Evaluation of a multisite weather generator on precipitation simulation in the Yangtze river basin

Yu Li, Xin-Min Zeng* and Jiali Guo

CMA-HHU Joint Laboratory for Hydrometeorological Studies, Hohai University, Nanjing 210098, China
College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China
College of Hydraulic & Environmental Engineering, China Three Gorges University, Yichang 443002, China
Hubel Key Laboratory of Intelligent Yangtze and Hydroelectric Science, China Yangtze Power Co., Ltd, Yi-chang Hubei, 443000, China
*Corresponding author. E-mail: zen_xm@yahoo.com

ABSTRACT

This paper presents the evaluation of a multisite statistical weather generator (MulGETS: Multisite weather Generator of École de Technologie Supérieure) based on its simulation effect of precipitation in the Yangtze River Basin. MulGETS effectively generates spatially correlated sequences of precipitation simultaneously, while maintaining their spatial and temporal distribution characteristics. On the spatial scales, the accuracy of the model varies from station to station, and in general, the errors are lower at stations in the middle and lower reaches of the Yangtze River Basin than in the upper reaches. This difference is likely to exist because of the lower amount of rainfall and more complex topography than those of the upper river basins. On the temporal scales, the simulated values are more precise on the annual scale than on the seasonal scale. Large relative errors occur more frequently in winter, ranging from −35% to 25%. MulGETS can consistently produce precipitation by considering the intensity, magnitude, and duration indices with sub-basin varied observations. However, the precipitation maxima were much lower than the observations. This work shows the general reasonability of the model in downscaling precipitation in the Yangtze River Basin.

Key words: daily precipitation, MulGETS, spatial distribution, stochastic hydrology, the Yangtze River Basin, weather generator

HIGHLIGHTS

• MulGETS is applied to generate daily precipitation for each subbasin in the Yangtze River Basin.
• The performance is evaluated by 9 indices in three categories over seasonal and annual scales.

INTRODUCTION

The general circulation model (GCM) is currently an important tool for simulating the basin’s climate and predicting future climate scenarios. However, the spatial resolution of the model’s output data is extremely coarse for application to meteorological and hydrological research at the basin scale. To solve this problem, scientists have extensively researched downscaling methods to convert rough GCM data into high-resolution regional or watershed scale data (Xu 1999; Wilby et al. 2002; Wood et al. 2004; Mehrotra et al. 2014; Pour et al. 2014; Singh et al. 2017; Araya-Osses et al. 2020; Li et al. 2020).

In general, downscaling methods can be divided into two categories: dynamical downscaling (Giorgi & Mearns 1999; Zeng et al. 2017; Liu et al. 2018; Shiferaw et al. 2018) and statistical downscaling (Chen et al. 2016; Yi et al. 2018) methods. Statistical downscaling is widely used because of its flexibility, simplicity, low computational complexity, and the ability to quickly simulate long-term regional climate sequence (Cheng et al. 2008; Chen et al. 2011; Chen et al. 2012; Wang et al. 2014; Verdin et al. 2015). Owing to the importance of fully understanding the secular variation of hydrological characteristics at the watershed scale, many scholars have applied different statistical downscaling methods to determine hydrological responses to climate change (Ahmed et al. 2013; Jeong et al. 2013; Wang et al. 2015; Lee & Bae 2016; Zheng et al. 2018). The climate change information needs to match the hydrological information on the basin scale; in other words, weather data corresponding to the existing stations should be used (Brisette et al. 2007).
Weather generators are an important statistical downscaling method, and they are widely used to generate synthetic sequences of weather variables that are statistically consistent with observed variables. Conventional weather generators are based on individual stations (or grid points) to generate single-point data or establish mutually independent regression equations (Brissette et al. 2007), however, hydrological simulation results are largely affected by the spatial distribution of regional precipitation. The above-mentioned single-site downscaling methods may neglect the correlation between adjacent points of the forecast elements. This deficiency partly restricts the application of the single-site statistical downscaling method (Chen et al. 2014). Therefore, in recent years, many studies have focused on spatially correlated downscaling methods (Kottegoda et al. 2003; Raje & Mujumdar 2009; Srikanthan & Pegram 2009; Breinl et al. 2013; Jeong et al. 2013; Chowdhury et al. 2019) or provided useful insights on the development of multisite downscaling methods (Wilks 1998; Harpham & Wilby 2005; Fang & Tacher 2010). Fu evaluated four multisite weather generators to simulate daily precipitation in the Gloucester Watershed in Australia. The results show that the four weather generators can satisfactorily simulate the extreme values of precipitation as well as the duration of wet and dry days (Fu et al. 2018).

