

A Two-Stage Multisite and Multivariate Weather Generator

Z. Li^{1*}, J. J. Li¹, and X. P. Shi¹

¹State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, College of Natural Resources and Environment, Northwest A&F University, Yangling, Shaanxi, 712100, China

Received 13 September 2016; revised 21 March 2017; accepted 26 October 2017; published online 27 November 2019

ABSTRACT. The spatial structure of climatic variables synthesized by a weather generator has considerable impact on the modeling of hydrological variability; however, in most cases, it needs computationally intensive work to reproduce multisite and/or multivariate correlations. This work proposed a two-stage weather generator (TSWG) to preserve intersite and intervariable correlations of daily precipitation, maximum and minimum temperatures. The first stage generates daily precipitation and temperature for each site and for each variable with, but not limited to, the Richardson-type approach. The second stage rebuilds the multisite multivariate correlation using a distribution-free shuffle procedure. The TSWG was applied to a network of 15 stations in the Jing River catchment (Northwest China). It reproduced the statistical parameters and multisite and multivariate correlations well. Furthermore, indirect validation by hydrological modeling showed TSWG outputs could be used satisfactorily for simulating streamflow variability. As a distribution-free method, the correlation reconstruction method can be applied to variables with different probability distributions. The TSWG can efficiently reconstruct the correlation with one optimization for all stations and all variables, which is superior to most current methods operated once for one station pair and one variable. The TSWG provides an option for improved multisite and multivariate weather generation.

Keywords: weather generator, multisite correlation, multivariate correlation, hydrological modeling

1. Introduction

Weather generators are powerful tools that can produce long-term climate data time series with statistical characteristics similar to those of observations (Richardson, 1981; Bárdossy and Plate, 1992; Semenov and Barrow, 1997; Wilks, 1998), and they have been used widely in impact assessment of climate change on hydrology (Wilby and Harris, 2006; Zhang et al., 2011). However, one important aspect related to the application of weather generators to hydrological modeling is that their spatial structure has considerable impact on the modeling of hydrological variability (Clark et al., 2004; Mehrotra and Sharma, 2007; Srikanthan and Pegram, 2009; Bárdossy and Pegram, 2012). For example, when using climate data produced by a single-site weather generator (SSWG) as the input to a conceptual and/or distributed hydrological model, the floods and variances of streamflow are substantially underestimated, whereas they are reproduced well when using the synthetic climate from a multisite weather generator (MSWG) (Khalili et al., 2011; Li, 2014; Chen et al., 2016; Li et al., 2017). This highlights the importance of multisite weather generation for the modeling of hydrological variability.

Three types of weather generators considering the spatial

structure of climatic variables have been developed: nonparametric, parametric, and semiparametric models (Mehrotra and Sharma, 2009; Li, 2014; Srivastav and Simonovic, 2015). A nonparametric MSWG mainly incorporates two approaches: a hidden Markov chain model and a resampling method. The hidden Markov chain model reproduces spatial correlation through relating some exogenous predictors to daily weather data at multiple sites via a finite number of hidden weather states (Rex, 1993; Boorman and Sefton, 1997; Thyer and Kuczera, 2003a, b; Mehrotra et al., 2006); however, it is comparatively complex because of the excessive verification and optimization required (Wilks, 1999; Mehrotra et al., 2006). The resampling method can simultaneously generate multisite and multivariate climate by conditionally resampling the observations, but it remains difficult to extrapolate the potential changes in climatic extremes and variability (Beersma and Buishand, 2003; Wilby et al., 2003; Caraway et al., 2014; King et al., 2014).

A parametric MSWG consists of two principal methods: a chain-dependent process and generalized linear models. The multisite chain-dependent process is extended from the single-site version using spatially correlated random numbers (Wilks, 1998, 1999). However, the random number generation is computationally intensive because a collection of $k \times (k-1)/2$ empirical curves has to be developed for all station pairs, for each precipitation process or variable, and for each month for a network of k stations (Qian et al., 2002; Brissette et al., 2007; Srikanthan and Pegram, 2009; Baigorria and Jones, 2010). The generalized linear models perform well in weather generation

* Corresponding author. Tel.: +86 29 87082069; fax: +86 29 87080055.
E-mail address: lizhibox@126.com (Z. Li).

and GCM downscaling, especially when large-scale predictors are introduced; however, they require considerable parameter estimation in connection with the relationships between the predictands and the predictors (Chandler and Wheeler, 2002; Yang et al., 2005; Furrer and Katz, 2007; Verdin et al., 2015; Asong et al., 2016). Although parameter estimation is not a challenge with the increase of computing power, simply structured and easy-to-implement approaches are highly desired.

Comparatively, semiparametric models simplify the algorithm substantially by combining the strengths of both parametric and nonparametric approaches (Angulo et al., 1998; Fowler et al., 2005; Mehrotra and Sharma, 2007; Srivastav and Simonovic, 2015), most of which are hybrid Markov-based models with a resampling method (Palutikof et al., 2002; Apipattanavis et al., 2007; Steinschneider and Brown, 2013). Among the semiparametric methods, the matrix recorrelation approach can efficiently reconstruct multisite correlations by adding a postprocessing step to the data obtained from the SSWG or other sources. For example, Iman and Conover developed a postprocessing, distribution-free method for pairing independent variables to induce the desired rank correlation (Iman and Conover, 1982). As a distribution-free method, its application is not limited to the probability distribution of data and thus, it has been used effectively in recorrelating precipitation amounts and durations (Zhang, 2005; Chen et al., 2009), and multiple series of precipitation from SSWG (Tarpanelli et al., 2012; Li, 2014). A similar method has been used successfully for the improvement of the correlation of precipitation obtained from regional climate models (Bárdossy and Pegram, 2012; Pegram and Bárdossy, 2013).

Overall, most of the above models can satisfactorily generate multisite and/or multivariate weather data; however, because of differences in the structures and algorithms of the models, their efficiencies differ considerably. Given that most models are computationally intensive, semiparametric models can efficiently reconstruct intersite or intervariable correlations, even for variables with different probability distributions. Furthermore, the performances of semiparametric methods are found comparable with those of parametric methods, both for weather statistics and hydrological modeling (Tarpanelli et al., 2012; Li, 2014). Therefore, semiparametric MSWGs have considerable potential in multisite and multivariate weather generation and studies of the impact of climate change.

A semiparametric model, e.g., a two-stage weather generator (TSWG), was developed for multisite precipitation generation in our earlier study (Li, 2014). In the first stage, the TSWG generates single-site precipitation and it then rebuilds the spatial coherency in the second stage using a distribution-free method. This article describes the extension of the TSWG to incorporate daily maximum and minimum temperatures to provide a useful tool for multisite and multivariate weather generation. The remainder of the paper is organized as follows. Section 2 provides a detailed description of the algorithm for the model development. Section 3 assesses the model performance through direct and indirect evaluation, and the final section summarizes and discusses the advantages and disadvantages of the TSWG.

2. Model Formulation

The method was developed and evaluated based on the Jing River catchment in northwest China, which encompasses an area of 45,421 km² (Figure 1). The Jing River catchment was selected because it is a typical catchment with high intra-annual and interannual variability of climate and runoff. The area-averaged annual precipitation is 542.1 mm, in which snow in winter and spring can be negligible and 55% of the precipitation falls in the flood season between July and September (1961 ~ 2005). Therefore, the variable climate and runoff are appropriate to investigate the model performance related to variability of climate and runoff. The daily precipitation, maximum and minimum temperatures obtained from 15 weather stations during 1961 ~ 2001 were used for the model development and direct evaluation. The monthly streamflow at the catchment outlet for the same period was used to calibrate a semidistributed hydrological model, e.g., the Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998), and for indirect evaluation of the performance of the TSWG.

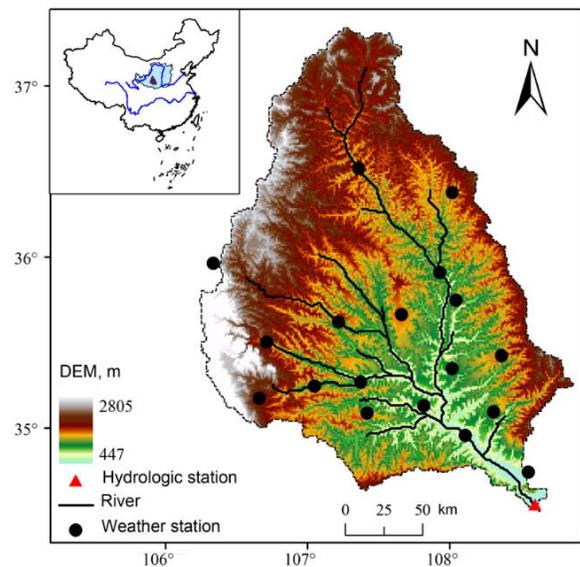


Figure 1. Location of the Jing River catchment (NW China) used as an example study area.

