ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 28, NO. 4, 2011, 809–820

A Comparison of Four Precipitation Distribution Models Used in Daily Stochastic Models

LIU Yonghe[1,5,6] (刘永和), ZHANG Wanchang*[2] (张万昌), SHAO Yuehong[3] (邵月红), and ZHANG Kexin[4] (张可欣)

[1] *Key Laboratory of Regional Climate-Environment Research for Temperate East Asia (RCE-TEA), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing* 100029

[2] *Center for Hydrosciences Research, Nanjing University, Nanjing* 210093

[3] *Applied Hydrometeorological Research Institute, Nanjing University of Information Science and Technology, Nanjing* 210044

[4] *Linyi Meteorological Bureau, Shandong Province, Linyi* 276004

[5] *Institute of Resources and Environment, Henan Polytechnic University, Jiaozuo* 454000

[6] *Graduate University of Chinese Academy of Sciences, Beijing* 100049

(Received 1 November 2009; revised 19 February 2011)

ABSTRACT

Stochastic weather generators are statistical models that produce random numbers that resemble the observed weather data on which they have been fitted; they are widely used in meteorological and hydrological simulations. For modeling daily precipitation in weather generators, first-order Markov chain–dependent exponential, gamma, mixed-exponential, and lognormal distributions can be used. To examine the performance of these four distributions for precipitation simulation, they were fitted to observed data collected at 10 stations in the watershed of Yishu River. The parameters of these models were estimated using a maximum-likelihood technique performed using genetic algorithms. Parameters for each calendar month and the Fourier series describing parameters for the whole year were estimated separately. Bayesian information criterion, simulated monthly mean, maximum daily value, and variance were tested and compared to evaluate the fitness and performance of these models. The results indicate that the lognormal and mixed-exponential distributions give smaller BICs, but their stochastic simulations have overestimation and underestimation respectively, while the gamma and exponential distributions give larger BICs, but their stochastic simulations produced monthly mean precipitation very well. When these distributions were fitted using Fourier series, they all underestimated the above statistics for the months of June, July and August.

Key words: weather generators, gamma distribution, mixed-exponential distribution, Markov chain, Fourier series

Citation: Liu, Y. H., W. C. Zhang, Y. H. Shao, and K. X. Zhang, 2011: A comparison of four precipitation distribution models used in daily stochastic models. *Adv. Atmos. Sci.,* **28**(4), 809–820, doi: 10.1007/s00376-010-9180-6.

1. Introduction

Stochastic weather generators used in meteorological and hydrological studies are mostly statistical models that produce random numbers resembling the observed data on which they have been fitted (Wilks, 1999). Stochastic weather generators can generate weather sequences that statistically resemble the real observed data (Kou et al., 2007). Scenarios of daily or monthly meteorological variables with long time series

*Corresponding author: ZHANG Wanchang, zhangwc@nju.edu.cn

are often needed in hydrological, ecological, or crop-yield models, but observed meteorological records in basin or regional scales are often incomplete or difficult to obtain. Stochastic weather generators are usually integrated into hydrological or ecological models to extend or interpolate the incomplete meteorological data series for the purposes of the simulation applications. Among all the meteorological variables that stochastic weather generators can generate, daily precipitation is the most difficult variable to model because of its spatial and temporal intermittence (Richardson, 1981). In stochastic weather generators, precipitation occurrence is usually initially described with first- or second-order Markov chain–dependent models (Richardson, 1981; Woolhiser and Roldan, 1982; Wilks, 1998; Wu and Wang, 2001; Zheng and Katz, 2008). With a first-order Markov chain, the probability of precipitation occurrence is determined based on the wet or dry status of the previous day; with a second-order Markov chain, the probability of precipitation occurrence is conditioned on the wet or dry status of the previous 2 days (Richardson, 1981; Wilks, 1989). On a wet day the precipitation amount is assumed independent of successive days and can be generated by stochastic precipitation models according to weather status, such as one-parameter exponential distribution, two-parameter gamma distribution (Buishand, 1978), or three-parameter mixed-exponential distribution (Woolhiser and Pegram, 1979). In order to simulate the seasonal or periodic variation of the Markov-chain transition probabilities and to determine the parameters of these models, Fourier series are often used (Richardson, 1981; Woolhiser and Roldan, 1982; Chapman, 1997). With Fourier series, every day of the year has its own unique model parameters. If Fourier series are not used, the days must be grouped into several periods of 14, 28, or 30 days each, for example, and the parameters must then be estimated for each period. When using Fourier series, the seasonal variation of model parameters for the whole year can be represented by only a few harmonies (Richardson, 1981).

Stochastic precipitation models have been widely used in meteorological and hydrological studies. Applications of such models can be found for downscaling precipitation from coarse spatio-temporal resolution simulated from the general circulation model (Katz and Parlange, 1996; Bardossy, 1997; Bates et al., 1998) in meteorology. When exponential, gamma, or mixed-exponential distributions have been used to model precipitation amount, their underestimation of the mean or variance of monthly precipitation has frequently been pointed out in the literature as a common disadvantage of such generators (Katz and Parlange, 1998; Katz and Zheng, 1999; Zheng and Katz, 2008). This shortcoming has contributed to the poor performance of these distribution-based models in reproducing the histograms of daily precipitation amounts of rainfall events. Some modifications have been made to these models to reduce this underestimation (Katz and Zheng, 1999), but further improvement is warranted. The main reason for inaccurate estimation may be that these presumed theoretical distributions cannot be perfectly fitted to the observed data. Woolhiser compared three distribution and found that mixed-exponential distribution performed best, followed by gamma distribution and exponential distribution, respectively (Woolhiser and Roldan, 1982). Lognormal distribution is another method that can be used to model precipitation amounts (Sansom, 1998; Apipattanavis et al., 2007; Gyasi-Agyei and Mahbub, 2007), but few tests have been performed on this method. In this study, the principles behind frequently used stochastic precipitation models were briefly reviewed. Then, based on the analyses of the frequencies of daily precipitation amounts recorded from 1961 to 2007 at 10 meteorological stations in the watershed of the Yishu River (an upstream branch of the Huaihe River basin in China), models based on four distribution, including lognormal distribution, were fitted to observation data. Their performances were examined and validated with the meteorological data collected. This study was conducted only on single-site stochastic models to test the four distributions of modeling precipitation amounts. The results may provide guidance for building spatio-temporal stochastic precipitation models in future research.