Chen proposed a MATLAB-based stochastic weather generator referred to as Multisite weather Generator of École de Technologie Supérieure (MuLGETS) and tested it by simulating precipitation and temperature at stations in the Jinghe Watershed in China, Peribonka Watershed in Canada, and Walnut Gulch Watershed in America (Chen et al. 2014). At these stations, MuLGETS was able to retain the spatial correlation of precipitation and temperature between stations. Many researchers have applied MuLGETS to areas of interest and compared it with other weather generators. This work has shown that MuLGETS often outperformed the other three weather generators when applied to a watershed of Eastern Ontario, Canada (Alodah & Seidou 2019). Tseng compared MuLGETS with three other weather generators on the task of producing precipitation at different spatial scales, and the results showed that MuLGETS exhibited the advantage of preserving most of the observed statistics and producing reasonable Q-Q plots (Tseng et al. 2020). Vallam and Qin compared MuLGETS with a multi-site rainfall simulator, a Neyman–Scott based Poisson cluster model, and a proposed modified k-nearest neighbour model for 30 years of rainfall in Singapore, and found that MuLGETS exhibited the lowest differences between prediction intervals (Vallam & Qin 2016). MuLGETS can also be coupled with other methods to act as a spatial downscaling method. Chen proposed a multi-site downscaling approach coupling a quantile mapping method with MuLGETS to simultaneously downscale precipitation for multiple locations within a watershed (Chen et al. 2018). He also assessed the effectiveness of MuLGETS in hydrological modeling over a Canadian watershed in the province of Québec, and the results showed that the monthly mean discharge was accurately represented by MuLGETS-generated precipitation and temperature (Chen et al. 2016).

The multisite weather generator effectively solves the challenging problem of retaining the spatial correlation of the input data. However, previous studies on MuLGETS have generally concentrated on the consistency between the daily results of model simulations and observations with respect to indices of precipitation intensity and precipitation frequency. Moreover, it is important to evaluate MuLGETS results on timescales larger than the daily scale due to the twofold reasons as follows. First, considering possible systematic deviations in some regions, these daily scale biases may be different when calculated on the seasonal or annual scales (e.g., the seasonal biases are likely to be reduced or amplified, or changed from positive to negative), therefore, it is necessary to evaluate MuLGETS on the annual and seasonal scales over multiple sub-basins. Second, the Yangtze River Basin is the third-largest river basin in the world, with a total drainage area of approximately 1.8 million km², accounting for 18.8% of China’s land area, where extraordinary floods or high river water levels of the trunk stream were observed to occur as extreme seasonal hydroclimate events with persistent heavy rainfalls over 1–2 months (e.g., the 1998 and 2020 summer floods in the basin). Meanwhile, the subbasins of the Yangtze River Basin are greatly influenced by frequent floods in which the most severe damage seems to result from specific episodes of extreme precipitation on a regional scale, such as rainstorms and heavy precipitation events that account for high percentages of the yearly total over a few rainy days (Zhang et al. 2009). Besides, precise precipitation data, which are sometimes unavailable, are needed for prediction of floods. Hence, in this study, we evaluated the performance of MuLGETS using multiple indices of extreme precipitation with MuLGETS’s ability to reproduce long-term precipitation characteristics examined at daily, annual, and seasonal scales on data collected from the Yangtze River Basin, which we hope will provide a reference for precipitation simulation and flood forecasting in future research.
MATERIALS AND METHODS

Study area
The Yangtze River originates from Tanggula Mountain in Qinghai Province and flows through 11 provinces, administrative regions, or municipalities in China from the west to the east. The river has a total length of 6,380 km and a drainage area of 1.8 million km². It is the third-largest drainage basin in the world, and the areas through which it flows contain mountains, plateaus, basins, plains, and other topographies. The entire basin area is subjected to diverse climate types; for example, some areas of the Qinghai-Tibet Plateau climate have a subtropical monsoon climate, the southern part has a tropical climate, and the northern part has a temperate climate. The annual mean precipitation in the Yangtze River Basin is approximately 1,067 mm. However, owing to the vast area and complex topography, the temporal and spatial distribution of annual precipitation and extreme events are very uneven; the annual mean precipitation in the west is between 300 and 500 mm, while that in the southeast can reach 1,600–1,900 mm. In addition, rainfall largely occurs during summertime (Gemmer et al. 2008). Based on the climate and underlying surface conditions, we roughly divided the study area into 11 sub-basins from the west to the east. The distribution of the sub-basins and the stations is shown in Figure 1.

