Research of Geostatistics on Spatial Distribution of Urban Surface Water Quality

Hongquan Sun\textsuperscript{1,a}, Jing Wang\textsuperscript{1,b}, Yachong Tian\textsuperscript{1,c}, Yiran Shi\textsuperscript{1,d}

\textsuperscript{1}Silicon Lake College of Vocational Technology, Jiangsu, China
\textsuperscript{a} email: hqsun@mail.usts.edu.cn, \textsuperscript{b} email: 214306733@qq.com
\textsuperscript{c} email: 779273334@qq.com, \textsuperscript{d} email: 2579992917@qq.com

Abstract: The principles and methods of geostatistics are introduced. With the help of variogram, the contents of the water quality: NH\textsubscript{4}+-N, TP and COD, in rivers in the city of south China are analyzed. By using the Kriging, the special distributions of the content of NH\textsubscript{4}+-N, TP and COD are simulated. The contours of the spatial distribution of the urban surface water quality are shown intuitively. The research results provide a theoretical basis for governing the water environment.

1. Introduction
Spatial information statistics (geostatistics) can not only study the randomness of data, but also the spatial correlation of data. Generally speaking, the data in nature related to spatial location have the characteristics of random and spatially correlated. With the deepening of people's understanding of spatial information statistics, it has been widely used in earth science, resource exploration, environmental science, civil engineering, traffic engineering, meteorology, agriculture, forestry, information processing and other disciplines. In this paper, the spatial distribution of urban surface water quality is simulated by computer using the principle of spatial information statistics, which provides a theoretical basis for tackling water pollution and protecting water environmental.

2. Principles of spatial information statistics
The core content of spatial information statistics includes two parts: one is variation function; Second, Kriging valuation.

2.1 Calculation of experimental variogram
Let (x\textsubscript{i}, y\textsubscript{i}, z\textsubscript{i}) \textsubscript{i=1, 2, … n} is a set of observed values, and (x\textsubscript{i}, y\textsubscript{i}) is the coordinate and (z\textsubscript{i}) is the observed value of the variable at the ith observed point. The following formula

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2$$

is called experimental variogram.

Where $\gamma^*(h)$ is the value of the experimental variogram, $x_i$ the coordinate at the observed point, which represent ($x_i, y_i$) in two-dimensional space. $z(x_i)$ is the observed value at point $i$; $z(x_i + h)$ is the observed value at the point whose distance from Point $i$ is $h$. $h$ is the distance between two observation points, which is called lag distance. $N(h)$ is the number of point pairs spaced $h$ apart.
The experimental variogram has directivity. If the lag distance $h$ is taken in a certain direction, the experimental variogram can be obtained in that direction [1].

2.2. Theoretical variogram model

In practical application, the experimental variogram points $[\gamma(h)]$ are fitted to the theoretical variogram model by using the least square method. The model of theoretical variogram is the spherical model. The mathematical expression of the variogram of the spherical model is [1, 2]

$$
\gamma(h) = \begin{cases} 
0 & h = 0 \\
C_0 + C \left( \frac{3}{2} \cdot \frac{h}{a} - \frac{1}{2} \cdot \frac{h^3}{a^3} \right) & 0 < h \leq a \\
C_0 + C & h > a
\end{cases}
$$

where: $a$ is the range; $C_0$ is the nugget value; $C$ is the arch height; $C_0 + C$ is the base station. The graph of the variogram of a spherical model can be expressed as

![Theoretical variogram of spherical model](image)

From Fig. 1, the variogram $\gamma(h)$ increases with the increase of $h$ within the range of $0 \sim a$ (Range). Therefore, the range of $a$ reflects the range of variable correlation. When $h > a$ the variogram no longer increases monotonically, but is stable near the base value $C_0 + C$, indicating that the variation of the variable is independent of the distance which reflects the random characteristics of the variable. Base station value $C_0 + C$ shows the variable variation. Nugget value $C_0$ gives the continuity of variables.