2. Methods

2.1 *Study area and data*

Observed daily total precipitation records for 46 years (from 1962 to 2007) were obtained for the 10 stations located in watershed of Yishu River, which is an important branch of Huaihe River in China. The 10 stations all lie within the administration area of Linyi City, Shandong Province. Because of the monsoon climate in eastern China, the precipitation of this area has obvious seasonal variation. In winter and early spring, very few days are wet, but many more wet days occur during summer, especially in July and August. For all stations, 46 years of data were used to estimate the parameters. The precipitation data of the last day in each leap year were ignored so that each year had 365 days of data.

2.2 *Markov chain–dependent model*

In stochastic weather generators, the probability of precipitation occurrence is usually generated with first- or second-order Markov chain–dependent models to determine the wet or dry climatic statuses of any specific day considered. For a wet day, the daily precipitation is assumed independent on successive days and can be generated by the precipitation intensity models.

2.2.1 *Precipitation occurrence model*

The first-order Markov chain–dependent model has been widely used to model daily precipitation occurrence (Richardson, 1981; Wilks, 1992). In this model, when the precipitation of a day is below a threshold (0.25 mm), the day is considered a dry day; otherwise, it is considered a wet day. Here, we denote p_{01} (or p_{11}) as the probability of a dry day (or wet day) conditioned on a preceding wet day, and p_{00} (or p_{10}) as the probability of a dry day (or wet day) conditioned on a preceding dry day, respectively. Thus we have

$$p_{01} = 1 - p_{00} \tag{1}$$

$$p_{11} = 1 - p_{10} \tag{2}$$

Transition probabilities, therefore, can be fully and easily defined by p_{00} and p_{10}.

In practice, second-order Markov chain–dependent models can also be found in some published literature (e.g., Ceo and Stern, 1982; Katz and Parlange, 1998; Hansen and Mavromatis, 2001), but these models are more complex and need more parameters for validation than first-order Markov chain–dependent models; therefore, they are not discussed here.

2.2.2 *Precipitation intensity models*

Given a wet day, daily precipitation can be modeled using several distributions such as exponential distribution (Buishand, 1978), gamma distribution (Woolhiser and Pegram, 1979), lognormal distribution, and mixed-exponential distribution (Richardson, 1981). If the curves of probability density functions of these four distributions are plotted, the general shape of the curves resembles the distribution of the daily precipitation amounts. The probability density functions of these distributions follow.

The probability density function of one-parameter exponential distribution is given by

$$f_1(y) = \lambda \exp(-\lambda y), \quad \lambda > 0 \tag{3}$$

where λ is its only parameter to be determined and y stands for the daily precipitation amounts.

The probability density function of two-parameter gamma distribution is given by

$$f_2(y) = \frac{y^{\delta-1}\exp(-y/\gamma)}{\gamma^{\delta}\Gamma(\delta)}, \quad \gamma, \delta > 0 \tag{4}$$

where δ and γ are the shape parameter and the scale parameter, respectively.

The probability density function of three-parameter mixed-exponential distribution is given by

$$f_3(y) = \alpha\lambda_1 \exp(-\lambda_1 x) + (1-\alpha)\lambda_2 \exp(-\lambda_2 x), \quad \lambda_1, \lambda_2 > 0, \quad 0 < \alpha \leqslant 1 \tag{5}$$

where α, λ_1, and λ_2 are the three distribution parameters that need to be estimated.

The probability density function of two-parameter lognormal distribution is given by

$$f_4(y) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln y - \mu)^2}{2\sigma^2}\right], \quad \mu, \sigma > 0 \tag{6}$$

where of μ and σ stand for mean and standard deviation of the logarithmic transformed precipitation amounts, respectively.

2.2.3 *Parameter estimation*

These parameters should first be estimated based on the precipitation dataset recorded in the meteorological stations of the study region. Seasonal variations of these parameters can be explained by the models using two methods.

With the first method, the precipitation datasets are split by month, and parameters are estimated for each calendar month. For precipitation in a given month, all unknown parameters are assumed to be constant, so maximum-likelihood estimates can be made directly, except p_{01} and p_{11}. For all kinds of distribution functions, probability density functions can be rewritten as $f(y|\boldsymbol{\theta})$, in which $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ denotes the parameter vector of a corresponding distribution. The maximum-likelihood estimation finds $\widehat{\theta}$ of $\boldsymbol{\theta}$ by maximizing the log-likelihood function $U=\log L(\boldsymbol{\theta})$.

The second method represents these parameters using Fourier series and estimates coefficients of the Fourier series. Thus the parameters become variables that vary each day of the year. The unknown parameters, including p_{01} and p_{11}, involved in all the models can also be estimated as variables that vary each day of the year, and their seasonal variations can be described by Fourier series. Let θ_n be the parameter in the nth day of the year, according to Woolhiser and Pegram (1979). It can be defined by finite Fourier series like this:

$$\theta_n = c_0 + \sum_{k=1}^{m}\left[c_k \sin\left(\frac{nk}{T} + d_k\right)\right] \tag{7}$$

where c_0, c_1, d_1, c_2, $d_2\ldots$ are Fourier coefficients that need to be estimated, m is the maximum number of harmonies concerned, n is the day number ranging from 1 to 365, and $T = 365/2\pi$. According to Richardson (1981) and Roldan and Woolhiser (1982), the first three harmonies are significant enough for use, and the Fourier series makes the parameter optimization more parsimonious because the parameters only need to be estimated once for the whole year.

