Multiobjective reservoir operating rules based on cascade reservoir input variable selection method

Guang Yang1, Shenglian Guo1, Pan Liu1, Liping Li1, and Chongyu Xu1

1State Key Laboratory of Water Resources and Hydropower Engineering Science, Hubei Provincial Collaborative Innovative Center for Water Resources Security, Wuhan University, Wuhan, China

Abstract

The input variable selection in multiobjective cascade reservoir operation is an important and difficult task. To address this problem, this study proposes the cascade reservoir input variable selection (CIS) method that searches for the most valuable input variables for decision making in multiple-objectivity cascade reservoir operations. From a case study of Hanjiang cascade reservoirs in China, we derive reservoir operating rules based on the combination of CIS and Gaussian radial basis functions (RBFs) methods and optimize the rules through Pareto-archived dynamically dimensioned search (PA-DDS) with two objectives: to maximize both power generation and water supply. We select the most effective input variables and evaluate their impacts on cascade reservoir operations. From the simulated trajectories of reservoir water level, power generation, and water supply, we analyze the multiobjective operating rules with several input variables. The results demonstrate that the CIS method performs well in the selection of input variables for the cascade reservoir operation, and the RBFs method can fully express the nonlinear operating rules for cascade reservoirs. We conclude that the CIS method is an effective and stable approach to identifying the most valuable information from a large number of candidate input variables. While the reservoir storage state is the most valuable information for the Hanjiang cascade reservoir multiobjective operation, the reservoir inflow is the most effective input variable for the single-objective operation of Danjiangkou.

1. Introduction

With the growth of the world's population, water demand has been increasing, which has led to great stress on the management of water resources. Mitigating the scarcity of water and balancing competing objectives require optimal rules that guide the operation of water reservoirs [Gleick and Palaniappan, 2010; Li et al., 2010; G. Yang et al., 2015; T. Yang et al., 2015]. Because of the combined effects of conflicting purposes, nonlinearity in the model, and strong uncertainties in flow forecasting, optimizing reservoir operating rules is both challenging and intriguing [Castelletti et al., 2008]. In recent decades, several studies have introduced, refined, and applied many kinds of methods and algorithms for reservoir operation [Chang and Chang, 2006; Chen et al., 2013; Guo et al., 2011; Labadie, 2004; Liu et al., 2015; Wei and Hsu, 2008; Yeh, 1985].

Researchers have proposed several types of reservoir operating rules. Revelle et al. [1969] assumed that reservoir water releases are linearly related to reservoir storage and decision parameters and introduced linear decision rules in reservoir operation. Karamouz et al. [1992] developed reservoir operating rules with the inputs and outputs of a stochastic optimization model. Chang and Chang [2006] used the adaptive network-based fuzzy inference system (ANFIS) to build a model for reservoir water level prediction. Consoli et al. [2008] derived monthly reservoir operating rules by regressing monthly reservoir releases against state variables. Goyal et al. [2013] derived reservoir operating rules for irrigation and power generation by using artificial neural network (ANN), fuzzy logic, and decision-tree algorithms. Giuliani et al. [2014] analyzed ANN and radial basis functions (RBFs) comparatively for solving high-dimensional state space problems in the reservoir operation under various sets of input and found that the RBFs outperform the ANN in policy parameterization.

Although the approaches for reservoir operation above are useful, they have limitations when one derives multiobjective cascade reservoir operating rules with complex nonlinearity and high dimensionality. For example, the linear decision rule is effective for obtaining multiobjective single reservoir operating rules by
using the explicit optimization [Ahmadi et al., 2014] but ineffective for developing rules for cascade reservoir operation in which state and decision variables are nonlinearly related [Cataláio et al., 2009; Cheng et al., 2012]. Because of outstanding performance in dealing with nonlinear problems, methods such as fuzzy logic, and decision-tree algorithms are suitable for deriving cascade reservoir operating rules [Huang and Townshend, 2003], but they are usually trained or calibrated based on a single-objective optimal reservoir operation [Goyal et al., 2013] and thus are unable to search for multiobjective operating rules. Although the RBFs method can approximate multiobjective operating rules in a nonlinear way, its optimization is time consuming and computationally intensive because of a great number of parameters involved in cascade reservoir operation. To reduce the dimension in optimization for cascade reservoir operating rules, selecting suitable input variables for reservoir decision making is crucial.

Input variable selection methods can be categorized into model-based and model-free approaches [Guyon and Elisseeff, 2003; Maier et al., 2010]. Model-based methods are effective because they are tuned to specific interactions between the model class and data, but they are computationally intensive and time consuming as a large number of calibration and validation processes must be performed to single out the best combination of inputs [Chow and Huang, 2005]. In the model-free approaches, input data sets are selected according to the information content of candidate input variables, as measured by interclass distance, statistical dependence, or information-theoretic measure [Peng et al., 2005]. Model-free methods, which are computationally efficient, have been used to solve water resources management problems with large data sets [Hejazi and Cai, 2009; Giuliani et al., 2015a, 2015b; Yang et al., 2016].

The crux of input variable selection in cascade reservoir operation is to determine the most sensitive or effective variables used for reservoir decision making in light of their nonlinear relationships as well as hydraulic connections in cascade reservoirs. Because a large number of candidate input variables are also involved in the operation of cascade reservoirs, the selection of input variables needs to be computationally efficient. So far, few studies in the literature have examined the selection of input variables and rules derivation for cascade reservoir operation. In this paper, we propose cascade reservoir input variable selection (CIS) method based on data-mining techniques to investigate relationships between reservoir input and decision variables and select the most suitable input data set for cascade reservoir operation. We use the RBFs method to parameterize reservoir operating rules and the Pareto-archived dynamically dimensioned search (PA-DDS) algorithm proposed by Asadzadeh and Tolson [2009] to optimize the parameters for two objectives.

