

An improved two-step parameter adjustment method for the optimization of a reservoir operation function model based on repeated principal component analysis and a genetic algorithm

Xuan Wang, Xiao Chen, Quan Cui and Zhifeng Yang

ABSTRACT

This paper presents an improved two-step parameter adjustment method for the construction of a reservoir operation function model, by using repeated principal component analysis (PCA) and a genetic algorithm (GA) to optimize parameters in conventional multiple regression models. The first step is to use repeated PCA, to exclude the co-linear parameters in a multiple regression expression reflecting relationships among possible impact factors of reservoir operation, so as to form an initial reservoir operation function model. The second step is to use a GA to optimize the model constructed in the first step, and to compare its effects with other regression methods. The results show that the proposed reservoir operation function model can produce better results, which correlate water volume for power generation, input discharge, water level, and ecological flow. Compared with established scheduling schemes, the optimized scheme increases the water volume for power generation by $1.06 \times 10^9 \text{ m}^3/\text{yr}$, and the optimized result generates an increase in economic benefits of $3.22 \times 10^7 \text{ yuan/yr}$ (i.e., $4.69 \times 10^6 \text{ USD/yr}$).

Key words | genetic algorithm, improved multiple regression, repeated principal component analysis, reservoir operation, two-step parameter adjustment

Xuan Wang (corresponding author)
Xiao Chen
Quan Cui
Zhifeng Yang
State Key Laboratory of Water Environment
Simulation,
Beijing Normal University,
Beijing 100875,
China
E-mail: wangx@bnu.edu.cn

Xuan Wang
Zhifeng Yang
Key Laboratory for Water and Sediment Sciences
of Ministry of Education,
Beijing Normal University,
Beijing 100875,
China

Zhifeng Yang
Beijing Engineering Research Center for Watershed
Environmental Restoration and Integrated
Ecological Regulation,
Beijing 100875,
China

INTRODUCTION

Owing to climate change and intensive human activities, river runoff faces reduction, and hydrological extremes occur more frequently. This causes water scarcity worldwide (Gain & Giupponi 2015), affecting the sustainable development of the social economy. Reservoirs have an important role in water conservation by regulating supply through storage, alleviating regional water shortages. They can prevent flooding, and can provide high volumes of water for irrigation and power generation. Reservoirs are particularly useful for runoff regulation; by storing floodwater for dry seasons, natural water can meet the requirements of the water sectors, allowing efficient regional water resource utilization and social economic growth. Thus, efficient reservoir

operation is of interest to governments and to researchers (Li *et al.* 2013).

Scheduling problems relate to reservoirs' ecological environment. Reservoir operation changes water level, which causes environmental damage to the water-level-fluctuating zone and its surroundings. Meanwhile, changes in the release of water from the dam, and instream flow downstream of the dam determine whether the ecological water requirement is adequate. The problems of reservoir operation are also socio-economic issues; the volume of incoming water, the water quantity required for irrigation and power generation directly affect the social and economic benefits and there is a trade-off of interests among

water sectors. Reservoir operation rules guide the use of reservoirs, and promote coordinated development of industrial and agricultural production, national economy, and ecological environment (Bai *et al.* 2015). Therefore, the study of optimized reservoir dispatching methods not only has general theoretical value, but also has practical significance.

A discrete stochastic dynamic programming model was first applied to reservoir operation in the 1940s (Liu 2006). Existing reservoir models can help to plan reservoir operation, providing efficient distribution of the water use and the water storage (Bai *et al.* 2015). Current methods for formulating reservoir operation rules mostly use a long-term runoff dataset. A deterministic optimization method is used to solve the optimal reservoir scheduling problem, then a variety of mathematical methods (e.g., regression analysis) are used to fit functions to the results, which then guide the operation of the reservoir (Liu *et al.* 2006). The reservoir operation function model is a scheduling tool which can incorporate a variety of scheduling information, and it usually reflects the relationship among some reservoir operation elements, such as surface runoff, reservoir storage capacity (or water level), and water discharge (or the output of hydropower stations), and so on. In general, the conventional regression analysis requires parameters to remain independent. However, parameters of the scheduling function often exhibit high linear correlation with each other, such as water volume for power generation and water discharge, ecological water demand and water discharge downstream of the dam. If all of these parameters are used in the regression analysis, then an unstable regression equation will be produced due to the existence of the co-linear parameters (Abdul-Wahab *et al.* 2005). Therefore, it has become the consensus that it is necessary to exclude unwanted interference from co-linear parameters when establishing a reservoir operation function model (Labadie 2004; Farrokhnia & Karimib 2016).

In recent years, the commonly used methods to solve co-linear problems in multiple regression have included ridge regression, partial least squares (PLS) regression, and principal component analysis (PCA) regression. Ridge regression, in essence, is a modified least squares estimation method. It can obtain relatively realistic and reliable regression coefficients at the expense of partial information loss and

accuracy reduction. In addition, it has certain subjectivity, since ridge parameters are determined according to experience (Xie & Hawkes 2015). The PLS regression is to find the best function relationship for a set of data by minimizing the sum of errors squared. It can be seen as a combination of the least squares regression and principal components (De Paulo *et al.* 2016). The outstanding characteristic of PLS is that it can implement regression modeling, predigest the data structure and analyze the correlation between two groups of parameters simultaneously with the same arithmetic. It brings great convenience to multiple linear regression analysis. However, ordinary PLS cannot get satisfactory results for dealing with the problem that the number of parameters are more than that of sample points (Song 2009). As a widely used dimensionality reduction technique, the PCA is a multiple statistical method, which detects the correlation among multiple parameters and has a data compression effect. It can reveal the simple data structure involved in the complex data (Çamdevýrena *et al.* 2005). In spite of the advantages of easy operation, repeated use of PCA may be needed in order to achieve a good effect for solving co-linear problems when there are linear correlations among regression parameters.

