

PAPER • OPEN ACCESS

A long-term intelligent operation and management model of cascade hydropower stations based on chance constrained programming under multi-market coupling

To cite this article: Jia Lu *et al* 2021 *Environ. Res. Lett.* **16** 055034

View the [article online](#) for updates and enhancements.

You may also like

- [Coordinated operation method of cascade hydropower stations considering runoff error](#)
Miao You and Kongfei Hu
- [A Load-Based Method for Peak Operations of Cascaded Hydropower Stations](#)
Sen Wang, Jianjian Shen and Shanzong Li
- [Research on the Influence of Ecological Flow on Cascade Short-term Power Generation Operation in the Baishui River Basin](#)
Jia Zhou, Chusheng Liang and Xiangming Tao



The Breath Biopsy® Guide
Fourth edition

FREE

DOWNLOAD THE FREE E-BOOK

BREATH BIOPSY

OWLSTONE MEDICAL

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

OPEN ACCESS

RECEIVED
28 October 2020REVISED
2 March 2021ACCEPTED FOR PUBLICATION
17 March 2021PUBLISHED
14 May 2021

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.

A long-term intelligent operation and management model of
cascade hydropower stations based on chance constrained
programming under multi-market couplingJia Lu^{*} , Gang Li^{*}, Chuntian Cheng and Benxi Liu

Institute of Hydropower System and Hydroinformatics, Dalian University of Technology (DUT), Dalian No. 2 Linggong Road, Ganjingzi District, Dalian City, Liaoning Province 116024, People's Republic of China

^{*} Authors to whom any correspondence should be addressed.E-mail: glee@dlut.edu.cn**Keywords:** cascade hydropower stations, multimarket, Copula-SARIMA, chance-constrained programming**Abstract**

Under the medium- and long-term electric markets, cascaded hydropower stations face a series of practical challenges due to the uncertainty of inflow and market price. For long-term dispatch scheduling, the allocation of power generation in multimarkets is critical, including clean energy priority consumption market, inter provincial market and intra provincial market in order to maximize the operator's expected revenue and reduce the market operation risks. Based on the hydro-dominant electricity market structure and settlement rules, we propose a long-term optimal operation method for cascade hydropower stations considering the uncertainty of multiple variables. First, a seasonal autoregressive integrated moving average model is used to handle the time-varying and seasonal characteristics of inflow series by using a copula connect function to fit the joint distribution of the monthly inflow, the clearing price of the intra provincial market and the delivery volume of the inter provincial market. Then, uncertain chance-constrained programming is established. Finally, a developed particle swarm optimization algorithm embedded in a Monte Carlo simulation is solved for hydropower operation policies, and the maximum revenue, resource allocation and scheduling strategy are obtained under the corresponding risk tolerance. Taking the actual data of cascaded hydropower stations in Yunnan Province, China, as an example, a simulation analysis is carried out. The results show that the proposed method can reasonably describe the uncertainty and correlation between the variables, realizing the optimal allocation of resources among multimarkets, and provide references for the long-term optimal operation of cascade hydropower stations in a multimarket environment. The results also show that the decision strategies should be determined considering the decision-maker's risk preference.

1. Introduction

Long-term optimal operations of cascaded hydropower stations, using a monthly scheduling period throughout one year, play a key role in the water level control of reservoir operation, energy transfer between the wet season and dry season, compensation for other power sources, etc (Allen and Bridgeman 1986, Barros *et al* 2003, Labadie 2004, Li *et al* 2010, Gu *et al* 2017), especially in hydro-dominated areas. With a new round of electricity marketization reform in China, however, the traditional operation methods of cascade hydropower plants have been greatly

impacted. Under the medium- and long-term electricity markets, the monthly generating capacity in multiple markets needs to be reasonably allocated (Cheng *et al* 2018, Liu *et al* 2019). This includes the clean energy priority consumption market, inter provincial market and intra provincial market, according to the influence of the market clearing price and transaction settlement rules, to maximize the revenue and reduce marketization risk.

When participating in the above markets, cascade hydropower stations need to formulate a long-term operation plan including a monthly operation schedule and trading capacity. Within this framework,

the coupling effect of the monthly trading capacity (generation capacity) and settlement rules should be considered, and reasonable allocation of electricity resources should be based on the settlement priority of each market. If the method of power allocation is unreasonable, this may lead to the failure of part of the contract, which will lead to the loss of a hydropower station's interests. On the other hand, there are several uncertainties in hydrological (reservoir inflow, water quality, etc) and market (clearing price, volume, etc) factors in operation planning and management. For example, the inflow distribution is uneven and random in the year. Hydropower stations need to reallocate them in time and space through long-term dispatching to achieve the goal of efficient utilization. Meanwhile, in order to obtain a greater benefit of electricity sales, hydropower stations need to adjust the operation scheme to track changes in the clearing price, and the uncertainty of price and inflow may lead to the failure of the contract.

The following sections present a thorough review of literature on relevant issues including different sources of uncertainty and their dependencies, decision-makers' risk preferences and optimization approaches, particularly with regard to chance-constrained programming (CCP). Regarding uncertainty in hydrological factors, many works choose the reservoir inflow as the primary and only source of uncertainty (Oliveira *et al* n.d., Ventosa *et al* n.d., Barforoushi *et al* 2006, Hansen 2010). Regarding the market sources of uncertainties, a review of the literature reveals the importance of the clearing price (Conejo *et al* 2003, Kang *et al* 2005, Zareipour *et al* 2010, Nojavan *et al* 2015) and volume uncertainty (Gianfreda 2010, Gianfreda *et al* 2010). Some studies believe there is a correlation between the two kinds of uncertainties (Faria and Fleten 2011, Saadaoui 2013, Saadaoui and Mrad 2017).

In addition to the aforementioned sources of uncertainty, the risk attitude at different levels of decision-makers plays a key role in reservoir operation. This has been analysed in different applications including optimal reservoir operation (García-González *et al* 2007, Bourry *et al* 2009, Moghaddam *et al* 2013, Liu *et al* 2015). Regarding the application of systems analysis and optimization techniques, a review of the literature reveals that both hydrological and market sources of uncertainty have been examined by similar uncertainty-based optimization models and solution approaches. In this line, stochastic linear programming (Grygier and Stedinger 1985), stochastic nonlinear programming (Shrestha *et al* 2005), stochastic dynamic programming (Thanos and Yeh 2010) and CCP (Liu *et al* 2018) are some of the approaches suitable for tackling large-scale optimal reservoir operation problems. Accounting for the interdependency of random parameters, copula-based CCP has recently been used in the

subject areas of waste management (Chen *et al* 2016) and power systems optimization (Chen *et al* 2017).

In this study, we propose a cascade hydropower station optimization model for optimal design and long-term operation considering multiple interdependent sources of uncertainty under multimarket coupling. Based on the presented review of the literature on the uncertain factors and their origins in these systems, the uncertainty sources considered are in the processes of reservoir inflow, clearing price of the intra provincial market and delivery volume of the inter provincial market. The seasonal autoregressive integrated moving average (SARIMA) model is adopted to consider the time-varying and seasonal characteristics of inflow series to ensure that the simulation results are more reasonable. The objective is to obtain the maximum revenue, resource allocation and scheduling strategy under the corresponding risk tolerance. In addition to ordinary constraints including characterizing the dynamics and the physical processes taking place in the system such as water balance equations over time and space, the model includes probabilistic constraints on meeting the maximum output requirement. The model is therefore a chance-constrained program in which the dependence structure of the mentioned random factors is considered using copula functions and covariances to consider dependencies in various random variables. The complex deterministic equivalent of the program, which is highly nonlinear and nonconvex, is solved by the particle swarm optimization (PSO) algorithm of a nested Monte Carlo simulation. The model framework is shown in figure 1.

Taking a cascade hydropower station in Southwest China as an example, it is verified that the proposed method can reasonably describe the uncertainty and correlation between the inflow, price of the intra provincial market and volume of inter provincial market. At the same time, it achieves optimal allocation of the resources among multiple markets and can better coordinate the relationship between profit and risk, to provide a reference for the long-term operation of cascade hydropower stations in a multicoupling market environment.

2. Correlation analysis and modelling of multivariate variables

The reservoir inflow, market clearing price and delivery volume of the inter provincial market are all time series, which have a series of typical characteristics such as time-varying volatility and volatility aggregation. These cannot be handled by ordinary simulation methods. Meanwhile, the mentioned random factors are interdependent, and their correlations need to be considered before the simulation. The SARIMA model (Guin 2006) has the advantages of considering the series autocorrelation, trend and seasonality at the same time. The copula function (Li 1999) can

