Application and Analysis of Bayesian Method and Grey Relational Analysis in Marine Water Quality Evaluation

WenchaoZhang[1], **HuiyingGao[1]**, **HaiSun[1]**

1 Ocean University of China, Qingdao, China
E-mail: 18306485875@163.com

Abstract. There are many factors that influence the quality of marine water. In order to make the evaluation process more efficient and accurate, based on the normal distribution principle and Bias formula, this article establishes the seawater quality evaluation model by Bayesian method. Taking the evaluation of water quality of a sea area in Qingdao as an example, the measured data of eight water quality monitoring points are selected. The results of the evaluation are compared with the grey relational analysis. It shows that the results of the two methods are the same, which are the I type of water quality. So the Bayesian method based on normal distribution is applicable to the evaluation of marine water quality and the Bayesian method has the characteristic of more integrated, suitable for both large and small samples, simple calculation and easily to be used widely.

1. Introduction

With the development of the marine economy, the environmental conditions of marine waters are becoming more and more severe. People's awareness of environmental protection has been gradually improved. The evaluation of marine water quality is of great significance to the improvement of the marine environment. Many researchers have studied the methods of marine water quality evaluation. There are many uncertainties. Considering the fuzziness and grey nature of evaluation, some evaluation methods such as fuzzy mathematics and grey relational analysis are put forward by some researchers, which have achieved certain results in the application of marine water quality evaluation. However, these methods generally have the characteristics of large calculation and complex calculation, and can not meet the requirements of high aging calculation for small samples [1].

In water quality evaluation, we should avoid the influence of subjective factors as far as possible. Bayesian method is widely used to solve the problem of uncertainty, this method can avoid the influence of excessive subjective factors.

In this article, the normal distribution principle and Bayesian method are combined to establish the marine water quality evaluation model, and compared with the grey relational analysis method. The two evaluation results are compared, in order to provide a more effective and convenient new method for marine environmental assessment.

2. The Principle of Bayesian Method and Grey Relational Analysis Method

2.1. Bayesian Method

Bayesian method is based on existing data, using probability theory and statistics method to predict and analyse an event, that is, to guess the probability of this event. Bayesian method originated in a paper in 1763, mainly based on Bayes' theorem, which can be used to solve the problem of statistics and the probability of events[2] [3]. Bayesian method is combined with probability theory and mathematical statistics method.

Bayesian method can be used in water quality evaluation. In order to avoid the sampling error, principle of normal distribution is used to calculate the likelihood probability. This method has the advantage of more integrated, suitable for both large and small samples and its requirements for data is low, with less effort[4].

2.2. Grey Relational Analysis Method

Grey relational analysis is a quantitative analysis of uncertain relations between different factors. This method analysis comparative sequence and reference sequences to judge whether associated closely. Grey relational analysis is often used in the environmental evaluation, the monitoring of sample data as a reference sequence, different standard of environmental indicators as a comparative sequence. The method compares correlation degree, a more similar trend means a closer relationship. This method can reduce the loss caused by the information asymmetry [5].

3. The Steps of Bayesian Method and Grey Relational Analysis method

3.1. The Steps of Bayesian Method

3.1.1. Improved the Bayesian formula
To solve the problem of water quality evaluation, the Bayesian formula can be improved.

$$P(y_{ij}|x_i)=\frac{P(y_{ij})P(x_i|y_{ij})}{\sum_{i=1}^{m}P(y_{ij})P(x_i|y_{ij})} \tag{1}$$

Where y_{ij} is sea water quality standards, i is the evaluation index, i=1,2······m; and j is seawater standard type, j= I, II, III, IV; x_k is monitoring indexes, the number of monitoring is n.
$P(y_{ij})$ is priori probability, it represents the possibility of the water quality level speculated by the experience.
$P(x_i|y_{ij})$ is likelihood function, it represents the water quality value of the monitoring point. According to the statistical theory, when water quality belongs to the grade j, there is a certain sampling error between the sample value and the standard value, and the distribution can be expressed in the normal distribution.
$P(y_{ij}|x_i)$ is posteriori probability, it indicates that water quality is the possibility of grade j under the condition of $P(y_{ij})$. The evaluation of water quality is usually based on the posterior probability. Based on this consideration, this paper uses the principle of normal distribution of sampling error to estimate the prior probability [6].

3.1.2. A Priori Probability Calculation
It is assumed that the water quality of each monitoring point is the same as the probability of a certain index

$$P(y_{i\,\mathrm{I}}) = P(y_{i\,\mathrm{II}}) = \cdots\ldots = P(y_{i\mathrm{IV}})=\frac{1}{4} \tag{2}$$

3.1.3. Calculation
The calculated likelihood function $P(x_i|y_{ij})$ is the probability that a single index is calculated to belong to a certain level. Estimation of likelihood function by the principle of normal distribution. The mean value of water quality standard value of i index is taken as an α_i of the normal distribution of the index, and the standard difference of water quality of i is σ_i, and C_{vi} is the coefficient of variation of i index. Estimate C_{vi} from $C_{vi}=\frac{\sigma_i}{\alpha_i}$.
Calculation of standard deviation α_{ij}, $\alpha_{ij}=C_{vi}\,y_{ij}$ for class j index.

