Proceedings of the 2017 International Conference on Advanced Mechatronic Systems, Xiamen, China, December 6-9, 2017

Evaluation of water quality for the Beilun Gulf and Zhenzhu Bay by principal component analysis

Minghui Ou,Shan Liang *,Ru Zhang
Key Laboratory of Dependable Service Computing
in Cyber Physical Society, Ministry of Education,
College of Automation,Chongqing University,
Chongqing 400044,China
Email: ouminghui@cqu.edu.cn, lightsun@cqu.edu.cn
201713021043@cqu.edu.cn

Qingyu Xiong
Key Laboratory of Dependable Service Computing
in Cyber Physical Society, Ministry of Education,
College of Software Engineering, Chongqing University,
Chongqing 401331,China
Email: cquxqy@163.com

***Abstract*—With the expansion of urbanization and economic development, the coastal water quality in mangrove ecosystem is becoming worse. This study applied the principal component analysis (PCA) technique to evaluate the water quality monitoring stations located in Beibu Gulf and Zhenzhu Bay of Guangxi and to evaluate the overall rating of stations with respect to the 14 water quality parameters. Results show a truth that the closer the distance between land and sea, the worse of the water quality, and vice versa. Our numerical analysis has verified the truth of the current paper is correct. In addition, our results can prove valuable if budget constraints require the refining of the current monitoring work.**

***Index Terms*—Mangrove ecosystem, Principal component analysis, Water quality.**

I. Introduction

It is well known that mangroves have global distribution within coastal tropical and subtropical regions, which are particularly sensitive to changes in the surrounding sea environment and provide many ecosystem services. It is also well known that the mangrove ecosystem possesses many ecological functions [1],[2]. In addition, mangroves and other coastal vegetations play a vital role in protecting the safety of human inhabitation and preventing wealth from furying by tsunami and typhoon. The idea that mangroves may protect coastal communities from coastal hazards (coastal erosion, typhoon, tidal bores and salt spray, tsunami, etc.) is well known in tropical and subtropical coastal ecology and increasingly by coastal managers [3]. Accordingly, much attention have been paid to discuss the ecological effects of mangrove ecosystem from researchers in recent years, such as attenuate wave energy [4],[5], provide optimal breeding, feeding and nursery grounds for many ecologically and economically important fish and shellfish species and the valuable sources of fuel wood, fodder, timber, tannin and other natural products for local people [6],[7].

However, with the expansion of urbanization and socioeconomic development, mangrove ecosystem is facing a serious threat. Indeed, the mangrove ecosystem is one of the most important biodiversity system and the global biodiversity decline constantly as a results of global climate change and various human activities, such as habitat destruction, pollution, introduction of alien species and overfishing [8],[9].

Mangrove forest is extremely sensitive to the health of the growing environment, such as the soil environment, water environment and microbial environment. Among them, the water quality has a significant impact on the growth of mangrove forests and is prone to change by the external environment. Many literatures have reported on the water quality parameter and monitoring stations by some statistical methods [10],[11],[12],[13]. For example, the principal component analysis (PCA) and principal factor analysis (PFA) techniques have been used to evaluate the effectiveness of the surface water quality-monitoring network in a river [10]. And two methods (PCA and PFA) are also used to evaluation the water quality for the Nakdong river watershed [12].

PCA and PFA are multivariable statistical techniques which are used to identify important components or factors that explain most of the variances of a complex system. They are designed to reduce the number of variables to a small number of indices and to preserve the relationships present in the original data. Those works are very valuable especially in the limited funds and resources. In recent years, the PCA and PFA techniques have been applied to research variety of environmental issues, including evaluation of ground water monitoring wells, classification and identification of the marine spatial [14],[15].

Accounting for the distribution of various mangroves in the coastline of the Beilunhekou Nature reserve in Guangxi. It is an important hot-spot to preserve the biodiversity of the mangrove ecosystem healthy. The study area (Fig.2) is close to the estuary of the coastal river basin, and its water quality can be changed by the pollution of human activity easily.

