INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **34**: 3089–3105 (2014)
Published online 2 January 2014 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/joc.3896

RMetS
Royal Meteorological Society

Comparison of five stochastic weather generators in simulating daily precipitation and temperature for the Loess Plateau of China

Jie Chen[a,b]* and François P. Brissette[b]

[a] *State Key Laboratory of Soil Erosion and Dryland Farming on Loess Plateau, Institute of Soil and Water Conservation, Northwest A&F University, Yangling, Shaanxi, China*

[b] *Department of Construction Engineering, École de technologie supérieure, Université du Québec, Montreal, QC, Canada*

ABSTRACT: Stochastic weather generators are computer models used to simulate synthetic weather time series based on statistical characteristics of observed weather time series for a location of interest. The performance of weather generators should always be evaluated when applied over a new region, especially in a different climate zone. The Loess Plateau of China suffers from extensive soil erosion problems, mostly resulting from extreme precipitation events in an otherwise dry climate. The generation of accurate synthetic climate data (including the extremes) is the key to adequately assess the effect of soil and water conservation measures, especially in a changing climate. Accordingly, this work compares the ability of five commonly used stochastic weather generators (WGEN, CLIMGEN, CLIGEN, WeaGETS and LARSWG) to simulate daily precipitation and temperature for the Loess Plateau. The results show that the first-order Markov chain-based model satisfactorily simulates the precipitation occurrence, even though slightly better results can be achieved by using the second- and third-order Markov chain-based models. The semi-empirical distribution-based model shows a slightly better performance than the higher-order Markov chain-based models at generating dry spells, while it gives a worse performance in generating wet spells. The semi-empirical distribution-based LARSWG is consistently better than the parametric distribution-based weather generators at simulating daily precipitation amounts, and the three-parameter distribution-based models are consistently better than two-parameter distribution-based models, especially for generating extreme precipitation events. Overall, the performance of WeaGETS is generally superior to that of the other weather generators; it can be used for simulating daily precipitation, $T_{\max}$ and $T_{\min}$ for the Loess Plateau. While CLIGEN is not consistently a top performer, its reasonable representation of extreme precipitation events makes it suitable for soil erosion studies.

KEY WORDS stochastic weather generator; precipitation; temperature; Markov chain; distribution

Received 3 May 2013; Revised 18 November 2013; Accepted 19 November 2013

1. Introduction

Stochastic weather generators are statistical models used to simulate synthetic weather sequences, which are expected to be statistically similar to the observed counterparts. They are usually used with hydrological and environmental models for water resource and environmental management (e.g. Semenov and Porter, 1995; Mavromatis and Hansen, 2001; Wheater *et al.*, 2005). More often, weather generators have been used as downscaling tools to produce high-resolution climate change projections by linking their parameters to climate model outputs (e.g. Semenov and Barrow, 1997; Wilks, 1992, 1999a; Pruski and Nearing, 2002; Zhang, 2005a, 2005b; Zhang and Liu, 2005; Qian *et al.*, 2005, 2010; Kilsby *et al.*, 2007; Chen *et al.*, 2012a, 2013). Compared to other statistical downscaling methods, such as regression-based approaches and weather typing schemes, the weather generator-based approach has the advantage of producing an ensemble of climate change projections for analysing risk-based environmental impacts.

The generation of precipitation and temperature are the two main components for most stochastic weather generators, especially for climate change impact studies. Two types of models are usually used to generate precipitation occurrence: the Markov chain-based model and the alternative renewal process. The most commonly used first-order Markov chain has been found to be adequate for simulating wet spells for various climates and dry spells for temperate climates (Katz, 1977; Richardson, 1981; Wilks, 1989, 1992; Nicks *et al.*, 1995; Wilks, 1999b; Zhang and Garbrecht, 2003; Chen *et al.*, 2009a). However, higher-order (second- and third-order, for example) or hybrid-order Markov chain-based models (for instance, the first-order Markov dependence for wet

* Correspondence to: J. Chen, Department of Construction Engineering, École de technologie supérieure, Université du Québec, 1100 Notre-Dame Street West, Montreal, QC H3C 1K3, Canada.
E-mail: jie.chen.1@ens.etsmtl.ca

spells and higher-order dependence for dry sequences) are usually necessary for adequately representing long dry spells, especially for dry climates (Wilks, 1999b; Chen *et al.*, 2012b). Instead of simulating precipitation occurrence day by day, the alternating renewal process considers the precipitation occurrence as a sequence of alternating wet and dry spells of varying lengths. Various distributions (e.g. logarithmic series, truncated negative binomial distribution, truncated geometric distribution and semi-empirical distribution) are then used to fit the wet and dry spells independently (Green, 1964; Buishand, 1978; Roldan and Woolhiser, 1982; Semenov and Barrow, 2002).

Given the occurrence of a wet day, a number of models have been proposed to generate daily precipitation amounts. Generally, these models include parametric, semi- and nonparametric distributions. The probability parametric distributions are the type most widely used to simulate daily precipitation amounts. They usually include single distributions, such as exponential (Todorovic and Woolhiser, 1975; Roldan and Woolhiser, 1982), gamma (Ison *et al.*, 1971; Richardson and Wright, 1984) and skewed normal (Nicks and Gander, 1994) distributions, and compound distributions, such as the mixed exponential distribution (Roldan and Woolhiser, 1982; Wilks, 1999b) and hybrid exponential and Pareto distributions (Li *et al.*, 2012). Daily precipitation amounts are also simulated using semi- or nonparametric distributions in several studies (e.g., Semenov *et al.*, 1998; Semenov and Barrow, 2002; Mehrotra and Sharma, 2007a, 2007b).

The generation of maximum and minimum temperatures ($T_{\max}$ and $T_{\min}$) is the other main objective of a stochastic weather generator. Compared to precipitation, the simulation of temperature is much simpler since it is continuous and approximately follows a normal distribution. However, $T_{\max}$ and $T_{\min}$ are conditioned on each other, and both are conditioned on precipitation status in the real world. Thus, the preservation of these correlations is very important for a weather generator.

Following the abovementioned theories, several stochastic weather generators have been developed over the last few decades, such as the Weather GENerator (WGEN) (Richardson, 1981; Richardson and Wright, 1984), the CLIMate GENerator (CLIMGEN) (Stöckle *et al.*, 1999), the CLImate GENerator (CLIGEN) (Nicks and Gander, 1994; Nicks *et al.*, 1995), École de technologie supérieure Weather Generator (WeaGETS) (Chen *et al.*, 2012b) and the Long Ashton Research Station-Weather Generator (LARSWG) (Semenov and Barrow, 2002). They have been widely used to simulate daily weather time series for impact studies (e.g. Semenov and Barrow, 1997; Wilks, 1992, 1999a; Zhang, 2005a, 2005b; Chen *et al.*, 2012a). However, different schemes are used by these weather generators to simulate precipitation and temperatures. Therefore, it is important to evaluate and compare their performances, especially on studies over a new region. A few studies have been conducted (e.g. Johnson *et al.*, 1996; Qian *et al.*, 2004; Semenov *et al.*, 1998) comparing the performance of different weather generators in simulating weather variables. However, these have been limited to the assessment of two weather generators. Other studies (Wilks, 1999b; Chen and Brissette, 2013; Li *et al.*, 2013) have compared a single weather generator component. However, it may be better to compare different weather generators as whole models, since the potential user is usually more interested in the overall performance of a weather generator than in that of a single component. For example, both precipitation and temperature are required as inputs to the Water Erosion Prediction Project (Flanagan and Nearing, 1995) model for predicting hillslope soil erosion. The hydrological response is also due to the joint effect of precipitation and temperature. This is especially true for snow-dominated watersheds, where temperature usually plays a more important role than precipitation for winter hydrology.

The Loess Plateau of China is one of the most severely eroded regions in the world due to its fine aeolian deposits, steep slopes, sparse vegetation cover and frequent heavy storms. The estimated mean annual soil loss rate is 5000–10 000 tonnes km^{-2} year^{-1} (Zhang *et al.*, 2008). Rainfall is heavily concentrated in summer and autumn, accounting for about 80% of annual precipitation. Rainfall amounts could be greater than 200 mm for a single event, even though the mean annual precipitation is often less than 600 mm. The heavy rainfall significantly impacts soil erosion in this region (Chen *et al.*, 2009a). Zhou and Wang (1992) reported that most soil erosion is attributable to a few heavy storms in the summer. Great effort has been taken to combat soil erosion in this region in the past several decades. For example, the 'Loess Plateau Watershed Rehabilitation Project' (http://www.worldbank.org/projects/) was launched in 1994 to mitigate desertification and increase agricultural production for the Loess Plateau. There is a need to generate synthetic climate data, especially for extreme rainfall events, for use in evaluating the effectiveness of soil and water conservation measures to control runoff and soil loss in ungauged areas, since the gauge density is relatively low for the Loess Plateau. In the context of global climate change, the other important issue for the Loess Plateau is assessing the impacts of climate change on soil erosion, hydrology and agriculture. Reliable stochastic weather generators would be very important tools for studying environmental-linked issues in this region. In particular, the simulation of extreme rainfall and temperature events is a key component of a weather generator applied to this region. In order to properly generate synthetic climate data for the Loess Plateau, the first and foremost step is to evaluate the performance of the most commonly used weather generators for this region.

Accordingly, this work compares the performance of five stochastic weather generators (WGEN, CLIMGEN, CLIGEN, WeaGETS and LARSWG) in simulating daily precipitation, $T_{\max}$ and $T_{\min}$ for the Loess Plateau. These five weather generators are widely used in simulating precipitation and temperature time series, and

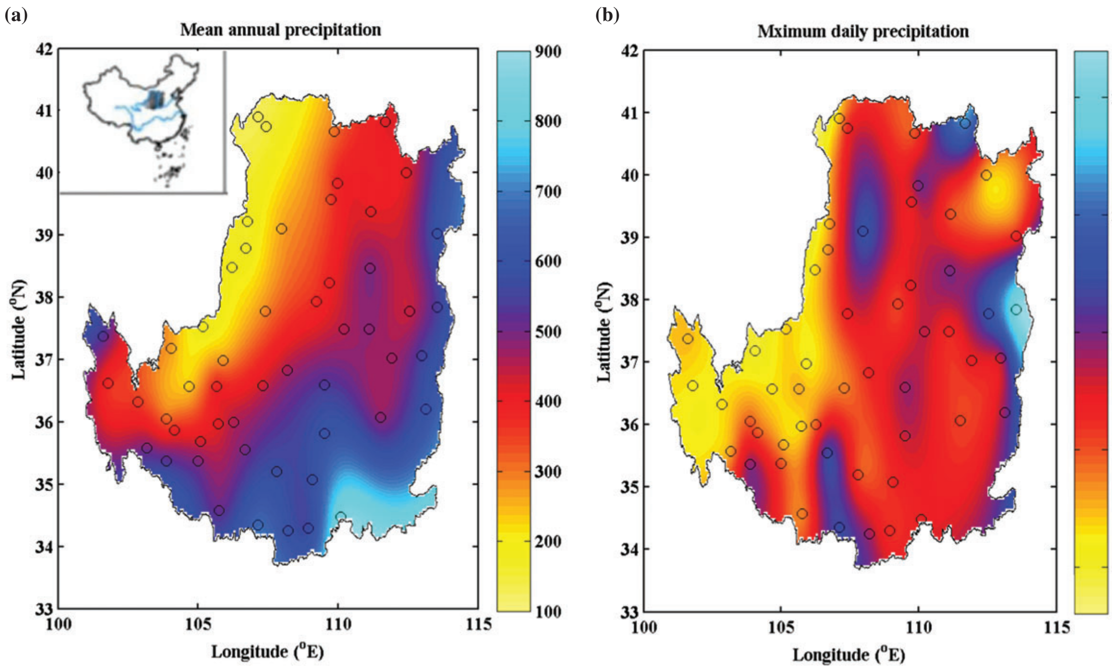

Figure 1. Selected study stations (denoted by ○), (a) mean annual precipitation and (b) maximum daily precipitation for the Loess Plateau.

as downscaling tools for climate change impact studies throughout the world.

