ORIGINAL PAPER

CrossMark

Impact of number of realizations on the suitability of simulated weather data for hydrologic and environmental applications

Tian Guo[1,2] · Sushant Mehan[1] · Margaret W. Gitau[1] · Qi Wang[3] · Thomas Kuczek[3] · Dennis C. Flanagan[4,1]

Abstract

Stochastic weather generators are widely used in hydrological, environmental, and agricultural applications to simulate weather time series. However, such stochastic models produce random outputs hence the question on how representative the generated data are if obtained from only one simulation run (realization) as is common practice. In this study, the impact of different numbers of realizations (1, 25, 50, and 100) on the suitability of generated weather data was investigated. Specifically, 50 years of daily precipitation, and maximum and minimum temperatures were generated for three weather stations in the Western Lake Erie Basin (WLEB), using three widely used weather generators, CLIGEN, LARSWG and WeaGETS. Generated results were compared with 50 years of observed data. For all three generators, the analyses showed that one realization of data for 50 years of daily precipitation, and maximum and minimum temperatures may not be representative enough to capture essential statistical characteristics of the climate. Results from the three generators captured the essential statistical characteristics of the climate when the number of realizations was increased from 1 to 25, 50 or 100. Performance did not improve substantially when realizations were increased above 25. Results suggest the need for more than a single realization when generating weather data and subsequently utilizing in other models, to obtain suitable representations of climate.

Keywords Stochastic weather generators · Simulation approaches · Climate realizations · Statistical properties · Statistical analysis

1 Introduction

Stochastic weather generators have been widely used for environmental, hydrological and agricultural assessment and applications (Wheater et al. 2005). These computer

✉ Margaret W. Gitau
mgitau@purdue.edu

1   Department of Agricultural and Biological Engineering, Purdue University, 225 S. University St, West Lafayette, IN 47907, USA

2   National Center for Water Quality Research, Heidelberg University, Tiffin, USA

3   Department of Statistics, Purdue University, West Lafayette, USA

4   USDA Agricultural Research Service, National Soil Erosion Research Laboratory, West Lafayette, USA

programs aim at producing synthetic time series of climate data—such as precipitation, maximum temperature (Tmax), minimum temperature (Tmin), solar radiation, and relative humidity—with statistical characteristics similar to those of observed climate data (Racsko et al. 1991; Semenov et al. 1998; Wilks and Wilby 1999). Simulated weather data is often used as model input, particularly where observed data may not be consistent or available in sufficient quantities. For example, existing data might have missing values, or may not be sufficient to allow application or estimation of the probability of extreme events (Semenov et al. 1998). Furthermore, weather generator simulations can provide as many realizations of the climate as needed, which may be necessary for Monte-Carlo analyses (Semenov et al. 1998; Wilks and Wilby 1999; Kou et al. 2007). Moreover, generated weather data can be used in modeling research based on future climate scenarios (Eames et al. 2012).

Daily precipitation, Tmax, and Tmin are the most common outputs from stochastic weather generators, in

🖄 Springer

which Markov chain approaches and the alternative renewal process are used to determine precipitation occurrence (Racsko et al. 1991; Chen and Brissette 2014a). Precipitation occurrence is typically simulated on a daily basis for Markov chain-based models, while precipitation occurrence is considered as a sequence of alternating wet and dry days simulated independently in the alternative renewal process (Roldán and Woolhiser 1982; Semenov and Barrow 1997). The most widely used approaches to simulate precipitation depth are parametric probability distributions, including single distributions, such as skewed normal (Nicks et al. 1995), exponential (Roldán and Woolhiser 1982), gamma distributions (Richardson and Wright 1984), compound distributions, such as hybrid exponential and Pareto distribution (Li et al. 2012), and mixed exponential distributions (Roldán and Woolhiser 1982; Wilks 1999). Daily Tmax and Tmin simulations are usually based on a normal distribution (Chen et al. 2012b).

Weather generators generally reproduce average characteristics of variables (means and variances) reasonably well, but neither parametric (Richardson 1981) nor resampling (Rajagopalan and Lall 1999) in generators performs particularly well in simulating extreme events (Wilks and Wilby 1999; Sharif and Burn 2006; Furrer and Katz 2008) particularly given that extreme values typically follow their own distributions suggesting the need for compound distributions. It is, therefore, important to evaluate performance of weather generators in simulating extreme weather events, such as wet and dry spells, large precipitation depths, and high and low temperatures. Extreme precipitation events (floods and droughts) can affect simulation of runoff volumes and peak runoff rates (Harmel et al. 2000). Water and temperature stresses may decrease crop yields (Semenov 2008; Guo et al. 2015), and increases in frequency and magnitude of extreme events may occur under climate change (Solomon et al. 2007).

The stochastic processes in weather generators produce random outputs and the climate might not be represented adequately if only one realization is used. Thus, several realizations may be required to approximate the statistical characteristics of weather data with acceptable accuracy (Hansen and Ines 2005; Qian et al. 2005). More commonly, however, one realization of generated weather data is used (Semenov and Barrow 1997; Semenov et al. 1998; Wilks 2002; Zhang and Garbrecht 2003; Chen et al. 2011, 2012a; Eames et al. 2012; Chen and Brissette 2014a, b; Chen et al. 2014). Some exceptions include, 10, 30, and 250 realizations of generated weather data used to: validate CLIGEN from sites in Uganda (Elliot and Arnold 2001); evaluate impacts on hydrologic modeling (Caron et al. 2008); and, simulate best management practice effects (Chaubey et al. 2010), respectively. Given the potential implications for hydrologic, agricultural, and environmental applications

(and the implications on computational resources), it is necessary to determine how many realizations of generated weather data are needed to capture essential statistical characteristics of observed data (such as precipitation occurrence, extremes, and variability). We hypothesized that one realization of simulated daily weather data is not adequate to capture essential statistical characteristics of observed weather data. The purpose of this study was, thus, to determine the impact of the number of realizations on the suitability of generated daily precipitation depth, Tmax, and Tmin. Three commonly used weather generators-CLImate GENerator, CLIGEN (Nicks and Gander 1994), Long Ashton Research Station-Weather Generator, LARSWG (Semenov and Barrow 1997), and Weather Generator of École de Technologie Supérieure, WeaGETS (Chen et al. 2012b) were used in this study. These three weather generators have been used successfully in a number of climate and climate change impact studies (Chen and Brissette 2014a; Kou et al. 2007; Semenov et al. 1998). A comparative assessment of the three generators is presented in Mehan et al. (2017), thus, this study is focused on the number of realizations.

CLIGEN uses a first-order, two-state Markov Chain and skewed normal distribution, to generate precipitation occurrence and amount, respectively (Chen and Brissette 2014b). A normal distribution is used to generate Tmax and Tmin. Parameters in CLIGEN are computed at the monthly scale. The standard deviation (SD) of Tmax and Tmin is reproduced based on two random numbers, the second one of which for one day is considered as the first random number for the next day. Tmax and Tmin depend on each other (Nicks and Gander 1994; Chen and Brissette 2014a). LARSWG uses a semi-empirical distribution to simulate precipitation amounts and lengths of alternating wet and dry spells. A normal distribution is used to generate Tmax and Tmin. A new residual series is generated based on the residuals of observed data using a first-order linear autoregressive model (Chen et al. 2014). A finite Fourier series is used to represent the seasonal cycles for the mean and SD of Tmax and Tmin which are conditioned on precipitation status, but they are not conditioned on each other (Chen and Brissette 2014a). Parameters in LARSWG are computed at the monthly scale (Semenov and Barrow 1997; Semenov et al. 1998). For WeaGETS, first-, second-, and third-order Markov chain-based models are incorporated to determine precipitation occurrence. Gamma, exponential, mixed exponential, and skewed normal distributions are included to calculate precipitation depths (Chen et al. 2011, 2012a; Chen and Brissette 2014a). Users are allowed to choose options for their specific needs (Chen et al. 2011). Parameters in WeaGETS are computed at the bi-weekly scale.

This study was based on the Fort Wayne, Adrian and Norwalk weather stations in the Western Lake Erie Basin (WLEB), each of which has a long record of consistent observed weather data. Statistical characteristics of generated precipitation occurrence, daily precipitation depths, Tmax and Tmin for 1, 25, 50, and 100 realizations of 50-year synthetic series were analyzed and compared to those of observed data. The impacts of the number of realizations on the suitability of generated data were then evaluated based on analysis results. Finally, the number of realizations needed to adequately capture statistical characteristics of the observed data was determined.

2 Methodology

2.1 Preliminary analysis of observed weather data

The Fort Wayne station (GHCND: USW00014827, Latitude: + 41.17, Longitude: − 85.13, Elevation: 241.1 m) in Indiana, Adrian station (GHCND: USC00200032, Latitude: + 41.92, Longitude: − 84.02, Elevation: 231.6 m) in Michigan, and Norwalk station (GHCND: USC00336118, Latitude: + 41.27, Longitude: − 82.62, Elevation: 204.2 m) in Ohio were selected for this study (Fig. 1). These stations were selected based on data availability and consistency, and considering the need for spatial coverage across the entire basin including Indiana, Ohio, and Michigan, consistent with Mehan et al. (2017). Daily precipitation, Tmax, and Tmin data from 01/01/1966 to 12/31/2015 at these stations were obtained from the National Climatic Data Center (NCDC) and used to create parameter sets specifically for all three generators. Annual precipitation at these stations during the study period ranged from 879 mm at Adrian to 957 mm at Norwalk. Mean maximum temperatures (summer) ranged from 27.1 °C at Norwalk to 28 °C at Fort Wayne, while mean minimum temperatures (winter) ranged from − 6.9 °C at Fort Wayne to − 7.9 °C at Adrian. Data availability was 100% for precipitation, Tmax and Tmin at Fort Wayne, 99.2% for precipitation, 98.3% for Tmax and 98.9% for Tmin at Adrian, and 99.7% for precipitation, and 99.0% for Tmax and Tmin at Norwalk.