The aims of this study are to demonstrate the application of the PCA technique to evaluate the water quality monitoring stations located in Beibu Gulf and Zhenzhu Bay of Guangxi and to evaluate the overall rating of stations with respect to the various water quality parameters. The study also provide some guidance for the water quality management.

The remainder of the paper is organized as follows. In section II, we express a research motivation and the occurred

978-1-5386-2602-3/17/$31.00/ ©2017 IEEE

times of typhoon are counted. Study area and methods are introduced in section III. In section IV, some results are presented. Finally, conclusions are presented in section V.

II. Research motivation

One of the ecologic services providing by mangroves is to buffer the impacts of waves, typhoon and storm surges on coastal property and infrastructure by dissipating incoming wave energy [5]. However, in the process of social development, many original ecosystems have suffered more or less damage. We have collected the typhoon data in the last decade from the internet (http://typhoon.weather.com.cn/, time from 2008 to September of 2017). And we also count the number of the typhoon landed in china. Those results have been shown in Fig.1.

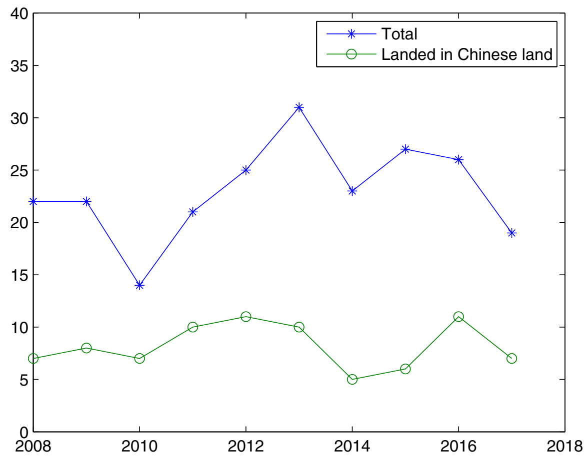

Fig. 1. The occurred times of typhoon and landed in china

Indeed, from statistical results we can know that the total number of typhoon occurrences was 230 times, which landed in the land of Chinese reached 82 times. There is no doubt that these misfortunes ill bring great losses to the lives and economies of the Chinese people. In addition, since from 2008 the times of typhoon landed in Guangxi of China are up to 14. Such as HATO and PAKHAR, they let coastal people's life more terrible.

Fortunately, there are more mangrove forests along the coast. And many previous studies shown that mangroves can buffer the impacts of waves, storm surges, and typhoon energy [2],[5]. Thus, it is indispensable to study the water quality of mangrove growth environment.

III. Study area and methods

The Beilunhekou National Nature Reserve is located in Beilun Gulf and Zhenzhu Bay of Guangxi Province. It is an area of approximately 3000 hectares and the corresponding coastline is 105 km. The protected area has a special geographical location, and Zhenzhu Bay has the largest area of mangrove forest, woody forest and mouse community in mainland China. There is an irreplaceable importance in the coastal mangroves of continental China.

We selected 22 monitoring stations which are located along with the Beilun Gulf and Zhenzhu Bay to sample the water quality data. And we used PCA technique to analysis the sampled data. Because of the dissolved oxygen (DO) is a positive index, and we use equation

$$\overline{x}_{DO} = \max x_{DO} - x_i$$

to deal with the each value of DO. And then the fourteen water quality parameters from the 22 stations were examined in this study(TABLE I). In this work, the research data is the average of the 2014 monitoring data.

Fig. 2. The distribution of the monitoring stations along with the coastal line

Mathematically, PCA is a multivariable statistical techniques used to identify important components that explain most of the variances of a complex system by linear transportation. The procedures of PCA are shown as follows.

(i) standardization of the measurements to ensure that they all have equal weights in the analysis,
(ii) calculate the covariance matrix C,
(iii) find the eigenvalues $\lambda_1, \lambda_1, \cdots, \lambda_n$ and the corresponding eigenvectors $\alpha_1, \alpha_2, \cdots, \alpha_n$ and extract the principal component,
(iv) find the principal component comprehensive evaluation function.