In addition, the results obtained from the regression methods for solving the reservoir operation function model may be inaccurate, because they omit some important constraints. Researchers have found that intelligent algorithms can be used to extract reservoir operation rules and solve optimizations, saving time and avoiding redundant work (Neelakantan & Pundarikanthan 2000; Chandramouli & Deka 2005; Liu *et al.* 2006). To be specific, intelligent algorithms can overcome the limitations of traditional optimization methods in dealing with complex problems, such as problems regarding multi-objective, uncertainty, non-linearity, discontinuousness, and discreteness in water resources and reservoir scheduling (Simonovic 2009). As a kind of widely used intelligent algorithm, the genetic algorithm (GA) directly uses the objective function value to globally search the problem space, and can improve solution accuracy and convergence; thus, it is suitable for solving reservoir operation function models. For example, through comparative analysis, Vasan & Srinivasa (2009) analyzed three methods including simulated annealing, simulated quenching, and genetic algorithms for optimal reservoir

operation, and found that all these non-traditional optimization techniques can be utilized for efficient planning of any irrigation system with suitable modifications. In addition, Chang *et al.* (2010) researched constrained genetic algorithms for optimizing multi-use reservoir operation and concluded that they would be powerful tools in searching for the optimal strategy. Considering the ease of using repeated PCA and the GA's ability to quickly approach optimal solutions for complex system problems, it is advantageous to jointly use these methods to optimize parameters in conventional multiple regression models and finally get a more accurate model.

Therefore, the objective of this research is to construct a new two-step parameter adjustment method to exclude the co-linear parameters in multiple regression expression, and get an optimized reservoir operation function model. This objective entails the following tasks: (1) combine repeated PCA with a multiple regression model, to produce an initial reservoir operation function model; (2) use GA to optimize parameters in that initial scheduling function to get a more accurate model for supporting reservoir operation management. After carrying out two-step adjustments to the reservoir operation function model, the co-linear factors in

the reservoir scheduling data can be effectively eliminated, and reasonable relationships among the decision variable and the impact factors are reflected more accurately. Therefore, this paper provides a scheduling model resulting in better economic efficiency and ecological benefits, producing optimal operation of the reservoir.

STUDY AREA AND DATA COLLECTION

Danjiangkou Reservoir is in the upper reaches of Hanjiang River in central China, and is located across the waters of Hubei and Henan provinces (see Figure 1). It is Asia's largest artificial freshwater lake. Danjiangkou Dam was constructed in early 1973. Covering an area of 745 km², the reservoir adjoins Danjiangkou City, Yun County, Yunxi County, and Shiyan City of Hubei Province and Xichuan County of Henan Province, covering a total area of 15,900 km². Danjiangkou's designed water level is 157 m, and the corresponding storage capacity is 17.45 billion m³. The place at which the reservoir is located has a north subtropical monsoon climate, with a mild climate and abundant rainfall. The average temperature is

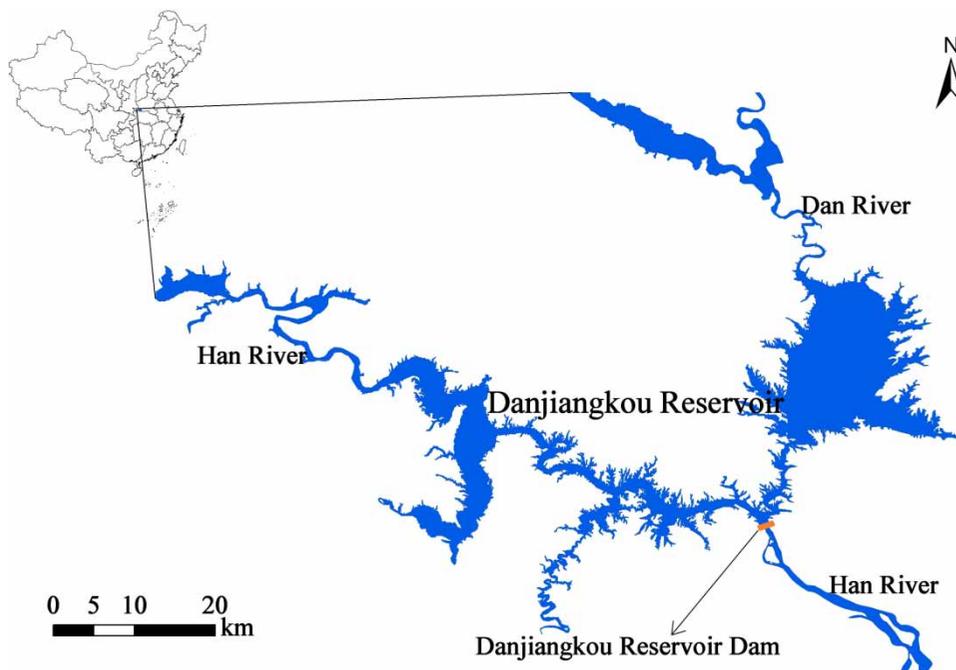


Figure 1 | Study site of Danjiangkou Reservoir, China.

15.9 °C, while the average annual rainfall of the reservoir area is 850–950 mm, and the annual water-surface evaporation is 860 mm. The annual water-inflow of the reservoir is an average of 39.48 billion m³; the water comes from the Han River and its tributary, Dan River. Danjiangkou Reservoir’s runoff frequency distribution and the dam’s

discharged water from 1980 to 2005 are shown in Figures 2 and 3. After the heightening project of the Danjiangkou Dam was built in 2005, both the top elevations of the concrete dam and the embankment dam were increased from 162 m to 176.6 m; the normal water level of the reservoir was increased from 157 m to 170 m. This resulted in an increase of the reservoir area from 745 km² to 1,050 km² and the reservoir capacity from 1.75 × 10⁸ m³ to 2.91 × 10⁸ m³.

Danjiangkou Reservoir is responsible for flood control, power generation, water supply, and shipping affairs. As a resource for the South–North Water Transfer Project in China, the reservoir will provide domestic water, industrial water, and agricultural water for more than 20 cities around Beijing City, Tianjin City, Henan and Hebei Provinces in the future. The total installed capacity of the hydroelectric power plants in Danjiangkou is currently 900,000 kilowatts. The cumulative energy generation was up to 141.82 billion kwh (at the end of 2011), and each kilowatt-hour of electricity could create 3.26 yuan (i.e., 0.48 USD) for Hubei Province’s gross domestic product (GDP) in 1990; the cumulative power generation can hit 429.70 billion yuan (i.e., 62.59 billion USD) of GDP.