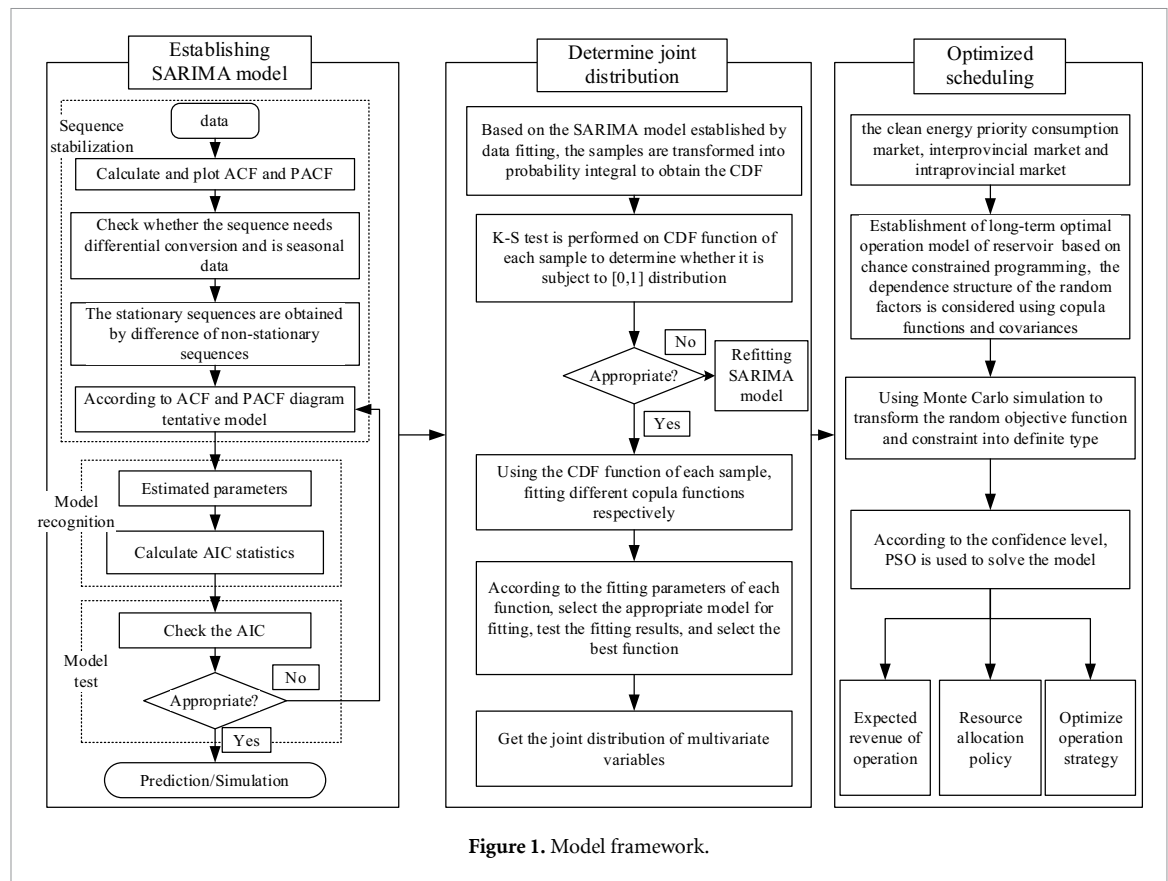


Figure 1. Model framework.

measure the dependence mechanism between variables which contains almost all the information of random variables. Therefore, combining the SARIMA model with a copula function can reasonably describe the dependence between multivariate time series without losing important details.

In summary, this article uses the SARIMA model to describe the marginal distribution of the time series, and then connects the series after a probability integral transformation using a copula function to describe the correlation structure between the multivariate variables. Different time series have different conditional marginal distributions. The resulting conditional marginal distribution functions can be connected by a copula function, which can completely describe the interdependence between the series. The Copula-SARIMA model has different marginal distributions, which is better than the traditional linear model, so it can effectively analyse the nonlinear relationship between multiple variables.

2.1. Establishing SARIMA model of inflow

2.1.1. Basic test of sample

To ensure that the fitting results are correct and reasonable when modelling the volatility of the time series, a series of tests need to be performed on the samples. The specific steps are as follows:

- (a) Normality test. If the time series does not obey the normal distribution, it is possible to use the SARIMA model to fit it. Therefore, using the

skewness and kurtosis joint test method (Jarque-Bera) to test whether the series obeys a normal distribution, the expression of the JB statistic is as follows:

$$JB = \frac{T_s}{6} \left[S^2 + \frac{1}{4}(K-2)^2 \right] \quad (1)$$

where T_s represents the sample size, S represents the skewness coefficient, and K represents the kurtosis coefficient. If the sample follows a normal distribution, then the JB statistic approximately follows a chi-square distribution with 2 degrees of freedom.

- (b) Stationarity test. Stationarity means that the unconditional expectation of the random variable y_t is constant, the variance is constant, and the covariance does not change over time. The specific expression is as follows:

$$\begin{aligned} E|y_t| &= a, \text{Var}|y_t| = \sigma^2, \\ \text{Cov}(y_t - y_{t-i}) &= \sigma_i. \end{aligned} \quad (2)$$

In an autoregressive process, $y_t = by_{t-1} + \alpha + \varepsilon_t$. If the lag coefficient b is 1, it is called the unit root. When the unit root exists, the relationship between the independent variable and the dependent variable is deceptive because any error in the residual series will not decay as the sample size increases, which means that the influence of the residual in the model is permanent. Use the ADF (Augmented Dickey-Fuller test) unit root to test the stationarity of

the time series. If the result of the ADF unit root test b does not equal 1, then reject the null hypothesis that the series has unit roots. If the sample has stationarity, then the series is stationary if there is no unit root, and vice versa.

- (c) Autocorrelation test. Draw the autocorrelation graph and partial autocorrelation graph of the sample. If the test data does not all fall within the acceptance area, then it gradually converges to the acceptance area, indicating that there is a tail in the series data. Then, the series has an autocorrelation and partial correlation.
- (d) White noise inspection. If a series is stationary, then it is necessary to judge whether the data are white noise. Use Ljung–Box statistics to test whether the series is a pure random series; if not, then the series is not white noise.

2.1.2. Identification and fitting of SARIMA model parameters

- (a) Estimation of SARIMA model. SARIMA(p, d, q) (P, D, Q)_S is a widely used method of time-series analysis and simulation that adds seasonal and periodic analysis to the original ARIMA model. p is the order of nonseasonal autoregression, d is the order of nonseasonal difference, q is the order of nonseasonal moving average, P is the order of seasonal autoregression, D is the order of seasonal difference, Q is the order of seasonal moving average and S is the order of seasonal period. The structure is as follows:

$$Y_t = \frac{\Theta(B)\Theta_S(B)}{\Phi(B)\Phi_S(B)\nabla^d\nabla_S^D}\varepsilon_t \quad (3)$$

where Y_t represents the time series to be fitted, ε_t is the residual of the fitted model, $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ is the moving average coefficient polynomial, $\Theta_S(B) = 1 - \theta_1 B^S - \dots - \theta_q B^{QS}$ is the seasonal moving average coefficient polynomial, $\Phi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$ is the autoregressive coefficient polynomial, $\Phi_S(B) = 1 - \varphi_1 B^S - \dots - \varphi_p B^{PS}$ is the seasonal autoregressive coefficient polynomial, and ∇^d and ∇_S^D are nonseasonal and seasonal difference items, respectively. The seasonal difference of the time series is eliminated by the method of series difference and seasonal difference.

The values of p, d, q, P, D and Q were preliminarily determined by the auto correlation function (ACF) and partial autocorrelation function (PACF). According to the literature, the values of P and Q should not exceed the second order. Therefore, the Akaike information criterion (AIC) values of different models are calculated by fitting the models one by one in the range of 0–2.

- (b) Fitting of SARIMA model. The AIC values obtained from the above steps are optimized.

The lower the AIC, the better the model fitting effect. After the values of p, d, q, P, D and Q are determined, the maximum likelihood estimation (MLE) is used to estimate the parameters of the model to obtain the marginal distribution function of time-series variables, and then the nonlinear correlation structure between the variables is connected by a copula function.

2.2. Copula function of variable selection

2.2.1. Marginal distribution transformation

Probability integral transformation assumes that the cumulative distribution function of variable x is $F(x)$, $F(x)$ is continuous, $u = F(x)$, and no matter what distribution x obeys, u obeys $[0, 1]$ in a uniformly distributed fashion. Since the copula function requires the independent variables of the marginal function to obey the uniform distribution of $[0, 1]$, the sample series $\{u_1, u_2, u_3\}$ of each variable is fitted through the above SARIMA model as the marginal distribution of the copula function, and then $\{u_x, u_y, u_z\}$ perform a probability integral conversion to obtain a uniform distribution series $\{u_1', u_2', u_3'\}$ obeying $[0, 1]$. Finally, a suitable copula function is chosen to connect.

2.2.2. Copula function selection

Since the samples fitted in this paper are ternary variables, t-copula and Gaussian-copula can be used for fitting. In this paper, t-copula function is selected to establish the joint distribution of variables. As the follow reasons: (a) t-copula function belongs to elliptic distribution family, which is easy to be extended to high dimension, and it can describe the correlation between variables; (b) the runoff of cascade reservoirs involves small probability floods such as 1000 year return period or 100 year return period, t-copula can describe the correlation of extremum. T-copula can describe a large range of upper tail correlation characteristics, which is suitable for extreme value analysis (Nikolouloupoulos *et al* 2009). The case study shows that t-copula is superior to Gaussian copula in the description of extremum correlation (Zong-Run *et al* 2010). The distribution function $C(\cdot, \dots, \cdot)$ and its density function $c(\cdot, \dots, \cdot)$ of the multivariate t-copula model is

$$\begin{aligned} C(u_1, u_2, u_3) &= T_{\rho, \nu}(T_{\nu}^{-1}(u_1), T_{\nu}^{-1}(u_2), T_{\nu}^{-1}(u_3)) \\ &= \int \int \int \frac{\Gamma(\frac{\nu+N}{2})|\rho|^{-\frac{1}{2}}}{\Gamma(\frac{\nu}{2})(\nu\pi)^{\frac{N}{2}}} \\ &\quad \times \left(1 + \frac{1}{\nu}x' \rho^{-1} x\right) dx_1 dx_2 dx_3 \end{aligned} \quad (4)$$

$$\begin{aligned} c(u_1, u_2, u_3) &= |\rho|^{-\frac{1}{2}} \frac{\Gamma(\frac{\nu+N}{2})\Gamma(\frac{\nu}{2})^{N-1}}{\Gamma(\frac{\nu+1}{2})^N} \frac{(1 + \frac{1}{\nu}\zeta' \rho^{-1} \zeta)^{-\frac{\nu+N}{2}}}{\prod_{n=1}^N (1 + \frac{\zeta_n^2}{2})^{-\frac{\nu+1}{2}}} \end{aligned} \quad (5)$$

where ρ is a symmetric positive definite matrix with a diagonal element of 1, $|\rho|$ is the determinant value corresponding to the matrix and $T_{p,v}(u_1, u_2, u_3)$ is the correlation coefficient matrix with ρ and the degree-of-freedom v multivariate standard t distribution. $T_v^{-1}(\cdot)$ is the inverse function of the univariate t distribution function $T_v(\cdot)$ with degree of freedom v , $x = (x_1, x_2, x_3)'$, $\xi = (\xi_1, \xi_2, \xi_3)'$, $\xi_n = T_v^{-1}(u_n)$ and $n = 1, 2, 3$.