The calculated value of the σ_{ij} into the $t_i=\frac{xi-yij}{\sigma ij}$ is standardized, and the standardized normal distribution is used to calculate the $P(x_i|y_{ij})$.

$$P(x_i|y_{ij})=2(1-\varphi(|t_i|)) \tag{3}$$

$$\varphi(|t_i|)=\int_{-\infty}^{t_i}\frac{1}{\sqrt{2\pi}}e^{\frac{u}{2}}du \tag{4}$$

According to the calculation results of $P(x_i|y_{ij})$, the posterior probability $P(y_{ij}|x_i)$ of a single index is calculated by $P(y_{ij}|x_i)=\frac{P(y_{ij})P(x_i|y_{ij})}{\sum_{i=1}^{m}P(y_{ij})P(x_i|y_{ij})}$.

3.1.4. Multi index comprehensive water quality test probability
$$Pj=\sum_{i=1}^{m}w_i\,P(y_{ij}|x_i) \tag{5}$$
W_i is the weight of different water quality index i, and can be determined according to the influence of various water quality indexes on the water quality.
In the comprehensive evaluation of water quality, the average weight is taken, that is,
$W_{i1}=W_{i2}=...=W_{i8}=\frac{1}{9}$.

3.1.5. Evaluate the water quality
The results of water quality evaluation are determined according to the maximum probability principle.
That is, the greater the probability, the greater the possibility of a certain category of water.

3.2. The Steps of Grey Relational Analysis method

3.2.1. Transform the original data, build a reference sequence and a comparison sequence
The monitoring data constitute the reference sequence x ,
$$x_i^0=\{\ x_i^0(1),x_i^0(2),\cdots\cdots,\ x_i^0(n)\},(i = 1, 2, \cdots\cdots, m).$$
Using the average method, eliminate the differences in order of magnitude

$$x_i=\{\ x_i(1),x_i(2),\cdots\cdots,\ x_i(n)\}=\{\frac{x_i^0(1)}{\overline{x_i^0}},\frac{x_i^0(2)}{\overline{x_i^0}},......,\frac{x_i^0}{\overline{x_i^0}}\},(i = 1, 2, \cdots\cdots, m) \tag{6}$$

$$\overline{x_i^0}=\frac{1}{n}\sum_{k=1}^{n}x_i^0(k) \tag{7}$$

3.2.2. Calculate the grey relational coefficient
Nondimensionalization the original data sequence to normalize the data before processing. (n is the number of evaluation factors, m representative sample quantity)
$x_j(k)$ represents the monitoring result of k at the monitoring stations of i
$$x_i=\{\ x_i(1),x_i(2),\cdots\cdots,\ x_i(n)\}_j=\{\ x_i(k)\mid i=1,2,\cdots\cdots,m\ ;\ k=1, 2, \cdots\cdots, n\},$$
The comparison sequence y is built by the data from the sea water quality standard,
$y_j(k)$ represents the standard values of k at the level of j
$$y_j=\{\ y_j(k)\mid j=\text{I, II, III, IV}\ ;\ k=1, 2, \cdots\cdots, n\}.$$
 Get the grey relational coefficient $L_{ij}(k)$,

$$L_{ij}(k)=\frac{min_j min_k \Delta ij(k)+\rho max_j max_k\ \Delta ij(k)}{\Delta ij(k)+\rho max_j max_k \Delta ij(k)} \tag{8}$$

$$\Delta ij(k) = \mid x_i(k) - y_j(k)\mid \tag{9}$$

Set ρ as the resolution coefficient, $\rho=0.5$[7].

3.2.3. Determine the weight of different indicators
The weight of each factor reflects its importance to the environment.

$$\text{Set } \overline{S_i}(k) = \sum_{j=1}^{m} \overline{S_i}(k)/m; W_i = C_i(k)/\overline{S_i}(k); W_i(k) = W_i/\sum_{k=1}^{n} W_i \qquad (10) \sim （12）$$

Here $W_i(k)$ indicates the weight of i , which reflects the impact degree of this indicator on other indicators. Through which we can get the relative weight of the indicator.

3.2.4. Determine the grey relational coefficient
The grey relational correlation degree of the reference sequence and comparative sequence is obtained:

$$r_{ij} = \sum_{k=1}^{n} L_{ij}(\text{k}) W_i(k) \qquad (13)$$

3.2.5. Evaluate the water quality
Calculate the water sample reference sequence and compare sequence correlation coefficient, determine the water quality evaluation results in accordance with the principle of correlation is the largest.