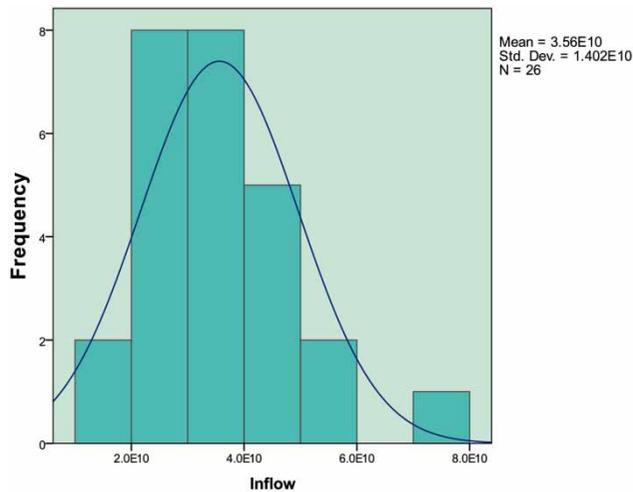


Figure 2 | Danjiangkou Reservoir frequency distribution from 1980 to 2005 (m³/yr).

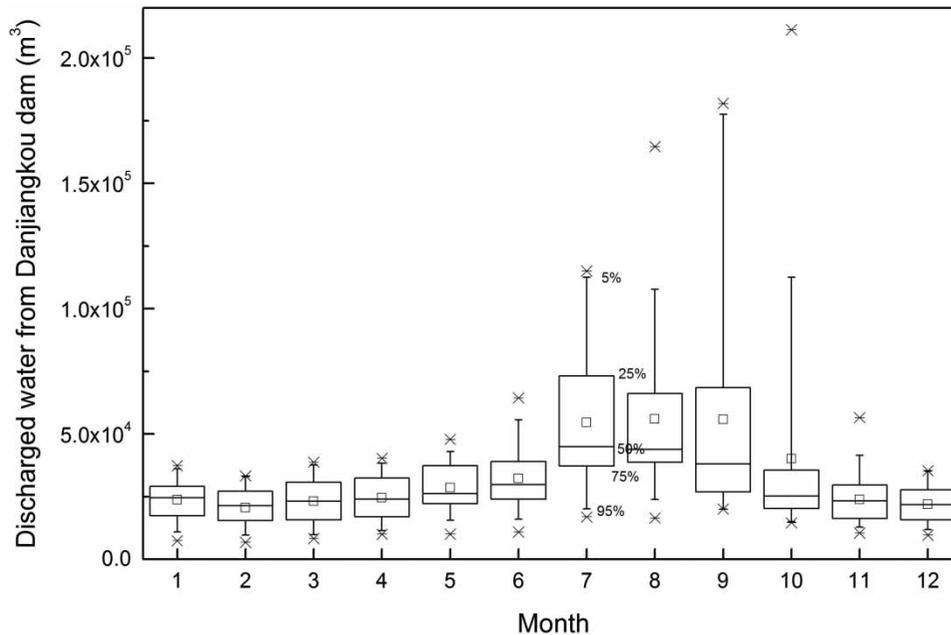


Figure 3 | Long-term monthly discharged water for the Danjiangkou Dam.

In this study, a total of 26 years of daily average inflow data of Danjiangkou Reservoir from 1980 to 2005 was collected as time series data for the optimization analysis of reservoir operation function model. Related hydrological, meteorological and reservoir data were derived from the South–North Water Transfer Project Construction Committee under China’s State Council, and local water authorities.

METHODOLOGY

Flowchart for establishing an optimized reservoir operation function model

The main task of reservoir operation is to increase economic and environmental benefits, and thus the utilization benefits of the reservoir as much as possible by coordinating the relations of upstream runoff, socio-economic water use, and the discharged ecological flow. Therefore, the water volume for power generation is chosen as the decision variable for reservoir operation, and other parameters (e.g., reservoir inflow, water level, and ecological flow) are used as its impact factors to build the reservoir operation function model. Then, the model parameters are optimized with the proposed two-step parameter adjustment method. The process is designed as follows (see Figure 4): (1) collect data, including historical hydrological conditions, and reservoir performance data over the years; (2) implement Step 1 to adjust parameters: apply repeated PCA to a multiple regression model to establish an initial reservoir operation function model; and (3) implement Step 2 to adjust parameters: use GA to optimize the initial reservoir operation function model, and achieve better economic benefits in the premise of ensuring the ecological water demand. Meanwhile, how the optimized scheduling rules can impact the objective function is compared.

First step of improved parameter adjustment method

By using repeated PCA combined with a multiple regression method, an initial reservoir operation function model can be established. First, through a multiple

regression method, a mathematical formula between the decision variable of reservoir operation and its impact factors can be set up:

$$V_D = \gamma_1 \cdot a_1 + \gamma_2 \cdot a_2 + \gamma_3 \cdot a_3 + \cdots + \gamma_{m-1} \cdot a_{m-1} + \gamma_m \quad (1)$$

where V_D is decision variable related to reservoir operation (e.g., the water volume for power generation), $a_1, a_2, a_3, \cdots, a_{m-1}$ are parameters relating to reservoir operation (e.g., reservoir inflow, water level, ecological water demand, reservoir outflow, surplus water, stored water, and the water volume for power generation), and $\gamma_1, \gamma_2, \gamma_3, \cdots, \gamma_{m-1}, \gamma_m$ are regression coefficients.

Then, a repeated PCA is used to eliminate collinearity of Equation (1). The PCA can extract several major components of the original parameters, including as much information as possible to preserve the original parameter information while keeping them independent (Petersen *et al.* 2001). There is a relationship between the parameters for the reservoir operation schedule. For example, storage capacity is a function of the water level; ecological water demand downstream of the dam is calculated by using the historical data for annually discharged water; and the water volumes for power generation and discharge are usually linear in the case of very small amounts of discarding water from the reservoir to the river. If all of these parameters are used in the regression analysis, it will produce an unstable regression equation, which is highly influenced by the co-linear parameters. Thus, all parameters in the regression equation are first of all analyzed by using PCA repeatedly to decide if there is collinearity, and to eliminate some co-linear parameters. The first step is to carry out first PCA on all of the system’s parameters (including both dependent and independent parameters) to get the principal component 1 (PC 1); then, to keep the most contributive parameter from the group of parameters that have the most significant contribution to PC 1 (i.e., excluding the remaining parameters that are highly relevant to PC 1 in this group, which are also called the excluded parameters). The second step is to carry out a second PCA on all parameters except for the excluded parameters to get the current PC 1, then to keep the most contributive parameter from the group of parameters that have the most significant contribution on this PC 1, excluding remaining parameters

