Identification of key factors affecting water quality concentration in the sluice-controlled river reaches of Shaying River in China via statistical analysis methods

Journal:	*Environmental Science: Processes & Impacts*
Manuscript ID:	EM-ART-03-2015-000130.R1
Article Type:	Paper
Date Submitted by the Author:	06-Jun-2015
Complete List of Authors:	Dou, Ming; Zhengzhou University, School of Water Conservancy & Environment Zhang, Yan; Zhengzhou University, School of Water Conservancy & Environment;Center for Water Science Research Zuo, Qiting; Zhengzhou University, School of Water Conservancy & Environment;Center for Water Science Research Mi, Qingbin; Zhengzhou University, School of Water Conservancy & Environment;Center for Water Science Research

Environmental impact statement

The Huaidian Sluice is located in Shenqiu County, Zhoukou City, which is in the middle reaches of the Shaying River and controls a catchment area of 28 150 km^2. This paper uses the river reach near the Huaidian Sluice in Shaying River, one of the badly polluted and highly regulated rivers of China, the key factors affecting the water quality concentration in the SCRRs were identified via statistical analysis methods, and the quantitative relationship between the water quality concentration change rate and key affecting factors was established, which could reduce water quality concentration disturbances due to SCRRs by adjusting these key affecting factors. The results provide guidance for preventing pollution and operating sluices in the Shaying River basin.

Identification of key factors affecting water quality concentration in the sluice-controlled river reaches of Shaying River in China via statistical analysis methods

*Ming Dou[1,2], Yan Zhang[1,2], Qiting Zuo[1,2],Qingbin Mi[1,2]

[1]School of Water Conservancy & Environment, Zhengzhou University, Zhengzhou, 450001, China

[2] Center for Water Science Research, Zhengzhou University, Zhengzhou, 450001, China

∗Corresponding Author: Ming Dou

douming@zzu.edu.cn

College of Water Conservancy & Environment
Zhengzhou University
No.100 of Kexue Road,
Zhengzhou, Henan, China

Abstract: The construction of sluices creates a strong disturbance in water environmental factors within a river. The change in water quality concentrations of sluice-controlled river reaches (SCRRs) is more complex than that of natural river segments. To determine the key factors affecting water quality concentration changes in SCRRs, river reaches near the Huaidian Sluice in the Shaying River of China were selected as a case study, and water quality monitoring experiments based on different regulating modes were implemented in 2009 and 2010. To identify the key factors affecting the change rates for the chemical oxygen demand of permanganate (COD_{Mn}) and ammonia nitrogen (NH_3-N) concentrations in the SCRRs of the Huaidian Sluice, a partial correlation analysis, principal component analysis and principal factor analysis were used. The results indicate four factors, i.e., the inflow quantity from upper reaches, opening size of sluice gates, water quality concentration from upper reaches, and turbidity before the sluice, are the common key factors for the COD_{Mn} and NH_3-N concentration change rates. Moreover, the dissolved oxygen before the sluice is a key factor for the COD_{Mn} concentration change rate, and the water depth before the sluice is a key factor for the NH_3-N concentration change rate. Multiple linear regressions between the water quality concentration change rate and key factors were established via multiple linear regression analyses, and the quantitative relationship between the COD_{Mn} and NH_3-N concentration change rates and key affecting factors was analyzed. Finally, the mechanism of action for the key factors affecting the water quality concentration changes was analyzed. The results reveal that the inflow quantity from upper reaches, opening size of sluice gates, COD_{Mn} concentration from upper reaches and dissolved oxygen before the sluice have a negative influence and the turbidity before the sluice has a positive influence for the COD_{Mn} concentration change rates and that the opening size of sluice gates, NH_3-N concentration from upper reaches, and water depth before the sluice have a negative influence and the inflow quantity from upper reaches and turbidity before the sluice have a positive influence for the NH_3-N concentration change rates, which provides a scientific grounding for pollution control and sluice operations in SCRRs.

Keywords: concentration change; key affecting factors; partial correlation analysis; principal component analysis; principal factor analysis; multiple linear analysis

1. Introduction

Dam and sluice construction impels social and economic development but has caused variations in hydrological elements in rivers and broken biological and chemical transport processes for river pollutants. Regulating dams and sluices can easily accumulate a large amount of sewage, which can seriously pollute the downstream water bodies when the gates are opened. These phenomena are particularly evident in sluice-controlled river reaches (SCRRs). Such water environmental deterioration has gained increasing attention with the continuous increase in the number and scale of dams and sluices, and many researchers have launched comprehensive and integrated interdisciplinary studies on the effects of dams and sluices.[1,2,3] In this study, the key factors affecting the water quality concentration in SCRRs in the Shaying River of China were identified via statistical analysis methods, and the quantitative relationship between the water quality concentration change rate and key affecting factors was established, which could help reduce water quality concentration disturbances due to SCRRs.

According to statistical data from the International Commission on Large Dams (ICOLD), 49248 dams (with dam heights above 15 m or storage capacities above 3000000 m^3) and 800000 sluices had been built in over 140 countries before 1998 with developing countries accounting for approximately two-thirds of the total number of dams.[4,5] There were approximately 2200 dams in China.[6,7] The increasing constructions of dams and sluices has weakened the natural quality and broken down the water environment and ecology, especially in basins with high population density, centralized production and living, and serious water pollution (i.e., Huaihe River Basin in China). Several paroxysmal water pollution incidents have occurred and could destroy the river ecological environment and affect the life and production of local residents if dam and sluice (reservoirs) scheduling was inappropriate.[8,9] Therefore, the influence that the polluted water environments has on rivers should be researched to better schedule dams and sluices, which will help improve dam and sluice dispatch management and reduce water disasters and accidents to provide a harmonious environment for both the people and the water.

The negative effects of sluice construction create a new challenge for exploiting and managing rivers and have placed great importance on preventing downstream pollution;[10] The water quality of Huai River Basin has mainly been influenced by point source pollution emission, flows regulated by dams, water temperature and land use variations, etc.[11] A modified geoaccumulation index, in which the regional background value was substituted with the predicted natural metal concentrations, was used to estimate heavy metal contamination in the Huaihe River.[12] Some research has detailed sluice stream discharge, channel structure, water environmental capacity, aquatic species and ecosystem diversity.[13,14,15,16] The influence of dam and sluice on the water environment in the river has been previously analyzed via numerical simulation techniques.[3,17] There are some paroxysmal water pollution incidents because of excessive construction and unreasonable dam and sluice scheduling, for example, the Aswan dam broke the ecological balance in the Nile basin.[18] Moreover, dam and sluice construction submerges a large amount of vegetation and land. The nitrogen, phosphorus and carbon content in the water increases because of the accelerated decomposition of organic matter; the greenhouse gas emission from the water surface also increases.[19,20]