The area of each sub-basin and their detailed information are summarized in Table 1, which shows that the stations are evenly distributed. The average density of the stations is approximately 15,000 km² per station; however, the density is relatively low at the Jinsha River, Mintuo River, and Jialing River. In addition, three stations in the northern part of the Jialing River were disregarded because of missing essential data, resulting in sparse station distribution in the basin.

Data
The precipitation data of the initial 175 stations in the Yangtze River Basin were obtained from the daily datasets provided by the China Meteorological Data Sharing Service System (http://cdc.cma.gov.cn/index.jsp). After quality control, 123 stations with better continuity and sufficient extent of data were selected, each with complete data from January 1961 to December 2005.
**Model: MulGETS**

The weather generator is a random model that generates a comprehensive weather time series with similar statistical characteristics as the input data. When the measured climate data is insufficient to meet the complete spatial coverage, the weather generator can supplement it. The generator can also be used to simulate past or future weather sequences and has become a commonly used tool in the research on climate change and hydrological response (Liao et al. 2004; Kilsby et al. 2007; Semenov 2008).

MulGETS is a MATLAB-based stochastic weather generator, developed by Chen et al. (2014). The MulGETS Matlab routine can be found on the Mathworks file exchange website (https://www.mathworks.com/matlabcentral/fileexchange/47537-multi-site-stochastic-weather-generator-mulgets). The steps and basic principles of MulGETS downscaling for precipitation are as follows:

1. Calculate two transition probabilities based on the historical precipitation data of each station.
2. The measured precipitation transition probabilities of each station are converted into values in a normal distribution with a mean value of 1 and a standard deviation of 0 by using the inverse complementary error function. Thus, the normalized transition probabilities matrix, norm_trans, is obtained. Then, the Cholesky method is used to generate a spatially correlated random number matrix, corr_random, with a size of \( m \times n \) (where \( m \) is the product of the number of days in a certain month and the number of years in the generated sequence and \( n \) is the number of stations in the basin).
3. Based on the stability ineffectiveness theory of Markov, assuming that the state of the day is state1, each value of corr_random generated in step 2 is compared with its corresponding parameter, norm_trans. According to the results, the precipitation states of next day, state2, are obtained.
4. \( K_m \) is introduced in the problem of spatial intermittence proposed by Brissette et al. (2007). Then, the daily precipitation occurrence coefficient \( K_m \) is calculated for each season as follows:

\[
K_m = \frac{O \times C^T}{U \times C^T}
\]

where \( O \) is the row vector of indicating whether precipitation occurs in the surrounding stations of a certain station on a certain day, and \( C^T \) is the column vector of the seasonal correlation coefficient of precipitation occurrence between a certain station and its surrounding stations. According to the magnitude of \( K_m \), the observed data and its corresponding \( K_m \) are divided into 9 intervals, an evenly distributed between 0 and 1, if a class contains fewer than 50 precipitation events, two or more neighboring classes are combined until there are more than 50 events in each of the new classes.
5. Parameters for exponential or gamma distribution are then calculated by devising a nonlinear regression equation of the \( K_m \) group median and the observed precipitation mean or standard deviation sequence of all points in the interval. Then, the simulated average precipitation is corrected with the observed values, and the parameters of the exponential or gamma distributions are calculated. For gamma distribution, MulGETS calibrates several pairs of \((\alpha, \beta)\) for each site and season.