The maximum-likelihood methods described by Woolhiser and Pegram (1979) can be used for estimating the parameters (found using the Fourier series) of the Markov chain–dependent models of precipitation occurrence and amount. All of these maximum log-likelihood equations can be fully estimated using genetic algorithms (GAs). Like any other maximum-likelihood solving method, GAs can only obtain approximations of real parameters, but it is an easy and powerful method that can calculate parameters with satisfying precision. To obtain credible parameter values, the GA devised for the parameter estimation of each model was run several times. In the results the parameter values obtained from each simulation were not the same, but the differences were minor, indicating that the GA estimations are stable and credible. For simple models with parameters not represented by Fourier series, like models based on exponential, gamma, and lognormal distributions, the parameters estimated using GA and those estimated using some statistical functions in Matlab software were almost the same. But for mixed-exponential distribution or all these models whose parameters are represented by Fourier series, no parameter estimation method is available in Matlab. So we developed a complex computer program for estimating parameters of all the GA-based models.

2.2.4 *Stochastic precipitation generation from the lognormal distribution*

Methods for producing daily precipitation amounts from Markov chain–dependent distribution models can be found in literature, such as Wilks (1998), and textbooks. Here only some problems with the lognormal distribution–based model are discussed. The stochastic generation from lognormal distribution–based model can be done using

$$r = \exp(g\sigma + \mu) \tag{8}$$

where g is the Gauss variable of $N(0, 1)$. With this model, extreme values that exceed the daily maximum historic record of precipitation were often generated, which led to gross overestimation of monthly mean and daily maximum precipitation for some specific months. To avoid this problem, threshold values were set prior to the Gauss random number of the lognormal stochastic generator for each specific month of a year as follows:

$$T_m = \frac{\ln M_m - \mu_t}{\sigma_t} \tag{9}$$

where T_m is the upper threshold of Gauss random number for the mth month, M_m is the maximum daily precipitation observed in the mth month, μ_t, σ_t are the mean and standard deviation of the log-transformed precipitation for each specific month of a year, respectively, and can be derived by Fourier series or calculated statistically. When a Gauss random number generated is larger than T_m, it is regarded as invalid, then a new Gauss random number is regenerated until the new Gauss random number becomes smaller than or equal to T_m.

3. Results

3.1 *Results of parameter estimation*

These chain-dependent models based on exponential, gamma, mixed-exponential and lognormal distributions were fitted to the observed precipitation for all months and all stations by monthly datasets. Results of the comparison of maximized likelihood for the models based on the four distributions showed that, for all months, the order of these models sorted from best fit to worst fit is lognormal, mixed-exponential distribution, gamma distribution, and exponential distribution. The fact that the mixed-exponential distribution fit better than gamma distribution and that gamma distribution fit better than exponential distribution agreed with Woolhiser and Roldan (1982). But in this study the lognormal distribution of precipitation amounts proved to fit better than any other distributions. Comparing the results listed in Tables 1 and 2, all maximum likelihoods and Bayesian information criteria (BIC) obtained by lognormal distribution are larger than those obtained by mixed-exponential distribution.

The estimated likelihood function can be regarded as a measure of good fit, but in terms of suitability, this measure should be penalized for the number of parameters it uses. The BIC is the criterion used most often (Schwarz, 1978), defined as follows:

$$\mathrm{BIC}(m) = -2 \ln M_m + k_m \ln n \tag{10}$$

where m is the model's order (its value can be 1,2,...), M_m is the maximized log-likelihood function for the mth model, k_m is the number of parameters of the mth model, and n is the total number of wet days of the data series being used to fit the model. A better model should have a lower BIC values. Tables 1 and 2 show the calculated BIC values of mixed-exponential distribution and lognormal distribution for a station. BIC values for all months and all stations were calculated.

Table 1. Parameters estimated for mixed-exponential distribution for Mengyin station. The α is the proportion parameter for mixing two exponential distributions which have rate parameters λ_1 and λ_2, respectively. The exponential distribution has a mean of the reciprocal of its rate parameter.

Month	α	$1/\lambda_1$	$1/\lambda_2$	Log-likelihood	BIC
1	0.4215	1.4767	5.3654	−274.2	554.7
2	0.4272	6.3422	2.0718	−375.7	758.1
3	0.2384	1.1791	6.6425	−481.7	970.2
4	0.4829	10.5829	2.852	−744.7	1496.6
5	0.3524	1.9565	12.7485	−848.1	1703.5
6	0.4709	24.8598	3.7324	−1267.2	2542.1
7	0.2915	2.1516	24.5473	−2243.4	4495.1
8	0.4001	2.0989	25.4563	−1706.1	3420.2
9	0.4338	2.2043	16.6109	−996.6	2000.7
10	0.4997	2.9158	11.769	−684.4	1375.9
11	0.0517	27.0899	4.4493	−457.4	921.4
12	0.1227	1.2871	3.9373	−314.9	636.3

Table 2. Parameters estimated for lognormal distribution for Mengyin station. μ and σ stand for mean and standard deviation of the logarithmic transformed precipitation amounts, respectively. BIC=Bayesian information criteria.

Month	μ	σ	Log-likelihood	BIC
1	0.6957	1.1254	−267.9	540.0
2	0.7687	1.1027	−368.0	740.4
3	0.9949	1.2716	−483.1	970.6
4	1.1569	1.2839	−740.4	1485.6
5	1.3738	1.4083	−852.8	1710.4
6	1.6441	1.5192	−1267.2	2539.5
7	1.9268	1.5597	−2247.6	4500.8
8	1.7168	1.6191	−1718.4	3442.1
9	1.4012	1.4691	−993.5	1991.9
10	1.2373	1.3034	−680.6	1366.0
11	1.018	1.2136	−452.5	909.4
12	0.7523	1.1003	−312.8	629.9

Table 3. Estimated maximum log-likelihood for the distributions parameterized by Fourier series.