4. The Marine Water Quality Assessment

Taking the evaluation of marine water environmental quality in Qingdao as an example. The method of Bayesian and grey relational analysis was used to evaluate the quality of sea water in a sea area in Qingdao. For the convenience of comparison, the water quality monitoring data of 8 sampling points in the sea area of Qingdao are adopted, and COD, DO, inorganic-nitrogen, PO_4-P, oil, Cu, Zn, Pb, Cd are selected as evaluation factors. The measured values of sea water sampling points are found in the literature [8][9] (Table 1). The criteria for evaluation of each index are "People's Republic of China sea water quality standard" (GB3097-1997) (Table 2).

Table 1. Monitoring results of seawater quality in Qingdao (mg/L)

Monitoring station	COD	DO	inorganic-nitrogen	PO_4-P	Oil	Cu	Zn	Pb	Cd
M1	0.655	7.678	0.108	9.050	0.051	0.003	0.041	0.002	0.0002
M2	0.705	7.620	0.121	6.350	0.049	0.004	0.074	0.002	0.0002
M3	0.680	7.612	0.131	8.100	0.027	0.005	0.039	0.005	0.0003
M4	0.700	7.502	0.101	6.550	0.034	0.006	0.057	0.003	0.0002
M5	0.750	7.553	0.088	6.550	0.029	0.003	0.049	0.002	0.0003
M6	0.935	7.635	0.128	6.300	0.100	0.003	0.046	0.003	0.0002
M7	0.835	7.635	0.122	7.900	0.049	0.004	0.034	0.002	0.0001
M8	0.600	7.519	0.097	5.950	0.062	0.003	0.039	0.002	0.0002

Table 2. Sea water quality standards (mg/L)

Level	COD≤	DO>	inorganic-nitrogen≤	PO_4-P≤	Oil≤	Cu≤	Zn≤	Pb≤	Cd≤
I	2	6	0.20	15	0.05	0.005	0.020	0.001	0.001
II	3	5	0.30	30	0.05	0.010	0.050	0.005	0.005
III	4	4	0.40	30	0.30	0.050	0.100	0.010	0.010
IV	5	3	0.50	45	0.50	0.050	0.500	0.050	0.010

4.1. The evaluation results of Bayesian method

According to the calculation steps, the standard deviation of the index is calculated. (table3)

Table 3. Calculation results of standard deviation of various indexes

Index	I	II	III	IV
COD	0.73771	1.1066	1.4754	1.8443
DO	1.7213	1.4344	1.1476	0.86066
inorganic-nitrogen	0.073771	0.11066	0.14754	0.18443
PO$_4$-P	6.1237	12.247	12.247	18.371
Oil	0.048432	0.048432	0.29059	0.48432
Cu	0.0042821	0.0085642	0.042821	0.042821
Zn	0.026759	0.066898	0.1338	0.66898
Pb	0.0013718	0.006859	0.013718	0.06895
Cd	0.0006706	0.003353	0.006706	0.006706

The calculation results in Table 3 are replaced by the formula $t_i = \frac{xi - yij}{\sigma ij}$ for the standard. Then according to the formula (3), (4) and (5), the comprehensive posterior probability and evaluation result of the water quality are calculated. (Table 4)

Table 4. Evaluation results of Bayesian method

Monitoring station	I	II	III	IV	Evaluation grade
M1	0.43032	0.24111	0.17656	0.15201	I
M2	0.42859	0.24608	0.17598	0.14934	I
M3	0.43945	0.23905	0.17409	0.14742	I
M4	0.43277	0.24258	0.17525	0.14939	I
M5	0.42359	0.23884	0.18090	0.15667	I
M6	0.43878	0.24560	0.17091	0.14470	I
M7	0.43148	0.24159	0.17548	0.15145	I
M8	0.41773	0.24574	0.18003	0.15650	I

From the calculation results of table 4, $p_{\mathrm{I}} > p_{\mathrm{II}} > p_{\mathrm{III}} > p_{\mathrm{IV}}$. According to the principle of maximum correlation, the sea water of 8 monitoring points can be judged to be Ⅰ type of water. And the order of water quality of each monitoring point is M3>M6>M4>M7>M1>M2>M5>M8.