There are a series of studies on model investigations, the impact of sluice scheduling and experiments for water quality have been carrying out in China. Lin (1995) [21] developed a new mathematical model based on the original mathematical model for the water quality of dams and sluices, considering the influence from different factors, such as the changes in water storage capacity and water quality along the river. Based on these sluice operation influences that including the changes of river runoff, flow velocity and water depth, the river water quality change was assessed according to setting different sluice operation scenarios.[22] Sewage drainage affects the downstream water quality, which was analyzed using the hydrological and water quality model;[23] a model based on the Soil and Water Assessment Tool was proposed to carry out the water quantity and quality simulation of the Huai River Basin by incorporating the operation rules of dams or sluices into the reservoir regulation module, and a multi-pollution source water quality model was integrated with Bayesian statistics to develop a robust method for supporting load reduction and effective water quality management in the Harbin City Reach of the Songhua River.[24,25] To determine the impact of dam and sluice scheduling from the actual operating condition of the Guazhou Sluice in Yangzhou city and the role of attenuation, the water quality was predicted from diffusion and pollutant excision under different sluice operating modes.[26] Experimental research includes model and field experiments, and the gate operation affects the water and pollutant migration and transformation when the water flow and gate change.[27] Spatial and temporal variations in polluted river water quality under various gate operating conditions were analyzed using several surveys and water quality monitoring at the Huaidian Sluice, and the operating mechanism for the water quality and quantity was explored.[28] Some research has explored theories, key topics, and methods to analyze the hydrological and environmental effects of dams and sluices on natural river characteristics (i.e., water quantity and quality) and proposed a quantitative framework to study and simulate water cycles on the river basin scale.[29] Chen et al. (2014)[30] also performed field experiments and found that dam operation had a different impact on water pollutants, suspended solids and sediments, in different media and promoted transformation of pollutants.

Overall, current studies have mainly focused on the effects of sluice operation on the water environment based on hydrodynamics and hydrology and usually considered SCRRs as an internal boundary condition. There are fewer reported research studies on how to identify key factors affecting the water quality concentration change rate, especially the factors driving water quality change in the complex hydrodynamic processes for SCRRs. In this study, the river reach near the Huaidian Sluice in the Shaying River, a seriously polluted and frequently regulated river in China, is selected as a typical area. This research first analyzes the partial correlation coefficients for the factors affecting the COD_{Mn} and NH_3-N concentration change rates via a partial correlation analysis before identifying key factors affecting the water quality concentration change rate in SCRRs via principal component analysis and principal factor analysis. Finally, the key factors are quantitatively related to the COD_{Mn} and NH_3-N concentration change rates via multiple linear regression analysis. The results provide guidance for preventing pollution and operating sluices in the Shaying River Basin.

2. Materials and Methods

2.1 Study area

The Huaidian Sluice is located in Shenqiu County, Zhoukou City, which is in the middle reaches of the Shaying River and controls a catchment area of 28 150 km^2. It contains shallow-hole gates (frequently regulated), deep-hole gates (regulated only during flooding season), and one ship-lock gate (rarely used). The shallow-hole gates were built in 1959 and collectively have eighteen 6-meter-wide holes. The deep-hole gates were built in 1969 and collectively have five 10-meter-wide holes. The design flood flow rate of the Huaidian Sluice is 3200 m^3/s (flow rate for a 20-year flood), and the checking flow is 3500 m^3/s (flow rate for a 200-year flood). The normal irrigation water levels range from 38.50 m to 39.50 m, and the highest irrigation water level is 40.00 m. The normal water storage is 3.0×10^7 to 3.7×10^7 m^3, and the maximum water storage is 4.5×10^7 m^3. The shallow-hole gates usually maintain a small-flow discharge to avoid accumulating polluted water in front of the sluice during the dry season; the deep-hole gates are used to release flood waters only during flood season, and the ship-lock gate is used for navigation.[10]

2.2 Experiments and data

Two water quality monitoring experiments were performed in the river reaches near the Huaidian Sluice to obtain first-hand data for this study. For the first experiment, performed in March 2010, the water quality and bottom sediment monitoring samples were obtained under three scheduling modes using the present scheduling, opening decreases and opening increases for the Huaidian Sluice. There were six total monitoring sections (I, II, III, IV, V and VII**)** and 12 sampling points, as shown in Fig. 1. Three systematic samplings and three supplementary samplings were performed and three bottom sediment samples and 39 water samples were collected to test the main water quality index data (i.e., pH and turbidity) during this experiment; there are six water samples in the section I and VII respectively, three water samples in the section II and III respectively, twelve water samples in the section IV, and nine water samples in the section V. A second experiment was conducted on October 2010. There were two scheduling modes, all gate opening and centralized discharge, which selected two different openings for each scheduling mode. There were seven total monitoring sections (I, II, III, IV, V, VI and VII**)** and 15 sampling points as shown in Fig. 1. Five systematic samplings and four supplementary samplings were conducted and 99 water samples were collected to test the main water quality index data (i.e., water temperature, pH and turbidity) during this experiment; there are nine water samples in the section I and VII respectively, nineteen water samples in the section II, IV, V and VI respectively, and five water samples in the section III. Additionally, some monitoring data were collected in 2009.

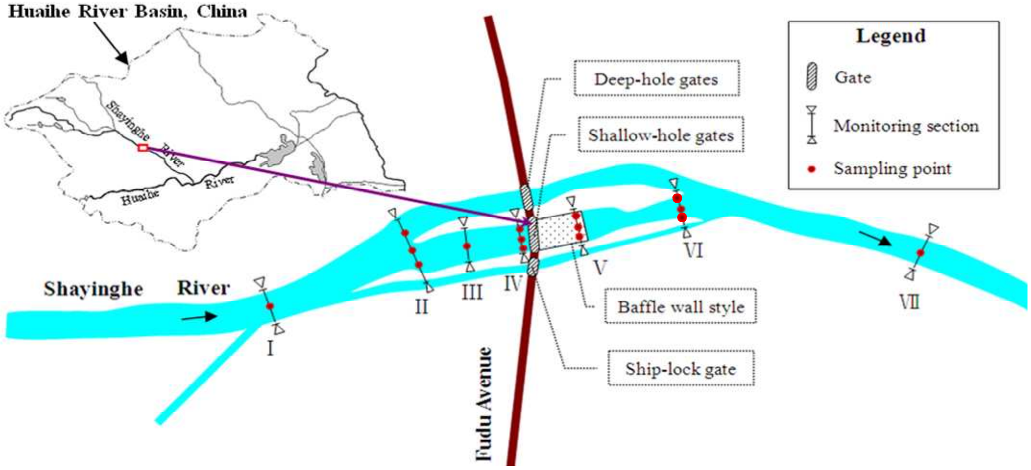


Fig. 1 Sketch of the river reaches near the Huaidian Sluice during the experiments

Three factor sets were used as analytical objects. The first factor set is the hydrological factors that mainly consider temporal and spatial variations in the river hydrology and includes inflow quantity from upper reaches, flow velocity before the sluice and water depth before the sluice that they can directly reflect the scouring action of the mud at the river bottom. The second factor set is the water environmental factors that mainly consider environmental influences on the migration and transformation of pollutants in the river and contains water quality concentration from the upper reaches, water temperature, pH, turbidity before the sluice and dissolved oxygen before the sluice. The water temperature and pH can indirectly reflect pollutant degradation in the water, and the turbidity can indirectly reflect pollutant resuspension from the bottom sediments, and the dissolved oxygen before the sluice can indirectly affect the chemical process of the pollutant. The last factor set is the regulating factors that mainly consider the comprehensive effect from the flow process caused by sluice operation and includes the opening sizes and the number of sluice gates. The factors affecting the water quality concentration and their abbreviations and units are shown in table 1.

