Research on water quality correlation facts based on grey relational and image analysis method

Ziyuan Li

Pinghe Bilingual School, G11 (3), Shanghai, China, Postcode: 200120

Email: 927254361@qq.com

Abstract. To understand the impact to the water quality of Huangpu River, a serious scientific research has been conducted. Using the questionnaire answered by the residence and samples of the water collected across different locations, a model has been setup to analyse the relationship between the water quality and residents’ knowledge of environmental protection. A novel Grey Relational Analysis method was proposed to calculate the impact factors from year 2013 to year 2018. The analysis results show that the residents' environmental awareness and their cognition of water quality is opposite to the actual water quality because of the level of education level of the residence around the river. As for the orientation of the government's environmental protection measures, the research displays a new focus from the perspective of social science: ways to popularize the status quo of environment protection to the general public.

1. Introduction

Water quality control is becoming more and more important to the governments globally. Government also setup standards to measure the water quality. A lot of the effort has been done to improve the water quality, but it seemed water quality (Water Pollution Index) of many locations collected in our recent year’s research still did not meet the standards for Class I water, and some barely even reached the standards for Class IV[1,2]. However, most residents are under the impression that not only the quality of the water is satisfactory, but the protection is also adequate [3]. Along the way, it was observed that some residents were swimming in the river and even washing clothes in its water. Do these residents realize the condition of the water they are associating with? What if they are using heavily polluted water unintentionally? Is there a relationship between residents' understanding of the water and the water quality itself? To address these questions, in this paper, a new model based on grey relational and image analysis method [4] has been built to evaluate the relationship between people’s environmental awareness and the actual water quality.

2. Method

To understand the relationship between the residence’s awareness and the actual water quality, furthermore to identify the areas need more effort to improve the water quality, a serious onsite researches have been conducted along a branch (Huangpu River) of the one of the world’s longest river Yangzi River to gather the resident’s environment awareness level by collecting feedbacks to designed questionnaires. Water samples were collected from 10 locations for the measurement of actual water quality. The research started from at the source of the Huangpu River, and created samples from the water collected at the following seven locations: Longwang mountain, Xiaofeng
bridge, Wujiaba bridge, Wuxiang bridge, Ancheng bridge, Meixi bridge and Baique bridge, which go downstream successively, and Questionnaires were distributed to the locals around the water body along the route of the investigation. In total, 88 copies of survey were distributed, and 100% valid feedback were collected. The survey questionnaire is listed as Table 1.

### Table 1. Survey questionnaire.

1. Do you think the water quality of your local river is good?
   A. Very good  B. Good  C. Fair  D. Not good
2. What do you think is the major source of pollution in the water environment in your area?
   A. Industrial Pollution  B. Agriculture Pollution  C. Domestic pollution  D. Don’t know
3. Do you think the pollution in your residence area is serious?
   A. Serious  B. Normal  C. Not serious  D. No pollution
4. Have you participated in any water protection activities in the past year?
   A. More than 3 times  B. One to two times  C. None
5. Do you know if there is a waste water processing system in your area?
   A. System running well  B. System partially running  C. No system  D. Others
6. What do you think of the water quality in your area compared with that of a few years ago?
   A. Improved  B. Same  C. Worse  D. Don’t care
7. At what level is your knowledge of environment-protection?
   A. Have solid knowledge  B. Not very clear  C. No knowledge
8. Do you know what measures the government has taken to improve the water quality?
   A. Very clear  B. Know about it  C. Hear about, not clear  D. Don’t know
9. What do you think of the impact of the water pollution on your life or work?
   A. Huge  B. Normal  C. No impact
10. Do you know any impacts of water pollution to the human body?
    A. Solid knowledge  B. Know about it  C. No knowledge

### 3. Data collection and processing

#### 3.1. Water quality data

As stated in National Standard gb3838-2002 [5], water can be classified based on the facts of temperature, pH value and Dissolved Oxygen level. Since there are no specific requirements for temperature and pH the respective grades (as shown in the Table 2), these two facts were removed from the calculation.

