Analysis and prediction of the dynamic change of water quality in the inflow rivers

Jingwei Chen

Ministry of Education Key Laboratory of Integrated Regulation and Resource Development on Shallow Lakes, Hohai University, Nanjing 210098, PR China
College of Environment, Hohai University, Nanjing 210098, PR China
1914050123@hhu.edu.cn (Jingwei Chen, Tel.: +86 19850419185)

Abstract. The Le’an River (Jiangxi Province) is polluted by multiple point sources; however, fluctuations in its water quality factors have not been analysed or predicted. This study investigated the Spatio-temporal variation of the river’s water quality. Nine monitoring points were analysed to reveal the past and predict future trends based on field measurements from 2012 to 2020. The autocorrelation coefficient and duration for each monitoring point was evaluated using a rescaled range analysis. Pearson coefficient was calculated to clarify the correlation among monitoring points. The results show that: (1) the water quality of the lower reaches of Le’an River was inferior to Class V quality from 2012 to 2015, but it greatly improved after that, and the entire basin was classified as Class II after 2018; (2) upstream and downstream monitoring sites showed significant correlation characteristics among themselves, and the excessive downstream pollution was attributed to point source pollution from industrial wastewater discharged in Leping City; and (3) the Hurst index of all monitoring points was > 0.5; future water quality characteristics are expected to present long-term memory and persistence. Our results are of significance for controlling the water quality of the Le’an river and nearby economic zones.

Keywords: R/S analysis; Pearson coefficient; long memory; trend prediction; Le’an River

1. Introduction

In recent years, activities to strengthen the economy and accelerate modernisation have led to the discharge of large amounts of industrial wastewater, domestic sewage, and agricultural non-point source pollution in the Le'an River, Leping City, China. The Le'an River is mainly polluted by point sources, including purse seine culture, domestic sewage discharge, agricultural irrigation, and over-exploitation of tourism resources and water resources. These activities discharge large amounts of nitrogen and phosphorus into the river, which increases its eutrophication rate. Eutrophication is mainly attributed to water pollution caused by excessive N, P, and other nutrients. This input of nutrients far exceeds the output, and consequently, algae grow excessively, which disrupts the material and energy flows in the aquatic system, thereby resulting in algal bloom and other harmful phenomena. Many scholars have investigated the nutrient characteristics in the Le'an River, including the correlation between water quality and diatoms in reservoir surface sediments. Zhang investigated the physiological response of heavy metals in sediments of the Le'an River by analysing the activity of benthic snail enzymes. Li et al. explored the influence of rainfall change on
non-point source pollution in the Le'an River Basin, Poyang Lake area. However, past studies have not analysed or predicted the fluctuation of water quality factors in the Le'an River.

This study analysed field measurements in the Le'an River Basin from 2012 to 2020. Nine water quality monitoring points in the upper, middle, and lower reaches of the river were investigated, and the concentrations and evolution trends of TP and $\text{NH}_4^+ - \text{N}$ in the river [1-3]. Based on rescaled range (R/S) and Pearson correlation analyses, the water quality of the Le'an River and the correlation among the results for different monitoring points were investigated to reveal the spatial and temporal distribution law of nutrients in the river [4-6]. The results show that the water quality of Le’an River has improved significantly in recent years, and there is a strong correlation between the monitoring points in time series and spatial distribution. This study aimed to investigate the correlation between water quality and economic development in the Le’an River Basin and provide a scientific background for effective water quality monitoring and protection measures. The results are expected to be of great significance for the control of the water quality of the Le’an River and nearby economic zones [7-10].