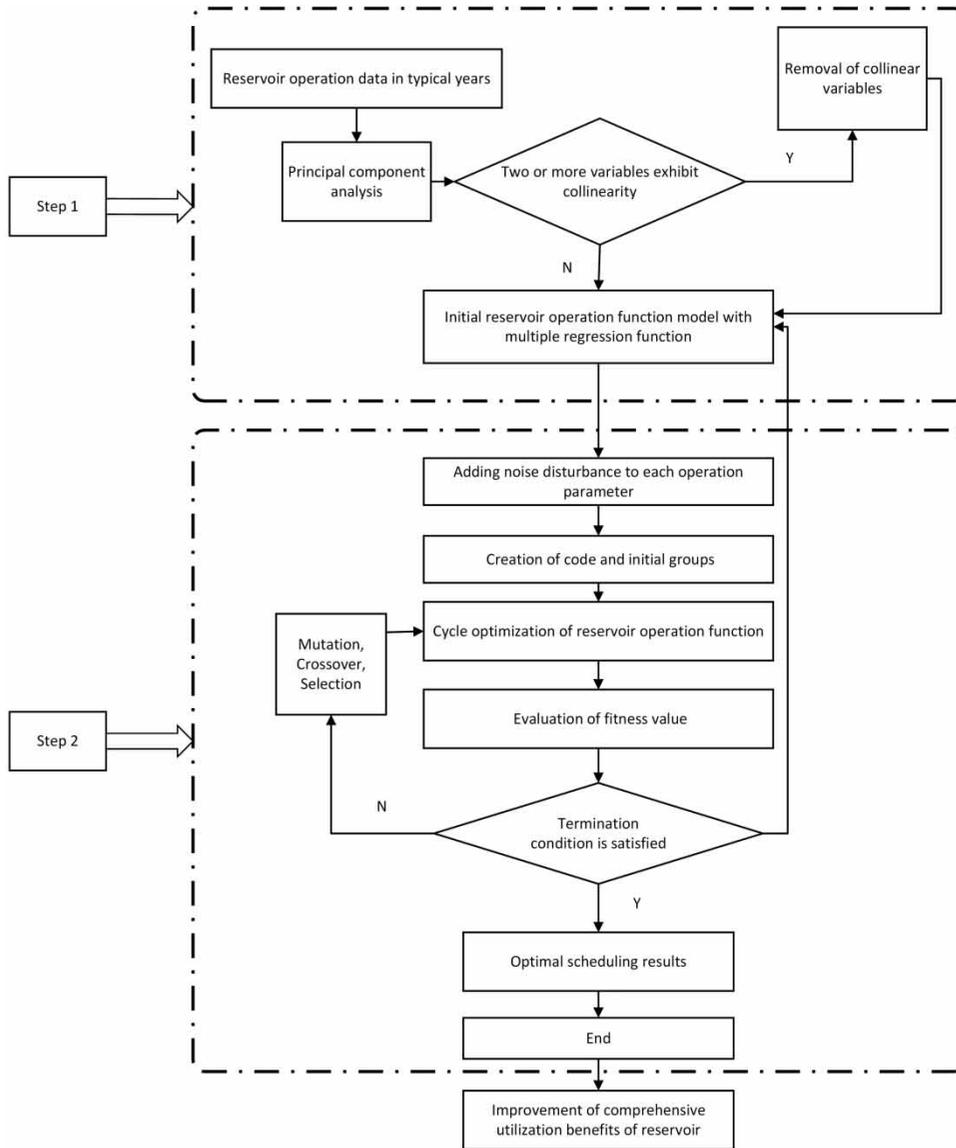


Figure 4 | Flowchart of an optimized reservoir operation function model by using the two-step parameter adjustment method.

that are highly relevant to this PC 1 in this group, and so on. These steps are repeated until all the parameters are uncorrelated. Then, the remaining parameters are used to do a scatterplot analysis for further testing correlation between every two parameters.

After the co-linear problem due to correlations among impact factors of decision variables is effectively treated, a mathematical expression reflecting the correlations between the independent and dependent parameters, called as an initial reservoir operation function model, can be

established as shown in Equation (2):

$$V_D = \beta'_1 \cdot x'_1 + \beta'_2 \cdot x'_2 + \beta'_3 \cdot x'_3 + \cdots + \beta'_{n-1} \cdot x'_{n-1} + \beta'_n \quad (2)$$

where $x'_1, x'_2, x'_3, \dots, x'_{n-1}$ are parameters without co-linear characteristics relating to reservoir operation (such as reservoir inflow, water level, ecological demand water, and so on), and $\beta'_1, \beta'_2, \beta'_3, \dots, \beta'_{n-1}, \beta'_n$ are regression coefficients.

The graphical method can be used to analyze whether the residuals are independent or normal distributed. The

more close to zero, or the more close to the normal distribution for the residuals, the less obvious co-linear characteristics for the data used for regression analysis, indicating it being better than the correlation between the calculated results with the proposed model and the actual data.

Second step of improved parameter adjustment method

A GA is further used to optimize the initial reservoir operation function model developed in Equation (2). The GA is an efficient, parallel, and global search method. It can automatically acquire and accumulate information related to the search space during the process, and adaptively control the search process to achieve the optimal solution (Zandieh & Karimi 2011). The users first determine the control parameters, set the number of chromosomes in the population, and then determine the probabilities of selection, crossover probability and mutation probability by a roulette method. Genetic Algorithms Toolbox in MATLAB software developed by the University of Sheffield in the UK is used to produce solutions of the model. The algorithm design process is shown in Figure 4. The GA parameters are set as follows: population size is 20; mutation probability is 0.2; crossover probability is 0.8; and the number of generations is 100, indicating that if generations are greater than 100 then the iteration terminates.

As for the proposed optimal reservoir operation function model, the objective function is the maximization of water volume for power generation. The following constraints are considered: (1) the water storage for the reservoir from a certain initial period to final period should meet the water balance relationship; (2) the value for the storage capacity should be between the dead storage capacity and the normal storage capacity (or the flood storage capacity in the flood season); (3) the ecological water demand should be maintained; (4) the power generation guarantee rate should be met. Since intelligent algorithms randomly generate the individuals in a certain space, part of individuals can inevitably not meet the constraint of power generation guarantee rate. Thus, the solutions that fail to meet the guarantee rate are handled by adding penalty coefficients.