For daily Tmax and Tmin from NCDC stations, biases related to changes in measurement techniques, station relocations and instrumentation had already been corrected in the downloaded data (NCDC 2017). Daily precipitation data were evaluated for observer biases, underreporting of amounts of less than 1.27 mm (0.05 in), and overreporting of amounts divisible by 5 and/or 10 (Schneider 2001; Daly et al. 2007). Observed daily precipitation from the three stations did not show observer biases. The Augmented

Dickey–Fuller (ADF) test was used to check for stationarity in annual total precipitation amounts, and average Tmax and Tmin of the observed data (Cheung and Lai 1995). To test the null hypothesis that a unit root is present in the annual observed data, the general linear regression equation with a constant and a linear trend was used, and the t-statistic for a first order autoregressive coefficient equals one was computed in the ADF test (Said and Dickey 1984; Banerjee et al. 1993; Fuller 2009). The lag order was selected as 1 based on Akaike's information criterion (AIC), which has been widely used to determine the lag order for small sample sizes (annual observed data as 50 in this case) (Ng and Perron 2001; Liew 2004). The Dickey–Fuller statistics were computed for annual total precipitation amounts, and average Tmax and Tmin of the observed data at the three stations (ranged from to − 6.2 to − 4.2). Serial autocorrelations of residuals of the general regression model of the observed data were also checked (p value < 0.05), and Dickey–Fuller statistics ranged from − 6.4 to − 4.2. Observed annual precipitation, Tmax and Tmin from the three stations were found to be stationary, based on the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots, results from which were consistent with ADF test results (p value < 0.05). The ACF and PACF plots are provided in supplemental materials (Fig. S1).

2.2 Generating realizations of weather data

Three commonly used stochastic weather generators, CLIGEN 5.3, LARSWG 5 and WeaGETS 1.6 were selected to determine their performance with different numbers of realizations in simulating 50-year time series of precipitation occurrence, daily precipitation depth, Tmax, and Tmin. Different random seed numbers were specified to generate 1, 25, 50 and 100 realizations in CLIGEN and LARSWG while in WeaGETS the random seed changed automatically for each simulation run/realization. The selection of the input parameters was based on previous studies on the three generators. For CLIGEN and LARSWG, the default setup in the generators was used for input parameters. This setup has successfully been used to reproduce the essential statistical characteristics of observed weather data in previous studies (Chen and Brissette 2014a; Kou et al. 2007; Mehan et al. 2017; Semenov et al. 1998). The Fourier series interpolation scheme was used in CLIGEN to generate continuous daily results. For WeaGETS, a common combination of a third-order Markov chain and a mixed exponential distribution was used to simulate precipitation occurrence and depth based on Chen et al. (2014). A conditional scheme was used to simulate Tmax and Tmin. The conditional scheme used to generate Tmax and Tmin is similar to the
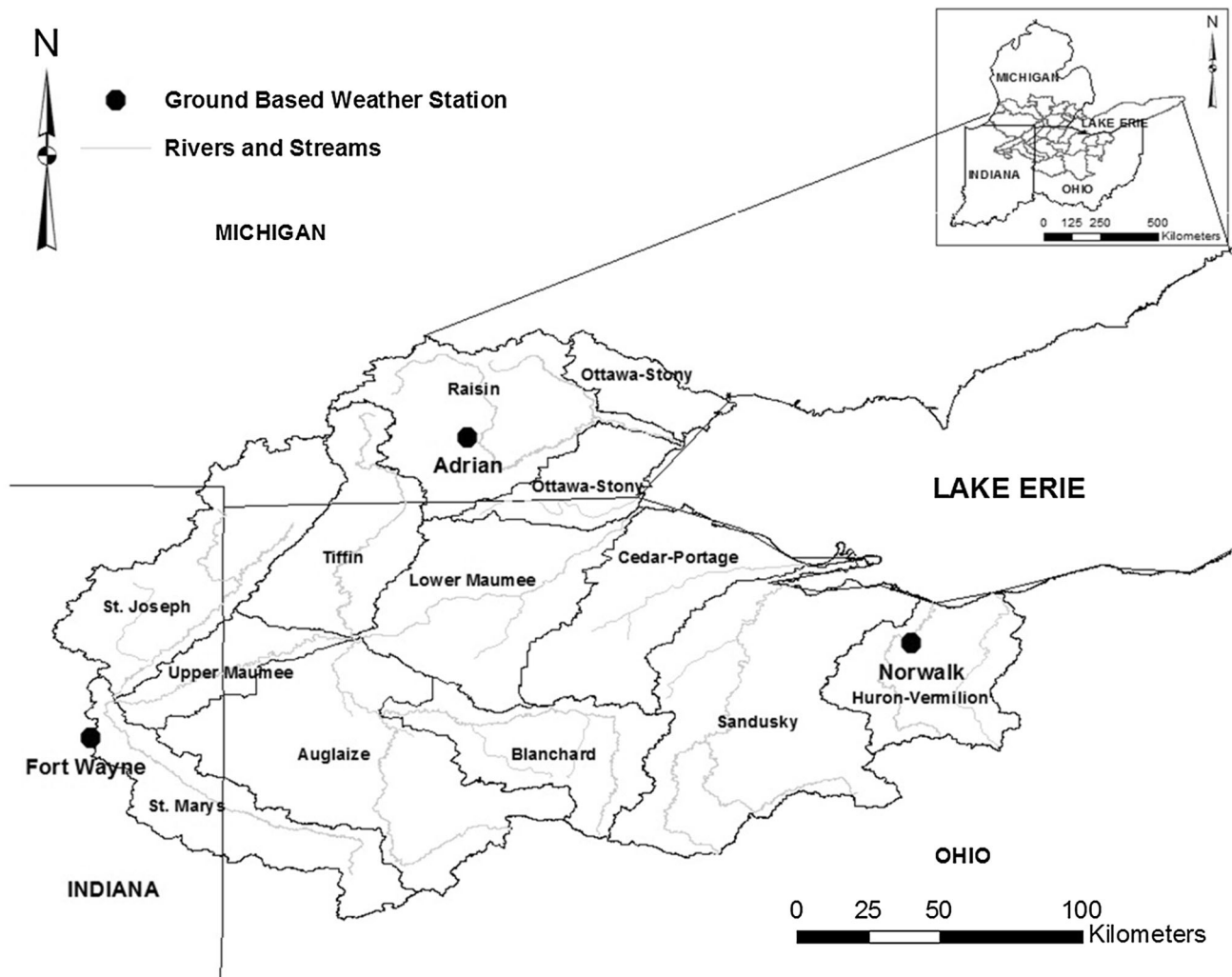
Fig. 1 Western Lake Erie Basin (WLEB) location showing the three selected National Climatic Data Center (NCDC) weather stations used in this study

normal distribution used in CLIGEN and the residual series of Tmax and Tmin conditioned on the wet and dry states are generated by a first-order linear autoregressive model (Chen et al. 2011; Chen and Brissette 2014a). Daily precipitation depths greater than 0.1 mm in CLIGEN or greater than 0.0 mm in LARSWG were used to define a wet day and, hence, to determine precipitation occurrence. A wet-day threshold of 0.1 mm is most commonly used for WeaGETS and was thus used in this study. Fifty years of daily precipitation, Tmax, and Tmin were generated for 1, 25, 50 and 100 realizations from each of the generators.

2.3 Statistical analysis

A rainfall amount of 0.1 mm per day was used as the threshold for wet and dry days since this value is usually used for precision of rain gauges (Dieterichs 1956; Moon

et al. 1994), which is consistent with the threshold for wet and dry days in CLIGEN and WeaGETS. A wet spell was identified as a period with at least four wet days provided a dry day did not occur within the first three days (Bai et al. 2007). A dry spell was identified as a period of at least 15 consecutive dry days (daily precipitation depth less than 0.1 mm) (Douguedroit 1987). Detailed analysis of temperature values was conducted considering 0 and 32 °C as extremes; corn (Zea mays), the primary crop grown in the study area, usually begins to be stressed when air temperature exceeds 32 °C during the tasseling-silking and grainfill stages under rainfed conditions, and cannot survive when air temperature is below freezing (0 °C) (Neild and Newman 1987). Table 1 provides a summary of statistical characteristics calculated for the observed and simulated data for 1, 25, 50 and 100 realizations.

Table 1 Summary of statistical characteristics calculated for the observed and simulated data for 1, 25, 50 and 100 realizations

Description/association	Statistical characteristic
Precipitation occurrence	Number of wet and dry spells
	Number of wet and dry days per year
	Percentage of no rainfall days (daily precipitation depth of 0.0 mm)
Precipitation depth variable	99th percentile of daily precipitation
Extreme temperature events variables	Number of days with Tmax greater than 32 °C
	Number of days with Tmin less than 0 °C
Lower frequency behavior	Interannual and decadal variability for yearly precipitation
	Interannual and decadal variability for average Tmax and Tmin during winter and summer

Boxplots of the aforementioned variables for 25, 50 and 100 realizations from the three generators were used to illustrate the spread of these variables for different realizations. Additionally, 95% confidence intervals (CI) were constructed for bias between the variables of the observed and simulated data, to indicate how biased/good the simulated results were and give an indication of the chance that the model captured the characteristics of the observed data.

Probability density plots were used to check for differences in distributions of daily Tmax and Tmin for 1, 25, 50 and 100 realizations and the observed data from the three generators. The Cohen's Effect Size (Cohen's d) was used to test the equality of means for the observed and simulated daily precipitation, Tmax, and Tmin. This statistic is suitable for large datasets (Cohen 1977, 1992; Bradley 1980; Royall 1986; Denis 2003) and has been used successfully in evaluating weather generator effectiveness (Mehan et al. 2017). A nonparametric Kolmogorov–Smirnov (K–S) test was used to test for equality of the distributions of observed and simulated daily precipitation, Tmax, and Tmin. All tests were two-tailed with a significance level of $p = 0.05$ consistent with Semenov et al. (1998) and Wilks (1999)'s studies. To investigate the reliability of the simulated data, cumulative probability plots for percent error between the statistical characteristics of the simulated data for 25, 50 and 100 realizations from three weather generators at three stations and those of the observed data were plotted. The probability of percent error between the essential statistical characteristics of the simulated data for 25, 50 and 100 realizations from the three generators and those of the observed data between -5 and 5% were then calculated (Bright et al. 2015).