According to the marginal distribution parameters of variables, the parameter values of the t-copula function are calculated, and then the distribution function of the t-copula function can be obtained by formulas (4) and (5). Finally, a Kendall rank correlation coefficient and Spearman correlation coefficient are calculated by t-copula random number points with the same sample number and compared with the statistics of the sample number. The advantages and disadvantages of fitting are tested in turn.

2.2.3. Generation method of simulation data

Similar to the simulation method of the binary copula function, the multivariate copula function simulation can generate a random number series (u_1, \dots, u_n) that obeys the specified n-valued copula function $C(\cdot, \dots, \cdot)$ through the conditional distribution of random variables. In fact, we need to simulate the interdependent three-dimensional random variables, we can only simulate the three components one-to-one, and the split pair of variables constitute interdependent two-dimensional random variables. The reason why it can be carried out in this way is that the ternary joint distribution can be obtained according to the binary conditional distribution and copula function. The specific steps are:

$$\begin{aligned} F(x_1, x_2, x_3) &= C(u_1, u_2, u_3) \\ &= \int_{x_3}^{-\infty} C(F(x_1|x_3), F(x_2|x_3)) dF(x_3) \\ &= \int_{u_3}^{-\infty} C(C(u_1|u_3), C(u_2|u_3)) du_3 \quad (6) \end{aligned}$$

where C is copula function; $F(x_1|x_3)$ and $F(x_2|x_3)$ are conditional probability distribution functions of variables X_1 and X_2 given X_3 .

Because the market development time of this paper is short, and the number of electricity price and electricity quantity in the sample is too small, only SARIMA fitting is carried out for the marginal distribution of runoff series (the runoff sample is from 1953 to 2019), and the kernel density estimation method is used to fit the marginal distribution of other variables.

- (a) According to $F(x_1, x_2, x_3) = C(u_1, u_2, u_3)$ which obtained from the direct statistics

of data, the binary conditional Copula functions $C(u_1|u_3) = \partial C(u_1, u_3)/\partial u_3$ and $C(u_2|u_3) = \partial C(u_2, u_3)/\partial u_3$ are calculated;

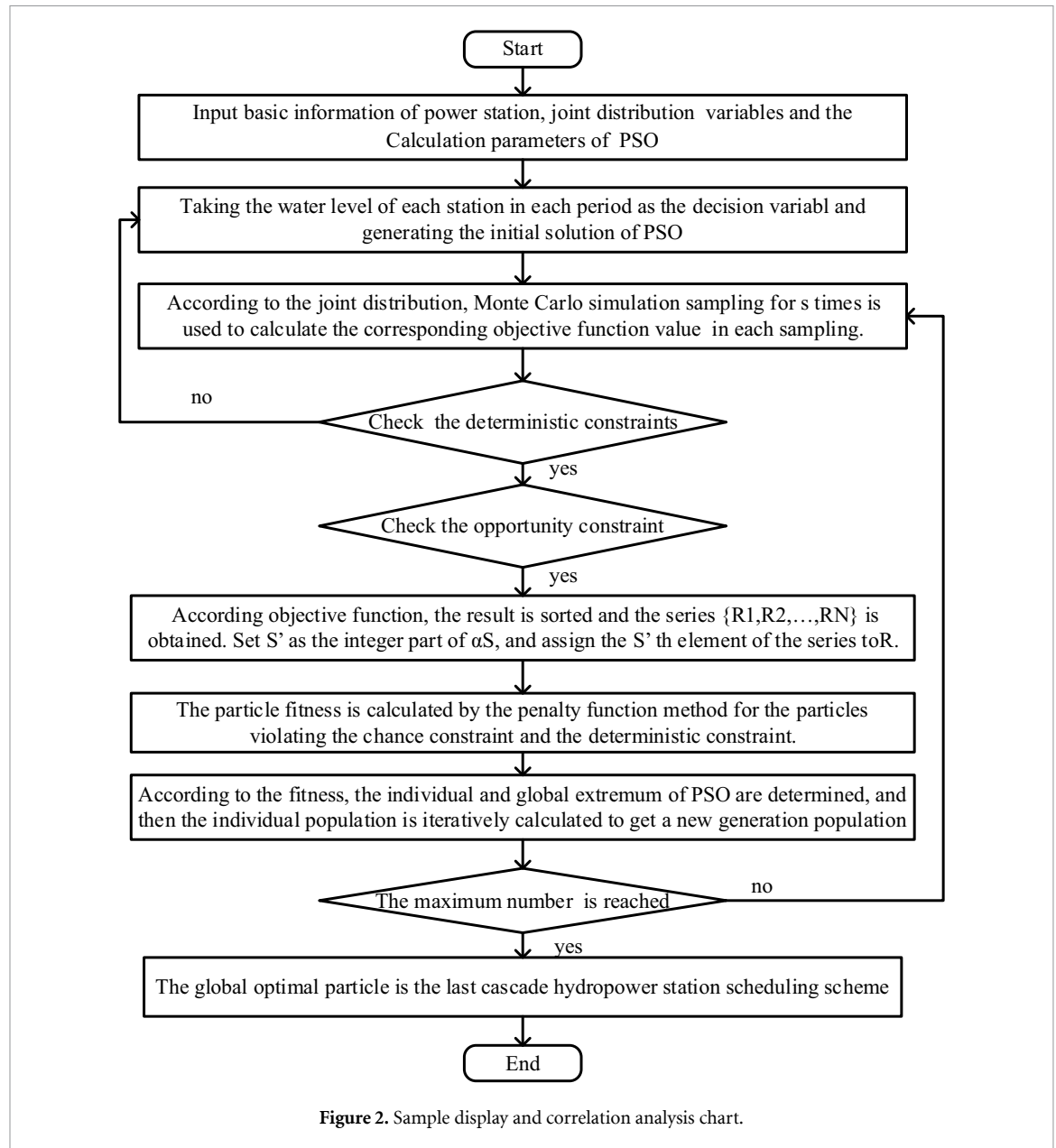
- (b) A random sequence (v_1, v_2, v_3) is generated, which contains three variables that obey $[0, 1]$ independent distribution, and then a random sequence that obeys the specified 3-element t-copula function $T_{p,v}$ is generated.
- (c) According to SARIMA model, m groups of run-off series are generated, and the corresponding u_3 is solved according to $x = F^{-1}(u)$;
- (d) According to the binary conditional copula function obtained in step 1, and Nelson theorem, the corresponding variables are obtained through the inverse functions of $C(u_1|u_3)$ and $C(u_2|u_3)$, that is, the random sequence that obeys the distribution of three variable t-copula function is obtained.
- (e) According to the distribution function $F(x_n)$ of each variable, the variable value corresponding to u_n is calculated.

3. Long-term optimization model based on CCP in multimarket

It is assumed that the cascade hydropower stations belong to the same stakeholder and that each station is a price taker. To make the model closer to the actual project, refer to the actual dispatch mode and introduce a dynamic assessment of the delivery default in the model. That is, dynamically adjust the assessment electricity price according to the proportion of the default power energy. Considering that the cost of cascade hydropower generation is mainly composed of fixed costs, it does not affect model optimization, which is ignored in this paper. Take the year as the dispatch cycle and one month as a period.

