Characterizing the Hydraulic Connection of Cascade Reservoirs for Short-Term Generation Dispatching via Gaussian Process Regression

XIAO CHEN¹,², JIANZHONG ZHOU¹,², BENJUN JIA¹,², XIN YANG¹,², AND CHAO ZHOU³

¹School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
²Hubei Key Laboratory of Digital Valley Science and Technology, Wuhan 430074, China
³Changjiang Institute of Survey, Planning, Design and Research, Wuhan 430010, China
Corresponding author: Jianzhong Zhou (jz.zhou@hust.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFC0402205 and Grant 2016YFC0402210, and in part by the National Natural Science Foundation Key Project of China under Grant 91547208 and Grant U1865202.

ABSTRACT The characterization of the mapping relationship (MR) between outflow of the upstream reservoir (OUR) and inflow of the downstream reservoir (IDR) in the short-term generation dispatching of cascade reservoirs (SGDCR) greatly impacts the safe and economic operation of hydropower plants. If this MR is not characterized properly, the operation process of hydropower stations will deviate from the planning dispatching schemes. Especially when the upstream reservoir (UR) undertakes peak load regulation tasks frequently and the downstream reservoir (DR) owns weak re-regulation capacity, the safety of cascade reservoirs will be seriously threatened. In this extreme system, the commonest characterizing models in SGDCR, such as the lag time (LT) model and the Muskingum model, may cause a huge deviation when simulating or forecasting the IDR. Given this dilemma, the Gaussian process regression (GPR) model, which is a representative for Bayesian regression methods, is firstly introduced to handle this MR and compared with the LT and the Muskingum model in this paper. The Three Gorges-Gezhouba cascade reservoirs (TGGCR) in China are selected as a typical case study. The performance indicators of four categories are adopted to evaluate the simulated IDR in a single period and a set of pivotal indicators are proposed to estimate the simulation dispatching in multiple periods. The results show that (1) The GPR model reduces the mean absolute deviation (MAD) about inflow of Gezhouba by 197 m³/s and 263 m³/s than the LT and the Muskingum model respectively; (2) The distribution characteristics of inflow deviation produced by the GPR model are most competitive. Meanwhile, the GPR model has the strongest ability when conducting the multi-period simulation dispatching and owns best applicability on the whole range of streamflow. (3) The Muskingum model is not recommended to characterize the hour-scale hydraulic connection in the extreme system of SGDCR.

INDEX TERMS Cascade reservoirs, short-term generation dispatching, hour-scale hydraulic connection, Gaussian process regression.

I. INTRODUCTION

Large reservoirs with large installed capacity are usually equipped with immediate downstream reservoirs to buffer the sharp fluctuation of outflow caused by the peak load regulation in a short period [1], [2]. In principle, the downstream reservoir formulates the power generation scheme according to the outflow process of the upstream reservoir, the forecasted local inflow process and the own re-regulation ability, so as to achieve the maximum power generation revenues of cascade reservoirs [1], [3]--[5]. In SGDCR, the hydraulic connection mainly consists of the MR of streamflow and the jacking effect between reservoirs[1], [2]. (In this paper, the hydraulic connection mainly refers to the MR of streamflow.) Take the TGGCR in China as an example. As the regeneration reservoir of the Three Gorges reservoir (TGR), the Gezhouba reservoir, which is a run-of-river hydropower plant, only owns a re-regulation capacity of 85 million m³ [1].
There are no large tributaries between the two dams in geography, and the local inflow can also be ignored [3]. Therefore, the dispatching scheme of Gezhouba is largely influenced by the outflow process of TGR in short-term or real-time joint dispatching. It’s worth mentioning that the frequent peak adjustments of TGR make the hour-scale and minute-scale generation dispatching of Gezhouba tougher and tougher [1]. In general, the hydraulic connection will be characterized roughly because of the aim of simplifying generation dispatching models and the difficulty in handling the aftereffect of streamflow in most studies on SGDCR [1]–[3], [5], [6]. But in this extreme system, the effective characterization on the hydraulic connection between the outflow of TGR and the inflow of Gezhouba becomes particularly important, which can avoid frequent adjustments on the developed power generation scheme and ensure the safety of cascade reservoirs [7], [8]. Unfortunately, when communicating with site dispatchers at the Three Gorges Cascade Reservoirs Communication and Control Center (TGCRCCC), we learn that this specific problem is mainly solved by manual experience of dispatchers. Hence, the characterization on the hour-scale hydraulic connection of TGGCR is fairly tricky.

Among all characterizing methods about the hydraulic connection, the hydrodynamic numerical simulation (HNS) model owns the highest accuracy and can accurately simulate the streamflow propagation process in the river channel even if the DR has a jacking effect [9]–[11]. Many studies have confirmed that most of the streamflow motion in the river channel is turbulent, and many physical parameters of streamflow, such as speed, temperature and streamflow, vary randomly with time and space [1], [9], [11], [12]. Although the turbulent motion is complex, the Saint-Venant equations can always capture the characteristics of the instantaneous turbulence motion [1], [13]. However, the HNS model generally requires high data integrity, such as topography data and the roughness of the river channel [11]. Meanwhile, the solving the Saint-Venant equations is time-consuming [1], [11]. Therefore, the HNS model rarely appears in short-term or real-time generation dispatching of cascade reservoirs. In general, the HNS models are utilized in the dispatching problem with low timeliness, such as the extraction of scheduling rules [1].

Considering the disadvantages of the HNS model, some simplified characterizing models on the hydraulic connection appear in vast studied on SGDCR. The LT model, which is equipped with simplicity and efficiency, is the commonest solution [14], [15]. The main point of the LT model is that the OUR is propagated to the downstream reservoir after a certain period, which is called the lag time [6], [16]. In this way, the existing classical algorithms such as dynamic programming (DP) algorithm and progressive optimality algorithm (POA) for solving the SGDCR problem can be utilized directly [3], [17]. Besides, some scholars regard LT as a piecewise function about streamflow, and consider the dynamic variation characteristics of LT, making the SGDCR a problem with aftereffect. Considering the difficulty in dealing with the aftereffect, the SGDCR problem can be transformed into a mixed integer linear programming problem [18]. But the key point of the two methods is the same, namely, the propagation of streamflow is steady. The LT model is acceptable when dealing with the problem with the small peak load regulation amplitude of the UR, the small streamflow in the river channel or the large re-regulation capacity of the DR. However, when the UR often undertakes heavy peak load regulation tasks and the streamflow in the river channel fluctuates greatly in a short period [13], [15], the LT model cannot reflect the dynamic and unstable propagation characteristics of the OUR. Especially when the DR has a weak re-regulation capacity, the power generation scheme developed by the LT model often deviates significantly from the actual operation process [1]. Even more, the reservoir may suffer the risk of exceeding the water level limits, which will threaten the safety of itself and downstream areas seriously [1]. In this extreme system, the hydropower stations have to negotiate with the power grid frequently and conduct new instructions.

In order to further improve the characterization accuracy of the hydraulic connection in SGDCR, some scholars try to introduce classical streamflow routing methods in hydrology (SRMH) to deal with this problem [17]. Changming Ji and Chuangang Li selected the Muskingum model, which is most widely used in SRMH, and combined it with multi-dimensional DP algorithm to solve the SGDCR of Jindong and Guandi hydropower stations successfully, which are located in Sichuan province in southwestern China [17]. The Muskingum model holds that the streamflow propagation between cascade reservoirs can be generalized as not only transposition but also attenuation [17] and assumes that there is a one-to-one relationship between the river channel storage and the streamflow at a random section [19]–[21]. As a simplified model of HNS, the Muskingum model remains efficient in characterizing the hydraulic connection as the LT model. However, whether it still works in the extreme system with sharp change of OUR and weak re-regulation capacity of DR needs to been further verified.