3 Results

3.1 Precipitation occurrence

3.1.1 Wet and dry spells

The number of wet spells for 1, 25, 50 and 100 realizations as generated by CLIGEN were overestimated and had systematic bias compared to the observed data at the three stations (Tables 2 and 3). The number of wet spells were reproduced reasonably well for 25, 50 and 100 realizations from LARSWG and WeaGETS at the Fort Wayne and Norwalk stations, but underestimated at Adrian (Tables 2, 3). Generally, the number of dry spells was overestimated in all realizations from all three generators at Fort Wayne and Norwalk, except that the number of dry spells was slightly underestimated for 25, 50 and 100 realizations from LARSWG at Norwalk (Tables 2, 3). The number of dry spells was overestimated for any realization from CLIGEN at Fort Wayne and Norwalk, but underestimated at Adrian. The number of dry spells was overestimated for 25, 50 and 100 realizations from LARSWG and WeaGETS at Fort Wayne and Norwalk, but reproduced well at Adrian (Tables 2, 3). Based on the results, the chance to capture the number of wet spells of the observed data in LARSWG and WeaGETS, and to capture the number of dry spells of the observed data in LARSWG improved when the number of realizations was increased from 1 to 25, 50 or 100. Details of each realization are as shown in Figures S2 and S3 included with supplemental materials. Overall, LARSWG showed the best performance in simulating the number of wet and dry spells, followed by WeaGETS irrespective of the number of the realizations (Tables 2, 3). CIs of bias between the number of wet and dry spells of the simulated data and those of the observed data for 25 realizations were similar to those for 50 and 100 realizations from each generator, respectively (Table 2).

Table 2 The 95% confidence intervals (lower limit, upper limit) of bias[†] between statistical characteristics of the simulated data and those of the observed data. All characteristics are aggregated over the 50-year simulation period for each realization

G*	R	Wet spell	Dry spell	Number of wet days per month	Number of dry days per month	Percentage of no rainfall days (%)	99th percentile precipitation	Number of days Tmax > 32 °C	Number of days Tmin < 0 °C
Fort Wayne									
C	25	58, 64	6, 8	− 2, − 0.1	0.1, 2	− 0.04, 0.1	− 1.61, − 1.20	8, 22	48, 77
	50	62, 66	4, 6	− 2, − 0.6	0.6, 2	− 0.003, 0.1	− 1.32, − 1.02	8, 19	62, 79
	100	62, 65	5, 7	− 2.0, − 1.0	1, 2	0.02, 0.1	− 1.28, − 1.08	10, 17	61, 74
L	25	− 11, − 0.3	8, 12	− 35, − 31	22, 27	1.4, 1.8	− 0.05, 0.86	− 373, − 352	221, 256
	50	− 8, − 0.4	14, 18	− 33, − 30	21, 24	1.4, 1.6	0.16, 0.71	− 372, − 359	316, 339
	100	− 7, − 2	11, 13	− 33, − 31	23, 25	1.5, 1.7	0.32, 0.70	− 364, − 355	226, 245
W	25	− 11, − 3	7, 15	− 23, − 16	15, 22	− 0.1, 0.3	− 0.86, − 0.11	275, 328	− 36, 26
	50	− 9, − 3	9, 13	− 21, − 18	17, 20	0.01, 0.3	− 0.63, − 0.13	277, 315	− 27, 15
	100	− 9, − 4	10, 13	− 20, − 18	17, 19	0.01, 0.2	− 0.54, − 0.21	281, 304	− 19, 15
Adrian									
C	25	28, 33	− 11, − 7	− 3, − 2	2, 3	3.6, 3.7	− 1.72, − 1.32	22, 37	− 87, − 57
	50	29, 32	− 11, − 9	− 3, − 2	2, 3	3.7, 3.8	− 1.83, − 1.55	28, 39	− 80, − 61
	100	30, 32	− 11, − 9	− 3, − 2	2, 3	3.7, 3.7	− 1.69, − 1.48	30, 37	− 81, − 69
L	25	− 24, − 12	− 4, 1	− 9, − 5	− 1, 3	0.2, 0.5	0.45, 1.05	− 203, − 186	198, 237
	50	− 20, − 12	− 5, − 1	− 8, − 5	− 1, 2	0.2, 0.4	0.65, 1.11	− 199, − 186	205, 232
	100	− 20, − 15	− 4, − 1	− 7, − 5	0, 2	0.3, 0.4	0.78, 1.10	− 197, − 187	208, 227
W	25	− 14, − 3	− 6, 0	− 11, − 5	5, 11	− 0.1, 0.3	− 0.34, 0.52	199, 238	− 248, − 176
	50	− 11, − 3	− 4, 2	− 11, − 7	7, 11	0.0, 0.3	− 0.06, 0.45	233, 261	− 267, − 223
	100	− 11, − 6	− 4, − 1	− 9, − 6	6, 9	0.0, 0.2	− 0.03, 0.32	220, 240	− 234, − 208
Norwalk									
C	25	43, 49	9, 13	− 3, − 1	1, 3	0.3, 0.5	− 2.40, − 1.87	3, 15	− 178, − 151
	50	46, 50	8, 10	− 3, − 2	2, 3	0.4, 0.4	− 1.91, − 1.60	4, 15	− 167, − 151
	100	47, 49	8, 10	− 3, − 2	2, 3	0.4, 0.4	− 1.97, − 1.73	6, 13	− 168, − 158
L	25	− 9, 3	1, 5	− 31, − 27	9, 13	0.7, 1.0	0.25, 0.85	− 240, − 222	158, 187
	50	− 8, − 1	1, 3	− 30, − 27	9, 12	0.7, 0.9	0.17, 0.69	− 234, − 222	163, 183
	100	− 10, − 4	2, 4	− 30, − 28	10, 12	0.8, 0.9	0.29, 0.64	− 230, − 222	168, 183

Table 2 continued

G*	R	Wet spell	Dry spell	Number of wet days per month	Number of dry days per month	Percentage of no rainfall days (%)	99th percentile precipitation	Number of days Tmax > 32 °C	Number of days Tmin < 0 °C
W	25	− 11, 1	11, 17	− 12, − 7	7, 12	0.0, 0.3	0.25, 0.97	287, 338	− 198, − 134
	50	− 11, 4	12, 14	− 13, − 9	9, 13	0.1, 0.3	0.18, 0.75	308, 337	− 202, − 161
	100	− 11, − 5	13, 16	− 14, − 10	10, 14	0.2, 0.4	0.40, 0.72	310, 328	− 206, − 172

*G generator, R realizations, C CLIGEN, L LARSWG, W WeaGETS

[†]Bias, statistical characteristic of the simulated data- corresponding statistical characteristic of the observed data. Bias between statistical characteristics of generated daily values have a 95% chance of falling in the CI represented in the table

Table 3 The second and third quartiles (25th, 75th) of extreme precipitation events for the simulated data from 25, 50 and 100 realizations and corresponding value of extreme events for the observed data and the simulated data from 1 realization. Values are aggregated over the 50-year simulation period

Realizations	Wet spell			Dry spell			99th percentile precipitation (mm)		
	C*	L	W	C	L	W	C	L	W
Fort Wayne									
Observed	213	213	213	18	18	18	33.7	33.7	33.7
1	277	199	222	24	40	32	32.3	34.6	32.9
25	267, 279	196, 219	198, 212	23, 26	25, 32	24, 32	32.0, 32.6	33.8, 34.5	32.5, 34.0
50	271, 281	198, 216	198, 212	21, 25	25, 32	22, 32	32.3, 32.9	33.8, 34.5	32.8, 33.9
100	271, 281	199, 217	198, 215	22, 26	27, 33	24, 33	32.2, 32.9	33.7, 34.8	32.9, 33.9
Adrian									
Observed	194	194	194	48	48	48	32.6	32.6	32.6
1	231	162	189	35	56	40	31.3	32.9	32.7
25	222, 227	166, 187	178, 192	37, 41	43, 50	40, 50	30.8, 31.3	32.9, 34.0	32.0, 33.4
50	222, 228	168, 188	179, 199	34, 40	39, 50	40, 51	30.6, 31.2	32.9, 34.1	32.2, 33.6
100	222, 229	170, 184	175, 195	36, 41	42, 50	40, 49	30.7, 31.4	33.0, 34.2	32.2, 33.4
Norwalk									
Observed	235	235	235	17	17	17	32.5	32.5	32.5
1	285	234	219	24	14	27	29.8	33.7	33.4
25	278, 284	224, 241	219, 241	24, 29	17, 25	26, 35	30.0, 30.8	32.6, 33.4	32.4, 33.8
50	279, 288	222, 234	216, 235	24, 28	16, 23	27, 34	30.3, 31.2	32.4, 33.7	32.3, 33.7
100	279, 287	219, 237	215, 238	24, 28	17, 23	27, 33	30.2, 31.3	32.5, 33.6	32.5, 33.6

*C CLIGEN, L LARSWG, W WeaGETS. The observed data used as input in CLIGEN included leap days whereas those used in LARSWG and WeaGETS did not

3.1.2 Number of wet and dry days

One, 25, 50 and 100 realizations from CLIGEN reproduced the number of wet and dry days per year reasonably well at Fort Wayne (Figs. 2a, 3a), while slightly underestimating the number of wet days per year (Fig. 2b, c) and overestimating dry days per year (Fig. 3b, c) at Adrian and Norwalk. For all realizations, LARSWG and WeaGETS underestimated the number of wet days per year (Fig. 2),

and overestimated the number of dry days per year at the three stations (Fig. 3), with all having systematic bias compared to the observed data (Table 2). However, LARSWG reproduced the number of dry days per year well at Adrian (Fig. 3b). Overall, CLIGEN had a greater chance to capture the number of wet and dry days per year observed, when the number of realizations was increased from 1 to 25, 50, or 100, whereas the number of realizations was immaterial for LARSWG and WeaGETS at Fort