Within the framework of the TSWG, a two-stage operation is undertaken for multisite and multivariate weather generation. The first stage generates daily precipitation (P) and maximum (T_{max}) and minimum (T_{min}) temperatures on a single-site and a single-variable basis. The second stage rebuilds the intersite and intervariable correlations. A flow chart of the model development and evaluation is presented in Figure 2.

2.1. First Stage: Single-Site Weather Generation

The daily data were generated for each station according to the Richardson-type approach (Richardson, 1981); however, the weather generation algorithm can be changed to use other methods, such as the semi-empirical methods in the LARS-

WG (Racsco et al., 1991; Semenov and Barrow, 1997). The Richardson-type approach has been documented in other literature (Richardson and Nicks, 1990; Wilks and Wilby, 1999; Li et al., 2013, 2014) and therefore, the procedure of weather generation is described briefly here. In the chain-dependent process, the precipitation occurrence and amounts are generated by the first-order Markov chain and the skewed normal distribution, respectively. Daily T_{max} and T_{min} are simulated stochastically using a standard normal distribution without consideration of the dependence on the dry/wet status and the intervariable correlation. The parameter estimation and generation of precipitation and temperature are all performed on a monthly basis. The 500-year daily data for each station and each variable are simulated. Each variable is generated independently and the data for each day are generated independent of others, which is not true in reality because they are actually cross- and serial correlated variables.

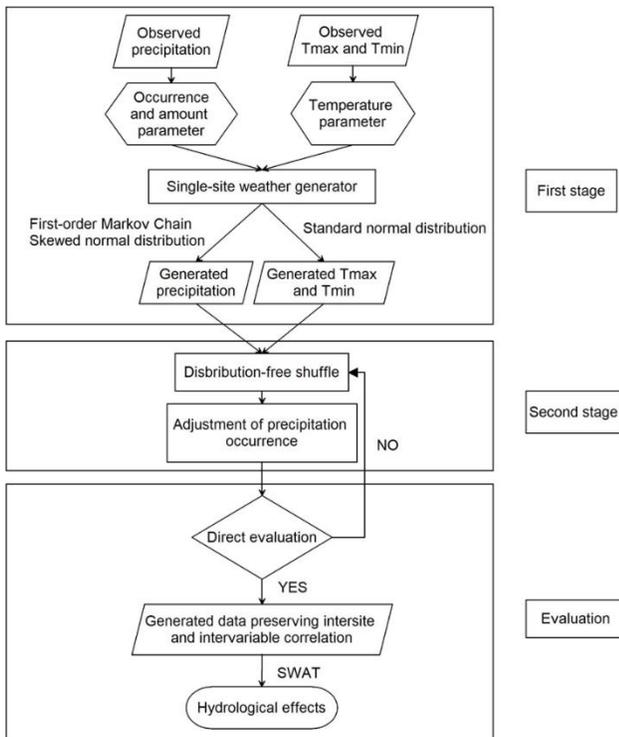


Figure 2. Flowchart of the development and evaluation of the two-stage weather generator.

2.2. Second Stage: Intersite and Intervariable Correlation Reconstruction

The second stage rebuilds the intersite and intervariable correlations of precipitation and temperature through the distribution-free method developed by Iman and Conover (1982). The theoretical basis can be described as follows. To assign a desired correlation matrix $[C]$ to a random row vector $[X]$, two steps should be undertaken. First, $[C]$ is decomposed as $[C] = [R][R']$ and then, the upper triangular matrix $[R']$ is used to multiply $[X]$ to generate a new matrix $[X][R']$ with the desired correlation matrix $[C]$.

In previous version of TSWG in Li (2014), $[X]$ is a data matrix for multisite precipitation only. However, because this study extended the TSWG to a multivariate model, $[X]$ becomes an $n \times k$ matrix with n rows of daily data for SSWG-generated precipitation and temperature for a certain month ($n = \text{years} \times \text{days}$ in a month, e.g., 15,500 for January in this study) at k columns of stations and variables ($k = \text{stations} \times \text{variables}$, e.g., 45 in this study). The matrix re-correlation can be accomplished in five steps, and example datasets of the SSWG-generated precipitation for 10 days at weather stations $G1$ and $G2$ with a correlation coefficient of 0.036 are presented in Table 1.

Step 1. Data transformation. The raw data cannot be used directly for matrix re-correlation and they have to be converted to scores following a certain probability distribution. For example, Bárdossy and Pegram (2012) converted the ranks of precipitation to cumulative probabilities. This study converted the ranks of $[X]$ to an $n \times k$ van der Waerden scores matrix $[S]$, which can be calculated using the inverse function of a standard normal distribution (Li, 2014). Because tied ranks increase the difficulty in the subsequent reordering step, two operations are used to assign a unique rank to each data element. First, additional decimal places are retained from the SSWG to differentiate nonzero values, which works well for temperature and wet-day precipitation amounts. Second, small values of less than the threshold definition of a wet event (0.1 mm in this study) are assigned to dry days. The raw data $[X]$ are first multiplied by 1,000 to make the precipitation threshold 100, which enlarges the range of trace values to be added to dry days. Then, a precipitation amount is assigned to the j^{th} dry day with the value of $j \times 99 \div m$, where $j = 1, 2, \dots, m$ and m is the total number of dry days. A new temporary matrix $[T1]$ is generated in which the dry days are arranged in increasing order and thus, they are serially dependent. A random permutation is performed for each column in $[T1]$ to obtain another temporary data matrix $[T2]$. Therefore, $[T2]$ is actually the matrix used for the data transformation to obtain the *Rank1* and *Score* values shown in Table 1.

Step 2. Decomposition of correlation matrix. If $[C_{obs}]$ is the $k \times k$ observed rank correlation matrix, it can be decomposed as $[C_{obs}] = [R][R']$ by the Cholesky factorization scheme. It should be noted that $[C_{obs}]$ might be a nonpositive correlation matrix, which makes Cholesky factorization impossible. In such cases, a spectral decomposition procedure is used to generate a new positive-definite matrix (Rebonato and Jäckel, 2000). For example, the observed correlation matrix $[C_{obs}]$ of [1, 0.605; 0.605, 1] can be decomposed to obtain an upper triangular matrix $[R']$ of [1, 0.605; 0, 0.796].

Step 3. Data reordering. A new $n \times k$ score matrix $[Score\ 2]$, shown in Table 1, can be obtained by multiplying the $n \times k$ $[Score\ 1]$ with the $k \times k$ $[R']$. $[Rank\ 2]$ are the ranks of $[Score\ 2]$ and they can be used to rearrange $[T2]$ to obtain $[X1]$ with the desired correlation matrix. However, it is possible that the actual correlation matrix of $[X1]$ will not be as expected. For example, the correlation matrix of $[X1]$ is actually [1, 0.663; 0.663, 1], where the values are greater than the target correlation.

Step 4. Improvement of intersite and intervariable correlations. The rebuilt correlation for $[X1]$ was 0.663, which

Table 1. Example Datasets for the Shuffle Procedure

Day	X		T1		T2		Rank 1		Score 1		Score 2	
	G1	G2	G1	G2	G1	G2	G1	G2	G1	G2	G1	G2
1	0	5.1	16.5	5135.3	49.5	10574.7	3	9	-0.60	0.91	-0.60	0.36
2	0	0	33.0	19.8	16.5	59.4	1	3	-1.34	-0.60	-1.34	-1.29
3	0	0	49.5	39.6	1938.3	39.6	7	2	0.35	-0.91	0.35	-0.51
4	0	48.6	66.0	48580.8	33.0	19.8	2	1	-0.91	-1.34	-0.91	-1.61
5	0	0	82.5	59.4	39943.8	1032.8	10	7	1.34	0.35	1.34	1.09
6	1.9	0	1938.3	79.2	2937.9	99.0	8	5	0.60	-0.11	0.60	0.27
7	31.5	0	31510.4	99.0	31510.4	48580.8	9	10	0.91	1.34	0.91	1.61
8	39.9	10.6	39943.8	10574.7	82.5	100.0	5	6	-0.11	0.11	-0.11	0.02
9	0	1.0	99.0	1032.8	66.0	5135.3	4	8	-0.35	0.60	-0.35	0.27
10	2.9	0.1	2937.9	100.0	99.0	79.2	6	4	0.11	-0.35	0.11	-0.21
Correlation	0.036								0.357		0.651	

Note: The desired correlation coefficient is 0.605. X is the SSWG-generated daily precipitation for stations G1 and G2. The description of the other variables can be found in Section 2.2.