### Table 1 | Primary information of sub-basins

<table>
<thead>
<tr>
<th>Region Categories</th>
<th>Sub-basin</th>
<th>Abbreviation</th>
<th>Area/10^4 km^2</th>
<th>No. of stations</th>
<th>Station density (10^4 km^2/station)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper region</td>
<td>Jinsha River</td>
<td>JSR</td>
<td>47.4</td>
<td>20</td>
<td>2.37</td>
</tr>
<tr>
<td></td>
<td>Mintuo River</td>
<td>MTR</td>
<td>16.3</td>
<td>10</td>
<td>1.63</td>
</tr>
<tr>
<td></td>
<td>Jialing River</td>
<td>JLR</td>
<td>16.3</td>
<td>7</td>
<td>2.33</td>
</tr>
<tr>
<td></td>
<td>Wujiang River</td>
<td>WJR</td>
<td>8.9</td>
<td>8</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td>Upper mainstream section</td>
<td>UMS</td>
<td>10.3</td>
<td>8</td>
<td>1.29</td>
</tr>
<tr>
<td>Middle region</td>
<td>Hanjiang River</td>
<td>HJR</td>
<td>15.8</td>
<td>14</td>
<td>1.23</td>
</tr>
<tr>
<td></td>
<td>Middle mainstream section</td>
<td>MMS</td>
<td>27.2</td>
<td>23</td>
<td>1.18</td>
</tr>
<tr>
<td></td>
<td>Poyang Lake</td>
<td>PYL</td>
<td>17.3</td>
<td>15</td>
<td>1.53</td>
</tr>
<tr>
<td>Lower region</td>
<td>Dongting Lake</td>
<td>DTL</td>
<td>10.0</td>
<td>7</td>
<td>1.43</td>
</tr>
<tr>
<td></td>
<td>Lower mainstream section</td>
<td>LMS</td>
<td>9.3</td>
<td>9</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>Tai Lake</td>
<td>THL</td>
<td>4.0</td>
<td>4</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>–</td>
<td>182.8</td>
<td>125</td>
<td>1.46</td>
</tr>
</tbody>
</table>
• Finally, a random precipitation amount is generated according to the precipitation occurrence coefficient and the related parameters by determining which $K_m$ interval falls on a certain day and generates a certain amount of precipitation based on the interval’s parameter and selected distribution.

Experimental design
For each sub-basin, observation data of the stations collected from 1961 to 2005 served as the input data. The MulGETS was used to generate the precipitation series on a daily scale. The only parameters required for the MulGETS operation are the daily precipitation occurrence threshold and type of precipitation distribution function. In this study, 0.1 mm/day is selected as the threshold and the gamma function is the distribution function. The mean and standard deviation of precipitation and occurrence index for each station of each sub-basin are provided in the Supplementary Material.

MulGETS is generally believed to generate a sequence that temporally coincides with the input values. When evaluating MulGETS, all sequences from 1961 to 2005 were used.

Most studies use explanatory variance, root mean squared error, standard deviation, and other indices to evaluate the simulation effect of the model (Huang et al. 2010; Liu et al. 2011). Some other scholars use more abundant indices, such as quantiles, median, maximum value, and minimum value, for evaluation (Wetterhall et al. 2006). In this study, we selected nine indices categorized into three types, which originated from the European Union STARDEX project (STAtistical and Regional dynamical Downscaling of EXtremes for the European region) (Goodess et al. 2012). In daily scale evaluation, we reserved standard deviation (SD) and 99th percentile of rain day amounts (Pq99) as extra indices. The abbreviations and definitions of extreme precipitation indices are shown in Table 2.

RESULTS AND DISCUSSION
Overview of simulated daily precipitation
This subsection presents an overview of simulated daily precipitation, in which the SD, 95% percentiles (Pq95), and 99% percentiles (Pq99) were calculated from the MulGETS daily precipitation over 1961–2005.

The spatial distribution of the SD of the observed and simulated precipitation and their relative errors are shown in Figure 2. The SD of simulated precipitation on the daily scale of MulGETS was generally underestimated, and it was 1–5% smaller in most areas, indicating that the dispersion of precipitation series may not be estimated well. However, the situation varied for each sub-basin. At most stations of the middle mainstream and stations near the borders between the Middle mainstream and Wujiang River and the borders between Tai Lake and Lower mainstream, the SD of the simulated values was overestimated by 1–5%, and for several stations was even overestimated by 5–10%. This may be attributed to the steep terrain and mountainous areas in the upper reaches of the Yangtze River, and the heterogeneity of the spatial distribution thereof. In the other sub-basins, there are a few stations with positive RE, which are different from most stations with negative RE. These stations are mostly near the boundary of the basin, which is usually a watershed formed by high mountains. Therefore, it is possible that MulGETS is more likely to produce a more widely distributed precipitation series in mountainous areas and more narrowly distributed ones in plains.