It is necessary to use PCA analysis because the study area was large and the monitoring stations are scattered and water quality system was highly non-linear and dynamic.

IV. Results

Software IBM SPSS Statistics 19 and MATLAB 2014a are used to perform the analysis. The value Kaiser-Meyer-Olkin (KMO) is 0.743, which indicate to analysis the correlation and other properties are reasonable. The variance distribution in PCA is shown in TABLE II. And the cumulative variance value is greater than 0.85 (or 85%) was considered significant.

TABLE I

WATER QUALITY DATA FROM THE MONITORING STATIONS LOCATED IN THE STUDY AREA

Water monitoring cstations	PH (standard units)	Salinity	DO (mg/L)	COD (mg/L)	Phosphate (mg/L)	Inorganic nitrogen (mg/L)	Crude oil (mg/L)	Chlorophyll$-a$ $(\mu g/L)$	Cu $(\mu g/L)$	Pb $(\mu g/L)$	As $(\mu g/L)$	Nitrogen, total(mg/L)	Phosphorus, total (mg/L)	Ammonia -Nitrogen (mg/L)
1	7.167	5.525	1.913	1.580	0.024	0.952	0.050	8.517	0.750	0.693	0.461	1.080	0.065	0.186
2	7.317	6.267	1.870	1.707	0.030	0.809	0.075	7.890	1.730	0.930	0.484	1.096	0.066	0.288
3	7.557	11.637	1.490	1.507	0.019	0.936	0.058	8.353	1.933	1.953	0.628	0.866	0.055	0.259
4	8.007	19.731	1.730	1.447	0.013	0.387	0.050	10.293	2.423	2.120	0.960	0.490	0.042	0.194
5	8.053	21.965	1.343	1.297	0.012	0.264	0.039	9.983	1.347	0.617	1.010	0.534	0.033	0.131
6	8.110	24.515	1.590	1.207	0.009	0.214	0.048	9.217	1.693	1.060	1.150	0.313	0.030	0.129
7	8.067	24.437	1.800	1.767	0.009	0.210	0.034	9.590	2.087	1.807	1.153	0.311	0.045	0.115
8	8.110	22.749	1.717	1.443	0.012	0.226	0.049	12.633	1.770	1.057	1.153	0.341	0.040	0.110
9	8.320	17.256	0.000	3.093	0.012	0.263	0.088	12.033	2.223	1.040	0.983	0.586	0.068	0.147
10	8.053	23.964	1.957	1.347	0.010	0.170	0.050	4.033	3.910	3.623	0.967	0.348	0.053	0.060
11	8.083	25.792	2.150	1.127	0.009	0.149	0.031	4.240	2.443	2.310	0.980	0.267	0.044	0.058
12	8.100	28.043	2.353	0.973	0.008	0.105	0.030	3.717	3.240	2.103	1.036	0.223	0.040	0.051
13	8.193	26.220	1.340	1.323	0.007	0.222	0.053	15.030	2.380	1.767	1.390	0.319	0.083	0.168
14	8.133	26.561	1.147	1.247	0.007	0.135	0.047	11.027	2.403	2.263	1.347	0.265	0.042	0.093
15	8.067	26.134	2.113	1.093	0.008	0.131	0.038	3.067	2.867	1.917	1.277	0.267	0.036	0.055
16	7.763	18.361	1.813	1.390	0.012	0.239	0.042	8.230	2.580	1.120	0.897	0.461	0.045	0.085
17	7.963	23.179	2.217	1.183	0.009	0.190	0.041	4.513	2.227	2.187	1.110	0.343	0.037	0.073
18	7.983	24.562	2.213	1.217	0.009	0.193	0.031	3.573	1.263	2.667	1.153	0.337	0.043	0.083
19	7.950	17.354	1.073	1.530	0.011	0.252	0.040	5.087	1.287	1.543	0.855	0.395	0.048	0.079
20	8.037	20.502	1.927	1.463	0.009	0.171	0.066	10.030	1.273	0.560	0.920	0.352	0.045	0.058
21	7.927	23.359	2.430	1.327	0.010	0.197	0.053	4.243	3.670	2.387	1.041	0.344	0.053	0.069
22	7.990	25.753	2.353	1.180	0.009	0.156	0.050	3.963	1.630	1.010	1.210	0.345	0.052	0.062

