

# An efficient stochastic algorithm for mid-term scheduling of cascaded hydro systems

Xiaolin GE<sup>1</sup>, Shu XIA<sup>2</sup>, Wei-Jen LEE<sup>3</sup>



**Abstract** Due to the stochastic and correlated attributes of natural inflows, the mid-term generation scheduling problem for cascaded hydro systems is a very challenging issue. This paper proposes a novel stochastic optimization algorithm using Latin hypercube sampling and Cholesky decomposition combined with scenario bundling and sensitivity analysis (LC-SB-SA) to address this problem. To deal with the uncertainty of natural inflows, Latin hypercube sampling is implemented to provide an adequate number of sampling scenarios efficiently, and Cholesky decomposition is introduced to describe the correlated natural inflows among cascaded stations. In addition, to overcome the difficulties in solving the objectives of all the scenarios, scenario bundling and sensitivity analysis algorithms are developed to improve the computational efficiency. Simulation results from both two-station and ten-station systems indicate that the proposed method has the merits in accuracy as well as calculation speed for the mid-

term cascaded hydro generation scheduling. The consideration of natural inflow correlation makes the formulated problem more realistic.

**Keywords** Cascaded hydro systems, Mid-term scheduling, Stochastic optimization algorithm, Correlation, Sensitivity

## 1 Introduction

Mid-term hydro scheduling (MTHS) aims to manage hydropower generation as well as water release with maximum profit while satisfying various system constraints over a yearly horizon [1, 2]. This problem plays a crucial role in power system operation and a good solution can provide significant economic benefit. Due to uncertainty and probability distribution factors (e.g., natural inflow [3]), it is intractable to formulate and solve MTHS problems without approximation or simplification. Historically, the mean of annual natural inflow is formulated as input and the problem is described by deterministic models [4–6]. It is difficult to use these models to investigate the impact of natural inflow fluctuation on system operation and potential revenue.

To study the impact of uncertainty of natural flow on the MTHS, stochastic models are proposed in the literature. In [7], a stochastic optimization problem with Markov decision process models was developed allowing for the constraints of governmental regulations. In [8], a novel stochastic risk-constrained model was discussed. To consider the seasonal character and uncertainty of reservoir inflow, a composite optimization model was presented in [9]. A stochastic model with consideration of random water inflow and price errors was described by scenarios in [10]. And an energy equivalent reservoir model [11] was applied

CrossCheck date: 12 March 2018

Received: 4 July 2017 / Accepted: 22 March 2018 / Published online: 8 May 2018

© The Author(s) 2018

✉ Xiaolin GE  
gexiaolin2005@126.com

Shu XIA  
870828@ncepu.edu.cn

Wei-Jen LEE  
wlee@uta.edu

<sup>1</sup> Electric Power College, Shanghai University of Electric Power, Shanghai, China

<sup>2</sup> Shibe Electric Supply Company of State Grid Shanghai Municipal Electric Power Company, Shanghai, China

<sup>3</sup> Energy Systems Research Center, University of Texas at Arlington, Arlington, TX 76019, USA



to the stochastic optimization of hydropower production decisions in the Brazilian power system. For cascaded hydropower systems, there are correlations among stations due to their dependence on a common river basin [12]. Typically, natural inflows of stations in the same river basin will increase or decrease synchronously under the influence of geographic factors. However, most reported literature focuses on the relationship between upstream and downstream reservoirs [13, 14], and considerations of correlation among natural inflows are limited. Therefore, it is necessary to consider the correlation of random natural inflows in MTHS problems.

The stochastic model can describe the mid-term hydro scheduling problem more accurately, while bringing great challenges of computation due to the fluctuation of variables and boundary conditions. Currently, stochastic dynamic programming (SDP) and the multi-scenario method are two major approaches to cope with this difficult issue.

SDP adopts discrete water inflow data and a transition probability matrix of adjacent periods to represent the randomness of water inflow [15–17], and maximizes the expected hydropower generation by finding the optimal solution for reservoir volume levels in each period. To reduce the computation time, stochastic dual dynamic programming (SDDP) [18–20] decomposes the multi-stage stochastic optimization problem into several one-stage sub-problems.

With the development of sampling and cutting technology, the multi-scenario method [21] is used to solve the MTHS problem. In this approach, uncertainty is formulated by random variables with known forecasting information, and many possible scenarios are generated. However, such a large number of scenarios will increase the computational burden. Scenario reduction methods, e.g., fast forward reduction [22], moment based reduction [23], and particle swarm algorithms [24], are employed to select representative discrete scenarios and bundle corresponding probabilities with the purpose of decreasing the computation load. However, most of these approaches can only obtain the mean value of the optimization objective, and cannot directly obtain other statistical information such as the standard deviation and probability distribution curves. One way to provide comprehensive statistical data is Monte Carlo simulation with a large enough sampling scale [21], but the computation time is too long. To overcome above mentioned deficiencies, it is imperative to develop an accurate and computationally efficient stochastic optimization algorithm to estimate the objective completely.

The main contributions of this paper include:

- 1) Latin hypercube sampling combined with Cholesky decomposition, enabling the detailed modeling of

stochastic natural inflows with the consideration of correlations among cascaded hydro stations. Cholesky decomposition has been successfully applied to solve the correlation of wind farms [25]. In this paper, for the first time, it is introduced to deal with correlations in the MTHS problem.

- 2) Scenario bundling and sensitivity analysis to obtain rapidly and accurately the optimal mid-term scheduling decisions for a large number of sampling scenarios. The mean, standard deviation, and probability distribution of the objective are captured, which allows the probability analysis of generation profits.

The rest of this paper is structured as follows. Section 2 describes MTHS mathematical formulation for a cascaded hydro system. Section 3 presents the stochastic optimization approach. Numerical results from case studies are provided in Section 4, and conclusions are summarized in Section 5.

## 2 Mathematical model

### 2.1 Objective function

The proposed model aims to maximize the power generation profits of cascaded hydropower stations over a 1-year horizon. The objective is formulated as:

$$Y = \max \sum_{t=1}^T \sum_{h=1}^{N_H} C_t w_{h,t} \quad (1)$$

where  $Y$  represents the total profit or generation for the cascaded hydro system;  $t$  and  $T$  are the index and number of months in the time horizon, respectively;  $C_t$  is the energy price in period  $t$ ;  $h$  and  $N_H$  are the index and number of hydropower units, respectively;  $w_{h,t}$  is the generation output of unit  $h$  in period  $t$ .