2. Study area and dataset

This study was conducted on the Loess Plateau of China (Figure 1). The Loess Plateau covers an area of about 626 800 km^2 in northwest China, and has some of the highest soil erosion rates in the world (Yang and Yu, 1992). The Loess Plateau generally has a semi-arid climate (Kottek *et al.*, 2006), with extensive monsoonal influence. Rainfall tends to be heavily concentrated in the summer, which accounts for about 65% of annual precipitation. Even though the annual precipitation is relatively low, the precipitation intensity is very high. One heavy storm in the summer can contribute up to 30% of the annual total precipitation. Winters are cold and dry, while summers are very warm in many places.

Meteorological data including precipitation, T_{max} and T_{min} from 54 stations (Figure 1) dispersed across the Loess Plateau were used in this study. The data were provided by the China Meteorological Data Sharing Service System (http://cdc.cma.gov.cn/home.do). The climate time series ranges between 32 and 52 years, with 2001 being the last year for all series. Less than 1.5% of data were missing for the period of record at each station. Previous studies (Chen *et al.*, 2009b; Li *et al.*, 2010) showed that the climate over the Loess Plateau does not manifest a significant change trend over the study period, and particularly in terms of precipitation. Thus, this dataset can be used to evaluate the performance of various stochastic weather generators. Figure1 presents the mean annual precipitation and maximum daily precipitation over the Loess Plateau calculated using daily time series from these 54 stations. The mean annual precipitation ranges between 130 and 845 mm and displays a clear decreasing trend from the southeast to the northwest (Figure 1(a)). However, there is no obvious trend regarding maximum daily precipitation, even though the northwest values are somewhat lower than those for the southeast (Figure 1(b)). Although the mean annual precipitation is only 545 mm, daily maximum precipitation can be as high as 262 mm at the Yangquan Station. Multi-year average T_{max} and T_{min} range from 0.8 to 19.3 °C and −6.6 to 9.0 °C, respectively, which adequately represent temperature variability over the Loess Plateau.

3. Methodology

3.1. Five stochastic weather generators

The performance of five commonly used stochastic weather generators is compared with respect to simulating precipitation occurrence, daily precipitation amounts, T_{max} and T_{min}, for all 54 stations. Since all five weather generators have already been presented in previous studies, Table 1 only presents the main algorithms for generating precipitation and temperature variables. More details can be found from the corresponding references listed in Table 1.

WGEN (Richardson and Wright, 1984), CLIMGEN (Stöckle *et al.*, 1999) and CLIGEN (Nicks and Gander, 1994) all use a first-order, two-state Markov chain to generate precipitation occurrence. However, in order to compare the performance of different orders (first-, second- and third-order) Markov chains in generating precipitation occurrence, CLIMGEN was modified to use a second-order Markov chain to simulate precipitation

Table 1. Comparison of five weather generators at simulating precipitation occurrence, daily precipitation amounts, and maximum and minimum temperatures (T_{max} and T_{min})

Model, version	Time scale	Precipitation occurrence	Precipitation amount	T_{max} and T_{min}	Key references
WGEN	Bi-weekly	Transition probabilities of a first-order, two-state Markov chain (two parameters)	Gamma distribution (two parameters)	T_{max} and T_{min} are generated using a normal distribution. The time series of observed data is reduced to a time series of residual elements at the daily scale. New residual series are then generated using the first-order linear autoregressive model with constant lag-1 auto-correlation and cross-correlation of and between T_{max} and T_{min}. The seasonal cycles of mean and standard deviation are modelled by finite Fourier series with two harmonics. T_{max} and T_{min} are not conditioned on each other, but rather, on dry and wet states	Richardson (1981), Richardson and Wright (1984)
CLIMGEN	Bi-weekly	Transition probabilities of a second-order, two-state Markov chain (four parameters)	Weibull distribution (two parameters)	Same procedure as WGEN	Stöckle *et al.* (1999), Mckague *et al.* (2005)
CLIGEN, v5.22564	Monthly	Same procedure as WGEN, but parameters are computed at the monthly scale	Skewed normal distribution (three parameters)	T_{max} and T_{min} are generated using a normal distribution. Parameters are calculated for each month using observed time series. Two random numbers are used to generate the standard deviate; the second random number for one day is reused as the first random number for the next day. T_{max} and T_{min} are conditioned on each other, but not on dry and wet states	Nicks and Gander (1994), Nicks *et al.* (1995)
WeaGETS, v1.1	Bi-weekly	Transition probabilities of a third-order, two-state Markov chain (eight parameters)	Mixed exponential distribution (three parameters)	Similar procedure as WGEN, but T_{max} and T_{min} are conditioned on each other	Chen *et al.* (2011, 2012b)
LARSWG, v5.5	Monthly	Lengths of alternate dry and wet sequences chosen from a semi-empirical distribution fitted to the observed series (21 parameters)	Semi-empirical distribution (21 parameters)	Similar procedure as WGEN, but instead of computing the cross-correlation from the observed residuals, LARSWG uses a pre-set cross-correlation. The seasonal cycles of mean and standard deviation are modelled by finite Fourier series with three harmonics	Semenov *et al.* (1998), Semenov and Barrow (2002)

occurrence in this study. Additionally, differently from the original CLIMGEN developed by Stöckle *et al.* (1999), which computes parameters on a monthly basis, this study modified CLIMGEN to compute parameters for every 2 weeks.

WeaGETS (Chen *et al.*, 2012a, 2012b) is a versatile weather generator for simulating daily precipitation and temperature. It brings together several options of other generators into one package, such as three Markov models (first-, second- and third-order Markov chain) to produce precipitation occurrence and four distributions (exponential, gamma, skewed normal and mixed exponential) to generate daily precipitation amounts. Additionally, a spectral correction approach was included in WeaGETS for correcting the underestimation of interannual variability, a problem common to all weather generators. However, this study only uses the combination of third-order Markov chain and mixed exponential distribution to simulate daily precipitation, while the spectral correction approach for correcting interannual variability is not used. A conditional scheme (Chen *et al.*, 2011, 2012b) is used to simulate daily $T_{\max}$ and $T_{\min}$.

Among the five weather generators selected, LARSWG (Semenov and Barrow, 2002) is the only one that uses a semi-empirical distribution to simulate precipitation occurrence and daily precipitation amounts. This semi-empirical distribution is a histogram with 11 bins (Semenov *et al.*, 1998). Random numbers from the distribution are chosen by first selecting one of the intervals and then selecting a value within that interval from the uniform distribution. The semi-empirical distribution is flexible enough to fit any shape of distribution, provided the observed time series is long enough. However, to use its semi-empirical distribution, 21 parameters must be estimated (the frequency of each bin and the 10 values separating them). LARSWG uses a similar technique as WGEN to simulate $T_{\max}$ and $T_{\min}$; however, unlike WGEN which computes the cross-correlation between $T_{\max}$ and $T_{\min}$ using observed residuals, LARSWG uses a pre-set cross-correlation coefficient. The seasonal cycles of mean and standard deviation are modelled by finite Fourier series, using two harmonics for WGEN and three for LARSWG.

3.2. Statistical analysis

The observed daily precipitation, $T_{\max}$ and $T_{\min}$ data, for all 54 stations were used to run the five weather generators to generate synthetic time series. The length of the generated series is 10 times that of the observed series. Statistics, including the mean, standard deviation and percentiles, were calculated for both observed and synthetic time series for all meteorological variables. The skewness and kurtosis coefficients were also calculated for daily, monthly and annual precipitation. The relative error (RE), computed as the difference between generated and observed values divided by the observed value, was calculated. Since daily $T_{\max}$ and $T_{\min}$ are approximately normal, t- and F-tests were used to test the equality of the means and standard deviations, respectively, for the generated *versus* observed daily temperatures at each station. As precipitation amounts are known to not follow normal distributions, instead of t- and F-tests, the nonparametric Mann–Whitney test and the squared ranks test (Conover, 1999), applicable to any type of distributions, were used to test the equality of the means and standard deviations of the observed and generated time series, respectively. A nonparametric Kolmogorov–Smirnov (K–S) test was used to test the equality of the two population distributions of the observed and generated data. All the tests were two-tailed, with a significance level of $P = 0.05$. Additionally, the observed and generated data were analysed for auto- and cross-correlation for and between daily $T_{\max}$ and $T_{\min}$.

4. Results

4.1. Precipitation occurrence

4.1.1. Wet day frequency

Figure 2 shows the scatter plots of the observed mean and standard deviation of the number of wet days per month *versus* the generated counterpart, respectively, for four occurrence models and 54 stations. The Markov chain-based models produce the mean and standard deviation of the number of wet days per month reasonably well. The third- and second-order Markov models show a slightly better performance than the first-order Markov model in terms of mean absolute errors (MAEs, mean(abs(sim-obs))) of mean and standard deviation, as presented in Table 2. These results indicate that all Markov chain-based models satisfactorily reproduce the wet-day frequency for the Loess Plateau. Both the mean and standard deviation of the number of wet days per month are generally underestimated by the semi-empirical distribution-based model. It is consistently slightly worse than the Markov chain-based model, especially at reproducing the standard deviation. The relatively poorer results indicate that the length of the time series may be insufficient to fit the wet spells for precipitation occurrence over the Loess Plateau, as the climate is very dry, especially in the winter when there may only be a few precipitation events at some stations. This point will be discussed later.

4.1.2. Dry and wet spells

4.1.2.1. Means and standard deviations of dry and wet spells: The Markov chain-based model with different levels of complexity is further compared with the semi-empirical distribution-based model with respect to reproducing the means and standard deviations of dry and wet spells. Overall, all four models perform well in reproducing the means and standard deviations of dry spells (Figure 3). The semi-empirical distribution-based model is worse than the Markov chain-based model at reproducing the means of dry spells. The nonparametric Mann–Whitney tests show that there is no significant

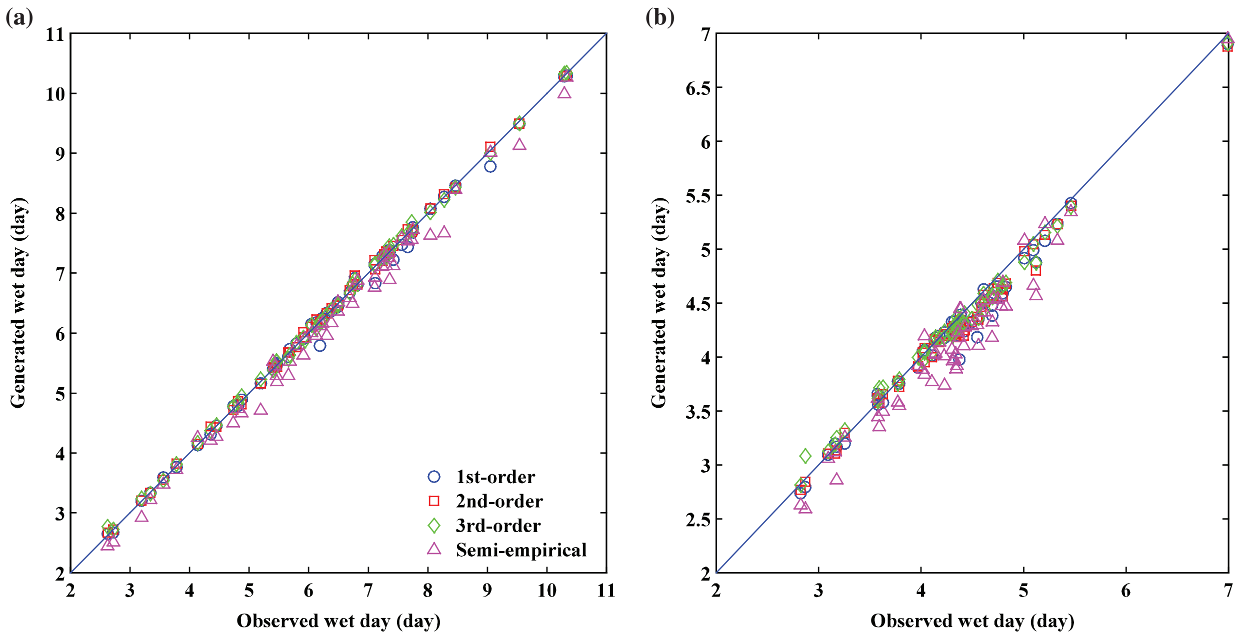

Figure 2. (a) Mean and (b) standard deviation (SD) of the observed *versus* the generated number of wet days per month from four models (first-, second- and third-order Markov chains and a semi-empirical distribution) for all 54 stations.

