Analysis of the main factors affecting salinity in Wuliangsuhai Lake

Xiaomin Chang1,2,3, Shaoli Wang1,2,3*, Haorui Chen1,2,3, Xiaoyan Guan1,2,3

1 State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing, 100038, China
2 Department of Irrigation and Drainage, China Institute of Water Resources and Hydropower Research, Beijing, 100048, China
3 National Center of Efficient Irrigation Engineering and Technology Research, Beijing, 100048, China

*Corresponding author’s e-mail: shaoliw@iwhr.com

Abstract. The identification and determination of the main influencing factors of lake salinity play a key role in the control measures and suggestions to prevent lake pollution. To ascertain the salinity status, its evolutionary trend in WL, and its impact on the lake, data describing water quality and the 10 major influencing factors for the period 2003–2012 were employed to analyze the evolutionary process and trend characteristics of the main factors affecting upper irrigation and drainage. The principal component analysis (PCA) method was used to identify the key influencing factors of the salinity of the lake; then, regression analysis was combined with PCA modeling to predict the salinity. Results showed that using a multivariate statistical technique is an effective and objective method for identifying the main influencing factors of salinity in WL, which can provide a scientific basis for the pollution control and environmental management of WL.

1. Introduction

The Wuliangsuhai Lake (WL) is an important part of the irrigation and drainage system in the HIR of Inner Mongolia. It is the only drainage channel from which local return flow from agricultural lands, domestic sewage, and industrial wastewater can be accessed, and the pollution level is very high[1, 2]. Lake salinization is one of the indicators reflecting the level of water pollution in lakes. In recent years, water-saving irrigation has been vigorously conducted in the Hetao Irrigation Region (HIR), which has greatly reduced the water diversion and water discharge into WL. Moreover, rapid economic development has resulted in considerable municipal sewage and industrial wastewater flowing into the lake from the upper channels, while the ecological water supply in WL is insufficient. This increasing salinity in WL will continue to deteriorate water quality, threatening the health of the lake ecosystem.

The water environment of WL is affected by many factors; thus, it is especially important to understand the influence of anthropogenic activities upstream of the irrigation area on the salinity of WL. The analysis of the WL water environment can provide historical environmental information, while the analysis of the main influencing factors can reveal the sources and key influencing factors for WL. Therefore, further investigation on the salinization trend and main influencing factors of WL is necessary.

In this study, with WL of the HIR used as a case study, we combined the multivariate regression method and PCA to obtain regression equations for the total salinity of WL. In this method, data covering
the water quality and influencing factors in WL for the period 2003–2012 are used to develop a regression equation for the calculation of salinity. This study aims to: (1) investigate and analyze the changes in water quantity, salt content, and ecological water supply in WL during the study period using statistical methods, (2) relate the salinity value in WL to the major influencing factors using historical monitoring data and identify the main influencing factors for WL, and (3) construct a salinity prediction model for WL.

2. Materials and Methods

WL is located in the upper reaches of the Yellow River in northern China and covers a surface area of 371 km², making it the largest lake in the Yellow River valley[3,4] (Figure 1) average air temperature is 7.3 °C with an average of 3185 h of sunshine, 224 mm of annual rainfall, and 1502 mm of evaporation per year[5]. It is the only channel for irrigation and drainage in the Hetao area, receiving irrigation return flow from the HIR, which is composed, in part, of industry wastewater and municipal sewage from cities and towns. The wastewater from upstream of the irrigation area drains into the main drainage channel, the eighth drainage channel, and ninth drainage channel but then flows into WL. A total of 90% of the wastewater is drained into the main drainage channel. Tongji channel, Tabu channel, and Changji channel directly divert water from the Yellow River into WL[6].

3. Results and discussion

3.1. Principal component analysis

To understand the impact of irrigation and drainage development in the upper reaches of the HIR on the degree of mineralization in WL, an indicator of water quality, PCA was performed based on the water salinity data in WL from 2003 to 2012 and the main influencing factors in the corresponding years. In this study, 10 influencing factors, namely, rainfall, evaporation, groundwater mineralization, groundwater depth, total water diversion, water discharge from WL to the Yellow River, total drainage of the upper reaches of the HIR, salinity of seawater discharged into WL, wastewater discharge, and ecological water supplement to WL were selected and named as X₁, X₂,…,X₁₀, respectively. Then, the PCA was used to extract the main factors affecting the salinity of WL.