In the deregulated electricity market, the energy price during period  $t$  ( $C_t$ ) is varying according to supply and demand [8]. However, this paper focuses on the vertically integrated utility in China, and the energy price is regulated and set by the government [7], hence, energy price parameters in the studied time horizon are set to be 1, and equivalently the objective of this problem is to maximize the total generation over the planning horizon.

### 2.2 Natural inflow probability model

Typically, natural inflow with correlation [26] is approximated as a normal distribution [27], so the probability density function for the natural inflow can be expressed as:

$$f(r_{h_p,t}) = \frac{1}{\sqrt{2\pi}\sigma_{h_p,t}} \exp\left[-\frac{(r_{h_p,t} - \mu_{h_p,t})^2}{2\sigma_{h_p,t}^2}\right] \quad (2)$$

where  $h_p$  is the index of reservoirs (or stations);  $r_{h_p,t}$  is the natural inflow of reservoir  $h_p$  in period  $t$ , and  $f(r_{h_p,t})$  is the corresponding probability density function;  $\mu_{h_p,t}$  and  $\sigma_{h_p,t}$  are the mean and the standard deviation, respectively.

### 2.3 Constraints on reservoirs and hydro units

The following constraints provide a complete description of the problem of minimizing (1):

$$v_{h_p,t} = v_{h_p,t-1} + \left[ r_{h_p,t} - \sum_{h \in h_p} q_{h,t} - s_{h_p,t} + \sum_{i \in U_P} \left( \sum_{h_1} I_{h_1,i} q_{h_1,t} + s_{i,t} \right) \right] \Delta t \quad (3)$$

$$\underline{w}_{h,t} \leq w_{h,t} \leq \bar{w}_{h,t} \quad (4)$$

$$\bar{w}_{h,t} = D_{d,h,t} \bar{P}_{h,t} \Delta t \quad (5)$$

$$\underline{w}_{h,t} = \underline{P}_{h,t} \Delta t \quad (6)$$

$$V_{\min,h_p} \leq v_{h_p,t} \leq V_{\max,h_p} \quad (7)$$

$$Q_{\min,h} \leq q_{h,t} \leq Q_{\max,h} \quad (8)$$

$$Q_{\min,h_p}^S \leq \sum_{h \in h_p} q_{h,t} + s_{h_p,t} \leq Q_{\max,h_p}^S \quad (9)$$

$$\begin{cases} v_{h_p,0} = V_{\text{ini},h_p} \\ v_{h_p,T} = V_{\text{term},h_p} \end{cases} \quad (10)$$

$$\sum_{r=1}^R V_{h_p,r-1} d_{h,h_p,t,r} \leq v_{h_p,t} \leq \sum_{r=1}^R V_{h_p,r} d_{h,h_p,t,r} \quad (11)$$

$$\sum_{r=1}^R d_{h,h_p,t,r} = 1 \quad (12)$$

$$d_{h,h_p,t,r} Q_{\min,h} \leq q_{h,t,r} \leq d_{h,h_p,t,r} Q_{\max,h} \quad (13)$$

$$\sum_{r=1}^R q_{h,t,r} = q_{h,t} \quad (14)$$

$$w_{h,t} = \sum_{r=1}^R K_{h,r} q_{h,t,r} \Delta t \quad (15)$$

where  $v_{h_p,t}$  is the water volume of reservoir  $h_p$  in period  $t$ ;  $q_{h,t}$  is the water discharged from unit  $h$  in period  $t$ ;  $s_{h_p,t}$  is the spillage from reservoir  $h_p$  in period  $t$ ;  $h_1$  is the index of upstream hydropower units, while  $i$  is the index of upstream reservoirs (or stations);  $I_{h_1,i}$  is the correspondence relationship parameter of  $h_1$  and  $i$ , such as if unit  $h_1$  belongs to station  $i$ , the value is 1, otherwise the value is 0;  $U_P$  denotes the set of upstream reservoirs for reservoir  $h_p$ ;  $q_{h_1,t}$  is the water discharged from upstream unit  $h_1$  in period  $t$ ;  $s_{i,t}$  is the

spillage from upstream reservoir  $i$  in period  $t$ ;  $\bar{w}_{h,t}$ ,  $\underline{w}_{h,t}$ ,  $\bar{P}_{h,t}$ ,  $\underline{P}_{h,t}$  are the upper and lower bounds of generation and power for unit  $h$  in period  $t$ , respectively;  $D_{d,h,t}$  is the load factor of hydropower unit  $h$  in period  $t$ ;  $\Delta t$  is the length of each time period;  $V_{\min,h_p}$ ,  $V_{\max,h_p}$ ,  $q_{\min,h}$ ,  $q_{\max,h}$ ,  $Q_{\min,h_p}^S$ ,  $Q_{\max,h_p}^S$  are the minimum and maximum storage volumes, water discharge and total release, respectively;  $V_{\text{ini},h_p}$  and  $V_{\text{term},h_p}$  are the initial and expected final volume of reservoir  $h_p$ ;  $r$  and  $R$  are the index and number of operating zones, respectively;  $K_{h,r}$  is the energy generation efficiency and  $V_{h_p,r}$  is the storage volume of unit  $h$  in operating zone  $r$ ; and  $d_{h,h_p,t,r}$  is the binary variable for unit  $h$  in station  $h_p$  at operating zone  $r$  during period  $t$ , such that if  $V_{h_p,r-1} \leq v_{h_p,t} \leq V_{h_p,r}$  the value is 1, otherwise the value is 0.

Constraint (3) describes the water balance of a reservoir between two consecutive time periods. Constraints (4–6) establish the upper and lower power generation limits. Constraints (7–9) impose the upper and lower bounds on storage, water discharge, and total release. Constraint (10) defines the initial volume and final volume, which are usually determined by long-term scheduling [27]. Constraints (11–15) formulate the links of the energy production to the water volume and the water inflow [28], as shown in Fig. 1. According to (11–13), only one of binary variables  $d_{h,h_p,t,r}$  has the value of 1. If  $v_{h_p,t}$  is being discharged in operating zone  $r$ , the binary variable  $d_{h,h_p,t,r}$  is forced to be 1, otherwise it is 0.  $V_{h_p,0}$  is assumed to be zero. Equation (15) indicates that if the water volume is in the operating zone  $r$ , the energy generation is the product of energy efficiency, water discharge in the operating zone  $r$ , and the time period.

