Quantitative remote sensing monitoring of water quality in Bohai Bay based on Landsat multispectral data

YE Wei [1], SONG Wei [1, *]

[1] Laboratory of Waterway Environmental Protection Technology, Tianjin Research Institute for Water Transport Engineering, M.O.T Tianjin 300456, China

Abstract. In this paper, through the principal component analysis of water quality survey data of Bohai Bay in 2006, 2009 and 2013, the main pollutant was selected, and the quasi-simultaneous Landsat multispectral remote sensing data are regressed to establish the quantitative inversion model of the sensitive band and the main pollutants in seawater. The accuracy of the model is determined to meet the requirements of quantitative inversion of water quality remote sensing through the significance test method of accuracy assessment, providing a basis for future multispectral remote sensing monitoring of water quality indicators.

1 INTRODUCTION

The Bohai Sea as a semi-enclosed inland sea, pollution buffer capacity is weak, marine biological disaster prevention and control is a long way to go[1]. Prevention and control of water pollution is the first to establish a complete water environment monitoring system, the water quality of the sea on a large scale, dynamic, continuous observation, and assessment of the status of water quality. Commonly used water quality evaluation methods are mainly single-factor index method, comprehensive index method, graded evaluation method and other methods, but due to the correlation between water quality indicators, the evaluation model is difficult to guarantee the effect, and in the case of more station data, the calculation is too large[2]. In particular, the single-factor evaluation method, although it is convenient to learn the status of seawater, but because of the lack of comprehensive evaluation of many indicators and it is difficult to grasp the pollution situation from a global perspective. Principal Component Analysis can incorporate multi-dimensional factors into the same system for quantitative research, using a small number of comprehensive indicators to reflect the information of multiple variables without losing the original variable data, avoiding the subjectivity of artificially determining weights, making the analysis more accurate and reliable[3].Multispectral remote sensing is a photoelectric detection technique that emerged in the 1980s to extract objects of interest through the spectral difference between the target and the background. It has been more widely used in agriculture and forestry monitoring, disaster assessment, urban planning, emergency monitoring, land and resources, etc[4]. In environmental monitoring, we can analyze the ecological status of water quality, hydrological distribution, chlorophyll distribution, suspended solids concentration, etc[5]. B. K.

Purandara, B. S. Jamadar et al. analysed the water quality of The Vembanad Lake in South India using remote sensing[6].Wang, J. , Zhanget al. used Landsat 8 OLI data to invert the potassium permanganate index of Songhua River in Harbin[7]. In this paper, principal component analysis was used to screen out the main evaluation factors in the water quality factor, and regression analysis was used to screen out the corresponding sensitive bands and establish a regression prediction model.

2 MATERIALS AND METHODS

2.1 Study Area

The Bohai Sea, formerly known as the Bohai Sea and the North Sea. The Bohai Sea is part of the western Pacific Ocean and is also China's inland sea. The Bohai Sea consists of five parts: the Liaodong Bay in the north, the Bohai Bay in the west, the Laizhou Bay in the south, the Central Shallow Sea Basin and the Bohai Strait[8]. TheBohai Bay is located at the bottom of the western part of the Bohai Sea.

2.2 Research data

2.2.1 Field survey data

Data from the Marine Water Quality Status Survey of 15 stations in April 2003, 20 stations in June 2009, and 22 stations in September 2013, with 10 water quality factors including COD, DO, PH, nitrate, inorganic nitrogen, petroleum, Pb, Zn, Cu, Cd, etc., are shown in Figure1foreach period.

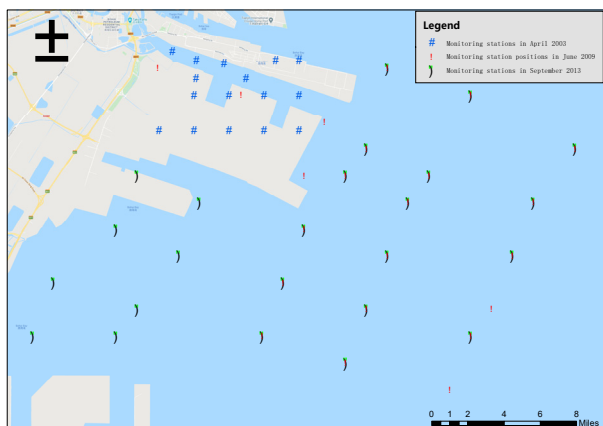* Corresponding author: wei_s56@gmail.com

Figure 1. Map of marine environmental field survey stations.