Besides, in pure and practical hydrological application, soft computing techniques under an artificial intelligence (AI) approach have been developing vigorously. For example, Yaseen, Z.M. utilized the enhanced extreme learning machine (EELM), extreme learning machine (ELM) and support vector regression (SVR) model to predict the daily-scale river flow of Guillemard station in the northeast region of the Malaysian Peninsula [22]. Cheng, C.T. forecasted the monthly discharges in Manwan hydropower by adaptive-network-based Fuzzy Inference System (ANFIS) and artificial neural network (ANN) successfully [23]. Similarly, Alexander Y. Sun utilized the GPR model to predict the monthly runoff [24]. Chau, K.W. adopted meta-heuristic techniques to depict Rainfall-Runoff modeling. But none of the above soft computing techniques have been specifically introduced into SGDCR to characterize the hydraulic connection [25]. Moreover, their studies only refer to the daily-scale streamflow.
In mathematics, the hydraulic connection between cascade reservoirs can be abstracted as a MR. Therefore, from the perspective of data mining, we can find out some effective information contained in historical data and develop the characterizing model about the hydraulic connection. Compared with common parametric models such as ANN, SVN, ELM and so on, the Bayesian regression methods are particularly prominent in seamless integration of model train, hyper parameters estimation and uncertainty estimation [26], [27]. Therefore, the Bayesian regression methods have gained sufficient attention in hydrological runoff prediction [24], [28], time series analysis [29] and so on. Inspired by previous studies, the GPR which is a representative for Bayesian regression methods is firstly introduced to characterize the hour-scale hydraulic connection in SGDCR.

The main outcomes of this study are as follows:
1) The GPR model is for the first time introduced to characterize the hour-scale hydraulic connection of cascade reservoirs for short-term generation dispatching.
2) A set of pivotal indicators are proposed to evaluate the effect of multi-period simulation dispatching.
3) The GPR model is adopted to characterize the hour-scale hydraulic connection, and shows its highest performance in single-period inflow simulation and multi-period simulation dispatching compared with the LT and Muskingum model.

The structure of the remainder of this paper is as follows. In Section II, the LT model, the Muskingum model, the GPR model, the simulation dispatching and the performance indicators are described. In Section III, the GPR which is a representative for Bayesian regression methods is briefly introduced to characterize the hydraulic connection. Finally, Section IV summarizes the relevant conclusions and points out future directions of this paper.

II. METHODOLOGIES
This section introduces two common methods to describe the hydraulic connection in SGDCR, which are the LT model and the Muskingum model respectively. And then, the GPR model is introduced in detail and simulation dispatching details are represented by a flow chart. Finally, the performance indicators utilized in single-period and multi-period simulations are explained.

A. LAG TIME (LT) MODEL
In the long-term joint dispatching of cascade reservoirs, the streamflow lag time between cascade reservoirs is often much smaller than the time span of the scheduling scheme. For this reason, the streamflow propagation time is often neglected [30]–[32]. But in short-term or real-time generation dispatching, the lag time and the time span are close [18]. In the LT model, the OUR is thought to propagate to the downstream reservoir after a certain period. In general, this lag time can be considered as a constant. Therefore, in this paper, we name this treatment method about the hydraulic connection as the lag time (LT) model, which can be characterized as follows:
\[ I_{i,t} = R_{i-1,t-\tau} + Q^0_{i,t}, \quad t = 1, 2, \ldots, T; \quad i = 2, 3, \ldots, N \] (1)
where \( I_{i,t} \) is the inflow at the \( t \)th period of the \( i \)th reservoir, \( R_{i-1,t-\tau} \) is the outflow at the \( t - \tau \)th period of the \( i - 1 \)th reservoir; \( Q^0_{i,t} \) is the local inflow at the \( i \)th period between the \( i - 1 \)th reservoir and the \( i \)th reservoir; \( T \) is the number of periods; \( N \) is the number of reservoirs. \( \tau \) is the LT, which is an integer multiple of one period. It can be seen from formula (1) that the LT method only considers transposition of the streamflow and ignores attenuation effects. In other words, the outflow of the \( i - 1 \)th reservoir at the \( t - \tau \)th period will reach the \( i \)th reservoir at the \( i \)th period absolutely.

B. MUSKINGUM MODEL
Many scholars have confirmed that the flow propagation in the natural channel not only has the feature of transposition but also attenuation [17]. In other words, the OUR belongs to unsteady flow, which can be describe the Saint-Venant equations. Due to the low solving efficiency and the high requirements for the topography and roughness data, many hydrologic streamflow routing methods, such as the Muskingum model which is the most extensively studied and utilized, have been put forward for the streamflow routing in the river channel [17], [21]. Muskingum was proposed by MC.Carthy in 1938 and named after its use in the Muskingum river in the United States [21].

The streamflow routing algorithm is to calculate the outflow process of the upper section into the inflow process of the lower section based on the water balance equation and the channel storage equation, which are represented respectively as follows:
\[ I - O = \frac{dS}{dt} \] (2)
\[ S = f(I, O) \] (3)
where \( I \) means the inflow; \( O \) means the outflow; \( S \) means the channel storage; \( t \) means the period. The Muskingum model holds that there is a one-to-one relationship between the river channel storage and the streamflow at a random section. So, the formula (3) can be rewritten further as follows:
\[ S = K[x + (1 - x)O] = KQ \] (4)
where \( K \) means the storage parameter; \( x \) means the weighting parameter; \( Q \) is the function of \( O, I, x \). The numerical solution of formula (2) and formula (4) results in the well-known Muskingum routing equation as:
\[ O_t = C_0I_t + C_1I_{t-1} + C_2O_{t-1} \] (5)
where \( C_0, C_1, C_2 \) are functions of \( K \) and \( x \), which are represented respectively as follows:
\[ C_0 = \frac{0.5\Delta t - Kx}{0.5\Delta t + K - Kx} \] (6)
\[ C_1 = \frac{0.5\Delta t + K - Kx}{0.5\Delta t + K - Kx} \] (7)
where $\Delta t$ is the duration of one period. After the $\Delta t$ is selected, certain parameters $K$ and $x$ can be derived to $C_0$, $C_1$ and $C_2$. Therefore, when the sequence $Q(t)$ of upstream reservoir and the initial inflow of the downstream reservoir are already known, the sequence $I(t)$ of downstream reservoir can be calculated by formula (5). The Muskingum model takes propagation and attenuation into account simultaneously and has no need to establish calculation grids, therefore it can combine accuracy with timeliness. The outflow of the $i$ - 1th reservoir at the $t - \tau$ th period is not considered to reach the $i$th reservoir at the $t$ th period absolutely, but is distributed in several periods after the $t - \tau$ th period. This is the significant difference between the Muskingum model and the LT model.

The key to the Muskingum model is to calibrate the values of $K$ and $x$ [21]. High-quality parameter calibration results will ensure a linear relationship between $S$ and $Q'$ in formula (4), and the physical meaning of $K$ is the propagation time of the stable streamflow. Trial and error methods, least squares estimation methods and intelligent algorithms are often utilized to estimate parameters of the Muskingum model [17].