Station	Mengyin	Pingyi	Feixian	Yishui	Yinan	Linyi	Ju'nan	Cangshan	Linshu	Tancheng
occur	−7421	−7329	−7545	−7367	−7534	−7645	−7632	−7578	−7630	−7633
logn_M	−10385	−10065	−10833	−10217	−10625	−11045	−10886	−10723	−11051	−11034
mixp_M	−10394	−10075	−10848	−10250	−10638	−11059	−10892	−10750	−11066	−11060
gmm_M	−10516	−10196	−10985	−10380	−10776	−11210	−11040	−10895	−11209	−11205
exp_M	−10703	−10358	−11158	−10547	−10951	−11435	−11261	−11103	−11421	−11428
logn_F	−10388	−10076	−10835	−10220	−10627	−11049	−10885	−10733	−11060	−11043
mixp_F	−10407	−10087	−10858	−10254	−10645	−11071	−10895	−10763	−11082	−11073
gmm_F	−10525	−10215	−10993	−10386	−10781	−11217	−11042	−10906	−11224	−11219
exp_F	−10713	−10373	−11166	−10551	−10951	−11442	−11262	−11116	−11441	−11447

Note: The abbreviation of "occur" represent the model for precipitation occurrence probabilities, the abbreviations of "logn", "mixp", "gmm" and "exp" are representing the models based on lognormal, mixed exponential, gamma and exponential distributions respectively, while "F" and "M" were used to distinguish the two methods (Fourier method and calendar-month method) for describing the seasonal variations of model parameter respectively.

Table 4. Total BIC of all distribution models at all stations.

Station	Mengyin	Pingyi	Feixian	Yishui	Yinan	Linyi	Ju'nan	Cangshan	Linshu	Tancheng
logn_M	20826	20186	21723	20491	21308	22147	21830	21504	22160	22126
mixp_M	20874	20235	21782	20584	21361	22205	21870	21586	22218	22207
gmm_M	21090	20449	22026	20816	21610	22477	22137	21847	22475	22467
exp_M	21435	20743	22344	21122	21931	22898	22551	22234	22870	22884
logn_F	20783	20159	21677	20447	21261	22105	21777	21473	22127	22093
mixp_F	20821	20181	21723	20515	21297	22149	21797	21533	22171	22153
gmm_F	21057	20437	21993	20779	21569	22441	22091	21819	22455	22445
exp_F	21433	20753	22339	21109	21909	22891	22531	22239	22889	22901

Note: The abbreviations in the first column have the same meaning as those in Table 3.

Comparisons show that all BIC values of lognormal models are lower than all those of mixed-exponential distribution in the same month. The BIC value of mixed-exponential distribution is lower than that of exponential distribution. Moreover, in this study the differences in BIC values among the 10 stations are minor. This means that the climate characteristics of the 10 stations are similar.

Fourier series for parameters of exponential, mixed-exponential, and lognormal distributions for the whole year were estimated using GAs. Notably, the first three harmonies were used to optimize the log-likelihoods functions first, then one more harmony was added, and the log-likelihoods functions were recalculated. The log-likelihoods were larger for the increased harmonies than for the original harmonies. Although this improvement was very minor, it consumed much more computer-time. Thus, the first three harmonies were used for all parameter estimation in this paper. Table 3 shows maximum-likelihood values of the precipitation occurrence and the four precipitation amounts for distributions whose parameters were represented by Fourier series. Similar to those estimated from monthly data, results showed that the order of distributions with best fit to worst fit were lognormal, mixed-exponential, gamma, and exponential distribution. The BIC values also support this conclusion (Table 4).

To compare the models whose parameters were represented by Fourier series, BIC values of models fitted from monthly data were summed up into yearly BIC (Table 4). Results show that, for lognormal, mixed-exponential, and gamma distribution, the models with parameters represented by Fourier series had smaller BIC values than the models fitted from monthly datasets, but for exponential distribution, the former have smaller BIC values than the latter only at six stations. Note that the maximum likelihoods estimated using Fourier series were smaller than those calculated using monthly datasets. This result resembles that of Woolhiser and Roldan (1982).

3.2 *Results of stochastic simulation*

To test the performance of these models, the monthly mean, monthly variance, and daily maximum for each month were calculated from generated daily precipitation data of 50 years.

Even the precipitation generators can only produce stochastic daily results; the daily mean precipitation of every month they produce should be very similar to the real monthly mean value. Figure 1 shows the plots of observed versus simulated monthly mean precipitation by models based on the four distributions whose parameters were estimated from monthly datasets. Obviously, the Markov models based on these four distributions can produce monthly mean precipitation data that resembles real observed data. Comparing the four plots, the correlations between observed data and data simulated by gamma distribution and exponential distribution are closer to 1.0, and regression interceptions are smaller than those of the other two distributions. This result seems contrary to the inference from estimated maximum-likelihood values, because exponential distribution and gamma distribution do not fit better than the other two distributions. This problem may result from the left tails of mixed-exponential and lognormal distributions because extreme daily values produced by these distributions come from their left tails. Figure 2 shows the plots of observed versus simulated monthly mean precipitation by models based on the four distributions whose parameters were estimated through Fourier series. All of the plots of these distributions showed underestimations when monthly mean precipitation exceeded 150 mm. The lognormal distributions did not under- or overestimate the monthly mean precipitation when these monthly means were $<$110 mm, but all of the other three distributions overestimated them when the monthly means were $<$140 mm. Almost all of the underestimations occurred in the months of June, July, and August.

An ideal stochastic weather generator should reproduce other statistics and characteristics of the real precipitation. Among the statistics, maximum daily