3.1. Objective function

The income of cascade hydropower stations mainly comes from clean energy priority consumption market, interprovincial market, intra provincial market and Penalty loss model for breach of contract. Among them, the priority of clean energy priority consumption market settlement is the highest and the price is set to a fixed value according to the flood and dry seasons. The second is the inter provincial market market, according to the statistical data of market operation, the volume of this market can be considered that the price in a year is a fixed value. Finally, the intra provincial market, the price is cleared according to the bidding price. According to the settlement rules, because of the uncertainty of inflow and intra provincial market clearing price, the allocated volume in the inter provincial market of the tap power station may not be completed. Therefore, the loss of default should be considered in the calculation of revenue. That the mathematical model of the long-term



optimal operation of cascade hydropower stations based on chance constraints under multiple market coupling is as follows:

$$\max \bar{R}$$

$$R = \max \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^J (r_{j,i,t}) \quad (7)$$

$$r_{1,i,t} = e_{1,i,t} \times p_{1,t} \quad (8)$$

$$r_{2,i,t} = e_{2,i,t} \times p_{2,t} \quad (9)$$

$$r_{3,i,t} = e_{3,i,t} \times p_{3,t} \quad (10)$$

$$r_{4,i,t} = e_{4,i,t} \times p_{4,t} \quad (11)$$

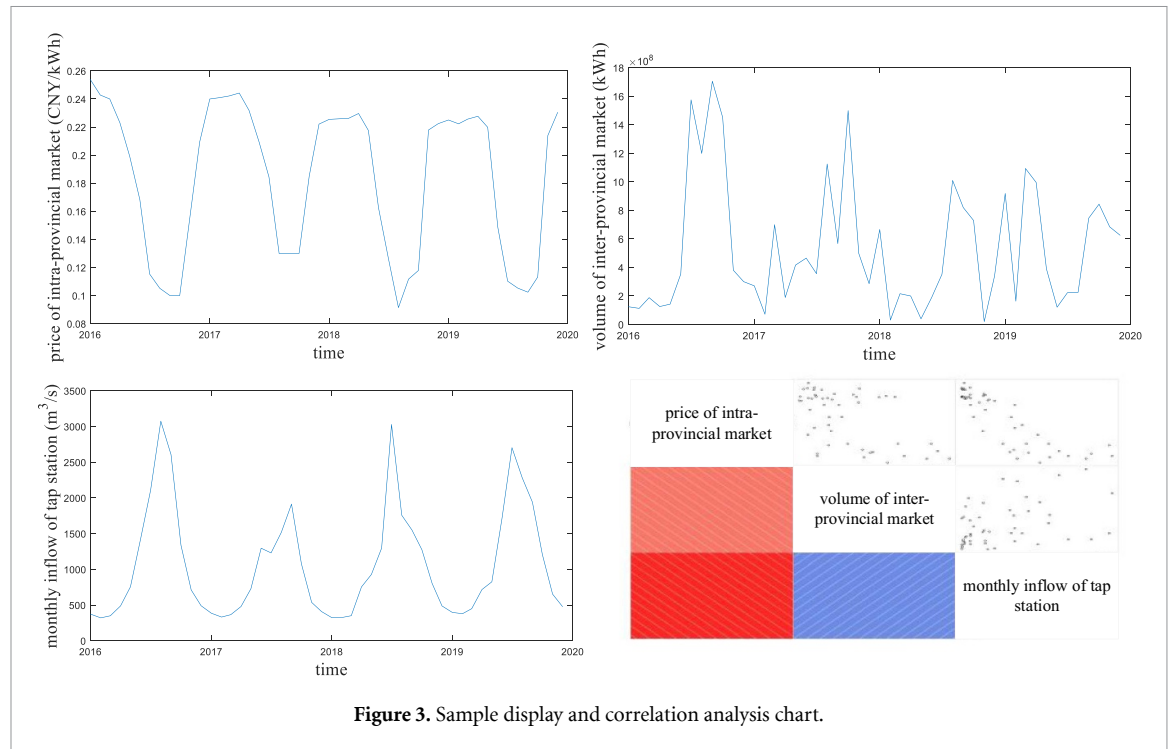
$$e_{4,i,t} = e'_{2,i,t} - e_{2,i,t} \quad (12)$$

$$p_{4,t} = e_{4,i,t} / e'_{2,i,t} \times p_{2,t} \times 0.5 \quad (13)$$

$$e_{i,t} = \eta_i \times Q_{i,t} \times H_{i,t} \times \Delta t \quad (14)$$

$$H_{i,t} = \frac{Z_{i,t-1} + Z_{i,t}}{2} - Z_{i,t}^d - H_{i,t}^d \quad (15)$$

where R is the total revenue of the cascade power station, $r_{j,i,t}$ represents the profit or loss of station i in the market in the t th month, $j = 1$ is the revenue of the clean energy priority consumption market, $j = 2$ is the revenue of the inter provincial market, $j = 3$ is the revenue of the intra provincial market, $j = 4$ is the penalty loss for breach of contract, and N is the number of cascade power stations. T is the



total dispatching period, and $r_{1,i,t}$, $e_{1,i,t}$ and $p_{1,t}$ are the priority consumption market revenue and settlement volume and price of station i in the t th month, respectively. $r_{2,i,t}$, $e_{2,i,t}$ and $p_{2,t}$ are the inter provincial market revenue, settlement volume and price of station i in the t th month, respectively, and $i = 1$. $r_{3,i,t}$, $e_{3,i,t}$ and $p_{3,t}$ are the intra provincial market revenue, settlement volume and price of station i in the t th month, respectively. where $r_{4,i,t}$, $e_{4,i,t}$, $e'_{2,i,t}$ and $p_{4,t}$ are the penalty loss, default amount and transaction volume in the inter provincial market and penalty price of the tap station in the t th month, $i = 1$. $e_{i,t}$, $Q_{i,t}$, $H_{i,t}$, $Z_{i,t}$, $Z^d_{i,t}$ and $H^d_{i,t}$ are the average generating capacity, average generating flow ($\text{m}^3 \text{s}^{-1}$), head, decision level, downstream tail water level and head loss of the hydropower station i in the t th month, respectively; η_i is the output coefficient of hydropower station i ; and Δt is the time step.

3.2. Constraints

In addition to the conventional hydraulic constraints, the constraints in the model also include revenue constraints, market constraints and opportunity constraints. Due to the limitation of paper space, see the appendix B for details.

3.3. Model solving method

The traditional method to solve a chance-constrained problem transforms the CCP model into its equivalent deterministic model according to the given confidence level in advance, and then solves it. However, in the mathematical model of the long-term optimal operation of cascade hydropower stations in the multimarket, there are many random variables, and the

opportunity constraints are too complex. This is usually difficult to achieve by traditional methods. Therefore, we use a Monte Carlo simulation method to deal with the random variables, and then use the PSO method to solve them. The specific steps are shown in figure 2.

4. Results

Based on the electricity market dominated by hydropower in a province of China, see appendix A1 for the main operation parameters of power stations A, B and C involved in the calculation. A is the tap station, and the output process directly affects the operation of the downstream stations. The long-term scheduling period is one year, the time period is one month, and the number of Monte Carlo sampling is 1000. The number of population is 100, the maximum number of iterations is 5000, the inertia factor is 0.9, and the learning factor is 2.

4.1. Multivariate correlation analysis

The random variables studied in this paper are the monthly inflow of the tap station, the clearing price of the intra provincial market and volume of the inter provincial market. The samples used in this paper involve 48 months from January 2016 to December 2019. The details of the samples are shown in figure 3. The three variables show obvious seasonality and volatility. At the same time, it can be seen from (d) that there are correlations among the variables. Red represents a negative correlation, blue represents a positive correlation; and the deeper the colour, the stronger the correlation (the Sperson coefficient of

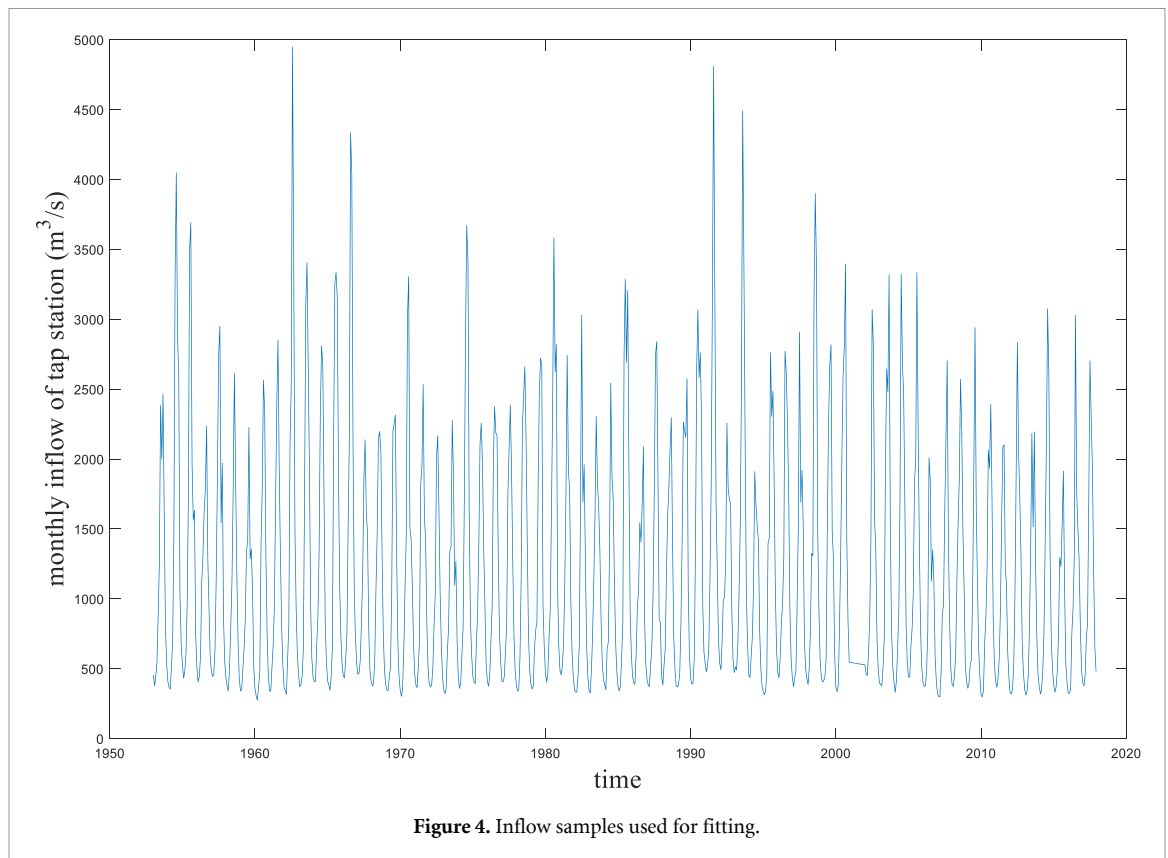


Figure 4. Inflow samples used for fitting.

price and volume is -0.55296 , that of the volume and inflow is 0.39326 and that of the inflow and price is -0.84908).