The above discussion, apart from probability expression of (2), establishes the mixed-integer linear programming (MILP) formulation for the MTHS problem. Since water inflows in (2) are modeled as random variables, the objective in (1) is stochastic. In order to estimate the profit and risk for the cascaded hydro system in the scheduling horizon, objectives for all of the possible natural inflow

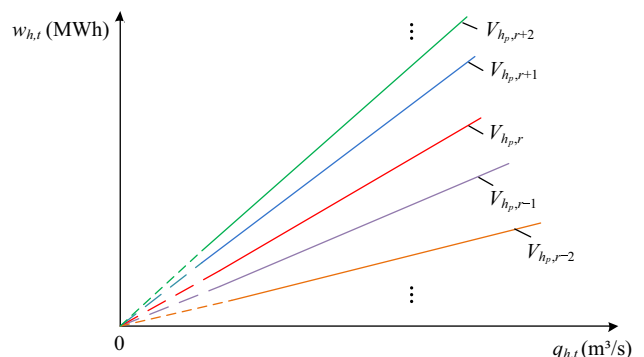


Fig. 1 Energy production curve



scenarios have to be calculated by the stochastic dispatching method.

### 3 Stochastic dispatching method

In order to solve MTHS problem with stochastic and correlated natural inflows, a new algorithm is developed. It consists of three parts, which are sampling, bundling and sensitivity analysis.

#### 3.1 Sampling with correlated random variables

In the proposed algorithm, natural inflows are the input random variables with correlation. Since Latin hypercube sampling is more efficient and robust than simple random sampling [29, 30], and correlations between variables can be considered by using the correlation matrix during the pairing process [31], Latin hypercube sampling with the Cholesky decomposition method is used to generate samples according to their correlation. The detailed steps are as follows.

- Step 1 Initialize  $t = 1$ , define the sample scale  $K$ , where  $K$  is a sufficiently large number of scenarios, such as 3000.
- Step 2 Use Latin hypercube sampling to obtain the independent standardized normal distribution sampling matrix  $R'_0$  with dimension  $N_H \times K$ .
- Step 3 Based on analysis of historical statistics, calculate the correlation matrix of natural inflows  $\rho_{H,t}$ . The matrix is symmetric and positive definite as:

$$\rho_{H,t} = \begin{bmatrix} 1 & \rho_{1,2}^t & \cdots & \rho_{1,N_H}^t \\ \rho_{2,1}^t & 1 & \cdots & \rho_{2,N_H}^t \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{N_H,1}^t & \cdots & \rho_{N_H,N_H-1}^t & 1 \end{bmatrix} \quad (16)$$

- Step 4 Decompose the correlation matrix using the Cholesky method:

$$\rho_{H,t} = \mathbf{H}\mathbf{H}^T \quad (17)$$

where  $\mathbf{H}$  is a lower triangular matrix.

- Step 5 Construct a new matrix  $R'_u$  including standard normal distribution variables with correlation:

$$R'_u = \mathbf{H}R'_0 \quad (18)$$

- Step 6 Update the matrix  $R'_u$  to  $R'_p$  with variables following the normal distribution  $[\mu_{hp,t}, \sigma_{hp,t}]$ , and save  $R'_p$  to the matrix  $R_s$  from rows  $N_H(t - 1) + 1$  to  $N_H t$ .
- Step 7 If  $t < T$ , set  $t = t + 1$ , and go to Step 2, otherwise stop and output the natural inflow sampling matrix  $R_s$  with dimension  $(N_H T)K$ .

#### 3.2 Natural inflow scenario bundling

Various scenarios have been produced by the sampling matrix  $R_s$ . Each column in the matrix represents a natural inflow input scenario, meaning that  $R_s = [R_1, R_2, \dots, R_k, \dots, R_K]$ . Due to the large number of scenarios, if the objectives for all of the scenarios are optimized directly, it is very time consuming. In order to reduce calculation time, a vector quantization algorithm is used to bundle scenarios [32]. Based on the distance between the data vectors, vector quantization assign each vector to a cluster through an optimization process. The required steps can be summarized as follows:

- Step 1 Determine the bundling distance parameter  $\Delta d$  based on the actual requirement.
- Step 2 Initiate the index and number of clusters as  $j = J = 1$ , set the cluster core  $Q_j$  to be the first scenario  $R_1$ , let the number of scenarios contained in cluster  $N_j = 0$ , and set the scenario index  $k = 1$ .
- Step 3 For scenario  $k$ , find out the minimum Euclidean distance  $d_k$  from the scenario to all cluster cores  $Q_j$  ( $j \in \{1, 2, \dots, J\}$ ), which may be expressed as:

$$d_k = \min_{j \in \{1, 2, \dots, J\}} d(R_k, Q_j) \quad (19)$$

- Step 4 If  $d_k$  is less than or equal to  $\Delta d$ , bundle scenario  $k$  to cluster  $j$  (where  $j = \arg \min_{j \in \{1, 2, \dots, J\}} d(R_k, Q_j)$ ), increase  $N_j = N_j + 1$ , update the cluster core  $Q_j$  as the average of all the included scenarios and go to Step 6; otherwise, go to Step 5.
- Step 5 If  $d_k$  is greater than  $\Delta d$ , create a new cluster, set  $J = J + 1$ , assign the scenario  $k$  to cluster  $J$ , save the scenario  $R_k$  as the core of the new cluster  $Q_J$ , and set  $N_j = 1$ .
- Step 6 If  $k$  is less than  $K$ , update  $k = k + 1$  and go to Step 3; otherwise, finish the bundling because all of the scenarios are distributed to clusters, and  $Q_j$  ( $j \in \{1, 2, \dots, J\}$ ) are the corresponding clustering cores.

### 3.3 Fast stochastic scheduling based on sensitivity analysis

In the framework of stochastic optimization, the objective of all the scenarios should be solved. Firstly, we can take one of the scenario clustering cores  $Q_j$  as the natural inflow input, hence the MTHS problem becomes single-scenario hydropower scheduling.

For brevity's sake, the single-scenario hydropower scheduling problem can be formulated as follows:

$$\max Y = f(x) \tag{20}$$

$$\text{s.t. } AR + Bx = 0 \tag{21}$$

$$g(x, y) \leq 0 \tag{22}$$

$$h(x, y) = 0 \tag{23}$$

$$\begin{cases} x \geq 0 \\ y \in \{0, 1\} \end{cases} \tag{24}$$

where  $R$  represents the natural inflows in the scenario of the clustering core  $Q_j$ ;  $x$  is a set of positive variables;  $y$  is a set of binary variables;  $A$  and  $B$  are coefficients. Equation (21) is the water balance constraint as modeled in (3), and  $g(x, y)$  and  $h(x, y)$  are the other inequality and equality constraints. The above formulation can be solved by MILP using, for example, the CPLEX solver, and the acquired objective function is denoted as  $Y_Q$ .