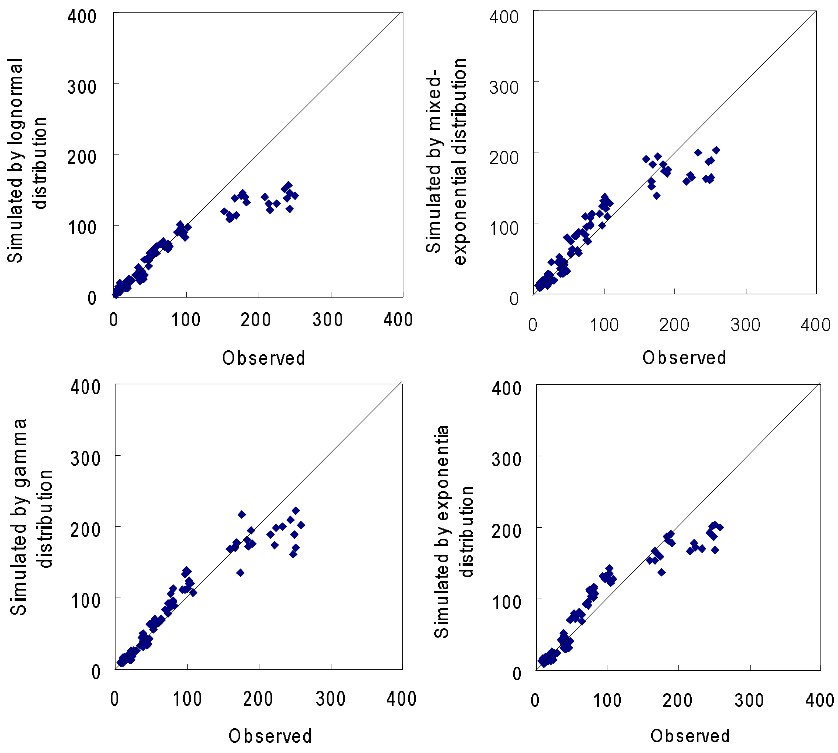

Fig. 1. Observed monthly mean precipitation (mm) versus that simulated by lognormal, mixed-exponential, gamma, and exponential distributions whose parameters were estimated for each month separately, for all station pairs.

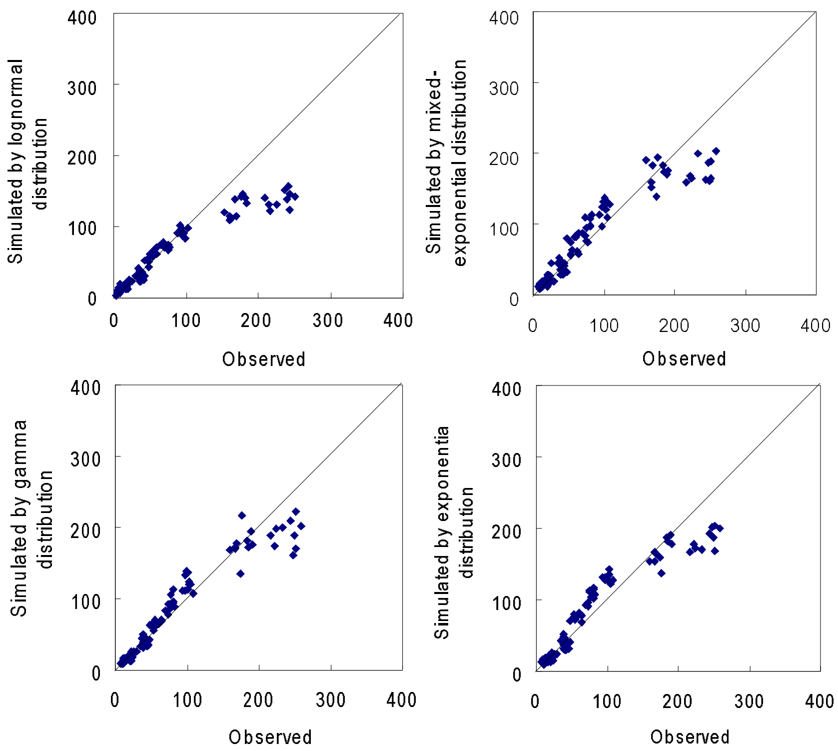

Fig. 2. Observed monthly mean precipitation (mm) versus that simulated by lognormal, mixed-exponential, gamma, and exponential distributions whose parameters were represented by Fourier series.

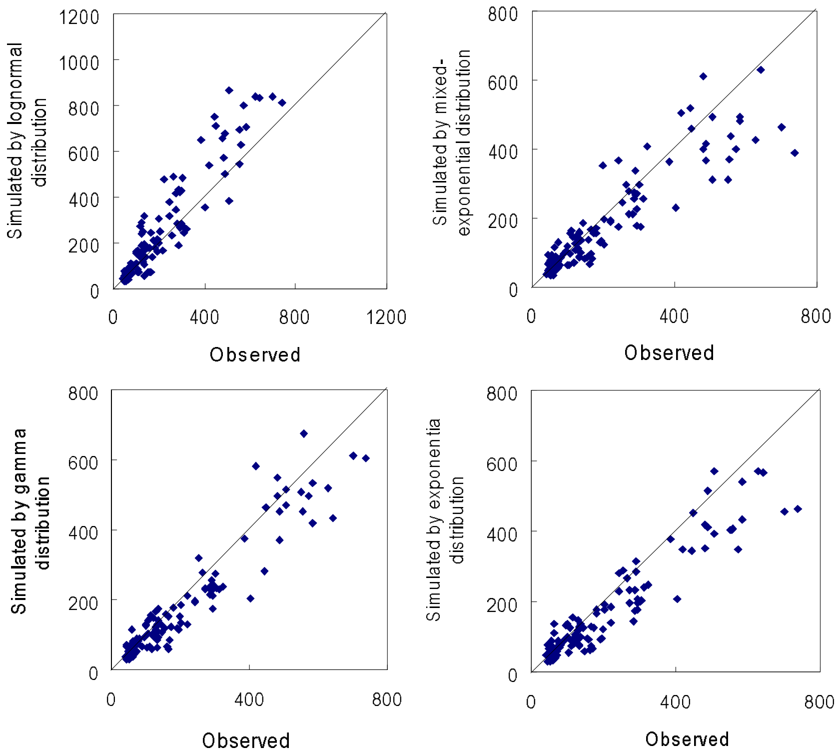

Fig. 3. Observed monthly maximum precipitation (mm) versus that simulated by lognormal, mixed-exponential, gamma, and exponential distributions whose parameters were estimated from monthly datasets.