Table 1. Continued

Rank 2		X1		X1-occ		Score 3		Rank 3		X2		X2-occ	
G1	G2	G1	G2	G1	G2	G1	G2	G1	G2	G1	G2	G1	G2
3	8	0	5.1	0	5.1	-0.60	0.37	3	8	0	5.1	0	5.1
1	2	0	0	0	0	-1.34	-1.28	1	2	0	0	0	0
7	3	1.9	0	0	0	0.35	-0.52	7	3	1.9	0	0	0
2	1	0	0	0	0	-0.91	-1.61	2	1	0	0	0	0
10	9	39.9	10.6	0	0	1.34	1.08	10	9	39.9	10.6	0	0
8	7	2.9	1	1.9	0	0.60	0.27	8	6	2.9	0.1	1.9	0
9	10	31.5	48.6	39.9	10.6	0.91	1.61	9	10	31.5	48.6	39.9	10.6
5	5	0	0	2.9	1	-0.11	0.02	5	5	0	0	2.9	0.1
4	6	0	0.1	0	0.1	-0.35	0.28	4	7	0	1.0	0	1.0
6	4	0	0	31.5	48.6	0.11	-0.21	6	4	0	0	31.5	48.6
		0.663		0.663		0.624				0.597		0.597	

is greater than the input value of 0.605 in $[C_{obs}]$ (Table 1). Obviously, the deviation is caused by the data transformation in Step 1 because Score 1 had a correlation coefficient of 0.357, which should be close to zero. The error further causes greater correlation coefficients of 0.651 for Score 2 and of 0.663 for $[X1]$. Therefore, the input correlation matrix $[C_{obs}]$ in Step 2 should be adjusted to a smaller value. After several trials, the correlation coefficient in $[C_{obs}]$ can be adjusted to 0.596 to generate Score 3 with a correlation coefficient of 0.624 and a resultant value of 0.597 for $[X2]$.

Similar problems exist in all the datasets (Figure 3). Taking March as an example, the intersite and intervariable correlations from the SSWG (Figure 3(a)) are substantially improved by the above three steps (Figure 3(b)), and the intersite correlation for T_{max} or T_{min} and the intervariable correlation for T_{max} & T_{min} are almost the same as the observations. However, the intersite correlation for P & P and the intervariable correlation for P & T_{max} and P & T_{min} are different. The previous multi-site TSWG fixed this problem by adjusting the input correlation matrices of precipitation $[C_{obs}]$ in Step 2 according to the strong linear relationships between the simulated and observed correlations. The equations developed between the observed and generated correlation coefficients for P & P , P & T_{max} , and P & T_{min} are $[C_{gen}] = 1.147 [C_{obs}] - 0.2906 (R^2 = 0.98)$, $[C_{gen}] =$

$0.861 [C_{obs}] + 0.042 (R^2 = 0.92)$, and $[C_{gen}] = 0.6661 [C_{obs}] - 0.002 (R^2 = 0.96)$, respectively. In this multisite and multivariate version, the intervariable correlation of P & T_{max} and P & T_{min} are also adjusted. It should be noted that the adjustment of the observed correlation matrices is performed separately for each pair of variables. The adjustment works well and the intersite and intervariable correlations are reproduced almost perfectly (Figure 3(c)).

Step 5. Adjustment of precipitation occurrence structure. The above four steps recorelate the input variables well; however, the precipitation occurrence structure is perturbed because the wet events are rearranged (Figure 4). To reduce the consequences of possible adverse effects on occurrence structure, the occurrence of the shuffled data $[X^*]$ can be adjusted according to that of the SSWG in $[X]$. For each month, one control station out of all the stations can be selected if it has the greatest mean cross-correlation coefficients. Then, the occurrence auto-correlation of $[X^*]$ is modified by reordering the rows so that occurrences for the control station in $[X^*]$ (e.g., $[X2]$ in Table 1) closely match those of the control station in $[X]$. That is, the occurrences of the other stations, as well as the temperatures, are rearranged simultaneously based on occurrences for the control station.

To explain the adjustment method more clearly, a larger

example dataset covering 10 days for 5 gauge stations is presented in Table 2. Let $G4$ be the control station, the desired occurrence for which can be obtained from $[X]$. Specifically, $G4$ has four wet days in the first column of Table 2, e.g., the 1st, 4th, 8th, and 9th days; however, the four wet days in the ‘occurrence after shuffle’ column are placed in the order of 3rd, 4th, 7th, and 9th. To match the two occurrence series, the 3rd and 7th lines are moved to the 1st and 8th lines, respectively. After the occurrence adjustment, the serial lag-1 correlation of TSWG-generated precipitation is very similar to that of the SSWG and better than without occurrence adjustment; the adjustment has little impact on the serial lag-1 correlation of temperature (Figure 4). These results imply that this step successfully reconstructs the occurrence structure. Returning to the dataset in Table 1 and using $G1$ in $[X]$ as a control station, the shuffled data $[X1]$ and $[X2]$ can be rearranged as $[X1\text{-occ}]$ and $[X2\text{-occ}]$, respectively. The correlation coefficient of the shuffled data $[X2]$ is the same as $[X2\text{-occ}]$, which implies that the occurrence adjustment has no impact on the intersite and intervariable correlations.

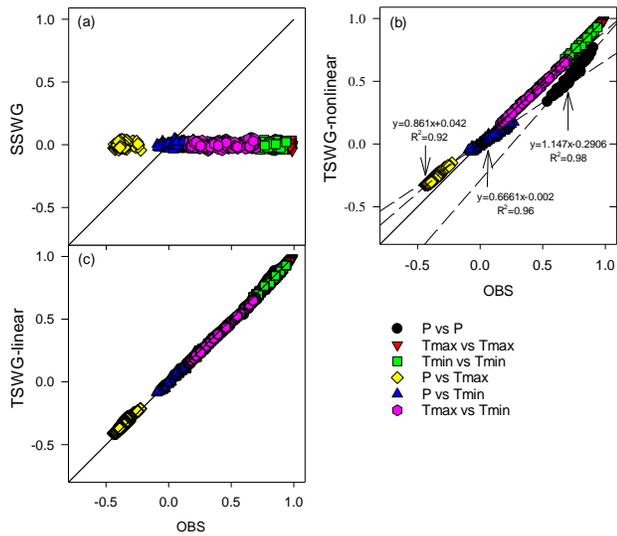


Figure 3. Intersite and intervariable correlations of daily precipitation (P) and temperature (T_{max} and T_{min}) for the observations (OBS) versus (a) SSWG-generated data, (b) TSWG-generated data without (TSWG-raw) and (c) with (TSWG-linear) linear adjustment of the input correlation matrices.

3. Model Evaluation

3.1. Evaluation Procedure

Direct and indirect evaluations were undertaken to assess the performance of the TSWG. The direct evaluation compared the observed versus simulated statistics related to precipitation and temperature, such as the averages and standard deviations of P , T_{max} , and T_{min} , joint probability and continuity ratios of precipitation occurrence and amount, and daily temperature range.

The indirect evaluation used the observed and generated climate data as inputs to the SWAT model to compare their

effects on streamflow, such as the averages, standard deviations, and maxima and minima of streamflow. Streamflow for the periods 1961 ~ 1965 and 1966 ~ 1970 were used for model calibration and validation, respectively. To minimize the effects of other climatic factors on streamflow, only precipitation and temperature were taken from the historical/generated data, whereas other variables (e.g., humidity, wind speed, and solar radiation) were obtained from one station generated by the SWAT. The Nash-Sutcliffe model efficiency coefficients of calibration and validation period were both 0.78 for monthly streamflow, implying the streamflow could be modeled satisfactorily by the SWAT. To avoid biases from the SWAT, the observed discharge during 1961 ~ 2001, used hereafter, reflects modeled data instead of actual observed data.