The aim of this paper is to identify relationships between input variables and optimal decisions in cascade reservoir operation and derive multiobjective operating rules by combing the CIS and RBFs methods. The rest of the paper is organized as follows: section 2 introduces the multiobjective reservoir operation model and the input variable selection method followed by the introduction of RBFs. Section 3 describes the reservoir characteristics and data used in this study. Section 4 discusses the application of the CIS and RBFs methods along with the analysis of the multiobjective operating rules with various input variable selections, and then it concludes.

2. Methodology

We can use a great number of candidate input variables to guide decision making for cascade reservoir operation. To select effective input variables and rank their importance for reservoir operation, we propose the CIS method, based on the Extra-Trees algorithm [Geurts et al., 2006]. After selecting input variables, we use the Gaussian RBFs method, which is capable of representing functions for a large class of problems [Busoniu et al., 2011; Giuliani et al., 2014, 2015a, 2015b] to incorporate the selected input variables into cascade reservoir operating rules. In the end, using the PA-DDS multiobjective optimization algorithm, which is robust in solving water resources problems [Asadzadeh and Tolson, 2013; Asadzadeh et al., 2010, 2014; G. Yang et al., 2015], we optimize these cascade reservoir operation rules, designed via the RBFs, and evaluated in terms of hypervolume indicator [Fleischer, 2003; Fonseca et al., 2006], water supply, and power generation. It is worth mentioning that the hypervolume indicator is a guidance criterion that evaluates solutions in multiobjective optimization [Knowles et al., 2003]. The procedure of the multiobjective cascade reservoir operation is shown in Figure 1.
2.1. Multiobjective Cascade Reservoir Operation Optimization

For the cascade reservoir operation, multiobjective functions that maximize water supply and power generation are expressed as follows:

\[
\text{Max } W = \sum_{i=1}^{N} \sum_{t=1}^{T} Q_{i,t}^i \cdot \Delta t \\
\text{Max } E = \sum_{i=1}^{N} \sum_{t=1}^{T} N_{i,t} \cdot \Delta t , \quad N_{i,t} = K_i \cdot Q_{i,t}^i \cdot H_{i,t} \tag{1}
\]

where \(W\) and \(E\) are the sum of water supply yield (m\(^3\)) and power generation (kWh), respectively; \(t\) and \(i\) are the number of operational periods and reservoirs, respectively; \(T\) and \(N\) are the total number of operational periods and reservoirs, respectively; \(Q_{i,t}^i\) is the water supply flow of the \(i\)th reservoir in period \(t\) (m\(^3\)/s); \(\Delta t\) is the time of a single period (s); \(N_{i,t}\) is the output of the \(i\)th reservoir in period \(t\) (kW); \(K_i\) is the hydropower generation efficiency of the \(i\)th reservoir; and \(Q_{i,t}^i\) and \(H_{i,t}\) are the release discharge for power generation and the average hydropower head of the \(i\)th reservoir in period \(t\) (m\(^3\)/s), respectively.

The constraints of cascade reservoir operation are listed as follows:

1. Water balance equation:

\[
V_{i,t+1} = V_{i,t} + \left( Q_{i,t}^i - Q_{i,t}^{\text{out}} - Q_{i,t}^{\text{int}} \right) \cdot \Delta t - E_{i,t} \tag{3}
\]

where \(V_{i,t}\) and \(V_{i,t+1}\) are the storages of the \(i\)th reservoir in period \(t\) and \(t + 1\) (m\(^3\)), respectively; \(Q_{i,t}^i\) is the inflow of the \(i\)th reservoir in period \(t\) (m\(^3\)/s); \(Q_{i,t}^{\text{out}}\) is the water discharge of the \(i\)th reservoir in period \(t\) (m\(^3\)/s); and \(E_{i,t}\) is the sum of evaporation and seepage from the \(i\)th reservoir in period \(t\) (m\(^3\)/s).

2. Connection of flow between reservoirs:

\[
Q_{i,t}^{\text{in}} = Q_{i-1,t}^{\text{in}} + Q_{i,t}^{\text{int}} \tag{4}
\]

where \(Q_{i,t}^{\text{int}}\) is the flow of the intervening basin between the \(i - 1\) and \(i\)th reservoirs in period \(t\) (m\(^3\)/s).
3. Reservoir water level limits:
The reservoir structural and operational constraints can be expressed as
\[ Z_{Li,t} \leq Z_{i,t} \leq Z_{Ui,t} \]  
where \( Z_{i,t} \) is the water level of the \( i \)th reservoir in period \( t \) (m), and \( Z_{Li,t} \) and \( Z_{Ui,t} \) are the minimum and maximum water levels of the \( i \)th reservoir in period \( t \), respectively (m).

4. Water discharge limits:
The outflow constraints needed for flood control and water demand are expressed as
\[ Q_{Li,t} \leq Q_{out}^{i,t} \leq Q_{Ui,t} \]  
where \( Q_{Li,t} \) and \( Q_{Ui,t} \) are the minimum and maximum water discharges of the \( i \)th reservoir in period \( t \) for all downstream uses, respectively (m³/s).

5. Power generation limits
\[ P_{Li,t} \leq N_{i,t} \leq P_{Ui,t} \]  
where \( P_{Li,t} \) and \( P_{Ui,t} \) are the minimum and maximum output limits of the \( i \)th reservoir in period \( t \) (kW).

6. Boundary constraint
\[ Z_{i,t} = \begin{cases} 
Z_{i}^{\text{begin}} & t = 1 \\
Z_{i}^{\text{end}} & t = T 
\end{cases} \]  
where \( Z_{i}^{\text{begin}} \) and \( Z_{i}^{\text{end}} \) are the water levels of the \( i \)th reservoir at the beginning and the end of the reservoir operating process (m), respectively.

