Analysis of Water Quality of Guangzhou Waterworks Based on Grey Correlation Method

Cang Chen¹,a, Zhiqin Zhao¹,b* and Suizhi He²,c

¹School of Economy and Trade, Guangzhou Xinhua University, Guangdong Guangzhou 510520, China
²Faculty of Medical Statistics and Epidemiology, School of Public Health, Sun Yat-Sen University, Guangdong Guangzhou 510080, China
*a-e-mail: zhhcc@xhsysu.edu.cn
*bCorresponding author’s e-mail: zzq@xhsysu.edu.cn

Abstract: The quality of urban drinking water has a great influence on the life and work of the residents. Only when the water quality reaches the National Standard, the safety of the residents and the development of society and economy can be guaranteed. In order to study the water quality of tap water in Guangzhou, 42 indicators from 8 water plants in Guangzhou were selected as sample data, and the grey correlation method was used to analyze and evaluate the indicators. The results showed that the water quality of each waterworks met the national sanitary standard for drinking water, among which the water quality of D, G and H waterworks was better than that of E and F waterworks. Based on the comparative analysis of the test indexes, some suggestions for improving the water quality of the factory are put forward. The grey correlation method can be used to evaluate the water quality of waterworks, which can not only make full use of the original data, but also be easy to calculate and the evaluation result is objective, it's very practical.

1. Introduction
The Grey System Theory was first put forward by Professor Deng Julong[1], a Chinese scholar, which is playing an increasingly important role in the field of prediction. The naming method of grey system is derived from cybernetics, in which the depth of color represents the amount of information. For example, black means no information, white means complete information, and grey means incomplete information. Grey system is a system with insufficient information. Its important characteristic is that the factors of the system are uncertain or there is no definite relationship between the factors. Grey system correlation method is a branch of it. It is a method to measure the degree of correlation between factors according to the degree of similarity or difference between the development trends of factors, namely “grey relational degree”. The grey system correlation method can be used to comprehensively evaluate the things and factors affected by many factors. This method not only can make full use of the information provided by the original data, but also can accurately reflect the differences of each evaluation object without complicated operation.

Water is an essential substance for human physiological activities and has a close relationship with life. If the chemical and microbiological properties of water are not good, it will damage human health and cause some diseases[2]. As the capital of Guangdong Province, Guangzhou has developed rapidly in industry. A large number of pollution water and waste water discharge can reduce the quality of
drinking water. Therefore, it is of great significance to understand the quality of drinking water in Guangzhou and to create a high quality and healthy drinking water environment.

2. Grey correlation

(1) According to the purpose of evaluation, the index system is determined and the evaluation data are collected

Let $n$ data sequences form the following matrix:
\[
(X'_1, X'_2, \ldots, X'_n) = \begin{bmatrix}
X'_1(1) & X'_2(1) & \cdots & X'_3(1) \\
X'_1(2) & X'_2(2) & \cdots & X'_3(2) \\
\vdots & \vdots & \cdots & \vdots \\
X'_1(m) & X'_2(m) & \cdots & X'_3(m)
\end{bmatrix}
\]

Where $n$ is the number of evaluated objects, $m$ is the number of indicators, $X'_1 = (x'_1(1), x'_2(2), \ldots, x'_m(m))^T$ \quad $i = 1, 2, \cdots, n$.

(2) Determine the reference data column

The reference data column is an ideal comparison standard, which can be constituted by the optimal value (or the worst value) of each index, or other reference values can be selected according to the purpose of evaluation. Remember as $X'_0 = (x'_0(1), x'_0(2), \ldots, x'_0(m))$.

(3) The index data is dimensionless

The dimensionless data sequence is formed into the following matrix:
\[
(X_1, X_2, \ldots, X_n) = \begin{bmatrix}
x_1(1) & x_2(1) & \cdots & x_3(1) \\
x_1(2) & x_2(2) & \cdots & x_3(2) \\
\vdots & \vdots & \cdots & \vdots \\
x_1(m) & x_2(m) & \cdots & x_3(m)
\end{bmatrix}
\]

Commonly used dimensionless methods include averaging method, initial value method, min-max normalization method and so on

\[
x_i(k) = \frac{x'_i(k)}{\frac{1}{m} \sum_{i=1}^{n} x'_i(k)} \quad x_i(k) = \frac{x'_i(k)}{x'_i(k)} \quad x_i(k) = \frac{x'_i(k) - x_{\min}(k)}{x_{\max}(k) - x_{\min}(k)} \quad i = 1, 2, \cdots, n ; k = 1, 2, \cdots, m
\]

$x_{\min}(k)$ is the minimum value of the K index of the original data, $x_{\max}(k)$ is the maximum value of the K index of the original data. Or use the interpolation method to make each index data value range (or order of magnitude) the same.