### Table 2. Basic unit limit of surface water environmental quality standard: mg/L.

| Number | Project          | Class I                                                                 | Class II                                                                | Class III                                                               | Class IV                                                                 | Class V                                                                 |
|--------|------------------|-------------------------------------------------------------------------|-------------------------------------------------------------------------|-------------------------------------------------------------------------|-------------------------------------------------------------------------|-------------------------------------------------------------------------|
| 1      | Water Temperature(˚C) | Human caused changes in water temperature should be limited to: Average maximum temperature rises per week <= 1 | Average maximum temperature drops per week <= 2                                                                         |                                                                        |                                                                        |                                                                        |
| 2      | pH(dimensionless) | 6 ~ 9                                                                   |                                                                        |                                                                        |                                                                        |                                                                        |
| 3      | Dissolved oxygen >= | 7.5                                                                     | 6                                                                       | 5                                                                       | 3                                                                       | 2                                                                       |

By using Dissolved oxygen level, water quality level can be calculated. The dissolved oxygen data collected from 2013 to 2018 is shown in Table 3.

### Table 3. Dissolved oxygen data from 2013-2018.

| Sampling locations | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|--------------------|------|------|------|------|------|------|
| Dissolved oxygen (mg/L) |      |      |      |      |      |      |
| Longwangshan       | 7.70 | 7.72 | 7.3  | 7.20 | 7.52 | 9.15 |
| Xiaofeng bridge    | 10.10| 10.36| 7.15 | 4.00 | 10.92| 7.15 |
| Wuxiang bridge     | 7.34 | 6.55 | 5.00 | 7.85 | 6.89 | 7.5  |
| Ancheng bridge     | 9.00 | 5.08 | 4.05 | 8.41 | 6.38 | 4.01 |
| Meixi bridge       | 5.70 | 4.60 | 4.07 | 3.70 | 6.23 | 3.65 |
| Gangkou bridge     | 5.25 | 3.93 | 3.94 | 7.51 | 6.38 | 4.1  |
| Baique bridge      | 7.65 | 5.30 | 3.72 | 5.30 | 5.92 | 3.36 |
As defined in National Standard 14848-9 of water quality measurement standards[6], the evaluation score of each water class can refer to Table 4.

**Table 4. Water quality quantification table.**

| Class | I | II | III | IV | V |
|-------|---|----|-----|----|----|
| $F_i$ | 0 | 1 | 3 | 6 | 10 |

As shown in Table 4, class I represents the high water quality. Using the Dissolved oxygen level in Table 3, the water quality of each location can be classified to various classes as per Table 2. Based on the classification of each local, using the score in Table 4, Table 3 can be converted to the standard score as shown in Table 5.

**Table 5. Quantitative conversion results of water quality over the years.**

| Location     | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|--------------|------|------|------|------|------|------|
| Longwang Shan | 0    | 0    | 1    | 0    | 1    | 1    |
| Xiaofeng bridge | 0 | 0    | 3    | 3    | 3    | 3    |
| Wuxiang bridge | 0 | 0    | 0    | 0    | 0    | 3    |
| Ancheng bridge | 0 | 3    | 0    | 0    | 3    | 0    |
| Meixi bridge | 0 | 0    | 0    | 0    | 0    | 1    |
| Gangkou bridge | 0 | 3    | 3    | 3    | 3    | 1    |
| Baique bridge | 0 | 0    | 0    | 3    | 3    | 3    |

The water quality index $F$ can be calculated according to equations (1) and (2)

$$F = \sqrt{\frac{F_1^2 + F_{\text{max}}^2}{2}}$$

$$\bar{F} = \frac{1}{n} \sum_{i=1}^{n} F_i$$

Key: $\bar{F}$ represents the average score value of each component $F_i$; $F_{\text{max}}$ - The maximum size of the score value of each component $F_i$.