4.2. Evaluation Results of Grey Relational Analysis
The measured values in Table 1 are dimensionless treated in accordance with the formula (6) and (7), and the obtained sequence is used as a reference sequence. The concentration limit of each standard in Table 2 is the comparison sequence. According to the calculation steps, the results of grey relational degree calculation and the evaluation of water quality are obtained. (Table5)

Table 5. Evaluation results of grey relational analysis

Monitoring station	I	II	III	IV	Evaluation grade
M1	0.94347	0.91884	0.91751	0.89325	I
M2	0.94506	0.92196	0.91452	0.89013	I
M3	0.94396	0.91974	0.91712	0.89229	I
M4	0.94396	0.92149	0.91487	0.89041	I
M5	0.94496	0.92154	0.91481	0.89036	I
M6	0.94536	0.92210	0.91532	0.89091	I
M7	0.94423	0.92008	0.91726	0.89246	I
M8	0.94546	0.92232	0.91339	0.889127	I

From the calculation results of table 5, $r_{1\,\mathrm{I}} > r_{1\,\mathrm{II}} > r_{1\mathrm{III}} > r_{1\mathrm{IV}}$; $r_{2\,\mathrm{I}} > r_{2\,\mathrm{II}} > r_{2\mathrm{III}} > r_{2\mathrm{IV}}$; $r_{3\,\mathrm{I}} > r_{3\,\mathrm{II}} > r_{3\mathrm{III}} > r_{3\mathrm{IV}}$; $r_{4\,\mathrm{I}} > r_{4\,\mathrm{II}} > r_{4\mathrm{III}} > r_{4\mathrm{IV}}$. According to the principle of maximum correlation, the sea water of 8 monitoring points can be judged to be Ⅰ types of water. And the order of water quality of each monitoring point is M8>M6>M2>M5>M7>M4>M3>M1.

5. Comparison and Analysis
The evaluation results of the two methods are basically the same. All the water quality evaluation results of the monitoring points are I. It shows that Bayesian method based on normal distribution is suitable for evaluating the water quality of the sea water, and the result is accurate. However, there are differences in the sorting of water quality at different monitoring points by the two methods. Taking monitoring station M2 and M4 as an example, in the Bayesian method, the water quality of M4 is better than that of M2, while the result of grey relational analysis is the opposite. It can be found that these two methods are similar in thought, the difference is mainly due to the different ways of weight assignment.

6. Conclusion
Based on the Bayesian formula and the normal distribution principle, this article constructs the Bayesian water quality evaluation model. Taking seawater monitoring point in a sea area of Qingdao as an example, water quality evaluation is carried out, and the evaluation results are compared with those of grey correlation analysis, results show that, with modified Bayesian methods for sea water quality assessment is reasonable and feasible. Compared with the grey correlation analysis, the Bayesian method is more comprehensive and also applicable to small sample problems.
In this paper, taking into account the sampling error of the principle of normal distribution, the effects of the removal of the sampling error index value and the standard value of the cause, can be used to estimate objectively; in addition to many different calculation methods of large, complex calculation method, Bayesian principle is clear and the calculation is convenient and fast. In a word, Bayesian method is a feasible new method of marine water quality evaluation, which is worth popularizing.

7. Reference
[1] Changgen F,Yanzhou L.2008 Application of comprehensive evaluation method in environmental evaluation *Journal of Safety and Environment (*Vol.8 No.5*)* pp 112–115
[2] Lingling S, Shuqian W, Baohong S. 2017 Comprehensive assessment of water quality based on Bayesian Method & Normal Distribution Theory *Journal of Hebei University of Engineering (Natural Science Edition) (*Vol.34 No.4*)* pp 70–74
[3] Likun Y , Xinhua Z , Sen P , Xia L. 2016 Water quality assessment analysis by using combination of Bayesian and genetic algorithm approach in an urban lake, China *Ecological Modelling (*Vol.339*)* pp 77–88

[4] Ying Zhao, Ashish Sharma, Bellie Sivakumar, Lucy Marshall, Peng Wang, Jiping Jiang 2014 A Bayesian method for multi-pollution source water quality model and seasonal water quality management in river segments *Environmental Modelling & Software (Vol.57)* pp 216–226

[5] Yuan Y, Jian W. 2017 Advanced Grey Relational Analysis Method and Its Application in Water Quality Evaluation of the Lake-type Wetland *Journal of Landscape Research (*Vol.9 No.4*)* pp 81–83,87

[6] Jianmin B,Rui W, Xiaohan L. 2011 Comprehensive Assessment of Yitong River Quality Based on Bayesian Method *water saving irrigation (*No.5*)* pp 31–36

[7] Hui F, Yinglan S, Lei S. 2007 Application of grey correlation analysis in quality evaluation of marine environment *Transactions of Oceanology and Limnology (*No.3*)* pp 127–131

[8] Lin Z, Wenlin C, Yonggang J. 2007 Evaluation on seawater quality by fuzzy comprehensive evaluation method in Qingdao dumping area *Marine Environmental Science (*vol.26,No.1*)* pp 38–41

[9] Dongliang Z. 2012 *Study on the influence of long-term dumping on the marine environment of Qingdao ocean dumping area* (China: Ocean University of China)