(4) Calculate the absolute difference value of the corresponding elements of the index sequence (comparison sequence) and reference sequence of each evaluated object one by one, i.e., $|x_0(k) - x_i(k)|$ (i = 0, 1, \cdots, n ; k = 1, 2, \cdots, m, n is the number of evaluated objects).

(5) Determine $\min_{i=1}^{n} \max_{k=1}^{m} |x_0(k) - x_i(k)|$ and $\max_{i=1}^{n} \min_{k=1}^{m} |x_0(k) - x_i(k)|$.

(6) Calculation of correlation coefficient $\xi_k(k) = \frac{\min_{i=1}^{n} \max_{k=1}^{m} |x_0(k) - x_i(k)| + \rho \max_{i=1}^{n} \min_{k=1}^{m} |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \max_{i=1}^{n} \min_{k=1}^{m} |x_0(k) - x_i(k)|}$, $k = 1, 2, \cdots, m$. In the formula, $\rho$ is the resolution coefficient, which is valued within $(0,1)$. The smaller $\rho$ is, the greater the difference between the correlation coefficients will be and the stronger the distinguishing ability will be. Generally, $\rho$ is 0.5.
When the optimal value (or the worst value) of each index is used to form the reference data column to calculate the correlation coefficient, an improved and simpler calculation method can also be used:

$$\zeta_i(k) = \frac{\min_{j=1}^{n} x_i^j(k) - x_i^j(k) + \rho \max_{j=1}^{n} x_i^j(k) - x_i^j(k)}{|x_i^j(k) - x_i^j(k) + \rho \max_{j=1}^{n} x_i^j(k) - x_i^j(k)|}$$

(1)

The improved method can not only omit the third step to make the calculation simple, but also avoid some negative effects of dimensionless on the function of indexes. If \( \{x_i(k)\} \) is the optimal value of the data column, the larger \( \zeta_i(k) \) is, the better; If \( \{x_i(k)\} \) is the worst value data column, the bigger \( \zeta_i(k) \) is, the worse it is.

(7) Calculate correlation order

For each evaluation object (comparison sequence), the mean value of the correlation coefficient of \( m \) indexes and corresponding elements of the reference sequence is calculated respectively to reflect the correlation relationship between each evaluation object and the reference sequence, which is called the correlation sequence. Remember to \( r_{ij} : \quad r_{ij} = \frac{1}{m} \sum_{k=1}^{m} \zeta_i(k) \quad i = 1, 2, \cdots, n \)

(8) According to the correlation order of each observed object, the comprehensive evaluation result is obtained.

### 3. Problem analysis

#### 3.1 Sample selection and data preprocessing

The sample data of 42 water quality indicators from 8 water supply plants in Guangzhou in April 2021 are selected. The data are from the official website of Guangzhou Water Supply Co., Ltd., which is true and reliable. The total coliform group, heat-resistant coliform group, Escherichia coli and total number of colonies were not detected; Detected values for indicators such as bromate (with ozone), formaldehyde (with ozone), chlorite (with chlorine dioxide disinfection), chlorate (with compound chlorine dioxide disinfection), ozone (with ozone) and chlorine dioxide (with chlorine dioxide) were not used; The detection values of arsenic, cadmium, chromium (hexavalent), lead, mercury, selenium, cyanide, carbon tetrachloride, chromaticity, turbidity, odor and taste, visible matter, iron, manganese, copper, zinc, volatile phenols (based on phenol) and anionic synthetic detergents are less than the standard limits and have no difference in the detection values of each water plant. Therefore, the detection values of 28 indexes in the above three categories are not included in the grey relational degree analysis and calculation. The test results of the remaining 14 indicators are shown in Table 1.