To explain the performance of the TSWG better, a counterpart model, e.g., the multisite chain-dependent model proposed by Wilks (1999), is introduced for context. Details of the algorithms of the chain-dependent model used in this study can be found in Li et al (2017). The precipitation occurrence and amounts were generated by the first-order Markov chain and the mixed exponential distribution, respectively, and the temperature was generated using a standard normal distribution with consideration of the dependence on the dry/wet status. For precipitation amount, one method used the untapered scale parameters to assign the smaller/greater of the two mixed exponential scale parameters directly to non-zero amounts with smaller/greater expected values, whereas the other method used tapered scale parameters to generate large precipitation amounts by continuously varying the larger of the two scale parameters. However, the precipitation amounts generated by the tapered scale parameters overestimated the averages, extremes, and variability of streamflow (Li et al., 2017); therefore, the precipitation amounts generated using the untapered scale parameters were used in this study.

3.2. Direct Evaluation

3.2.1. Averages and Standard Deviations

The averages and standard deviations of the observed daily P , T_{max} , and T_{min} are reproduced well by the TSWG since the slopes and determination coefficients (R^2) between the observed and generated values are close to ones while the root mean square errors and intercept are very small or equal to zeros (Table 3). The TSWG-generated statistics are actually the same as the SSWG because they are generated in the first stage of the operation and retained during the shuffle procedure in the second stage. Furthermore, no difference is detected in the abilities of the SSWG, TSWG, and chain-dependent model to reproduce the averages and standard deviations of the daily climate (Table 3).

The statistics of the watershed-averaged daily precipitation and temperature are used to assess the spatial structure of the climate. The averages and standard deviations of the SSWG-generated daily precipitation (Figures 5(a) and 5(d)) and the standard deviations of the SSWG-generated daily temperature (Figures 5(e) and 5(f)) are all smaller than the observations; however, all the statistics calculated from the TSWG-generated daily

Table 2. Example of Occurrence Structure Adjustment

WG occurrence	Occurrence after shuffle						Adjusted occurrence					
	Old order	G1	G2	G3	G4	G5	New order	G1	G2	G3	G4	G5
1	1	1	1	1	0	0	3	1	1	1	1	1
0	2	0	0	0	0	0	1	1	1	1	0	0
0	3	1	1	1	1	1	2	0	0	0	0	0
1	4	1	1	1	1	1	4	1	1	1	1	1
0	5	0	0	0	0	0	5	0	0	0	0	0
0	6	0	0	0	0	0	6	0	0	0	0	0
0	7	1	0	0	1	0	8	0	0	0	0	0
1	8	0	0	0	0	0	7	1	0	0	1	0
1	9	1	1	1	1	1	9	1	1	1	1	1
0	10	0	0	0	0	0	10	0	0	0	0	0

Note: Suppose G4 is the control station. The occurrence after the shuffle is rearranged according to that of the control station from the SSWG. 0 and 1 represent dry and wet status, respectively. WG occurrence refers to the occurrence from single-site weather generator (SSWG); control refers to the data from control station.

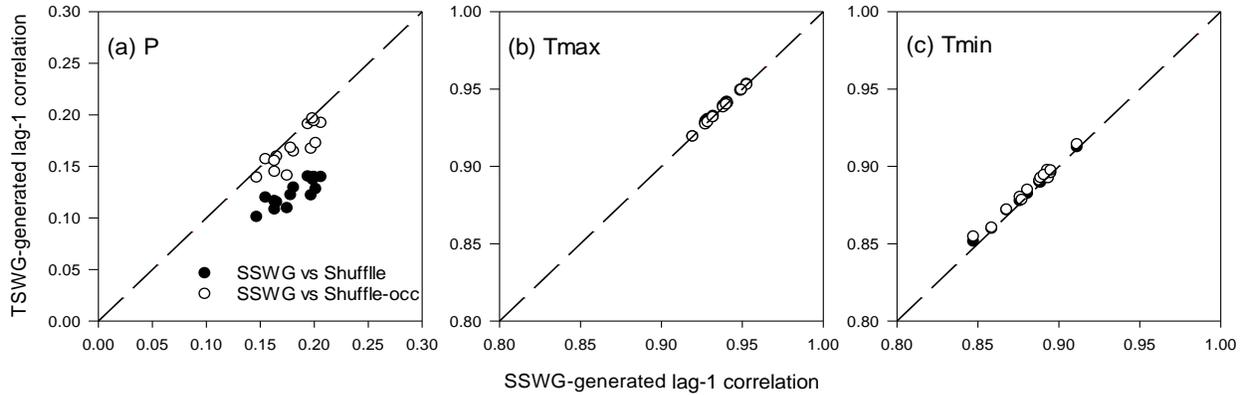


Figure 4. SSWG-generated versus TSWG-generated serial lag-1 correlation for (a) daily precipitation P , (b) daily maximum temperature T_{max} , and (c) daily minimum temperature T_{min} . Shuffle and Shuffle-occ mean the data are shuffled with or without occurrence adjustment, respectively.

Table 3. Example Datasets for the Shuffle Procedure

Variables	Weather generator	Average				Standard deviation			
		Slope	Intercept	R^2	RMSE	Slope	Intercept	R^2	RMSE
P	SSWG	0.9961	0	0.9891	0.2681	1.0156	0	0.9833	0.5066
	TSWG	0.9946	0	0.9890	0.2690	1.0163	0	0.9833	0.5067
	Wilks	0.9896	0	0.9988	0.0947	0.9866	0	0.9973	0.2166
T_{max}	SSWG	0.9999	0	0.9999	0.0827	1.0018	0	0.9921	0.0544
	TSWG	0.9992	0	0.9999	0.0815	1.0098	0	0.9917	0.0566
	Wilks	0.9955	0	0.9998	0.1232	0.9911	0	0.9936	0.0508
T_{min}	SSWG	1.0003	0	0.9999	0.0651	0.9985	0	0.9933	0.0446
	TSWG	1.0005	0	0.9999	0.0649	1.0035	0	0.9920	0.0490
	Wilks	0.9992	0	0.9999	0.0822	0.9948	0	0.9949	0.0392

Note: Desired correlation coefficient is 0.605. X is the SSWG-generated daily precipitation for stations G1 and G2. The description of the other variables can be found in Section 2.2. SSWG refers to single-site weather generator; TSWG refers to two-stage weather generator; Wilks refers to multisite chain-dependent process proposed by Wilks (1999).

weather data match the observations well. The sharp differences indicate that the TSWG performs well when incorporating multisite correlations in large catchments. In addition, the performance of the TSWG is similar to or better than that of the chain-dependent model.

3.2.2. Precipitation Occurrence

The precipitation occurrence structure is perturbed in the second stage; however, the serial lag-1 correlation is reconstructed well (Figure 4), which implies that the TSWG performs satisfactorily in simulating the temporal dependence between

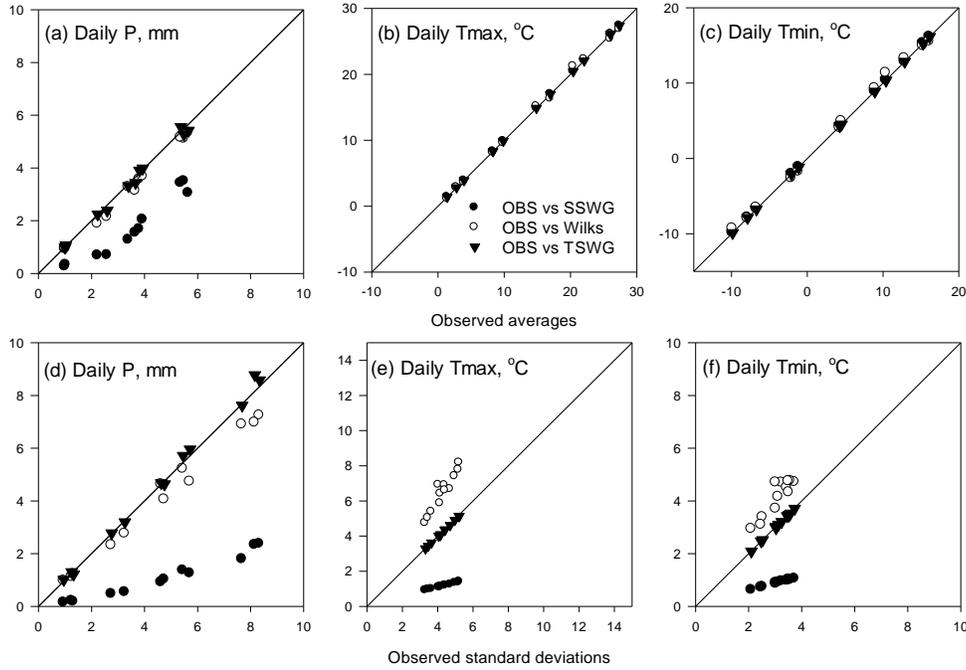


Figure 5. Averages and standard deviations of the observed watershed-averaged versus SSWG-generated, Wilks-generated, and TSWG-generated daily precipitation and temperature.