### C. GAUSSIAN PROCESS REGRESSION (GPR) MODEL

The hydraulic connection in SGDCR can be regarded as a regression problem in mathematics. The regression model can be characterized as follows:

\[ y = f(x) + \varepsilon \]  

(10)

where $y$ is the independent variable vector, $y$ means the dependent variable, $f(x)$ means the mapping function, $\varepsilon$ means the observed noise. In supervised learning, two kinds of methods are usually used to determine the mapping relationship. The first method is the parametric regression, which holds that $y$ is determined by the parameter $\omega$ and $x$ simultaneously:

\[ y = f(x, \omega) \]  

(11)

However, the parametric regression is easy to overfit and its generalization ability is poor. The second method is Bayesian regression. The GPR is a typical Bayesian regression method, which is suitable for high-dimensional, small-sample and nonlinear regression problems [26]. Because of the adaptive optimization of the hyperparameters in the process of model training, the GPR has been widely used [24], [26], [27], [33], [34]. Next, we will introduce principles and conclusions of GPR from the perspective of function space. Gaussian process (GP) refers to a set of random variables, any finite number of which will obey a common joint gaussian distribution. Its properties are completely determined by the mean function and the covariance function [26]. In GPR, the $f(x)$ obeys the GP, namely:

\[ f(x) \sim GP[m(x), k(x, x')] \]  

(12)

where $x$ and $x'$ mean random vectors in $R_d$; $m(x)$ means the mean function; $k(x, x')$ means the covariance function. The essence of the GPR model is to determine the conditional distribution of the target output for the given input vector. Now, it is assumed that observation noise $\varepsilon$ exists in historical data and obey the gaussian distribution, namely, $\varepsilon \sim N(0, \sigma^2)$ [26], [27]. According to the Bayesian principle, For the given training data set $D = \{(x_i, y_i)\}_{i=1}^{N}$, the dependent variable $y$ is also a GP, namely:

\[ y \sim GP[m(X), K(X, X) + \sigma^2 I] \]  

(13)

where

\[ y = [y_1, y_2, \cdots, y_N]^T \]

\[ X = [x_1, x_2, \cdots, x_N]^T \]

\[ m(X) = [m(x_1), m(x_2), \cdots, m(x_N)]^T \]

and $K(X, X)$ describes the correlation between $X$ and $X$, and is a symmetric positive definite covariance matrix of $N \times N$. Now, for the test data $(x, y_a)$, the prior joint distribution can be expressed as follows:

\[
\begin{bmatrix}
  y \\
  y_a \\
\end{bmatrix} \sim GP
\left[
\begin{bmatrix}
  m(X) \\
  m(x_a) \\
\end{bmatrix},
\begin{bmatrix}
  K(X, X) + \sigma^2 I & K(X, x_a) \\
  K(x_a, X) & k(x_a, x_a) \\
\end{bmatrix}
\right]
\]  

(14)

where $K(X, x_a) = (K(x_a, X))^T$ is the covariance matrix between the training data and test data of $N \times 1$; $K(x_a, x_a)$ is the covariance of a test input vector $x_a$. According to the properties of the conditional distribution of GP, we get the prior distribution, namely:

\[ y_a | X, y, x_a \sim N(\mu_a, \Sigma_a) \]  

(15)

\[ \mu_a = K(x_a, X) \left[ K(X, X) + \sigma^2 I \right]^{-1} \times (y - m(X)) + m(x_a) \]  

(16)

\[ \Sigma_a = k(x_a, x_a) - K(x_a, X) \left[ K(X, X) + \sigma^2 I \right]^{-1} \times K(X, x_a) \]  

(17)

where $\mu_a$ means the mean function of $y_a$, $\Sigma_a$ means the covariance function of $y_a$. The covariance function quantifies the similarity between samples, which is very important for the performance of the GPR model.

### D. SIMULATION DISPATCHING

Overall, to intuitively evaluate the impact of different characterizing models on SGDCR in multiple periods, there are three modes to be selected: 1) Fixing the upstream water level process of Gezhouba as the actual one, we simulate the power generation deviation of Gezhouba. 2) Fixing the outflow process of Gezhouba as the actual one, we simulate the upstream level deviation of Gezhouba. 3) Fixing the power generation process of Gezhouba, we simulate the upstream level deviation of Gezhouba. In generation dispatching, outflow can be converted into power generation. Hence, this third mode is removed. The first and second mode are called the fixing water level scheme (FWLS) and the fixing outflow
scheme (FOS) for short respectively. (The reason why these two modes are necessary will be displayed in Section III-E after the independent variables of the GPR model have been determined.) The flowchart of the simulation dispatching in multiple periods is shown in FIGURE 1.

It should be mentioned that the deviation of outflow or upstream water level is calculated by the water balance equation in formula (18) and (19). Considering the short-term generation dispatching accuracy of Gezhouba has been a tough problem so far, we utilize the deviation of storage (multiplying outflow by periods) divided by the actual rate of power generation of Gezhouba.

FIGURE 1. The flowchart of simulation dispatching in multiple periods with different characterizing models.

where $V_{gcb}(t-1)$ means the initial storage of Gezhouba at the $r$th period. $Z_{gcb}$ means the upstream water level of Gezhouba. $f$ is the corresponding relationship between upstream water level and storage. $\Delta Q(t)$ means the difference between the simulated outflow and the actual outflow at the $r$th period. $R_{actual}$ means the actual rate of mean water consumption on the calculating day. $AG$ means the total deviation of the power generation of Gezhouba.

E. PERFORMANCE EVALUATION

At present, scholars generally evaluate the regression performance of prediction models by introducing a series of statistical indicators [17], [32], [34]–[38]. In this paper, four categories indicators are adopted to measure the performance of different characterizing methods for the single-period inflow simulation, including the mean absolute deviation (MAD) [17], [34], the threshold statistic ($TS_a$) which is related to the relative MAD [36], the Nash–Sutcliffe model efficiency coefficient ($E$) [35], [37] and the determination coefficient ($R^2$) [22], [37]. The definitions of performance indicators are listed below:

$$MAD = \frac{1}{T} \sum_{t=1}^{T} |I_t^o - I_t^s|$$

$$TS_a = \frac{n_a}{T} \times 100$$

$$E = 1 - \frac{\sum_{t=1}^{T} (I_t^o - I_t^s)^2}{\sum_{t=1}^{T} (I_t^o - T\bar{I})^2}$$

$$R^2 = \frac{\sum_{t=1}^{T} (I_t^o - T\bar{I})(I_t^s - T\bar{I})}{\left(\sum_{t=1}^{T} (I_t^o - T\bar{I})^2 \sum_{t=1}^{T} (I_t^s - T\bar{I})^2\right)^{1/2}}$$

where $I_t^o$ means the observed inflow of the DR at the $r$th period. $I_t^s$ means the simulated or forecasted inflow of the DR at the $r$th period. $T$ is the number of period; $n_a$ is the number of data simulated having relative deviation (the absolute difference between $I_t^o$ and $I_t^s$ divide by $I_t^o$) in simulating less than $a%$. In this paper, $TS_a$ is listed for eight levels of 1%, 2%, 3%, 4%, 5%, 10%, 20% and 30%. $T\bar{I}$ is the mean value of the observed inflow. $T\bar{I}$ is the mean value of the simulated inflow. MAD measures the deviation between $I_t^o$ and $I_t^s$. The smaller the MAD is, the higher accuracy the characterizing model owns. $TS_a$ which is related to the relative deviation evaluates the distribution characteristics of simulation deviation. Under the same relative simulation deviation $a$, the larger the $n_a$ is, the larger the $TS_a$ is, which indicating that the simulation accuracy from the specific characterizing model is better [36]. The Nash–Sutcliffe model efficiency coefficient is used to assess the predictive power of characterizing models [35]. An efficiency of one corresponds to a perfect match of the simulated data to the observed data [35]. As for the determination coefficient, it measures the degree of correlation among the observed and simulated values [35]. Legates indicated that a model can be sufficiently evaluated by these indicators. Because these indicators are sensitive to additive
and proportional differences between model simulations and observations and not oversensitive to extreme values [37].