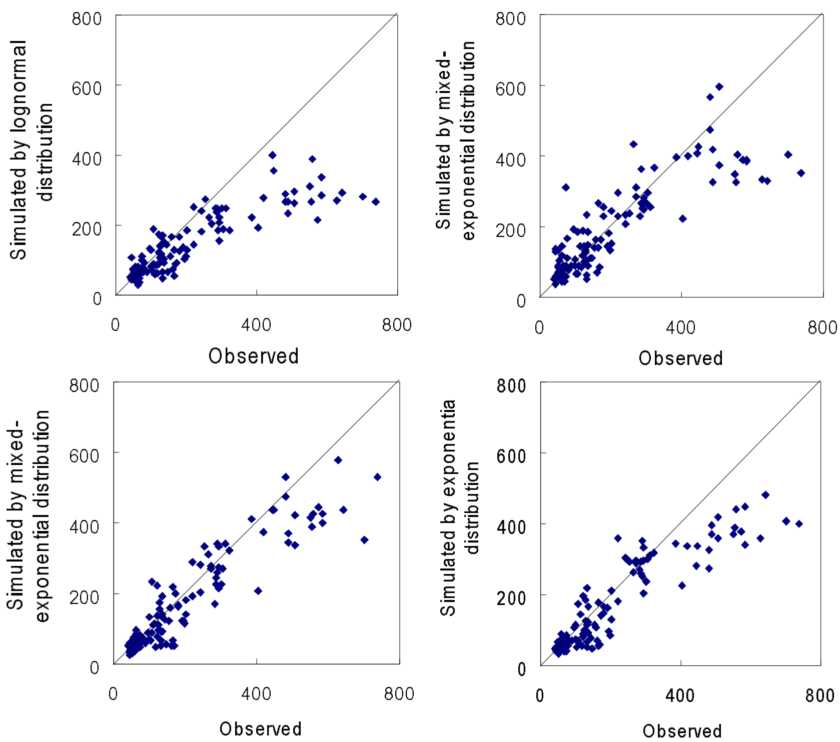

Fig. 4. Observed monthly maximum precipitation (mm) versus that simulated by lognormal, mixed-exponential, gamma, and exponential distributions whose parameters were represented by Fourier series.

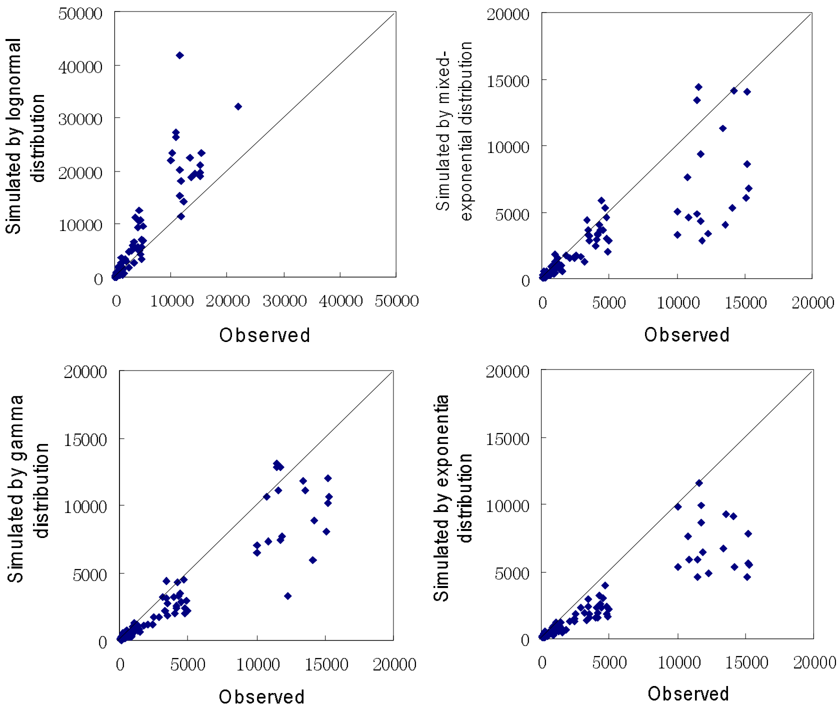

Fig. 5. Observed monthly variance versus that simulated by lognormal, mixed-exponential, gamma, and exponential distributions whose parameters were estimated from monthly datasets.

precipitation and monthly variance are very important for ecological or hydrological assessment and simulation. These two statistics produced by stochastic weather generators may not be very close to the observed data, but the smaller the differences are, the better the weather generators. Figure 3 shows the plots of observed monthly maximum daily precipitation versus those simulated by models based on the distributions whose parameters were estimated from monthly datasets. The plots show that the lognormal distribution produced larger extreme precipitation in most cases, and the other three distributions underestimated them, except for mixed-exponential distribution. This underestimation was not as obvious as that of the gamma and exponential distributions. When these models' parameters were estimated using Fourier series, all of the distributions, including lognormal distributions, underestimated maximum daily precipitation of each summer month (Fig. 4), and this underestimation seems more evident than the distributions with parameters estimated from monthly datasets. Figure 5 shows the plots of observed versus simulated monthly variance of daily precipitation by models based on the distributions whose parameters were estimated from monthly datasets. Similarly, the plots of those distributions with parameters estimated by Fourier series are shown in Fig. 6. These results indicate that, on the whole, lognormal distribution overestimated monthly variances, but the other three distributions underestimated them. However, all four distributions underestimated monthly variance.

4. Discussions

For the models estimated from monthly datasets, lognormal distribution and mixed-exponential distribution fitted better than did those of gamma and exponential distribution, but they under- or overestimated monthly mean precipitation and monthly maximum daily precipitation. This may be due to their special ability to generate extreme values, because monthly maximum and monthly mean precipitation are very sensitive to extreme values; lognormal distribution can more frequently produces extreme values than observed data, while mixed-exponential distribution produces less extreme values than observed data. This under- or overestimation is difficult to eliminate because these distributions were unable to be fitted more perfectly to observed data. Gamma and exponential distribution may also be underfitted, but this underfitting does not affect the monthly mean precipitation.