| Table 2 | Abbreviations and definitions of the indices |
| Category | Index | Unit | Definition |
| Precipitation intensity | Pint | mm/day | Simple daily intensity (rain per rain day) |
| Ppdd | % | Mean dry-day persistence |
| Pnl90 | day | No. of events $>\text{long-term } 90^{th} \text{ percentile}$ |
| Pq95 | mm | 95th percentile of rain day amounts |
| Precipitation magnitude | Pf90 | mm | Total rainfall from events $>\text{long-term } 90^{th} \text{ percentile}$ |
| Px5d | mm | Greatest 5-day total rainfall per year |
| Pxccd | day | Max no. of consecutive dry days |
| Precipitation persistence | Pn10 mm | day | No. of days with precipitation $\geq 10$ mm |
| Pdsav | day | Mean dry-day spell lengths |
The simulated values of $P_{q95}$ and $P_{q99}$ were relatively accurate, and the errors in most regions were within 5% (Figure 3). Large negative REs of $P_{q95}$ were distributed mainly in the Jinsha River and northern Hanjiang River, while positive REs were located mainly near the borders between the middle mainstream section, Mintuo River, and Jinsha River. Compared with $P_{q95}$, $P_{q99}$ showed fewer large negative REs, but the spatial distribution of REs was more scattered. In general, in most sub-basins, MulGETS can capture precipitation maximums and reproduce their spatial distribution characteristics.

MulGETS is characterized by its ability to generate precipitation sequences with spatial correlations. The correlation coefficient of precipitation occurrence between stations in the middle and lower reaches of the Yangtze River Basin is larger than that in the upper reaches of the Yangtze River Basin (Figure 4(b1)–4(b11)). MulGETS was able to generate precipitation occurrence sequences consistent with the observed values for all sub-basins, in which all the scattered points were near the 1:1 line. Although with respect to precipitation amounts, the scattered points were farther from the 1:1 line, overall they were still along the line. In the Tai Lake, lower mainstream section, and Poyang Lake areas, the spatial correlation coefficients of simulated values were often lower than that of observation, while those of Jinsha River and Wujiang River were higher than observation. On a daily scale, a good simulation of precipitation occurrence cannot guarantee a good spatial correlation of precipitation amounts. The interstation correlation is induced at the monthly scale for precipitation occurrence and at the seasonal scale for precipitation amounts, so the accuracy of spatial correlation on the daily scale may not be very high. When evaluating the effect of the weather generator, the indices calculated on different time scales may exhibit different performance characteristics, which are determined by the construction and properties of the model.
Annual-scale precipitation

In the 11 sub-basins of the Yangtze River Basin, MulGETS was used to downscale precipitation from 1961 to 2005. The simulation results were evaluated using 9 indices and analysed at both annual and seasonal scales. At the seasonal scale, summer and winter were selected as representatives.

Spatial distribution of pint

The Pint of the Yangtze River Basin has an increasing pattern from the west to the east (Figure 5(a)). The observed precipitation amount in the lower reaches is greater than that in the middle reaches, and it is higher in the middle reaches than in the upper reaches. Moreover, stations on the south and the north sides of the basin were rainier. The comparison of Figure 5(a) and 5(b) shows that MulGETS can reasonably simulate the Pint of precipitation at most stations in the Yangtze River Basin and retain the increasing pattern of low and high values in the west and east, respectively. The spatial distribution of the simulated values was also largely consistent with the observed values. The simulation effects of each sub-basin were different.
As shown in Figure 5(c), the simulation errors of most stations were between −5 and 5%. The stations with errors below −5% were mainly located in the Jinsha River. In comparison, the stations with errors more significant than 5% were mainly located in the Chongqing and Sichuan area, eastern Mintuo River, southern Jialing River, western Wujiang River, and western upper mainstream. This region is the junction of mountains and plains; therefore, precipitation at different stations in the same sub-basin varies greatly, leading to precipitation homogenization at some stations and sizeable relative errors of simulation values.

**Performance of various indices**

The boxplots of each index are shown in Figure 6(a)–6(i). Overall, MulGETS has a good simulation effect on each index, and the box line of the simulated value is generally consistent with the observations.