(a) Fort Wayne, IN

(b) Adrian, MI

Number of wet days in an year

(c) Norwalk, OH
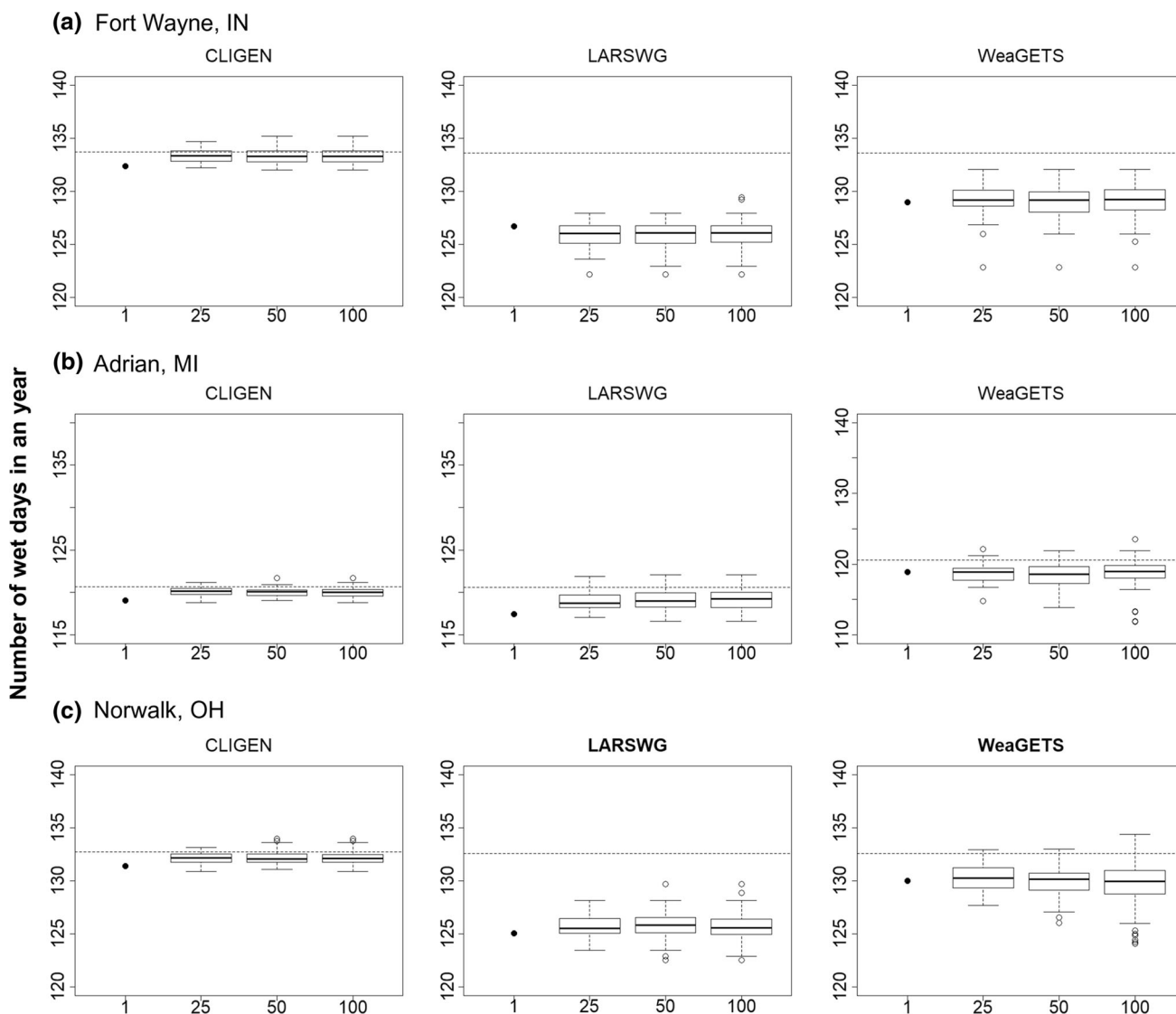
Fig. 2 Number of wet days per year calculated from generated precipitation showing the results for 1 realization (dots) and boxplots for 25, 50 and 100 realizations compared to the number calculated from observed data (straight lines) in the period 1966–2015 at the three selected stations

Wayne station as, in general, these generators could not capture the number of wet and dry days per year observed at Fort Wayne station regardless of the number of realizations. CIs of bias between the number of wet and dry days per month of the simulated data and those of the observed data for 25 realizations were similar to those for 50 and 100 realizations from each generator, respectively (Table 2).

3.1.3 Percentage of no rainfall days

One, 25, 50 and 100 realizations from WeaGETS reproduced the percentage of no rainfall days reasonably well at three stations (Fig. 4), while those from CLIGEN and LARSWG were slightly overestimated or underestimated at the three stations, and had a systematic bias with the

observed data (Fig. 4 and Table 2). All three generators had a greater chance to capture the percentage of no rainfall days as obtained from the observed data, with the chances being improved when the number of realizations was increased from 1 to 25, 50 or 100 at the three stations. Generally, 25, 50, and 100 realizations from the three generators could capture the percentage of no rainfall days fairly well at the three stations (Table 2). However, LARSWG could not capture the percentage of no rainfall days of the observed data regardless of the number of realizations at Fort Wayne (as detailed in Figure S4). CIs of bias between the percentage of no rainfall days of the simulated data and those of the observed data for 25 realizations were similar to those for 50 and 100 realizations from each generator (Table 2).
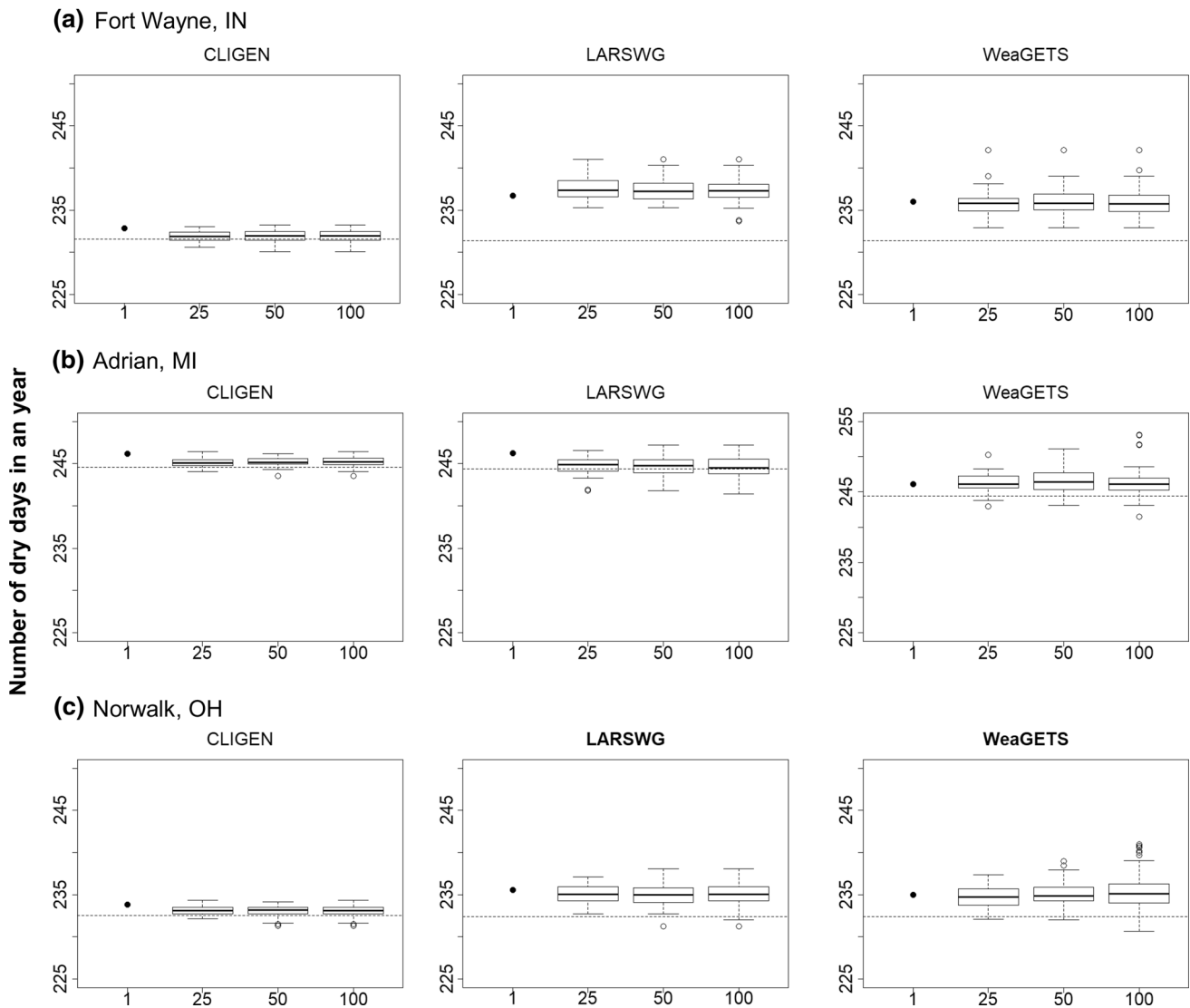
Fig. 3 Number of dry days per year calculated from generated precipitation showing the results for 1 realization (dots) and boxplots for 25, 50 and 100 realizations compared to the number calculated from observed data (straight lines) in the period 1966–2015 at the three selected stations
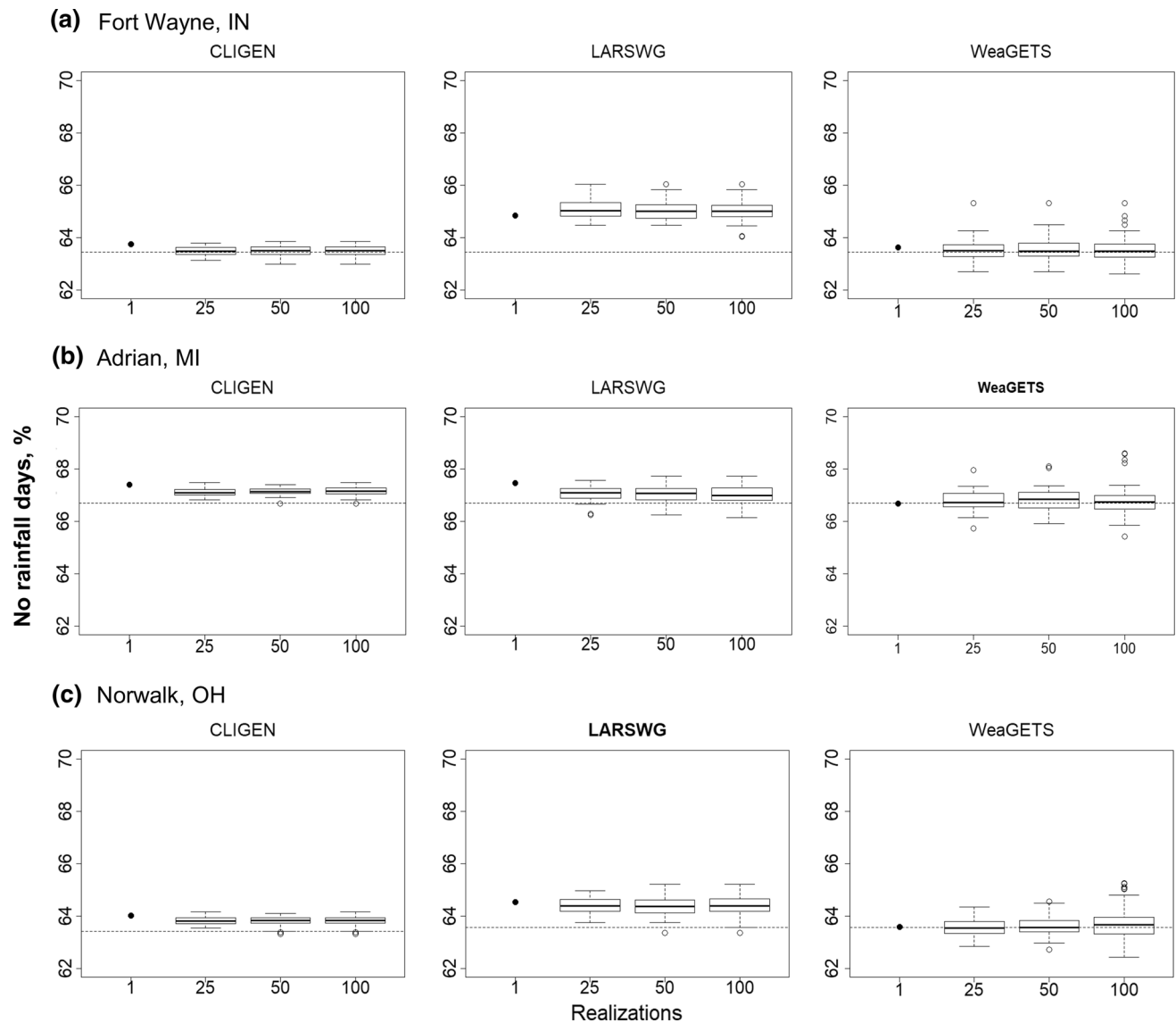
3.2 Precipitation depth