| Indicators       | Unit | A  | B  | C  | D  | E  | F  | G  | H  | Standard Limit |
|------------------|------|----|----|----|----|----|----|----|----|---------------|
| fluoride         | mg/L | 0.2| 0.21| 0.37| 0.37| 0.29| 0.32| 0.21| 0.22| 1.0           |
| nitrate          | mg/L | 1.9| 1.85| 2.04| 1.89| 1.76| 2.42| 1.87| 1.91| 10            |
| chloroform       | mg/L | 0.011| 0.01| 0.018| 0.007| 0.006| 0.011| 0.01| 0.008| 0.06         |
| PH               | —    | 7.8| 7.7| 7.38| 7.52| 7.38| 7.42| 7.72| 7.68| 6.5-8.5       |
| oxygen consumption| mg/L | 0.8| 0.6| 1 | 0.7| 0.9| 0.8| 0.8| 0.7| 3.0          |
| chloride         | mg/L | 13.4| 12.6| 40.4| 37.8| 60.1| 23.7| 11.9| 11.3| 250           |
| sulfate          | mg/L | 30 | 29.6| 20.4| 25.6| 32.3| 34.2| 29.6| 29.7| 250           |
| free chlorine    | mg/L | 0.71| 0.79| 0.91| 0.73| 0.76| 0.71| 0.63| 0.55| 0.3-4         |
Because the comprehensive evaluation of water quality of waterworks was carried out\cite{3}, \(\{x_n\} = \{0,0,0,0,0,0,0,0,0,0\}\) was taken as the reference sequence. On the basis of qualitative analysis of the research problem, one dependent variable factor and several independent variable factors are determined. Let the data of dependent variables constitute the reference sequence \(x_n\), and the data of respective variables constitute the comparison sequence \(x'_i\), \(i = 1,2,\cdots,8\). Eight data sequences form a matrix.

In general, the original variable sequences have different dimensions or orders of magnitude. In order to ensure the reliability of the analysis results, the variable sequences need to be dimensionless. After dimensionless, each factor forms the following matrix: (Keep the data in three decimal places)

\[
X = \begin{bmatrix}
0 & 0.212 & 0.417 & 1 & 0.785 & 0.043 & 0.696 & 0.26 & 1 & 0.5 & 0.295 & 0.011 & 0.444 & 0.32 \\
0.059 & 0.136 & 0.333 & 0.762 & 1 & 0.027 & 0.667 & 0.26 & 1 & 0 & 1 & 0.094 & 0.667 & 0.46 \\
1 & 0.424 & 1 & 0 & 0 & 0.596 & 0 & 0.164 & 0.015 & 1 & 0.152 & 0.881 & 1 & 1 \\
1 & 0.197 & 0.083 & 0.333 & 0.215 & 0.543 & 0.377 & 0 & 0 & 0.25 & 0.026 & 0 & 0.5 & 0.52 \\
0.529 & 0 & 0 & 0 & 0.076 & 0 & 0.862 & 1 & 0.204 & 0.75 & 0.331 & 1 & 0.583 & 0.5 \\
0.706 & 1 & 0.417 & 0.095 & 0.342 & 0.254 & 1 & 0.616 & 0.909 & 0.5 & 0 & 0.563 & 0.444 & 0.22 \\
0.059 & 0.167 & 0.333 & 0.81 & 0.684 & 0.012 & 0.667 & 0.123 & 1 & 0.5 & 0.329 & 0.165 & 0.222 & 0.2 \\
0.118 & 0.227 & 0.167 & 0.714 & 0.684 & 0 & 0.674 & 0.26 & 0.97 & 0.25 & 0.427 & 0.02 & 0 & 0
\end{bmatrix}
\]

3.2 Calculate correlation coefficient and correlation order

The absolute difference values of the corresponding elements of the index sequence (comparison sequence) of each evaluated object and the reference sequence are calculated one by one. Since the reference sequence is \(\{x_n\} = \{0,0,0,0,0,0,0,0,0,0,0\}\), the absolute difference matrix is the same as the dimensionless matrix mentioned above.