For the multivariate variables with strong correlation in this paper, a copula function can be used to analyse the correlation between variables and contain almost all the details without determining whether the traditional linear correlation coefficient can correctly measure the correlation between variables. For marginal distributions, the traditional method assumes that the variance of time-series variables is fixed, which is not in line with reality. For example, inflow series show obvious seasonality. The ARIMA model can only capture the linear relationship of a time series but cannot capture the nonlinear relationship, which leads to an error in the simulation results and deviations from reality.

However, the SARIMA model takes into account the seasonal factors and carries out a 12-step difference simultaneously with the first-order difference of time series, thus eliminating the seasonal effect and solving the above problems. Due to the short development time of the market studied in this paper, the numbers of price and volume in the sample are too small, so only the marginal distribution of the inflow series is fitted (inflow samples from 1953 to 2019). The inflow details are shown in figure 4. The marginal distribution of other variables was fitted by the kernel density estimation method.

Before fitting the SARIMA model, it is necessary to conduct a basic test of the inflow data. The result of the JB test is 1. The original hypothesis that the series has a normal distribution is rejected, which means that the distribution is not a normal distribution. The result of the ADF unit root test does not equal 1, the original hypothesis of the unit root exists in the rejection series, and the sample is stationary. ACF and PACF were drawn to test the autocorrelation. The results of the Ljung box test were 1, which rejected the original hypothesis of white noise, and the series was not white noise. The specific test results are shown in figure 5. The basic test shows that the SARIMA model can be used to fit the inflow series.

The values of model p , d , q , P , D and Q are identified and processed by the R language, and the optimal model is SARIMA (1, 0, 1) (2, 1, 1) [12]. Then, the parameters of the model are estimated by the MLE method. The fitting results are listed in table 1. Then, 100 sets of inflow processes are generated, and the results are shown in figure 6. The model SARIMA (1, 0, 1) (2, 1, 1) [12] was used to simulate the runoff series from 1953 to 2019, the root mean square error is 11.945%, which indicates that the fitting runoff process is in good agreement with the actual runoff.

According to the method described in section 2.2.1, the marginal distribution of each variable is transformed into a uniform distribution subject to $[0, 1]$ through probability integral

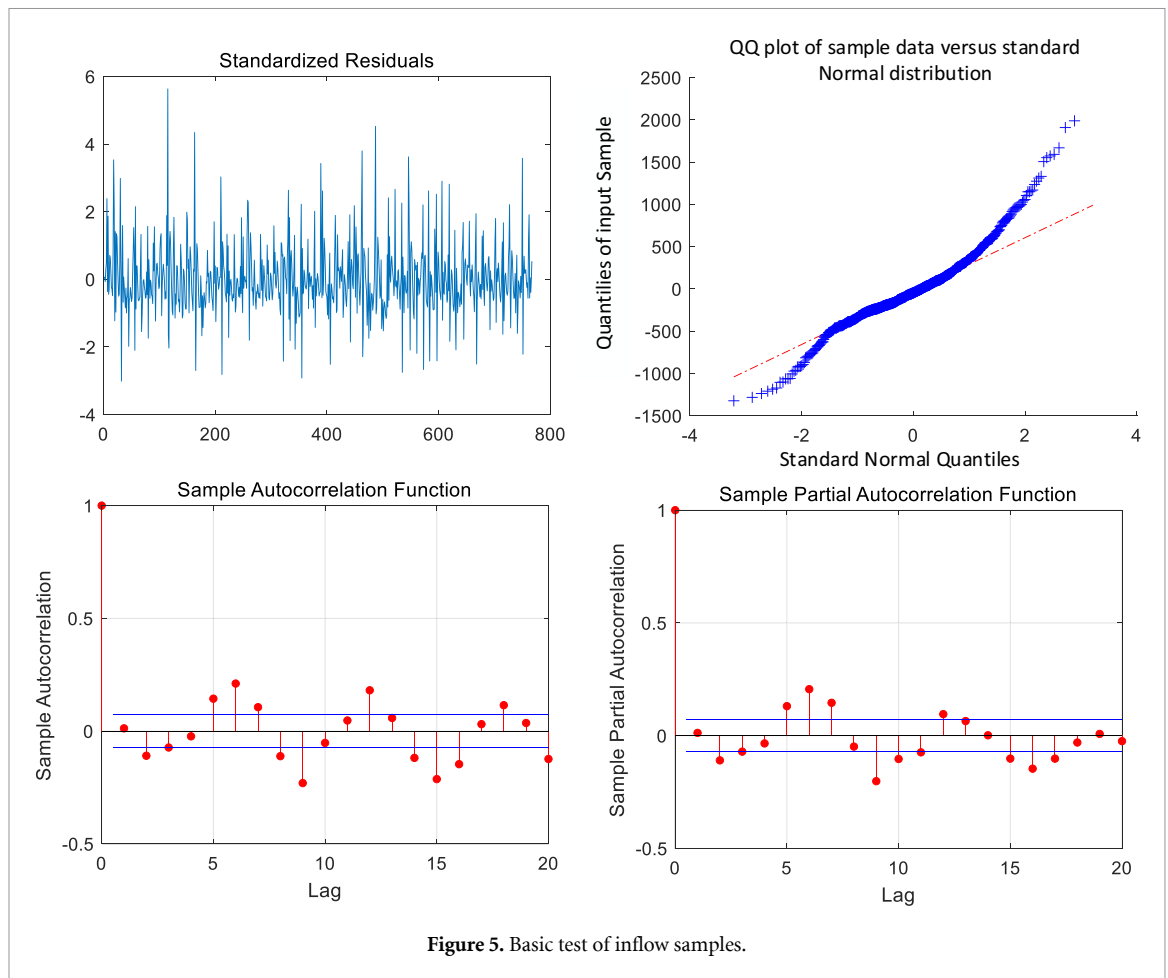


Figure 5. Basic test of inflow samples.

Table 1. Parameter estimation and test comparison of SARIMA.

Model	AR(1)	MA(1)	SAR(1)	SAR(2)	SMA(1)
Parameter	0.4977	0.0172	0.0193	0.0937	−0.9486
s.e.	0.0746	0.0877	0.0483	0.0478	0.0288

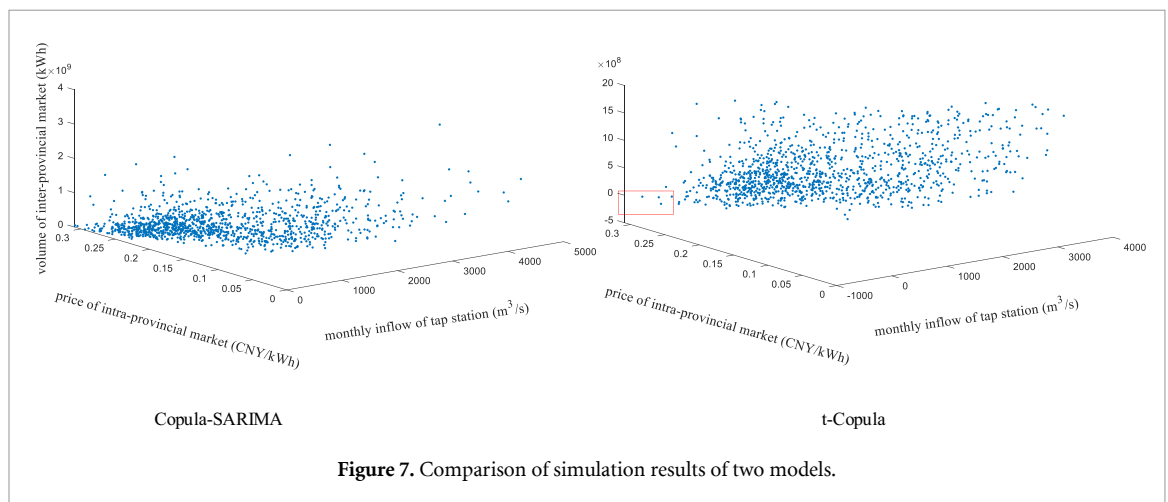
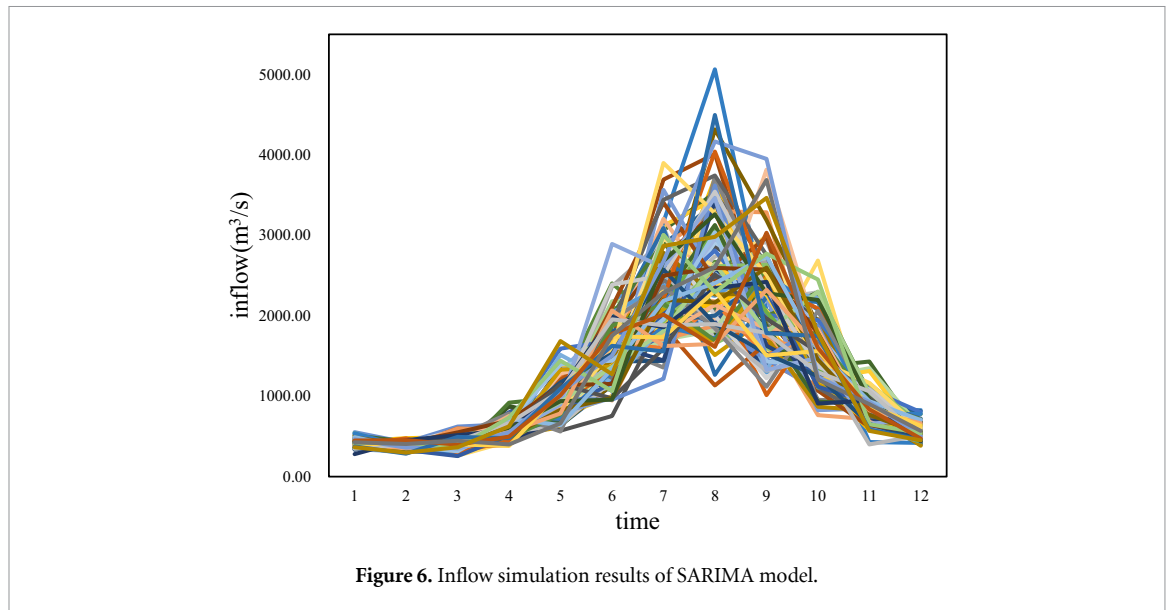
transformation, in which the inflow series uses some data after marketization from 2016 to 2019, and uses the marginal distribution of the whole sample for probability integration to obtain a uniform distribution. Then, the t-copula function proposed in section 2.2.2 is obtained by the two-stage maximum likelihood method of nonparametric estimation. One hundred sets of scenarios were randomly generated, and the correlation coefficients between variables were calculated for verification (the Sperson coefficient of price and electricity quantity is -0.45609 , electricity quantity and inflow are 0.346654 , and inflow and price are -0.87613). Finally, a comparison is made with the t-copula function without the SARIMA model, and the results are shown in figure 7. It can be seen from the figure that inflow and price have negative value (red box) in the scenarios generated by the conventional method, while the method proposed in this paper can generate more reasonable scenarios. The reason for this situation is that the fluctuation