MulGETS can simulate the precipitation intensity indices accurately on an annual scale (Figure 6(a)–6(c)). It can also reasonably reproduce each station’s Pint characteristics (Figure 6(a)). For example, the observed and simulated values of each station are in good agreement, and the characteristics of different stations (even adjacent ones) were also well distinguished. In most sub-basins, Ppdd shows a slight difference between stations; MulGETS can accurately capture the differences between stations in other basins, such as the Ppdd in some stations in the Mintuo and Jinsha River. However, at stations where Ppdd has high interannual variation, the simulated Ppdd shows lower variability; this indicates that the error will be larger in years when drought time is overconcentrated or scattered. In general, Ppdd in these basins is above 50%, and most of them reach 80%, indicating a high proportion of consecutive dry days. Pnl90 reflects the total number of precipitation days. As shown, Pnl90 has an inverse correlation with Ppdd. MulGETS slightly underestimated the interannual variance in Pnl90.

Clearly, the precipitation magnitude indices of the simulated values are mostly consistent with the observations; however, the range of boxplots at some stations is slightly different from the actual values (Figure 6(d)–6(f)). Specifically, there was a good correspondence between Pq95 and Pint. As shown in Figure 6(a) and 6(d), stations with high Pint are more likely to have a higher Pq95, and the interannual variation of Pq95 is greater than that of Pint. MulGETS also have good simulating effect on Pq95 (Figure 8(d)). For stations where precipitation varies widely from year to year, the simulated values show the same features and vice versa. However, the simulation of Pnl90 is much larger overall, especially in the Jinsha River, Mintuo River, Jialing River, Wujiang River, Hanjiang River, and Dongting Lake, most of which are in the middle and upper reaches of

**Figure 6 | Comparison between indices of annual scale observation and MulGETS simulated precipitation.**
the Yangtze River Basin. The errors of Pnl90 seem to indicate that overestimating the 10% largest part of precipitation is one of the reasons for the systematic errors. However, Px5d was underestimated in most basins. These two inverse errors may cancel each other to ensure a more accurate overall mean, such as Pint.

Most of the annual scale precipitation persistence indices are well simulated; however, in some basins, a systematic error was observed (Figure 6(g)–6(i)). Stations with a relatively medium Pxcdd had a more accurate Pxcdd (Figure 6(g)–6(i)). For example, the simulated Pxcdd values for some stations in the Jinsha River and Mintuo River are slightly overestimated; for the Jialing River, Hanjiang River, the lower mainstream, and Tai Lake they are accurate; for Wujiang River, Dongting Lake, and the middle mainstream basins, they are all low. Because Pxcdd is underestimated in the sub-basins with lower observed Pxcdd, the influence of the simulated errors is comparably small. However, for the stations where Pxcdd is high, overestimation will result in more drought events. Pn10 mm is, to some extent, consistent with Pq95, which has a good relationship with Pint. MulGETS's simulation value for Pn10 mm is accurate at most stations and slightly low at some of the Jinsha River, Mintuo River, and Hanjiang River stations. Pdsav is also well simulated at most stations, and MulGETS precisely captures its interannual variation and interstation differences. Moreover, Pdsav has a positive relationship with Ppdd (Figure 6(b)). However, the Ppdd simulated values are not as precise as that of Pdsav for MulGETS, thus underestimating its interannual variation.

**Seasonal-scale precipitation**

**Spatial distribution of pint**

On a seasonal scale, the simulated Pint is in good agreement with the observed ones, and the differences between each sub-basin and the overall distribution can be well simulated. In the Yangtze River Basin, summer precipitation accounted for more than half of the annual precipitation and the Pint is approximately 7.0–24.5 mm/day; in winter, it is less than 10 mm/day in general. The interannual mean precipitation has a spatial distribution of gradually increasing from the upper reaches to the lower reaches, especially in the winter (Figure 7(d)).

In summer, the Sichuan area receives a considerable amount of precipitation; at some stations, the amount can be equivalent to the lower reaches (Figure 7(a)). As shown in Figure 7(a) and 7(b), the precipitation conditions of some stations differ from those of the surrounding stations, forming an evident local closure; for example, some stations in the Jinsha River and Hanjiang River. MulGETS can capture such differences (Figure 7(c) and 7(d)), indicating that the model does not provide a completely average precipitation.