2.3. Kriging valuation and Kriging variance

Let $(z_i, i=1, 2, \ldots n)$ be the observed values at the observation points. To predict (estimate) a value at an unknown point using the values at the known points is called an estimate. The usual linear estimation formula is

$$
Z_0^* = \sum_{i=1}^{n} \lambda_i Z_i
$$

where $Z_0^*$ is the estimated value at the unknown point; $\lambda_i$ is the weight coefficient of $Z_i$. Obviously, the accuracy of the estimates depends entirely on the weight coefficient $\lambda_i$. The weight coefficient of Kriging is derived on the premise of "Unbiased" and "minimum variance", so it is the Best Linear Unbiased Estimator.
3. Spatial variability analysis on the quality parameters of urban surface water

The water quality parameter is related to the spatial location, which has both randomness and correlation. Therefore, by using the spatial information statistics method to study the spatial variability of water quality parameters, the correlation and random characteristics can be obtained, and the estimated values at unknown points have higher estimated accuracy. In this research, the water data are taken from a city district in south China. The contents of ammonia nitrogen (NH4+ -N), total phosphorus (TP) and permanganate index (COD) in 136 water samples were analyzed and studied, the parameters of variogram are obtained, and the isoline map of the water quality distribution was drawn.

3.1. Experimental variation function

Data acquisition. The author selected 136 water sample collection points in the research area, and obtained the content data of ammonia nitrogen (NH4+-N), total phosphorus (TP) and permanganate index (COD) of the 136 water sample points through analysis and test. The coordinates of each water sample point are \((x_i, y_i)\) \(i=1,2,\ldots,136\), contents \((z_i)\) \(i=1,2,\ldots,136\).

Calculation of variogram. According to formula (1), the ammonia nitrogen content data of the water sample points can be obtained through analysing and testing in the east-west direction and north-south direction \([3~5]\) respectively. And the experimental variogram value can be calculated for each hi. hi is set as \((50, 100, 150, 200, 250)\) (the data units are meters). Based on of the experimental variogram data pairs \([hi, \gamma^*(h_i)]\), the Least Square Method was used to fit the experimental variogram points into the theoretical variogram model, and the corresponding theoretical variogram parameters were obtained (see Table 1).

| Water Quality Index | Directions  | Range (km) | \(a\) | Nugget \(C_0\) | Base Value \(C_0+C\) | Ratio \(C_0/(C_0+C)\) |
|---------------------|-------------|------------|-------|----------------|-----------------|-----------------|
| NH4+ -N             | East-West   | 3.2        | 0.25  | 1.15           | 0.2174          |
|                     | North-South | 1.9        | 0.48  | 1.50           | 0.3200          |
| TP                  | East-West   | 2.0        | 0.6   | 1.80           | 0.3333          |
|                     | North-South | 2.4        | 0.62  | 2.78           | 0.2230          |
| COD                 | East-West   | 2.8        | 0.53  | 1.40           | 0.3786          |
|                     | North-South | 3.1        | 0.40  | 1.60           | 0.2500          |

From Table 1, the three water quality indexes all have spatial correlation within a certain range, but the correlation of different water quality index in different directions is slightly different. Nugget value \(C_0\) usually caused by experimental error and less than experimental sampling scale. That is, spatial variability caused by random. The base station value \(C_0+C\) gives the total variation within the system. For the study area, the spatial variability of each water quality parameter can be described by the ratio of nugget value \(C_0\) and base station value \(C_0+C\), which includes the random factors in the spatial variation components and the spatial structure factors. The ratios obtained from the Table 1 are all less than 0.5, indicating that the variation caused by the spatial structure factors is greater than that caused by random factors. To some extent, this is also related to the spatial distribution pattern of the data in research area.

3.2. Computer simulation of water quality

In order to describe the distribution of NH4+ -N content in surface water more accurately, kriging estimation of research data can be carried out based on the establishment of the Kriging estimation is an unbiased estimation of minimum variance. According to kriging estimation principle, the kriging method is used to estimate 136 water sample points.