Table 2. Mean absolute errors (MAEs) of the wet day per month, and dry and wet spells for four models-generated precipitation occurrence

Model	Wet day per month		Dry spells			Wet spells		
	Mean	SD	Mean	SD	Q99	Mean	SD	Q99
First order	0.054	0.096	0.034	0.554	4.559	0.017	0.272	0.220
Second order	0.038	0.086	0.042	0.430	3.681	0.017	0.273	0.341
Third order	0.040	0.073	0.040	0.374	3.489	0.018	0.264	0.197
Semi-empirical	0.173	0.215	0.204	0.181	1.637	0.045	0.297	0.410

Mean = mean value; SD = standard deviation; Q99 = 99th percentile.

difference between observed and generated dry spells for all Markov chain-based models. However, with the order increasing from the first to the third, the mean P values of Mann–Whitney tests increase from 0.504 to 0.724. The semi-empirical distribution-generated dry spells are significantly different from the observed value for 17 out of 54 stations at the $P = 0.05$ level with a mean P value of 0.280.

In terms of the MAE (Table 2), the semi-empirical distribution-based model is slightly better than the Markov chain-based model at reproducing the standard deviation of dry spells, and the higher-order Markov model is better than the first-order Markov model. However, the squared ranks test shows non-significant differences between the observed and generated data of the four models for all 54 stations at $P = 0.05$.

Mean wet spells are reasonably reproduced by all four models and 54 stations (Figure 3). Although the MAE (Table 2) indicates a similar performance for all four models, the statistical tests provide a different result. The Mann–Whitney tests show significant differences in mean wet spells generated by the first-order Markov chain and the semi-empirical distribution-based model at $P = 0.05$ for 7 and 17 out of 54 stations, respectively. However, there is no significant difference between the observed and the second- and third-order Markov chain-generated wet spells for all stations at $P = 0.05$. The mean P values of the Mann–Whitney tests are 0.331, 0.716, 0.813 and 0.311 for the first, second- and third-order Markov chain-based and the semi-empirical distribution-based models, respectively.

The standard deviations of wet spells are well reproduced by all models for all 54 stations (Figure 3). In terms of the MAE (Table 2), the semi-empirical distribution-based model performs slightly worse than the Markov chain-based models. However, the squared ranks test shows non-significant differences between the observed and generated data for all models and all 54 stations at $P = 0.05$.

4.1.2.2. Distribution of dry and wet spells: The K–S test is used to test the equality of the distribution between observed and generated dry and wet spells. The results show that the observed and the Markov chain-based models generate wet spell distributions that are not statistically different for all 54 stations at $P = 0.05$. However, the mean P value of the K–S test increases

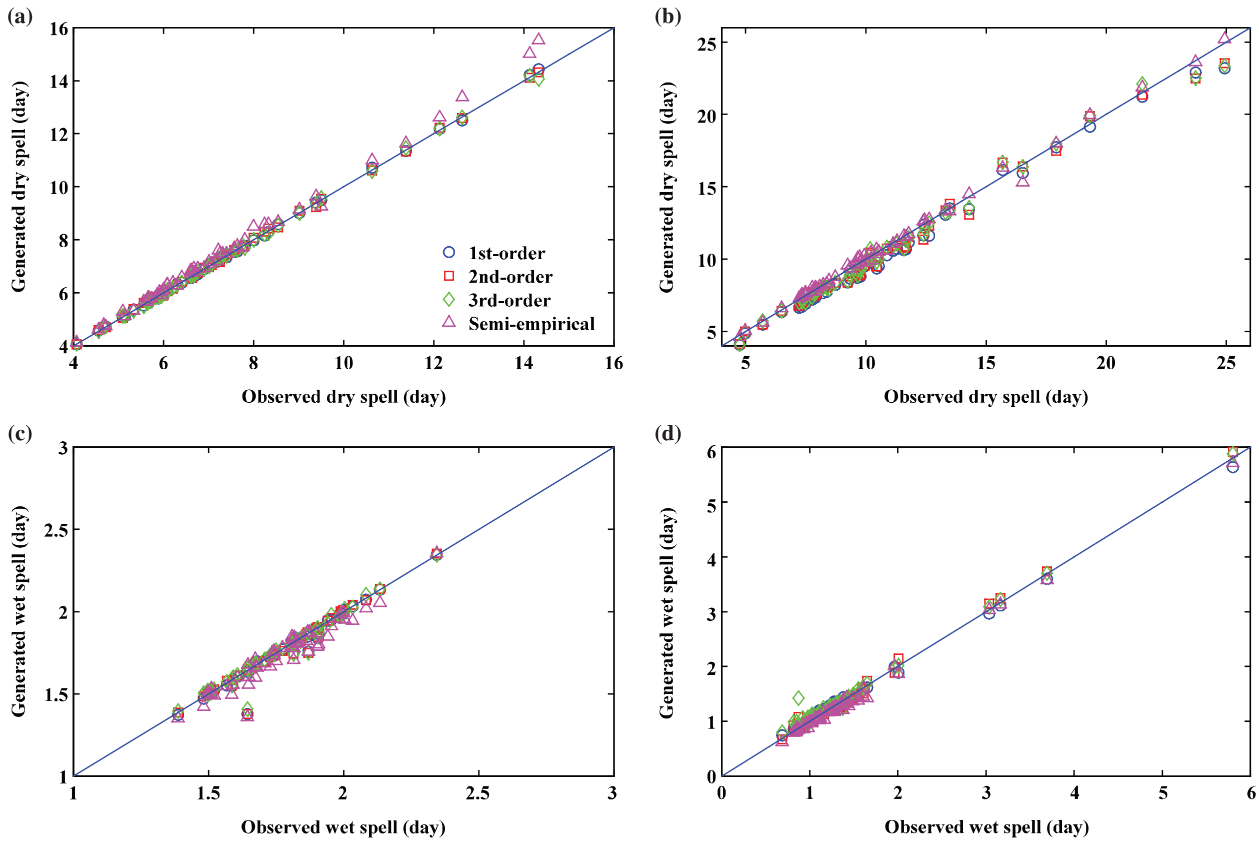

Figure 3. Mean and standard deviation (SD) of observed *versus* generated dry and wet spells from four models (first-, second- and third-order Markov chain and semi-empirical distribution) for all 54 stations. (a) Dry spell, mean, (b) dry spell, SD, (c) wet spell, mean and (d) wet spell, SD.

from 0.691 to 0.885 when the Markov chain order increases from the first to the third. Even though the semi-empirical distribution-based model performs better at reproducing the long dry spells, the K–S test shows a significant difference between observed and generated data for 18 out of 54 stations with the mean P value of 0.271, because the semi-empirical distribution-based model is not as good at reproducing short dry spells.

In terms of statistical tests, the second- and third-order Markov chain-based models also show a better performance than the first-order Markov chain and semi-empirical distribution-based model. The K–S tests indicate a significant difference between observed and first-order Markov chain and semi-empirical distribution-generated wet spells for 11 and 18 out of 54 stations at $P = 0.05$ with mean P values of 0.393 and 0.412, respectively. However, there is no significant difference between observed and second- and third-order Markov chain generated wet spells at $P = 0.05$ with mean P values of 0.940 and 0.980, respectively, across all 54 stations.

4.1.2.3. Extreme dry and wet spells: Scatter plots of generated and observed extreme dry and wet spells (99th percentile) are presented in Figure 4. Even though all Markov chain-based models underestimate extreme dry spells, the higher-order models consistently perform better than the first-order model (Table 2). Additionally, the semi-empirical distribution-based model performs better than the Markov chain-based model.

The long wet day spells are less well-reproduced by all models. Especially, the semi-empirical distribution-based model performs somewhat worse than the Markov chain-based models (Table 2). Similarly, the ability of the Markov chain-based models to reproduce wet spells improves with an increase in the model order.

4.2. Precipitation amount

The performance of five weather generators are compared with respect to simulating the mean, standard deviation and skewness and kurtosis coefficients of daily, monthly and annual precipitation.

4.2.1. Mean and standard deviation of precipitation amount

4.2.1.1. Mean precipitation amount: Figure 5 shows the scatter plots of the observed daily (precipitation amount $R \geq 0.1$ mm), monthly and annual precipitation *versus* the generated counterpart, respectively, for all five weather generators and 54 stations. Overall, all five models reproduce the mean daily, monthly and annual precipitation reasonably, as indicated by the fact that all values are close to the 1 : 1 line. This is further shown by the small

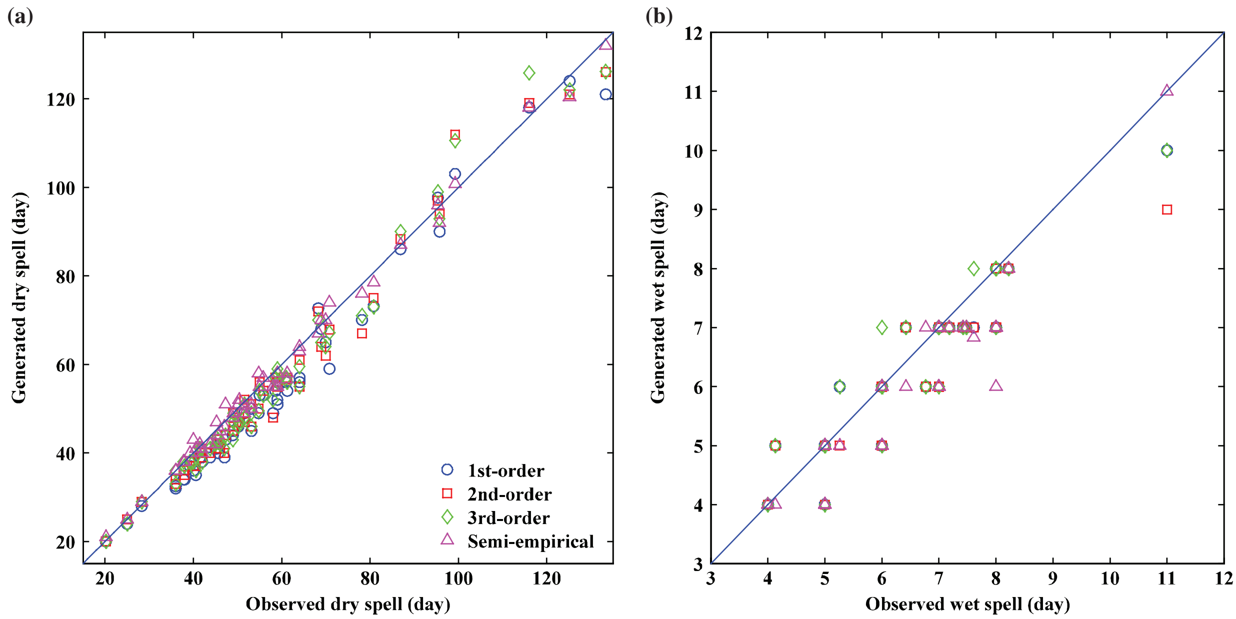

Figure 4. Scatter plots of the extreme (a) dry and (b) wet spells (Q99) between observed and generated from four models (first-, second- and third-order Markov chain and semi-empirical distribution) for all 54 stations.

mean absolute relative error (MARE, mean(abs((sim-obs)/obs))) of the generated mean precipitation (Table 3). However, the Mann–Whitney tests show that the mixed distribution-based WeaGETS performs clearly better than all other models at $P=0.05$, with a significant difference between observed and generated daily precipitation ($R \geq 0.1$ mm) for 2 out of 54 stations, followed by the semi-empirical distribution-based LARSWG, with significant differences for 25 out of 54 stations (Table 4). The worst performance is observed for the gamma distribution-based WGEN, with significant differences between observed and generated daily precipitation for all stations. The two-parameter Weibull distribution-based CLIMGEN performs better than the three-parameter skewed normal distribution-based CLIGEN and the two-parameter gamma distribution-based WGEN in simulating daily precipitation amounts. However, when the precipitation threshold is increased to 5 mm, LARSWG shows the best performance in the Mann–Whitney tests. This indicates that the semi-empirical distribution is more effective at reproducing the heavy precipitation events for the Loess Plateau. Moreover, CLIGEN is better than CLIMGEN at simulating daily precipitation amounts for a threshold greater than 5 mm. All five weather generators reproduce the monthly and annual total precipitation very well. The Mann–Whitney tests show non-significant differences between observed and generated data for all the weather generators at each of the 54 stations (Table 4).