PCA and correlation analysis were used to identify the sources of WL salinity. The results of the varimax rotation on the 10 PCs, together with the amount of variance accounted for by each component, are summarized in Tables 1 and 2. With a higher loading of a certain variable, the variation in the other variables caused on a certain day by a particular PC is greater. As seen from Table 1, the first three PCs account for more than 82.9% of the total variation, indicating that they basically contain all the information of the 10 influencing factors. Therefore, the first three PCs were selected. The first PC can explain 44.4% of the information of the original variable, the most informative of all the PCs. The second PC can explain 27.0%, and the third principal component can explain 11.48%. The corresponding PC load matrix is shown in Table 2.
Table 1. Eigenvalue and variance contribution table

| No. | Initial eigenvalue (total variance) | Extract sum-of-squares load (total variance) |
|-----|------------------------------------|---------------------------------------------|
|     | variance % | accumulate % | variance % | accumulate % |
| 1   | 4.45       | 44.4         | 4.45       | 44.4         |
| 2   | 2.70       | 27.0         | 2.70       | 27.0         |
| 3   | 1.15       | 11.5         | 1.15       | 11.5         |
| 4   | 1.05       | 10.5         | 1.05       | 10.5         |

Table 2. Principal component load matrix

| variable | Ingredients | f₁ | f₂ | f₃ |
|----------|-------------|----|----|----|
| X₁       | 0.240       | -0.894 | 0.172 |
| X₂       | -0.282      | 0.698  | 0.248 |
| X₃       | -0.826      | -0.040 | 0.350 |
| X₄       | 0.202       | 0.584  | -0.662 |
| X₅       | -0.339      | 0.790  | 0.480 |
| X₆       | 0.979       | 0.035  | 0.148 |
| X₇       | 0.905       | -0.218 | 0.206 |
| X₈       | -0.932      | -0.199 | -0.125 |
| X₉       | 0.669       | 0.265  | 0.399 |
| X₁₀      | 0.613       | 0.535  | -0.165 |

According to the value of the eigenvector after centralization, the expressions for the PCs are listed as follows:

\[ f₁ = 0.1138ZX₁ - 0.1335ZX₂ - 0.3917ZX₃ + 0.0960ZX₄ + 0.1606ZX₅ + 0.4643ZX₆ + 0.4294ZX₇ - 0.4423ZX₈ + 0.3173ZX₉ + 0.2906ZX₁₀ \]  
\[ f₂ = -0.5442ZX₁ + 0.4248ZX₂ - 0.0243ZX₃ + 0.3557ZX₄ + 0.4809ZX₅ + 0.021ZX₆ + 0.1329ZX₇ - 0.1211ZX₈ + 0.1614ZX₉ + 0.3257ZX₁₀ \]  
\[ f₃ = 0.1609ZX₁ + 0.2315ZX₂ + 0.3262ZX₃ - 0.6177ZX₄ + 0.4480ZX₅ + 0.1385ZX₆ + 0.1919ZX₇ - 0.1162ZX₈ + 0.3723ZX₉ + 0.1541ZX₁₀ \]

In these expressions, \( ZX₁, ZX₂, \ldots, ZX₁₀ \) are the standardized variables of \( X₁, X₂, \ldots, X₁₀ \), respectively. According to the above linear expression and PC molecular load matrix (Table 2), the first PC has a greater load on the groundwater salinity (\( X₃ \)), discharge of WL into the Yellow River (\( X₆ \)), irrigation area drainage (\( X₇ \)), mineralization discharged into WL (\( X₈ \)), sewage discharge (\( X₉ \)), and supplemental ecological water (\( X₁₀ \)). Meanwhile, the second PC has a greater load on the rainfall (\( X₁ \)), evaporation capacity (\( X₂ \)), and water diversion volume (\( X₅ \)). The third PC has a greater load on groundwater depth (\( X₄ \)).