2.2. Extra-Trees Algorithm
The Extra-Trees algorithm, a nonparametric tree-based regression method based on an ensemble of decision trees, has already been applied in many fields such as environmental modeling [Jung et al., 2009] and water reservoir operation [Castelletti et al., 2010]. Because of its efficient computational performance in problems with very high dimensionality [Geurts et al., 2006; Geurts and Wehenkel, 2005], we selected the Extra-Trees algorithm to determine the relationships between variables in cascade reservoir operation.

Tree-based regression models are composed of decision nodes, branches, and leaves, which form a cascade of rules leading to numerical values. Each decision tree in the model is obtained by partitioning at the top decision node with a proper splitting criterion, and the splitting process is repeated until all subsets fall into a certain criterion. Like in many other tree-based regression methods such as the random subspace algorithm [Ho, 1998] and the random forests algorithm [Breiman, 2001], the splitting process in Extra-Trees algorithm is performed randomly. Throughout these processes, the hierarchical structure of subset partitions is presented by tree branches, and the leaf is the smallest subset associated with terminal branches. Each leaf in the hierarchical structure is finally labeled with a numerical value. More details of this method appear in Geurts et al. [2006].

The procedure of the Extra-Trees algorithm is shown in Figure 2. Moreover, given a regression problem with an output variable \( y \), \( K \) inputs \( \{a_1, \ldots, a_K\} \) and a sample subset \( S \) composed of \( N \) input-output observations, the splitting procedure for numerical attributes in Extra-Trees algorithm is described below:

**Step 1:** Let \( a_{i}^{\text{max}} \) and \( a_{i}^{\text{min}} \) denote the maximal and minimal values of \( a_i \) in \( S \), and then split the decision node in the learning subset \( S \) by drawing a random cut-point \( s_i \) uniformly in \( [a_{i}^{\text{min}}, a_{i}^{\text{max}}] \).

**Step 2:** Calculate the variance reduction \( \Delta_{var}(a_i) \) associated with \( i \)th attribute \( a_i \) from the equation below:
\[
\Delta_{var}(a_i) = \frac{\text{var}(y|S) - \frac{1}{S} \sum_{i=1}^{S} \text{var}(y|S_i(a_i))}{\text{var}(y|S)}
\]  

(9)
where the term $\text{var}\{y|S\}$ is the variance of output $y$ in sample $S$; $S_l(a_i)$ and $S_r(a_i)$ are two subsets of $S$ that satisfy conditions $a_i < s_i$ and $a_i \geq s_i$, respectively, in which $s_i$ is the randomly selected cut point.

Step 3: Select attribute $a_i$ with the maximal values of $D_{\text{var}}(a_i)$ from $\{a_1, \ldots, a_k\}$, and then use the corresponding $s_i$ as the ultimate cut point to generate the new sample subsets, $S_l(a_i)$ and $S_r(a_i)$, from current subset $S$.

Step 4: Repeat Steps 1–3 until all attributes are constant in $S$, the output is constant in $S$, or the number of members in $S$ is less than the minimum permitted number $n_{\text{min}}$.

Step 5: Produce the estimated outputs by the established tree-based regression structure according to $K$ inputs $a_1, \ldots, a_K$, and then aggregate the estimates obtained by each single tree with an arithmetic average over the ensemble of $M$ trees.

After the regression according to the $N$ input-output observations, the response relationships between the inputs and outputs are established, which means that the outputs can be predicted by certain inputs based on the ensemble tree-based structure.

Suppose that the $\Delta_{\text{var}}(\text{nod}^j)$ is the variance reduction associated with the $j$th node in a tree composed of $\Omega$ nonterminal nodes; the variance reduction for the $M$ ensemble trees [Hatzargyriou, 1999] can be calculated by the following equation:

$$
\Delta_{\text{var}} = \sum_{i=1}^{M} \sum_{j=1}^{\Omega} (\Delta_{\text{var}}(\text{nod}^j) \cdot |S^j|)
$$

where $|S^j|$ is the number of samples associated with the $j$th node for the $r$th tree. This equation calculates the total variance reduction in the form of Extra-Trees, in which the higher $\Delta_{\text{var}}$ implies a more evident difference between the data sets divided by the nodes and more information used for the regression.

Based on variance reduction in the Extra-Trees algorithm, Wehenkel [1998] proposed an approach that ranks the importance of input variables in explaining output behavior. More specifically, relevance $G(a_i)$, which describes the importance of the $i$th input variable $a_i$, can be calculated by

$$
G(a_i) = \frac{\sum_{i=1}^{M} \sum_{j=1}^{\Omega} (\delta(\text{nod}^j, a_i) \cdot \Delta_{\text{var}}(\text{nod}^j) \cdot |S^j|)}{\Delta_{\text{var}}}
$$

where $\delta(\text{nod}^j, a_i)$ is equal to 1 if attribute $a_i$ is used to split the $j$th node (and 0 otherwise).

### 2.3. Cascade Reservoir Input Variable Selection

Many candidate input variables, including state variables such as the reservoir water level and inflow in the present period and exogenous variables such as recent rainfall can be used for decision making in cascade reservoir operation. This work proposes the CIS method, which extracts suitable input variables from these candidate input variables for each reservoir. The flowchart of the CIS method is shown in Figure 3, and the main steps are summarized below.

**Figure 2.** Procedure of the Extra-Trees algorithm.
Step 1: Initially include the basic variables: the time period, reservoir inflow and storage in a selected data set, and then establish the Extra-Trees model with all of the input variables, except the selected data set, to predict the output and rank candidate input attributes by indicator $G_i(a_i)$ in equation (11).