4.2.1.2. Standard deviation of precipitation amount: LARSWG slightly overestimates the standard deviation of daily precipitation, with a MARE of 3.95%. However, the standard deviation of daily precipitation amounts is underestimated for all four other weather generators (Figure 5). In terms of the MARE (Table 3), WeaGETS performs better than all the other models, followed by CLIGEN. WGEN and CLIMGEN consistently perform worse than the others. The squared ranks tests further prove these conclusions regardless of the precipitation threshold (0.1 mm and 5 mm in this study; Table 4). All the weather generators underestimate the standard deviation of monthly and annual precipitation, indicating the underestimation of low-frequency variability for precipitation. Different performances are observed in terms of preserving the low-frequency component, even though none of them can absolutely remove the overdispersion. LARSWG shows the best performance, followed by CLIGEN. CLIGEN performs even better than the compound distribution-based WeaGETS. This conclusion is in line with the study of Chen and Brissette (2013).

The squared ranks test for monthly and annual precipitation shows slightly different results from using the MARE as a criterion. CLIGEN and WeaGETS consistently give a better performance than the other models. LARSWG presents the best performance in preserving the standard deviation of annual precipitation, with significant differences for only 7 out of 54 stations at $P=0.05$. However, it gives a relatively poor performance in preserving the standard deviation of monthly precipitation, with significant differences for 20 out of 54 stations at $P=0.05$. WGEN consistently shows the worst performance compared to the other models.

4.2.2. *Skewness and kurtosis of precipitation amount*

4.2.2.1. Skewness of precipitation amount: Compared to the mean and standard deviation of the observed *versus* generated precipitation being close to the 1 : 1 line, the skewness coefficients of observed precipitation (daily, monthly and annual scale) *versus* generated data are

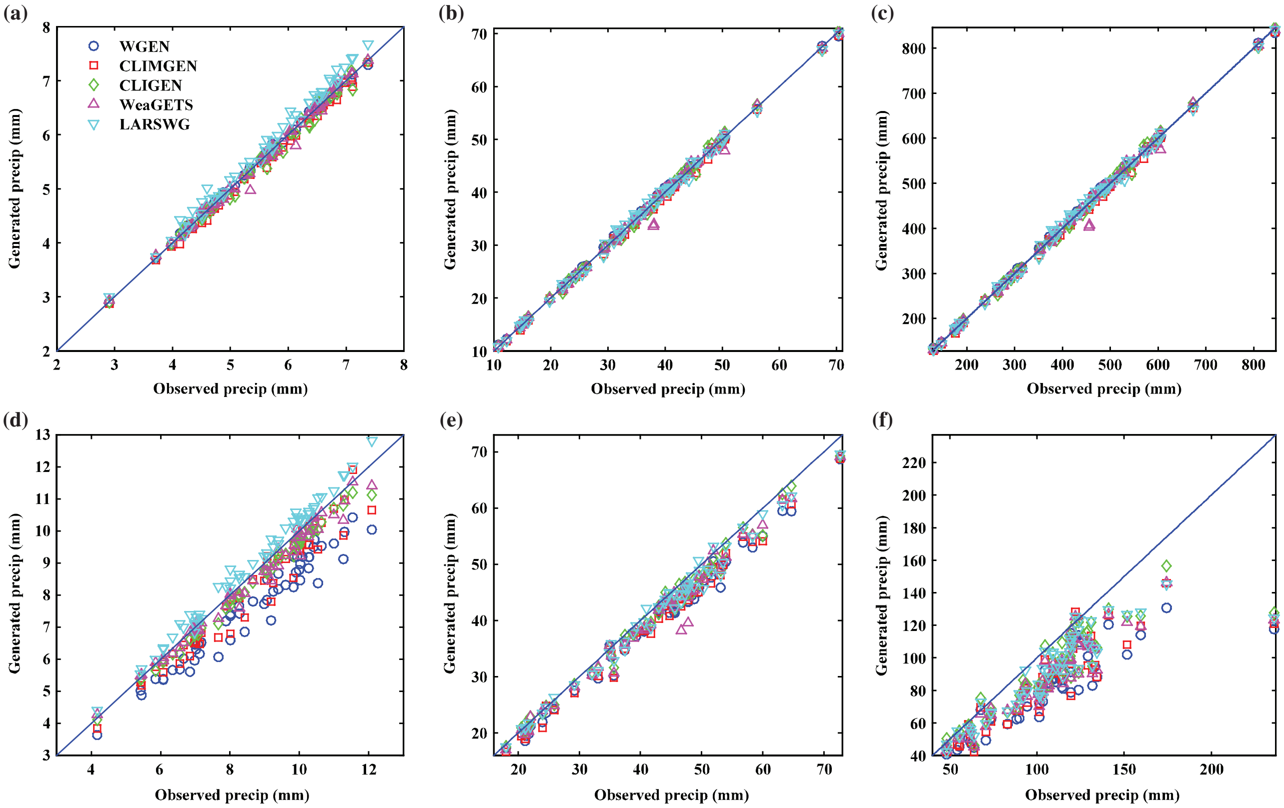

Figure 5. Mean and standard deviation (SD) of observed *versus* generated daily, monthly and annual precipitations from five models (WGEN, CLIMGEN, CLIGEN, WeaGETS and LARSWG) for all 54 stations. (a) Mean daily precipitation, (b) mean monthly precipitation, (c) mean annual precipitation, (d) SD of daily precipitation, (e) SD of monthly precipitation and (f) SD of annual precipitation.

Table 3. Mean absolutely relative errors (MAREs, %) of daily (DP0.1 = daily precipitation is greater than 0.1 mm), monthly (MP) and annual precipitation (AP) generated by five models

Model	Mean			SD			Skewness coefficient			Kurtosis coefficient		
	DP0.1	MP	AP	DP0.1	MP	AP	DP0.1	MP	AP	DP0.1	MP	AP
WGEN	0.583	0.784	0.784	11.748	6.456	21.958	14.771	14.194	174.350	26.498	19.453	17.604
CLIMGEN	1.497	1.478	1.478	5.540	6.091	18.870	14.608	9.289	231.078	46.794	15.659	20.811
CLIGEN	1.178	1.253	1.253	3.072	3.388	13.382	4.970	9.270	318.604	19.255	23.509	30.996
WeaGETS	0.913	1.358	1.358	2.526	4.663	17.059	10.237	10.887	161.998	20.839	17.369	20.730
LARSWG	2.994	1.701	1.701	3.954	3.262	13.623	3.027	8.520	257.728	7.279	15.387	21.332

Mean = mean value; SD = standard deviation.

much more scattered, as presented in Figure 6. This indicates that all the weather generators perform less well at simulating the skewness of precipitation distribution. The distribution of daily precipitation is extremely skewed to the left, as almost all the skewness coefficients are greater than three. WGEN considerably underestimates the skewness coefficient for most stations, while CLIMGEN remarkably overestimates the skewness coefficient for most stations, followed by WeaGETS. The overestimation or underestimation of the skewness coefficients by WGEN, CLIMGEN and WeaGETS indicates that the gamma, Weibull and mixed exponential distributions poorly preserve the shape of the daily precipitation distribution. In terms of MARE (Table 4), LARSWG and CLIGEN consistently perform better than the other weather generators. The best performance of LARSWG is as expected, since the semi-empirical distribution is flexible enough to fit any precipitation distribution shape.

Although the distribution of monthly precipitation is still skewed to the left, the skewness coefficient is much smaller than that of daily precipitation. Again, LARSWG shows the best performance, followed by CLIGEN and CLIMGEN. WGEN and WeaGETS perform worst at preserving the skewness of monthly precipitation. The skewness coefficients of observed and generated annual precipitation range between 0 and 1 for most stations, indicating that the annual precipitation approaches a normal distribution. The skewness coefficient of the generated annual precipitation is overestimated for small values, while it is underestimated for large values. The differences among the five models are not remarkable.

Table 4. The percentage (%) of stations (54 total) for daily (DP0.1 = daily precipitation is greater than 0.1 mm and DP5 = daily precipitation is greater than 5 mm), monthly (MP) and annual precipitation (AP) that rejected the Mann–Whitney, squared ranks and Kolmogorov–Smirnov (K–S) tests at the $P = 0.05$ level

Model	Mann–Whitney test				Squared ranks test				K–S test			
	DP0.1	DP5	MP	AP	DP0.1	DP5	MP	AP	DP0.1	DP5	MP	AP
WGEN	100	94.4	0	0	79.6	100	18.5	46.3	100	85.2	0	0
CLIMGEN	55.6	88.9	0	0	98.1	98.1	44.4	35.2	100	81.5	0	1.9
CLIGEN	85.2	35.2	0	0	72.2	83.3	29.6	24.1	100	24.1	3.7	0
WeaGETS	3.7	7.4	0	3.7	37.0	18.5	25.9	22.2	100	9.3	0	3.7
LARSWG	46.3	0	0	0	85.2	42.6	37.0	13.0	100	0	0	0