The variance of PC scores with time is shown in Figure 2. The first PC, \( f₁ \) shows a clear increasing trend, which may be related to changes in the irrigation area drainage and supplemental ecological water. With the implementation of the water-saving project, the drainage amount of the irrigation area decreases, and the discharge into WL decreases. Meanwhile, the supplemental ecological water amount of WL increases, and the influence of the mineralization of WL also gradually increases. However, the annual variation of the \( f₂ \) score fluctuates greatly, which can be attributed to the randomness of annual rainfall. For example, the significant change of the \( f₂ \) score in 2012 is due to the occurrence of a once-in-50-year rainstorm in Hetao in 2012, which caused the increase in the amount of WL discharged from the irrigation area; meanwhile, the larger rainfall had a certain dilution effect on WL, leading to a decrease in the salinity of WL.
3.2. Principal component regression prediction model

The PCA results of the PCA were used for PC regression analysis, and a stepwise regression option was applied to select which PCs to include in the regression equation with WL salinity as the dependent variable. The high loading variables were selected on the rotated PCs that were then used for inclusion in the ultimate regression model\(^7\).

According to the above Equations, the scores of \(f_1\)–\(f_3\) from 2003 to 2012 were calculated, respectively, and a regression analysis was conducted between \(f_1\)–\(f_3\) and the standardized dependent variable \(Z_Y\), finally converted into a relationship between the original independent variables and the dependent variable. First, \(Z_Y\) was stepwise regressed with the three PCs \(f_1, f_2,\) and \(f_3\). Only \(f_1\) was introduced into the model through a significance test of significance level \(a = 0.05\). According to the regression between the standardized dependent variable and the selected principal component \(f_1\), coefficient \(R\) is 0.708, indicating that 70.8\% of the variation of the degree of mineralization of WL can be explained by this model. The standardized principal component regression coefficient and test results are shown in Table 3.

\[
Z_Y = -0.353 - 0.462 f_1
\]

where \(Z_Y\) is the normalized value of the degree of mineralization of WL.

| Model       | Coefficient of non-standardization | Standard error | Standard coefficient | t     | Sig.    |
|-------------|------------------------------------|----------------|----------------------|-------|---------|
| (constant)  | -0.353                             | 0.267          | -1.318               | 0.224 |
| \(f_1\) score | -0.462                             | 0.163          | -0.708               | -2.837| 0.022   |

First, the coefficient matrix of the regression equation expressed by the standardized independent variables is obtained by a matrix composed of the coefficient vector of the PC and the coefficient vector estimator of PC regression.

\[
a = \begin{bmatrix}
0.1138 \\
-0.1335 \\
-0.3917 \\
0.0958 \\
-0.1606 \\
0.4644 \\
0.4294 \\
-0.4423 \\
0.3173 \\
0.2906
\end{bmatrix} \cdot [-0.462] = \begin{bmatrix}
-0.0526 \\
0.06169 \\
0.1809 \\
-0.0443 \\
0.0742 \\
-0.2146 \\
-0.1984 \\
0.2044 \\
-0.1466 \\
-0.1343
\end{bmatrix}
\]

(4)

The regression equation expressed by the standardized independent variable is

\[
Z_Y = -0.0526Z_{X_1} + 0.06169Z_{X_2} + 0.1809Z_{X_3} - 0.0443Z_{X_4} + 0.0742Z_{X_5} - 0.2146Z_{X_6} - 0.1984Z_{X_7} + 0.2044Z_{X_8} - 0.1466Z_{X_9} - 0.1343Z_{X_{10}} - 0.353
\]

(5)

Then, the standardized PC equation is transformed into a regression model of the original variable \(Y\) to \(X\).
using the PC regression model to restore the regression model of the original variable for prediction. Figure 3 shows the scatter diagram of the measured and predicted salinity in WL. To more intuitively express the prediction of water salinity, the measured value is compared with the predicted value, with an R^2 of 0.50, average absolute error (MAE) of 0.445 g/l, and root mean square error (RMSE) of 0.38 g/l. Overall, the change trend of the predicted value is consistent with the measured value. The results show that the prediction equation established by the PC regression model is representative.

\[
Y = -0.000778X_1 + 0.000348X_2 + 0.467717X_3 - 0.436615X_4 + 0.015844X_5
- 0.11429X_6 - 0.11624X_7 + 0.49703X_8 - 0.85967X_9 - 0.130921X_{10} + 2.231
\]  

Figure 3 Scatter diagram of the measured and predicted salinity in WL

In summary, the PC regression model can reduce the dimensionality of the problem and reduce the amount of information loss while retaining all the influencing factors. The PC regression analysis method can be used to establish the prediction model of the degree of mineralization in WL. Meanwhile, the degree of influence of each influencing factor on the water salinity can be intuitively understood through each factor coefficient of the model, and the main driving factors influencing the change in water quality of WL can be further analyzed to provide a relevant basis for the environmental protection of the downstream WL.

