correlation coefficient $r < -0.62$, the wavelength related is from 450 nm to 510 nm, corresponding to the main absorption valley in the spectrum curve.

In Fig. 4, r_de1, r_de2 and r_de3 are correlation coefficient between lnChla and reflectance processed by spectrum derivation with order 1, 2 and 3. r_de1 changes obviously compared with r, the max negative

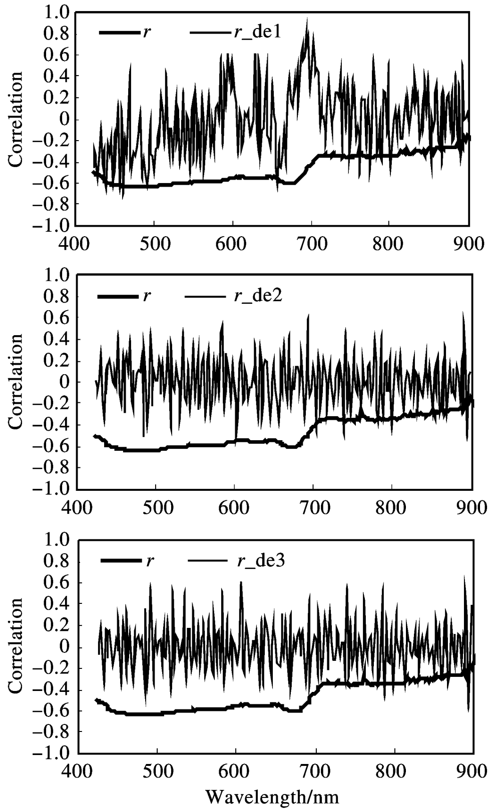

Fig. 4 Comparison of correlation coefficient
r: correlation coefficients in original reflectance;
r_de1, r_de2, r_de3: correlation coefficients in reflectance changed by spectral derivation with order 1, 2 and 3

Table 2 Max and min correlation coefficient between lnChla and reflectance

Correlation	r	r_de1	r_de2	r_de3
Min	−0.6389	−0.6969	−0.5137	−0.5681
Wavelength	486	462	894	892
Max	−0.1474	0.8109*	0.5906	0.6089
Wavelength	897	693	691	604

*: $p = 0.001$

correlation coefficient −0.69 ($p = 0.002$) appears in 462 nm where reflectance curve is the absorption valley in spectrum, the max positive correlation coefficient 0.81 ($p = 0.001$) appears in 693 nm where curve is the reflection peak. r_de2 and r_de3, compared with r_de1, does not increase.

Therefore, to the reflectance in the paper, the spectrum derivation with the first order is suitable, the derivation with the second and the third order are not useful.

4.3 Correlation analysis between band combination and lnChla

In the following sections, rd_{21}, rr_{21} and $rrde_{21}$ are used to represent the correlation coefficient between lnChla and difference (B2-B1), ratio (B2/B1) of two different wavelength's reflectance, and ratio (dB2/dB1) of two different wavelength's reflectance handled by spectral derivation with the first order, respectively. The main results listed in Table 3.

Table 3 Max and min correlation coefficient between lnChla and band combination

Band combination	rd_{21}	rr_{21}	$rrde_{21}$
Min	−0.7332*	−0.8655**	−0.9356***
Bands	453−446	666/655	701/663
Max	0.9185***	0.9507***	0.9364***
Bands	709−660	696/661	693/538

*: $p = 0.001$; **: $p = 0.0001$; ***: $p = 0.000000$

rd_{21} changes in [−0.7332, 0.9185], and rr_{21} changes in [−0.8655, 0.9507]. Band combination produced higher correlation than single band.

According to the max correlation coefficient rule, the band ratio combination is better than the band difference. The significance of max positive rr_{21} is same with that of rd_{21}, but the significance of max negative rr_{21} increases one level.

$rrde_{21}$ changes in [−0.9356, 0.9364]. The max negative $rrde_{21}$ is higher than that of rr_{21}, and the max positive correlation coefficient is bigger than rd_{21} and is lower than rr_{21}.

The sensitive wavelengths change with spectrum's processing method. According to the max correlation rule, 696 nm and 661 nm can be regard as the best.