Table 1 Factors affecting water quality concentration and their abbreviations and units

Affecting factors	Abbreviation	Unit
COD_{Mn} concentration change rate	$COD_{Mn}CR$	%
NH_3-N concentration change rate	NH_3-NCR	%
Inflow quantity from upper reaches	*IQU*	m^3/s
Opening size of sluice gates	*OSG*	cm
Opening number of sluice gates	*ONG*	number units

COD_{Mn} concentration from upper reaches	$COD_{Mn}CU$	mg/L
NH_3-N concentration from upper reaches	NH_3-NCU	mg/L
Dissolved oxygen before the sluice	*DOS*	mg/L
Water temperature	*WT*	℃
Turbidity before the sluice	*TS*	NTU
pH value	pH	pH units
Flow velocity before the sluice	*FVS*	cm/s
Water depth before the sluice	*WHS*	m

Note: COD_{Mn} is chemical oxygen demand of permanganate; NH_3-N is ammonia nitrogen

The COD_{Mn} and NH_3-N concentration change rates in the SCRRs were used to indicate the influence on the water quality concentrations. The COD_{Mn} and NH_3-N concentration change rates were calculated as follows:

$$\lambda = \frac{(C_{low} - C_{up})}{C_{up}} \tag{1}$$

where C_{up} and C_{low} are the inflow and outflow COD_{Mn} and NH_3-N concentrations, respectively, for the SCRRs.

2.3 Methods

In this study, first, the partial correlation analysis between the water quality concentration change rate and the affecting factors was analyzed, and the partial correlation coefficients were obtained. Second, the key affecting factors for the water quality concentration change rate were identified via principal component analysis and principal factor analysis. Finally, the quantitative relationship between the water quality concentration change rate and key factors were established via multiple linear regression analysis.

2.3.1 Partial correlation analysis

A partial correlation analysis is based on the partial correlation coefficients for observation data and determines which independent variables are important to the dependent variable. A partial correlation relies on all of the usual assumptions for Pearson correlations: quantitative variables, linear relationships, same relationship degree throughout the independent variable range, and data with an untruncated range.[31] The recursion formula for partial correlation coefficients is as follows:

$$r_{0i,12\cdots(i-1)(i+1)\cdots p} = \frac{r_{0i,12\cdots(i-1)(i+1)\cdots(p-1)} - r_{0p,12\cdots(p-1)}\, r_{ip,12\cdots(i-1)(i+1)\cdots(p-1)}}{\sqrt{1-r^2_{0p,12\cdots(p-1)}}\sqrt{1-r^2_{ip,12\cdots(i-1)(i+1)\cdots(p-1)}}} \qquad (2)$$

where $r_{0i,12\cdots(i\text{-}1)(i+1)\cdots p}$ is the p-1 order partial correlation coefficient, $r_{0i,12\cdots(i\text{-}1)(i+1)\cdots(p\text{-}1)}$ is the p-2 order partial correlation coefficient, $r_{0p,12\cdots(p\text{-}1)}$ is the p-2 order partial correlation coefficient between dependent variables and independent variables, $r_{ip,12\cdots(i\text{-}1)(i+1)\cdots(p\text{-}1)}$ is the p-2 order partial correlation coefficient for independent variables.

The partial correlation analysis method was used to analyze the correlation between two variables and ignore the third variable when both variables were simultaneously associated to the third variable. Independent variables with greater influences on the dependent variable must be selected as the essential factors; however, those with less influence can be eliminated. After determining the levels of influence, we only considered the dominant factors and described the dependent variable changes using as few independent variables as possible.[10] According to this principle, we selected one of ten influences and controlled the other factors to analyze its correlation to the COD_{Mn} and NH_3-N concentration change rates in this study.

2.3.2 Principal component analysis (PCA) and principal factor analysis (PFA)

PCA is a useful multivariate statistical method for reducing, manipulating, and visualizing complex data systems when patterns and data similarities are poorly understood.[32,33,34] The principal components (PCs) are uncorrelated variables obtained by multiplying the original correlated variables with an eigenvector (loadings or weightings). The eigenvalues for the PCs measure their associated variance. The participation of the original PC variables is given by the loading, and the individually transformed observations are called scores.[35,36,37,38] PCA was performed on normalized (z-scale transformation) variables for the parameters after sorting highly correlated variables from the data sets. PCs with eigenvalues above 1 were retained.[39] The principal component (PC) can be expressed as

$$z_{ij} = pc_{i1}x_{1j} + pc_{i2}x_{2j} + \cdots + pc_{im}x_{mj} \qquad (3)$$

where z is the component score, pc is the component loading, x is the measured value, i is the component number, j is the sample number and m is the total number of variables.

PFA follows PCA. The main purpose of PFA is to reduce the contribution of less significant variables to further simplify the PCA data structure by rotating the axis defined by PCA according to well-established rules and constructing new variables called varifactors (VF).[40] PCA was performed on the normalized variables to extract significant PCs and further reduce the contribution of variables with minor significance; these PCs were subjected to varimax rotation to generate VFs.[41,38,42] The PFA can be expressed as

$$y_{ij} = f_{j1}z_{i1} + f_{j2}z_{i2} + \cdots + f_{jm}z_{im} + e_{ij} \quad (4)$$

where y is the measured variable, f is the factor loading, z is the factor score, e is the residual term accounting for errors or other variations, i is the sample number and m is the total number of factors.[43]

2.3.2 Multiple linear regression analysis (MLR)

To model the relationships between the dependent and independent variables, a multiple liner regressions model was applied to examine the impact of the COD_{Mn} and NH_3-N concentration change rates for the key affecting factors. Multiple linear regression analysis is a statistical tool for relating two or more variables. Multiple linear regressions examine the relationship between a single dependent variable and a set of independent variables to best represent their relationship in the population.[40] This technique is used for both predictive and explanatory purposes via experimental or nonexperimental designs.[43] For an arbitrary number of explanatory variables, the linear regression model takes the following form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m + e_{ij} \quad (5)$$

where Y is the dependent variable, X_1, X_2,…X_m are the different independent variables, β_0, β_1, … β_m are the regression coefficients, and e is the random error.[44]

3. Results and Discussion

3.1 The partial correlation analysis results

The partial correlation analysis results for the COD_{Mn} and NH_3-N concentration change rates (dependent factors) and ten affecting factors (independent variables) are shown in Table 2.

Table 2 The partial correlation analysis results for the $COD_{Mn}CR$ and NH_3-NCR and affecting factors

Affecting factors for $\mathrm{COD_{Mn}}$*CR*	*IQU*	*OSG*	*ONG*	$\mathrm{COD_{Mn}}$*CU*	*DOS*	*WT*	*TS*	pH	*FVS*	*WHS*
Partial correlation coefficient	-0.354	-0.406	-0.396	-0.861	-0.863	0.738	0.537	0.49	0.11	-0.647
Significance level	0.107	0.245	0.257	0.01	0.01	0.02	0.11	0.15	0.762	0.043
Affecting factors for $\mathrm{NH_3}$-N*CR*	*IQU*	*OSG*	*ONG*	$\mathrm{NH_3}$-N*CU*	*DOS*	*WT*	*TS*	pH	*FVS*	*WHS*
Partial correlation coefficient	0.257	-0.608	0.246	-0.535	-0.581	0.555	0.855	0.736	0.686	-0.373
Significance level	0.623	0.2	0.638	0.275	0.227	0.253	0.03	0.095	0.132	0.467