The objective function to be optimized with GA is the regression model shown as Equation (2). Let the parameters in the regression model ($x'_1, x'_2, x'_3, \dots, x'_{n-1}$) in Equation (2) be expressed as the following with scientific notation:

$$\begin{aligned}x_1 &= x'_1 \times 10^{k_1}; \\x_2 &= x'_2 \times 10^{k_2}; \\x_3 &= x'_3 \times 10^{k_3}; \\&\dots \\x_{n-1} &= x'_{n-1} \times 10^{k_{n-1}}\end{aligned}\quad (3)$$

To prevent the algorithm from falling into a local optimum, it is necessary to screen the new individuals generated by the crossover operation. Thus, noise disturbance coefficients are added to each parameter, respectively. Then, Equation (2) is turned into Equation (4):

$$\begin{aligned}V_D &= (\beta_1 + n_1) \cdot x_1 + (\beta_2 + n_2) \cdot x_2 + (\beta_3 + n_3) \cdot x_3 + \dots \\&\quad + (\beta_{n-1} + n_{n-1}) \cdot x_{n-1} + (\beta_n + n_n) \\n_i &= [-1 \times 10^{k_i+c}, 1 \times 10^{k_i+c}], \quad i = 1, 2, 3 \dots n\end{aligned}\quad (4)$$

where n_i ($i = 1, 2, 3 \dots n$) is the noise disturbance coefficient to reservoir operation parameters; c is the noise disturbance range index, and it depends on the workload and the model parameter optimization range; and $\beta_1, \beta_2, \beta_3, \dots, \beta_{n-1}, \beta_n$ are regression coefficients.

RESULTS AND DISCUSSION

Establishment of the initial reservoir operation model by using repeated PCA combined with multiple regression method

The purpose of the Danjiangkou Reservoir is flood control, power generation, water supply, shipping, and aquaculture. The maximized power generation is set as a target with full consideration of ecological water demand in this research. Ecological water demand is a key component of the river water allocation, and is taken as an important basis to limit the maximum development and utilization of water

resources. China's State Environmental Protection Administration regulated that the minimum ecological water demand should serve as the mandatory minimum discharged water for reservoirs (Chen et al. 2005). With the Tennant method, ecological water demand can be divided into stable ecological water demand and pulsed ecological water demand with the function of providing stable runoff and flood pulses for the biological life processes, respectively (Yang et al. 2009). The outflow from the reservoir, which is recorded after the completion of the Danjiangkou hydro project, is used as a calculation basis for the minimum discharged water of the dam. According to the daily upstream inflow data, the monthly water flow of Danjiangkou Reservoir from 1980 to 2005 is calculated, and the non-zero minimum monthly average flow is used as the minimum ecological water demand. Specifically, the ecological water demand can be calculated with the following formulae:

$$Q_{All} = Q_{State} + Q_{Pulse} \quad (5)$$

$$Q_{State} = Q_{d_{min}} \times RW \quad (6)$$

$$Q_{Pulse} = Q_{OF} \times RM \quad (7)$$

where Q_{All} is the total ecological water demand; Q_{State} is the stable ecological water demand; Q_{Pulse} is the pulsed ecological water demand; $Q_{d_{min}}$ is the minimum discharge flow; RW is the stable ecological water demand coefficient, $RW = 1.0$ in food season and $RW = 0.9$ in non-flood season; Q_{OF} is the discharge water when the frequencies are 25%, perennial mean, 75% and 95%, $M = 0.3$ in food season and $RM = 0.2$ in non-flood season.

According to results of the frequency analysis for the runoff, 1993 is found to be a typical year where the annual runoff guarantee rate is most close to 50%. Thus, the historical data of this year are selected as the data of 50% runoff guarantee rate. The water volume for power generation (VF) is used as the dependent parameter, and reservoir inflow (IF), water level (H), stored water (V), surplus water (VQ), reservoir outflow (VO), and the ecological water demand under 95% runoff guarantee rate ($E_{95\%}$) are chosen as the independent parameters. The above indicators of daily mean values are selected for PCA, and the results are shown in Table 1. The correlation between operation parameters of the reservoir is shown in Figure 5. It can be seen

Table 1 | Matrix for principal component analysis

Parameters	1st PCA		2nd PCA		3rd PCA
	PC 1	PC 2	PC 1	PC 2	PC 1
IF	0.597	0.482	-0.561	0.600	0.826
VO	0.876	0.180	-	-	-
VQ	-0.454	-0.053	-0.815	0.284	0.765
$E_{95\%}$	0.227	0.879	-0.100	0.906	0.733
H	-0.836	0.500	0.918	0.357	-
V	-0.832	0.503	0.916	0.360	-
VF	0.922	0.182	-	-	-

that the upper triangular area is a scatter plot, which gives a visual representation about whether there is a correlation between two variables, and the strength of the correlation. For example, the scatter plot of $E_{95\%}$ and VO can be found at the intersection of these two variables. The lower region is a smooth curve fitting, and confidence ellipses. The confidence ellipse can be used to describe the confidence region with its long axis and short axis as the parameters of the confidence region. The semi-major axis and the semi-minor axis of the confidence ellipse represent the standard deviation of the two-dimensional coordinate component, respectively. The main diagonal panel in Figure 5 contains parameters with maximum and minimum values; the rank of the matrix is reordered using PCA. As can be seen in Table 1, the contributions of the dependent parameter VF and the reservoir outflow (VO) to the PC 1, were both greater than 0.85, when using all the parameters to do PCA for the first time. Thus, collinearity is exhibited, and in the next step of regression analysis the independent parameter VO should be excluded. Then, it can be seen that the contributions of the stored water (V) and the water level (H) to the PC 1 are both greater than 0.90 when doing PCA for the second time. It means that collinearity is also exhibited. Then, the independent parameter V is excluded while keeping H . Then, the remaining parameters are used for the third PCA, and the reservoir inflow (IF) exhibits collinearity with the surplus water (VQ). It also can be seen that $E_{95\%}$ made a significant contribution to the PC 2 alone for each PCA analysis, and is left alone in the process of repeated analysis for the PC 1, so it is kept. Finally, the water level (H), reservoir inflow (IF), and ecological demand water ($E_{95\%}$) are