2.2.2 Landsat data

Due to the failure of the Landsat 7 ETM+ sensor on 31 May 2003, in order to maximize the consistency of the satellite data, Landsat 5 TM data from 16 April 2003, Landsat 5 TM data from 27 June 2009 and Landsat 8 OLI data from 26 September 2013 were used as close as possible to the date of the field water quality survey. The correspondence between the Landsat multispectral data and the field survey data is shown in Table 1.

Table 1. Correspondence table between Landsat multispectral data and field survey data.

serial number	Date of on-site investigation	Landsatdata date
1	April10,2003	April16,2003
2	June18,2009	June27,2009
3	September20,2013	September26,2013

2.3 Research method

2.3.1 Principal component analysis

Ten pollutants from 15 stations in the April 2003 field survey were selected for principal component analysis. From the analysis results (Table2), it can be seen that COD, DO and other previous indicators have a strong direct correlation. Through the variance decomposition principal component extraction analysis, the first three principal components with eigenvalues greater than 1 were selected to replace the original 10 pollutant variables. Divided by the corresponding eigenvalues of the principal components open square root, the two principal components in each indicator corresponds to the coefficient to get the data in the principal component load matrix. It can be seen that COD accounts for more than 80% of the first principal component.

Table 2. Load matrix of the initial factor in the principal component

water quality factor	composition		
	1	2	3
COD	-0.877	-0.143	-0.024

DO	0.89	0.055	0.064
PH	0.819	-0.119	-0.009
potassium nitrate	0.136	0.651	0.201
inorganic nitrogen	-0.529	0.726	0.039
petroleum	0.461	-0.512	0.144
Pb	0.45	0.435	-0.606
Zn	0.091	-0.724	0.143
Cu	0.195	0.376	0.763
Cd	-0.77	-0.326	0.043

2.3.2 Regression analysis

COD has the highest proportion in the principal component and can be used as an index reflecting water pollution, so COD is the dependent variable and the other water pollution indicators are the independent variables to create a scatter plot and do a preliminary impression analysis of COD.
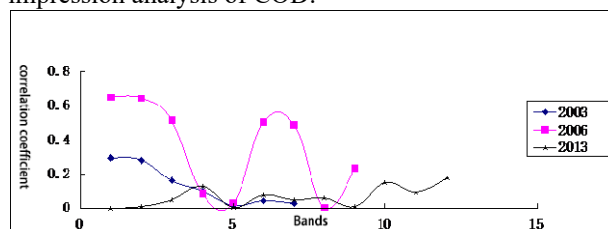

Figure 2. Scatter plot of COD correlation coefficient.

The analysis results show that COD as an important indicator of water pollution analysis, it has a clear correlation in the figure. From the above figure shows that some of the trend lines are satisfied, and there are three trend lines of the R^2 is greater than 0.5. This shows that the COD value for the band has an obvious relationship, correlation is significant, so in the future we analyze water pollution problems, we can use the COD value to make a preliminary prediction of water pollution status.

Similarly, in order to further investigate the relationship between COD and pollutants, we use regression analysis of COD and band values.

2.3.3 Stepwise regression

In order to verify the accuracy of the equations, weselected3out of10data sets for multiple regression and build stepwise regression equations, we used the data to assess the accuracy and significance of the equations and models we built. B1, B2, B3 3 bands were tested.

Test variables for X,Y, select the independent and dependent variables that you want to analyze, the choice of variables is important, if there is no obvious linear relationship between the independent and predicted variables, then linear regression cannot be used for the analysis, in the previous analysis, we analyzed that there is an obvious linear relationship between the COD and the band pixel values, so it was determined that a linear relationship can be established. X, test data as above. Try to perform two-two correlation analysis. X, Y establish matching format data file, do normality test for X,Y, so

that they all obey normal distribution. X, Y selection should meet the following requirements:

(1) whether to meet the requirement of variance congruence.

(2) whether to meet the requirement of the normal. Based on the output correlation data, we eliminated the other bands and let B1, B2, and B3 data participate in the regression analysis.

3 RESULTS AND DISCUSSION

3.1 Model Construction

There is a clear linear relationship between the COD and the pixel values of the remote sensing image bands, so it was determined that a linear relationship could be established. In the stepwise regression, the variables B2 and B3 are excluded from the regression equation due to their low significance because stepwise regression includes significant independent variables in the model and excludes insignificant variables from the regression analysis. The remaining variable B1 participates in the regression analysis, and the final results of the stepwise regression are.

$$Y = 8.183 - 0.061B1 \quad (1)$$

Where: B1 is the pixel value of the B1 band.

3.2 Accuracy assessment

Randomly take five indicators in the remaining water pollutants to assess the accuracy of the model, as shown in Table5, it can be seen that the stepwise regression equation has a strong correlation, the error value is very small. The scatter plot of the calculated and measured values is shown in Figure 2, from which it can be seen that the five sets of data are basically in full compliance with the equation established by stepwise regression, and their R^2 is 0.9344, which meets the requirements of the model.