The $\mathrm{COD_{Mn}}$*CR* has a better partial correlation coefficient for $\mathrm{COD_{Mn}}$*CU* and *DOS*, at -0.861 and -0.863, respectively, followed by *WT* and *WHS*, whose partial correlation coefficients are 0.738 and -0.647, respectively. The significance of these four factors for the $\mathrm{COD_{Mn}}$*CR* are below 0.05, which passes the significance tests. The next five most correlated affecting factors, *IQU*, *OSG*, *ONG*, *TS* and pH, have partial correlation coefficients for the $\mathrm{COD_{Mn}}$*CR* of -0.354, -0.406, -0.396, 0.537 and 0.49, respectively, but their significance levels are above 0.05, which indicates that the difference between these five affecting factors and the $\mathrm{COD_{Mn}}$*CR* are considerable, and the correlations are thus weaker. The last affecting factor, *FVS*, has a poor correlation to the $\mathrm{COD_{Mn}}$*CR*, 0.11, and the significance fails the significance test. Moreover, *IQU*, *OSG*, *ONG*, $\mathrm{COD_{Mn}}$*CU*, *DOS* and *WHS* are negatively correlated to the $\mathrm{COD_{Mn}}$*CR*. The other factors exhibit positive correlations.

The $\mathrm{NH_3}$-N*CR* has the highest partial correlation coefficient with *TS* at 0.855, with a significance below 0.05. The next three most correlated affecting factors, *OSG*, pH and *FVS*, exhibit partial correlation coefficients of -0.608, 0.736 and 0.686, respectively. The next three most correlated affecting factors, $\mathrm{NH_3}$-N*CU*, *DOS* and *WT*, also have nonzero correlations to the $\mathrm{NH_3}$-N*CR*, -0.535, -0.581 and 0.555, respectively, which indicates that the differences between these three factors and $\mathrm{NH_3}$-N*CR* is considerable, and the correlations are thus weaker. The other affecting factors, *IQU*, *ONG* and *WHS*, are poorly correlated to the $\mathrm{NH_3}$-N*CR*, 0.257, 0.246 and -0.373, respectively. The significance levels do not pass the significance test except for *TS*. Furthermore, *OSG*, $\mathrm{NH_3}$-N*CU*, *DOS* and *WHS* are negatively correlated to the $\mathrm{NH_3}$-N*CR*. The other affecting factors exhibit positive correlations.

These factors, such as *OSG*, $\mathrm{COD_{Mn}}$*CU*, $\mathrm{NH_3}$-N*CU*, *DOS* and *WHS*, indicate a negative and immediate influence on water quality concentration in SCRRs via the partial correlation analysis, which is most likely because the SCRRs are short, and thus, the time for pollutants to biochemically react is limited. The partial correlation analysis was a preliminary analysis of the affecting factors. The key factors affecting the water quality concentration and the quantitative relationship between the water quality concentration and these key factors required further analysis.

3.2 The principal component analysis and principal factor analysis results (PCA / PFA)

According to the PCA/PFA, the Kaiser-Meyer-Olkin (KMO) is 0.722 for the factors affecting the COD_{Mn} concentration change rate, and the KMO is 0.649 for those affecting the NH_3-N concentration change rate. The PCA/PFA is very suitable when the KMO is above 0.8; that is suitable when the KMO is above 0.6 and less than 0.8; that is less suitable when the KMO is above 0.5 but below 0.6; and is not suitable when the KMO is below 0.5. The concomitant probability is 0.000 for the Bartlett's test of sphericity, and the significance is at 0.05, which is suitable for PCA/PFA.[45] Therefore, these results are suitable for the PCA/PFA on the factors affecting the COD_{Mn} and NH_3-N concentration change rates.

3.2.1 The correlation of affecting factors

Table 3 Correlation matrices for the factors (bold figures indicate significance at $p < 0.05$)

Affecting factors for COD_{Mn}*CR*	*IQU*	*OSG*	*ONG*	COD_{Mn}*CU*	*DOS*	*WT*	*TS*	pH	*FVS*	*WHS*
IQU	1									
OSG	**0.863**	1								
ONG	0.242	0.034	1							
COD_{Mn}*CU*	**-0.756**	**-0.661**	0.242	1						
DOS	**0.698**	**0.627**	-0.315	**-0.984**	1					
WT	**0.596**	**0.605**	-0.308	**-0.832**	**0.874**	1				
TS	**0.767**	**0.769**	-0.221	**-0.894**	**0.887**	**0.759**	1			
pH	0.174	0.013	**0.421**	0.139	-0.198	**-0.469**	-0.091	1		
FVS	**0.705**	**0.641**	0.167	-0.558	**0.518**	**0.398**	0.561	0.244	1	
WHS	**0.571**	0.509	**-0.440**	**-0.834**	**0.844**	**0.800**	**0.725**	-0.111	0.419	1

Affecting factors for NH_3-N*CR*	*IQU*	*OSG*	*ONG*	NH_3-N*CU*	*DOS*	*WT*	*TS*	pH	*FVS*	*WHS*
IQU	1									
OSG	**0.830**	1								
ONG	-0.069	-0.312	1							

NH_3-N*CU*	**-0.817**	**-0.752**	**0.506**	1						
DOS	**0.828**	**0.740**	**-0.492**	**-0.997**	1					
WT	**0.825**	**0.733**	**-0.491**	**-0.990**	**0.996**	1				
TS	**0.817**	**0.839**	**-0.443**	**-0.875**	**0.869**	**0.864**	1			
pH	**-0.755**	**-0.631**	0.310	**0.789**	**-0.787**	**-0.791**	**-0.795**	1		
FVS	**0.577**	**0.540**	-0.149	**-0.569**	**0.576**	**0.569**	**0.538**	**-0.530**	1	
WHS	**0.535**	**0.456**	**-0.714**	**-0.854**	**0.868**	**0.886**	**0.693**	**-0.650**	0.349	1

Table 3 shows the correlation matrices for the factors affecting the COD_{Mn} and NH_3-N concentration change rates based on the principal component analysis. In general, the strongest factors are obviously significant (bold figures indicate significance for $p < 0.05$ in table 3), and the correlation is closer for each affecting factor.

For factors affecting the COD_{Mn}*CR*, *OSG*, COD_{Mn}*CU*, *DOS*, *TS* and *WHS* exhibited correlation coefficients of 0.863, -0.756, 0.698, 0.767 and 0.705 for *IQU*, respectively, there was a negative correlation between *IQU* and COD_{Mn}*CU*, and there was a positive correlation between *IQU* and the other parameters. Increasing *IQU* rapidly increases the water before the sluice, which strongly disturbs the water before the sluice; *DOS*, *TS* and *WHS* exhibit increasing trends. Such correlations change slightly with a positive correlation for *DOS* (0.627), *TS* (0.769), and *FVS* (0.641) and a negative correlation for COD_{Mn}*CU* (-0.661) with *OSG*. However, a strong negative correlation existed for *DOS* (-0.984), *WT* (-0.832), *TS* (-0.894), and *WHS* (-0.834) with COD_{Mn}*CU*, which indicates a negative influence between COD_{Mn}*CU* and *DOS*, *WT*, *TS* and *WHS.* The same but positive phenomenon was observed for *DOS* and *WT*, *TS*, *WHS*, which exhibited correlation coefficients of 0.874, 0.887 and 0.844. A good correlation was found for *WT* and *WHS* (0.800) and *TS* (0.759). The correlation coefficient between *DOS* and *WHS* was better, 0.725, which indicated that *DOS* increased as *WHS* increased.