\[
\min_{i=1}^{n} \min_{k=1}^{m} |x_n(k) - x'_i(k)| = 0 \quad \max_{i=1}^{n} \max_{k=1}^{m} |x_n(k) - x'_i(k)| = 1
\]
Take $\rho$ as 0.5 to calculate the correlation coefficient, i.e.,

$$
\zeta_i(k) = \frac{\min_{i=1}^{n} x_i'(k) - x_i'(k) + \rho \max_{i=1}^{n} x_i'(k) - x_i'(k)}{|x_i'(k) - x_i'(k) + \rho \max_{i=1}^{n} x_i'(k) - x_i'(k)|} \quad k = 1, 2, \ldots, 14, \quad n = 1, 2, \ldots, 8 \quad (3)
$$

the results are shown in Table 2.

| Tap Water | Correlation Coefficient |
|-----------|--------------------------|
| A         | 0.89                     |
| B         | 0.702                    |
| C         | 0.545                    |
| D         | 0.333                    |
| E         | 0.61                    |
| F         | 0.658                    |
| G         | 0.753                    |
| H         | 0.333                    |

| Tap Water | Sorting |
|-----------|---------|
| A         | 6       |
| B         | 5       |
| C         | 7       |
| D         | 3       |
| E         | 1       |
| F         | 8       |
| G         | 2       |
| H         | 1       |

The correlation order $r_i = \frac{1}{14} \sum_{k=1}^{14} \zeta_i(k)$, $i = 1, 2, \ldots, n$ was calculated and sorted according to the size of the correlation order data. The results are shown in Table 3.

| Tap Water | Correlation | Sorting |
|-----------|-------------|---------|
| A         | 0.610       | 4       |
| B         | 0.607       | 6       |
| C         | 0.608       | 5       |
| D         | 0.705       | 7       |
| E         | 0.6         | 8       |
| F         | 0.551       | 3       |
| G         | 0.632       | 2       |
| H         | 0.691       |         |

### 3.3 Results analysis

According to the correlation order calculated by the grey correlation method, Table 3 can be obtained, from which it can be seen that the water quality of the three water plants $D$, $G$ and $H$ is better and the water quality of $E$ and $F$ is worse. Combined with Table 1, it can be seen that the contents of chloride, sulfate and total $\beta$ in the ex-factor water of $E$ and $F$ waterworks are higher than those of other water works; the content of dissolved total solid in the ex-factor water of $E$ waterworks is higher than that of other waterworks; and the content of nitrate in the ex-factor water of $F$ waterworks is higher than that of other waterworks. The difference of water quality between $D$ and $F$ plants is mainly reflected in the contents of sulfate, nitrate, fluoride and total hardness. Total hardness mainly refers to the content of $\text{CaCO}_3$ in water. $\text{CaCO}_3$ in water mainly comes from soil and rocks. The hardness of water is high.
enough to cause scale, which will affect the taste of water and even affect digestion and absorption. The content of inorganic non-metallic indicators which is sulfate, nitrate, chloride and fluoride mainly comes from the sea into the dip, sewage, industrial wastewater and agricultural seepage. The high content of inorganic non-metallic index will produce certain harm to human body, and the situation of sewage pollution in water can be inferred according to the change of nitrate and chloride content in water.

The grey relation method can be used to comprehensively evaluate the water quality of waterworks[4,5], which can not only make full use of the original data, but also be easy to calculate and can accurately reflect the differences among evaluation objects. Grey correlation analysis has the characteristics of unlimited sample size, simple principle, wide application range and intuitive results. This method is used to compare the proximity between the evaluation object and a specific object. The closer the two are, the greater the correlation of the evaluation object is[6]. Through the grey correlation analysis, it is shown that the water quality testing of the water supply company in Guangzhou in April has reached the national standard of drinking water sanitation. It shows that the management of water works is becoming more and more strict with the rapid development of industry in Guangzhou.

4. Suggestions
The quality of water supply is one of the important signs to measure the degree of civilization of a country and a region. At present, industrial wastewater, domestic sewage and other pollution problems to drinking water sources still exist. According to the analysis results of the water quality of the factory in Guangzhou, the primary task is to reduce the content of sulfate, nitrate, fluoride and total hardness in the water. In order to improve the quality of plant water and improve the stability of water quality, we can start from the following aspects:

(1) Improve the stability of water quality of factory water and the stability of water quality is in connection with water bicarbonate, calcium carbonate and carbon dioxide. When the PH value in the water is less than 6.5 and the content of iron is more than 3mg/L or the pipe is a metal pipe, it will lead to autotrophic iron bacteria and metal corrosion, and then cause the increase of bacteria, turbidity, chroma, iron and other indicators. Therefore, the PH value of water must be adjusted to 7-8.5 when it leaves the factory. The general practice is to add quicklime.