Ratio band combination used to estimate Chla are different in the different water body in previous studies. For example, Thiemann chose R705/R678[3]. In Qiandao Lake where water is clear, R701/R516 is related closely with Chla[7]. In Chaohu Lake that is similar to Taihu Lake, R705/R680 is chose[6]. In Taihu Lake, the best ratio band combination selected was R705/R675[5,9], R702/R685[12] and R706/R682[8].

Correlation coefficients between lnChla and ratio index that are listed above were calculated. The results showed that these indexes are related with lnChla closely (r is from 0.90 to 0.91) besides R701/R516 (r=0.82). Water environment in Qiandao Lake is different from that in Taihu Lake, so the ratio R701/R516 can not be used in Taihu Lake. Because of the difference in measuring date, measuring method and the water environment, the best band combination cannot be completely same with each other.

Ratio band combination was used widely in remote sensing. However, this expression may bring the misunderstanding. Take R696/R661 as an example, high correlation with lnChla may be regarded that lnChla increases with increase of R696 and decrease of R661. In fact, we can see from the scatter chart, with the increase of Chla, both R696 and R661 decreased. On the other hand, the ratio is sensitive to the data change. The water reflectance is often lower, the smaller change of the denominator in ratio may result in the bigger variation of ratio value.

The detailed analysis indicated that it is the difference between two bands reflectance to relate with lnChla. Therefore, following the form of vegetation index NDVI, a normalization band index is made: NR21 = (B2 − B1) / (B2 + B1).

Fig. 5 shows relation between NR21 and lnChla where B2 is 696 nm and B1 is 691 nm. In distributions of sample, the value of ratio R691/R661 is higher and has great variation, NR21 is lower and has lower variation. To the change of lnChla, NR21 is more stable.

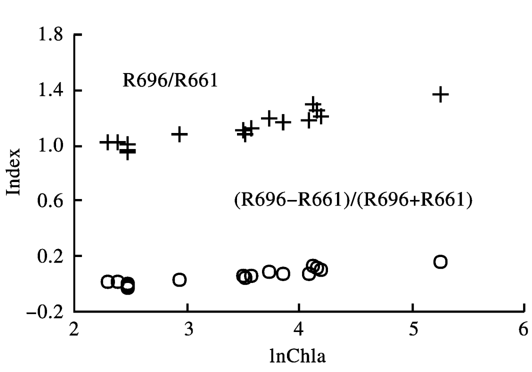

Fig. 5 Relation between lnChla and band index

The correlation coefficient between NR21 and lnChla is denoted by rNR21. The max negative rNR21 is −0.8666, in which B2 and B1 are 666 nm and 655 nm. The max positive rNR21 is 0.9518, in which B2 and B1 are 709 nm and 663 nm. rNR21 of 696 nm and 661 nm is 0.9513, which is very close to the max r21 in Table 3.

RNR21 of two groups wavelength (706, 682) and (705, 675) that used in Taihu Lake[5,8] were calculated, they are higher than r21 too.

According to the above rule, the best band index should be NR21 of (709, 663). Because the change of correlation coefficient is small, NR21 of (696, 661) is also taken into account. These two band Indices were used to build model.

4.4 Regression model

Table 4 is the result of regression analysis where dependent variable is lnChla, explanatory variable in Model 1 is NR21 of (696, 661), and that in Model 2 is NR21 of (709, 663). Higher F value and R^2 suggested the obvious linear relation between lnChla and NR21.

Table 4 Model expression of lnChla in July **2004**, Taihu Lake

	Model 1		Model 2	
	Coefficients	t	Coefficients	t
Intercept	2.4782	24.4	2.9274	37.5
NR21	16.378	11.9	10.5535	12.0
R^2	0.9051		0.9060	
F	143		145	
p	4.53E-09		4.21E-09	
Residuals				
Min	−0.4549		−0.5820	
1Q	−0.1948		−0.1334	
Median	0.0672		0.0607	
3Q	0.2439		0.1977	
Max	0.3904		0.4336	

According to F value and R^2, model 2 is the best. However, the min and max residuals in model 1 are smaller, shows model 1 can fit better to extreme value. Therefore, model 1 is accepted to estimate Chla in August finally.

4.5 Model validation

Data from August 2004 were used to test the model 1 and result showed in Fig.6—Fig.8. In Fig.6, the computed result was compared with the original Chla without consideration of water surface condition and weather condition. ChlaNR66 is the result calculated by model parameters. ChlaR66, which used as a reference, is the result calculated by the model parameters based on the relation between lnChla and ratio of (696, 661) in July. When Chla is low, ChlaNR66 and ChlaR66 are similar. When Chla is high, the difference between ChlaNR66 and ChlaR66 is obvious and error between ChlaNR66 and Chla is lower.