Meanwhile, we introduce a set of pivotal indicators to judge the effect of different characterizing models when conducting the multi-period simulation dispatching. Inflow, outflow, upstream water level and power generation are the most concerned parameters in SGDCR, among which inflow is an intermediate indicator in formula (18) and outflow can be converted into power generation by formula (20). In FWLS, the upstream water level of DR is within the feasible range. Hence, the safety of DR is basically guaranteed, and the power generation becomes the key parameter. However, in FQS, the upstream water level of DR is uncertain, so the safe operation of reservoirs takes precedence over the power generation indicator. Therefore, the indicators of simulation dispatching are defined in TABLE 1.

### TABLE 1. The indicators of parameters of simulation dispatching.

| Mode | Definition | Abbreviation |
|------|------------|--------------|
| FWLS | max positive deviation in a single period | MPDSP |
| | min negative deviation in a single period | MNDSP |
| | max positive cumulative deviation per day | MPCDPD |
| | min negative cumulative deviation per day | MNCDPD |
| | deviation at the end of each day | DED |
| | mean absolute deviation per day | MADPD |
| FQS  | max positive deviation | MPD |
| | min negative deviation | MND |
| | deviation at the end of each day | DED |
| | mean absolute deviation per day | MADPD |

Note: max and min are abbreviations for their respective adjectives. In FWLS, all indicators are about power generation. In FQS, all indicators are about upstream water level.

In FWLS, the indicators in TABLE 1 can be divided into three categories: 1) power generation extreme deviation indicators in a single period including MPDSP and MNDSP; 2) daily cumulative power generation extreme deviation indicators including MPCDPD and MNCDPD; 3) daily average power generation accuracy indicator, including MADPD. In addition, the indicator DED close to zero does not indicate that the performance of the characterizing model is good, and it is necessary to combine the above indicators of three categories to judge the pros and cons of the simulation effect.

In FQS, resulting from the water balance calculations by formula (18) and (19) for continuous periods, the extreme deviation of upstream water level in a single period is consistent with the accumulated extreme deviation in a day. Therefore, this simulation mode has only four indicators. Similarly, the MPD and MND in TABLE 1 are the extreme values of the upstream level deviation, which limit the swing range of deviation. The MADPD is the daily average simulation accuracy indicator. The more it is closed to zero, the better the simulated upstream level process perform. The DED indicator need to be combined with the one-day simulating process to evaluate the pros and cons of results.

## III. CASE STUDY AND RESULTS

### A. STUDY AREA AND DATASET

The Three Gorges-Gezhouba Cascade Water Conservancy Project is located in the main stream of the Yangtze River, China. The TGR has the largest installed capacity in the world. The Gezhouba is the re-regulation reservoir of the TGR and located in 38 km downstream from the dam site of the TGR [39]. The joint dispatching of TGR and Gezhouba plays an irreplaceable role in flood control, power generation, navigation, ecology and so on in the Yangtze region [8], [40]. The mainly parameters of the TGR and the Gezhouba are listed in TABLE 2. Their locations are demonstrated in FIGURE 2.

### TABLE 2. The main parameters of TGR and Gezhouba.

| Parameter                          | TGR   | Gezhouba |
|------------------------------------|-------|----------|
| Total storage (billon m$^3$)       | 39.3  | 1.58     |
| Regulation ability                 | seasonal | daily |
| Flood control storage (billon m$^3$) | 22.15 | -        |
| Installed capacity (MW)            | 22500 | 2715     |
| Normal water level (m)             | 175   | 66.5     |
| Flood limited water level (m)      | 145   | -        |
| Annual generation (billon kWh)     | 84.7  | 15.7     |
| re-regulating capacity (million m$^3$) | -    | 85       |

Due to the huge installed capacity, the TGR undertakes peak load regulation of power system frequently. According to the historical data, when all units in TGR undertake the rated output, the whole streamflow for power generation is about 31000m$^3$/s. The peak load regulation amplitude of the TGR can reach 7000MW. Therefore, it is very common for the outflow variation amplitude of TGR to reach 5000m$^3$/s or even more within 2 hours. (Since the streamflow data of the TGGCR is done once every 2 hours in the actual dispatching, the period length is determined as 2 hours in this paper [1]). There is no large tributary between the TGR and the Gezhouba, and the local inflow is almost negligible. Therefore, the inflow of Gezhouba is mainly determined by the outflow of the TGR. But under the current calculation way, filling or emptying the Gezhouba Reservoir in the course of reservoir application means that the difference between daily inflow and outflow of the Gezhouba reaches or exceeds 1000 m$^3$/s [1]. This level of deviation seriously threatens the safe and economic operation of Gezhouba. The situation is even worse when TGR and Gezhouba conduct the 2-hour or 15-minute joint dispatching. Hence, the TGGCR is a quite typical case of studying the hour-scale hydraulic connection in SGDCR.

13140 sets of 2-hour historical operation data during 2013-2015 are adopted to determine the parameters of the Muskingum model and the hyperparameters of the GPR model. 4392 sets of 2-hour historical operation data in 2016 are used to conduct single-period inflow simulation with different characterizing models. Four typical daily operation processes of TGR throughout 2016 are selected to conduct the multi-period simulation dispatching.

### B. GPR MODEL DEVELOPMENTS

The development of GPR model mainly includes three parts: 1) the selection of the mean function; 2) the selection of...
The kernel function plays an important role in enhancing the GPR performance. Based on the previous publications, the square exponential function is seen as one of the most commonly used kernel functions because it has better generalization ability compared with other kernel functions. Hence, the square exponential function is selected as the kernel function.

As we know, the machine learning methods such as ANN, SVN, GPR and so on generally require data preprocessing before training. A large number of studies tell us that 0-mean function can simplify the derivation of the GPR model and the calibration of hyperparameters. Hence, the 0-mean function is selected to develop the GPR model, which can be rational by the Z-Score normalization of data.

The kernel function plays an important role in enhancing the GPR performance. Based on the previous publications, the square exponential function is seen as one of the most commonly used kernel functions because it has better generalization ability compared with other kernel functions. Hence, the square exponential function in formula (25) is selected as the kernel function.

\[ k(x_i, x_j) = \sigma_f^2 \exp\left(-\frac{(x_i - x_j)^T(x_i - x_j)}{2l^2}\right) \]  

where \( \sigma_f \) means the kernel function signal variance, which is used to control the local correlation of input variables; \( l \) means the scale of the equation. To sum up, there are three hyperparameters, including \( \sigma \), \( \sigma_f \) and \( l \). According to the existing training data set \( D \), the maximum likelihood estimation is used to determine the hyperparameters when training the GPR model.

As for the third part, the dependent variable of three characterizing models is particularly explicit, namely \( I_{\text{gzb}}(t) \). (the inflow of Gezhouba at the \( t \)th period) Therefore, it is necessary to identify the independent variables of the GPR model and determine its dimensions. The hydraulic connection between TGR and Gezhouba has multivariable and nonlinear characteristics. At the beginning, we can preliminarily determine the dimension of the independent variables through analysis.