range of runoff is large, when only using t-copula to generate simulation data, we need to use the marginal distribution function obtained by kernel density estimation to transform. In the face of extreme cases, the stochastic runoff process obtained by transformation is prone to negative values. The runoff stochastic process generated by SARIMA model can avoid the above situation, and t-copula can better fit the tail dependence, so t-copula with SARIMA adjustment can work better.

4.2. Analysis of long-term optimal scheduling results

4.2.1. Analysis of optimal operation results of cascade hydropower stations under multimarket coupling

A scenario was generated by random sampling according to the joint distribution fitted in this paper. This was calculated by the PSO algorithm and compared with the results of the maximum generating model. The calculation results are shown in figure 8. First, with regard to the water level, the operation results of the downstream reservoir water level with poor regulation capacity are the same, and the water level of the tap power station varies greatly in the flood season. In terms of generation, the conventional dispatching mode generates more power in the flood season and less in the dry season. The model in this paper



does not follow this law, and the total power is lower than in the conventional model.

The resource allocation results in the above multi-market coupling model are shown in figure 9 (since B and C do not participate in the delivery market, only B is shown). It can be seen from the figure that when participating in the coupling market, stations tend to generate more electricity when the price is high and less when the price is low, and allocate more volume in the market with high price and less volume in the market with low price, or break even on the contract. Given the factors of price, head and inflow, the month with the largest generating capacity does not appear when the price is the highest. The influence of settlement price and assessment price is taken into account when A allocates resources. In May, for example, the settlement price in the clean energy priority consumption market is the highest. To increase revenue, all power resources should be allocated to this market. However, due to the default assessment mechanism of the inter provincial market, some of the resources need to be completed to avoid additional losses. B

allocates power resources in the inter provincial market and clean energy priority consumption market according to the price.

4.2.2. Analysis of dispatching results under specified risk tolerance

According to the model described in section 3, change the risk α , β_1 and β_2 that the power station is willing to bear, and then calculate and analyse the scheduling schemes of situation 1 (considering the uncertainty of all variables), situation 2 (considering the uncertainty of price and inflow), situation 3 (considering the uncertainty of price and volume) and situation 4 (considering the uncertainty of inflow and volume). The results are listed in table 2. It can be seen from table 2 that with an increase in the willingness of stations to bear risk, the possible generation power and target profit also increase. Further analysis shows that the change in α has an obvious effect on the generation revenue.

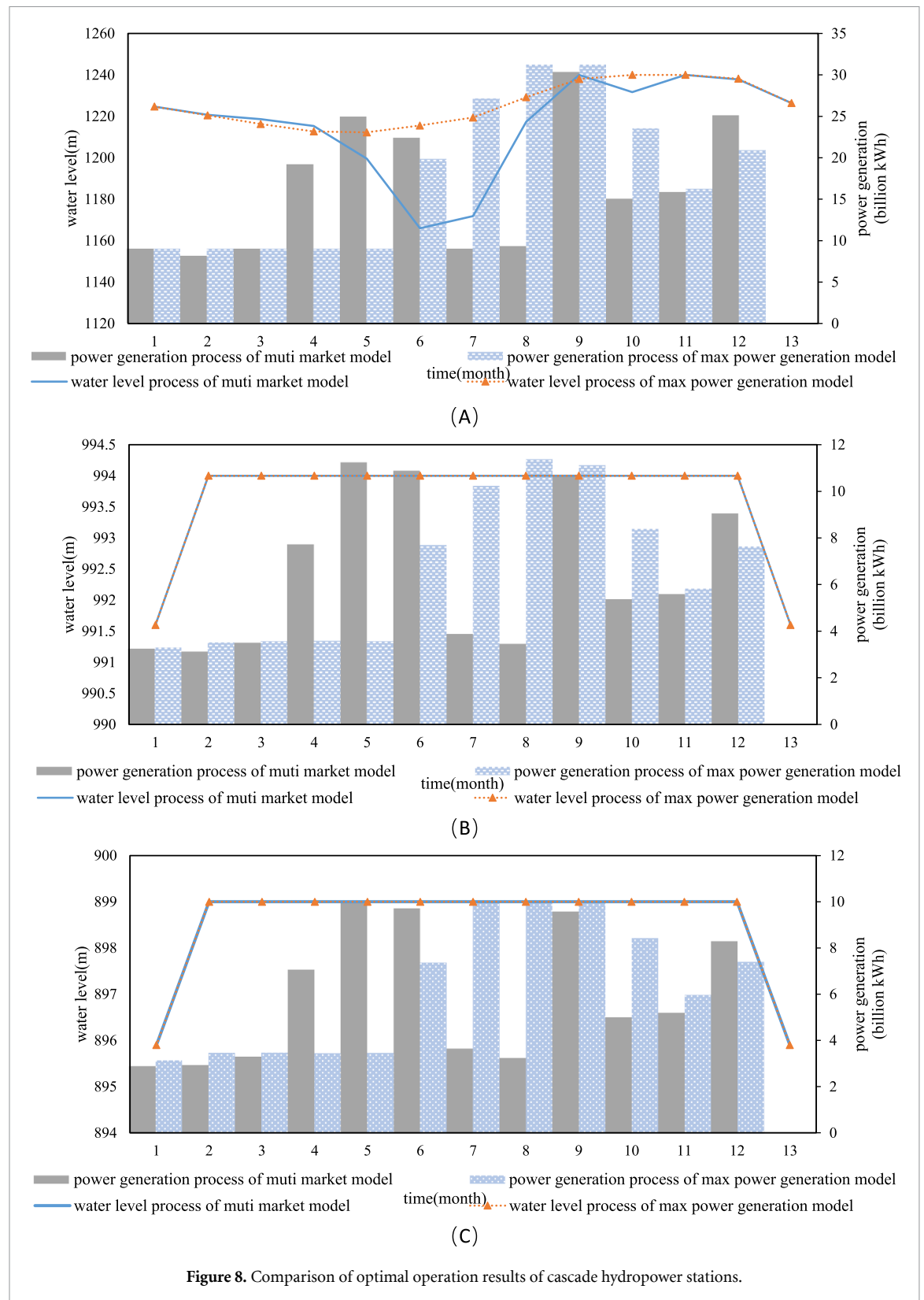


Figure 8. Comparison of optimal operation results of cascade hydropower stations.

To analyse the impact of risk tolerance α in the above four situations objective functions on the long-term operation income of cascade hydropower stations, results corresponding to different α in different situations are plotted in figure 10. When the risk tolerance is high (risk preference) and medium (risk neutral), situation 4 has the

highest income. Situation 1 comes next. Situation 2 and situation 3 are similar. When the risk tolerance is low (risk aversion), situation 3 has the highest income. Situation 4 comes next, and situation 2 and situation 1 are similar. When the risk is high, compared with fluctuations when the electricity price, inflow and power are more volatile,