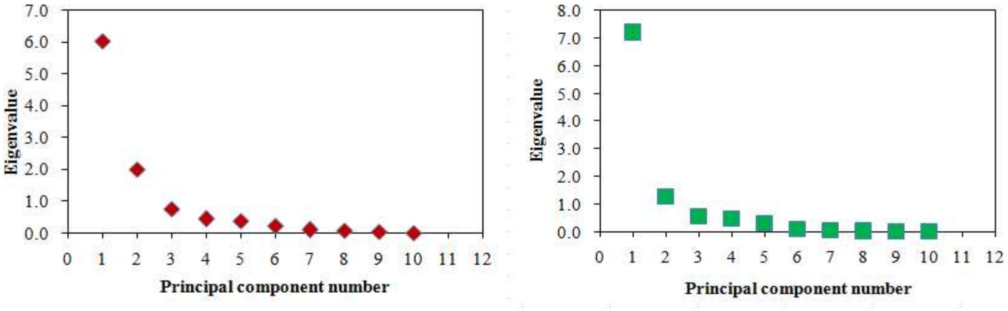
For the factors affecting the NH_3-N*CR*, *IQU* had a stronger correlation coefficient than the other parameters, i.e., *OSG*, NH_3-N*CU*, *DOS*, *WT*, *TS* and pH, whose correlation coefficients were 0.830, -0.817, 0.828, 0.825, 0.817 and -0.755, respectively, with a negative correlation between *IQU* and NH_3-N*CU* and pH; however, the opposite influence occurred between *IQU* and the other parameters. *OSG* exhibited the highest correlation coefficient with *TS*, 0.839. Such correlations changed slightly with a positive correlation for *DOS* (0.740) and *WT* (0.733) and a negative correlation for NH_3-N*CU* (-0.752) and pH (-0.631) for *OSG*. *ONG* had a negative correlation for

WHS, -0.714. There was a strong correlation between NH_3-N*CU* and *DOS*, NH_3-N*CU* and *WT*, NH_3-N*CU* and *TS*, and pH and *WHS* were, -0.997, -0.990, -0.875, 0.789 and -0.854, respectively, which were all negative influences except for pH. *DOS* had the best correlation coefficient with *WT*, *TS*, pH and *WHS*, at 0.966, 0.869, -0.787 and 0.868, respectively, which indicated that *DOS* increased when *WT*, *TS* and *WHS* increased; however, *DOS* decreased with increasing pH. A strong positive correlation was found between *WT* and *TS* (0.864) and *WHS* (0.759); however, a negative correlation (-0.791) existed between *WT* and pH. The *DOS* changed slightly with a negative correlation for pH (-0.795) and a positive correlation for *WHS* (0.693). The correlation coefficient between pH and *WHS* was -0.650.

3.2.2 The principal factor analysis results for the affecting factors

Eigenvalues are normally used to determine the number of principal components (PCs) to retain for further PCAs.[46] A scree plot for the eigenvalues from this study shows a pronounced slope change after the second eigenvalue (Fig. 2). Therefore, the first two PCs were analyzed further. These two PCs have eigenvalues above 1 and explain 80.15% and 84.69% of the total variances in the original dataset for the COD_{Mn} and NH_3-N concentration change rates, respectively. These PCs contain more information than the original affecting factors, which can reduce the affecting factor dimensions.

The component loadings are linear combinations for each principal component and correlate the original affecting factors and newly formed components. The component loadings can be used to determine the relative importance of an affecting factor relative to other affecting factors in a PC and do not reflect the importance of the component itself.[46]



(a) affecting factors of COD_{Mn}*CR* (b) affecting factors of NH_3-N*CR*

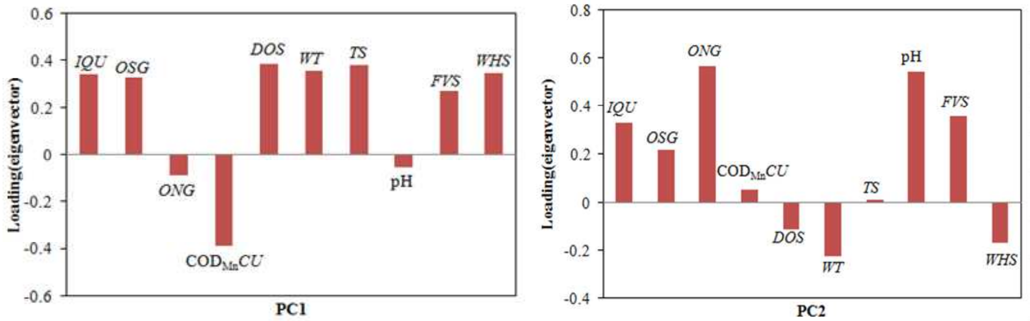
Fig. 2 Scree plot of principal component eigenvalues for the factors affecting the COD_{Mn}*CR* and NH_3-N*CR*

Component loadings for the first two retained PCs for COD_{Mn}*CR* and NH_3-N*CR* are shown in Fig. 3. The principal component (PC1) shown in Fig. 3 (a) explained 60.16% of the total affecting factors and was positive, largely controlled by hydrological factors (i.e., *IQU* and *WHS*), a regulation factor (i.e., *OSG*) and water environmental factors (i.e., *DOS*, *WT* and *TS*), and negatively affected by a water environmental factor (i.e., COD_{Mn}*CU*). Therefore, this component contains three affecting factors, such as the hydrological factor, regulation factor and water environmental factor. This component also indicated that the *ONG*, pH and *FVS* were less important for PC1; however, *ONG* and pH exhibited slight positive influences, and *FVS* exhibited a slight negative influence.

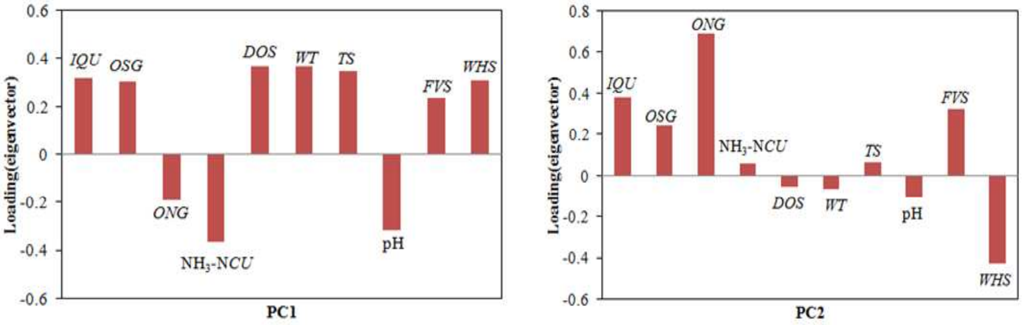
PC2 in Fig. 3 (b) explained 19.99% of the total affecting factors and was positive, largely controlled by *IQU*, *ONG*, pH and *FVS*, and negatively controlled by *WT*. This component indicates the less important hydrological (i.e., *WHS*), regulation (i.e., *OSG*) and water environmental (i.e., COD_{Mn}*CU*, *DOS* and *TS*) factors; there was a positive influence for *OSG*, COD_{Mn}*CU* and *TS*; however, *DOS* and *WHS* exhibited negative influences.