As mentioned in Section III-A, the inflow of Gezhouba depends greatly on the outflow of TGR, and there is a 30-minute lag time between TGR and Gezhouba. Therefore, when we depict a 2-hour mapping function, \( Q_{\text{sx}}(t-1) \) and \( Q_{\text{sx}}(t) \) (\( \text{sx} \) means TGR) should be included in the independent variables. From the perspective of hydrodynamic simulation, the water level process of upper and lower sections can affect the propagation of streamflow in the river channel, so \( Z_{\text{sx}}(t-1), Z_{\text{sx}}(t), Z_{\text{gzb}}(t-1) \) and \( Z_{\text{gzb}}(t) \) (\( Z_{\text{sx}} \) means the tail water level of TGR while \( Z_{\text{gzb}} \) means the upstream water level of Gezhouba) should be included in the independent variables. Besides, the formula (5) in the Muskingum model tells us the inflow of the downstream section in the current period is related to the inflow of the previous one period. Therefore, the independent variables should contain \( I_{\text{gzb}}(t-1) \). To sum up, the hydraulic connection function can be preliminarily expressed as follows:

\[ I_{\text{gzb}}(t) = f(Q_{\text{sx}}(t-1), Q_{\text{sx}}(t), Z_{\text{sx}}(t-1), Z_{\text{sx}}(t), Z_{\text{gzb}}(t-1), Z_{\text{gzb}}(t), I_{\text{gzb}}(t-1)) \]  

In formula (26), the dimension of the independent variables is seven. Trial and error methods are adopted to eliminate redundant independent variables. 13140 sets of 2-hour historical data during 2013-2015 are adopted used for the GPR model training, and 4392 sets of 2-hour historical data in 2016 are used for simulation evaluation, testing the simulation results with the GPR model under different combinations of independent variables.

In order to display the simulation results clearly, the independent variables in formula (26) are abbreviated as \( [x_1, x_2, x_3, x_4, x_5, x_6, x_7] \). The four categories performance indicators introduced in Section II-E are listed in TABLE 3.

Firstly, we compare column 2 with column 7 in TABLE 3. The scheme excluding \( I_{\text{gzb}}(t-1) \) does not show obvious disadvantages in terms of four categories indicators compared with the scheme with complete independent variables.
variables. In the actual dispatching, there is a strong demand for simulating the inflow of Gezhouba for several consecutive periods. For example, dispatchers often forecast the upstream water level process of Gezhouba in one day. If we include the \( I_{gzb}(t - 1) \) with simulation deviation into the independent variables, it is bound to lead to the accumulated deviation in simulating the subsequent inflow process. Meanwhile, simulating \( I_{gzb}(t) \) by \( I_{gzb}(t - 1) \) is only a simplified method in hydrology. Therefore, the \( I_{gzb}(t - 1) \) can be regarded as a redundant variable. Then, we focus on column 6 and column 7 in TABLE 3. The performance of the two schemes is similar. In the actual dispatching, the upstream water level of Gezhouba is not only affected by inflow, but also affected by the operation mode (such as fixing the outflow or conducting propagation time between TGR and Gezhouba, the flood flattening effect is not obvious. Considering there is no local inflow, the change of the flood peak is very small. Therefore, this paper chooses the parameter setting results with high \( E \) as the priority. The values of the three parameters in formula (5) are \( c_0=0.522, \ c_1=0.268 \) and \( c_0=0.210 \) respectively. 4392 sets of test data (the 2-hour historical outflow of TGR) in 2016 are used to verify the accuracy of the Muskingum model. The simulation results are shown in TABLE 4 and the distribution of simulation deviation is shown in FIGURE 3.

![FIGURE 3. The distribution characteristics of deviation obtained by different methods in 2016.](image)

Then, 13140 sets of 2-hour historical data (the outflow of TGR and the inflow of Gezhouba) during 2013–2015 are used to calibrate the Muskingum model parameters with MOSCDE algorithm, whose aims are to maximize \( E \) and minimize the relative error of flood peak [41]. Due to the short propagation time between TGR and Gezhouba, the flood flattening effect is not obvious. Considering there is no local inflow, the change of the flood peak is very small. Therefore, this paper chooses the parameter setting results with high \( E \) as the priority. The values of the three parameters in formula (5) are \( c_0=0.522, \ c_1=0.268 \) and \( c_0=0.210 \) respectively. 4392 sets of test data (the 2-hour historical outflow of TGR) in 2016 are used to verify the accuracy of the Muskingum model. The simulation results are shown in TABLE 4 and the distribution of simulation deviation is shown in FIGURE 3.

Finally, after determining the independent variables of the GPR model in Section III-B, we start to train and test the GPR model. The independent variables of the GPR model include 5 variables, namely,

\[
x_t = [Q_{sx}(t - 1), \ Q_{sx}(t), \ Z_{sx}(t - 1), \ Z_{sx}(t), \ Z_{gzb}(t - 1)], \quad t = 1, 2, \cdots T \tag{29}
\]

Therefore, the simulated inflow of Gezhouba is rewritten as follows based on formula (10) in Section II-C.

\[
I_{gzb}(t) = E(f_t^{GPR}(x_t)|X, y), \quad t = 1, 2, \cdots T \tag{30}
\]

### TABLE 3. Comparison of prediction performance with different independent variables in GPR.

| Indexes | all | \([x_5, x_6]\) | \([x_5, y]\) | \([x_5, x_7]\) | \([x_5, y, x_7]\) | \([x_5, y]\) |
|---------|-----|----------------|----------------|----------------|----------------|----------------|
| \(MAD(m^3/s)\) | 287 | 319 | 300 | 325 | 291 | 291 |
| \(E\) | 0.9974 | 0.9968 | 0.9971 | 0.9967 | 0.9973 | 0.9973 |
| \(R^2\) | 0.9981 | 0.9979 | 0.9979 | 0.9979 | 0.9981 | 0.9981 |
| \(a=1\) | 1075 | 1075 | 1059 | 978 | 1067 | 1100 |
| \(a=2\) | 2144 | 1936 | 2119 | 1902 | 2160 | 2147 |
| \(a=3\) | 3018 | 2805 | 2968 | 2746 | 3011 | 2992 |
| \(a=4\) | 3580 | 3438 | 3560 | 3411 | 3609 | 3599 |
| \(a=5\) | 3961 | 3885 | 3945 | 3862 | 3966 | 3984 |
| \(a=10\) | 4387 | 4384 | 4383 | 4381 | 4389 | 4389 |
| \(a=20\) | 4391 | 4391 | 4391 | 4391 | 4391 | 4391 |
| \(a=30\) | 4392 | 4392 | 4392 | 4392 | 4392 | 4392 |

Note: \(a\) is defined in formula (22); all means the independent variables is complete, namely seven variables; the symbol \(\lceil\) means exclusion, so \(\lceil x_i \) means excluding the variable \( x_i \) from the independent variables in formula (26).

### TABLE 4. Comparison of simulation performance with different methods in 2016.

| Indexes | LT | MUSKINGUM | GPR |
|---------|----|-----------|-----|
| \(MAD(m^3/s)\) | 488 | 554 | 291 |
| \(E\) | 0.9910 | 0.9885 | 0.9973 |
| \(R^2\) | 0.9971 | 0.9924 | 0.9981 |
| \(a=1\) | 731 | 714 | 1067 |
| \(a=2\) | 1302 | 1455 | 2160 |
| \(a=3\) | 2237 | 2185 | 3011 |
| \(a=4\) | 2917 | 2817 | 3609 |
| \(a=5\) | 3521 | 3312 | 3966 |
| \(a=10\) | 4342 | 3968 | 4389 |
| \(a=20\) | 4391 | 4334 | 4391 |
| \(a=30\) | 4392 | 4390 | 4392 |

Note: All indicators in this table are mentioned in Section II-E.
where \( E() \) means the expectation function; \( X \) and \( y \) are described in Section II-C; \( f_{GPR}(x_t) \) obey following Gaussian process:

\[
f_{GPR}(x_t) \sim GP(m(x_t), k(x_t, x_t')) \quad t = 1, 2, 3, \ldots T \tag{31}
\]

13140 sets of 2-hour historical data from TGR and Gezhouba are adopted as the training data to optimize the hyperparameters of the GPR model with the maximum likelihood estimation method [26], [27]. The three hyperparameters of GPR are \( \sigma = 276.8058 \), \( l = 3.0646 \), \( \sigma_1 = 14776 \), respectively. Then, the GPR model is used to simulate the 2-hour inflow of Gezhouba with 4392 sets of test data in 2016. The simulation results are shown in TABLE 4 and the distribution of simulation deviation is shown in FIGURE 3.