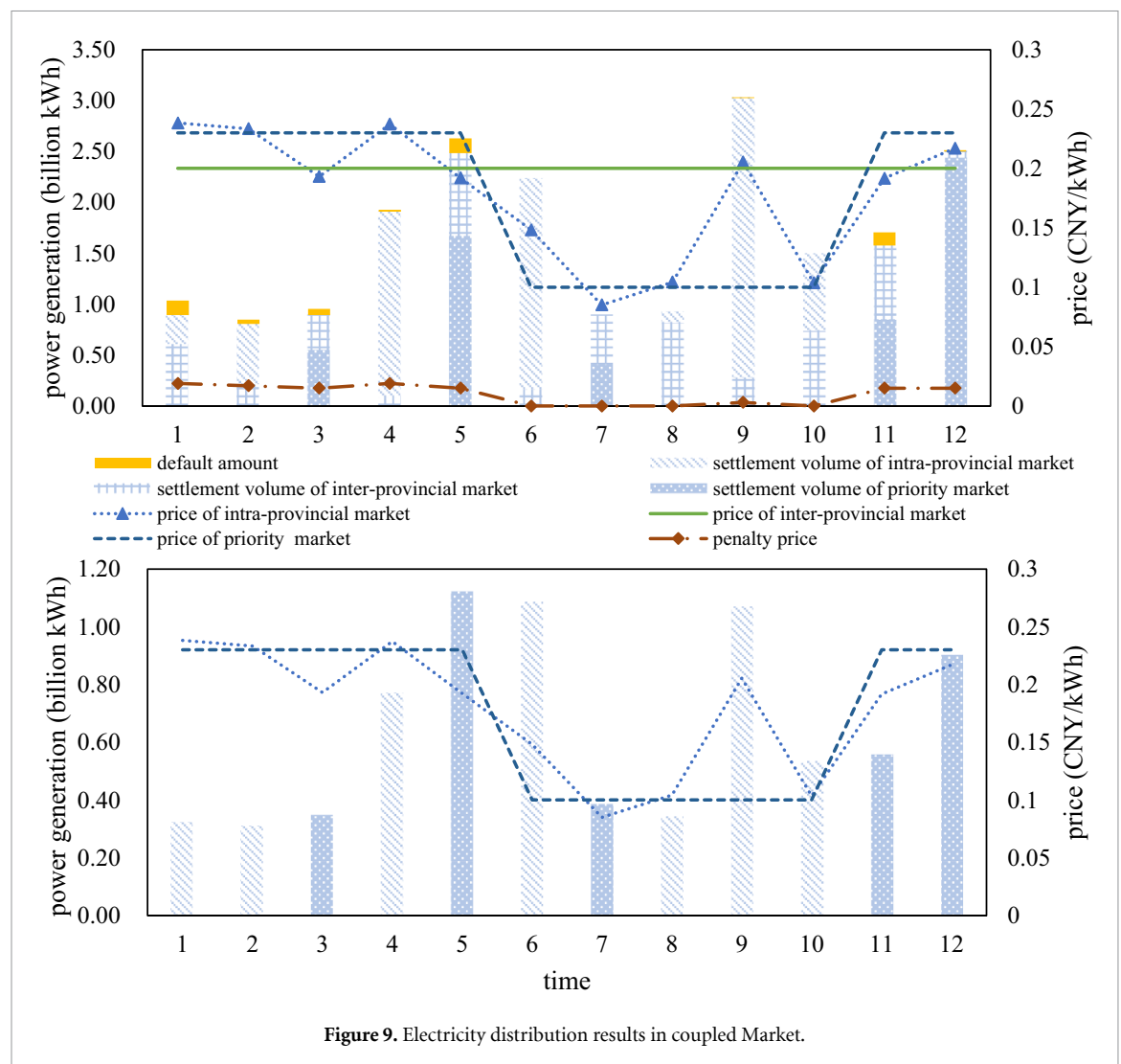


Figure 9. Electricity distribution results in coupled Market.

Table 2. The influence of willing to take risks on scheduling schemes in different situations.

Risk/ (1- α)	Risk/ (1- β_1)	Risk/ (1- β_2)	Generation revenue (billion CNY)			
			1	2	3	4
0.2	0.8	0.8	6.79270	6.64658	6.88530	6.88258
0.5	0.8	0.8	7.12256	6.95439	7.02261	7.24354
0.8	0.8	0.8	7.38562	7.29536	7.27267	7.58284
0.9	0.9	0.9	7.57898	7.36454	7.33506	7.90049
0.95	0.9	0.9	7.66697	7.40683	7.35844	7.92991
0.95	0.95	0.95	7.68509	7.46514	7.43070	7.93850
0.95	0.99	0.99	7.70521	7.48250	7.51599	8.01121
0.99	0.99	0.99	7.76063	7.72186	7.53619	8.02226

large flows bring more power generation and can allocate more delivery resources to obtain more revenue. When the risk is low, the inflow tends to dry up, and the allocated amount of volume in the inter provincial market is less. Although the intra provincial market price increases, the generating capacity greatly reduces, leading to an overall decrease of income. Situation 3 does not consider the uncertainty of inflow, which has little impact on the increase in power generation, so the revenue is the highest. Situation 4 does not consider

the uncertainty of the intra provincial market price, which has little impact on the price fluctuation in the flood season, so the revenue comes next. Situation 1 and situation 2 take into account the variables that play a major role in revenue, resulting in lower price during the flood season and reduced water supply, so they have the lowest revenue. The results of the model are consistent with the actual situation, so it can be seen that the copula-SARIMA model proposed in this paper can be used to reasonably evaluate the uncertainty factors brought

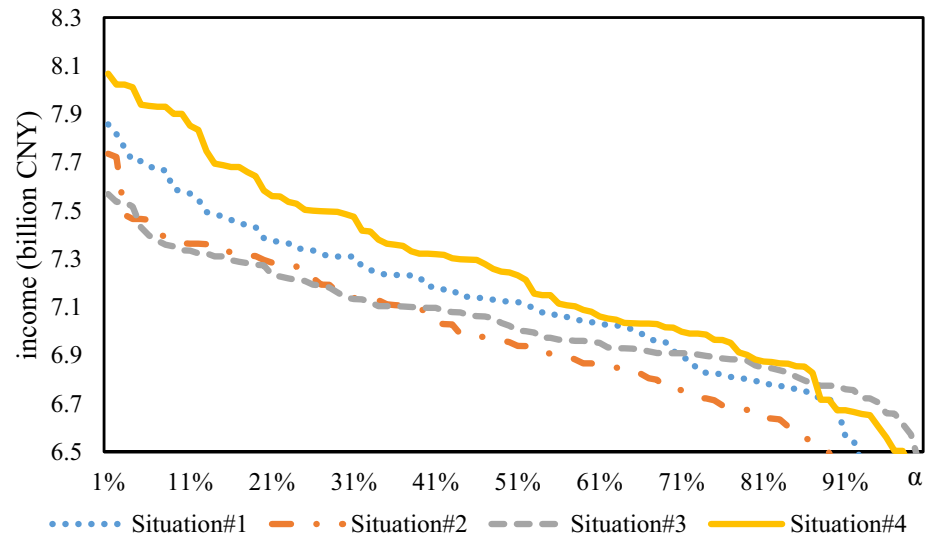


Figure 10. Income corresponding to different risk tolerance in different scenarios.

by cascade stations participating in multimarket coupling.

5. Conclusions

- (a) There are many uncertain factors in the market operation of cascade hydropower: inflow, price, etc. These are all time series. By using a Copula-SARIMA model, we combined the research of the correlation degree and correlation mode of a multivariate situation, and considered the volatility and time-varying nature of the time series at the same time. Compared with the constant correlation model, which considers the uncertainty of variables alone, the fitting data generated by this method maintains the correlation between samples, making the simulation experiment more reasonable.
- (b) Large-scale cascade hydropower stations play a role in regulating and storing natural inflow due to their large storage capacity. In the medium- and long-term market environment, water resources can be redistributed in time and space according to the situation of incoming water and market clearing, to achieve the efficient utilization of water resources, improve income and reduce the

default assessment caused by the randomness of inflow.

- (c) As a signal to guide the resource allocation of cascade hydropower stations in a variety of coupled markets, considering the uncertainty of inflow, volume and other factors at the same time, the price can encourage the market participants to evaluate income more reasonably and can provide an effective reference for the cascade hydropower stations to arrange the operation scheme of each time scale, formulate the transaction plan, optimize the resource allocation and obtain more benefits.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants 51879030 and 51709035. We also thanks to the anonymous reviewers whose comments have helped improve the quality of this paper.

Appendix A

Table A1. Main characteristic parameters of cascade hydropower stations.

Hydropower station	A	B	C
Operation	Multi-year	Seasonal	Daily
Normal water level (m)	1240	994	899
Dead water level (m)	1166	988	882
Maximum outflow ($\text{m}^3 \text{s}^{-1}$)	20 683	16 805	5000
Minimum outflow ($\text{m}^3 \text{s}^{-1}$)	0	0	0
Output coefficient	9.4	8.5	8.5
Starting water level (m)	1224.7	991.6	895.9
End water level (m)	1226.4	991.6	895.9
Guaranteed power generation (million kW)	4.2	1.67	1.35
Minimum power generation (million kW)	0	0	0
Maximum power generation flow ($\text{m}^3 \text{s}^{-1}$)	2202	2233	2085

Appendix B

$$Z_{i,0} = Z_{i,\text{begin}} \quad (20)$$

(a) Profit constraint.

$$Z_{i,T} = Z_{i,\text{end}} \quad (21)$$

$$P(R \geq \bar{R}) \geq \alpha. \quad (16)$$

(b) Hydraulic connection of upstream and downstream reservoirs.

$$q_{i,t} = Q_{i-1,t} + SP_{i-1,t} + q'_{i,t} \quad (17)$$

where $q_{i,t}$ is inflow of station i in the t th month ($\text{m}^3 \text{s}^{-1}$); $Q_{i-1,t}$ and $SP_{i-1,t}$ are the generation flow ($\text{m}^3 \text{s}^{-1}$) and wastewater flow ($\text{m}^3 \text{s}^{-1}$) of hydropower station $i - 1$ in the t th month, respectively; and $q'_{i,t}$ is the interval inflow between station $i - 1$ and station i in the t th month ($\text{m}^3 \text{s}^{-1}$). Note that due to the large regulating effect of the tap station, the downstream interval discharge $q'_{i,t}$ is treated as constant.

(c) Water balance constraint.

$$V_{i,t} = V_{i,t-1} + 3600 \times (q_{i,t} - Q_{i,t}) \times \Delta t \quad (18)$$

where $V_{i,t}$ is the storage capacity of station i at the end of the t th month (m^3).