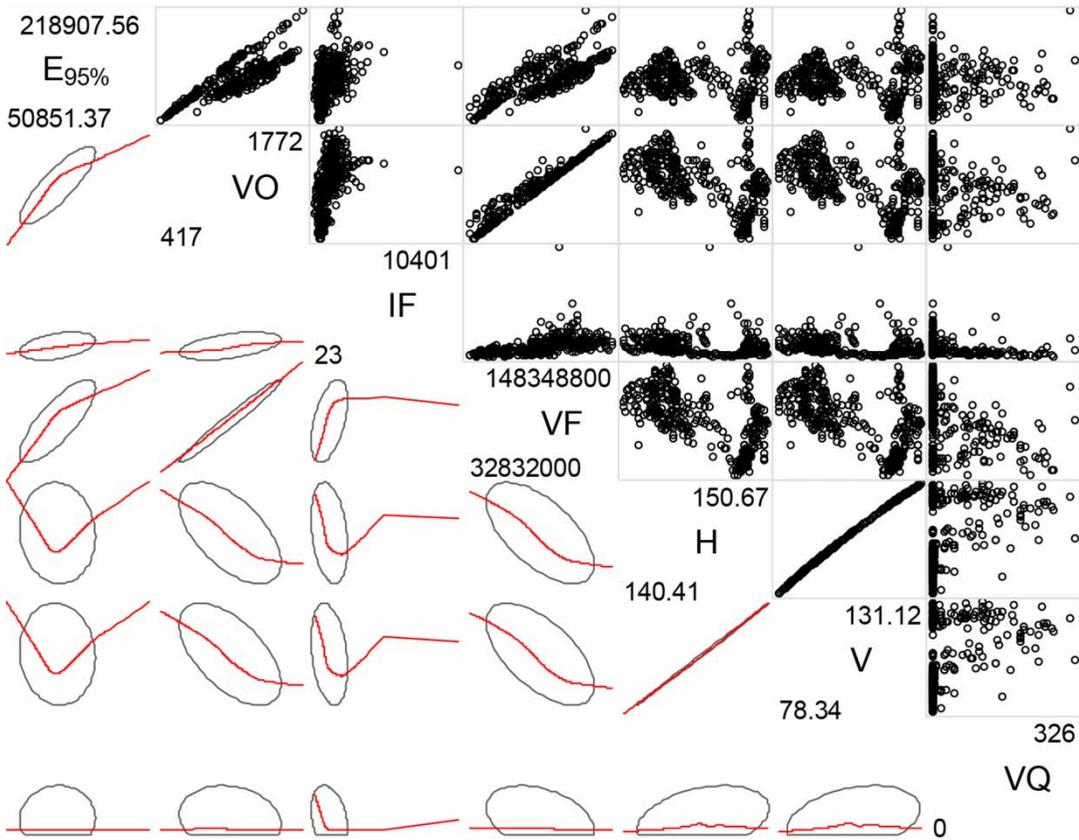


Figure 5 | Scatterplot and oval figure for the correlation between operation parameters of the reservoir.

chosen to conduct multiple regression analysis. It should be noted that because of fewer parameters for the reservoir operation, a relationship between the parameters is exhibited. Under this circumstance, if all of the parameters are selected directly to do regression analysis, it is inevitable that incorrect results would be produced, regardless of the choice of method used to propose parameters (such as perfect fit, forward, backward, enter). Repeated PCA is used to avoid the occurrence of this problem, and more accurate results are produced by correctly excluding parameters with collinearity.

According to Equation (1) and several PCA analyses, reservoir inflow, water level, and ecological demand water under 95% runoff guarantee rate are chosen to perform the regression and obtain the following equation (i.e., initial reservoir operation model):

$$VF = 0.061 \cdot IF - 5923100.72 \cdot H + 1.001 \cdot E_{95\%} + 923717838.0 \quad (8)$$

where VF is the optimized water volume for power generation (m^3), IF is the reservoir inflow (m^3), H is water level (m), and $E_{95\%}$ is ecological demand water under 95% runoff guarantee rate (m^3/s). The significance levels ($Sig. = 0.000$) of the three regression coefficients are all less than 0.05, so it can be concluded that the above independent parameters all have significant effects on the dependent parameter VF .

Using a graphical method to analyze the residuals' statistical distribution, the results are shown in Figures 6–8. As can be seen in Figure 6, the mean residual 1.87×10^{-14} is very close to 0; the standard deviation is 0.996; most of the regression residuals are normally distributed, and they are in the range $(-3, +3)$, with no significant deviation. Figure 7 presents the regression analysis between observed residuals and expected normal residuals. It can be seen that the observed residuals' distribution is well under the assumption of a normal distribution regression, and the correlation between the calculated values of the model and the

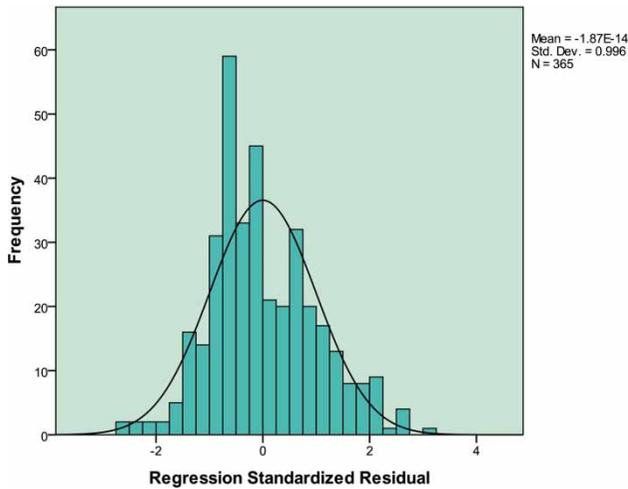


Figure 6 | Regression standardized residual distribution for dependent parameter VF (water volume for power generation).