It is worth emphasizing that the tests in this section all belong to a single-period simulation, not continuous simulations in chronological order. Specifically speaking, if we want to simulate the \( I_{c,b}(t) \), \( t = 1, 2, \ldots, 4392 \), we selected the \( I_{c,b}(t) \) of three models are all historical data.

Meanwhile, to determine the relationship between the simulation accuracy of different characterizing models and the streamflow range, we divide the test data in 2016 into three subsets based on the relative deviation and calculate some indicators listed in TABLE 5.

### TABLE 5. The relationship between simulation accuracy and streamflow range of different models in 2016.

| Indexes/RMAD | (0%,5%) | (5%,20%) | (20%,30%) |
|--------------|---------|----------|-----------|
| LT_MAX       | 32124/33461 | 31588/33569 | 4999/6490 |
| LT_MIN       | 4697/4541  | 4739/4475  | 4999/6490 |
| LT_MED       | 9408/9405  | 16372/17582 | 4999/6490 |
| MUS_MAX      | 32061/33461 | 31586/33569 | 15213/17528 |
| MUS_MIN      | 4679/4541  | 4826/4475  | 4999/6490 |
| MUS_MED      | 10344/10499 | 10207/10745 | 9879/8034 |
| GPR_MAX      | 32526/33461 | 31776/33569 | 4929/6490 |
| GPR_MIN      | 4552/4475  | 4574/4485  | 4929/6490 |
| GPR_MED      | 10873/10976 | 7669/8039  | 4929/6490 |

Note: RMAD means the relative deviation of inflow mentioned in Section II-E; MUS means the abbreviation of Muskingum. MAX, MIN and MED are the maximum, minimum and median value of inflow respectively. The simulated value and actual value are separated by \( / \). The unit of all data in this table is \( m^3/s \).

### D. COMPARISON RESULTS WITH DIFFERENT MODELS

Combined TABLE 4 with FIGURE 3, two conclusions can be drawn easily: 1) The \( E \) and \( R^2 \) indicators of three characterizing models are very close to one, which shows that the overall performance of the three models is great; 2) But in terms of absolute (MAD) and relative deviation distribution (TSa), there arise significant differences between three models. The GPR model owns the best performance while the Muskingum model owns the worst performance. Meanwhile, from the perspective of the streamflow range, the third conclusion can be abstracted based on the data in TABLE 5: 3) the GPR model owns best applicability on the whole range of streamflow. For the above three conclusions, these can be explained as follows:

1) According to the analysis in Section II-E, the closer the values of \( E \) and \( R^2 \) are to one, the higher prediction ability and the better matching degree of the observed and predicted data the model owns respectively. We test 4392 sets of 2-hour data in 2016 with the LT model, the Muskingum model and the GPR model respectively. The average inflow of Gezhouba in 2016 is 13300m\(^3/s\), and the largest MAD in TABLE 4 is 554m\(^3/s\), accounting for only 4.16%. Therefore, the deviation of simulated inflow has little influence on \( E \) and \( R^2 \). This phenomenon also indicates that the MAD and TS indicator selected in Section II-E are necessary.

2) The LT method directly takes the outflow of TGR as the inflow of Gezhouba at the same period and holds that the streamflow is propagated in a steady pattern. However, the Muskingum model only considers the influence of streamflow between adjacent periods, but does not consider the effect of reservoir water level. The GPR model has more dimensions of independent variables, and more valuable information can be used to simulate the inflow of Gezhouba. Therefore, GPR model perform best.

Although the Muskingum model uses the streamflow of adjacent periods at upper and lower sections to simulate the inflow of Gezhouba, it works worse than the LT method. The TGR often undertakes the peak load regulation task whose amplitude may reach 7000MW, so the outflow of TGR jumps dramatically in a short time (hours or minutes level), resulting in the unstable propagation of the outflow and aggravating the difficulty in simulating inflow of Gezhouba. The motion of the flood wave in the river channel can be described by the Saint-Venant equations. Based on formula (4), the Muskingum model simplifies the channel storage equation and can quickly calculate the streamflow of the lower section. But this simplification must be based on the one-to-one relationship between the river channel storage and the streamflow. Considering TGR and Gezhouba are only 38 kilometers apart and the propagation lag time is about 30 minutes, the barrier and jacking effects of Gezhouba have significantly changed the propagation pattern of streamflow. Therefore, the river reach from TGR to Gezhouba cannot be regarded as a natural channel in a real sense. At the same time, due to the large-amplitude change in outflow of TGR in a short time and low the re-regulating capacity (85 million \( m^3 \)) of Gezhouba, the upstream water level of Gezhouba will fluctuate rapidly in the range of 63.5m to 66.5m. Therefore, the Muskingum model does not work well in this extreme system. According to the test results in 2016, the performance of the Muskingum model is the lowest. As for the LT model, it only uses outflow of TGR to simulate inflow of Gezhouba, so the performance is not as good as the GPR model. Meanwhile, it doesn’t need the one-to-one relationship in the Muskingum model which are inapplicable in this extreme system, so it works better than Muskingum model.

In terms of the simulation accuracy, compared with the LT model and the Muskingum model, the MAD of the GPR model is reduced by 197m\(^3/s\) and 263m\(^3/s\) respectively. Therefore, the GPR model significantly improves the inflow
accuracy of Gezhouba and reduces the risk of modifying the established dispatching scheme. As for the distribution characteristics of simulation deviation, we can group the simulations into three categories. The simulations with relative deviation less than 5% are grouped into the high-quality simulations set (HQSS). The simulations with relative deviation between 5% and 20% are grouped into the neutral simulations set (NSS) and the simulations with relative deviation greater than 20% are grouped into the low-quality simulations set (LQSS). It can be calculated by the TABLE 4 that the number of HQSS in the GPR model is 445 and 654 larger than that of the LT model and the Muskingum model respectively. There arises only one low-quality simulation in the GPR model and the LT model. While, the Muskingum model produces 58 low-quality simulations. Moreover, for the same relative deviation $a$, the number of $n_a$ (defined in Section II-E) in the GPR model is also significantly larger than the other two models. That is to say that the GPR model transforms many neutral simulations in the LT model and the Muskingum model into high-quality simulations.

3) In TABLE 5, the high and neutral-quality simulations of the LT and the GPR model have covered the whole streamflow range and there is no positive correlation between the relative MAD and streamflow. This phenomenon also indicates that the 3-year training data in 2013-2015 is relatively comprehensive and representative, basically covering the scope of inflow of Gezhouba in flood and non-flood season. However, for a particular relative deviation interval where the value of $a$ defined in formula (22) is large, the larger the inflow span (MAX minus MIN in TABLE 5) is, the worse performance the characterization model owns. Because, generally speaking, the large span inflow indicates that there are more simulation periods located in this relative deviation interval. For example, the number of periods in LQSS of the Muskingum model is 56 while other models only produce one. Hence, in the interval with $a$ from 20 to 30, the inflow span of the Muskingum model is the largest. As for the LT and the GPR model, they perform similarly in existing three intervals in TABLE 5. But if we further focus on the interval with $a$ from 10 to 20, their inflow spans by statistics are 23079m$^3$/s and 1292m$^3$/s respectively. Hence, the applicability of the GPR model is the best.