Table 4. Statistical Parameters of the Bbserved and Simulated Streamflows

	Statistics	OBS	SSWG	TSWG	Wilks
Annual streamflow	Average, mm	39.3	39.5	38.3	40.2
	STD, mm	13.9	6.9	14.8	12.2
	Maxima, mm	76.4	55.0	82.0	75.9
	Minima, mm	20.2	24.6	13.1	14.9
Monthly streamflow	Average, mm	3.28	3.29	3.19	3.35
	STD, mm	4.10	2.98	3.75	3.80
	Maxima, mm	28.02	15.54	22.76	25.40
	Minima, mm	0.02	0.04	0.04	0.04

Note: OBS refers to observation; SSWG refers to single-site weather generator; TSWG refers to two-stage weather generator; Wilks refers to multisite chain-dependent process proposed by Wilks (1999).

different dry and/or wet events. To evaluate further the capability of the TSWG to reproduce the characteristics of precipitation occurrence, the multisite correlation of occurrence, joint probability, and continuity ratios are compared with the observations. As shown in Figure 6, the multisite correlation of precipitation occurrence is simulated well, and better than by the chain-dependent model. The above results suggest that the shuffle procedure can effectively reconstruct the simultaneous occurrences of precipitation within a large watershed.

The joint probabilities that station pairs are both dry or both wet on a given day are compared between the observed and SSWG-generated or TSWG-generated values for all months. Obviously, the SSWG cannot reproduce the observed joint probabilities because multisite correlation is not taken into account (Figures 7(a) and 7(d)). However, the shuffled SSWG-generated data reconstruct the joint probabilities almost perfectly (Figures 7(c) and 7(f)), similar to those from the chain-

dependent model (Figures 7(b) and 7(e)), suggesting that the spatiotemporal coherence is reconstructed well by the TSWG.

Continuity ratios are used to quantify the links between precipitation occurrence and amount. For each station pair, continuity ratio is the ratio of the precipitation mean at station k depending on whether station 1 is wet or dry. The ratio will be small for highly correlated stations (where there is a link between occurrences and amounts), and large for stations that are not correlated (no link between occurrences and amounts). As the occurrence and amount processes are generated independently at each station by the SSWG, the continuity ratios are close to one for all station pairs (Figure 8(a)). However, the more widely separated station pairs of the continuity ratios of the TSWG-generated precipitation indicated an effective reproduction of the dependences between the occurrence and the amount processes (Figure 8(c)). The continuity ratios for the TSWG are better than for the untapered chain-dependent model

(Figure 8(b)). Although they are similar to the TSWG and the tapered chain-dependent model (not used in this study), the precipitation from the tapered chain-dependent model overestimates streamflow (Li et al., 2017). Therefore, the TSWG is considered to outperform the chain-dependent model in reproducing the link between precipitation occurrence and amount.

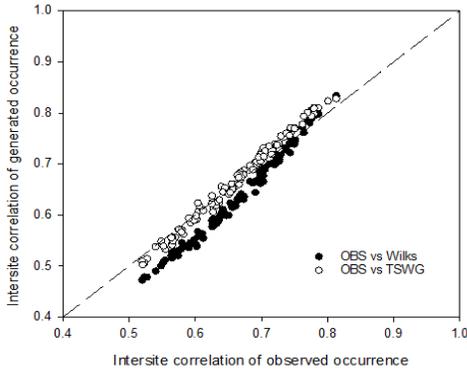


Figure 6. Intersite correlation of daily precipitation occurrence.

As daily temperature is generated stochastically by SSWG without a range check in this study, the SSWG-generated T_{max} is smaller than T_{min} for some days. The days with a negative daily temperature range ($T_{max} - T_{min} < 0$) account for 1.7% ~ 4.7% of the total data length for the 15 stations. However, no negative daily temperature ranges are detected after the shuffle, which indicates that the TSWG reproduces the intervariable correlation between T_{max} and T_{min} well.

3.2.4. Multisite and Multivariate Correlation

Although the multisite and multivariate correlations for four months are presented in the section of model development (Figure 3), the correlations for all months and all stations are presented in Figure 9 to show the overall performance of the TSWG. Similar to its performance for the four months, the TSWG can satisfactorily simulate either the multisite correlation or the multivariate correlation. The performance of the TSWG is better than that of the chain-dependent model, especially for the multivariate correlation, because this is not taken into account by the chain-dependent model.

3.2.3. Daily Temperature Range

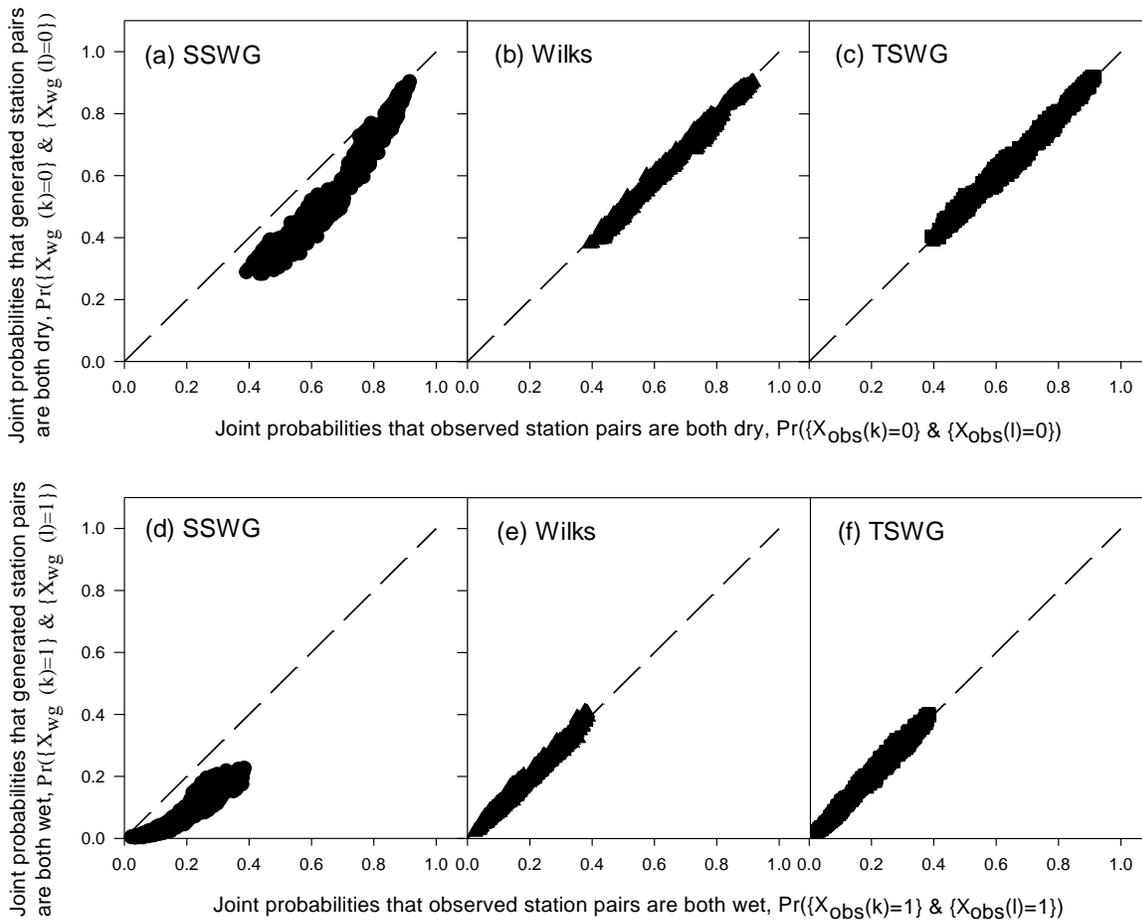


Figure 7. Joint probabilities that station pairs are both dry or both wet on a given day between the observed (X_{obs}) and the generated (X_{wg}) series from (a) and (d) SSWG, (b) and (e) Wilks, or (c) and (f) TSWG.