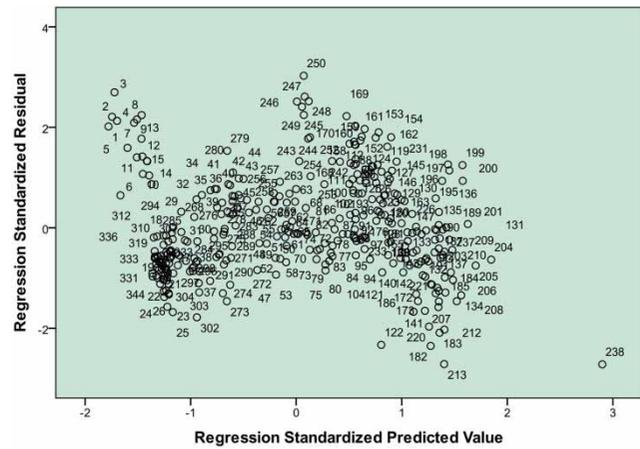


Figure 8 | Regression standardized predicted values for dependent parameter VF (water volume for power generation).

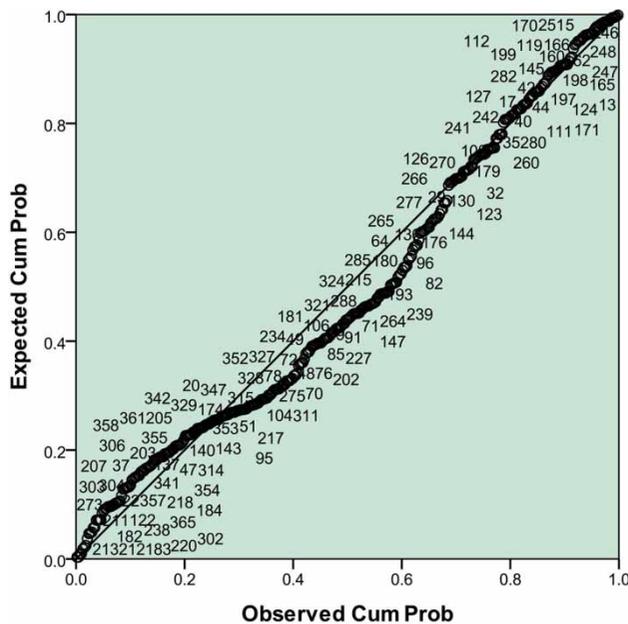


Figure 7 | Normal P-P plot of regression standardized residuals for dependent parameter VF (water volume for power generation).

actual data is good. Figure 8 shows the distribution of residuals, taking the standardized predictive value as the x-axis, the standardized residuals as the y-axis, and the number of days as the label. Most of the residuals scatter within the vicinity of the origin and they are within two standard deviations. Large deviation values are very small, and only the data on the 238th day show a deviation within the vicinity of three standard deviations.

To investigate the accuracy of the above methods, the regression results and the actual values are used to calculate the Nash coefficient, and the result is 0.599. Also, the principal component regression and the ridge regression are used to do a united regression, and the Nash coefficients are calculated as 0.464 and 0.475, respectively. Thus, it can be concluded that the repeated PCA combined with multiple regression can better reflect the correlations among the decision variable for reservoir operation (power generation) and impact factors such as the reservoir inflow, water level, ecological demand water, etc., and it can guide the actual operation of the reservoir.

Optimization of reservoir operation parameters by GA to establish a reservoir operation model

GA is used to optimize the parameters of reservoir operation based on the initial reservoir operation model in the typical year. After adding noise disturbance coefficients to each operation parameter in Equation (8), we get:

$$VF = (0.061 + n_1) \cdot IF - (5923100.72 + n_2) \cdot H + (1.001 + n_3) \cdot E_{95\%} + (923717838.0 + n_4) \quad (9)$$

For Equation (4), $c=2$ is recommended; the scope of the noise disturbance coefficient is: $n_1 = -1$ to 1 , length of step is 0.001 ; $n_2 = -10^7$ to 10^7 , length of step is 0.01 ; $n_3 = -10^2$ to 10^2 , length of step is 0.001 ; $n_4 = -10^9$ to 10^9 ,

length of step is 0.1, establishing four optimization loops. Then, Equation (9) is transformed into the following optimized function:

$$VF = 0.050 \cdot IF - 5375848.179 \cdot H + 0.935E_{95\%} + 84939335.5 \quad (10)$$

The optimized calculation results with Equation (10) are compared with the actual outflow water, the actual daily average water volume for power generation, and the water volume for power generation in the regression model. As can be seen in Figure 9, by using GA to optimize the scheduling model, the water volume for power generation varies more gradually than the actual amount of water, reducing the impact on the power grid. The water volume for power generation after optimization has increased by $1.06 \times 10^9 \text{ m}^3/\text{yr}$ compared with Equation (8), while the economic benefits from the optimized results has increased by $3.22 \times 10^7 \text{ yuan/yr}$ (i.e., $4.69 \times 10^6 \text{ USD/yr}$) over the regression results.

It should be noted here that the year of 1993 with annual runoff guarantee rate most close to 50% is selected as a typical year for the analysis in this research, with the aim to verify the application effects of the proposed optimal operation function model. Actually, the scheduling model can be combined with other models for water resources planning of local development. For

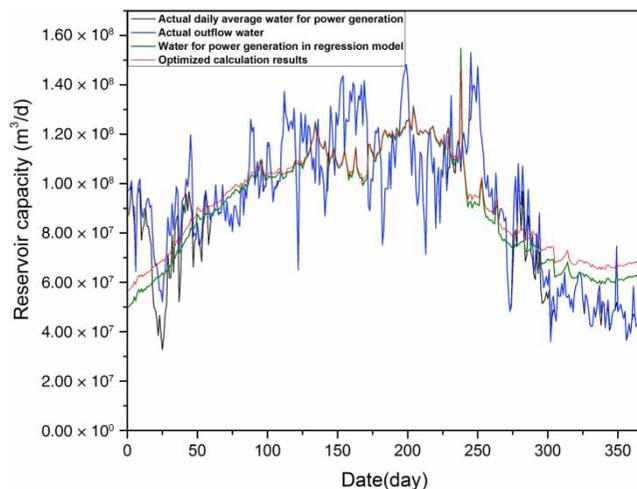


Figure 9 | Optimized results of the water volume for power generation compared with the original operation rules.

example, it can be combined with runoff prediction models such as the improved Thomas–Fiering and wavelet neural network models (Cui et al. 2016) to work out the reservoir operation planning under a certain amount of runoff in the next year, or it could be combined with an ecological planning model to determine the ecological water demand guarantee programs. With the joint operation of the above model, water resources can be used rationally. Therefore, it is a useful technical tool for reservoir management.