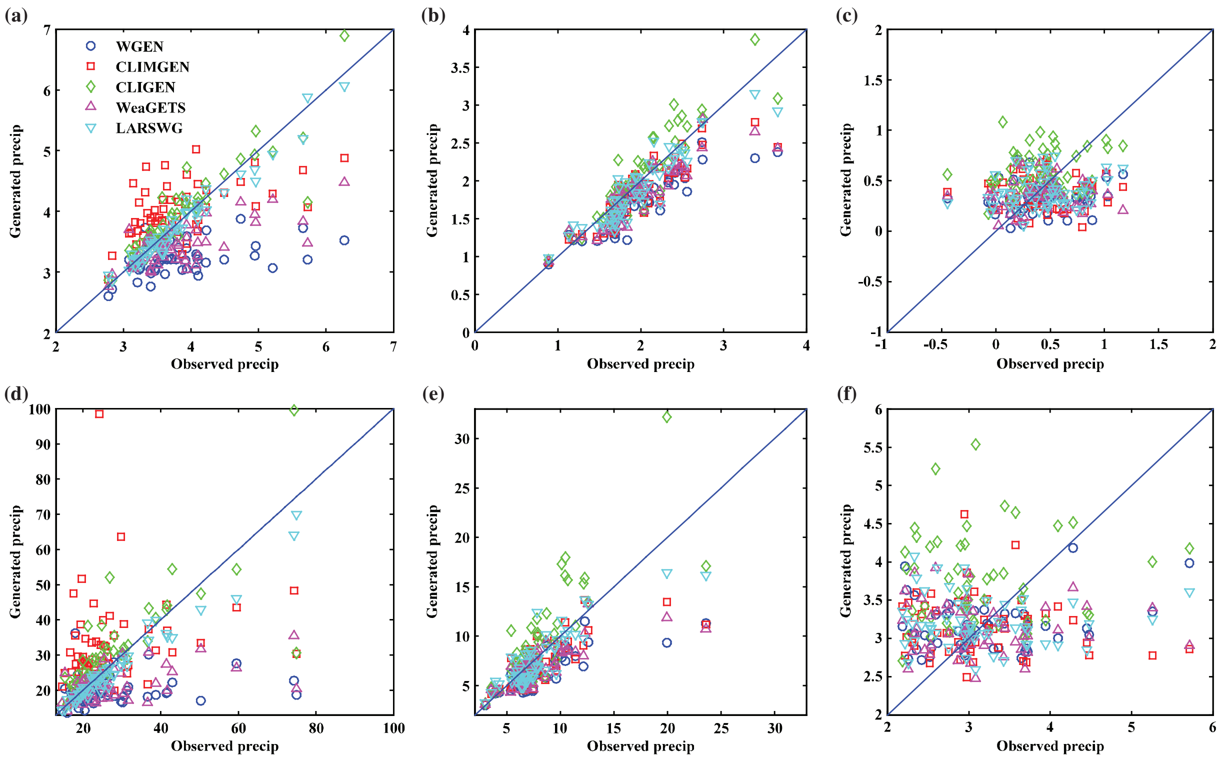

Figure 6. Skewness and kurtosis coefficients of observed *versus* generated daily, monthly and annual precipitations from five models (WGEN, CLIMGEN, CLIGEN, WeaGETS and LARSWG) for all 54 stations. (a) Skewness coefficient of daily precipitation, (b) skewness coefficient of monthly precipitation, (c) skewness coefficient of annual precipitation, (d) kurtosis coefficient of daily precipitation, (e) kurtosis coefficient of monthly precipitation and (f) kurtosis coefficient of annual precipitation.

4.2.2.2. Kurtosis of precipitation amount: The daily precipitation distribution is extremely peaked, as indicated by the large kurtosis coefficients for both observed and generated data (Figure 6). Unsurprisingly, LARSWG is consistently better than the other models in terms of the MARE (Table 3), followed by CLIGEN and WeaGETS. The kurtosis coefficient is less well-preserved by WGEN and CLIMGEN.

Similarly, LARSWG is better than all other models at preserving the kurtosis of monthly precipitation. However, the advantage is non-significant. The kurtosis coefficients of annual precipitation are consistently smaller than those of daily and monthly precipitation. The kurtosis coefficients of annual precipitation are poorly simulated by all the models. The semi-empirical distribution-based LARSWG is no longer better than the other models; instead, it is slightly worse than three of the five models.

4.2.3. *Distribution of precipitation amount*

Scatter plots of the generated *versus* observed percentiles (15th, Q25th, 50th, 75th, 95th and 99th, denoted by Q15, Q25, Q50, Q75, Q95 and Q99, respectively) of daily precipitation ($R \geq 0.1$ mm) are presented in Figure 7 to represent the overall distribution of daily precipitation. WGEN consistently overestimates the light and medium daily precipitation (Q15, Q25, Q50 and Q75), while it underestimates the heavy precipitation (Q95 and Q99), indicating that the most commonly used gamma distribution is not suitable for simulating the daily precipitation distribution for the Loess Plateau. The Weibull distribution-based CLIMGEN is consistently better than WGEN, especially for Q75, which is reasonably well-reproduced. However, percentiles smaller than Q75 are still overestimated and those greater than Q75 are underestimated. CLIGEN

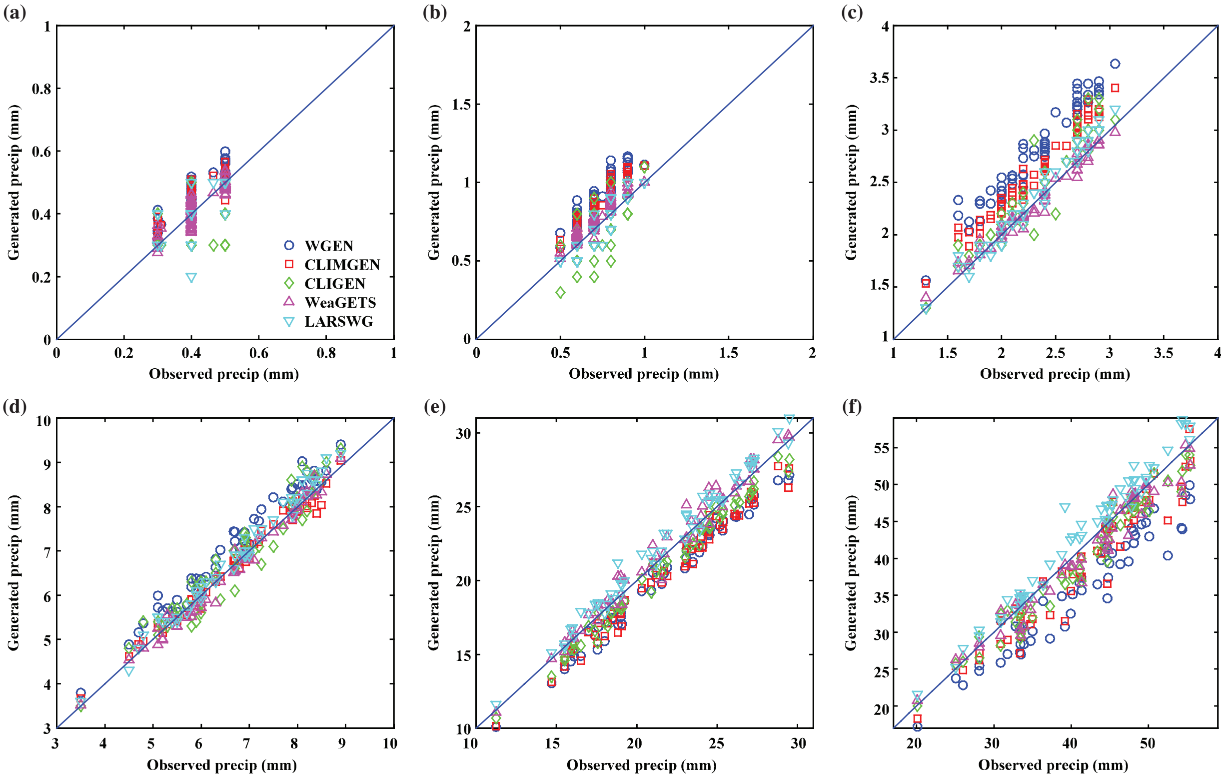

Figure 7. Scatter plots of the quantiles (a: Q15, b: Q25, c: Q50, d: Q75, e: Q95 and f: Q99) between observed and generated daily precipitation from five models (WGEN, CLIMGEN, CLIGEN, WeaGETS and LARSWG) for all 54 stations.

poorly reproduces the distribution of light precipitation, while reproducing the medium and heavy precipitation effectively. The poor reproduction of light precipitation may be caused by a random number quality control used by CLIGEN. The skewed normal distribution can be used for generating negative values, when the skewness coefficient is greater than 4.5 (Meyer, 2011). The random number control is used to ensure that the generated daily precipitation will always be positive. WeaGETS is better than the other parametric distribution-based models (WGEN, CLIMGEN and CLIGEN) at preserving the distribution of daily precipitation. This is because WeaGETS specifically takes into account the upper tail of the daily precipitation distribution by using an exponential distribution. LARSWG shows the best performance in reproducing the observed daily precipitation distribution.

The K–S test is further used to test the similarity of the two population distributions of observed and generated daily, monthly and annual precipitation (Table 4). The K–S tests reject the null hypothesis that the observed and generated daily precipitations are from the same population for all five models and 54 stations at $P = 0.05$, with a 0.1 mm precipitation threshold. These results may be due to the large sample size. When the sample size is very large, as in this study, the K–S test becomes excessively stringent (Zhang, 2013). However, when the precipitation threshold increases to 5 mm, a deviation of different models can be observed. LARSWG shows the best performance. WeaGETS is better than all the other parametric distribution-based models, followed by CLIGEN. WGEN and CLIMGEN reproduce the overall distribution of daily precipitation poorly.

The distributions of monthly and annual precipitation are reasonably reproduced by all five models for all 54 stations. The K–S tests for each station show that the cumulative distributions of generated and observed monthly and annual precipitation are not different at $P = 0.05$, suggesting that they are likely from the same distribution. The average P values of the K–S test for monthly precipitation are 0.487, 0.596, 0.451, 0.710 and 0.675 for WGEN, CLIMGEN, CLIGEN, WeaGETS and LARSWG, respectively, across all 54 stations, indicating the better performance of WeaGETS and LARSWG. All of the models indicate a similar performance in reproducing the distribution of annual precipitation, with average P values of 0.504, 0.576, 0.621, 0.598 and 0.575 for WGEN, CLIMGEN, CLIGEN, WeaGETS and LARSWG, respectively, across all 54 stations.

4.3. Maximum and minimum temperatures

Since CLIMGEN uses the same technique as WGEN to simulate $T_{\max}$ and $T_{\min}$, only four weather generators (WGEN, CLIGEN, WeaGETS and LARSWG) are compared with respect to reproducing the mean, standard deviation and overall distribution of $T_{\max}$ and $T_{\min}$.

4.3.1. Mean value of maximum and minimum temperatures

All four weather generators reproduce the mean $T_{\max}$ and $T_{\min}$ very well (Figure 8). The t-tests show that neither

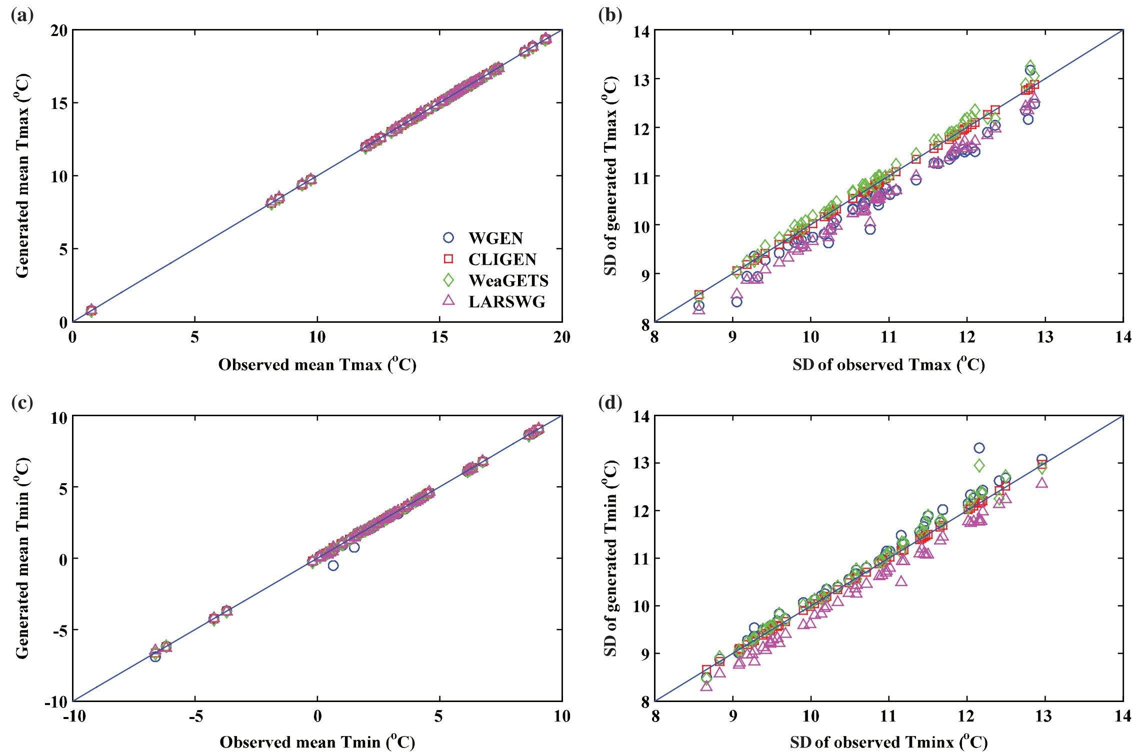

Figure 8. Mean and standard deviation (SD) of observed *versus* generated maximum and minimum temperatures (T_{max} and T_{min}) from four models (WGEN, CLIGEN, WeaGETS and LARSWG) for all 54 stations. (a) Mean T_{max}, (b) SD of T_{max}, (c) Mean T_{min} and (d) SD of T_{min}.