Using formulas (1) and (2), the water quality index over the years can be calculated, as shown in Table 6.

**Table 6. Water quality index over the years.**

| Year | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|------|------|------|------|------|------|------|
| F    | 0.769| 2.27 | 3.699| 1    | 0.714| 2.44 |

Table 7 shows the mapping of water quality index to the actual water quality.

**Table 7. Water quality table.**

| Water Quality | Excellent | Good | Normal | Poor | Worst |
|---------------|-----------|------|--------|------|-------|
| $F$           | <0.80     | 0.80~<2.50 | 2.50~<4.25 | 4.25~<7.20 | >7.20 |

The actual water quality is shown in Table 8.

**Table 8. Actual water quality.**

| Year | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|------|------|------|------|------|------|------|
| Water Quality | Excellent | Good | Normal | Good | Excellent | Good |

3.2. Residence survey data
The statistical results from the survey in 2018 are shown in Table 9.
Table 9. Statistics of survey results in 2018.

| Question | A      | B      | C      | D      |
|----------|--------|--------|--------|--------|
| 1        | 14.77% | 47.72% | 28.40% | 9.09%  |
| 2        | 30.68% | 6.81%  | 44.31% | 18.18% |
| 3        | 6.81%  | 38.63% | 45.45% | 9.09%  |
| 4        | 1.13%  | 26.13% | 72.72% |        |
| 5        | 32.95% | 32.95% | 7.95%  | 26.13% |
| 6        | 44.31% | 29.54% | 15.90% | 10.22% |
| 7        | 13.63% | 64.77% | 21.59% |        |
| 8        | 15.90% | 37.50% | 28.40% | 18.18% |
| 9        | 61.36% | 22.72% | 15.90% |        |
| 10       | 22.72% | 71.59% | 5.68%  |        |

Different weights were used to obtain the water quality cognition coefficient of residents, and the questions in the questionnaire survey were quantified. The quantization table is shown in Table 10.

Table 10. Quantitative table of residents' understanding of water quality.

| Answer | A | B | C | D |
|--------|---|---|---|---|
| 𝐸 | 0 | 1 | 3 | 6 |

Set $q_i$ as the weight of the questionnaires, $\sum_{i=1}^{10} q_i=1$. The weight of each question is shown in Table 11.

Table 11. Weight table of questionnaire questions.

| Q | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 |
|---|----|----|----|----|----|----|----|----|----|-----|
| $q_i$ | 0.06 | 0.1 | 0.1 | 0.12 | 0.1 | 0.06 | 0.12 | 0.1 | 0.12 | 0.12 |

The average residential can be calculated as formula (3)

$$\bar{E} = \frac{1}{n} \sum_{i=1}^{10} q_i E_i \quad (n=10)$$

The environmental protection index in 2018 was calculated to be 0.1765. When evaluating people's understanding of nearby water quality, their beliefs on water quality is defined as the residents' cognition index of water quality. E is defined as the "understanding coefficient" of locals. The higher the value of E, the better the water quality; on the contrary, E decreases as residents believe that the water quality is worse. According to the questions in the annual survey, the quantity of the influencing factors in the residents' understanding of water quality along the river are determined by adopting different weights on the main ideas of each of the questions. By weighting in this way, the "understanding coefficient" is obtained. The weighting is determined by each main idea's level of influence on the locals' recognition of their water body's condition.

Using the similar method and the historical data collected, the results of comprehensive statistics over the years are shown in Table 12.