TABLE II
THE VARIANCE OF THE COMPONENTS

Components	Total	Inital eigenvalue variance %	Cumulative value %
1	7.2525	51.802	51.802
2	2.828	20.203	72.005
3	1.492	10.656	82.661
4	0.756	5.403	88.064

From TABLE II, we selected four components and their variance together accounted for about 88.064% of the total variance. Therefore, our discussions should focus only on the first four components.

The loading matrix of four components is shown in TABLE III. Compared the loading absolute value of four components that we can know the fourteen water quality parameters information is reflected by the four components properly.

TABLE III
THE LOADING MATRIX OF COMPONENTS ABOUT PARAMETERS

Water quality parameters	1	2	3	4
PH	-0.289	-0.349	0.013	-0.053
Salinity	-0.359	-0.089	-0.003	0.154
DO	-0.118	0.492	0.005	0.112
COD	0.188	-0.403	0.172	-0.455
Phosphate	0.345	0.167	0.043	-0.009
Inorganic nitrogen	0.342	0.159	0.039	0.187
Crude oil	0.231	-0.318	0.222	-0.223
Chlorophyll$-a$	0.127	-0.434	-0.173	0.535
Cu	-0.177	0.034	0.633	0.018
Pb	-0.185	0.164	0.564	0.157
As	-0.309	-0.222	-0.054	0.311
Nitrogen, total	0.361	0.103	0.034	-0.002
Phosphorus, total	0.202	-0.180	0.401	0.253
Ammonia-Nitrogen	0.316	-0.048	0.065	0.444

From the loading matrix in the PCA, the four components, Z_1, Z_2, Z_3, Z_4, can be given as follows, respectively.

$$\begin{aligned} Z_1 =& -0.289x_1-0.359x_2-0.118x_3+0.188x_4+0.354x_5 \\ &+0.342x_6+0.231x_7+0.127x_8-0.177x_9-0.185x_{10} \\ &-0.309x_{11}+0.361x_{12}+0.202x_{13}+0.316x_{14} \end{aligned} \quad (1)$$

$$\begin{aligned} Z_2 =& -0.349x_1-0.089x_2+0.492x_3-0.403x_4+0.167x_5 \\ &+0.159x_6-0.318x_7-0.434x_8+0.034x_9+0.164x_{10} \\ &-0.222x_{11}+0.103x_{12}-0.180x_{13}-0.048x_{14} \end{aligned} \quad (2)$$

$$\begin{aligned} Z_3 =& 0.013x_1-0.003x_2+0.005x_3+0.172x_4+0.043x_5 \\ &+0.039x_6+0.222x_7-0.173x_8+0.633x_9+0.564x_{10} \\ &-0.054x_{11}+0.034x_{12}+0.401x_{13}+0.065x_{14} \end{aligned} \quad (3)$$

$$\begin{aligned} Z_4 =& -0.053x_1+0.154x_2+0.112x_3-0.455x_4-0.009x_5 \\ &+0.187x_6-0.223x_7+0.535x_8+0.018x_9+0.157x_{10} \\ &+0.311x_{11}-0.002x_{12}+0.253x_{13}+0.444x_{14} \end{aligned} \quad (4)$$

In the views of above, it is obviously that the four components Z_1, Z_2, Z_3 and Z_4 are linear function about the water quality parameters from (1), (2), (3) and (4), respectively. Where x is the water quality parameter, the subscripts denote the corresponding parameter in TABLE I (PH as $1,\cdots,$ Ammonia-Nitrogen as 14). And we displayed the diagram of component loadings for the first component and the second component (Fig. 3).