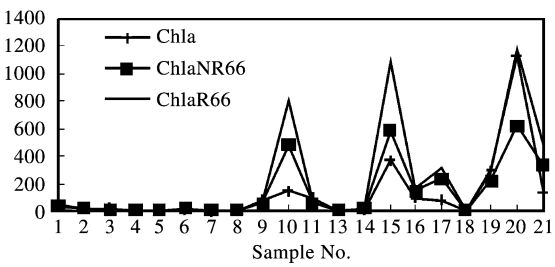

Fig.6　Estimated and measured Chla in August 2004, Taihu Lake

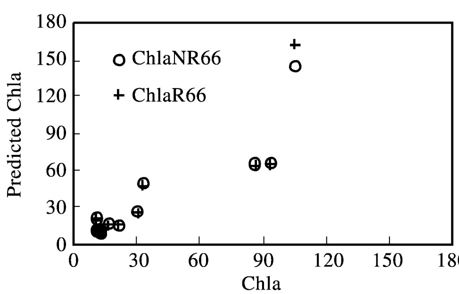

Fig.7　Estimated and measured Chla after remove deviant samples in August 2004

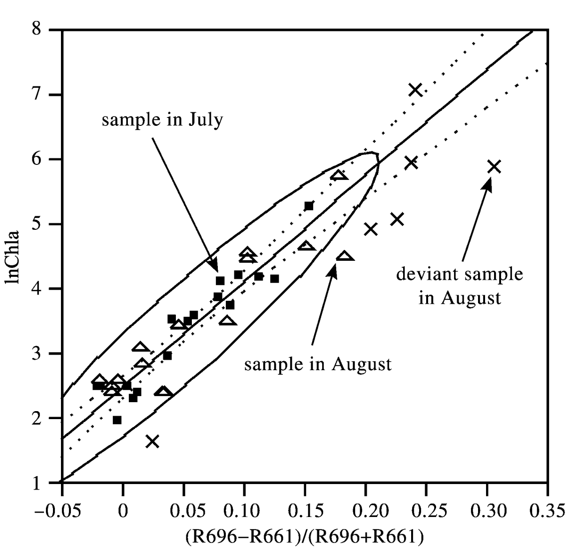

Fig.8　Model fit line and sample distributions

Under the normal measurement condition (removing samples in the cloudy day and water surface with blue-green algae bloom), the total number of samples in August is 13. In these samples, error between ChlaNR66 and Chla is lower than that between ChlaR66 and Chla too (Fig.7).

Fig.8 shows all samples in July and August, fit line and confident fit curve of Model. Based on double variable normal probability ellipse with confident 0.99, the model has the better forecast ability under the normal condition. This also suggests that model built in July 2004 can be used to estimate Chla in August 2004.

By model, the lowest absolute error is 1μg/L, while its relative error is 6%. The highest absolute error is 42μg/L and its relative error is 22%. The error value changes with location of samples. Samples with the absolute error smaller than 5μg/L appeared in the center of Taihu Lake (sample No. 4—8). The sample (sample No. 19) with the max absolute error, which is 42μg/L in July and 78μg/L in August, appeared in the main channel of the lake, but its relative error is smaller than 25% in both months.

5　CONCLUSIONS

Correlation between the single band reflectance and the chlorophyll-a concentration is low and can not be used to build model. Higher correlation can be found between lnChla and reflectance after spectrum derivation with the first order. Band combination can highlight the difference in reflectance. Normalized band index is better than the ratio index. Model based on normalized band index has lower error and is easy to explanation.

Regression model of Chla in July 2004 is Chla(μg/L) = exp(2.4782 + 16.378 * (R696 − R661)/(R696 + R661)), where R661 and R696 are remote sensing reflectance at wavelength 661 nm and 696 nm. Validation by data in August 2004 shows that this model can be used to estimate Chla. Therefore, based on hyperspectral data, building a model to estimate Chla in July and August in Taihu Lake is possible.

Both expression and selection of sensitive bands are important in model building. When the goodness R^2 of model is near, distributions of residuals may be different. It is necessary to validate model by different data from different date.

Linear relation between Chla and hyperspectral data is obvious and useful, but model precise need to increase further. Study of variation of Chla and suspended sediment and their optics characteristics in different months and spatial will be the next works.