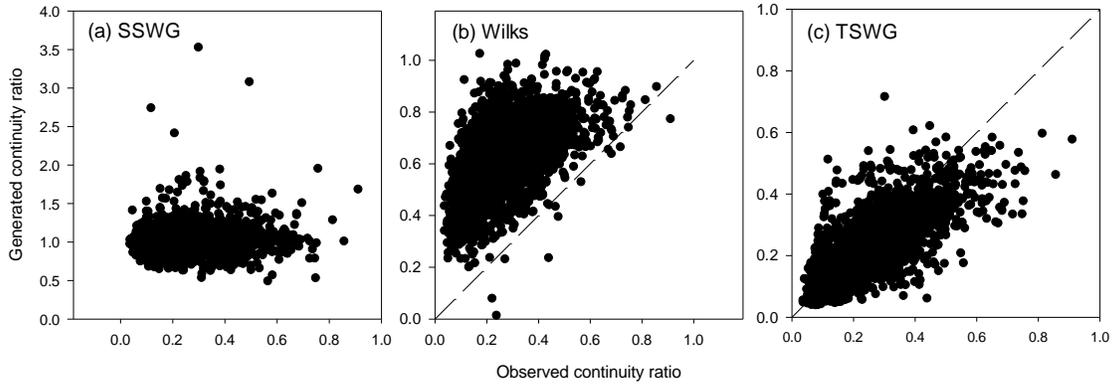


Figure 8. Observed continuity ratios versus (a) SSWG-generated, (b) Wilks-generated, and (c) TSWG-generated continuity ratios.

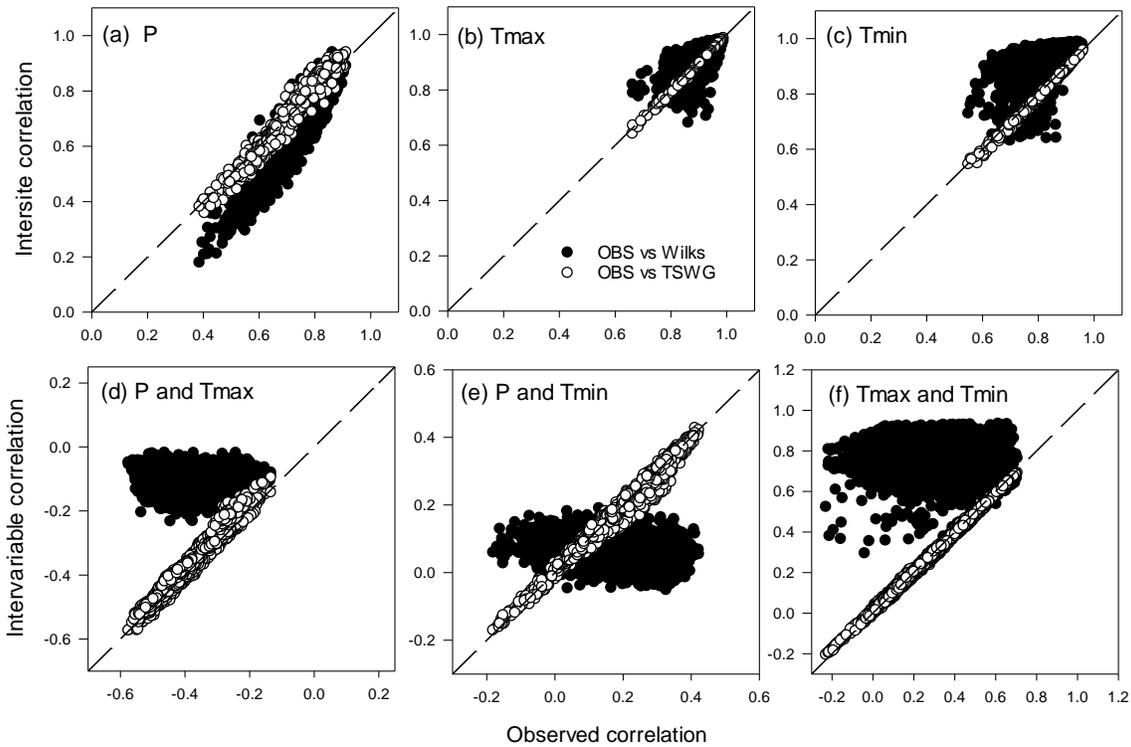


Figure 9. Observed versus Wilks-generated and TSWG-generated intersite and intervariable correlations.

3.3. Indirect Evaluation by Hydrological Modeling

The mean monthly and annual streamflows from the SSWG, chain-dependent model, and the TSWG are very similar to the observations (Table 4), which implies that the multisite and multivariate correlations of climate have little effect on the averages of streamflow. However, the maxima and standard deviations of streamflow are greatly underestimated by the SSWG, while those from the chain-dependent model and the TSWG are very close to the observations. The above results suggest that the TSWG-generated climate can be used satisfactorily for the modeling of hydrological variability, and that its performance is comparable with that of the untapered chain-dependent model.

To further validate the performance of the TSWG in the

modeling of the hydrological regime, the statistics of monthly streamflow are presented in Figure 10. Obviously, the chain-dependent model and the TSWG perform much better than the SSWG in reproducing the standard deviations and the maxima and minima of streamflow. Therefore, it is concluded that the TSWG is a powerful tool that could be used to provide climate inputs for the modeling of hydrological variability.

4. Summary

Similar to Wilks (1998, 1999), a semiparametric TSWG, originally designed for multisite precipitation generation, was extended in this study to incorporate daily maximum and mini-

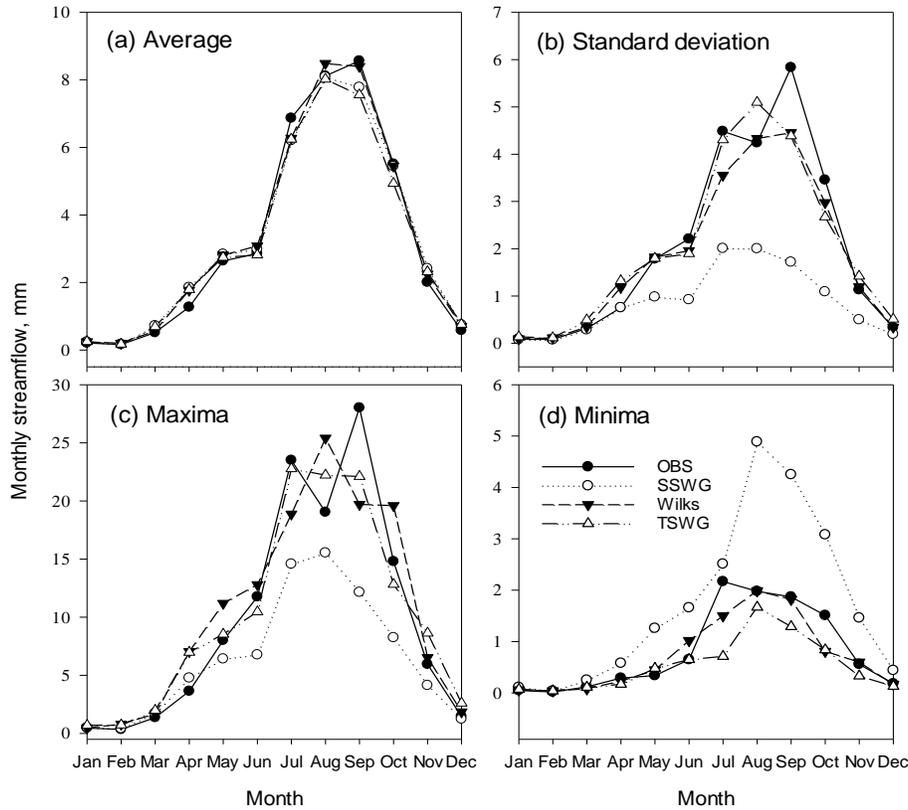


Figure 10. Averages, standard deviations, maxima and minima of observed, SSWG-generated, Wilks-generated, and TSWG-generated monthly streamflow.

imum temperatures. The TSWG generates single-site and single-variable data in the first stage, and it then reconstructs the multi-site and multivariate correlations in the second stage. The extension of the TSWG is straightforward and involves using a larger data matrix in the first stage and then adjusting the input correlation matrix for different pairs of variables in the second stage. Although the matrix size is larger than that in the previous multi-site version, the model efficiency is unaffected according to a comparison of the running times.