Step 2: Add the first-ranked candidate input variable into the selected data set.

Step 3: Establish the Extra-Trees model with the selected data set to predict the output and compare the model in terms of a suitable distance metric between the optimal output and model prediction. The distance metric can be a suitable accuracy index such as the mean-squared error, or a more sophisticated metric that balances accuracy and parsimoniousness such as the Bayesian information criterion and the Young identification criterion [Young, 2011].

Step 4: Repeat Steps 2–3 until the performance of the underlying prediction does not significantly improve.

Step 5: Remove each of the variables from the selected data set and establish the Extra-Trees model by the rest of the variables and select the first $p$-ranked variables according to the indicator $G_i(a_i)$.

Step 6: Assemble all of the selections in Step 5 and choose the most commonly selected variables as the optimal set. It is worth mentioning that we process the selections in the reservoir from upstream to downstream and account for the relationships in the operation of various reservoirs by adding decision making predicted by selected variables for the $r$th reservoir upstream into the selected data set to determine the input variables for the $r + 1$th reservoir downstream.

2.4. Radial Basis Functions (RBFs)

Because of the scalability of the policy search method with respect to the state-decision space and capability to deal with a great number of variables, we adopt this method, based on RBFs [Deisenroth et al., 2013], to determine the water release decision according to input variables. It is worth noting that the method can be combined with evolutionary algorithms in the case of reservoir operation characterized by high-dimensional decision spaces and multiple competing objectives [Giuliani et al., 2015a, 2015b]. In the RBFs method, the reservoir water release decision can be defined as

$$Q_{r}^{\text{out}} = \sum_{u=1}^{U} \omega_u \varphi_u(X_t) \quad t=1, \ldots, T \quad 0 \leq \omega_u \leq 1$$

(12)
\[
\phi_u(X_t) = \exp \left[ -\sum_{j=1}^{M} \left( \frac{(X_t)_j - c_{ju}}{b_{ju}} \right)^2 \right] \quad c_{ju}, b_{ju} \in [-1, 1]
\]  

(13)

where \( U \) is the number of RBFs \( \phi(\cdot) \) and \( \alpha_u \) are the weights of the \( u \)th RBF; \( M \) is the number of input variables \( X_t \); and \( c_{ju} \) and \( b_{ju} \) are the \( j \)th-dimensional center and radius of the \( u \)th RBF, respectively. It is worth noting that each radial basis function can be regarded as one pattern of decision making in reservoir operation based on \( X_t \), and ultimate decision making is generated by the combination of four patterns \( (U = 4) \).

3. Study Area and Data

3.1. Hanjiang Cascade Reservoirs

The Hanjiang River, the largest tributary of the Yangtze River in China, has a basin area of about 159,000 km\(^2\). The basin has a subtropical monsoon climate, so most of its precipitation occurs from May to October. The river originates in the southwest of Shaanxi province and meets the Yangtze River in Wuhan, a city with more than 10 million inhabitants. The Hanjiang cascade reservoirs in our case study include two reservoirs in the basin: Ankang and Danjiangkou (see in Figure 4).

The Ankang and Danjiangkou Reservoirs are located at the upper and middle reaches of Hanjiang River, respectively. They serve many uses such as power generation, flood control, and navigation. In particular, the Danjiangkou Reservoir is also used as a water source for the south-to-north water transfer project in China. The characteristic parameter values of Ankang and Danjiangkou Reservoirs are listed in Table 1. The main reasons that we selected the Hanjiang cascade reservoirs for this study are because they are used for many...
purposes and because the length of its inflow data (more than 30 years) is accessible for the investigation of operating rules.

### 3.2. Reservoir Operation Rules

In Hanjiang cascade reservoir operation, the water release in period $t$ can be determined through equation (12), in which the input variables are selected by the CIS method. In the equation, the released water of the Ankang Reservoir is only used for power generation while that of Danjiangkou Reservoir is also used for the south-to-north water transfer project. The water supply of Danjiangkou Reservoir is controlled by operating rule curves shown in Figure 5. According to the rule curves, the reservoir water level should be controlled above 145 m during the entire operating period, but it should not exceed 160, 163.5, or 170 m during the summer flood season (from mid-June to mid-August), the autumn flood season (from late-August to early-October), and the nonflood season (from mid-October to early-June in the following year), respectively. If the reservoir water level falls between the water level for flood control or normal uses and the critical limit of the water supply, the water supply flow $Q_d$ is 420 and 135 m$^3$/s, respectively. Similarly, the water supply flow $Q_d$ is 350 m$^3$/s when the reservoir water level is between the water level for flood control or normal uses and the first lower limit of the water supply; 300 m$^3$/s when reservoir water level is between the first and the second lower limit of water supply; and 260 m$^3$/s when the reservoir water level is between the second lower and the critical limit of the water supply.

### 3.3. Data and Parameter Settings

To extract the operating rules for the Hanjiang cascade reservoirs, we collected observed inflow from 31 years of data from 1980 to 2010, provided by Changjiang Water Resources Commission of China. The data from each year consisted of 36 ten day averaged inflow (3 ten day periods/month). We plotted the 10 day average of the Ankang reservoir inflow and the intervening basin flow between Ankang and Danjiangkou Reservoirs, shown in Figure 6. The plot shows that the variations of the Ankang reservoir inflow and the intervening basin flow are similar and that both kinds of flow provide relatively more water from July to October but much less water from December to March of the following year.

To prevent the regression model from overfitting the data, we evaluated the output prediction in the Extra-Trees algorithm by using the K-fold cross-validation method ($K = 5$) [Kohavi,
We conducted a sensitivity analysis of the number of trees parameter $M$ in the output prediction for the Ankang and Danjiangkou Reservoirs. We calculated the mean coefficient of determination value for each value of $M$ in calibration and validation in Figure 7 and set the $M$ at 10; and we set the neighborhood perturbation size parameter and the maximum of the function evaluation for the PA-DDS algorithm at 0.2 and 5000, respectively.