Thereafter, to acquire the objective functions for all of the scenarios in cluster  $j$ , sensitivity analysis is introduced. Figure 2 shows that all the scenarios in the same cluster are in close proximity to the clustering center. Due to sensitivity analysis theory [33], if the sensitivity of the objective function to the input random variable is  $S_Y$ , the objective function  $Y_R$  of any scenario  $R_k$  in the cluster can be restored by:

$$Y_R = Y_Q + S_Y(R_k - Q_j) \tag{25}$$

In (25), the key task is to address the sensitivity. When there are binary variables in the proposed formulation, it is very difficult to acquire the sensitivity. In order to

overcome this difficulty, the optimal binary variable  $\tilde{y}$  is used in the formula, and the scheduling problem is transformed into a linear programming problem. Some auxiliary variables  $\lambda, \mu, \gamma$  are introduced to convert (20–24) into an unconstrained optimization problem as follows:

$$\max L = f(x) - \lambda(AR + Bx) - \mu(g(x, \tilde{y})) - \gamma(h(x, \tilde{y})) \tag{26}$$

When the optimal solution is found, the gradient (partial derivative) of  $L$  for all variables will be equal to 0. So:

$$\nabla_R L|_{\substack{x=\tilde{x} \\ y=\tilde{y}}} = \nabla_R f(x)|_{\substack{x=\tilde{x} \\ y=\tilde{y}}} - \lambda A = 0 \tag{27}$$

The sensitivity can be represented by the gradient of  $f(x)$  with respect to  $R$ , so according to (27) the sensitivity of the primitive objective function can be formulated as:

$$S_Y = \nabla_R f(x)|_{\substack{x=\tilde{x} \\ y=\tilde{y}}} = \lambda A \tag{28}$$

It can be seen that the physical interpretation of  $\lambda$  is the dual solution of the constraints (21). According to the dual principle, the sensitivity can be reached, and on the basis of (25) the objective functions for the scenarios included in cluster  $j$  may be obtained with high calculation speed.

Finally, if all of the clusters have been calculated by this method, the objectives for the original sampling scenarios ( $K$  scenarios) are captured, and then the expectation and probability distribution of the scheduling objective are able to be analyzed.

Consequently, by considering random and correlated natural inflows, the proposed approach can provide an operator a probability analysis of generation profits for cascaded hydro systems.

## 4 Case studies

The models and methods presented are applied to two actual cases in China, a two-station system and a ten-station system. The systems have been implemented on the GAMS and MATLAB platform. The tests are executed on an IBM dual processors computer operating at 2.5 GHz with 4 GB RAM under 32-bit Windows 7. The time horizon of the mid-term scheduling is 1 year, and the time step is 1 month.

### 4.1 Two-station system

The parameters of the hydropower stations are given in Table 1. The mean values of water inflow forecast information are listed in Table 2. The water inflow correlations of the two stations over the 12-month horizon are all set to be 0.6. In this system, there are 24 water inflow input

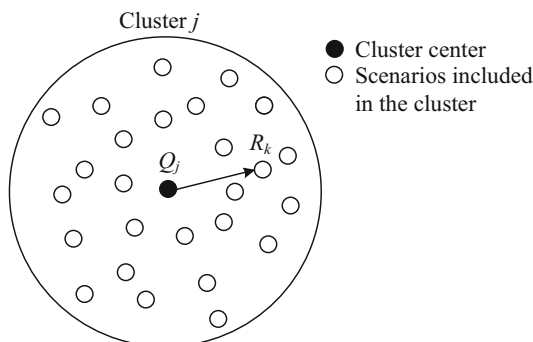


Fig. 2 Diagrammatic sketch of scenario cluster

**Table 1** Characteristics of hydro stations in two-station system

Station No.	Capacity (MW)	$V_{max,h_p}$ ( $10^6 m^3$ )	$V_{min,h_p}$ ( $10^6 m^3$ )	$QS_{max,h_p}$ ( $m^3/s$ )	$V_{ini,h_p}$ ( $10^6 m^3$ )	$V_{term,h_p}$ ( $10^6 m^3$ )	Hydro units No.
1	320	455	133	8386	364	364	1–3
2	84	348	100	5260	278	278	4–6

**Table 2** Predicted values of natural inflow in two-station system

Period (month)	Station 1 ( $m^3/s$ )		Station 2 ( $m^3/s$ )	
	Mean	Standard deviation	Mean	Standard deviation
1	8.16	0.41	20.83	1.04
2	7.68	0.38	19.04	0.95
3	7.04	0.35	16.58	0.83
4	11.36	0.57	25.50	1.28
5	34.72	1.74	85.76	4.29
6	68.16	6.07	233.57	20.79
7	70.40	6.27	261.26	23.25
8	55.04	4.90	179.99	16.02
9	43.52	3.87	138.46	12.32
10	33.28	2.96	101.78	9.06
11	19.36	0.97	52.38	2.62
12	10.88	0.54	27.04	1.35

random variables (multiply 2 stations by 12 months), and the sample scale  $K$  is 3000 scenarios.

In order to evaluate the performance of the proposed method, a series of studies of Latin hypercube sampling with Cholesky decomposition (LC), LC using scenario bundling (LC-SB) and LC combined with scenario bundling and sensitivity analysis (LC-SB-SA) are carried out on the test system. Since the sample scale of LC is 3000, the sampling generates 3000 natural inflow scenarios. In LC-SB, the bundling distance parameter is set to be 30 and the number of scenarios is reduced to 46. The mid-term cascaded hydropower scheduling problem is optimized by the three methods in turn. It is assumed that the results of LC are very accurate and can be treated as the reference case. Statistical analyses of the objective ( $Y$ ) by the three methods are shown in Table 3. With respect to the result of LC with 3000 scenarios, the errors of LC-SB-SA are much smaller than that of LC-SB. Due to the reduction in scenarios, the solution of LC-SB is less accurate, and it gives a poorer estimate of the standard deviation than of the mean value.