(d) Reservoir water level constraint.

$$Z_{i,\min} \leq Z_{i,t} \leq Z_{i,\max} \quad (19)$$

where $Z_{i,\min}$ and $Z_{i,\max}$ are the lowest and highest water levels in front of the dam of reservoir i , respectively.

(e) Water level at the beginning and end stage constraint.

where $Z_{i,\text{begin}}$ is the water level of station i at the beginning of the month (m). $Z_{i,\text{end}}$ is the water level of station i at the end of the month (m).

(f) Water discharge constraint.

$$P(Q'_{i,\min} \leq Q_{i,t} + SP_{i,t} \leq Q'_{i,\max}) \geq \beta_1 \quad (22)$$

where $Q'_{i,\min}$ and $Q'_{i,\max}$ are the minimum and maximum water discharge of station i , respectively ($\text{m}^3 \text{s}^{-1}$).

(g) Generating discharge constraint.

$$Q_{i,\min} \leq Q_{i,t} \leq Q_{i,\max} \quad (23)$$

where $Q'_{i,\min}$ and $Q'_{i,\max}$ are the minimum and maximum generation discharge of station i , respectively ($\text{m}^3 \text{s}^{-1}$).

(h) Output constraint.

$$P(P_{i,\min} \leq \eta_i \times Q_{i,t} \times H_{i,t} \leq P_{i,\max}) \geq \beta_2. \quad (24)$$

(i) Water level storage capacity relation constraint.

$$V_{i,t} = f_{i,zv}(Z_{i,t}) \quad (25)$$

where $f_{i,zv}(\bullet)$ is the relationship function of the water level and storage capacity of station i .

(j) Tail water level discharge relation constraint.

$$Z^d_{i,t} = f_{i,zq}(Q_{i,t}) \quad (26)$$

where $f_{i,zq}(\bullet)$ is the relationship function of the tail water level and discharge of station i .

(k) Restriction of settlement electricity quantity in each market.

$$e_{1,i,t} \leq e_{i,t} \quad (27)$$

$$e_{2,i,t} \leq \min(e_{i,t} - e_{1,i,t}, e'_{2,i,t}) \quad (28)$$

$$e_{3,i,t} \leq e_{i,t} - e_{1,i,t} - e_{2,i,t} \quad (29)$$

$$e_{i,t} = e_{1,i,t} + e_{2,i,t} + e_{3,i,t} \quad (30)$$

ORCID iD

Jia Lu  <https://orcid.org/0000-0002-7903-2579>

References

- Allen R B and Bridgeman S G 1986 Dynamic Programming in Hydropower Scheduling *J. Water Resour. Plan. Manage.* **112** 339–53
- Barforoushi T, Moghaddam M P, Javidi M H and Sheikholeslami M K 2006 A new model considering uncertainties for power market *Iran. J. Electr. Electron. Eng.* **2** 71–81 (available at: <https://ijeee.iust.ac.ir/article-1-16-en.html>)
- Barros M T, Tsai FT-C, Yang S, Lopes J E and Yeh WW-G 2003 Optimization of Large-Scale Hydropower System Operations *J. Water Resour. Plann. Manage.* **129** 178–88
- Bourry F, Costa L M and Kariniotakis G 2009 Risk-based strategies for wind/pumped-hydro coordination under electricity markets *Powertech IEEE Bucharest* (www.researchgate.net/publication/224601999_Risk_based_strategies_for_windpumped_hydro_coordination_under_electricity_market)
- Chen F, Huang G H, Fan Y R and Chen J P 2017 A copula-based fuzzy chance-constrained programming model and its application to electric power generation systems planning *Appl. Energy* **187** 291–309
- Chen F, Huang G H, Fan Y R and Wang S 2016 A copula-based chance-constrained waste management planning method: an application to the City of Regina *J. Air Waste Manage. Assoc.* **66** 307
- Cheng C, Chen F, Li G, Bora R, Ali M, Tu Q and Kaveh M 2018 Reform and renewables in China: the architecture of Yunnan's hydropower dominated electricity market *Renew. Sustain. Energy Rev.* **94** 682–93
- Conejo A J, Nogales F J and Arroyo J M 2003 *Price-taker Bidding Strategy under Price Uncertainty* (available at: <https://ieeexplore.ieee.org/xpl/abstractKeywords.jsp?arnumber=1270446>)
- Faria E and Fleten S-E 2011 Day-ahead market bidding for a Nordic hydropower producer: taking the Elbas market into account *Comput. Manage. Sci.* **8** 75–101
- García-González J, Parrilla E and Mateo A 2007 Risk-averse profit-based optimal scheduling of a hydro-chain in the day-ahead electricity market *Eur. J. Oper. Res.* **181** 1354–69
- Gianfreda A 2010 Volatility and volume effects in European electricity spot markets *Econ. Notes* **39** 47–63
- Gianfreda A, Grossi L and Olivieri D 2010 Volatility structures of the Italian electricity market: an analysis of leverage and volume effects (available at: www.researchgate.net/publication/251949675_Volatility_structures_of_the_Italian_electricity_market_An_analysis_of_leverage_and_volume_effects)
- Grygier J C and Stedinger J R 1985 Algorithms for optimizing hydropower system operation *Water Resour. Res.* **21** 1–10
- Gu W, Shao D, Tan X, Shu C and Wu Z 2017 Simulation and Optimization of Multi-Reservoir Operation in Inter-Basin Water Transfer System *Water Resour. Manage.* **31** 3401–412
- Guin A 2006 Travel time prediction using a seasonal autoregressive integrated moving average time series model *IEEE Intelligent Transportation Systems Conf.* (available at: www.researchgate.net/publication/224650691_Travel_Time_Prediction_Using_a_Seasonal_Autoregressive_Integrated_Moving_Average_Time_Series_Model)
- Hansen P V 2010 Inflow uncertainty in hydropower markets *Scand. J. Econ.* **111** 189–207
- Kang C, Bai L, Xia Q, Jiang J and Zhao J 2005 Incorporating reliability evaluation into the uncertainty analysis of electricity market price *Electr. Power Syst. Res.* **73** 205–15
- Labadie J W 2004 Optimal Operation of Multireservoir Systems: State-of-the-Art Review *J. Water Resour. Plann. Manage.* **130** 93–111
- Li D X 1999 On default correlation: a copula function approach *SSRN Electron. J.* **9** 43–54
- Li X, Guo S, Liu P and Chen G 2010 Dynamic control of flood limited water level for reservoir operation by considering inflow uncertainty *J. Hydrol.* **391** 124–32
- Liu B, Cheng C, Wang S, Liao S, Chau K W, Wu X and Li W 2018 Parallel chance-constrained dynamic programming for cascade hydropower system operation *Energy* **165** 752–67
- Liu P, Lin K and Wei X 2015 A two-stage method of quantitative flood risk analysis for reservoir real-time operation using ensemble-based hydrologic forecasts *Stochastic Environ. Res. Risk Assess.* **29** 803–13
- Liu S, Qiang Y, Huaxiang C, Minghui Y and Xie M 2019 Market reform of Yunnan electricity in southwestern China: practice, challenges and implications *Renew. Sustain. Energy Rev.* (<https://doi.org/10.1016/j.rser.2019.109265>)
- Moghaddam I G, Nick M, Fallahi F, Sanej M and Mortazavi S 2013 Risk-averse profit-based optimal operation strategy of a combined wind farm–cascade hydro system in an electricity market *Renew. Energy* **55** 252–9
- Nikolouloupoulos A K, Harry J and Haijun L 2009 Extreme value properties of multivariate t copulas *Extremes* **12** 129–48
- Nojavan S, Mohammadi-Ivatloo B and Zare K 2015 Robust optimization based price-taker retailer bidding strategy under pool market price uncertainty *Int. J. Electr. Power Energy Syst.* **73** 955–63
- Oliveira G C, Costa J P, Rio I P and Campos D 2016 Hydrothermal scheduling policy stability under inflow uncertainty (available at: www.researchgate.net/publication/228402186_Hydrothermal_scheduling_policy_stability_under_inflow_uncertainty)
- Saadaoui F 2013 The price and trading volume dynamics relationship in the EEX power market: a wavelet modeling *Comput. Econ.* **42** 47–69
- Saadaoui F and Mrad M 2017 Stochastic modelling of the price-volume relationship in electricity markets: evidence from the Nordic energy exchange *Eur. Trans. Electr. Power* **27** 1–15
- Shrestha G B, Pokharel B K, Lie T T and Fleten S E 2005 Medium term power planning with bilateral contracts *IEEE Trans. Power Syst.* **20** 627–33
- Thanos T and Yeh W W-G 2010 Use of stochastic dynamic programming for reservoir management *Water Resour. Res.* **23** 983–96
- Ventosa M, García-Alcalde A, Mencia A, Rivier M and Ramos A 2016 Modeling inflow uncertainty in electricity markets: a stochastic MCP approach (available at: www.researchgate.net/publication/228957581_Modeling_Inflow_Uncertainty_in_Electricity_Markets_A_Stochastic_MCP_Approach)
- Zareipour H, Canizares C A and Bhattacharya K 2010 Economic impact of electricity market price forecasting errors: a demand-side analysis *IEEE Trans. Power Syst.* (<https://doi.org/10.1109/TPWRS.2009.2030380>)
- Zong-Run W, Xiao-Hong C, Yan-Bo J and Yan-Ju Z 2010 Estimating risk of foreign exchange portfolio: using VaR and CVaR based on GARCH–EVT-copula model—scienceDirect *Physica A* **389** 4918–28