2. Research Area and Data Source

The research area included the Le’an River and the Le’an Jiang stream, which is the main stream of the Rao River, a tributary of Poyang Lake in the Yangtze River basin. The study area has a total length of 279 km and a drainage area of 8,989 km². The main tributaries are Lianshi, Fuchun, Lianshui, Changle, Jianshui, and Zhuxi. The Le’an River source section is located in Wuyuan County, which belongs to a subtropical monsoon climate zone, with warm and humid climate, abundant rainfall, abundant vegetation on both banks above the dam site, and high forest coverage. The study area is crucial for agricultural irrigation, animal husbandry, and tourism. As the backbone of the Raohe River, Le’an River has an average annual runoff of 12.6 billion m³, and it is the main water resource in Shangrao, with a theoretical reserve of nearly 240,000 KWh. By the 1990s, small hydropower stations had been built in the basin, with a total installed capacity of 43,000 KWh. The Le’an River is also an important waterway in the basin. It is 0.9–1 m deep in the dry season, and it can be navigable by ships with 30–50 tonnes throughout the year.

In this study, nine water quality monitoring points in the Le’an River Basin were investigated. There were monitoring points in the upper, middle, and lower reaches, namely, Duanxin (DX), Sandu (SD), and Qiukou Town (QKZ) in the upper reaches; Detong Water Plant (DTSC), Xiangtun (XT), and Daicun (DC) in the middle reaches; and Zhenqiao (ZQ), Jiyang Bridge (JYQ), and Hanjiadu (HJ) in the lower reaches. The ammonia nitrogen and total phosphorus data were obtained from Jiangxi Hydrological Monitoring Station. Monthly water quality data of nine monitoring stations from 2012 to 2020 were obtained, with a monitoring frequency of once a month. Every point was sampled and monitored monthly, and 2 L samples were collected by a plexiglass water sampler at 0.5 m underwater. The water samples were separately packed in 500-ml polyethylene bottles, which were previously disinfected with 70% alcohol after acid washing. Subsequently, the water sample was acidified with sulfuric acid, and the contents of $\text{NH}_4^+ - \text{N}$ and TP were determined within 72 h using Nessler's reagent colorimetry and ammonium molybdate spectrophotometry, respectively.
3. Materials and Methods

3.1 R/S Analysis

This research was divided into two main parts: the analysis of pollutants at each monitoring point and the correlation analysis for different monitoring points, that is, the temporal and spatial distribution of pollutants. The trend and forecast of nutrient concentration change in each monitoring point of the Le’an River were investigated in an R/S analysis [11, 12]. The R/S analysis is commonly used to describe and calculate the Hurst index based on a rescaled range. This method is often used to analyse the fractal characteristics and long-term memory process of a time series. The Hurst index represents long memory and is used to describe the autocorrelation of a time series. Memory is reflected in the non-independence of autocorrelation functions, whereas long memory is reflected in the slow decay of autocorrelation coefficients [13-16].

For two non-overlapping time spans with lengths $t_1$ and $t_2$, the correlation coefficient between the increments in these time spans is

$$C(t, t_1, t_2) = \frac{1}{2} \left( \frac{t + t_1 + t_2}{t_1} \right)^{H} + \left( \frac{t + t_1}{t_2} \right)^{H} - \left( \frac{t + t_2}{t_1} \right)^{H} - \left( \frac{t + t_1 + t_2}{t_2} \right)^{H}$$

When $H > 0.5$, the correlation coefficient is greater than 0, regardless of the values of $T$, $t_1$, and $t_2$. When $H < 0.5$, the correlation coefficients are all less than 0. The criteria for the calculation of the Hurst coefficient was then obtained (Table 1) [17-19].

<table>
<thead>
<tr>
<th>$H$ value</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H = 0.5$</td>
<td>This time series is not correlated</td>
</tr>
<tr>
<td>$H &gt; 0.5$</td>
<td>Changing time series has long-term memory</td>
</tr>
<tr>
<td>$0 &lt; H &lt; 0.5$</td>
<td>Time series shows anti-persistence (likely to show opposite trend in the future)</td>
</tr>
</tbody>
</table>

The specific steps for the calculation of the Hurst index were as follows:

1. The water quality standard sequence was used as input data, with the length $M$ given in metres. A logarithmic difference sequence of length $N = M - 1$ was calculated, and the input price sequence was transformed into a logarithmic growth rate sequence:

$$N_{i} = \log \left( \frac{M_{i+1}}{M_{i}} \right), i = 1, 2, ..., M - 1$$

2. The logarithmic growth rate sequence with length $n$ was equally divided into $A$ subsets, and the length of each subset was $n = N/A$. The average of each subset was calculated and noted as $e_{a}, a = 1, 2, ..., M - 1$