CONCLUSIONS

A reservoir operation function model was developed through applying repeated PCA to exclude co-linear parameters in multiple regression. Then, GA is used to optimize the reservoir scheduling parameters. The results show that the proposed parameter optimization method for reservoir operation can better reflect the correlation between the reservoir operation elements, and show that the accuracy is higher than other regression methods. Compared with former operation planning methods, the optimized approach not only increases the amount of water used to generate electricity, but also economic benefits, by assuring downstream ecological flow. Therefore, the constructed model is practicable and feasible, and can provide a scientific basis for reservoir operation.

In fact, the relationship between the outflow, inflow, and the water level is usually connected by a water balance equation; the regression model does not reflect this point. In future reservoir scheduling models, planning for water diversion and the effects that the water transfer project have on the amount of water need to be further considered in Danjiangkou Reservoir, one of the most important water resources for the South–North Water Transfer Project in China. In addition, some runoff forecast models and ecological planning models will be combined with the proposed two-step parameter adjustment method to increase economic or ecological benefits during the reservoir operation in the future. Nevertheless, this paper presents a simple method to optimize reservoir operation, which can also be applied to other reservoirs.

ACKNOWLEDGEMENTS

This research was financially supported by the National Natural Science Foundation of China (Grant No. 51439001, 51721093, 51679008), Chinese National key research and development program (Grant No. 2016YFC0401302), and National Science and Technology Support Program (Grant No. 2011BAC12B02). We would like to extend special thanks to the editor and the anonymous reviewers for their valuable comments in greatly improving the quality of this paper.

REFERENCES

- Abdul-Wahab, S. A., Bakheit, C. S. & Al-Alawi, S. M. 2005 Principle component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environmental Modelling & Software* **20**, 1263–1271.
- Bai, T., Chang, J. X., Chang, F. J., Huang, Q., Wang, Y. M. & Chen, G. S. 2015 Synergistic gains from the multi-objective optimal operation of cascade reservoirs in the Upper Yellow River basin. *Journal of Hydrology* **523**, 758–767.
- Čamdevýřena, H., Demýřa, N., Kanika, A. & Keskýnc, S. 2005 Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs. *Ecological Modelling* **181** (4), 581–589.
- Chandramouli, V. & Deka, P. 2005 Neural network based decision support model for optimal reservoir operation. *Water Resources Management* **19**, 447–464.
- Chang, L. C., Chang, F. J., Wang, K. W. & Dai, S. Y. 2010 Constrained genetic algorithms for optimizing multi-use reservoir operation. *Journal of Hydrology* **390** (1–2), 66–74.
- Chen, K. Q., Cao, X. H., Chen, L. Q. & Peng, F. 2005 Some considerations of key technologies on environmental protection in the construction of water conservancy and hydropower. *Proceedings of Environmental and Aquatic Ecosystem Protection Technology Policy Symposium Related to Water Conservancy and Hydropower Projects* **7**, 44–60 (in Chinese).
- Cui, Q., Wang, X., Li, C. H., Cai, Y. P. & Liang, P. Y. 2016 Improved Thomas-Fiering and wavelet neural network models for cumulative errors reduction in reservoir inflow forecast. *Journal of Hydro-Environment Research* **13**, 134–143.
- De Paulo, J. M., Barros, J. E. M. & Barbeira, P. J. S. 2016 A PLS regression model using flame spectroscopy emission for determination of octane numbers in gasoline. *Fuel* **176**, 216–221.
- Farrokhnia, M. & Karimi, S. 2016 Variable selection in multivariate calibration based on clustering of variable concept. *Analytica Chimica Acta* **902**, 70–81.
- Gain, A. K. & Giupponi, C. 2015 A dynamic assessment of water scarcity risk in the Lower Brahmaputra River Basin: an integrated approach. *Ecological Indicators* **48**, 120–131.
- Labadie, J. 2004 Optimal operation of multireservoir systems: state-of-the-art review. *Journal of Water Resources Planning and Management* **130** (2), 93–111.
- Li, B., Xiao, H. Y., Yuan, X. Z., Willison, J. H. M., Liu, H., Chen, Z. L., Zhang, Y. W., Deng, W. & Yue, J. S. 2013 Analysis of ecological and commercial benefits of a dike-pond project in the drawdown zone of the Three Gorges Reservoir. *Ecological Engineering* **61**, 1–11.
- Liu, H. 2006 *A new Method for Reservoir Optimal Operation*. PhD thesis, Xi'an University of Technology, Xi'an, China (in Chinese).
- Liu, P., Guo, S. L., Xiong, L. H., Li, W. & Zhang, H. G. 2006 Deriving reservoir refill operating rules by using the proposed DPNS model. *Water Resources Management* **20**, 357–357.
- Neelakantan, T. R. & Pundarikanthan, N. V. 2000 Neural network-based simulation-optimization model for reservoir operation. *Journal of Water Resources Planning and Management* **126**, 57–64.
- Petersen, W., Bertino, L., Callies, U. & Zorita, E. 2001 Process identification by principal component analysis of river water-quality data. *Ecological Modelling* **138**, 193–213.
- Simonovic, S. P. 2009 *Managing Water Resources: Methods and Tools for A Systems Approach*. UNESCO, Paris and Earthscan James & James, London.
- Song, G. Y. 2009 *Study on Partial Least Squares Regression*. MA thesis, Zhejiang University, Hangzhou, China (in Chinese).
- Vasan, A. & Srinivasa Raju, K. 2009 Comparative analysis of Simulated Annealing, Simulated Quenching and Genetic Algorithms for optimal reservoir operation. *Applied Soft Computing* **9** (1), 274–281.
- Xie, C. P. & Hawkes, A. D. 2015 Estimation of inter-fuel substitution possibilities in China's transport industry using ridge regression. *Energy* **88**, 260–267.
- Yang, Z. F., Sun, T., Cui, B. S., Chen, B. & Chen, G. Q. 2009 Environmental flow requirements for integrated water resources allocation in the Yellow River Basin, China. *Communications in Nonlinear Science and Numerical Simulation* **14**, 2469–2481.
- Zandieh, M. & Karimi, N. 2011 An adaptive multi-population genetic algorithm to solve the multi-objective group scheduling problem in hybrid flexible flowshop with sequence-dependent setup times. *Journal of Intelligent Manufacturing* **22**, 979–989.

First received 30 July 2016; accepted in revised form 12 April 2017. Available online 23 October 2018