In order to view the accuracy of all the optimization results, the optimized total generation output for 3000 scenarios by LC-SB-SA are compared with LC, and the relative errors are illustrated in Fig. 3. The errors are

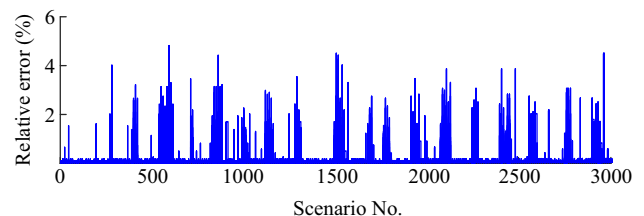
**Table 3** Comparisons of optimization results for total generation output with different methods in two-station system

Optimization results		LC	LC-SB	LC-SB-SA
$Y$ (MWh)	Mean	1220626	1220429	1220429
	Standard deviation	23577	15755	23709
	Maximum	1292018	1282422	1293962
	Minimum	1148110	1172157	1147966
$\varepsilon_Y$ (%)	Mean	/	0.02	0.02
	Standard deviation	/	33.18	0.56
	Maximum	/	0.743	0.15
	Minimum	/	2.09	0.01

mainly caused by nonlinear constraints in the model. For scenarios further from the cluster center, the probability of a large error is higher. In Fig. 3, the maximum error is 4.81% and the average error is only 0.32%. It can be concluded LC-SB-SA has a high precision in solving the MTHS problem.

Figure 4a, b compare the probability distribution of total generation obtained by LC-SB and LC-SB-SA against LC. The simulation results of LC-SB have a larger deviation. In contrast, the probability density curve of LC-SB-SA can better match that of LC, and the cumulative distribution curves of the two approaches are almost identical. This means that the proposed LC-SB-SA can accurately indicate the probability distribution of objectives in the optimal scheduling problem including uncertain factors.

The cumulative probability distribution curves show the likelihood of achieving different levels of total generation. If only considering the predicted mean value of natural inflow, the objective function is 1220800.29 MWh, and the probability of achieving less than or equal to this target is 51.27%. This indicates the high risk of focusing only on the



**Fig. 3** Relative errors for all the sampled scenarios in two-station system

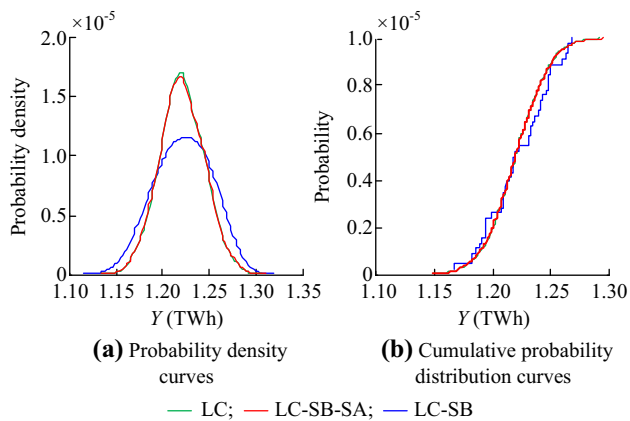


Fig. 4 Comparison of probability distribution curves

scenario with predicted mean value. Thus, being able to obtain the cumulative distribution curve quickly can help operators to evaluate the profit and risks of the cascaded hydropower system.

Table 4 summarizes the computing time for the three mentioned methods. The total computing time is composed of four parts, which are sampling, bundling, optimizing, and restoring. As shown in Table 4, the most time is consumed on the optimization calculation for each scenario, and the sampling, bundling, and other processes do not increase the time greatly. Since the size of scenario clusters for LC-SB is reduced, LC-SB consumes much less time than LC. Compared to the LC-SB, LC-SB-SA obtains more accurate dispatching results with a minor increment on computation time.

The sensitivity of probabilistic results and computing time for LC-SB-SA to different bundling distances is presented in Table 5. As the bundling distance increases, the size of scenario clusters tends to decrease, and the total computing time is also reduced. However, compared with LC, the relative errors of mean and standard deviation of the objective ( $\varepsilon_Y$ ) demonstrate an upward tendency. It can be found that, when 30 is selected as the bundling distance parameter, both of the errors and the computing time are moderate, which means that a reasonable bundling distance parameter can balance accuracy and computation time.

To study the impact of inflow correlation on the simulation results, two-station test system has been optimized with different correlation coefficients. The correlation parameters are set to be 0.1, 0.3, 0.5, 0.7, 0.9, respectively, and corresponding probability distributions of the objective are shown in Fig. 5. As the water inflow correlation coefficient increases from 0.1 to 0.9, the range of variation of the total generation presents an increasing trend. The reason is that with the rising inflow correlation, the synchrony of the adjacent hydropower plants is enhanced, the

Table 4 Comparison of computing time with different methods for the two-station system

Methods	Sampling time (s)	Bundling time (s)	Optimizing time (s)	Restoring time (s)	Total computing time (s)
LC	0.678	0	2269.601	0	2270.279
LC-SB	0.678	0.807	55.212	0	56.696
LC-SB-SA	0.678	0.807	55.212	0.361	57.057

Table 5 Analysis of distance parameters for LC-SB-SA

$\Delta d$	Size of scenario clusters	Relative error $\varepsilon_Y$ (%)		Total computing time (s)
		Mean	Standard deviation	
10	2462	0.006	0.083	2854.88
20	243	0.013	0.258	415.05
30	46	0.016	0.560	57.06
40	15	0.122	0.954	18.16
50	8	0.217	1.267	10.17

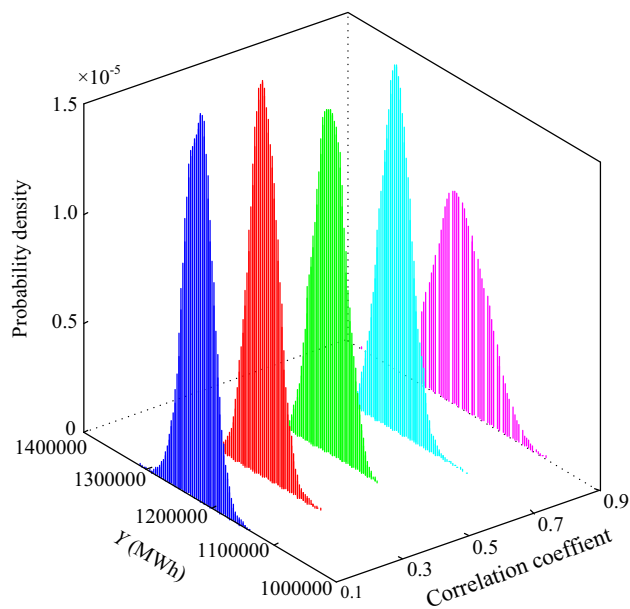


Fig. 5 Probability distributions of objective function with different correlation coefficients

probability of the generators being in the state of low output or high output at the same time is increased, and the total generation is more volatile.