3.2.1 99th percentile of daily precipitation depth

For all realizations, the 99th percentile of daily precipitation depth was underestimated by CLIGEN at the three stations (Table 3 and as detailed in Figure S5). The 99th percentile daily precipitation depths for 1, 25, 50 and 100 realizations from WeaGETS were reproduced reasonably well at Fort Wayne and Adrian (Table 3), but slightly overestimated at Norwalk (Table 3). The 99th percentile daily precipitation depths for 1, 25, 50 and 100 realizations from LARSWG were slightly overestimated at Fort Wayne, Adrian and Norwalk (Table 3). WeaGETS and LARSWG had a chance to capture the 99th percentile of the daily precipitation depth of the observed data when the

number of realizations was increased from 1 to 25, 50 or 100. CIs of bias between the 99th percentile of daily precipitation depth of the simulated data and those of the observed data for 25 realizations were similar to those for 50 and 100 realizations from each generator (Table 2). The 25th and 75th quartile of the 99th percentile of daily precipitation depth from 25 realizations were similar to those from 50 and 100 realizations, and the spread for the 99th percentile of daily precipitation depth did not change a lot when the realization number increased from 25 to 50 and 100 (Table 3).

The null hypothesis that there were no differences in distributions between generated daily precipitation and observed data was rejected by K–S tests for all three generators (p value < 0.05). However, the K–S test may be biased as it tends to be excessively stringent for very large

(a) Fort Wayne, IN

(b) Adrian, MI

No rainfall days, %


(c) Norwalk, OH

Realizations

Fig. 4 Percentage of no rainfall calculated from generated precipitation showing the results for 1 realization (dots) and boxplots for 25, 50 and 100 realizations compared to the percentage calculated from observed data (straight lines) in the period 1966–2015 at the three selected stations

sample sizes (Zhang and Garbrecht 2003; Zhang 2013; Chen and Brissette 2014a), as was the case in this study. Cohen's d values provide a more robust measure than K–S tests for large datasets (Cohen 1977; Bradley 1980; Royall 1986; Denis 2003). In this study, Cohen's d values for daily precipitation at all three stations were small (< 0.2), indicating small differences between means of the simulated and observed daily precipitation (Cohen 1977, 1992).

3.2.2 Interannual and decadal variability of total precipitation amounts

One, 25, 50 and 100 realizations from three generators underestimated interannual and decadal SD of total precipitation amounts at Fort Wayne and decadal SD of

precipitation amounts at Norwalk, while somewhat overestimating interannual SD of total precipitation amounts at Adrian (Fig. 5). CLIGEN somewhat underestimated decadal SD of total precipitation amounts at Adrian, and annual SD of total precipitation amounts at Norwalk (Fig. 5). LARSWG and WeaGETS reproduced decadal SD of total precipitation amounts reasonably well at Adrian, and interannual SD of total precipitation amounts reasonably well at Norwalk (Fig. 5). All three generators had a chance to capture interannual and decadal SD of total precipitation amounts of the observed data when the number of realizations was increased from 1 to 25, 50 or 100 at the three stations, except that CLIGEN underestimated decadal SD of total precipitation amounts of the observed data regardless of the number of realizations
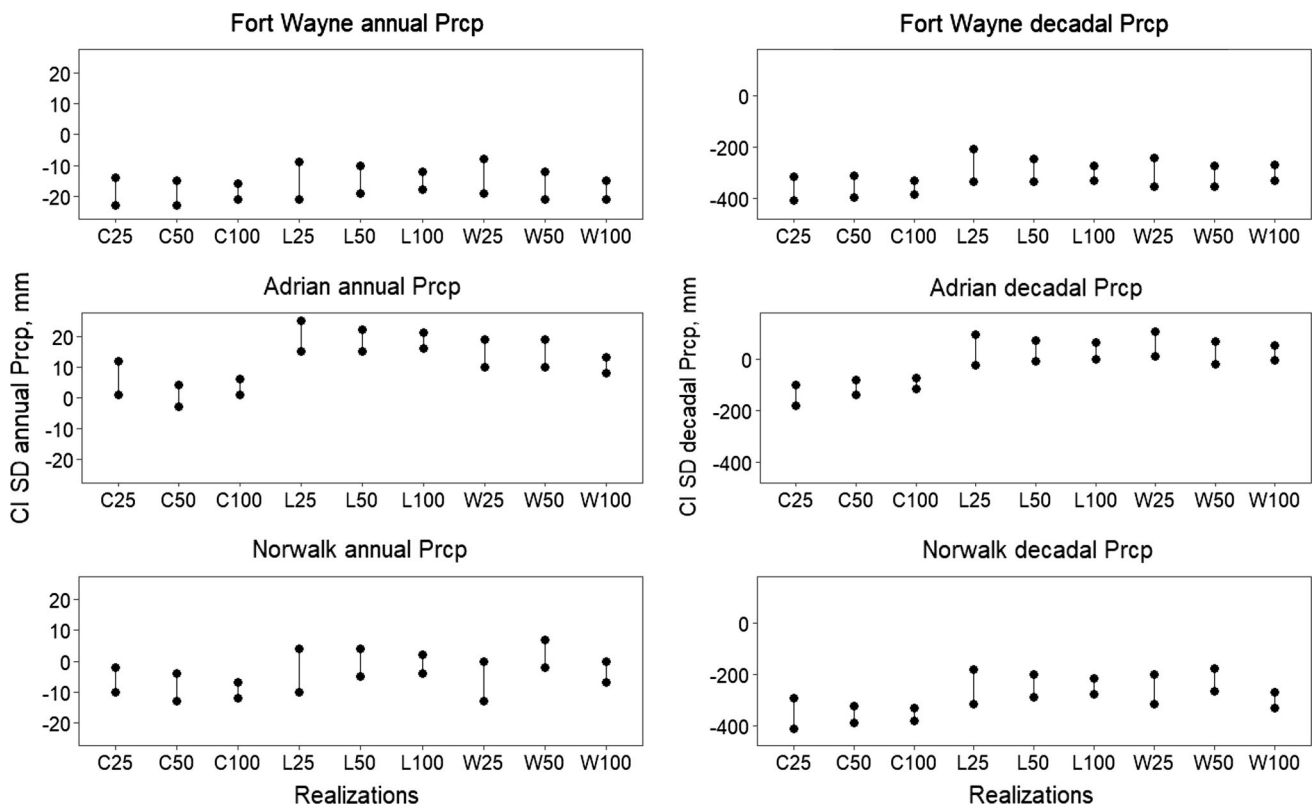
Fig. 5 The 95% confidence intervals (CIs) of bias between standard deviation of interannual and decadal total precipitation amounts of the simulated data and those of the observed data. The lines with dots represent confidence intervals. C25, C50, and C100 represent 25, 50, and 100 realizations from CLIGEN and similarly for LARS-WG (L) and WeaGETS (W). SD and Prcp represent standard deviation and total precipitation, respectively

(Fig. 5). CIs of bias between interannual and decadal SD of the total precipitation amounts of the simulated data and those of the observed data for 25 realizations were similar to those for 50 and 100 realizations from each generator (Fig. 5).