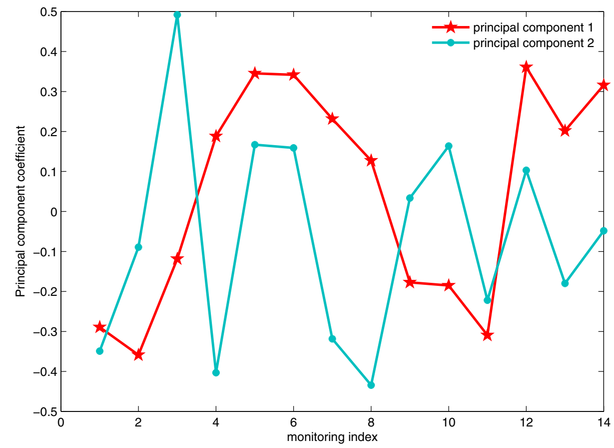

Fig. 3. Component loadings for the first component and the second component

And then, in order to understand the comprehensive level of the water quality in the monitoring stations of the study area. we use the four components corresponding variance value as the coefficient to construct a comprehensive evaluation function (5), where $Z_i, (i = 1, 2, 3, 4)$ represents the four components.

$$S = 0.51802Z_1 + 0.20203Z_2 + 0.10656Z_3 + 0.05403Z_4 \quad (5)$$

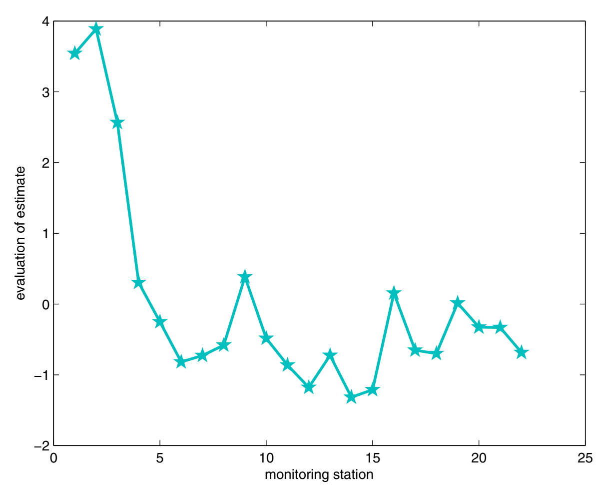

Fig. 4. The score of each monitoring station

According to the comprehensive evaluation function, the water quality of 22 stations in study area in 2014 was ranked. As shown in TABLE IV. And in order to show the results more clearly, the corresponding diagram was shown in Fig. 4.

TABLE IV
THE RANK OF THE WATER QUALITY

Monitoring stations	overall rating
14	-1.317
15	-1.211
12	-1.179
11	-0.863
6	-0.819
7	-0.727
13	-0.725
18	-0.700
22	-0.684
17	-0.653
8	-0.580
10	-0.485
21	-0.332
20	-0.326
5	-0.251
19	0.014
16	0.153
4	0.304
9	0.383
3	2.567
1	3.542
2	3.887

Because of all the variables values are reverse index, i.e., the variable reflects that the water quality deteriorate as the values increase.

From the overall rating, we can know the water quality of the fourteenth station is the best, while, the second station is the worst. Fig. 4 shows that with the serial number of the station increasing, the overall rating shows a downtrend. Indicating that the closer the distance between land and sea, the worse of the water quality, and vice versa. At the same time, this reflects the fact that the water quality is more susceptible to the impact of the people's activity with distance decreased.

Some guidance can be found from these results to improve the coastal water quality when the funds and resources are limited. For example, it is a prior task for us to improve water environmental of second, first and third station if the management with a limited budget.

V. CONCLUSION

This paper has demonstrate the application of the PCA technique to evaluate the water quality monitoring stations located in Beibu Gulf and Zhenzhu Bay of Guangxi and to evaluate the overall rating of stations with respect to the 14 water quality parameters is reasonable. And the imformation of the fourteen water quality parameters can be reflected by the four components. The connection between water quality and the distance of land and sea is acquired. According to the results of this study can provide some guidance if budget constraints require the refining of the current monitoring work.