Table 5. The percentage (%) of stations (54 total) for maximum and minimum temperatures (T_{max} and T_{min}) that rejected the t-, F- and Kolmogorov–Smirnov (K–S) tests at the $P = 0.05$ level

Model	T_{max}			T_{min}		
	t-test	F-test	K–S test	t-test	F-test	K–S test
WGEN	0 (0.780)	96.3 (0.017)	96.3 (0.009)	7.4 (0.646)	46.3 (0.125)	88.9 (0.008)
CLIGEN	0 (0.901)	0 (0.926)	57.4 (0.078)	0 (0.896)	0 (0.953)	18.5 (0.167)
WeaGETS	0 (0.844)	38.9 (0.197)	61.1 (0.087)	0 (0.784)	33.3 (0.227)	94.4 (0.006)
LARSWG	0 (0.708)	100 (<0.0001)	100 (0.0001)	0 (0.608)	100 (0.0002)	100 (0.0001)

The mean P value is also presented (in parentheses) for all three tests.

the observed T_{max} nor T_{min} are significantly different from the generated data at $P = 0.05$ for all 54 stations, with the exception of significant differences between observed and WGEN-generated T_{min} for 4 out of 54 stations (Table 5). However, the average P value of the t-tests shows that CLIGEN and WeaGETS consistently perform better than WGEN and LARSWG. In particular, CLIGEN consistently performs better than all others. WGEN and LARSWG give similar performances in reproducing the mean T_{max} and T_{min}, as expected, since both use a very similar scheme.

4.3.2. *Standard deviation of maximum and minimum temperatures*

With the exception of CLIGEN, all the other three weather generators poorly reproduce the standard deviation of T_{max} and T_{min} (Figure 8). In terms of F-tests (Table 5), WeaGETS is consistently better than WGEN and LARSWG at preserving the standard deviation of T_{max} and T_{min}. WGEN underestimates the standard deviation of T_{min}, while it slightly overestimates that of T_{max} for most stations. LARSWG consistently underestimates both T_{max} and T_{min}, with significant differences in the F-tests between observed and generated T_{max} and T_{min} for all stations at the $P = 0.05$ level.

4.3.3. *Distribution of maximum and minimum temperatures*

The selected percentiles (Q15, Q25, Q50, Q75, Q95 and Q99) of observed and generated daily T_{max} are plotted in Figure 9 for distribution comparison purposes. Overall, the probability distribution of T_{max} is adequately reproduced by all four models for all stations, as indicated by their percentiles, which fall close to the 1 : 1 line. However, the average P values of the K–S tests show that WeaGETS and CLIGEN perform slightly better than the other two weather generators. The performance in almost all of the stations rejects the premise that the

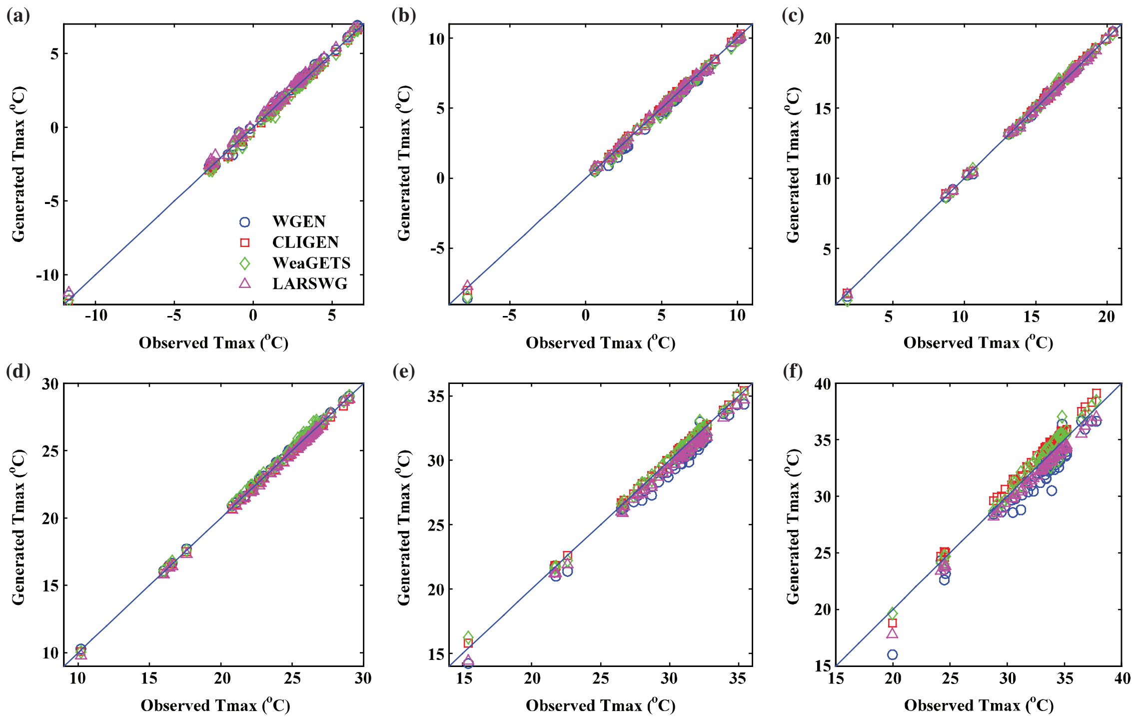

Figure 9. Scatter plots of the quantiles (Q15, Q25, Q50, Q75, Q95 and Q99) between observed and generated maximum temperature ($T_{\max}$) from four models (WGEN, CLIGEN, WeaGETS and LARSWG) for all 54 stations. (a) $T_{\max}$, Q15, (b) $T_{\max}$, Q25, (c) $T_{\max}$, Q50, (d) $T_{\max}$, Q75, (e) $T_{\max}$, Q95 and (f) $T_{\max}$, Q99.

observed and WGEN- and LARSWG-generated daily $T_{\max}$ come from the same probability distribution. The large number of stations that reject the K–S tests may be due to the large sample size in this study. However, the varying numbers of stations that reject the K–S tests demonstrate the different performances of the four models. Compared to other percentiles, the extremely high $T_{\max}$ (Q99 in this study) is less well reproduced by all four weather generators. Specifically, WeaGETS and CLIGEN slightly overestimate the extremely high $T_{\max}$, while WGEN and LARSWG slightly underestimate extreme values.

$T_{\max}$ symmetrical percentiles (Q85, Q75, Q50, Q25, Q5 and Q1) of the observed and generated daily $T_{\min}$ are plotted in Figure 10. All of the weather generators adequately reproduce the distribution of $T_{\min}$ for all stations. CLIGEN shows the best performance among the four models, as indicated by the average P value of the K–S test in Table 5. However, WGEN-, WeaGETS- and LARSWG-generated $T_{\min}$ are significantly different from the observed counterparts for most of the stations at $P = 0.05$. In particular, LARSWG consistently performs worse than the other three weather generators in reproducing extremely low $T_{\min}$.

4.3.4. Auto- and cross-correlations

Auto- and cross-correlations for and between daily $T_{\max}$ and $T_{\min}$ are computed for both the observed and generated datasets. Since similar results are obtained from all stations, only those from the Xi'an and Yinchuan Stations are shown in Figure 11 for illustration. Xi'an is located in the southern Loess Plateau, with a mean annual precipitation of 570.1 mm, and Yinchuan is located in the northwestern Loess Plateau, with a mean annual precipitation of 192.8 mm. These two stations represent the average and dry climates, of the Loess Plateau, respectively.

The auto-correlation level is a measure of the persistence of temperature trends (Chen *et al.*, 2008). The observed data show a clear day-to-day persistence. The observed lag 1 auto-correlation of $T_{\max}$ is reasonably reproduced by WGEN, while lags greater than one day are consistently overestimated. LARSWG is slightly better than WGEN at preserving the auto-correlation of $T_{\max}$. However, lags greater than one day are also overestimated. CLIGEN consistently underestimates the day-to-day persistence. WeaGETS reproduces not only the day-to-day persistence, but also the month-to-month persistence, as shown by the 30-day lag results. Similar to the $T_{\max}$ results, CLIGEN consistently underestimates the auto-correlation of $T_{\min}$ and the cross-correlation between $T_{\max}$ and $T_{\min}$. The performance of CLIGEN is consistently worse than that of the other three models. WGEN and LARSWG underestimate the lag 1 cross-correlation between $T_{\max}$ and $T_{\min}$, but reasonably reproduce lags greater than one day. In contrast, the lag 1 cross-correlation is well reproduced by WeaGETS, but lags greater than one day are slightly underestimated. These results indicate that the two random numbers used by CLIGEN are unable to preserve the auto- and cross-correlations for and between $T_{\max}$ and $T_{\min}$. In contrast, the auto- and cross-correlations can be much better

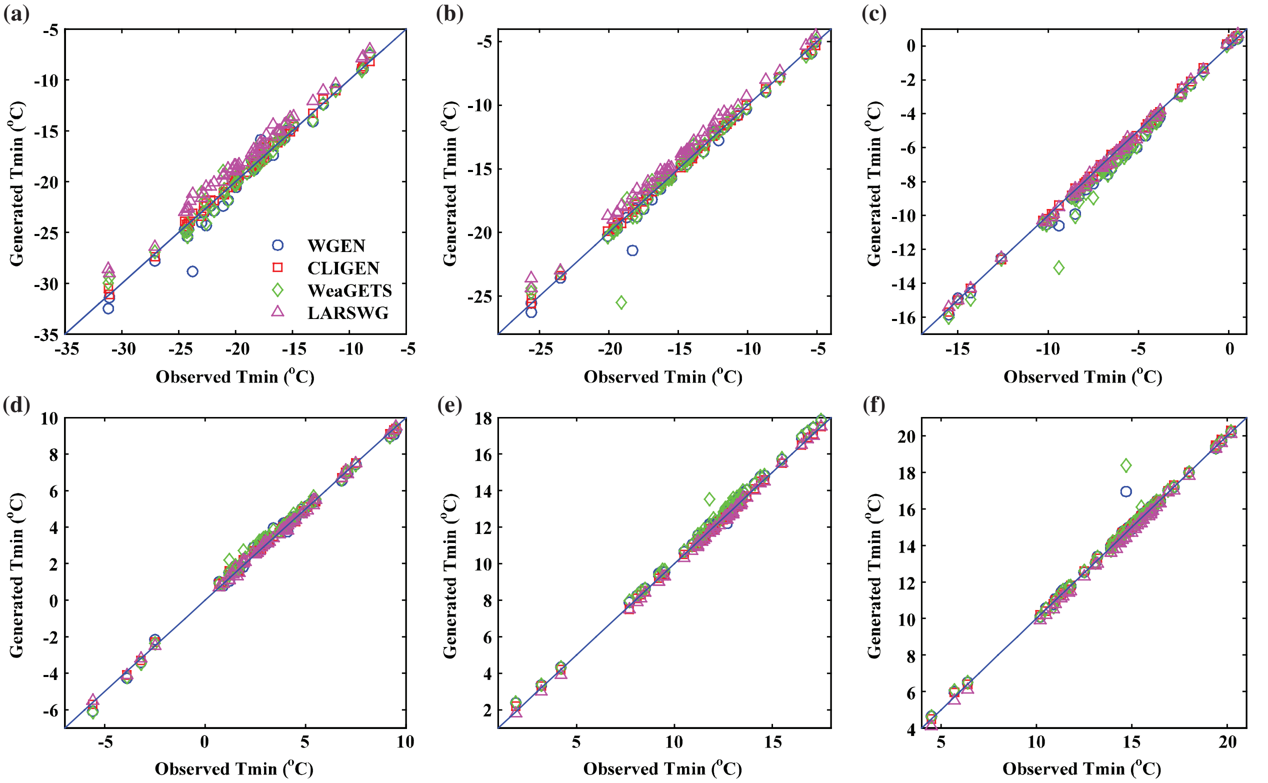

Figure 10. Scatter plots of the quantiles (Q1, Q15, Q25, Q50, Q75 and Q95) between observed and generated minimum temperature (T_{min}) from four models (WGEN, CLIGEN, WeaGETS and LARSWG) for all 54 stations. (a) T_{min}, Q15, (b) T_{min}, Q25, (c) T_{min}, Q50, (d) T_{min}, Q75, (e) T_{min}, Q95 and (f) T_{min}, Q99.

reproduced by the first-order linear autoregressive model. However, the auto- and cross-correlations are affected by the temperature scheme. The conditional scheme used by WeaGETS shows an overall better performance.