### 4.2 Ten-station system

A ten-station system with 10 reservoirs and 34 hydropower units is used to test the proposed approach. The



**Table 6** Predicted mean values of natural inflow in ten-station system test

Period	Station No. (m <sup>3</sup> /s)									
	1	2	3	4	5	6	7	8	9	10
1	4	21	11	40	42	2	21	12	37	21
2	4	21	11	40	41	2	19	12	40	21
3	4	21	11	38	43	3	17	10	44	23
4	6	40	17	52	84	10	26	19	90	41
5	17	93	42	108	189	25	86	63	158	87
6	34	145	87	234	326	30	234	134	204	164
7	35	130	76	269	292	19	261	110	158	161
8	28	83	48	218	183	18	180	72	102	100
9	22	63	37	172	122	12	138	46	73	79
10	17	50	32	122	121	8	102	38	78	61
11	10	38	20	70	87	6	52	24	67	43
12	5	26	14	47	52	3	27	15	43	28

schematic layout and parameters of hydropower stations in [34] are adopted. The predicted mean values of natural inflow over a 12-month horizon are presented in Table 6. In the rainy season (June to October), standard deviations are assumed to be 6.67% of the mean, while in the dry season (November to May) standard deviations are assumed to be 3.33% of the mean. The water inflow correlation parameters among stations are shown in Table 7. The scenario bundling distance parameter is set to be 30 as for the two-station system. There are 120 natural inflow values (multiply 10 stations by 12 months) which are input random variables for the MTHS problem.

To check the effectiveness of the LC-SB-SA on large-scale stochastic scheduling, the LC with 3000 scenarios is selected as the baseline for comparison because of its high accuracy. The results of the two methods are given in

**Table 7** Water inflow correlation parameters among stations

Station No.	Station No.									
	1	2	3	4	5	6	7	8	9	10
1	1	0.58	0.33	0.02	0.19	0.11	0.46	0.36	0.01	0.05
2	0.58	1	0.7	0.33	0.33	0.2	0.24	0.54	0.05	0.1
3	0.33	0.7	1	0.73	0.36	0.25	0.06	0.25	0.14	0.15
4	0.02	0.33	0.73	1	0.61	0.41	0.19	0.23	0.18	0.26
5	0.19	0.33	0.36	0.61	1	0.65	0.07	0.03	0.24	0.6
6	0.11	0.2	0.25	0.41	0.65	1	0.03	0	0.16	0.21
7	0.46	0.24	0.06	0.19	0.07	0.03	1	0.82	0.14	0.14
8	0.36	0.54	0.25	0.23	0.03	0	0.82	1	0.23	0.11
9	0.01	0.05	0.14	0.18	0.24	0.16	0.14	0.23	1	0.69
10	0.05	0.1	0.15	0.26	0.6	0.21	0.14	0.11	0.69	1

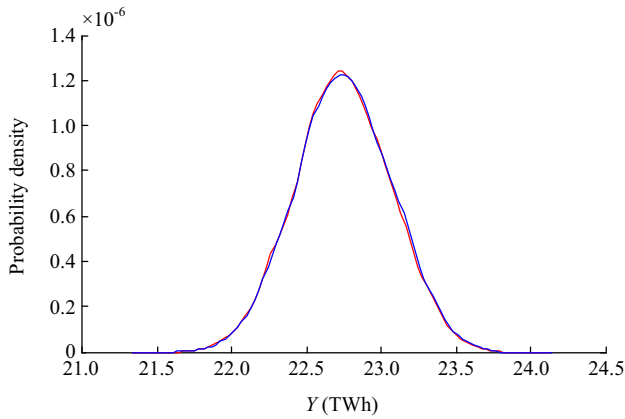
**Table 8** Comparison of optimization results using different methods in ten-station system

Methods	Optimization results (MWh)			
	Mean value	Standard deviation	Maximum	Minimum
LC	22745954.94	316343.84	23916060.70	21541920.97
LC-SB-SA	22751224.08	318074.77	23933036.68	21538074.23

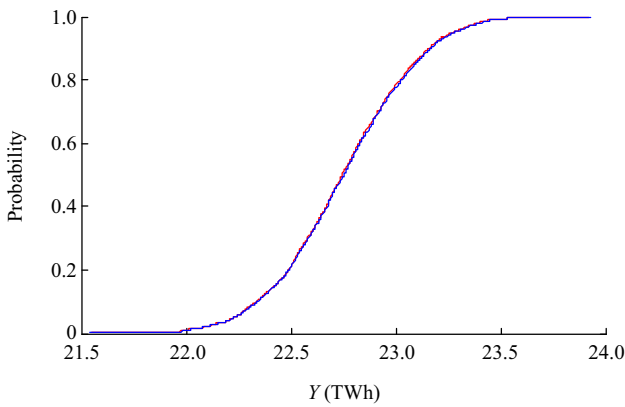
Table 8. The relative error of mean, standard deviation, maximum, and minimum are 0.023%, 0.547%, 0.071%, 0.018%, respectively. Figure 6 shows probability distribution curves of total generation output corresponding to the dispatching schemes obtained by different methods, and the curves are almost coincident. Hence the LC-SB-SA method can yield the results with similar accuracy as LC method, and maintain its good performance when number random variables is increased to 120.

The comparison of the computing time for the two approaches is studied in Table 9. The total computing time of LC-SB-SA is only 1.3% of that of LC. This is because LC-SB-SA uses scenario bundling to reduce the number of scenarios from the original 3000 scenarios to 37. Therefore, only a few scenarios need to be optimized. The optimization results of all the original 3000 scenarios can be restored through sensitivity analysis. The restoration step uses simple linear calculations and consequently takes a small amount of time. For a large-scale hydropower system, the optimizing step is more time consuming, so reducing the number of scenarios will lighten the computation burden significantly, with the restoring step to ensure precision of calculations for each scenario. As a result, the proposed LC-SB-SA has clear advantages in providing an accurate solution with a small amount of calculation.





(a) Probability density curves



(b) Cumulative probability distribution curves  
— LC; — LC-SB-SA

Fig. 6 Comparison of probability distribution of objective  $Y$

Figure 7 presents the optimal total reservoir storage levels in each month, where the highlighted line is the result under the mean value scenario of natural inflow, and the other lines are solutions for all the scenario cluster centers. It illustrates that the proposed method can provide many candidate proposals corresponding to various operating circumstances.