At most stations, the relative error of Pint was between −5 and 5%, and the simulation effect was good (Figure 7(e) and 7(f)). The negative values were mainly concentrated in the upstream basins. In summer, the simulated Pint is overestimated in most sub-basins, whereas in winter, only a few stations have overestimated simulation values in the Jialing River, Hanjiang River, and Wujiang River. Unlike the annual scale simulation, the positive relative errors in summer and winter are scattered rather than concentrated in the Sichuan area. Because there is less precipitation in winter, the denominator of the relative error calculation is smaller. If the absolute error is the same, the relative error will be much larger than that in summer. Therefore, the distribution area of stations with significant errors in winter was larger than that in summer throughout the year (Figure 8(f)).

MulGETS can closely simulate the spatial distribution of Pint in the Yangtze River Basin; however, the relative error at some stations is still difficult to ignore. The accuracy of the model should be increased, especially in stations with less precipitation.

**Performance of various indices**

The boxplots of the seasonal scale indices are shown in Figures 8 and 9. It was found that MulGETS can precisely simulate precipitation at a seasonal scale. For the precipitation intensity indices, the overall effect is good, especially for Ppdd and Pnl90, which are highly consistent with the observations. In the summer, the Pint is larger than the annual mean; in the winter, it is rather small and mostly below 10 mm/day. MulGETS has a good simulation effect in both seasons (Figures 8(a) and 9(a)). Nevertheless, it faces certain problems, in several sub-basins, the interannual variation of simulated values at various stations tends to be consistent. For the Jialing River in Figure 8(a), the maximum interannual observations were not captured. In winter, the Jinsha River as well as the Mintuo River, where the observations are low, have inaccurate Pint’s simulated values; the difference between the stations is not captured as well. The simulated summer Ppdd was lower than that of the annual scale, but its interannual variation range was higher than that of the annual scale (Figure 8(b)), which agrees with
the observations. This is probably because summer precipitation in the Yangtze River Basin is affected by large-scale ocean circulation and sea surface temperature anomalies such as Hardley circulation and ENSO (El Niño-Southern Oscillation) (Zhou & Wang 2006; Luo et al. 2018), which lead to large interannual fluctuations during the summer precipitation. The simulated Pnl90 results are very consistent with the observation (Figure 8(c)); this is probably because the model refers to the observed rainy/no-rain days and only activates rainfall simulation on rainy days. Therefore, with a steady total number of days, the number of rain days will be close to the observation, as will the Pnl90. The shorter the period, the more concentrated is Pnl90, and thus the boxplot is compressed into a single line (even during winter). This index shows interannual differences on an annual scale (with a longer time range), although its interannual variation is underestimated. In winter, MulGETS can reproduce Ppdd and Pnl90 in most basins (Figure 9(b) and 9(c)). Overall, there is a need to improve the simulation effect of MulGETS for winter precipitation.

The range of seasonal scale precipitation magnitude indices is wide, but MulGETS can simulate each sub-basin’s characteristics and the differences between stations. Compared with the annual scale Pq95, the simulated Pq95 in the summer was not significantly lower. However, in some sub-basins, the interannual variation in simulated values is underestimated. For example, in the Jialing River (Figure 8(d)), the interannual maximum values of some stations were not captured accurately.
The simulation results of Pq95 in the winter are in good agreement with those of Pint, and there is a significant error in some of the Jinsha River and Mintuo River stations. For a basin with less precipitation in the winter, most rain days may be concentrated in a few short periods so that the error of Pq95 will lead to a corresponding error of Pint. In the summer, the Pf90 is slightly overestimated in each sub-basin, indicating that MulGETS simulation values have a systematic error in this index (Figures 8(e) and 9(e)). Pfl90 is larger than Pnl90, but Pnl90 is more accurate, indicating that the simulation value of MulGETS is not accurate enough to capture the distribution of daily precipitation in the summer. In summer, Px5d was generally underestimated in each sub-basin (Figure 8(f)). In winter, the effect was not good. Specifically, the Px5d and its interannual variation are underestimated in the Jinsha River, Mintuo River, Jialing River, Wujiang River, and the mainstream basins of the upper reaches of the Yangtze River. In summary, on a seasonal scale, MulGETS can simulate the differences between stations but is still weak in heavy rainfall simulations. This is a disadvantage when applying the model to precipitation prediction and flood risk management in the Yangtze River Basin.