To sum up, compared to the LT and the Muskingum model, the proposed GPR model improves the accuracy of simulated inflow. Meanwhile, it possesses the best distribution characteristics of simulation deviation and best applicability on the whole range of streamflow.

### E. SIMULATION DISPATCHING WITH DIFFERENT CHARACTERIZING METHODS

In order to evaluate the actual dispatching effect of the three characterizing methods and the ability of multi-period simulation dispatching, we select four typical daily operation processes of TGR throughout 2016, two of which have large regulation amplitude (greater than 3500MW) and the other two have small regulation amplitude (less than 2000MW). Meanwhile, considering the representativeness of scenarios, two (April 17 and June 16) of them are selected in the flood season while the other two (January 31 and November 12) are selected in the non-flood season. Here, we mainly focus on the MR between the outflow of TGR and the inflow of Gezhouba in this paper. In order to evaluate the characterization effect clearly, under the premise that the operation process of the TGR is consistent with the actual one, three different characterizing methods are used to simulate the inflow process of Gezhouba for one day.

It is worth emphasizing that there is a cumulative effect caused by the simulated inflow process for the Muskingum model and the GPR model. For the Muskingum model, the $O_{t-1}$ (the inflow of Gezhouba in $t$ − 1th period) in first period in formula (5) is the initial value, and the remainder of the inflow process is calculated by the Muskingum model. Therefore, the inflow of the later period is affected by the simulated inflow of one previous period. For the GPR model, if the upstream level process of Gezhouba is fixed, then the independent variables in formula (29) are all actual data, so there is no cumulative effect. If fixing the outflow process, the simulating inflow process of Gezhouba will deviate from the actual one. According to the principle of water balance, the upstream level process of Gezhouba is different from the actual one. In other words, the independent variables of the GPR model has changed. In this case, there is also a cumulative effect. Hence, FWLS and FQS are all necessary. The flow chart of simulation dispatching can be seen from the FIGURE 1 in Section II-D.

The results of FWLS are shown in FIGURE 4-7. The abbreviation GPR single in the figures represents the power generation deviation in a single period of the GPR model. The abbreviation GPR cumulation in the figures represents the cumulative power generation deviation of the GPR model as the period goes by. The rest of the abbreviations have similar coding ways. The statistical results of the power generation deviation are shown in TABLE 6.

![FIGURE 4. Comparison of power generation deviation simulated by different characterizing models on June 16 with 4100MW peaking amplitude.](image-url)
indicators of the GPR model in TABLE 6 performs significantly better than the other two models on June 16. From FIGURE 5 and TABLE 6, the GPR model still has a clear advantage on single-period and cumulative power generation deviation. It should be noted that the DED of the Muskingum model is closer to zero. Due to the multi-period simulated inflow, there arises a large swing in power generation deviation in a single period of Muskingum model. It happens that at the end of the day, the accumulated power generation is closer to the actual one. Besides, The LT model is weaker than the GPR model on all indicators. Overall, the GPR model performs best on January 31. Similarly, on April 17, although the MNDSP indicator of the Muskingum model is slightly better than that of the GPR model, the GPR model is still dominant on other performance indicators and the degree of fit to the actual process. The characteristics on November 12 are similar to those on January 31, here we not repeat again. Overall, when fixing the upstream level process of Gezhouba and simulating the deviation of power generation in one day, the GPR model produces the most consistent power generation process compared to the actual one and the smallest deviation swing amplitude of power generation.

The results of FQS are shown in FIGURE 8-11. The statistical results of the upstream water level deviation are shown in TABLE 7. The abbreviation GPR in the following figures represents the upstream water level process of Gezhouba with the characterization of the GPR model. The rest of the abbreviations have similar coding ways.

The simulation results on June 16, April 17 and November 12 are quite similar. Although the GPR has a slightly large positive level deviation (MPD), the upstream level process of the GPR model fluctuates within a small level bandwidth based on MPD and MND in TABLE 7. More importantly, the trends of GPR and real process in FIGURE 8-11 are consistent. As for the result on January 31, the GRP model performs much better than the other two models.
methods in all aspects. Overall, when fixing the outflow process as the actual one, the simulating the upstream level process of Gezhouba with GPR model in one day is consistent with the actual process in trends and sizes. Since the accumulated inflow simulation deviation of the GPR model is not linearly proportional to the period length, the characterization ability of the GPR model is the best when simulating the inflow process for multiple periods. In this simulation mode, the hydraulic connection characterized by the GPR method will greatly reduce the risk of exceeding water level the limit of Gezhouba.

To sum up, the above two kinds of simulation dispatching results indicate that the GPR model possesses the strongest ability when conducting the multi-period simulation dispatching.

Finally, it is worth emphasizing that the TGCRCCC mainly adopts two common methods when executing the hour-scale joint scheduling of the Three Gorges cascade power stations at present: 1) The dispatchers multiply the outflow of TGR by a conversion factor as the inflow of Gezhouba. 2) A sequence of outflow in previous periods is multiplied by conversion factors respectively and added up linearly. The two methods can be expressed by the formulas (30) and (31) respectively.

\[ I_{gzb}(t) = a_t \cdot Q_{sx}(t) \]  

(32)

\[ I_{gzb}(t) = \sum_{n=0}^{N} a_t \cdot Q_{sx}(t-n) \]  

(33)

where \( N \) means the length of the outflow sequence. The conversion factor is fitted with daily average outflow of TGR and the daily average inflow of Gezhouba. When we use this conversion relationship, the outflow of TGR is the independent variable, namely:

\[ a_t = f(Q_{sx}(t)) \]  

(34)

The scatter plot between the daily average outflow of TGR and the conversion factor during 2013-2015 is shown in FIGURE 12.

As you can see in FIGURE 12, when the \( Q_{sx} \) is less than 35000 \( m^3/s \), each outflow will correspond to a wide bandwidth of conversion factor \( a \), so it’s difficult to find a reasonable curve to fit the formula (34). We calculate the linear correlation coefficient between \( Q_{sx} \) and conversion factor \( a \),
which is only 0.65. Therefore, when we utilize formula (32) to simulate the hour-scale inflow of Gezhouba, the conversion relationship fitted by daily-scale data in FIGURE 12 cannot meet the needs of actual production operation of TGGCR. As for N in formula (33), how to determine its value is also tricky and the TGCRCCCC has also been studying for this dilemma. The current solution is mainly based on the experience of dispatchers, which is relatively arbitrary. Therefore, Gezhouba has to adjust its operating schemes frequently. Strictly speaking, these two methods are merely expedient and cannot guarantee the safe, efficient and stable operation of TGGCR.

IV. CONCLUSIONS

In this paper, the GPR model, which is a representative for Bayesian regression methods, is firstly introduced to characterize the hour-scale hydraulic connection in generation dispatching of cascade reservoirs. To verify the performance of the GPR model in a single period, we use the 2-hour historical data of the TGGCR during 2013-2015 as training data, the 2-hour historical data in 2016 as test data, and compare the inflow of Gezhouba simulated by the LT method, the Muskingum model and the GPR model respectively. Meanwhile, four typical daily peak load regulation processes of TGR are selected in 2016 to conduct the multi-period simulations with two simulation dispatching modes under different characterizing models. In addition, a set of pivotal indicators are proposed to evaluate the effect of multi-period simulation dispatching. According to the simulation results, following conclusions can be draw:

1) The GPR model has the most accurate characterization on the hour-scale hydraulic connection between TGR and Gezhouba. The MAD of the GPR model with 2-hour in 2016 decreases by 197m$^3$/s and 263m$^3$/s than the LT and Muskingum model, respectively.