Figure 2, Figure 3 and Figure 4 show the contour map of ammonia nitrogen (NH4+ -n), total phosphorus (TP) and permanganate index (COD) content distribution respectively.
Fig. 2  Contour of spatial distribution of NH4+-N

From figure 2, the variation trend of NH4+-N content in this region is significant, higher in the north and lower in the south of the middle range. The content of NH4+-N in the middle is extremely high. In the north range, and this indicates that NH4+-N pollution is relatively serious. The reason lies in the disordered discharge of industrial enterprise wastewater and household sewage in the area. The two small "hills" in the southwest and southeast in figure 2 show the location of sewage treatment stations and concentrated residential areas respectively. According to the analysis and research results, the content distribution law of NH4+-N is visually displayed, which provides a reliable theoretical basis for the environmental protection department to determine the treatment measures and governance order.

Fig. 3  Contour of spatial distribution of TP

Figure 3 shows that there are regions with high TP content in the south, the north and the central part of the study area, especially in the south-west. The high TP content reflects the High-Tech Zones and Industrial Zones in the city, where industrial enterprises are concentrated and sewage is discharged, causing great harm to river water quality. The middle part is the ancient urban area, showing that the
impact of domestic sewage on the rivers in the region can’t be ignored.

Fig. 4 Contour of spatial distribution of COD

From the contour map in figure 4, you can see that the COD concentration in the Middle East river distributes even relatively. The value of the COD concentration is under the “Surface Water Environment Quality Standard” between IV and V. The COD contents of the western rivers in the study area are significantly higher than that of the above parts. According to the investigation, the main reason is that the sewage pipe network in this area is not perfect enough, so the river channel is seriously affected by the domestic sewage. With the impact of some urban construction projects on the river, the river retention and sludge deposition are serious. The river contains a large number of organic matter, resulting in the COD concentration higher than other areas.

4. Conclusions
1) Geostatistics is an effective tool and method to study the spatial variability of variables related to spatial location (regionalized variables). The variogram and its parameters can quantitatively describe the structural characteristics of spatial distribution of regionalized variables, so as to reveal the spatial variation law of regionalized variables.

2) Three-dimensional simulation of surface water quality by using geostatistics can intuitively display the spatial variation rules of water quality, which is conducive to the analysis of abnormal polluted areas and main pollution sources of rivers in the study area, so as to provide corresponding prevention and control measures for river pollution.

3) In this paper, the ammonia nitrogen content index of water is analyzed. And other indexes of water quality can also be statistically analyzed with geostatistics. If GIS, GPS and other auxiliary means are used to determine and record the location of sampling points, a water sample management information system can be developed, or a series of maps such as the distribution map and evaluation map of water quality parameters can be drawn, so as to provide consulting services for water environment management.

Reference
[1] SUN Hongquan. Geostatistics and its Application [M]. Xuzhou: China University of Mining and Technology Press, 1990.
[2] Sun Hongquan, Kang Yongshang, Du Huizhi. Practical Geostatistics Assembly [M].Beijing: Geology Press, 1997.
[3] Xing Jun, SUN Hongquan, LI Weitao. Spatial Information Statistical Analysis of Geotechnical Properties [J]. Geotechnical Engineering, 2004(1): 5-7.

[4] Zhang Zheng, Zhao Junlin, Wang Hongqi, Chen Jiajun. Analysis Principle and Method of spatial variability of migration parameters of water environmental pollutants. Journal of Beijing Normal University (Natural Science edition) 2004. 34(4).

[5] Yuan Zheng Ming, Chen Tongbin, Chen Huang, Wu Hong, Tao Zhou, Jian Li, Luo Jin, Fa Huang, Ze Chun. Spatial structure and distribution characteristics of soil in Beijing suburb. Geomatics Bulletin. 2003. 58(3) 8.

[6] Xu Xiaotong, Tao Yue, Xi Daoying. Geostatistics method was used to evaluate the impact of point pollution sources on river water quality. Water resources protection. 2005. 21(4). 42-46.