5. Discussion and conclusion

This article evaluates the performance of five commonly used stochastic weather generators in simulating daily precipitation, T_{max} and T_{min} for the Loess Plateau. In terms of the precipitation amounts generation, these five weather generators can be classified into parametric distribution-based models (WGEN, CLIMGEN, CLIGEN and WeaGETS) and a semi-empirical distribution-based model (LARSWG). For parametric distribution-based models, different schemes are used to simulate the precipitation occurrence and amount. These include three Markov chain-based models for simulating precipitation occurrence and five different probability distributions to simulate daily precipitation amounts. In theory, semi-empirical distributions are more flexible than a Markov chain-based model, and probability distribution in terms of closely fitting the distribution of wet and dry spells, and daily precipitation amounts, respectively. This is indicated by the performance of LARSWG being consistently better than that of other models in simulating the overall distribution of daily precipitation for the Loess Plateau, and especially in preserving the skewness and kurtosis coefficients. LARSWG also performs better than WGEN and CLIMGEN in simulating the mean and standard deviation of precipitation at all daily, monthly and annual scales. However, the semi-empirical distribution-based model is not always better than the parametric distribution-based models over the Loess Plateau. For example, it is consistently worse than the Markov chain-based models at reproducing the statistics of wet spells. This is because, compared to the parametric distribution-based model, LARSWG requires that more parameters are recorded. The higher number of parameters means that LARSWG requires longer observed time series for a good parameterization. However, there may be only a few precipitation events occurring each winter in the Loess Plateau. For some extreme cases, there is no precipitation occurring for a 1-month period. In particular, consecutive precipitation events for a few days are even rarer. Thus, the precipitation time series may not be long enough to parameterize the semi-empirical distribution for wet spells. This is especially true for the winter season and northwestern stations over the Loess Plateau, since precipitation events are very infrequent in both cases.

More work is needed to establish baselines as to the minimum data requirement for achieving stable parameters for LARSWG. This is especially true for the Loess Plateau, where precipitation is scarce and the historical record is short. Users should be aware of the limitations of any weather generator they use and ensure that they select a weather generator which is appropriate for their study, paying particular attention to its data

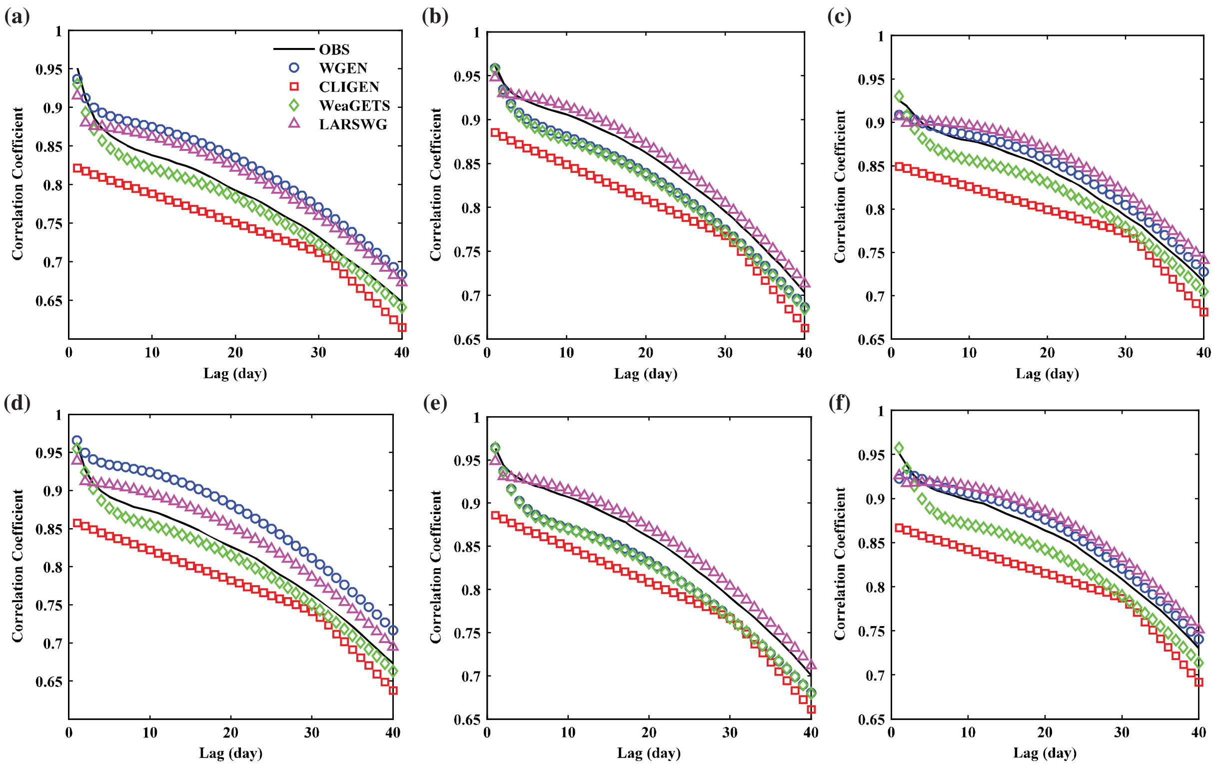

Figure 11. Auto- and cross-correlations (1 to 40 days lag) of and between observed (OBS) and WGEN, CLIGEN, WeaGETS, and LARSWG generated data for maximum and minimum temperatures (T_{max} and T_{min}) for the Xi'an and Yinchuan Stations. (a) Xi'an, auto T_{max}, (b) Xi'an, auto T_{min}, (c) Xi'an cross $T_{max}-T_{min}$, (d) Yinchuan, auto T_{max}, (e) Yinchuan, auto T_{min} and (f) Yinchuan cross $T_{max}-T_{min}$.

requirements and the effect that restricted calibration datasets may have on the results. This study further indicates that the performance of a weather generator is location-dependent, and that careful validation should always be performed, especially for studies over new regions.

The Markov chain-based models are more suitable for generating precipitation occurrence for the Loess Plateau. The overall performance of the first-order Markov chain satisfactorily simulates the precipitation occurrence. However, the second- and third-order Markov chain-based models are slightly better than the first-order Markov chain-based model, especially at simulating long wet and dry spells, and there is no obvious difference between the second- and third-order Markov chain-based models. Considering the fact that the number of parameters is increased exponentially with the increase of orders, it is more appropriate to use the first- and second-order Markov chain-based models to simulate precipitation occurrence for the Loess Plateau.

For simulating daily precipitation amounts, the three-parameter distribution-based weather generators (CLIGEN and WeaGETS) generally perform better than the two-parameter distribution-based ones (WGEN and CLIMGEN), especially at simulating extreme precipitation events. In particular, WeaGETS is even better than the semi-empirical distribution-based LARSWG at generating daily precipitation amounts for some cases. This is because WeaGETS specifically takes into account extreme precipitation using an additional exponential distribution. CLIGEN is capable of simulating extreme precipitation events, even though light precipitation is slightly underestimated. The other advantage of CLIGEN is that the parameters of the skewed normal distributions (mean, standard deviation and skewness coefficient) are easy to link to climate change signals. Considering that the simulation of extreme precipitation is vital in studying soil erosion for the Loess Plateau, it is more reasonable to use a three-parameter distribution-based model to simulate the daily precipitation amounts. The exploration of heavy-tailed distributions for adequately representing extreme precipitation is still being researched. For example, Furrer and Katz (2008) proposed a four-parameter hybrid gamma and generalized Pareto (GP) distribution to improve the simulation of extreme precipitation events. Li *et al.* (2012) proposed a compound distribution combining exponential and GP distributions to model the full spectrum of daily precipitation amounts. However, they are prone to generating outliers outside of the physically-plausible range.

All the weather generators underestimate the low-frequency variability of precipitation, even though WeaGETS, LARSWG and CLIGEN are generally better than the other two. The use of simple stationarity models cannot fully reproduce the variability of a non-stationary climate (Zhang and Garbrecht, 2003; Chen *et al.*, 2009a, 2010). However, WeaGETS provides a component to correct the underestimation of low-frequency variability for both precipitation and temperature (Chen *et al.*, 2010, 2011). To ensure a fair comparison of these five weather generators, the low-frequency variability was not corrected for WeaGETS-generated precipitation and temperature.

CLIGEN is consistently better than the other three weather generators (WGEN, WeaGETS and LARSWG) at reproducing the mean and standard deviation of daily $T_{\max}$ and $T_{\min}$. However, the temperature is generated independently from the precipitation status. This is not coincident to real cases, since temperature should generally be high for dry days and low for wet days. In particular, the two random numbers used by CLIGEN to preserve the auto- and cross-correlation for and between $T_{\max}$ and $T_{\min}$ are unsatisfactory. In contrast, the first-order linear autoregressive model used by WGEN, LARSWG and WeaGETS performs much better at preserving the observed auto- and cross-correlation. However, $T_{\max}$ and $T_{\min}$ generated by WGEN and LARSWG are not conditioned on each other, even though both of them are conditioned on precipitation status. WeaGETS makes use of the advantages of both WGEN and CLIGEN and improves the simulation of $T_{\max}$ and $T_{\min}$.

Overall, WeaGETS shows the best performance with respect to generating both precipitation (precipitation occurrence and amounts) and temperature ($T_{\max}$ and $T_{\min}$) for the Loess Plateau. Therefore, it is recommended for use in evaluating the effectiveness of soil and water conservation measures over the Loess Plateau. More importantly, WeaGETS provides a spectral approach to correcting the underestimation of low-frequency variability. This brings the synthetic data closer to real-world data. However, the accurate representation of extreme precipitation events is perhaps more important than the simulation of any other variables over the Loess Plateau for soil erosion studies. In light of this, CLIGEN could also be suitable over the Loess Plateau, since its skewed normal distribution reproduced the extreme precipitation reasonably well. The spectral approach of WeaGETS can easily be incorporated into CLIGEN.

The evaluation of commonly used weather generators is a first step prior to their use over the Loess Plateau. In order to accurately generate synthetic data to study soil erosion, the following should be included in future studies.

- Additional work to explore other distributions to better represent extreme precipitation events: The GP-based compound distribution could be one of the options, since it specifically considers the heavy tail of daily precipitation distribution. However, a reasonable strategy for screening outliers should be applied when using such a distribution as it is known to generate values outside of the physical realm. Additionally, one major objective of using a stochastic weather generator is to quantify the climate change impacts on soil erosion, through the perturbation of its parameters according to the climate change signal. As such, the ease of parameter modification should be a factor when considering a new distribution.
- Even though the interannual variability of precipitation amounts has been addressed by Chen *et al.* (2010) and other studies (e.g. Dubrovsky *et al.*, 2004; Wang and Nathan, 2007), ongoing work indicates that transition probabilities of precipitation occurrence also display interannual variability. Thus, additional studies should be undertaken on correcting the underestimation of interannual variability for precipitation occurrence.
- Since both precipitation and temperature values exhibit a significant annual cycle, performance may be somewhat affected by this annual cycle, and, as a result, weather generator may perform differently depending on the season. Weather generators should also be compared on a seasonal basis. In this work, the overall performance in terms of the seasonal scale (results not shown) was very similar to those at the annual scale.
- This study only compared the performance of five weather generators with respect to reproducing the observed precipitation and temperature characteristics. A more comprehensive comparison should also look at results from erosion models using the outputs of the various weather generators.