2) The distribution characteristics of simulation deviation of the GPR model is most competitive compared to the LT and the Muskingum model. Meanwhile, the GPR model shows the strongest applicability to the full range of streamflow and the strongest ability for multi-period dispatching.

3) The Muskingum model is not recommended to characterize the hour-scale hydraulic connection in the extreme short-term generation dispatching system, where the UR often undertakes heavy peak load regulation tasks and the DR owns weak re-regulation capacity.

To sum up, the study in this paper shows that the GPR model is a powerful tool for characterizing the hour-scale hydraulic connection of cascade reservoirs, which is usually treated roughly or ignored in most non-extreme short-term generation dispatching system of cascade reservoirs.

Besides, due to the difficulty of data acquisition, we only used 3-year data for model training. But the amount of reservoir operation data will increase with the passage of time which directly affects the development of GPR model, such as the determination of hyperparameters. Although it has been mentioned in Section III-D that the selected data is relatively comprehensive and representative, we will consider the time variability of model development in the following studies if data conditions permit. Meanwhile, some important hydrological parameters of the river channel itself, such as roughness and so on, are important in hydrodynamic and hydrological simulation of streamflow. Hence, we will try to introduce hydrological parameters into the GPR model, which can improve the interpretation of the characterizing model from the aspect of physical mechanism.

REFERENCES

[1] Y. Shang, Y. Xu, L. Shang, Q. Fan, Y. Wang, and Z. Liu, “A method of direct, real-time forecasting of downstream water levels via hydropower station reregulation: A case study from gezhouba hydropower plant, China,” J. Hydrol., vol. 573, pp. 895–907, Jun. 2019.

[2] S. Lu, Y. Shang, W. Li, Y. Peng, and X. Wu, “Economic benefit analysis of joint operation of cascaded reservoirs,” J. Cleaner Prod., vol. 179, pp. 731–737, Apr. 2018.

[3] S. Guo, J. Chen, Y. Li, P. Liu, and T. Li, “Joint operation of the multi-reservoir system of the three gorges and the qingjiang cascade reservoirs,” Energies, vol. 4, no. 7, pp. 1036–1050, Jul. 2011.

[4] Z. He, J. Zhou, M. Xie, B. Jia, Z. Bao, H. Qin, and H. Zhang, “Study on guaranteed output constraints in the long-term joint optimal scheduling for the hydropower station group,” Energy, vol. 185, pp. 1210–1224, Oct. 2019.

[5] C. Ji, H. Yu, J. Wu, X. Yan, and R. Li, “Research on cascade reservoirs’ short-term optimal operation under the effect of reverse regulation,” Water, vol. 10, no. 6, p. 808, Jun. 2018.

[6] K. K. Mandal and N. Chakraborty, “Short-term combined economic emission scheduling of hydrothermal power systems with cascaded reservoirs using differential evolution,” Energy Convers. Manage., vol. 50, no. 1, pp. 97–104, Jan. 2009.

[7] L. Mo, P. Lu, C. Wang, and J. Zhou, “Short-term hydro generation scheduling of three gorges–Gezhouba cascaded hydropower plants using hybrid MACS-ADE approach,” Energy Convers. Manage., vol. 76, pp. 260–273, Dec. 2013.

[8] C. Ma, J. Lian, and J. Wang, “Short-term optimal operation of three-gorge and Gezhouba cascade hydropower stations in non-flood season with operation rules from data mining,” Energy Convers. Manage., vol. 65, pp. 616–627, Jan. 2013.

[9] D. Carpentier, J. Haas, M. Olivares, and A. de la Fuente, “Modeling the multi-seasonal link between the hydrodynamics of a reservoir and its hydropower plant operation,” Water, vol. 9, no. 6, p. 367, May 2017.

[10] G. Ibarra, A. de la Fuente, and M. Contreras, “Effects of hydropoeaking on the hydrodynamics of a stratified reservoir: The Rapel reservoir case study,” J. Hydraulic Res., vol. 53, no. 6, pp. 760–772, Nov. 2015.

[11] G. Savant, C. Berger, T. O. McAlpin, and J. N. Tate, “Efficient implicit finite-element hydrodynamic model for dam and levee breach,” J. Hydraulic Eng., vol. 137, no. 9, pp. 1005–1018, Sep. 2011.

[12] A. R. Shaw, H. S. Sawyer, E. J. LeBouef, M. P. McDonald, and B. Hadjerioua, “Hydropower optimization using artificial neural network surrogate models of a high-fidelity hydrodynamics and water quality model,” Water Resour. Res., vol. 53, no. 11, pp. 9444–9461, 2017.

[13] Y. Shang, X. Li, X. Gao, Y. Guo, Y. Ye, and L. Shang, “Influence of daily regulation of a reservoir on downstream navigation,” J. Hydrol. Eng., vol. 22, no. 8, pp. 1–8, 2017.

[14] J. Zhang, S. Lin, and W. Qiu, “A modified chaotic differential evolution algorithm for short-term optimal hydrothermal scheduling,” Int. J. Electr. Power Energy Syst., vol. 65, pp. 159–168, Feb. 2015.

[15] M. Xie, J. Zhou, C. Li, and P. Lu, “Daily generation scheduling of cascade hydrop power plants considering peak shaving constraints,” J. Water Resour. Plan. Manag., vol. 142, no. 4, 2016, Art. no. 04015072.

[16] X. Fu, A. Li, L. Wang, and C. Ji, “Short-term scheduling of cascade reservoirs using an immune algorithm-based particle swarm optimization,” Comput. Math. Appl., vol. 62, no. 6, pp. 2463–2471, Sep. 2011.

[17] C. Ji, C. Li, B. Wang, M. Liu, and L. Wang, “Multi-stage dynamic programming method for short-term cascade reservoirs optimal operation with flow attenuation,” Water Resour. Manage., vol. 31, no. 14, pp. 4571–4586, Nov. 2017.
X. Chen et al.: Characterizing the Hydraulic Connection of Cascade Reservoirs for Short-Term Generation Dispatching

IEEE Access

XIAO CHEN was born in Enshi, China, in 1993. He received the B.S. degree from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2017, where he is currently pursuing the Ph.D. degree in hydraulic and hydropower engineering.

His research interests include modeling and operation theory in water resources management.

JIANZHONG ZHOU was born in Wuhan, China, in December 1959. He received the B.S. degree in automatic control from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1982. He is currently a Professor at the School of Hydropower and Information Engineering, HUST. His research interests include optimal operation and control of hydropower energy systems.

BENJUN JIA was born in Zunyi, China, in July 1994. He received the B.S. degree from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2017, where he is currently pursuing the Ph.D. degree in hydraulic and hydropower engineering.

His research interests include modeling and operation theory in water resources management.

XIN YANG was born in Shanxi, China, in 1995. He received the B.S. degree from Northwest A&F University (NWSUAF), Yangling, China, in 2016. He is currently pursuing the Ph.D. degree with the School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan, China.

His research interests include modeling and operation theory in water resources management.

CHAO ZHOU was born in Wuhan, China, in June 1986. He received the Ph.D. degree from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2015. He is currently a Researcher at the Changjiang Institute of Survey, Planning, Design and Research. His research interests include reservoir (group) optimal dispatch, hydropower dispatching decision support systems, and hydro-informatization construction.