ACKNOWLEDGMENT

The authors would like to thank experts for their helpful comments and suggestions, and this work is supported by the Science and Technology Major Special Project of Guangxi (GKAA17129002).

REFERENCES

[1] I. Nagelkerken, S. Blaber, S. Bouillon, P. Green, M. Haywood, L. Kirton, J.-O. Meynecke, J. Pawlik, H. Penrose, A. Sasekumar *et al.*, "The habitat function of mangroves for terrestrial and marine fauna: a review," *Aquatic Botany*, vol. 89, no. 2, pp. 155–185, 2008.

[2] K. Zhang, H. Liu, Y. Li, H. Xu, J. Shen, J. Rhome, and T. J. Smith, "The role of mangroves in attenuating storm surges," *Estuarine, Coastal and Shelf Science*, vol. 102, pp. 11–23, 2012.

[3] S. Wells and C. Ravilious, *In the front line: shoreline protection and other ecosystem services from mangroves and coral reefs*. UNEP/Earthprint, 2006, no. 24.

[4] S. Massel, K. Furukawa, and R. Brinkman, "Surface wave propagation in mangrove forests," *Fluid Dynamics Research*, vol. 24, no. 4, pp. 219–249, 1999.

[5] B. Blankespoor, S. Dasgupta, and G.-M. Lange, "Mangroves as a protection from storm surges in a changing climate," *Ambio*, vol. 46, no. 4, pp. 478–491, 2017.

[6] M. V. Rasolofo, "Use of mangroves by traditional fishermen in madagascar," *Mangroves and Salt Marshes*, vol. 1, no. 4, pp. 243–253, 1997.

[7] F. Spaninks and P. van Beukering, "Economic valuation of mangrove ecosystems: potential and limitations," 1997.

[8] E. J. Gregr, A. L. Ahrens, and R. I. Perry, "Reconciling classifications of ecologically and biologically significant areas in the worlds oceans," *Marine Policy*, vol. 36, no. 3, pp. 716–726, 2012.

[9] C. Martin, M. Tolley, E. Farmer, C. Mcowen, J. Geffert, J. Scharlemann, H. Thomas, J. van Bochove, D. Stanwell-Smith, J. Hutton *et al.*, "A global map to aid the identification and screening of critical habitat for marine industries," *Marine Policy*, vol. 53, pp. 45–53, 2015.

[10] Y. Ouyang, "Evaluation of river water quality monitoring stations by principal component analysis," *Water research*, vol. 39, no. 12, pp. 2621–2635, 2005.

[11] H. Lee, K. Oh, D. Park, J. Jung, and S. Yoon, "Fuzzy expert system to determine stream water quality classification from ecological information," *Water science and technology*, vol. 36, no. 12, pp. 199–206, 1997.

[12] K. Y. Jung, K.-L. Lee, T. H. Im, I. J. Lee, S. Kim, K.-Y. Han, and J. M. Ahn, "Evaluation of water quality for the nakdong river watershed using multivariate analysis," *Environmental Technology & Innovation*, vol. 5, pp. 67–82, 2016.

[13] S. E. Kim, I. W. Seo, and S. Y. Choi, "Assessment of water quality variation of a monitoring network using exploratory factor analysis and empirical orthogonal function," *Environmental Modelling & Software*, vol. 94, pp. 21–35, 2017.

[14] T. Winter, S. Mallory, T. Allen, and D. Rosenberry, "The use of principal component analysis for interpreting ground water hydrographs," *Groundwater*, vol. 38, no. 2, pp. 234–246, 2000.

[15] H. Hao, C. Bin, and L. Jinlan, "The marine spatial classification and the identification of priority conservation areas (pcas) for marine biodiversity conservation–a case study of the offshore china," *Ocean & Coastal Management*, vol. 116, pp. 224–236, 2015.