In contrast to the PC1 for COD_{Mn}*CR*, the PC1 for factors affecting the NH_3-N*CR* are shown in Fig. 3 (c), explained 72.07% of the total affecting factors, was positively controlled by hydrological (i.e., *IQU* and *WHS*), regulation (i.e., *OSG*) and water environmental (i.e., *DOS*, *WT* and *TS*) factors and was negatively controlled by water environmental factors (i.e., NH_3-N*CU* and pH). This component also indicates that *ONG* and *FVS* were less important to the total effect on the NH_3-N*CR* because their loading coefficients were low; *ONG* exhibited a negative influence, but *FVS* exhibited a positive influence.

PC2 in Fig. 3 (d) accounted for 12.62% of the total affecting factors, was positively influenced by *IQU*, *ONG* and *FVS* and was negatively affected by *WHS*, which were important factors for PC2. PC2 shows the less important regulation (i.e., *OSG*) and water environmental (i.e., NH_3-N*CU*, *DOS*, *WT*, *TS* and pH) factors. *OSG*, NH_3-N*CU* and *TS* exhibited positive influences; however, *DOS*, *WT* and pH exhibited negative influences.



(a) PC1 for $COD_{Mn}CR$ (b) PC2 for $COD_{Mn}CR$



(c) PC1 for NH_3-N*CR* (d) PC2 for NH_3-N*CR*

Fig. 3 Component loading for the first component (PC1) and second component (PC2) for factors affecting the $COD_{Mn}CR$ and NH_3-N*CR*

3.2.3 Identification of key factors affecting the $COD_{Mn}CR$ and NH_3-N*CR*

As shown in Fig. 3, the PC1 and PC2 for all of the $COD_{Mn}CR$ and NH_3-N*CR* were influenced (negatively and positively) by most of the affecting factors. There are some affecting factors that are more important than the others in influencing the change rates of the COD_{Mn} and NH_3-N concentration are hidden. Therefore, the PFA must circumvent the factor data ambiguity.

Data exhibiting rotated correlation coefficients for the first two factors for the $COD_{Mn}CR$ and NH_3-N*CR* are shown in Table 4. The first two analyzed factors accounted for 80.2% and 84.7% of the total affecting factors for the COD_{Mn} and NH_3-N concentration change rates, respectively. The other 8 factors accounted for lower percentages and exhibited low or insignificant correlation coefficients. Any factors with an absolute correlation coefficient value >80% was considered a key factor that contributes to the COD_{Mn} and NH_3-N concentration change rates in this study.

Table 4 shows the key factors affecting water quality concentration. The hydrological (i.e., *IQU*), regulatory (i.e., *OSG*), and water environmental (i.e., $COD_{Mn}CU$, *DOS* and *TS*) factors are the key factors affecting the COD_{Mn} concentration change rate. The hydrological (i.e., *IQU* and

WHS), regulatory (i.e., *OSG*) and water environmental (i.e., NH_3-N*CU* and *TS*) factors were identified as the key factors affecting the NH_3-N concentration change rate.

Table 4 Rotated factor correlation coefficients for the factors affecting the COD_{Mn}*CR* and NH_3-N*CR*

Affecting factors for COD_{Mn}*CR*	Factor1	Factor2	Affecting factors for NH_3-N*CR*	Factor1	Factor2
IQU	0.937	0.336	*IQU*	0.987	0.082
OSG	0.821	0.136	*OSG*	0.807	0.217
ONG	-0.053	0.720	*ONG*	-0.069	-0.764
COD_{Mn}*CU*	-0.920	0.286	NH_3-N*CU*	-0.802	-0.603
DOS	0.896	-0.383	*DOS*	0.784	0.601
WT	0.778	-0.501	*WT*	0.776	0.611
TS	0.901	-0.183	*TS*	0.812	0.433
PH	0.001	0.597	*PH*	-0.738	-0.367
FVS	0.697	0.302	*FVS*	0.610	0.14
WHS	0.751	-0.416	*WHS*	0.410	0.894

3.3 The results of multiple linear regression analysis (MLR)

3.3.1 The MLR analysis between the COD_{Mn}*CR* and the key affecting factors

The multiple linear regression equation (Equation 6) relates the COD_{Mn}*CR* and key affecting factors. The R^2 and significance level were 0.525 and 0.058, respectively. The adjusted R^2 and significance level are not high; however, Equation 6 quantitatively relates the COD_{Mn}*CR* and the key affecting factors. The relationship between the actual and fitted curves and the algebraic difference for the actual and fitted value (D-value) of the water quality concentration change rate are shown in Fig. 4.

$$Y = 208.12 - 0.112X_1 - 0.011X_2 - 24.913X_3 - 17.258X_4 + 1.667X_5 \quad (6)$$

where Y is the COD_{Mn}*CR*; X_1 is *IQU*; X_2 is *OSG*; X_3 is COD_{Mn}*CU*; X_4 is *DOS*; and X_5 is *TS*.

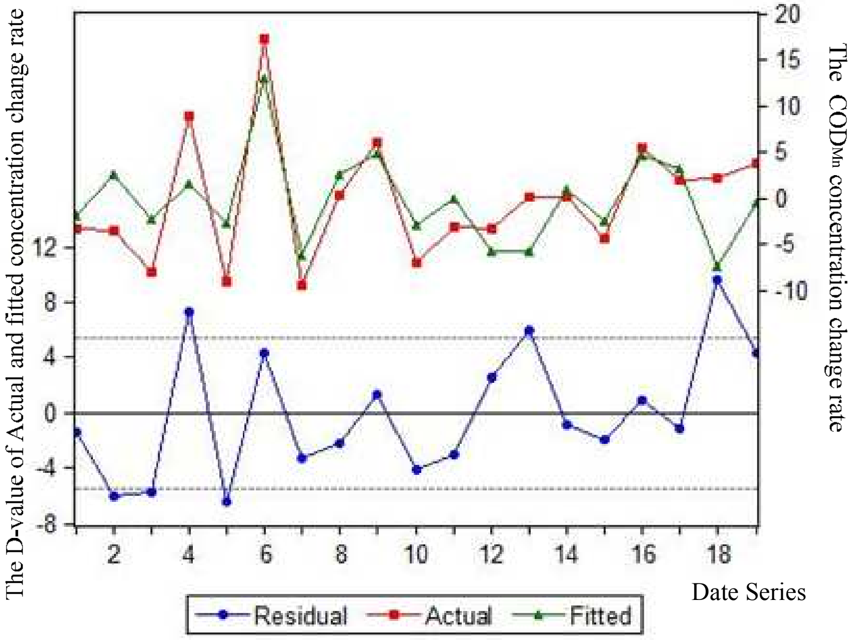


Fig. 4 The fitting figure for the MLR between the key factors and COD_{Mn}*CR*

The key factors affecting the COD_{Mn} concentration change rate contain three factor sets, such as hydrological, regulation and water environmental factors, which indicates that the process change for the COD_{Mn} concentration in the SCRRs was disturbed because of the combined influence of the local hydrological conditions, water quality conditions and sluice operation. Equation 6 indicates that the COD_{Mn} concentration change rate decreased when *IQU*, *OSG*, COD_{Mn}*CU* and *DOS* increased and increased when *TS* increased, which is the same as the partial correlation analysis result. The scouring action strengthens the mud at the river bottom when *IQU* and *OSG* increase, and resuspension dominates the COD_{Mn} in the water, which resuspends endogenous pollution, increases the COD_{Mn} concentration behind the sluice, and decreases the COD_{Mn} concentration change rate. The COD_{Mn} concentration in front of the sluice increases when the COD_{Mn}*CU* increases; however, the combined influence is smaller, which decreases the COD_{Mn} concentration change rate. The aerobic role strengthens and the degradation rate is enhanced when *DOS* increases, which decrease the COD_{Mn} concentration behind the sluice and the COD_{Mn} concentration change rate. Increasing the *TS* decreases the COD_{Mn} concentration before the sluice and increases the COD_{Mn} concentration change rate, and vice versa.