3.3 Tmax and Tmin

3.3.1 Distributions of Tmax and Tmin

Generally, the distributions of daily Tmax and Tmin for 1, 25, 50 and 100 realizations from the three generators were similar to those of the observed data at three stations, The distributions of daily Tmax and Tmin for 50 realizations and for the observed data at three stations have been presented in a previou study (Mehan et al. 2017) The trend of the distribution of daily Tmax (Tmin) was similar for all three generators (Mehan et al. 2017). LARSWG performed best in simulating distribution of daily Tmax and Tmin for 1, 25, 50 and 100 realizations at three stations; the probability density of Tmax between 0 and 5 °C, and Tmin between − 5 and 0 °C for 1, 25, 50 and 100 realizations for LARSWG matched well with those of the observed data, while CLIGEN and WeaGETS did not capture these well

consistent with Mehan et al. (2017)'s study. The K–S tests for daily Tmax and Tmin for all realizations from the three generators all rejected the null hypothesis that there was no significant difference between the distribution of generated daily Tmax and Tmin and those of the observed data ($p < 0.05$). Cohen's d values for daily Tmax (Tmin) at all three stations were small (< 0.2), indicating small differences between means of the simulated and observed daily Tmax (Tmin).

3.3.2 Extreme temperature events

One, 25, 50 and 100 realizations from CLIGEN and WeaGETS somewhat overestimated the number of days with Tmax greater than 32 °C (Tables 2, 4), while those from LARSWG were underestimated at three stations (Table 4). CLIGEN had a chance to capture the number of days with Tmax greater than 32 °C of the observed data, when the number of realizations was increased from 1 to 25, 50 or 100 whereas LARSWG and WeaGETS could not for any of the realizations at three stations. Details of each realization are shown in Figure S6.

The number of days with Tmin less than 0 °C for 1, 25, 50 and 100 realizations from LARSWG were

Table 4 The second and third quartiles (25th, 75th) of extreme temperature events for the simulated data from 25, 50 and 100 realizations and corresponding value of extreme events for the observed data and the simulated data from 1 realization. Values are aggregated over the 50-year simulation period

Realizations	Number of days with Tmax > 32 °C			Number of days with Tmin < 0 °C		
	C	L	W	C	L	W
Fort Wayne						
Observed	752	752	752	5940	5930	5930
1	782	393	949	6039	6104	6185
25	757, 779	371, 403	1003, 1081	5972, 6020	6136, 6196	5889, 5969
50	753, 780	374, 405	1004, 1081	5992, 6024	6144, 6196	5889, 5971
100	752, 778	381, 405	1007, 1083	5990, 6023	6131, 6197	5875, 5997
Adrian						
Observed	613	613	613	6975	6964	6964
1	663	439	947	6923	7146	6772
25	635, 637	409, 434	805, 859	6875, 6928	7146, 7213	6697, 6792
50	637, 658	407, 439	825, 898	6877, 6934	7148, 7215	6671, 6759
100	634, 658	408, 436	812, 871	6881, 6922	7150, 7213	6695, 6786
Norwalk						
Observed	625	625	625	6320	6309	6309
1	662	404	990	6174	6461	6080
25	625, 644	381, 405	878, 976	6132, 6184	6461, 6512	6116, 6200
50	624, 648	383, 413	915, 987	6142, 6182	6460, 6512	6082, 6169
100	625, 647	387, 413	907, 975	6141, 6175	6459, 6511	6074, 6172

*C CLIGEN, L LARSWG, W WeaGETS. The observed data used as input in CLIGEN included leap days whereas those used in LARSWG and WeaGETS did not

overestimated at all three stations (Tables 2, 4). This number of days was overestimated for the 1, 25, 50 and 100 realizations from CLIGEN at Fort Wayne, but underestimated at Adrian and Norwalk. The number of days was reproduced reasonably well for the ensemble of 25, 50 and 100 realizations from WeaGETS at Fort Wayne, but underestimated at Adrian and Norwalk (Tables 2, 4). CIs of bias between the number of days with Tmax greater than 32 °C and Tmin less than 0 °C of the simulated data and those of the observed data for 25 realizations were similar to those for 50 and 100 realizations from each generator, respectively (Table 2). Details are as shown in Figure S7. The 25th and 75th quartile of Tmax greater than 32 °C and Tmin less than 0 °C from 25 realizations were similar to those from 50 and 100 realizations, and the spread for Tmax greater than 32 °C and Tmin less than 0 °C did not change much when the realization number increased from 25 to 50 and 100 (Table 4).

3.3.3 Interannual and decadal variability of Tmax and Tmin

One, 25, 50 and 100 realizations from CLIGEN and LARSWG underestimated interannual and decadal SD of average Tmax and Tmin during winter and summer at three stations. CIs of bias between interannual SD of average Tmax during winter simulated from CLIGEN and LARSWG for 25, 50 and 100 realizations and those from

the observed data all ranged between − 1.3 and − 1.2 °C. However, 25, 50 and 100 realizations from WeaGETS somewhat overestimated interannual SD of average Tmin during summer at Fort Wayne (CIs of bias from 0.1 to 0.2 °C), and decadal SD of average Tmax during summer at Adrian (CIs of bias from 0.0 to 0.1 °C). WeaGETs realizations were able to reproduce decadal SD of average Tmax (CIs of bias from − 0.1 to 0.0 °C), Tmin (CIs of bias from 0.0 to 0.1 °C) during summer at Fort Wayne, interannual SD of average Tmin during summer at Adrian (CIs of bias from − 0.1 to 0.0 °C), interannual SD of average Tmin during summer (CIs of bias from − 0.1 to 0.0 °C), and decadal Tmax during summer (CIs of bias from − 0.1 to 0.0 °C) at Norwalk reasonably well. WeaGETS had a chance to capture interannual and decadal SD of average Tmax and Tmin during winter and summer of the observed data when the number of realizations was increased from 1 to 25, 50 or 100 at the three stations, except for interannual SD of average Tmax during winter (CIs of bias from − 0.5 to − 0.3 °C), decadal SD of average Tmin during winter at Adrian (CIs of bias from − 0.7 to − 0.6 °C), and decadal SD of average Tmin during summer at Norwalk (CIs of bias from − 0.6 to − 0.5 °C). CLIGEN and LARSWG could not capture interannual and decadal SD of average Tmax and Tmin during winter and summer at three stations regardless of the number of realizations.

3.3.4 Assessment of 10 realizations

Based on the results as discussed, performance of 25 realizations did not seem appreciably different than that of 50 or 100 realizations. Thus, there did not seem to be an advantage of generating more than 25 realizations. To determine whether there was an advantage to generating 25 realizations, ten realizations of generated daily results for CLIGEN, LARSWG and WeaGETS were analyzed (Fig. 6). This was done to determine whether the essential statistical characteristics of the observed data could be captured just as well with ten realizations. The aforementioned statistical characteristics that could not be captured by any realization were not taken into consideration. For the ten realizations: CLIGEN could not capture days with Tmax higher than 32 °C of the observed data at three station, the 99th percentile of daily precipitation depth of

the observed data at Fort Wayne, the number of dry days per month of the observed data at Adrian, or the number of wet days per month of the observed data at Norwalk; WeaGETS could not capture the 99th percentile of the daily precipitation depth of the observed data at Fort Wayne, the number of dry spells at Adrian, or the number of wet days per month of the observed data at Norwalk (Fig. 6). However, these characteristics were captured when the number of realizations was increased to 25, 50 or 100, as previously discussed.