3. In each subset $a$, the prior $k$ points were calculated individually $(k = 1, 2, ..., A$) relative to the cumulative deviation of subset $e_{a}$:

$$X_{k,a} = \sum_{i=1}^{k} (N_{i,a} - e_{a}), k = 1, 2, ..., n$$

4. The specific steps for the calculation of the Hurst index were as follows:

$$H = \frac{\sum_{i=1}^{n} X_{k,a}^{H}}{\sum_{i=1}^{n} (N_{i,a} - e_{a})^{H}}$$
The fluctuation range $R_a$ of the logarithmic growth rate series in each subset $a$ was calculated, and it was equal to the difference between the maximum and minimum cumulative deviation:

$$R_a = \max \left( X_{k,a} \right) - \min \left( X_{k,a} \right), \quad 1 \leq k \leq n$$

The standard deviation $S_a$ of the logarithmic growth rate series was calculated for each subset $a$.

Within each subset, the fluctuation range $R_a$ was standardised based on its standard deviation $S_a$, and the rescaling range $R_a/S_a$ was obtained. For the selected length $n$, calculations from step 2 onwards were conducted, and $A$ subsets were obtained, for an $A$ rescaling range[20]. Their mean value $n$ was obtained as the rescaling range of the original logarithmic sequence in the time span of the given length, and it was recorded as $(R/S)_n$.

$$\left( R/S\right)_n = \frac{1}{A} \sum_{a=1}^{A} \frac{R_a}{S_a}$$

The $n$ value was increased, and the first six steps were repeated to obtain the rescaling range $(R/S)_n$ of the logarithmic series for different time spans $n$.

The Hurst index, $H$, was calculated as a direct ratio of $(R/S)_n$ and $n^H$, that is $(R/S)_n = C \times n^H$.

Therefore, double logarithmic regression is conducted for $n$ and $(R/S)_n$, that is, linear regression was performed using pairs, and $\log(n)$ was used for the linear regression of $\log((R/S)_n)$. The intercept of the regression equation in the above relation was the constant $C$, and the slope is the Hurst index $H$. The future trend was then estimated according to the criterion of the Hurst coefficient (Table 1) [21-23].

### 3.2 Correlation Analysis

The Pearson correlation analysis was used to investigate the correlation between the monitoring points of the Le’an River. P value is the probability, reflecting the possibility of an event, which means the probability that the difference between samples is caused by sampling error [24-26]. Three significance levels (0.05, 0.01, and 0.001) were set to investigate statistically significant correlations between the monitoring points. The specific formula for calculating the Pearson coefficient was as follows ($p < 0.05$; $p < 0.01$; $p < 0.001$):

$$r = \frac{l_{xy}}{\sqrt{l_{xx} l_{yy}}}; \quad l_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n};$$

$$l_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}; \quad l_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n}$$

where $x_i$ and $y_i$ are the contents of ammonia nitrogen or total phosphorus index in two monitoring points; $n$ is the total number of monitoring times from 2012 to 2020; $l_{xx}$ and $l_{yy}$ are the sums of squared deviation of variables $x$ and $y$; $l_{xy}$ is the sum of deviation from average products of variables $x$ and $y$; and $r$ is the correlation coefficient of two monitoring points[27-30]. The larger the absolute value of correlation coefficient is, that is, the closer it is to 1 or $-1$, the stronger the correlation is, and the closer the correlation coefficient is to 0, the lower the correlation degree is. Correlation was extremely strong, strong, medium, weak, and extremely weak when $0.8 \leq |r| \leq 1$, $0.6 \leq |r| < 0.8$, $0.4 \leq |r| < 0.6$, $0.2 \leq |r| < 0.4$, and $|r| < 0.2$, respectively.
4. Results and Discussion

4.1 Evolution and Trend of Water Quality

The annual average indexes of $\text{NH}_4^+ - \text{N}$ (Figure 1) and TP (Figure 2) in nine monitoring points in the Le’an River from 2012 to 2020 were analysed. To determine the water quality, we referred to the Surface Water Environmental Quality Standard (GB3838-2002) (Table 2).