For the models using Fourier series, lognormal distribution produced monthly mean well; other models

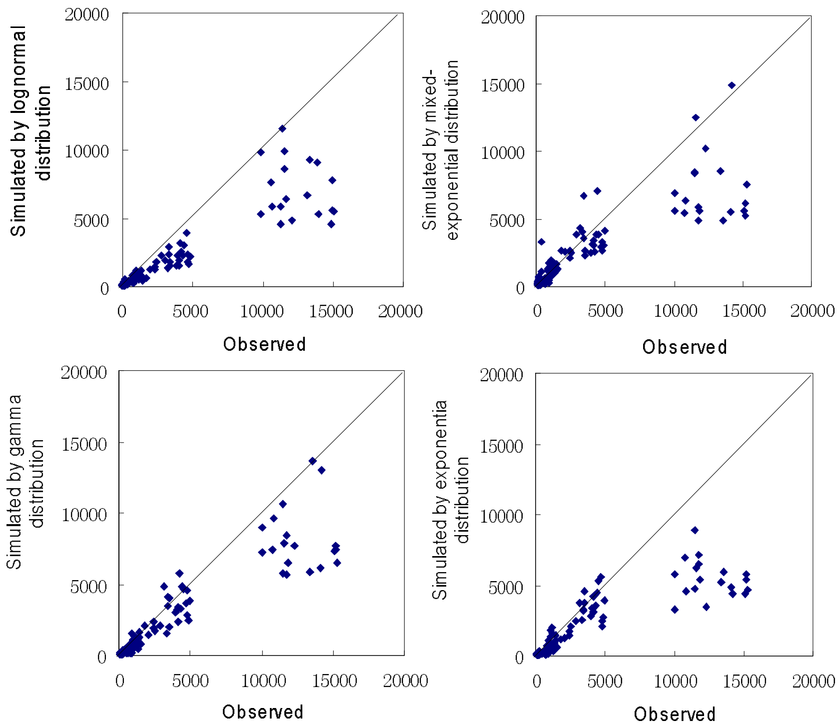

Fig. 6. Observed monthly variance versus that simulated by lognormal, mixed-exponential, gamma and exponential distributions whose parameters were represented by Fourier series.

overestimated monthly means when the monthly mean were small, but these models utterly underestimated results in the months of June, July, and August. As for monthly maximum precipitation, all models that used Fourier series produced extreme underestimations. All underestimations of monthly mean precipitation and monthly maximum daily precipitation were due to the smoothing effects of Fourier series that erode the parameter values that produce extreme precipitation. This can be seen in Fig. 7b: the curve of simulated monthly mean precipitation was smoothed over the whole year. For the months of June, July, and August, monthly mean precipitation was obviously underestimated. This tendency contrasts with the monthly mean precipitation data produced by models whose parameters were estimated from monthly datasets (Fig. 7a). The underestimation caused by smoothing added to the underestimation in the underfitted models, thus the model underestimated monthly mean and monthly maximum precipitation to an even greater extent. The smoothing caused by Fourier series stems from two sources. The first source is the smoothed probability of precipitation occurrence, illustrated by comparison of the lines in Fig. 8. Another source is the smoothed distribution parameters of precipitation amounts. These two smoothing factors can be alleviated by adding more harmonies to estimating parameters, but our tests showed that adding one or two harmonies does not improve the model much, and adding harmonies for estimating Fourier coefficients consumes too much additional computer time.

5. Conclusions

The Markov-chain model and four distribution-based models (i.e., exponential, mixed exponential, gamma, and lognormal) were used to describe daily precipitation. In this study, the performances of these models were compared; all of the models were fitted to observed time series of daily precipitation of 46 years at 10 observation stations. The fitting process was handled in two ways: (1) by estimating parameters for each calendar month and (2) by estimating Fourier coefficients for an entire year. Both methods were conducted using GAs.

The estimated maximum log-likelihoods and BIC criteria for both estimation methods showed that the lognormal distribution fitted observation data with the most accuracy, followed by the mixed-exponential distribution, the gamma distribution, and the exponential distribution, respectively.

For the distributions fitted to monthly datasets, the stochastic simulation results are as follows: (1) lognormal distribution overestimated the monthly mean,

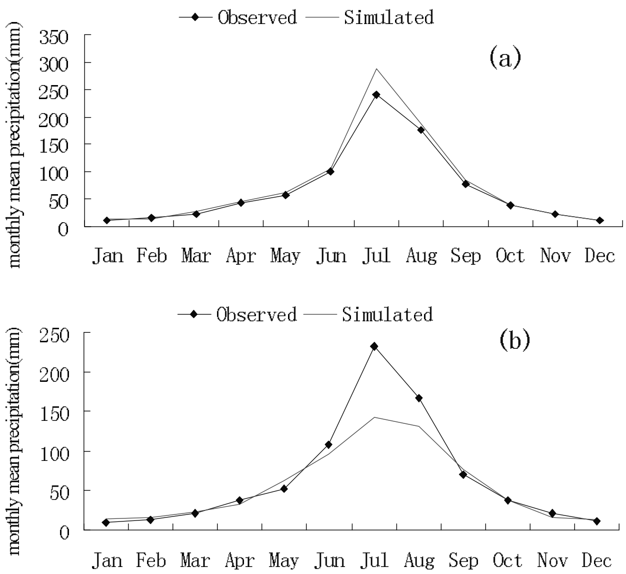

Fig. 7. Observed means of monthly total precipitation compared with those simulated by lognormal distribution: (a) the parameters of lognormal distribution were estimated by monthly datasets, (b) the parameters of lognormal distribution were estimated using Fourier series. The means of monthly total precipitation were calculated by averaging the monthly total precipitation at all stations. Because the simulated data were stochastic, the plots of (a) or (b) are not the same at each generation. Notably, the monthly mean precipitation was overestimated by model based on lognormal distribution in the month of July, but for models based on other distributions, results may not be similar.