4 Discussion

Generally, when the number of realizations was increased from 1 to 25, 50 or 100: all three generators had a higher chance to capture the percentage of no rainfall days and
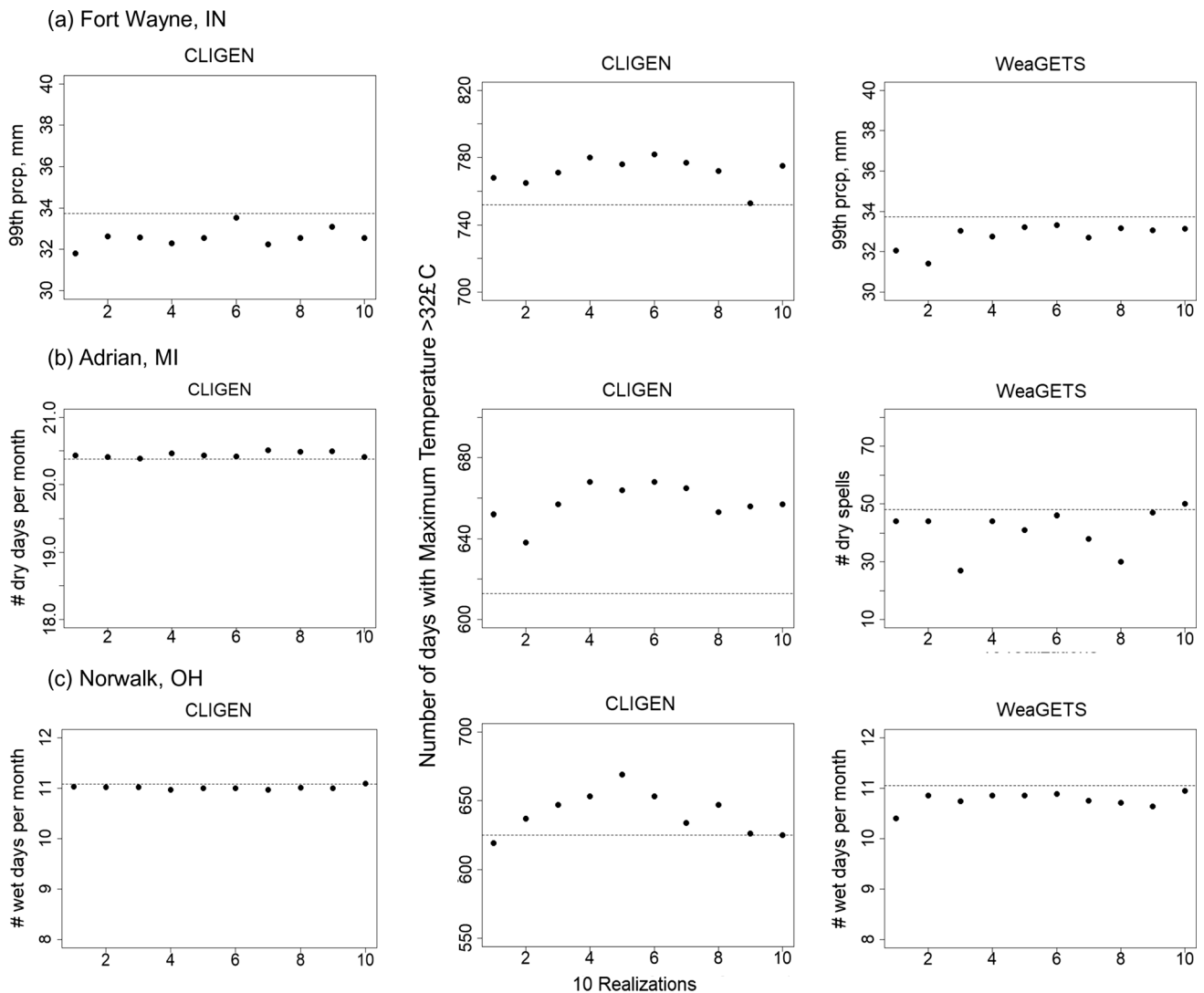
Fig. 6 Statistics of the generated daily precipitation, maximum temperature (Tmax), and minimum temperature (Tmin) results for 10 realizations from the three generators compared to that of the observed data from 1966 to 2015 at the three selected stations

had a chance to capture interannual and decadal SD of total precipitation amounts of the observed data; CLIGEN had a higher chance to capture the average number of wet and dry days per month, and the number of days with Tmax greater than 32 °C; LARSWG and WeaGETS had a chance to capture the number of wet spells of the observed data; and, WeaGETS had a higher a chance to capture the 99th percentile of the daily precipitation depth and the interannual and decadal SD of average Tmax and Tmin during winter and summer months. The probability of percent error between the essential statistical characteristics of the simulated data for 25 realizations from the three generators and those of the observed data between − 5 and 5% were similar to those for 50 and 100 realizations (Table S1). For example, the probability of percent error between the number of days with Tmax greater than 32 °C of the simulated data for 25, 50, 100 realizations from CLIGEN at Fort Wayne and those of the observed data were 0.96, 0.96 and 0.95, respectively (Table S1). The probability of percent error between the number of dry spells of the simulated data for 25, 50, 100 realizations from LARSWG at Adrian and those of the observed data were 0.36, 0.26 and 0.21, respectively (Table S1). These were consistent with the cumulative probability plots for percent error between the essential statistical characteristics of the simulated data for 25, 50 and 100 realizations from three generators and those of the observed data at the three stations (Figure S8). Distributions of daily Tmax and Tmin for 1, 25, 50 and 100 realizations from the three generators were similar to those of the observed data. The trend of the distribution of daily Tmax (Tmin) did not change substantially when the number of realizations was increased from 1 to 25, 50 and 100 for all generators at the three stations.

In some cases, discrepancies between statistical characteristics of the simulated and observed data were mainly caused by the generator itself rather than the number of realizations. Every generator has its pros and cons. By incorporating a semi-empirical distribution to account for and quantify precipitation depths, and a first order linear auto regressive model to preserve auto- and cross- correlations (Srikanthan and McMahon 2001; Figueiredo Filho et al. 2013), LARSWG showed the best performance in simulating the number of wet and dry sequences and distribution of daily Tmax and Tmin. This generator could not, however, reasonably reproduce the average number of wet and dry days per month, nor the number of days with Tmax greater than 32 °C or Tmin less than 0 °C. CLIGEN uses two random variables to preserve the auto- and cross-correlations for and between Tmax and Tmin, which might result in discrepancies causing it to reproduce numbers of wet and dry sequences that are overestimated (Koutsoyiannis and Manetas 1996; Mehan et al. 2017). The third-order Markov chain-based model in WeaGETS

resulted in better estimates of wet and dry sequences than the first-order Markov chain-based model in CLIGEN, consistent with Chen and Brissette (2014a). Simulated precipitation occurrences from second- and third-order Markov chain-based models are slightly better than those simulated by the first-order Markov chain-based model, especially for long wet and dry spells Additionally, Markov chain-based models have been found to underestimate extreme dry spells (Chen and Brissette 2014a). The mixed exponential distribution represents extreme precipitation events relatively well whereas the gamma distribution underestimates these extremes (Wilks 1999). Both distributions are built into WeaGETS. The choice of a distribution function should, therefore, depend to the characteristics of observed precipitation which are location- and even season-specific.

Without consideration of the low-frequency component of climate data variability, CLIGEN and LARSWG both underestimated interannual and decadal variances of precipitation, Tmax, and Tmin. In contrast, WeaGETS preserves the autocorrelation and uses a spectral correction approach to correct the underestimated frequency variability of climate data at monthly and interannual levels (Chen et al. 2010); this generator, thus, showed the best performance in simulating interannual and decadal SD of average Tmax and Tmin during winter and summer at all three stations. Additionally, the mixed exponential distribution in WeaGETS was better at representing extreme precipitation events than the skewed normal distribution in CLIGEN and the semi-empirical distribution in LARSWG.

An alternative approch to significance testing, Cohen's Effect Size, provides a robust measure for large datasets (Cohen 1977; Bradley 1980; Royall 1986; Denis 2003). Moreover, a 95% CI can illustrate how biased/good the generated values are and the chance that the model will capture the characteristics of the observed data. As the true value may be more readily captured within a CI, it is often better to examine the CI of bias between the simulated and observed value for various realizations than base the evaluation on a single number. In this study, the 95% CIs of bias between the aforementioned statistical characteristics of generated data from the three generators and those of the observed data did not change appreciably when the realization number was increased from 25 to 50 or 100.

Overall, results showed that one realization from CLIGEN, LARSWG and WeaGETS might not be representative enough to capture characteristics of the observed data, since these climate generators are stochastic. In order to reduce sampling error caused by randomness of the weather generator process, only one realization with results generated for a much longer synthetic series than the observed data have been widely used in previous studies (Wilks 2002; Zhang and Garbrecht 2003; Chen et al.

2010, 2011, 2012a, b; Eames et al. 2012; Chen and Brissette 2014a, b). However, generating time series that are longer than the observed data could result in bias by insufficient sampling of distribution and cause inflated probability and distorted distribution (Mithen and Black 2011). In contrast, generating time series with the same length or shorter than the observed data could avoid biases caused by fitting a theoretical distribution to the observed data (Semenov and Barrow 1997; Mithen and Black 2011).

An ensemble of multiple realizations could incorporate varying parameter sets for different realizations, rather than carry forward one single parameter set as included in one realization. For example, WeaGETS incorporates various sub-models of first-, second-, and third-order of Markov chain-based models for determination of precipitation occurrence, gamma, exponential, mixed exponential, and skewed normal distributions for precipitation depth calculation, an option for smoothing precipitation parameters and low frequency correction, and conditional and unconditional schemes to simulate temperature—different combinations of which could be considered to generate weather data for multiple realizations. An ensemble of multiple realizations would represent model randomness and quantify a sampling variability in simulated results consistent with Kalnay et al. (2006).

A wide range of probability intervals can be reproduced using multiple realizations. Where a pairwise relationship exists or is assumed to exist between simulated and observed data, simulations can be evaluated using accuracy plots and other probability assessments as detailed in Jolliffe and Stephenson (2012). Such a relationship was not assumed in this study, given that time series produced by weather generators are not necessarily matched in a pairwise fashion with the base input data (Gitau et al. 2017). The probability of percent error between the statistical characteristics of the simulated and observed data between − 5 and 5%, and the cumulative probability plots of percent error between the statistical characteristics of the simulated and observed data for 50 and 100 realizations from three generators at three stations were similar to those for 25 realizations (Table S1, Figure S8).

Generally, the advantage of generating more than 25 realizations of weather data was not obvious from the analysis; while improvements were observed when the number of realizations was increased to 25, statistical characteristics did not improve appreciably when the number of realizations was increased from 25 to 50 or 100. Instead, generated values for 50 and 100 realizations only became more condensed around the values for 25 realizations from each generator. Twenty-five realizations were, however, found to have distinct advantages over 10 realizations, suggesting that 25 could be an optimal number for the realizations. Generating too many realizations is time-consuming and computationally expensive, especially if these realizations are to be applied in hydrological and environmental models.

Biases between statistical characteristics of simulated daily precipitation and temperature data and those of observed data affect hydrologic, water quality, and crop growth model simulations (Hansen and Jones 2000), due to their impacts on soil water balance dynamics, plant water and temperature stresses and soil erosion (Hansen and Ines 2005; Gitau 2016). Thus, regardless of the number of realizations, any weather generator should be tested to ensure that generated data have statistical characteristics similar to those of observed data (Semenov et al. 1998; Wilks and Wilby 1999; Soltani and Hoogenboom 2003).