Table 2. Maximum concentrations for classification of surface water environmental quality (unit: mg/L).

<table>
<thead>
<tr>
<th>Serial number</th>
<th>Classification</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ammonia nitrogen $\text{NH}_3 - \text{N}$</td>
<td>0.15</td>
<td>0.5</td>
<td>1.0</td>
<td>1.5</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Total phosphorus (calculated as P)</td>
<td>0.02</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td></td>
</tr>
</tbody>
</table>

The results show that, except for downstream points (Hanjiadu, Jiyang Bridge, and Zhenqiao), the $\text{NH}_4^+ - \text{N}$ values were all lower than 0.5 mg/L, that is, the water quality was classified as at least Class II, and in most upstream points the water quality was of Class I. From 2012 to 2015, the $\text{NH}_4^+ - \text{N}$ values of Zhenqiao and Hanjiadu in the lower reaches of the river were far above 2.0 mg/L, that is, they presented a quality inferior to Class V. In 2013 and 2014, Hanjiadu reached the peaks of 10.46 and 10.25 mg/L, and Zhenqiao reached 8.89 and 8.79 mg/L. The reason for these high values was the discharge of polluted wastewater from the industrial and economic belt near Leping City, in addition to the lack of supervision and control. After 2015, the pollution was clearly controlled, and in 2018, it reached 0.309 and 0.280 mg/L in Hanjiadu and Zhenqiao, respectively, which represented Class II quality. Overall, the concentration of $\text{NH}_4^+ - \text{N}$ in each monitoring point decreased, and the water quality constantly improved. After 2018, the water quality in the Le’an River has reached Class II and above.
Advances in Engineering Technology Research

Figure 1. (a) NH$_4^+$ – N and (b) TP concentration of nine main monitoring points in the Le’an River from 2012 to 2020. 3D maps of (c) NH$_4^+$ – N and (d) TP concentrations.

The pollution pattern of total phosphorus is often similar to that of ammonia nitrogen. Except for the three downstream points, the water quality of the other six monitoring points remained above Class II. The downstream Hanjiadu monitoring point was the most polluted place, and its total phosphorus concentration peaked at 1.17 mg/L in 2013. From 2012 to 2015, the total phosphorus of the three monitoring points in the lower reaches was higher than 0.4 mg/L, which was inferior to Class V water. Similarly, after 2018, the total phosphorus in the Le’an River was within Class II values.

The discharge of industrial pollution from the Leping City economic belt into the Le’an River led the water quality of the lower reaches of the river to be inferior to Class V before 2015. After the 18th National Congress of the Communist Party of China, the water quality of Le’an River clearly improved, and after 2018, the water quality of the entire basin was within Class II, which demonstrated clear improvement. The 3D maps (Figure 3) of pollutant concentrations illustrate the changes in water quality.

The changes in NH$_4^+$ – N and TP concentrations at different monitoring of Le’an River from 2012 to 2020 were investigated. The line chart and 3D maps directly reflect the improvement in water quality since the 18th National Congress. To predict future nutrient time series, the Hurst index of each point was calculated by the R/S method (Figure 4 shows fitting curves) to explore the "long memory" of water quality [23, 31, 32].

Hanjiadu, which presented large fluctuation of ammonia nitrogen concentration, and Jiyang Bridge, which presented stable concentrations, were selected for the linear regression using log(n) pairs log((R/S)$_n$). The slope of the regression equation is the Hurst exponent H. The H values
were 0.921 and 0.795, respectively, which indicated that the nutrient sequence of each point showed strong persistence. Similarly, the total phosphorus indexes of Zhenqiao and Sandu were investigated, and their H values were 0.889 and 0.935, respectively. Therefore, the Hurst index values were greater than 0.5, that is, there was a strong autocorrelation in the time series. In the future, the water quality is expected to remain the same, at Class II or above, and the concentration of various pollutants is expected to fluctuate slightly.