3.3.2 The MLR analysis between NH_3-N*CR* and the key affecting factors

The multiple linear regression equation (Equation 7) relates the NH_3-N*CR* and key affecting factors. The adjusted R^2 and significance are 0.646 and 0.058, respectively. The adjusted R^2 and significance are not much higher, but equation 7 also quantitatively relates the NH_3-N concentration change rate and the key affecting factors. The actual and fitted relationship curves and the algebraic difference for the actual and fitted value (D-value) of the water quality concentration change rate are shown in Fig. 5.

$$Y = 50.84 + 0.032X_1 - 0.371X_2 - 2.179X_3 + 4.538X_4 - 25.219X_5 \tag{7}$$

where Y is the NH_3-N*CR*; X_1 is *IQU*; X_2 is *OSG*; X_3 is NH_3-N*CU*; X_4 is *TS*; X_5 is *WHS*.

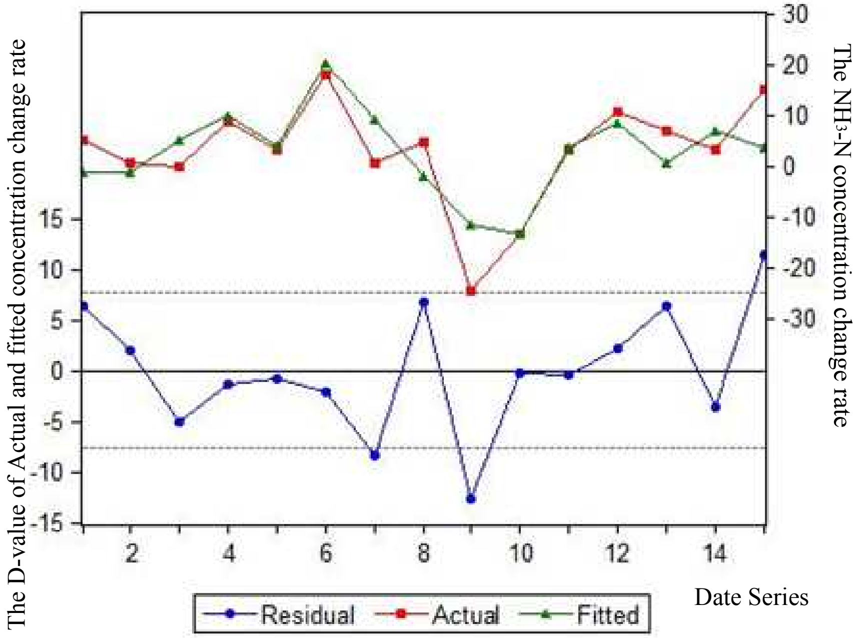


Fig. 5 The fitting figure of the MLR between the key factors and NH_3-N*CR*

The key factors affecting the NH_3-N*CR* also contain three factor sets, including the hydrological, regulation and water environmental factors. Equation 7 indicates that the NH_3-N concentration change rate positively increases with *IQU* and *TS* and is negatively correlated to *OSG*, NH_3-N*CU* and *WHS*, which is the same result obtained from the partial correlation analysis. Water discharged through the sluice increases in the SCRRs, and self-purification before the sluice is enhanced with increasing *IQU*, which decreases the NH_3-N concentration before the sluice and increases the NH_3-N concentration change rate. The NH_3-N concentration before the sluice decreases with *TS*, which increases the NH_3-N concentration change rate. Moreover, because water discharge via the sluice increases in the SCRRs and self-purification behind the sluice is enhanced with increasing *OSG*, the NH_3-N concentration decreases behind the sluice, and the NH_3-N concentration change rate decreases. The NH_3-N concentration before the sluice increases when the NH_3-N*CU* increased; however, the combined influence decreases, which decreases the NH_3-N concentration change rate. The scouring action strengthens the mud in the bottom of the river when more water discharges through the sluice due to *WHS* increasing, and the NH_3-N concentration behind the sluice decreases because the NH_3-N is more soluble in water, which decreases the NH_3-N concentration change rate, and vice versa.

4. Conclusions

This study used monitoring data from two field experiments and other periods in 2009. The relationships between affecting factors and water quality concentration change were analyzed for the sluice-controlled river reaches of the Shaying River using partial correlation, principal component, principal factor and multiple linear regression analyses, and the key affecting factors were identified. Several concluding remarks from this research are as follows. (1) The COD_{Mn} concentration change rate exhibited higher partial correlation coefficients with COD_{Mn}*CU* and *DOS*, -0.861 and -0.863, respectively, which indicates that the COD_{Mn}*CU* and *DOS* influenced the COD_{Mn} concentration change rate more. The NH_3-N concentration change rate exhibited the highest partial correlation coefficient with *TS*, 0.855, which indicates that *TS* most influenced this rate for the partial correlation analysis. (2) *IQU*, *OSG*, COD_{Mn}*CU*, *DOS* and *TS* were the key factors affecting the COD_{Mn} concentration change rate; *IQU*, *OSG*, NH_3-N*CU*, *TS* and *WHS* were the key factors affecting the NH_3-N concentration change rate based on the principal component and principal factor analyses. (3) The multiple linear regression equation relates the water quality concentration change rate to the key affecting factors, as shown by Equation 6 and 7. The action mechanism for the key factors changing the water quality concentration was analyzed.

Acknowledgements

The research was supported by the Natural Sciences Foundation of China (No. U1304509 and 51279183), the Program for Innovative Research Team (in Science and Technology) in

University of Henan Province (No. 13IRTSTHN030), and the Scientific and Technological Project of Zhengzhou City, China (No. 083SGYG26122-6 and 0910SGYS33389-2).

References

1 Geoffrey, E.P., Angela, M.G., Dams and geomorphology: Research progress and future directions, Geomorphology, 2005, 71 (1-2), 27-47.

2 Mallika, A.U., Richardson, J.S., Riparian vegetation change in upstream and downstream reaches of three temperate rivers dammed for hydroelectric generation in British Columbia, Canada, Ecological Engineering, 2009, 35 (5), 810-819.

3 Newham, L.T.H., Letcher, R.A., Jakeman, A.J., et al., Integrated water quality modeling: Ben Chifley Dam Catchment, Australia, International Environmental Modelling and Software Society, 2002, (1), 275-280.

4 McCully P., Silenced Rivers: The Ecology and Politics of Large Dams, London and New Jersey: ZED books, 1996.

5 Tharme R.E., A global perspective on environmental flow assessment: Emerging trends in the development and application of environmental flow methodologies for rivers, River Research and Applications, 2003, 19 (5-6), 397-441.

6 Postel S., Richter B., Rivers for life: Managing water for people and nature, Washington D.C.: Island Press, 2003.

7 The World Commission on Dams (WCD), Dams and Development: A New Framework for Decision-Making, London and Sterling, VA: Earthscan Publications Ltd, 2000.