5 Conclusions

In this study, 1, 25, 50, and 100 realizations of 50 years of daily precipitation, and Tmax and Tmin for the Fort Wayne, Adrian, and Norwalk weather stations in the WLEB were generated from CLIGEN, LARSWG and WeaGETS. The generated results were compared with the observed data and analyzed to investigate how many realizations of generated weather data are needed to capture the statistical characteristics of the climate. Our results supported the hypothesis that one realization of simulated daily precipitation, Tmax and Tmin from stochastic weather generators may not be representative enough to capture essential statistical characteristics of observed weather data. Weather generators produce random results and an ensemble of multiple realizations are required to present the range of variability of statistical characteristics of climate data. Generally, all three generators had a better chance to capture essential statistical characteristics when the number of realizations was increased to 25, 50 or 100. Generally, there was no obvious advantage of generating more than 25 realizations, as results did not improve appreciably beyond that number. Furthermore, increasing the number of realizations did not improve performance if the generator either tended to overestimate or underestimate a particular weather variable by a large margin, as this likely meant that the generator was not able to capture the variable at all. Results provide guidance for selecting the number of realizations to enable more accurate and efficient weather data simulation for use with hydrologic, agricultural, and environmental applications. Results of this study are based on three data-rich weather stations in the WLEB, and might not be directly applicable in data sparse regions or those with arid or semi-arid climates. Likewise, the potential to vary parameters for each realization would be of benefit and, thus, warrants further

study. Statistical methodologies presented herein are, however, widely applicable.

Compliance with ethical standards

Conflict of interest The authors have no conflicts of interest to declare.

References

Bai A, Zhai P, Liu X (2007) Climatology and trends of wet spells in China. Theoret Appl Climatol 88:139–148

Banerjee A, Dolado JJ, Galbraith JW, Hendry D (1993) Co-integration, error correction, and the econometric analysis of non-stationary data. Advanced texts in econometrics. Oxford University Press, Oxford

Bradley JV (1980) Nonrobustness in Z, t, and F tests at large sample sizes. Bull Psychon Soc 16:333–336

Bright J, Smith C, Taylor P, Crook R (2015) Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data. Sol Energy 115:229–242

Caron A, Leconte R, Brissette F (2008) An improved stochastic weather generator for hydrological impact studies. Can Water Resour J 33:233–256

Chaubey I, Chiang L, Gitau MW, Mohamed S (2010) Effectiveness of best management practices in improving water quality in a pasture-dominated watershed. J Soil Water Conserv 65:424–437

Chen J, Brissette FP (2014a) Comparison of five stochastic weather generators in simulating daily precipitation and temperature for the Loess Plateau of China. Int J Climatol 34:3089–3105

Chen J, Brissette FP (2014b) Stochastic generation of daily precipitation amounts: review and evaluation of different models. Clim Res 59:189–206

Chen J, Brissette FP, 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 ASABE 54:1627–1637

Chen J, Brissette F, Leconte R, Caron A (2012a) A versatile weather generator for daily precipitation and temperature. Trans ASABE 55:895–906

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

Chen J, Brissette FP, Zhang XJ (2014) A multi-site stochastic weather generator for daily precipitation and temperature. Trans ASABE 57:1375–1391

Cheung Y-W, Lai KS (1995) Lag order and critical values of the augmented Dickey–Fuller test. J Bus Econ Stat 13:277–280

Cohen J (1977) Statistical power analysis for the behavioural sciences, Rev edn. Academic. New York, NY, New York

Cohen J (1992) Statistical power analysis. Curr Dir Psychol Sci 1:98–101

Daly C, Gibson WP, Taylor GH, Doggett MK, Smith JI (2007) Observer bias in daily precipitation measurements at United States cooperative network stations. Bull Am Meteorol Soc 88:899–912

Denis DJ (2003) Alternatives to null hypothesis significance testing. Theory Sci 4:21

Dieterichs H (1956) Frequency of dry and wet spells in san salvador. Geofisica Pura e Applicata 33:267–272

Douguedroit A (1987) The variations of dry spells in Marseilles from 1865 to 1984. J Climatol 7:541–551

Eames M, Kershaw T, Coley D (2012) A comparison of future weather created from morphed observed weather and created by a weather generator. Build Environ 56:252–264

Elliot W, Arnold C (2001) Validation of the weather generator CLIGEN with precipitation data from Uganda. Trans ASAE 44:53–58

Figueiredo Filho DB, Paranhos R, Rocha ECd, Batista M, Silva Jr JAd, Santos MLWD, Marino JG (2013) When is statistical significance not significant? Braz Polit Sci Rev 7:31–55

Fuller WA (2009) Introduction to statistical time series, vol 428. Wiley, Hoboken

Furrer EM, Katz RW (2008) Improving the simulation of extreme precipitation events by stochastic weather generators. Water Resour Res 44:W12439. https://doi.org/10.1029/2008WR007316

Gitau M (2016) Long-term seasonality of rainfall in the southwest Florida Gulf coastal zone. Clim Res 69:93–105

Gitau MW, Mehan S, Guo T (2017) Weather generator utilization in climate impact studies: implications for water resources modeling. Eur Water. Accepted 05 Sept 2017

Guo T, Engel BA, Shao G, Arnold JG, Srinivasan R, Kiniry JR (2015) Functional approach to simulating short-rotation woody crops in process-based models. Bio Energy Res 8:1598–1613

Hansen JW, Ines AV (2005) Stochastic disaggregation of monthly rainfall data for crop simulation studies. Agric For Meteorol 131:233–246

Hansen J, Jones J (2000) Scaling-up crop models for climate variability applications. Agric Syst 65:43–72

Harmel R, Richardson C, King K (2000) Hydrologic response of a small watershed model to generated precipitation. Trans ASAE 43:1483

Jolliffe IT, Stephenson DB (eds) (2012) Forecast verification: a practitioner's guide in atmospheric science, 2nd edn. John Wiley & Sons Ltd, Chichester

Kalnay E, Hunt B, Ott E, Szunyogh I (2006) Ensemble forecasting and data assimilation: two problems with the same solution. In: Palmer T, Hagedorn R (eds) Predictability of weather and climate. Cambridge University Press, Cambridge, pp 157–180. https://doi.org/10.1017/CBO9780511617652.008

Kou X, Ge J, Wang Y, Zhang C (2007) Validation of the weather generator CLIGEN with daily precipitation data from the Loess Plateau, China. J Hydrol 347:347–357

Koutsoyiannis D, Manetas A (1996) Simple disaggregation by accurate adjusting procedures. Water Resour Res 32:2105–2117

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. https://doi.org/10.1029/2011WR011446

Liew VK-S (2004) Which lag length selection criteria should we employ? Econ Bull 3:1–9

Mehan S, Guo T, Gitau M, Flanagan DC (2017) Comparative study of different stochastic weather generators for long-term climate data simulation. Climate 5:26. https://doi.org/10.3390/cli5020026

Mithen S, Black E (2011) Water, life and civilisation: climate, environment and society in the Jordan Valley, vol Cambridge. University Press, Cambridge

Moon SE, Ryoo SB, Kwon JG (1994) A Markov chain model for daily precipitation occurrence in South Korea. Int J Climatol 14:1009–1016

NCDC (2017) NCDC. https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets

Neild RE, Newman JE (1987) Growing season characteristics and requirements in the Corn Belt. Iowa State University, Cooperative Extension Service, Ames

Ng S, Perron P (2001) Lag length selection and the construction of unit root tests with good size and power. Econometrica 69:1519–1554

Nicks AD, Gander GA (1994) CLIGEN: a weather generator for climate inputs to water resource and other models. In: Proceedings of the fifth international conference on computers in agriculture, pp 903–909

Nicks AD, Lane LJ, Gander GA (1995) Chapter 2. Weather Generator. In USDA-Water Erosion Prediction Project: Hillslope Profile and Watershed Model Documentation; NSERL Report #10; USDA-ARS National Soil Erosion Research Laboratory: West Lafayette, IN, USA, 1995. https://www.ars.usda.gov/

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

Racsko P, Szeidl L, Semenov M (1991) A serial approach to local stochastic weather models. Ecol Model 57:27–41

Rajagopalan B, Lall U (1999) A k-nearest-neighbor simulator for daily precipitation and other weather variables. Water Resour Res 35:3089–3101

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. In. US Department of Agriculture, Agricultural Research Service Washington, DC

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

Royall RM (1986) The effect of sample size on the meaning of significance tests. Am Stat 40:313–315

Said SE, Dickey DA (1984) Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika 71:599–607

Schneider T (2001) Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. J Clim 14:853–871

Semenov MA (2008) Simulation of extreme weather events by a stochastic weather generator. Clim Res 35:203–212

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, Brooks RJ, Barrow EM, Richardson CW (1998) Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. Clim Res 10:95–107

Sharif M, Burn DH (2006) Simulating climate change scenarios using an improved K-nearest neighbor model. J Hydrol 325:179–196

Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt K, Tignor M, Miller H (2007) Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, 2007. Cambridge University Press, Cambridge

Soltani A, Hoogenboom G (2003) A statistical comparison of the stochastic weather generators WGEN and SIMMETEO. Clim Res 24:215–230

Srikanthan R, McMahon T (2001) Stochastic generation of annual, monthly and daily climate data: a review. Hydrol Earth Syst Sci Dis 5:653–670

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

Wilks DS (1999) Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. Agric For Meteorol 93:153–169

Wilks D (2002) Realizations of daily weather in forecast seasonal climate. J Hydrometeorol 3:195–207

Wilks DS, Wilby RL (1999) The weather generation game: a review of stochastic weather models. Prog Phys Geogr 23:329–357

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 33:326–342

Zhang X, Garbrecht JD (2003) Evaluation of CLIGEN precipitation parameters and their implication on WEPP runoff and erosion prediction. Trans ASAE 46:311