(2) Add chlorine properly because adding chlorine in the factory water can eliminate bacteria, viruses and other microorganisms in the water[2,7]. However, adding too much chlorine will lead to a great increase in the residual chlorine content in the factory water. The residual chlorine will react with the secondary polluted organic matter to form chloroform and carbon tetrachloride, which will seriously affect the water quality. Residual chlorine is useful for disinfection, but it also has side effects. Therefore, under the premise of ensuring the elimination of bacteria, viruses and other microorganisms in the water, the dosage of chlorine should be reduced as far as possible. Chlorine point as far as possible back to move, as far as possible to achieve more chlorine; In order to maintain the residual chlorine in the pipe network, the residual chlorine in the factory water can be made in the state of chloramine which is well mixed with the factory water for more than 30min. Chlorination disinfection can eliminate bacteria, viruses and other microorganisms in water, but it can not effectively kill parasites. Ultraviolet disinfection has a better disinfection effect on parasites, so the combination of ultraviolet disinfection and chlorination disinfection can effectively eliminate bacteria, viruses, other microorganisms and parasites in the water.

(3) Increase activated carbon adsorption to improve the quality of factory water. The biological activated carbon adsorption technology can adsorb microorganisms in water, and has a good removal rate of chloride and heavy metals in water. The removal rate of chloride is about 65%, and the removal rate of heavy metals such as iron, manganese and copper is about 80%-92.5%.

(4) Adding ultrafiltration membrane filtration technology can intercept colloid[8], bacteria and relatively high molecular weight substances in the water, remove a lot of organic matter, reduce turbidity, and improve the stability of water quality.

(5) Ion exchange method can be used to reduce the content of sulfate and nitrate in factory water and improve the quality of factory water.
5. conclusion
As the drinking water of people's life, the water quality of urban tap water has a direct impact on people’s health. Poor water quality and even mild water pollution can lead to people in physical discomfort symptoms after drinking. If it is a serious water pollution, it may threaten people's life safety[7]. It can be seen that it is an urgent task to improve the water quality of waterworks. Therefore, according to the analysis results of factory water quality in Guangzhou, this paper uses quicklime to adjust pH value, chlorination and ultraviolet combined disinfection method, biological activated carbon adsorption method, ultra-membrane filtration method and ion exchange method to improve the factory water quality. In order to further ensure the safety of drinking water, it is necessary to monitor the change of water quality, and constantly improve the measures to improve water quality.

Acknowledgments
This article is supported by the research project :Advanced mathematics(2019HHJG001) on the reform of Internet + Hybrid Teaching Model in Guangzhou Xinhua University.

Reference
[1] Deng, J.L. (1990) Grey system prediction and decision making. Huazhong University of Science and Technology Publishing, Wuhan.
[2] Yan, Z.H., Zhou, J.J., Zeng, G.Q. (2021) Analysis on the testing results of urban drinking water quality in Fuzhou from 2014 to 2018. Chin J Health Lab Technol, 31:880-883.
[3] Di, W.D. (2007) On the guarantee of water quality of water supply network. Sci-Tech Information Development & Economy, 17:278.
[4] Li, Q. Z., He, J. S. (2007) Discussion on comprehensive evaluation method of water quality. Yellow River, 29:50.
[5] Fang, J. Q. (2009) Health statistics. People's Medical Publishing House, Beijing. 391-393.
[6] Li, J.L. (2020) Study on the application of grey correlation analysis in evaluating enterprise value by market method. Yunnan University of Finance and Economics, Kunming.
[7] Nie, X. J. (2016) Detection and analysis of sanitary problems of urban tap water and related water quality indexes. The world's latest medical information digest, 16:211-212.
[8] Wang, R., Lu, K., Feng, B.J., Zhang, L., Zhang, D.D., Hou, C.C. (2018) Sanitary monitoring analysis and control measures of drinking water made and sold in Tianjin. Public health and preventive medicine, 29:32-34.