Table 3. Calculation of stepwise regression errors for the equation.

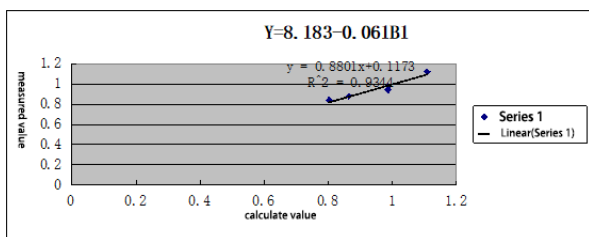| serial number	calculated value	measured value	errors	relative error
1	1.107	1.12	-0.013	-0.011743451
2	0.863	0.88	-0.017	-0.019698725
3	0.802	0.84	-0.038	-0.047381546
4	0.985	0.98	0.005	0.005076142
5	0.985	0.94	0.045	0.045685279

Figure 3. Stepwise regression accuracy assessmentofthe equation.

4 CONCLUSION

In this paper, on the basis of fully reading a large number of aspects about the progress of multispectral at home and abroad, by 10 sets of band data were initially organized, analyzed, and then 10 water pollutants to establish the principal component analysis, and then later, to find the sensitive bands, through step-by-step regression. Further analysis of multispectral investigation of the quantitative monitoring model of sewage who knows.

(1) The analysis shows that we when the multispectral data for multi-order multi-variation processing, it is difficult to find the relationship between them through a simple method, need to resort to other methods.

(2) Due to water pollution factors in the COD in the water body evaluation system has an important role, when the COD analysis, we have a COD, the greater the more serious water pollution. By evaluating the accuracy of the stepwise regression and multiple regression equations, we found that the stepwise regression has a higher R-value and its correlation is higher, so we can use the stepwise regression equation Y=8.183-0.061B1 as the optimal equation in the measurement of COD content.

There are some shortcomings in this paper and some issues that deserve further investigation and improvement.

(1) So the multispectral data used for the analysis is only 10 sets, which is overall small and somewhat influences the conclusion.

(2) In the principal component analysis, we randomly selected several water body indicators, and there is a certain degree of chance.

(3) Therefore, some of the data are abnormal and can only be selected for elimination as anomalous data.

Further outlook, we can see that multispectral data has a great advantage in water monitoring In the future time, multispectral will develop faster and faster, its application will be more and more extensive, it can be involved in the field of forestry, agriculture, water conservancy, civil construction, atmosphere, geography, biology and so on. Ecological and environmental monitoring is an arduous and long-term task,

In recent years, with the rapid development of 3S technology (RS, GPS, RS), ecological monitoring has become more efficient, scientific and convenient. In the past, ecological monitoring tended to focus on monitoring pollution and disasters, describing them qualitatively and in a very unspecific way. And the future trend of ecological environmental monitoring and evaluation of the combination of ways, 3S technology will be more and more applied to the ecological environment among the monitoring, so that ecological monitoring results are more time-sensitive and specific, with scientific basis.

REFERENCES

1. Jiang Bing , & Song Guangjun.(2019). Marine biological hazards in the Bohai Sea and management strategies for prevention and control. Hebei Fisheries, 000(006), 51-53.

2. Dong-Yang, F. U. , Ying, Z. , Da-Zhao, L. , You-Zhuan, D. , Hong, L. , & Feng, Y. , et al. (2015). Evaluation model of coastal water quality and application research based on principal component analysis--a case of leizhou peninsula waters. Journal of Marine ences.

3. Gao Jinrong, Zhao Zechun. Application of principal component analysis in the evaluation of water quality in Maowei Sea[J]. Guangdong Chemical Industry, 2019, 046(006):98-100,76.

4. Plamenka Borovska, Milena Lazarova. Parallel software framework for high-performance multispectral analysis for earth monitoring. Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems, 2011,1,pp.203-207.

5. Wang, Z. , Li, B. , & Li, L. . (2019). Research on water quality detection technology based on multispectral remote sensing. IOP Conference Series: Earth and Environmental Science.

6. B. K. Purandara, B. S. Jamadar, T. Chandramohan, M. K. Jose, & B. Venkatesh. (2018). Water quality assessment of a lentic water body using remote sensing: a case study.

7. Wang, J. , Zhang, J. , Li, T. , & Wang, X. . (2018). Water Quality Analysis of Remote Sensing Images Based on Inversion Model. IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE.

8. Ye, W. , & Song, W. . (2018). The application of remote sensing image sea ice monitoring method in Bohai bay based on c4.5 decision tree algorithm. IOP Conference Series: Earth and Environmental ence, 113, 012218.