By extending a single-site chain-dependent process into a multisite model, the TSWG and the multisite chain-dependent model (Wilks, 1999) are very similar in some respects. For the step of weather generation, the mixed exponential distribution is used in the chain-dependent model to obtain better multisite correlation of precipitation through alternately choosing one of the two mixed exponential scale parameters. In this study, the TSWG is applicable to any probability distribution because the correlation reconstruction does not depend on the method for precipitation simulation. Therefore, either a parametric method such as the Richardson-type approach (Richardson, 1981), the semi-empirical method used in LARS-WG (Semenov and Barrow, 1997), or the circulation-based method used in SDSM (Wilby et al., 2002) could be used for the weather generation of the TSWG. This advantage presents more options for precipitation and/or temperature simulation in weather generation. For exam-

ple, the skewed normal distribution used in this study is found to be the most appropriate model for the simulation of precipitation amount on the Loess Plateau of China (Li et al., 2014).

For the step of correlation construction, the multisite chain-dependent model recorrelates the random numbers through developing empirical curves for all station pairs, each precipitation process or variable, and each month. Using a distribution-free method, the TSWG simultaneously reconstructs the multi-site and multivariate correlations through one shuffle for each month, but for all sites and all variables despite their probability distributions. The TSWG outperforms the chain-dependent model for reconstruction of the intervariable correlation because the latter method does not consider this aspect. Although it appears that the multivariate correlation has little impact on hydrological modeling in the Jing River catchment, it is of great importance for some other applications. In particular, the correlation between precipitation and temperature greatly influences the snowmelt runoff in cold regions, which has been detected in our previous study for two catchments in Canada (Li et al., 2013). Similarly, the intervariable correlation between precipitation and evapotranspiration derived from temperature is essential to calculate irrigation requirements. As a postprocessing method, the TSWG affects the occurrence structure, which is an inherent weakness of the post-processing method (Wilks, 1998; Clark et al., 2004; Bárdossy and Pegram, 2012). How-

ever, the occurrence adjustment according to the control station from the SSWG effectively recovers the related statistics of the observations (Figures 6 ~ 8). The indirect evaluation showed that the TSWG-generated climate is acceptable for the simulation of hydrological variability, at least on monthly and annual scales. Therefore, the TSWG is considered a useful tool for multisite and multivariate weather generation.

Acknowledgments. This study was jointly funded by the National Natural Science Foundation of China (U1703124 and 41761144060) and the International Partnership Program of Chinese Academy of Sciences (161461KYSB20170013).

References

- Angulo, J.M., Gonzalez-Manteiga, W., Febrero-Bande, M., and Alonso, F.J. (1998). Semi-parametric statistical approaches for space-time process prediction, *Environ. Ecol. Stat.*, 5(4), 297-316. <http://dx.doi.org/10.1023/A:1009670920927>
- Apipattanasri, S., Podestá, G., Rajagopalan, B., and Katz, R.W. (2007). A semiparametric multivariate and multisite weather generator, *Water Resour. Res.*, 43(11), W11401. <http://dx.doi.org/10.1029/2006WR005714>
- Arnold, J.G., Srinivasan, R., Muttiah, R.S., and Williams, J.R. (1998). Large area hydrologic modeling and assessment - Part 1: Model development, *J. Am. Water Resour. As.*, 34(1), 73-89. <http://dx.doi.org/DOI10.1111/j.1752-1688.1998.tb05961.x>
- Asong, Z.E., Khaliq, M.N., and Wheeler, H.S. (2016). Multisite multivariate modeling of daily precipitation and temperature in the Canadian Prairie Provinces using generalized linear models, *Clim. Dyn.*, 47(9-10), 2901-2921. <http://dx.doi.org/10.1007/s00382-016-3004-z>
- Bárdossy, A. and Pegram, G. (2012). Multiscale spatial recorelation of RCM precipitation to produce unbiased climate change scenarios over large areas and small, *Water Resour. Res.*, 48(9), W09502. <http://dx.doi.org/10.1029/2011WR011524>
- Bárdossy, A. and Plate, E.J. (1992). Space-time model for daily rainfall using atmospheric circulation patterns, *Water Resour. Res.*, 28(5), 1247-1259. <http://dx.doi.org/10.1029/91WR02589>
- Baigorria, G.A. and Jones, J.W. (2010). GiST: A Stochastic Model for Generating Spatially and Temporally Correlated Daily Rainfall Data, *J. Climate.*, 23(22), 5990-6008. <http://dx.doi.org/10.1175/2010jcli3537.1>
- Beersma, J.J. and Buishand, T.A. (2003). Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation, *Clim. Res.*, 25(2), 121-133. <http://dx.doi.org/10.3354/cr.025121>
- Boorman, D.B. and Sefton, C.E.M. (1997). Recognising the uncertainty in the quantification of the effects of climate change on hydrological response, *Clim. Change*, 35(4), 415-434. <http://dx.doi.org/10.1023/A:1005372407881>
- Brissette, F.P., Khalili, M., and Leconte, R. (2007). Efficient stochastic generation of multi-site synthetic precipitation data, *J. Hydrol.*, 345(3-4), 121-133. <http://dx.doi.org/10.1016/j.jhydrol.2007.06.035>
- Caraway, N.M., McCreight, J.L., and Rajagopalan, B. (2014). Multisite stochastic weather generation using cluster analysis and k-nearest neighbor time series resampling, *J. Hydrol.*, 508, 197-213. <http://dx.doi.org/10.1016/j.jhydrol.2013.10.054>
- Chandler, R.E. and Wheeler, H.S. (2002). Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland, *Water Resour. Res.*, 38(10), 10-1-10-11. <http://dx.doi.org/10.1029/2001WR000906>
- Chen, J., Brissette, F.P., and Zhang, X.J. (2016). Hydrological Modeling Using a Multisite Stochastic Weather Generator, *J. Hydrol. Eng.*, 21(2), 04015060. [http://dx.doi.org/10.1061/\(asce\)he.1943-5584.001288](http://dx.doi.org/10.1061/(asce)he.1943-5584.001288)
- Chen, J., Zhang, X.C., Liu, W.Z., and Li, Z. (2009). Evaluating and Extending CLIGEN precipitation Generation for the Loess Plateau of China, *J. Am. Water Resour. As.*, 45(1), 1-19. <http://dx.doi.org/10.1111/j.1752-1688.2008.00296.x>
- Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., and Wilby, R. (2004). The Schaake Shuffle: A Method for Reconstructing Space-Time Variability in Forecasted Precipitation and Temperature Fields, *J. Hydrometeorol.*, 5(1), 243-262. [http://dx.doi.org/10.1175/1525-7541\(2004\)005<0243:TSSAMF>2.0.CO;2](http://dx.doi.org/10.1175/1525-7541(2004)005<0243:TSSAMF>2.0.CO;2)
- Fowler, H.J., Kilsby, C.G., O'Connell, P.E., and Burton, A. (2005). A weather-type conditioned multi-site stochastic rainfall model for the generation of scenarios of climatic variability and change, *J. Hydrol.*, 308(1-4), 50-66. <http://dx.doi.org/10.1016/j.jhydrol.2004.10.021>
- Furrer, E.M. and Katz, R.W. (2007). Generalized linear modeling approach to stochastic weather generators, *Clim. Res.*, 34(2), 129-144. <http://dx.doi.org/10.3354/cr034129>
- Iman, R.L. and Conover, W.J. (1982). A distribution-free approach to inducing rank correlation among input variables, *Commun. Stat-Simul. C.*, 11(3), 311-334
- Khalili, M., Brissette, F., and Leconte, R. (2011). Effectiveness of Multi-Site Weather Generator for Hydrological Modeling, *J. Am. Water Resour. As.*, 47(2), 303-314. <http://dx.doi.org/10.1111/j.1752-1688.2010.00514.x>
- King, L.M., McLeod, A.I., and Simonovic, S.P. (2014). Simulation of historical temperatures using a multi-site, multivariate block resampling algorithm with perturbation, *Hydrol. Process.*, 28(3), 905-912. <http://dx.doi.org/10.1002/hyp.9596>
- Li, Z. (2014). A new framework for multi-site weather generator: a two-stage model combining a parametric method with a distribution-free shuffle procedure, *Clim. Dyn.*, 43(3-4), 657-669. <http://dx.doi.org/10.1007/s00382-013-1979-2>
- Li, Z., Brissette, F., and Chen, J. (2013). Finding the most appropriate precipitation probability distribution for stochastic weather generation and hydrological modelling in Nordic watersheds, *Hydrol. Process.*, 27(25), 3718-3729. <http://dx.doi.org/10.1002/hyp.9499>
- Li, Z., Brissette, F., and Chen, J. (2014). Assessing the applicability of six precipitation probability distribution models on the Loess Plateau of China, *Int. J. Climatol.*, 34(2), 462-471. <http://dx.doi.org/10.1002/joc.3699>
- Li, Z., Lü, Z., Li, J., and Shi, X. (2017). Links between the spatial structure of weather generator and hydrological modeling, *Theor. Appl. Climatol.*, 128(1), 103-111. <http://dx.doi.org/10.1007/s00704-015-1691-8>
- Mehrotra, R. and Sharma, A. (2007). A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability, *J. Hydrol.*, 335(1-2), 180-193. <http://dx.doi.org/10.1016/j.jhydrol.2006.11.011>
- Mehrotra, R. and Sharma, A. (2009). Evaluating spatio-temporal representations in daily rainfall sequences from three stochastic multi-site weather generation approaches, *Adv. Water. Resour.*, 32(6), 948-962. <http://dx.doi.org/10.1016/j.advwatres.2009.03.005>
- Mehrotra, R., Srikanthan, R., and Sharma, A. (2006). A comparison of three stochastic multi-site precipitation occurrence generators, *J. Hydrol.*, 331(1-2), 280-292. <http://dx.doi.org/10.1016/j.jhydrol.2006.05.016>
- Palutikof, J.P., Goodess, C.M., Watkins, S.J., and Holt, T. (2002). Generating rainfall and temperature scenarios at multiple sites: Examples from the Mediterranean, *J. Climate.*, 15(24), 3529-3548. [http://dx.doi.org/10.1175/1520-0442\(2002\)015<3529:Gratsa>2.0.Co;2](http://dx.doi.org/10.1175/1520-0442(2002)015<3529:Gratsa>2.0.Co;2)
- Pegram, G. and Bárdossy, A. (2013). Downscaling Regional Circulation Model rainfall to gauge sites using recorelation and circulation pattern dependent quantile-quantile transforms for quantifying