4. Results and Discussion

4.1. Input Variable Selection for Cascade Reservoir Operation

Before selecting input variables, we need to obtain optimal cascade reservoir operation trajectories. Since the optimization of the cascade reservoir operation model is a multistage problem, and its objective function is time separable, it can be solved by using the deterministic dynamic programming (DDP) algorithm [Bellman, 1957]. However, the curse of dimensionality may occur with the use of the DDP algorithm in cascade reservoir operation, in which options for optimization dramatically increase with the number of reservoirs increases [Elmaghraby, 1993]. To overcome the curse of dimensionality, we use the modified dynamic programming algorithm: discrete differential dynamic programming (DDDP) [Heidari et al., 1971], which searches for near-optimal solutions.

Based on optimized trajectories for cascade reservoirs, we use the CIS method to select the most relevant input variables from a sample data set of nine candidate exogenous variables. These nine candidate

![Figure 7. Mean coefficient of determination value for each value of $M$ in calibration and validation.](image)

![Figure 8. Selection frequency of the 1st, 2nd, and 3rd input variables using the CIS and IIS methods.](image)
variables are composed of time period $t$, Ankang reservoir storage ($V_{A,t}$), and Danjiangkou reservoir storage ($V_{D,t}$) in the current period, and Ankang reservoir inflow and intervening basin flow between Ankang and Danjiangkou in the current period ($Q_{A,t}^{in}$, $Q_{D,t}^{in}$) and in the previous period ($Q_{A,t-1}^{in}$, $Q_{D,t-1}^{in}$, $Q_{A,t-2}^{in}$, and $Q_{D,t-2}^{in}$). To reduce the impact of stochastic factors when using the CIS, we run the CIS for 50 times, in which the input data are shuffled randomly for each experiment in this study.

Recently, Galelli and Castelletti [2013] proposed a tree-based iterative input variable selection (IIS) method capable of selecting the most significant and nonredundant inputs in water resources management [Giuliani et al., 2015a,b]. To evaluate the performance of the CIS method, we also apply the IIS method in the selection of input variables for cascade reservoir operation. After 50 independent runs, the input variable selection frequencies of the CIS and IIS methods for the Hanjiang cascade reservoir are shown in Figure 8. In the figure, “1st,” “2nd,” and “3rd” indicates the first, second, and third input variables, respectively, for each reservoir selected by either the CIS or IIS method.

According to Figure 8, the selection results of the first input variable are more concentrated than those of the second and third ones, indicating that the uncertainty in the results becomes evident as the selection processes. For the Danjiangkou Reservoir used for water supply and power generation, its operation is much more complicated than that of the Ankang Reservoir, which leads to more uncertainty in selecting the second and third input variables both for the CIS and IIS methods. In the selection of input variables for the Ankang Reservoir, both methods can select the first and second input variables with a frequency of about 100%, but the CIS evidently outperforms the IIS in the selection of the third input variable. In the input variable selection for the Danjiangkou Reservoir, the selection frequency of IIS decreases dramatically as the number of input variables increases. The number of variables selected by using IIS reaches to seven in the third selection. In contrast, the CIS selects the variables with fewer options (no more than four in the third selection) and higher frequency, which indicates more stable performance in the selection.

The results of input variable selection after 50 runs of the CIS method are shown in Figure 9, which represents each variable as a line crossing the two axes at values of the corresponding selection frequency and average position. Each line in Figure 9 is composed of three points: the first represents the name of the candidate input variable, the second the selection frequency and the third the average ranked position in 50 independent experiments. In general, variables with higher selection frequency and a lower average position are more relevant to the optimal decision-making process in reservoir operation. Selection results show that two variables ($Q_{A,t}$ and $Q_{D,t}^{in}$) provide the highest relative contributions in the optimal operation of the Ankang Reservoir. For the Danjiangkou Reservoir, we select $t$ and $Q_{D,t}^{in}$ as the most important input variables for decision making. Apart from these relevant variables, displayed by green and red solid lines in Figure 9, another two potentially relevant variables ($Q_{A,t-1}^{in}$ and $V_{D,t}$ for the Ankang Reservoir and $V_{D,t}$ and $Q_{D,t}^{in}$ for the Danjiangkou Reservoir), represented by blue and light blue lines are also beneficial for reservoir operation.

Figure 9. Input variable selection frequency and average position after 50 runs of the CIS method for the Ankang and Danjiangkou Reservoirs.
Since decision making in reservoir operation is generally related to the reservoir inflow and time period, we select the $t, Q_{in}^A$, and $Q_{in}^{AD}$, to guide the operation of Hanjiang cascade reservoirs. As the storage of the Ankang Reservoir is much smaller than that of the Danjiangkou Reservoir, previous inflow $Q_{in}^{AD}$ at time $t-1$ rather than current reservoir storage $V_{D,t}$ is selected as the third input variable. It is noted that $Q_{in}^{AD}$ at time $t-1$ is selected because optimal reservoir operation requires accurate inflow forecasting, and information about $Q_{in}^A$ and $Q_{in}^{AD}$ can help us predict the subsequent inflow series. In Danjiangkou reservoir operation, storage $V_{D,t}$ is essential to determining the water supply and thus selected as the third input variable. According to the selection, the most relevant variables for the Ankang and Danjiangkou Reservoirs in the cascade reservoir operation are the inflow and time period, respectively.