Acknowledgements

This work was partially funded by the Open Research Fund Program of the State Key Laboratory of Soil Erosion and Dryland farming on Loess Plateau, Institute of Soil and Water Conservation, Northwest A&F University (No. 10501-1205). The authors thank the Natural Science and Engineering Research Council of Canada (NSERC), Hydro-Québec and the Ouranos Consortium on Regional Climatology and Adaption to Climate Change for their support and contributions to this project. The authors also thank the China Meteorological Data Sharing Service System and the Soil Water Conservation Database of the Loess Plateau for providing climate data.

References

Buishand TA. 1978. Some remarks on the use of daily rainfall models. *J. Hydrol.* **36**: 295–308.

Chen J, Brissette F. 2013. Stochastic generation of daily precipitation amounts: review and evaluation of difference models. *Clim. Res.*, DOI: 10.3354/cr01214.

Chen J, Zhang XC, Liu WZ, Li Z. 2008. Assessment and improvement of CLIGEN non-precipitation parameters for the Loess Plateau of China. *Trans. Am. Soc. Agric. Biol. Eng.* **51**(3): 901–913.

Chen J, Liu WZ, Wang WL, Li Z. 2009a. The characteristics of precipitation and soil erosive rainfall in Changwu tableland-gully area of the Loess Plateau. *Sci. Soil Water Conserv.* **7**(1): 27–31 (In Chinese with an English abstract).

Chen J, Zhang XC, Liu WZ, Li Z. 2009b. Evaluating and extending CLIGEN precipitation generation for the Loess Plateau of China. *J. Am. Water Resour. Assoc.* **45**(2): 378–396.

Chen J, Brissette F, Leconte R. 2010. A daily stochastic weather generator for preserving low- frequency of climate variability. *J. Hydrol.* **388**: 480–490.

Chen J, Brissette F, Leconte R. 2011. Assessment and improvement of stochastic weather generators in simulating maximum and minimum temperatures. *Trans. Am. Soc. Agric. Biol. Eng.* **54**(5): 1627–1637.

Chen J, Brissette F, Leconte R, Caron A. 2012a. A versatile weather generator for daily precipitation and temperature. *Trans. Am. Soc. Agric. Biol. Eng.* **55**(3): 895–906.

Chen J, Brissette FP, Leconte R. 2012b. Downscaling of weather generator parameters to quantify the hydrological impacts of climate change. *Clim. Res.* **51**(3): 185–200, DOI: 10.3354/cr01062.

Chen J, Zhang XC, Brissette F. 2013. Assessing scale effects for statistically downscaling precipitation with GPCC model. *Int. J. Climatol.*, DOI: 10.1002/joc.3717.

Conover WJ. 1999. *Practical Nonparametric Statistics*, Third edn. Wiley: New York; 270–274, 300–301, 456–461.

Dubrovsky M, Buchteke J, Zalud Z. 2004. High-frequency and low-frequency variability in stochastic daily weather generator and its effect on agricultural and hydrologic modeling. *Clim. Change* **63**: 145–179.

Furrer EM, Katz RW. 2008. Improving the simulation of extreme precipitation events by stochastic weather generators. *Water Resour. Res.* **44**: W12439, DOI:10.1029/2008WR007316.

Flanagan DC, Nearing MA 1995. USDA-Water Erosion Prediction Project (WEPP) Hillslope Profile and watershed model documentation. NSERL Report No. 10, National Soil Erosion Research Laboratory, USDA-Agricultural Research Service, West Lafayette, Indiana.

Green JR. 1964. A model for rainfall occurrence. *J. Roy. Stat. Soc.* **B26**: 345–353.

Ison NT, Feyerherm AM, Bark LD. 1971. Wet period precipitation and the Gamma distribution. *J. Appl. Meteorol.* **10**(4): 658–665.

Johnson GL, Hanson CL, Hardegree SP, Ballard EB. 1996. Stochastic weather simulation: overview and analysis of two commonly used models. *J. Appl. Meteorol.* **35**: 1878–1896.

Katz RW. 1977. Precipitation as a chain-dependent process. *J. Appl. Meteorol.* **16**(7): 671–676.

Kilsby CG, Jones PD, Burton A, Ford AC, Flower HJ, Harpham C, James P, Smith A, Wilby RL. 2007. A daily weather generator for use in climate change studies. *Environ. Model. Softw.* **22**: 1705–1719.

Kottek M, Grieser J, Beck C, Rudolf B, Rubel F. 2006. World Map of the Köppen-Geiger climate classification updated. *Meteorol. Z.* **15**(3): 259–263.

Li Z, Zheng FL, Liu WZ, Flanagan DC. 2010. Spatial distribution and temporal trends of extreme temperature and precipitation events on the Loess Plateau of China during 1961-2007. *Quat. Int.* **226**: 92–100.

Li C, Singh VP, Mishra AK. 2012. Simulation of the entire range of daily precipitation using a hybrid probability distribution. *Water Resour. Res.* **48**, W03521, DOI: 10.1029/2011WR011446.

Li Z, Brissette F, Chen J. 2013. Assessing the applicability of six precipitation frequency distribution models on the Loess Plateau of China. *Int. J. Climatol.*, DOI: 10.1002/joc.3699.

Mavromatis T, Hansen JW. 2001. Interannual variability characteristics and simulated crop response of four stochastic weather generators. *Agr. Forest. Meteorol.* **109**: 283–296.

Mckague K, Rudra R, Ogilvie J, Ahmed I, Gharabaghi B. 2005. Evaluation of weather generator ClimGen for Southern Ontario. *Can. Water Resour. J.* **30**(4): 315–330.

Mehrotra R, Sharma A. 2007a. A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability. *J. Hydrol.* **335**: 180–193.

Mehrotra R, Sharma A. 2007b. Preserving low-frequency variability in generated daily rainfall sequences. *J. Hydrol.* **345**: 102–120.

Meyer C. 2011. *General Description of the CLIGEN Model and Its History*. West Lafayette, IN: USDA-ARS National Soil Erosion Laboratory.

Nicks AD, Gander GA. 1994. CLIGEN: a weather generator for climate inputs to water resource and other model. In *Proceedings of the 5th International Conference on Computers in Agriculture*. American Society of Agricultural Engineers, St. Joseph, MI, 3–94.

Nicks AD, Lane LJ, Gander GA. 1995. Weather generator, Ch. 2. In *USDA-Water Erosion Prediction Project: Hillslope Profile and Watershed Model Documentation*, Flanagan DC, Nearing MA (eds). NSERL Report No. 10, USDA-ARS-NSERL, West Lafayette, IN.

Pruski FF, Nearing MA. 2002. Climate-induced changes in erosion during the 21st century for eight U.S. locations. *Water Resour. Res.* **38**(12): 341–3411.

Qian BD, Gameda S, Hayhoe H, Jong RD, Bootsma A. 2004. Comparison of LARS-WG and AAFC-WG stochastic weather generators for diverse Canadian climates. *Climate Res.* **26**: 175–191.

Qian BD, Hayhoe H, Gameda S. 2005. Evaluation of the stochastic weather generators LARS-WG and AAFC-WG for climate change impact studies. *Climate Res.* **29**: 3–21.

Qian BD, Gameda S, Jong R, Fallon P, Gornall J. 2010. Comparing scenarios of Canadian daily climate extremes derived using a weather generator. *Climate Res.* **41**(2): 131–149.

Richardson CW. 1981. Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resour. Res.* **17**: 182–190.

Richardson CW, Wright DA. 1984. WGEN: a model for generating daily weather variables. US Department of Agriculture, Agricultural Research Service, ARS-8, 83.

Roldan J, Woolhiser DA. 1982. Stochastic daily precipitation models, 1. A comparison of occurrence processes. *Water Resour. Res.* **18**: 1451–1459.

Semenov MA, Barrow EM. 1997. Use of a stochastic weather generator in the development of climate change scenarios. *Clim. Change* **35**: 397–414.

Semenov MA, Barrow EM. 2002. *LARS-WG, A Stochastic Weather Generator for Use in Climate Impact Studies, User Manual*. http://www.rothamsted.ac.uk/mas-models/download/LARS-WG-Manual.pdf.

Semenov MA, Porter JR. 1995. Climatic variability and the modelling of crop yields. *Agr. Forest. Meteorol.* **73**: 265–283.

Semenov MA, Brooks RJ, Barrow EM, Richardson CW. 1998. Comparison of the WGEN and LARS-WG stochastic weather generators in diverse climates. *Climate Res.* **10**: 95–107.

Stöckle CO, Campbell GS, Nelson R. 1999. *ClimGen Manual*. Pullman, WA: Biological Systems Engineering Department, Washington State University.

Todorovic P, Woolhiser DA. 1975. A stochastic model for n-day precipitation. *J. Appl. Meteorol.* **14**(1): 17–24.

Wang QJ, Nathan RJ. 2007. A method for coupling daily and monthly time scales in stochastic generation of rainfall series. *J. Hydrol.* **346**: 122–130.

Wheater H, Chandler R, Onof C, Isham V, Bellone E, Yang C, Lekkas D, Lourmas G, Segond ML. 2005. Spatial-temporal rainfall modelling for flood risk estimation. *Stoch. Environ. Res. Risk Assess.* **19**: 403–416.

Wilks DS. 1989. Conditioning stochastic daily precipitation models on total monthly precipitation. *Water Resour. Res.* **25**: 1429–1439.

Wilks DS. 1992. Adapting stochastic weather generation algorithms for climate change studies. *Clim. Change* **22**: 67–84.

Wilks DS. 1999a. Multisite downscaling of daily precipitation with a stochastic weather generator. *Climate Res.* **11**: 125–136.

Wilks DS. 1999b. Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agr. Forest. Meteorol.* **93**: 153–169.

Yang WZ, Yu CZ. 1992. *Regional governance and evaluation for the Loess Plateau*. Science Press: Beijing, China; 1–100 (In Chinese).

Zhang XC. 2005a. Spatial downscaling of global climate model output for site-specific assessment of crop production and soil erosion. *Agr. Forest. Meteorol.* **135**: 215–229.

Zhang XC. 2005b. Spatial downscaling of global climate model output for site-specific assessment of crop production and soil erosion. *Agr. Forest. Meteorol.* **135**: 215–229.

Zhang XC. 2013. Verifying a temporal disaggregation method for generating daily precipitation of potentially non-stationary climate change for site-specific impact assessment. *Int. J. Climatol.* **33**(2): 326–342, DOI: 10.1002/joc.3425.

Zhang XC, Garbrecht JD. 2003. Evaluation of CLIGEN precipitation parameters and their implication on WEPP runoff and soil loss prediction. *Trans. Am. Soc. Agric. Eng.* **46**(2): 311–320.

Zhang XC, Liu WZ. 2005. Simulating potential response of hydrology, soil erosion, and crop productivity to climate change in Changwu tableland region on the Loess Plateau of China. *Agr. Forest. Meteorol.* **131**: 127–142.

Zhang XP, Zhang L, Zhao J, Rustomji P, Hairsine P. 2008. Responses of streamflow to changes in climate and land use/cover in the Loess Plateau, China. *Water Resour. Res.* **44**(7): W00A07.1–W00A07.12.

Zhou PH, Wang ZL. 1992. A study on rainstorm causing soil erosion in the Loess Plateau. *J. Soil Water Conserv.* **6**: 1–5 (In Chinese with an English Abstract).