Table 12. Water quality recognition coefficient.

| Water quality recognition coefficient | 2014 | 2015 | 2017 | 2018 |
|--------------------------------------|------|------|------|------|
|                                      | 0.17523 | 0.2221 | 0.1207 | 0.1765 |

4. Grey relational data analysis

The calculation results of the above experimental data over time are compared with the actual water quality, as shown in Figure 1. It can be seen from the figure that the increase and decrease of the actual water quality situation index are synchronized with the increase and decrease of the residents' water quality awareness coefficient. That is, the change trend between the two is synchronous.
In order to further verify the above conclusions, the relationship between the two was analysed by grey relational analysis. The degree of correlation is essentially the degree of difference between geometric shapes of curves. Therefore, the difference between curves can be used as a measure of the degree of correlation. The correlation coefficient of the comparison series with the reference series at various moments \( i \) (i.e., points in the curve) \( \xi \).

\[
\xi = \frac{\Delta (\min) + \rho \Delta (\max)}{\Delta (\max)}
\]

(4)

Where:

- \( \rho \) is the identification coefficient, between 0~1, usually takes 0.5.
- \( \Delta \) is the absolute difference value between two curves at certain point. \( \Delta (\min) \) is the minimum difference between the nodes, \( \Delta (\max) \) is the maximum difference between the nodes.

Based on the data collected in section 3, two data reference sequences were constructed. Sequence 1 is from the water quality recognition coefficient, another is the water quality index. The task is to calculate the grey correlation coefficient between these two sequences. The sequences are listed in Table 13.

### Table 13. Data reference sequence.

| #   | Data Reference Value               |
|-----|-----------------------------------|
| 1   | Sequence 1 0.17523, 0.2221, 0.1207, 0.1765 |
| 2   | Sequence 2 2.27, 3.699, 0.714, 2.44   |

Therefore, the correlation coefficient of each moment (that is, each point in the curve) are calculated as Table 14.

### Table 14. Correlation coefficient result.

| Sequence point | Correlation coefficient \( \xi \) |
|----------------|-----------------------------------|
| (0.17523,2.27) | 0.608                             |
| (0.2221,3.699) | 0.4471                            |
| (0.1207,0.714) | 1                                 |
| (0.1765,2.44)  | 0.5827                            |

The overall correlation degree \( r \) formula is as follows (the closer the \( r \) value is to 1, the better the correlation is):

\[
r = \frac{1}{4} \sum_{k=1}^{4} \xi(k)
\]

(5)

Using the result in table, \( r = (0.608+0.4471+1+0.5827)/4 = 0.659438 \).
The value of $r$ is close to 1, indicating a high correlation between the two. The correlation degree value reflects that there is a certain correlation between the people's understanding of water quality around water body and the actual situation of water body.

5. Root cause analysis

The comparison chart of the water quality cognition coefficient and the actual water quality index was analysed, and the increase and decrease were observed simultaneously. Residents believe that the better the water quality, the worse the actual water quality is. This observation not only initially verified the conjecture of the research on this topic -- the water quality that residents think is related to the actual water quality, and the correlation is different; more importantly, combined with the grey correlation analysis method, it can be used to judge whether the correlation degree is different or similar.

According to the relevant data of China statistical network, currently, there are 24,344 people with a college degree or above in Anji County, who are education recipients. Population with a high school (technical secondary) diploma: 56,676; There are 57,936 people with junior high school education, 15,2,360 people with primary school education, and 26,459 people who are illiterate (those aged 15 and over). Refer to Figure 2 for the detail distribution.

Since 2013, the former National Environmental Protection Bureau and the China Insurance Regulatory Commission have issued a number of policies on the implementation of an environmental pollution liability insurance, including measures of encouragement and fines. The government will enforce industrial assessment, strictly examine and approve the business of centralized industrial platforms and eliminate the introduction of polluting enterprises. In addition to the effective exploration of the environmental pollution liability insurance, Huzhou has also implemented the continuous improvement of local villages and the comprehensive treatment of secondary water source protection zones. This shows that the government has been consistently enforcing many measures in terms of water pollution legislations. In summary, the government's environmental protection policies are indeed sound. Therefore, the reasons for the residents' lack of understanding of the Huangpu River's water quality and environment can be inferred as follows:

- According to Figure 2, the number of non-educated people in the population of Anji is comparatively high, and according to the annual surveys, the education and policies on environmental protection isn't popularized; the population with a low education degree is large, so it is reasonable to suggest that it would be illogical for these people to pay too much attention to the education on sciences and environmental protection.
- Although many environmental protection measures were introduced by the government, few people understand or are concerned about these contents, hence naturally, they do not pay much attention to water pollution.
- The locals' habits of using water in their daily lives are not appropriate. For example, people who wash and bathe under Wuxiang Bridge do not know whether the water quality there is up to standard. Admittedly, the inevitable acts of restaurant owners dumping used water into the Huangpu River suggest their lack of understanding on the consequences of water pollution.