To analyze the performance of the CIS method, we compare the scatterplots of the optimal and predicted 10 day average outputs, shown in Figure 10. The use of reservoir inflow $Q_{in}^A$ in Figure 10a (time period $t$ in Figure 10b) describes the main trend in the prediction of reservoir output while the introduction of $t$ and $Q_{in}^{AD}$ in Figure 10a ($Q_{in}^{AD}$, and $V_{D,t}$ in Figure 10b) enhances the predictive capability of the model. The increase of input variables improves the predictive capability, which indicates that decision making for both the Ankang and Danjiangkou Reservoirs approaches the optimal trajectories when more information is provided.

![Figure 10. Comparison between the optimal and predicted 10 day average outputs.](image)

![Figure 11. Comparison of the cascade reservoir operating rules with several input variables in the distribution of nondominated solutions.](image)
4.2. Multiobjective Optimization With Different Information
According to the input variable selection, we select three variables, \( Q_{\text{AD}}^n(t), t, \) and \( Q_{\text{VA}}^n(t) \) which describe the operating rules for the Ankang (Danjiangkou) Reservoir. To evaluate information pertaining to reservoir operation, these variables are incrementally added into the operating rules of the Hanjiang cascade reservoirs. In this study, to describe the operating rules for these reservoirs, we use four input variable data sets. Input variable \( X_t \) is composed of variables \( X_t^A \) and \( X_t^D \) for the Ankang and Danjiangkou Reservoirs, respectively. These four input variable data sets can be expressed as \( X_t(X_t^A(Q_{\text{AD}}^n(t), X_t^D(t)), X_t^A(Q_{\text{VA}}^n(t), t), X_t^D(t, Q_{\text{AD,1}}^n)), X_t(X_t^A(Q_{\text{AD}}^n(t), t, Q_{\text{VA,1}}^n), X_t^D(t, Q_{\text{AD,1}}^n, V_{\text{AD}})), \) and \( X_t(X_t^A(Q_{\text{AD}}^n(t), t, V_{\text{VA}}), X_t^D(t, Q_{\text{AD,1}}^n, V_{\text{AD}})) \), with which the operation results are optimized by PA-DDS “Series 1,” “Series 2,” “Series 3,” and “Series 4,” respectively, shown in Figure 11.

To avoid the impact of randomness, we aggregate the optimized solution sets in Figure 11 from the results of 30 random optimization trials. We compare the nondominated solutions obtained by the decision rule with results are optimized by PA-DDS “Series 1,” “Series 2,” “Series 3,” and “Series 4,” respectively.

As shown in Figure 11, the optimized results of the water supply in “Series 2” with input variables \( X_t(X_t^A(Q_{\text{AD}}^n(t), t), X_t^D(t, Q_{\text{VA,2}}^n)) \) is slightly higher than that in “Series 1” with \( X_t(X_t^A(Q_{\text{AD}}^n(t), X_t^D(t)) \), which indicates that introducing \( Q_{\text{VA}}^n \) increases the water supply for cascade reservoir operation. When information about reservoir storage \( V_{\text{AD}} \) and \( V_{\text{VA}} \) is further incorporated into the operating rules, optimization results exhibit greater water supply and power generation, which indicates that states of reservoirs (both for the upstream and downstream reservoir) are crucial to cascade reservoir operation. It is evident that the performance of optimal cascade reservoir operation further improves when the storage of Ankang Reservoir \( V_{\text{AD}} \) is replaced by the inflow in the previous period, \( Q_{\text{AD,1}}^n \).

According to the above analysis, the water supply and power generation in cascade reservoir operation increase if more information is available and accurate information generally decreases uncertainty in reservoir operation. The distribution of solution sets with higher water supply are more concentrated than those with higher power generation, and solution sets display more uncertainty as power generation increases. Each solution set in Figure 11 represents one kind of cascade reservoir operating rules, and the solution set close to the water supply and power generation can be regarded as a kind of operating rules that mainly focuses on the water supply and power generation, respectively.

4.3. The Impact of Input Variables on Cascade Reservoir Operation
To illustrate the impact of input variables on cascade reservoir operation, we use the hypervolume indicator, water supply and power generation to evaluate Hanjiang cascade reservoir operation, illustrated in Figure 12, in which each number represents the mean value of all optimized solutions. It is worth noting that the relatively greater hypervolume indicator indicates better performance in multiobjective optimization. In Figure 12, four series of data set shown on the x axis represent results from the input variables: \( X_t(X_t^A(Q_{\text{AD}}^n(t), t), X_t^D(t, Q_{\text{VA}}^n(t))), X_t(X_t^A(Q_{\text{AD}}^n(t), t, V_{\text{VA}}), X_t^D(t, Q_{\text{AD,1}}^n, V_{\text{AD}})), \) and \( X_t(X_t^A(Q_{\text{AD,2}}^n(t), t, Q_{\text{AD,1}}^n), X_t^D(t, Q_{\text{AD,2}}^n, V_{\text{AD}})). \) The figure shows that the average hypervolume of nondominated solutions increases steadily with the introduction of \( Q_{\text{AD,1}}^n, V_{\text{AD}}, V_{\text{VA}} \) and \( Q_{\text{AD,2}}^n, Q_{\text{AD,1}}^n \), which indicates that the selected input variables are accurate and important for the multiobjective operation of the Hanjiang cascade reservoirs. The selected input variables are \( t, V_{\text{VA}}, \) and \( Q_{\text{AD}}^n \) for the Ankang Reservoir and \( Q_{\text{AD,1}}^n, V_{\text{AD}}, \) and \( Q_{\text{AD,2}}^n \) for the Danjiangkou Reservoir.