**Figure 2.** H value fitted by some monitoring points.

### 4.2 Spatial Characteristics of Water Quality Indicators

After investigating the time scale characteristics of pollutants in the Le’an River, the spatial distribution characteristics of pollutants in the monitoring points were analysed, and a correlation test was conducted by calculating the Pearson correlation coefficient among monitoring points. The results are shown in Figures 5 and 6.
Figure 3. (a) $\text{NH}_4^+ - \text{N}$ and (b) TP correlation among nine monitoring points in the Le’an River (* $p$<0.05, ** $p$<0.01, *** $p$<0.001), DX and so on are short for each monitoring points.

The following conclusions were obtained from this analysis: (1) The correlation of $\text{NH}_4^+ - \text{N}$ was more significant than that of TP among the monitoring points [33, 34]. The correlation of ammonia nitrogen in upstream points was very strong, at above 0.8 ($p \leq 0.05$), and the correlation among Duan Xin, Sandu, and Qiukou Town was higher than 0.9 ($p \leq 0.001$). Although the correlation of total phosphorus was lower, the correlation was still significant. The reason for this was the narrow upper reaches of the Le’an River, which are characterised by fast water flow speed, which leads to good transmission. Moreover, there were not several major pollution sources nearby. (2) There was a strong correlation between downstream points. This likely occurred because the pollution in downstream points was serious, and the pollution sources are mostly point sources in the industrial economic belt, whereas non-point sources are less common. Moreover, there is a lack of monitoring and control measures for pollutants, and therefore, pollutants flow down the river basin, strongly contaminating downstream points. (3) Because there are many branches in the middle reaches, an accurate analysis is complex, and pollution derives from many different sources, rather than fixed point source pollution such as in the lower reaches; therefore, no strong correlation was observed. Adjacent monitoring points often presented a high correlation; the pollution in the Le’an River can therefore be reduced by controlling and monitoring pollutant discharge [35, 36].

5. Conclusions and Recommendations

In this study, the historical evolution of the water quality in Le’an River from 2012 to 2020 was visualised through line maps and 3D maps. In addition, the trend of future water quality change was reasonably predicted using the rescaling range method.

The results show that the concentrations of ammonia nitrogen and total phosphorus in downstream points of the Le'an river from 2012 to 2015 were all lower than the requirements of Class V quality. However, the water quality remarkably improved from 2015 to 2017, and the entire Le’an River Basin has been stable above Class II water quality since 2018. (2) The Hurst index fitted by each monitoring point was greater than 0.5, that is, there is a strong autocorrelation in the time series, and this trend is likely to continue in the future. (3) The correlation analysis of each monitoring point indicated that upstream and downstream points showed significant correlation among themselves.

These results demonstrate that the water pollution in the Le’an River is mainly related to the discharge of industrial wastewater and domestic sewage in Leping City [37-39]. This strong correlation also indicates that the quality of the entire Le’an River can be improved by controlling the discharge of exogenous pollutants. However, a time series with higher Hurst index processing frequency would be more accurate, as a long-time span log [(n)] decreases the accuracy of results. Therefore, pollutant source intensity should be calculated for a more accurate prediction and analysis.

Acknowledgments

Funding: This research was funded by the Fundamental Research Funds for the Central Universities (grant number 2018B53314) and the National Natural Science Foundation of China (grant numbers 51708543 and 51738013).

The authors would like to thank the Key Laboratory of Integrated Regulation and Resource Development on Shallow Lakes for its support.
References


[30] Labidi; Taher; Sakhrawi; Zaineb; Sellami; Asma; Mtibaa; Achraf; Bouassida; Nadia, On the use of OLS regression algorithm and Pearson correlation algorithm for improving the SLA establishment process in cloud computing. Innovations in Systems and Software Engineering 2022, (prepublish).


482