The simulation effects were varied for the precipitation persistence indices. In general, the simulation effect of Pxcdd in the middle and lower reaches of the Yangtze River Basin is better than that in the upper reaches of the Jinsha River, Jialing River, Mintuo River, and Wujiang River (Figure 8(g)). For example, at the Jialing River Basin’s first station, the simulated value was approximately 50% higher than the maximum of inputs and, at the seventh station, 50% lower than the maximum value. Pxcdd was underestimated in most basins in winter; however, the difference was not significant (Figure 9(g)). Pn10 mm is generally underestimated in the Jinsha and Mintuo River and partly underestimated in the Hanjiang River and Dongting Lake. The minimum interannual values were all lower than the observation minima (Figure 8(h)). Some simulated values reached zero in the middle and upper reaches in the winter. Furthermore, when the observation is zero, the simulated result must be zero but not vice versa (Figure 9(h) and 9(i)). This inclination indicates that although the MulGETS simulation result at one station is affected by its surrounding stations, its own precipitation condition still plays a decisive role. The Pdsav of the simulated value is in good agreement with the observation. However, the interannual variance is overestimated at some stations, such as stations in the Jialing River, Hanjiang River, and Dongting Lake (Figure 8(i)). In summary, MulGETS tends to underestimate the number of days with precipitation of more than 10 mm in the summer and overestimate the most extended length without rain, which is also one reason for some other indices’ errors.

In summary, owing to the underestimation of heavy rain in summer and the overestimation of rainless days, MulGETS tends to produce a precipitation series that is drier than the observation. However, the extent is not the same in the sub-basins.
Specifically, the simulation effect at the stations at the middle and lower reaches, with more precipitation, is better than that at the upper reaches with less precipitation.

CONCLUSIONS

The primary purpose of the weather generators is to reconstruct data with statistical features close to the input observation data, where one crucial and challenging factor is to reconstruct the time autocorrelation and space correlation of the data. We evaluated MulGETS, a multisite stochastic weather generator, for its simulation effect on precipitation in the Yangtze River Basin from the perspective of both time and space. The results showed that MulGETS can effectively generate spatially correlated variables. The simulated sequences retained the spatial distribution and temporal variation of the input data to a great extent. The simulation effect of the model varied for each sub-basin. The simulation effect of the model for the basins at the middle and lower reaches was better than that for the upper reaches. A possible explanation is the scanty rainfall conditions and the influence of mountainous topography in the upper reaches. On the time scale, the model performs better on the annual scale than on the seasonal scale, and it performs better in summer than in winter.

MulGETS provides satisfactory results for the simulation of intensity, magnitude, and duration indices and is more accurate at simulating the indices reflecting rainless days. However, its ability to capture maximum values must be improved.

In addition, we proposed a method to evaluate multisite downscaling models by analysing the effects of simulations in terms of both spatial distribution and temporal variation. This study provides a reference not only to apply MulGETS in the Yangtze River Basin but also to understand and improve the MulGETS. This study only examined the simulation effect of the MulGETS model. The next step will be to add results from other models for comparison, which will help to further understand the characteristics of the multisite downscaling model.

ACKNOWLEDGEMENTS

We gratefully acknowledge the source code of the MulGETS shared by Professor Chen. The authors would like to thank two anonymous reviewers for their helpful comments and suggestions.
AUTHOR CONTRIBUTIONS
Conceptualization, Yu Li, X. M. Zeng, J. L. Guo; methodology, Yu Li, X. M. Zeng, J. L. Guo; formal analysis, Yu Li; writing – original draft preparation, Yu Li; writing – review and editing, Yu Li, X. M. Zeng; visualization, Yu Li; supervision, X. M. Zeng; funding acquisition, X. M. Zeng. All authors have read and agreed to the published version of the manuscript.

FUNDINGS
This research was funded by the National Key Research and Development Program of China, grant number 2018YFC1508200, the National Natural Science Foundation of China, grant number 42005023, the Fundamental Research Funds for the Central Universities, grant number B210203067 and Open Funding of Hubei Key Laboratory of Intelligent Yangtze and Hydroelectric Science, China Yangtze Power Co., Ltd. (NO. ZH2002000103).

COMPETING INTEREST
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY STATEMENT
All relevant data are available from an online repository or repositories (http://data.cma.cn/).

REFERENCES


First received 29 April 2021; accepted in revised form 24 November 2021. Available online 10 December 2021.