The probability density curves and cumulative probability distribution curves of the objective function are compared in Fig. 8 with and without consideration of water inflow correlation. It can be observed that water inflow correlation has little influence on the mean value of the generation. However, the correlation increases the range of

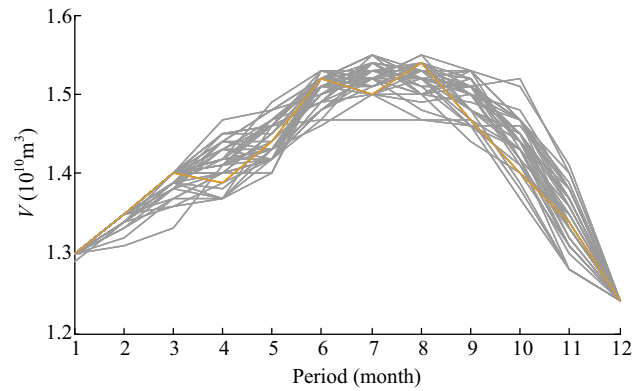
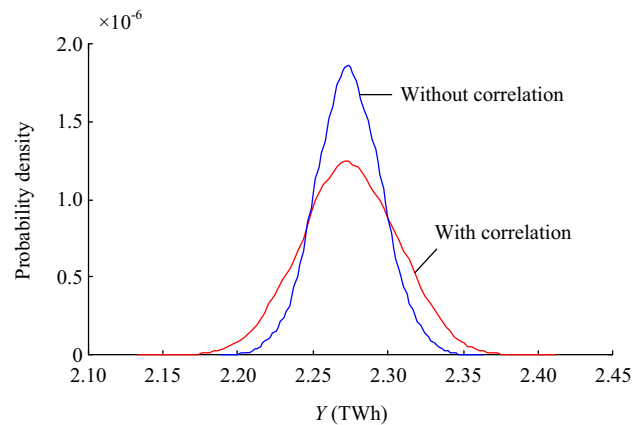
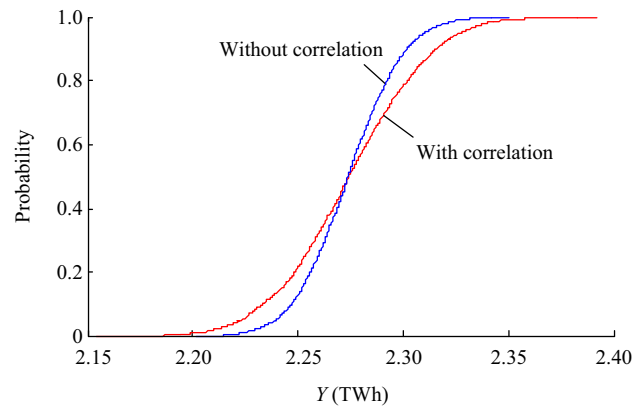


Fig. 7 Optimal reservoir volumes for all the scenario cluster centers



(a) Probability density curves



(b) Cumulative probability distribution curves

Fig. 8 Comparison showing the impact of natural inflow correlation

Table 9 Comparisons of computing time with different methods in ten-station system

Methods	Sampling time (s)	Bundling time (s)	Optimizing time (s)	Restoring time (s)	Total computing time (s)
LC	2.09	0	45375.15	0	45377.24
LC-SB-SA	2.09	2.32	585.03	1.49	590.93



variation of the total generation output, resulting in a higher probability of small or large total generation. The correlation of natural inflow correlation should therefore be considered in the assessment of total generation.

## 5 Conclusion

This paper proposes a computationally efficient stochastic optimization algorithm to solve the MTHS problem for cascaded hydropower systems. Latin hypercube sampling and the Cholesky decomposition method are employed to formulate the random natural inflows accounting for correlation. Scenario bundling and sensitivity analysis are used to obtain the objective function solutions of all scenarios, and to accelerate the calculation. Two-station and ten-station cascaded hydropower systems are used to investigate the proposed method. Conclusions are summarized as follows:

- 1) The proposed Latin hypercube sampling with Cholesky decomposition combined with SB and LC-SB-SA to solve the stochastic MTHS problem has the merits both in calculation speed and precision compared with LC alone and LC-SB. Thus LC-SB-SA is able to yield the probability distribution of the objective function solution rapidly and reliably, and operators can make decisions in a better way to reduce risks.
- 2) The appropriate bundling distance parameter can be selected according to the requirements of actual operating conditions. Generally, if the parameter is smaller, it is possible to achieve higher solution accuracy with the expense of longer computing time.
- 3) Natural inflow correlation has a significant impact on the estimated generation. Scheduling without consideration of natural inflow correlation significantly misjudges the fluctuation of total energy production. Consequently, the solutions of MTHS problem become more realistic by adding such correlation to simulations.
- 4) While coping with many random and correlated variables as in a ten-station system, this approach can still provide an accurate probability curve of total power generation for assessment of mid-term hydro scheduling.

Further work may include pumped hydro storage facilities to adapt to grids with high penetration of variable renewable generation.

**Acknowledgements** This work was supported in part by National Natural Science Foundation of China (No. 51507100), in part by Shanghai Sailing Program (No. 15YF1404600), and in part by “Chen