4. Conclusion

Salinity is considered as one of the main pollutants degrading the water environment. The objective of this study was to obtain accurate prediction models (i.e., models that depend on as few variables as necessary) for salinity in WL, with major influencing factors as predictor variables. Data spanning 10 years (2003–2012) were employed to analyze the effects of the upstream development of the irrigation area on the salinity of WL and develop regression equations for calculating the salinity of WL using both multiple linear and PC regression methods. The practical implication of this study is to identify factors that influence the salinity in WL such that these values can be described and predicted using mathematical models. In this study, the main factors influencing the degree of mineralization of WL were determined by PCA method; then, the PC regression prediction model of the degree of mineralization in WL was constructed. The main conclusions of this study are as follows:

1. Through PCA, four main controlling factors were obtained. The salinity of groundwater, amount of water discharged into the Yellow River from WL, drainage amount of the irrigation area, salinity discharged into WL, sewage discharge, and supplemental amount of ecological water have the greatest influence on the salinity of WL, followed by rainfall, evaporation, and water diversion.

2. The PC regression model of the degree of mineralization in WL was established, and R = 0.708. The MAE was 0.445 dS/m, and the RMSE was 0.38 dS/m. Generally, the change trends of the predicted and measured values were relatively consistent.

3. Using a multivariate statistical technique is an effective and objective method for identifying the main influencing factors of salinity in WL, which can provide a scientific basis for the pollution control and environmental management of WL.

Overall, these results clearly indicate the advantage of using irrigation and drainage data as well as correlation analysis to investigate the impact of the major influencing factors on the salinity in WL. The
validation and comparison conducted in this study successfully demonstrate the high accuracy of the regression equation, supporting the research.

**Acknowledgments**

This work was supported by the National Key R&D Program of China (Grant Nos. 2019YFC0409203, 2016YFC0502301, 2017YFC0403302), Basic Research Fund of China Institute of Water Resources and Hydropower Research (Grant No. ID0145B072020), China Postdoctoral Science Foundation funded project (Grant No. 2020M670386), Major Project in Key Research and Development Program of the Ningxia Hui Autonomous Region (Grant No. 2018BBF02022).

**References**

[1] Deng, Y.M.; Wang, Y.X.; Ma, T.; Yan, H.; He, J. Arsenic associations in sediments from shallow aquifers of northwestern Hetao Basin, Inner Mongolia. Environ. Earth. Sci. 2011, 64, 2001-2011.

[2] Feng, Z.Z.; Wang, X.K.; Feng, Z.W. Soil N and salinity leaching after the autumn irrigation and its impact on groundwater in Hetao Irrigation District, China. Agr. Water. Manage. 2005. 71, 131-143.

[3] Zhu, D.N.; Ryan, M.C.; Sun, B.; Li, C.Y. The influence of irrigation and Wuliangsuhai Lake on groundwater quality in eastern Hetao Basin, Inner Mongolia, China. Hydrogeol. J. 2014, 22, 1101-1114.

[4] Quan, D.; Shi, X.H.; Zhao, S.N. Zhang, S.; Liu, J.J. Eutrophication of Lake Ulansuhai in 2006-2017 and its main impact factors. J. Lake. Sci. 2019, 5, 71-79.

[5] Yu, R.H.; Liu, T.X.; and Shang, M.H. The Environment Variation and Analysis of Wuliangsuhai Wetland in Inner Mongolia Autonomous Region. Adv. Materials Res. 2012, 518-523, 1348-1354.

[6] Guo, X.J.; He, L.S.; Li, Q.; Yuan, D.H.; Deng, Y. Investigating the spatial variability of dissolved organic matter quantity and composition in Lake Wuliangsuhai. Ecol.Eng. 2014, 62, 93-101.

[7] Abdul-Wahab, S. A.; Bakheit, C.S.; and Al-Alawi, S.M. Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. Environ. Modell. Softw. 2005, 20, 1263-1271.