To describe uncertainty in multiobjective optimization, 10%
and 90% percentiles of hypervolume for 30 sets of nondominated solutions are also calculated and shown in Figure 12. In general, uncertainty in multiobjective reservoir operation decreases with the increase of the number of input variables if the added information is accurate. The range of hypervolume between 10% and 90% percentiles becomes smaller when input variables change from \( X_t(\chi^{m}_A(t), \chi^{m}_D(t)) \) to \( X_t(\chi^{m}_A(t), \chi^{m}_D(t)) \), \( X_t(\chi^{m}_A(t), \chi^{m}_D(t)) \) and from \( X_t(\chi^{m}_A(t), \chi^{m}_D(t)) \) to \( X_t(\chi^{m}_A(t), \chi^{m}_D(t)) \). But the range is widened when input variables change from \( X_t(\chi^{m}_A(t), \chi^{m}_D(t)) \) to \( X_t(\chi^{m}_A(t), \chi^{m}_D(t)) \). Therefore, the \( Q_{AD}^{m-1} \) selected by the CIS method provides more accurate information and obtains nondominated solutions with less uncertainty than conventional input variable, \( V_A \).

Decision makers need to not only select important input variables but also assess the monetary value provided by these input variables. To evaluate monetary values of input variables selected by the CIS method, cascade reservoir operation results with the rules using these input variables are presented by economic signals in Table 2. The economic profits of power generation and water supply are calculated based on electrovalence (0.21 Yuan RMB per kWh) and water price for the south-to-north water transfer project (0.13 Yuan RMB per m^3) in China. Evidently, all selected input variables increase economic profits for cascade reservoir operation, and the economic profits of \( (Q_{AD}^{m}, V_A, V_D) \) and \( (Q_{AD}^{m-1}, V_A, V_D) \) are calculated as 3075, 4996, and 1610 \( \times 10^4 \) Yuan (RMB), respectively. Therefore, the input variables with the greatest economic profits are reservoir storage states \( (V_A, V_D) \), and then are \( Q_{AD}^{m} \) and \( Q_{AD}^{m-1} \), for the cascade reservoir operation.

### 4.4. Cascade Reservoir Operating Rules With Different Input Variables

The trajectories simulated by using reservoir operating rules with selected input variables are displayed and analyzed. As shown in Figure 13, the water level of the Ankang and Danjiangkou Reservoirs, the cascade reservoir power generation and water supply process are represented by the mean value over the period from 1980 to 2010. In Figure 13, “1 input,” “2 inputs,” “3 inputs,” and “3 inputs 2” represent the results of operating rules with input variables: \( X_t(\chi^{m}_A(t), \chi^{m}_D(t)) \), \( X_t(\chi^{m}_A(t), \chi^{m}_D(t)) \), \( X_t(\chi^{m}_A(t), \chi^{m}_D(t)) \), \( X_t(\chi^{m}_A(t), \chi^{m}_D(t)) \), and \( X_t(\chi^{m}_A(t), \chi^{m}_D(t)) \), respectively.

According to trajectories simulated by using the reservoir operating rules mainly for power generation (in Figures 13a, 13c, 13e, and 13g), all operating rules tend to keep a high water level of the Ankang Reservoir during the whole year. Because a higher water level generally leads to more power generation, when the reservoir storage is not considered in decision making as “1 input” and “2 inputs,” the operating rules will keep relatively a high water level for the Danjiangkou Reservoir. However, the Danjiangkou Reservoir is operated not only for power generation but also for water supply which is controlled by the reservoir water level (more water supplies with higher water level). Thus, the water supply instead of power generation may increase if the operating rules keep a higher reservoir water level. As it shown in Figures 13a and 13c, the operating rules using conventional input variables (“3 inputs”) take the reservoir storage in consideration and keep the lowest water level for the Danjiangkou Reservoir. In contrast, the operating rules using input variables selected by the CIS method (“3 inputs 2”) can prevent excessive water supply and generate more hydropower by keeping a moderate water level.

In Figures 13b and 13f, the operating rules focusing on water supply increase the water supply yield mainly by raising the Danjiangkou reservoir water level in the nonflood season, which will unavoidably decrease the Ankang reservoir water level and cascade reservoir power generation. The operating rules with more input variables effectively raise the Danjiangkou reservoir water level, especially in the nonflood season, and thus increase water supply for cascade reservoirs. Also, compared with conventional input variables (“3 inputs”), input variables selected by the CIS method (“3 inputs 2”) use the historical reservoir inflow instead
of the reservoir storage to control reservoir water level. Thus, the “3 inputs 2” can precisely recognize the nonflood season, leading to a lower water level in the Ankang Reservoir and more available water for the Danjiangkou Reservoir.

Figure 13. Trajectories of reservoir water level, power generation, and water supply simulated by using the reservoir operating rules mainly for power generation and water supply with different input variables.
The operation results obtained by using the operating rules mainly for power generation and water supply are listed in Table 3. Compared with the conventional input variables, the input variables selected by the CIS method lead to more water supply and power generation. When power generation is the main target, information about $Q_{in}^{AD} (t), (VA; t), (VD; t)$, and $Q_{in}^{A2} (t2, 1)$ can increase the power generation by 1.70%, 1.30%, and 0.37%, respectively. The introduction of $Q_{in}^{AD} (t)$ and $Q_{in}^{A2} (t2, 1)$ can increase the power generation and water supply simultaneously, but the use of reservoir storage $(VA; t), (VD; t)$ tends to increase the power generation by reducing the water supply. For the operating rules focus on water supply, $Q_{in}^{AD} (t), (VA; t), (VD; t)$, and $Q_{in}^{A2} (t2, 1)$ can increase the water supply by 6.98%, 1.59%, and 2.94%, respectively. The use of $Q_{in}^{AD} (t)$ increases the water

Table 3. Hanjiang Cascade Reservoir Operation Using Operating Rules Mainly for Power Generation and Water Supply

<table>
<thead>
<tr>
<th>Input Variables</th>
<th>Power Generation $(10^8 \text{ kWh})$</th>
<th>Water Supply $(10^8 \text{ m}^3)$</th>
<th>Power Generation $(10^8 \text{ kWh})$</th>
<th>Water Supply $(10^8 \text{ m}^3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(Q_{in}^{AD} (t), (VA; t), (VD; t))$</td>
<td>$30.12$</td>
<td>$28.62$</td>
<td>$55.03$</td>
<td>$31.76$</td>
</tr>
<tr>
<td>$(Q_{in}^{A2} (t2, 1))$</td>
<td>$28.84$</td>
<td>$60.19$</td>
<td>$28.84$</td>
<td>$48.56$</td>
</tr>
</tbody>
</table>

The bold values significance that the water supply or power generation increases because of the change of input variables.