8 Zuo, Q., Gao, Y., Liu, Z., Analysis and discussion about the mechanisms of the impacts of dams on water quality and quantity of heavily polluted rivers, Resources Science, 2010, 32 (2), 261-266.

9 Cui, K., Gao, J., Zuo, Q., et al., Impact assessment of dams and water gates on water quality and water quantity of rivers, Journal of Yangtze University (Nat Sci Edit), 2011,8 (6), 12-14.

10 Dou, M., Li, G., Li, C., Quantitative relations between chemical oxygen demand concentration and its influence factors in the sluice-controlled river reaches of Shaying River, China, Environ Monit Assess, 2015, 187, 4139.

11 Zhai, X., Xia, J., Zhang, Y., Water quality variation in the highly disturbed Huai River Basin, China form 1994 to 2005 by multi-statistical analyses, Science of the Total Environment, 2014, 496, 594-606.

12 Wang, J., Liu, G., Lu, L., et al., Geochemical nomalization and assessment of heavy metals (Cu, Pb, Zn and Ni) in sediments from the Huaihe River, Anhui, China, Gatena,2015, 129, 30-38.

13 Hayes, D.F., Labadie, J.W., Sanders, T. G., et al., Enhancing water quality in hydropower system operations, Water Resources Research, 1998, 34 (3), 471-483.

14 Brandt, S. A., Classification of geomorphological effects downstream of dams, Catena, 2000, 40 (4), 375-401.

15 Petts, G.E., Gurnell, A.M., Dams and geomorphology: Research progress and future directions, Geomorphology,

2005, 71 (1 /2), 27-47.

16 Mallika, A.U., Richardson, J.R., Riparian vegetation change in upstream and downstream reaches of three temperate rivers dammed for hydroelectric generation in British Columbia, Canada, Ecological Engineering. 2009, 35 (5), 810-819.

17 Marcé, R., Moreno-Ostos, E., García-Barcina, J.M., et al., Tailoring dam structures to water quality predictions in new reservoir projects: assisting decision-making using numerical modeling, Journal of Environmental Management, 2010, 91 (6), 1255-1267.

18 Powell K., Open the floodgates, Nature, 2002, 420, 356-358.

19 Petts G., 1984. Impounded Rivers: perspectives for ecological management. New York: Wiley, Chichebster.

20 Fearnside, P.M., Greenhouse gas emissions from hydroelectric dams: Controversies provide a springboard for rethinking a supposedly “clean” energy source, Climatic Change, 1984, 66, 1-8.

21 Lin, W., Water quality simulation of dam-Controlled river with case study in Huaihe river, Pollution Control Technology, 1995, 8 (4), 233-236.

22 Zheng, B., Dou, M., Huang, L., et al., Analysis of sluice operation impact on diversification in water quality, Environmental Science and Technology, 2012, 35 (2), 14-18.

23 Zhang, Y., Xia, J., Wang, G., et al., Research on influence of dams’ union dispatch on water quality in Huaihe River Basin, Engineering Journal of Wuhan University, 2007, 40 (4), 31-35.

24 Zhang, Y., Xia, J., Shao, Q., et al., Water quantity and quality simulation by improved SWAT in highly regulated Huai River Basin of China, Stochastic Environmental Research & Risk Assessment, 2013, 27(1), 11–27.

25 Zhao, Y., Ashish S., Bellie S., et al., A Bayesian method for multi-pollution source water quality model and seasonal water quality management in river segments, Environmental Modelling & Software, 2014, 57, 216-226.

26 Zhu, W., Zheng, X., Zhu,W., Water quality prediction for channals controlled by sluices and dams, Advances in Science and Technology of Water resource, 1998, 18 (1), 49-51.

27 Ruan, Y., Zhang, X., Zhang, Y., Experimental research on the impaction of gate operation on contamination transfer in stream channel, China Rural Water and Hydropower, 2009, (7), 52-54.

28 Liu, Z., Zuo, Q., Zhao, G., et al., Experiment of impacts of gate dispatching on water quality of polluted river, Journal of Water Resources and Water Engineering, 2011, 22 (5), 34-37.

29 Zhang, Y., Xia, J., Zhai, X., The hydrologic-environmental effects of dams and sluices and the assessment frameworks, Progress in Geography,2013, 32 (1), 105-113.

30 Chen, H., Zuo, Q., Dou, M., et al., Comprehensive experimental research on impacts of dam operation on water environment of polluted river, Acta Scientiae Circumstantiae, 2014, 4 (3), 763-771.

31 Garson, G.D. Partial Correlation on StatNotes. Online: http://faculty.chass.ncsu.edu/garson/PA765/partialr.htm#assume. Downloaded: May 1 (2010).

32 Melloul, A., Collin, M., The ‘principal components’ statistical method as a complementary approach to

geochemical methods in water quality factor identification; application to the coastal plain aquifer of Israel, J Hydrol, 1992, 140 (1), 49–73.

33 Stauffer, D.F., Garton, E.O., Steinhors,t R.K., A comparison of principal components from real and random data, Ecology ,1985, 66 (6), 1693–1698.

34 Valder, J.F., Long, A.J., Davis, A.D., et al., Multivariate statistical approach to estimate mixing proportions for unknown end members, J Hydrol, 2012, 460, 65–76.

35 Vega, M., Pardo, R., Barrado, E., et al., Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis, Water Research, 1998, 32, 3581–3592.

36 Helena, B., Pardo, R., Vega, M., et al., Temporal evolution of groundwater composition in an alluvial (Pisuerga river, Spain) by principal component analysis, Water Research, 2000, 34, 807–816.

37 Wunderlin, D.A., Diaz, M.P., Ame, M.V., et al., Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (Cordoba- Argentina), Water Research, 2001, 35, 2881–2894.

38 Singh, K. P., Malik, A., Mohan, D., et al., Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study, Water Research, 2004, 38, 3980–3992.

39 S.Y. Chung, S. Venkatramanan, N. Park, et al., An assessment of selected hydrochemical parameter trend of the Nakdong river water in South Korea, using time series analyses and PCA, Environ Monit Assess, 2015, 187, 4192.

40 Rabia, K., Bulent, S., Bayram, T., Water quality assessment using multivariate statistical methods-A case study: Melen River System (Turkey), Water Resour Manage, 2010, 24, 959-978.

41 Brumelis, G., Lapina, L., Nikodemus, O., et al., Use of an artificial model of monitoring data to aid interpretation of principal component analysis, Environ Model Softw, 2000, 15, 755-763.

42 Singh, K.P., Malik, A., Sinha, S., Water quality assessment and apportionment of pollution sources of Gomti River (India) using multivariate statistical techniques-a case study, Anal Chim Acta, 2005, 538, 355-374.

43 Tinsley, E.A.H., Brown, S.D., Handbook of applied multivariate statistics and mathematical modeling. Academic, San Diego, 2000.

44 Freund, R.J., Wilson, W.J., Regression analysis—statistical modeling of a response variable, Academic Press, 1998.

45 Du, Q., Jia, L., SPSS statistical analysis from entry to the master, Beijing, Posts and Telecom Press, 2009.

46 Y. Ouyang, P., Nkedi-Kizza, Q.T., WU, et al., Assessment of seasonal variations in surface water quality, Water Research,2006, 40, 3800-3810.