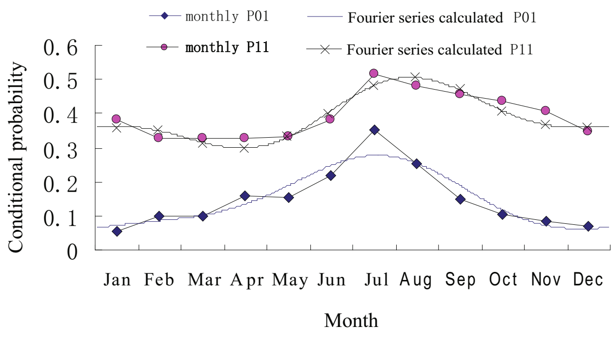

Fig. 8. Comparison of the timely series of conditional precipitation occurrence probabilities calculated from Fourier series with those counted directly from observations.

the maximum daily precipitation in month and the variance in month; (2) mixed-exponential distribution underestimated the monthly mean, while gamma and exponential distributions produced monthly mean very well; (3) the mixed-exponential, gamma, and exponential distributions models underestimated the monthly maximum daily precipitation and monthly variance.

For the distributions fitted using Fourier series, all distribution methods underestimated the monthly mean, the monthly maximum, and the variance for the months of June, July, and August. These underestimations may have been caused by the smoothing effects of Fourier series, which affected the model's ability to produce extreme daily values.

The results indicate that a better fitted model does not mean a better model in all aspects. The model fitting of lognormal distribution and mixed-exponential distribution achieved the best log-likelihoods and the best BICs, but they slightly over- or underestimated monthly mean and monthly maximum precipitation. When choosing a good distribution, therefore, maximum likelihoods and BIC may not be the only important criteria, and the selected distribution method should be further tested using other characteristics.

Acknowledgements. We are very much indebted to the staff members of the Linyi Meteorological Bureau, Shandong Province for their help and support with meteorological data collection throughout this research. This study is financially supported by the National Key Developing Program for Basic Sciences of China (Grant No. 2010CB951404), Chinese Nature Science Foundation (Grant No. 40971024) and the Special Meteorology Project [GYHY(QX)2007-6-1].

REFERENCES

Apipattanavis, S., G. Podesta, B. Rajagopalan, and R. W. Katz, 2007: A semiparametric multivariate and multisite weather generator. *Water Resour. Res.*, **43**(11), W11401, doi: 10.1029/2006WR005714.

Bardossy, A., 1997: Downscaling from GCMs to local climate through stochastic linkages. *Journal of Environmental Management*, **49**(1), 7–17.

Bates, B. C., S. P. Charles, and J. P. Hughes, 1998: Stochastic downscaling of numerical climate model simulations. *Environmental Modelling & Software*, **13**(3–4), 325–331.

Buishand, T. A., 1978: Some remarks on use of daily rainfall models. *J. Hydrol.*, **36**(3–4), 295–308.

Ceo, R., and R. D. Stern, 1982: Fitting models to daily rainfall data. *J. Appl. Meteor.*, **21**(**7**), 1024–1031.

Chapman, T. G., 1997: Stochastic models for daily rainfall in the Western Pacific. *Mathematics and Computers in Simulation*, **43**(3–6), 351–358.

Gyasi-Agyei, Y., and S. M. Parvez Bin Mahbub, 2007: A stochastic model for daily rainfall disaggregation into fine time scale for a barge region. *J. Hydrol.*, **347**(3–4), 358–370.

Hansen, J. W., and T. Mavromatis, 2001: Correcting low-frequency variability bias in stochastic weather generators. *Agriculatural and Forest Meteorology*, **109**(4), 297–310.

Katz, R. W., and M. B. Parlange, 1996: Mixtures of stochastic processes: Application to statistical downscaling. *Climate Research*, **7**(2), 185–193.

Katz, R. W., and M. B. Parlange, 1998: Overdispersion phenomenon in stochastic modeling of precipitation. *J. Climate*, **11**(4), 591–601.

Katz, R. W., and X. G. Zheng, 1999: Mixture model for overdispersion of precipitation. *J. Climate*, **12**(8), 2528–2537.

Kou, X. J., J. P. Ge, Y. Wang, and C. J. Zhang, 2007: Validation of the weather generator CLIGEN with daily precipitation data from the Loess Plateau, China. *J. Hydrol.*, **347**(3–4), 347–357.

Richardson, C. W., 1981: Stochastic simulation of daily precipitation, temprature, and solar-radiation. *Water Resour. Res.*, **17**(1), 182–190.

Roldan, J., and D. A. Woolhiser, 1982: Stochastic daily precipitation models.1. A comparison of occurrence processes. *Water Resour. Res.*, **18**(5), 1451–1459.

Sansom, J., 1998: A hidden Markov model for rainfall using breakpoint data. *J. Climate*, **11**(1), 42–53.

Schwarz, G., 1978: Estimating dimension of a model. *Annals of Statistics*, **6**(2), 461–464.

Wilks, D. S., 1989: Conditioning stochastic daily precipitation models on total monthly precipitation. *Water Resour. Res.*, **25**(6), 1429–1439.

Wilks, D. S., 1992: Adapting stochastic weather generation algorithms for climate change studies. *Climatic Change*, **22**(1), 67–84.

Wilks, D. S., 1998: Multisite generalization of a daily stochastic precipitation generation model. *J. Hydrol.*, **210**(1–4), 178–191.

Wilks, D. S., 1999: Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agricultural and Forest Meteorology*, **93**(3), 153–169.

Woolhiser, D. A., and G. Pegram, 1979: Maximum likelihood estimation of Fourier coefficients to describe seasonal-variations of parameters in stochastic daily precipitation models. *J. Appl. Meteor.*, **18**(1), 34–42.

Woolhiser, D. A., and J. Roldan, 1982: Stochastic daily precipitation models. 2. A comparison of distributions of amounts. *Water Resour. Res.*, **18**(5), 1461–1468.

Wu, J. D., and S. L. Wang, 2001: Incorporating stochastic weather generators into studies on climate impacts: Methods and uncertainties. *Adv. Atmos. Sci.*, **18**(5), 937–949.

Zheng, X. G., and R. W. Katz, 2008: Mixture model of generalized chain-dependent processes and its application to simulation of interannual variability of daily rainfall. *J. Hydrol.*, **349**(1–2), 191–199.