The key to reducing illiteracy and the low education population is the popularization of legislations and the guarantee of education. Core subjects such as language and arithmetic should undoubtedly be dominant, but the education on environmental application should strengthen its weight in future student's curriculums. Environmentalism is closely related to the quality of life and development of communities. In school biology classes, the proportion of environmental study should be increased to have a further influence.

Secondly, the government should increase focus on publicizing policies related to the Huangpu River, and moreover, the environment. Policies should be published and promoted in headlines of local newspapers and media to be recognized by a large population; Posters and billboards should be set up in both urban and rural areas; free lectures and organized activities on the local water body and environment protection should be accessible to all residents regularly.
In particular, relevant departments should conduct regular inspections on restaurant owners and those living near the sewage discharge points, at the same time popularizing education on river protection to ensure their water use standards. A good atmosphere should be established for the people to carry out environment protection, supervise the implementation of environmental protection policies, and utilize water resources properly; tourist spots along the Huangpu River should have signs set up, introducing the state of water quality.

6. Conclusions

Based on the research above, this topic determines the difference between residents' understanding of water quality and the actual situation of water quality, and the meaning behind it. As for the orientation of the government's environmental protection management, the research displays a new focus from the perspective of social science: ways to popularize the status quo of environment protection to the general public. Many current water environment protection policies and laws often focus only on large companies and factories. Hence, if ordinary people become familiar with the environment, realizing the importance of environmental problem such as poor conditions of water quality, this form of education will not only aid people to begin subconsciously devoting in the deed of water protection, but also reflect in the locals' habit of using water, standardizing and eliminating unhealthy water usage. Residents along the river should also voluntarily learn and become familiar with the water body beside their homes, knowing to love it and protect it.

The dissimilar grey relational degree greater than 0.65 shows that the popularization of water quality still needs to be improved. The method used in this paper is a preliminary exploration and analysis of the correlation between people's understanding of water quality and the actual situation of water quality along the river. For example, by adding richer time node data and influence factor data, residents' understanding curve and water quality curve can be more convergent and conform to the actual situation. In addition, the correlation between the two factors before and after time can be further discussed for further causal correlation analysis. It is also possible to study the known variation trend of the coefficient curve to predict and analyse the future variation trend.

References

[1] Chen Zhenglou, Xu Shiyan 2001 Resource and Environment in the Yangtze River Basin, Law of surface water pollution in the Yangtze river delta and countermeasures 10(4) 353-359
[2] Dai Shугui 2001 Environmental Chemistry, Higher Education Press, ISBN: 7040199564
[3] Deng Nansheng, Wu Feng 2000 Environment Chemistry Course, Wuhan University Press, ISBN: 9787307047280
[4] Zhang Ying, Gao Qianqian 2015 Research on Comprehensive Water Quality Prediction model based on Grey Model and Fuzzy Neural Network
[5] China environmental science press 2002 Surface water environmental quality standard (national standard of the People's Republic of China gb3838-2002) [S]
[6] State administration of technical supervision 1993 Groundwater quality standard (national standard of the People's Republic of China GB/T 14848-9) [S]
[7] Chen huayu, Zhao jiabao, Liu chunlin 2004 Properties of combined prediction model based on grey relational degree [J] Journal of southeast university (natural science edition) 01
[8] Li wanxu 1990 Clustering analysis method based on grey relational degree and its application [J] Systems engineering 03