- ing climate change, *J. Hydrol.*, 504, 142-159. <http://dx.doi.org/10.1016/j.jhydrol.2013.09.014>
- Qian, B.D., Corte-Real, J., and Xu, H. (2002). Multisite stochastic weather models for impact studies, *Int. J. Climatol.*, 22(11), 1377-1397. <http://dx.doi.org/10.1002/joc.808>
- Racsko, P., Szeidl, L., and Semenov, M. (1991). A serial approach to local stochastic weather models, *Ecol. Model.*, 57(1), 27-41. [http://dx.doi.org/10.1016/0304-3800\(91\)90053-4](http://dx.doi.org/10.1016/0304-3800(91)90053-4)
- Rebonato, R. and Jäckel, P. (2000). The most general methodology to create valid correlation matrix for risk management and option pricing purposes, *J. Risk*, 2, 17-28. <http://dx.doi.org/10.1.1.16.49>
- Rex, J.F. (1993). The dynamics of uncertainty: application to parameterization constants in climate models, *Clim. Dynam.*, 8(3), 135-150. <http://dx.doi.org/10.1007/BF00208093>
- Richardson, C.W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation, *Water Resour. Res.*, 17(1), 182-190. <http://dx.doi.org/10.1029/WR017i001p00182>
- Richardson, C.W. and Nicks, A.D. (1990). *Weather generator description Rep*, Technical Bulletin, USDA ARS, pp: 235
- Semenov, M.A. and Barrow, E.M. (1997). Use of a stochastic weather generator in the development of climate change scenarios, *Clim. Change*, 35(4), 397-414. <http://dx.doi.org/10.1023/a:1005342632279>
- Srikanthan, R. and Pegram, G.G.S. (2009). A nested multisite daily rainfall stochastic generation model, *J. Hydrol.*, 371(1-4), 142-153. <http://dx.doi.org/10.1016/j.jhydrol.2009.03.025>
- Srivastav, R. and Simonovic, S. (2015). Multi-site, multivariate weather generator using maximum entropy bootstrap, *Clim. Dyn.*, 44(11-12), 3431-3448. <http://dx.doi.org/10.1007/s00382-014-2157-x>
- Steinschneider, S. and Brown, C. (2013). A semiparametric multivariate, multisite weather generator with low-frequency variability for use in climate risk assessments, *Water Resour. Res.*, 49(11), 7205-7220. <http://dx.doi.org/10.1002/wrcr.20528>
- Tarpanelli, A., Franchini, M., Brocca, L., Camici, S., Melone, F., and Moramarco, T. (2012). A simple approach for stochastic generation of spatial rainfall patterns, *J. Hydrol.*, 472-473, 63-76. <http://dx.doi.org/10.1016/j.jhydrol.2012.09.010>
- Thyer, M. and Kuczera, G. (2003a). A hidden Markov model for modelling long-term persistence in multi-site rainfall time series 1. Model calibration using a Bayesian approach, *J. Hydrol.*, 275(1-2), 12-26. [http://dx.doi.org/10.1016/S0022-1694\(02\)00412-2](http://dx.doi.org/10.1016/S0022-1694(02)00412-2)
- Thyer, M. and Kuczera, G. (2003b). A hidden Markov model for modelling long-term persistence in multi-site rainfall time series. 2. Real data analysis, *J. Hydrol.*, 275(1-2), 27-48. [http://dx.doi.org/10.1016/S0022-1694\(02\)00411-0](http://dx.doi.org/10.1016/S0022-1694(02)00411-0)
- Verdin, A., Rajagopalan, B., Kleiber, W., and Katz, R.W. (2015). Coupled stochastic weather generation using spatial and generalized linear models, *Stoch. Res. Risk A.*, 29(2), 347-356. <http://dx.doi.org/10.1007/s00477-014-0911-6>
- Wilby, R.L., Dawson, C.W., and Barrow, E.M. (2002). SDSM - a decision support tool for the assessment of regional climate change impacts, *Environ. Modell. Softw.*, 17(2), 145-157. [http://dx.doi.org/10.1016/S1364-8152\(01\)00060-3](http://dx.doi.org/10.1016/S1364-8152(01)00060-3)
- Wilby, R.L. and Harris, I. (2006). A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK, *Water Resour. Res.*, 42(2), W02419. <http://dx.doi.org/10.1029/2005WR004065>
- Wilby, R.L., Tomlinson, O.J., and Dawson, C.W. (2003). Multi-sitesimulation of precipitation by conditional resampling, *Clim. Res.*, 23(3), 183-194. <http://dx.doi.org/10.3354/cr023183>
- Wilks, D.S. (1998). Multisite generalization of a daily stochastic precipitation generation model, *J. Hydrol.*, 210(1-4), 178-191. [http://dx.doi.org/10.1016/S0022-1694\(98\)00186-3](http://dx.doi.org/10.1016/S0022-1694(98)00186-3)
- Wilks, D.S. (1999). Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain, *Agr. Forest Meteorol.*, 96(1-3), 85-101. [http://dx.doi.org/10.1016/S0168-1923\(99\)00037-4](http://dx.doi.org/10.1016/S0168-1923(99)00037-4)
- Wilks, D.S. and Wilby, R.L. (1999). The weather generation game: a review of stochastic weather models, *Prog. Phys. Geog.*, 23(3), 329-357. <http://dx.doi.org/10.1177/030913339902300302>
- Yang, C., Chandler, R.E., Isham, V.S., and Wheeler, H.S. (2005). Spatial-temporal rainfall simulation using generalized linear models, *Water Resour. Res.*, 41(11), W11415. <http://dx.doi.org/10.1029/2004WR003739>
- Zhang, X.C. (2005). Generating correlative storm variables for CLIGEN using a distribution-free approach, *T. Asabe.*, 48(2), 567-575. <http://dx.doi.org/10.13031/2013.18331>
- Zhang, X.C., Liu, W.Z., Li, Z., and Chen, J. (2011). Trend and uncertainty analysis of simulated climate change impacts with multiple GCMs and emission scenarios, *Agr. Forest Meteorol.*, 151(10), 1297-1304. <http://dx.doi.org/10.1016/j.agrformet.2011.05.010>