Guang” project supported by Shanghai Municipal Education Commission and Shanghai Education Development Foundation (No. 14CG55).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- [1] Arce A, Ohishi T, Soares S (2002) Optimal dispatch of generating units of the Itaipu hydroelectric plant. *IEEE Trans Power Syst* 17(1):154–158
- [2] Wang J, Huang W, Ma G et al (2015) An improved partheno genetic algorithm for multi-objective economic dispatch in cascaded hydropower systems. *Int J Electr Power Energy Syst* 67:591–597
- [3] Quintana VH, Chikhani AY (1981) A stochastic model for mid-term operation planning of hydro-thermal systems with random reservoir inflows. *IEEE Trans Power App Syst PAS* 100(3):1119–1127
- [4] Wang C, Zhou J, Lu P et al (2015) Long-term scheduling of large cascade hydropower stations in Jinsha River, China. *Energy Convers Manag* 90:476–487
- [5] Yu Z, Sparrow FT, Bowen BH (1998) A new long-term hydro production scheduling method for maximizing the profit of hydroelectric systems. *IEEE Trans Power Syst* 13(1):66–71
- [6] Zhao TTG, Zhao JS, Liu P et al (2015) Evaluating the marginal utility principle for long-term hydropower scheduling. *Energy Convers Manag* 106:213–223
- [7] Zhao YJ, Chen X, Jia QS et al (2010) Long-term scheduling for cascaded hydro energy systems with annual water consumption and release constraints. *IEEE Trans Auto Sci Eng* 7(4):969–976
- [8] Wu L, Shahidehpour M, Li ZY (2008) GENCO’s risk-constrained hydrothermal scheduling. *IEEE Trans Power Syst* 23(4):1847–1858
- [9] Wu L, Shahidehpour M (2011) Optimal coordination of stochastic hydro and natural gas supplies in midterm operation of power systems. *IET Gener Transm Distrib* 5(5):577–587
- [10] Helseth A, Gjelsvik A, Mo B et al (2013) A model for optimal scheduling of hydro thermal systems including pumped-storage and wind power. *IET Gener Transm Distrib* 7(12):1426–1434
- [11] Matos VL, Finardi EC (2012) A computational study of a stochastic optimization model for long term hydrothermal scheduling. *Int J Electr Power Energy Syst* 43(1):1443–1452
- [12] Mahor A, Rangnekar S (2012) Short term generation scheduling of cascaded hydro electric system using novel self adaptive inertia weight PSO. *Int J Electr Power Energy Syst* 34:1–9
- [13] Guedes LSM, Vieira DAG, Lisboa AC et al (2015) A continuous compact model for cascaded hydro-power generation and preventive maintenance scheduling. *Int J Electr Power Energy Syst* 73:702–710
- [14] Mantawy AH, Soliman SA, El-Hawary ME (2013) The long-term hydro-scheduling problem—a new algorithm. *Electr Power Syst Res* 64:67–72
- [15] Li C, Yan R, Zhou J (1990) Stochastic optimization of interconnected multi-reservoir power systems. *IEEE Trans Power Syst* 5(4):1487–1496

- [16] Sherkat VR, Campo R, Moslehi K et al (1985) Stochastic long-term hydrothermal optimization for a multireservoir system. *IEEE Trans Power App Syst PAS* 104(8):2040–2050
- [17] Bruno HD, Marcelo AT, André LMM (2013) Parallel computing applied to the stochastic dynamic programming for long term operation planning of hydrothermal power systems. *Eur J Oper Res* 229(1):212–222
- [18] Flach BC, Barroso LA, Pereira MVF (2010) Long-term optimal allocation of hydro generation for a price-maker company in a competitive market: latest developments and a stochastic dual dynamic programming approach. *IET Gener Transm Distrib* 4(2):299–314
- [19] Geoffrey P (2015) Stochastic inflow modeling for hydropower scheduling problems. *Eur J Oper Res* 246:496–504
- [20] Pereira MVF, Pinto LMVG (1991) Multi-stage stochastic optimization applied to energy planning. *Math Progr* 52:359–375
- [21] Baslis CG, Bakirtzis AG (2011) Mid-Term stochastic scheduling of a price-maker hydro producer with pumped storage. *IEEE Trans Power Syst* 26(4):1856–1865
- [22] Feng YH, Ryan SM (2014) Scenario reduction for stochastic unit commitment with wind penetration. In: *Proceedings of IEEE PES general meeting, National Harbo, USA, 27–31 July 2014*, 5 pp
- [23] Gil E, Aravena I, Cardenas R (2015) Generation capacity expansion planning under hydro uncertainty using stochastic mixed integer programming and scenario reduction. *IEEE Trans Power Syst* 30(4):1838–1847
- [24] Pappala VS, Erlich I, Rohrig K et al (2009) A stochastic model for the optimal operation of a wind-thermal power system. *IEEE Trans Power Syst* 24(2):940–950
- [25] Delgado C, Domínguez-Navarro JA (2014) Point estimate method for probabilistic load flow of an unbalanced power distribution system with correlated wind and solar sources. *Int J Electr Power Energy Syst* 61:267–278
- [26] Pereira MVF, Oliveira GC, Costa CCG et al (1984) Stochastic streamflow models for hydroelectric systems. *Water Resour Res* 20:379–390
- [27] Xie M, Zhou J, Li C et al (2015) Long-term generation scheduling of Xiluodu and Xiangjiaba cascade hydro plants considering monthly streamflow forecasting error. *Energy Convers Manage* 105:368–376
- [28] Li X, Li TJ, Wei JH et al (2014) Hydro unit commitment via mixed integer linear programming: a case study of the three gorges project. *China. IEEE Trans Power Syst* 29(3):1232–1241
- [29] Shukla A, Singh SN (2016) Clustering based unit commitment with wind power uncertainty. *Energy Convers Manage* 111:89–102
- [30] Díaz G, Casielles PG, Coto J (2014) Simulation of spatially correlated wind power in small geographic areas—sampling methods and evaluation. *Int J Electr Power Energy Syst* 63:513–522
- [31] Azizipanah-Abarghooee R, Niknam T, Malekpour M et al (2015) Optimal power flow based TU/CHP/PV/WPP coordination in view of wind speed, solar irradiance and load correlations. *Energy Convers Manage* 96:131–145
- [32] Cooper HJ, Goodwin GC, Feuer A et al (2012) Design of scenarios for constrained stochastic optimization via vector quantization. In: *Proceedings of American control conference, Montreal, USA, 27–29 June 2012*, 5 pp
- [33] Greene S, Dobson I, Alvarado FL (2002) Sensitivity of transfer capability margins with a fast formula. *IEEE Trans Power Syst* 17(1):34–40
- [34] Ge XL, Zhang LZ, Shu J et al (2014) Short-term hydropower optimal scheduling considering the optimization of water time delay. *Electr Power Syst Res* 110:188–197

**Xiaolin GE** received the M.S. degree and Ph.D. in Electrical Engineering from North China Electric Power University, Beijing, China, in 2010 and 2013, respectively. In 2013, she joined the Shanghai University of Electric Power, where she is currently an associate professor. Her research interests include optimization of large-scale systems and scheduling of cascaded hydro power systems.

**Shu XIA** received Ph.D. in Electrical Engineering from North China Electric Power University, Beijing, China, in 2014. Currently, he joined the State Grid Shibe Power Supply Company, where he is an engineer. His research interests include scheduling of power systems and mixed-integer nonlinear programming problems.

**Wei-Jen LEE** received the B.S. and M.S. degrees from National Taiwan University and the Ph.D. degree from the University of Texas, in 1978, 1980, and 1985, respectively, all in Electrical Engineering. In 1985, he joined the University of Texas, Arlington, where he is currently a professor of the Electrical Engineering Department and the director of the Energy Systems Research Center. He has been involved in research on power flow, transient and dynamic stability, voltage stability, short circuits, relay coordination, power quality analysis, renewable energy, and deregulation for utility companies. He is a registered Professional Engineer in the State of Texas.