REFERENCES

[1] Qing B Q, Hu W P, Cheng W M, et al. Process and Mechanism of Environmental Changes of the Taihu Lake[M]. Science Press (Chinese). 2004

[2] Han L H, Rundquist D C. Comparison of NIR/RED Ratio and First Derivative of Reflectance in Estimating Algal-chlorophyll Concentration: a Case Study in a Turbid Reservoir[J]. Remote Sensing of Environment, 1997, 62: 253—261.

[3] Fraser R N. Hyperspectral Remote Sensing of Turbidity and Chlorophyll a Among Nebraska Sand Hill Lakes[J]. Int. J. Remote Sensing, 1998, 19: 1579—1589.

[4] Thiemann S, Kaufmann H. Determination of Chlorophyll Content and Trophic State of Lakes Using Field Spectrometer and IRS-1C Satellite Data in the Mecklenburg Lake District[J]. Germany Remote Sensing of Environment, 2000, 73: 227—235.

[5] Shu X Z, Yin Q, Kuang D B. Relationship between Algal Chlorophyll Concentration and Spectral Reflectance of Inland Water[J]. J. Remote Sensing (Chinese), 2000, 4(1): 41—45.

[6] Li S J, Wu J, Wang X J, et al. Correlations between Reflectance Spectra and Contents of Chlorophyll a in Chaohu Lake[J]. Journal of Lake Sciences (Chinese), 2002, 14(3): 228—234.

[7] Liu Y, Wang K, Zhou B, et al. Preliminary Study on Hyperspectral Remote Sensing of Qiandao Lake Chlorophyll-a Concentration[J]. Journal of Zhejiang University (Agric. & Life Sci.) Chinese, 2003, 29(6): 621—626.

[8] Ma R H, Dai J F. Chlorophyll-a Concentration Estimation with Field Spectra of Water Body Near Meiliang Bayou in Taihu Lake[J]. Journal of Remote Sensing, Chinese, 2005, 9(1): 78—86.

[9] Liu K, Huang J Z, Zhang Q. Study on Hyperspectral Monitoring of Concentrations for Chlorophyll-a of Alga in Taihu Lake[J]. Journal of Nanjing Normal University (Natural Science), Chinese, 2005, 28(3): 97—101.

[10] Tang J W, Tian G L, Wang X Y, et al. The Methods of Water Spectra Measurement and Analysis I: above-water Method[J]. J. Remote Sensing, Chinese, 2004, 8(1): 37—44.

[11] Pu R L, Gong P. Hyperspectral Remote Sensing and Its Applications[M]. Higher Educations Press (Chinese), 2003.

[12] Yang D T, Pan D L, Zhang X Y, et al. Retrieval of Water Quality Parameters by Hyperspectral Remote Sensing in Lake TaHu, China. Optical Technologies for Atmospheric, Ocean, and Environmental Studies, Pts 1 and 2[A]. Proceedings of The Society of Photo-Optical Instrumentation Engineers (Spie)[C]. Bellingham, Spie-Int Society Optical Engineering, 2005, 5832: 431—439.

夏季太湖叶绿素a浓度的高光谱数据监测模型

韦玉春，黄家柱，李云梅，光 洁

（南京师范大学 虚拟地理环境教育部重点实验室，南京 210046）

摘 要：本文依据2004年7月的实测数据构建了太湖夏季叶绿素a浓度的实测光谱数据估计模型，并使用2004年8月的数据对模型进行了验证。调查样点覆盖了太湖内的典型水域，水样数据由无锡太湖环境监测站采集。样点的光谱数据用ASD FieldSpec野外光谱仪获取，每个样点测量10次，测量结果被转换为遥感反射率。对不同的波段组合进行比较分析后，从可解释性出发，最终选择了归一化指数表达式作为最佳波段组合，所建立的模型为：Chla(μg/L) = EXP(2.478 + 16.378 * N66)，其中，N66为(R696 − R661)/(R696 + R661)。模型的 R^2 为0.9051，显著性 $P<0.0001$。与其他模型相比，本文的模型比较稳健，用于估计8月的叶绿素a浓度具有较小的绝对误差。本文的工作同时表明，在太湖的夏季相邻月份，可以使用实测光谱数据模型进行水体叶绿素a浓度的估计。

关键词：绿素a；高光谱数据；太湖；水环境