The operation results obtained by using the operating rules mainly for power generation and water supply are listed in Table 3. Compared with the conventional input variables, the input variables selected by the CIS method lead to more water supply and power generation. When power generation is the main target, information about $Q_{in}^{AD} (VA; t), (VD; t)$, and $Q_{in}^{A2} (t2, 1)$ can increase the power generation by 1.70%, 1.30%, and 0.37%, respectively. The introduction of $Q_{in}^{AD} (VA; t), (VD; t)$ and $Q_{in}^{A2} (t2, 1)$ can increase the power generation and water supply simultaneously, but the use of reservoir storage $(VA; t), (VD; t)$ tends to increase the power generation by reducing the water supply. For the operating rules focus on water supply, $Q_{in}^{AD} (VA; t), (VD; t)$, and $Q_{in}^{A2} (t2, 1)$ can increase the water supply by 6.98%, 1.59%, and 2.94%, respectively. The use of $Q_{in}^{AD} (t)$ increases the water

Figure 14. Historical distribution of water supply and power generation for the cascade reservoir operating rules for different targets (the first, second, and third letter in the legend means the target, number of input variables, and the plotted variable, in which “P” denotes power generation and “W” denotes water supply).
supply in both flood and nonflood seasons (5.02% and 8.13%), but the introduction of \((V_{A,t}, V_{O,t})\) and \(Q_{i,D,t-1}^o\) improve the water supply only in the nonflood and flood season, respectively, which indicates that the reservoir inflow is the most important input variable for water supply during the whole year. It is worth mentioning that the most valuable input variable for the multiobjective optimization in reservoir operation does not have to be the most important one for the operation focusing on water supply or power generation. For example, reservoir storage, the most valuable input variables in multiobjective cascade reservoir operation, is not the most effective variable in the operation of the Danjiangkou Reservoir for water supply.

Uncertainty in the optimization results for the cascade reservoir operating rules focusing on water supply and power generation are analyzed in Figure 14. To illustrate the uncertainty, we calculate the power generation and water supply by using operating rules for about 1000 times. Figure 14 shows the historical distribution of these results, in which the first, second, and third letter in the legend mean the target, the number of input variables, and the plotted variable ("P" denotes power generation; "W" denotes water supply), respectively. For example, the "P2W" means the water supply by optimizing reservoir operating rules mainly for power generation with two input variables. For all reservoir operating rules, the results with one input variable show more uncertainty than those with two input variables and the use of more input variables substantially increases the water supply or power generation.

5. Conclusions

In this paper, the CIS method was proposed to select the most valuable and useful input variables for cascade reservoir operation, and it was compared with the IIS method, which has been proven to be effective in input variable selection. We used the RBFs method to describe reservoir operating rules and optimized the parameters by the PA-DDS meta-heuristic method. We also evaluated the impacts of input variables on the multiobjective cascade reservoir operation in terms of economic profits and uncertainty analysis on selection results. The operating rules using different input variables were analyzed by their simulated trajectories of reservoir water level, power generation, and water supply. The main conclusions of this study are summarized as follows:

1. The selected input variables have much less uncertainty than those of IIS, and the difference between the predicted and optimal reservoir output becomes less with the increase of selected input variables. The \(t, Q_{i,D,t-1}^o, Q_{i,O,t}, Q_{i,DD,t},\) and \(V_{O,t}\) are selected for the Hanjiang cascade reservoirs to derive the multiobjective operating rules.

2. For the input variables selected by CIS method, the reservoir storage states \((V_{A,t}, V_{O,t})\) are the input variables with the greatest hydropower profit for cascade reservoir operation, and the \(Q_{i,D,t-1}^o\) with inflow forecasting information is beneficial to improve the nondominate solutions and decrease the uncertainty in multiobjective optimization. According to the operation trajectories of the Ankang and Danjiangkou Reservoirs, which were simulated based on different input variables, the rules with more information are more effective in recognizing the storing and spilling opportunities.

In summary, the proposed CIS method is a practical and valuable technique for selecting the most suitable input variables with less uncertainty in multiobjective cascade reservoir operation. The CIS method could provide effective and suitable operating rules for water resources systems consisting of multireservoirs or multisectors by using the information from pervasive sensor networks, which is becoming increasingly available for decision makers.

Acknowledgments

This study was financially supported by the National Natural Science Foundation of China (51539009, 51422907) and the National Key Research and Development Plan of China (2016YFC0402206). The authors really thank the editor and the anonymous reviewers for their valuable comments, which helped improve the quality of the paper. We extend our sincere gratitude to Jane Chisholm for proofreading and grammatical corrections. The data used in this paper can be requested by contacting the corresponding author S. Guo at slguo@whu.edu.cn. Conflict of interest statement: The authors declared that they have no conflicts of interest to this work.

References


YANG ET AL. CASCADE RESERVOIR OPERATION BASED ON CIS

Water Resources Research

10.1002/2016WR020301


