ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 28, NO. 6, 2011, 1357–1366

An Improved, Downscaled, Fine Model for Simulation of Daily Weather States

JIANG Zhihong[1] (江志红), DING Yuguo*[1](丁裕国),
ZHENG Chunyu[1,2] (郑春雨), and CHEN Weilin[1] (陈威霖)

[1]*Key Laboratory of Meteorological Disaster of Ministry of Education,*

Nanjing University of Information Science & Technology, Nanjing 210044

[2]*Environment Assessment, Northeast Electric Power Design Institute, Shenyang* 110000

(Received 23 September 2010; revised 17 February 2011)

ABSTRACT

In this study, changes in daily weather states were treated as a complex Markov chain process, based on a continuous-time watershed model (soil water assessment tool, SWAT) developed by the Agricultural Research Service at the U.S. Department of Agriculture (USDA-ARS). A finer classification using total cloud amount for dry states was adopted, and dry days were classified into three states: clear, cloudy, and overcast (rain free). Multistate transition models for dry- and wet-day series were constructed to comprehensively downscale the simulation of regional daily climatic states. The results show that the finer, improved, downscaled model overcame the oversimplified treatment of a two-weather state model and is free of the shortcomings of a multistate model that neglects finer classification of dry days (i.e., finer classification was applied only to wet days). As a result, overall simulation of weather states based on the SWAT greatly improved, and the improvement in simulating daily temperature and radiation was especially significant.

Key words: stochastic simulation, daily weather state series, Markov chain, state vector.

Citation: Jiang, Z. H., Y. G. Ding, C. Y. Zheng, and W. S. Chen, 2011: An improved, downscaled, fine model for simulation of daily weather states. *Adv. Atmos. Sci.*, **28**(6), 1357–1366, doi: 10.1007/s00376-011-0086-8.

1. Introduction

In recent years, widespread attention has been given to statistical downscaling methods and applications that use a low-resolution atmospheric-oceanic general circulation model (AOGCM) output to produce finer, spatial-scale, regional and local surface climate information (e.g., temperature and precipitation). Although the development of statistical downscaling techniques is still in progress, they can be organized into three broad categories: regression models, weather classification schemes, and weather generators (Chin, 1977; Ding and Zhang, 1989; Bardossy et al., 1991; Bouraoui et al., 2002;). Weather generators comprise a set of statistical models that can construct stochastic processes for climate variables. When the evolution of daily weather events is treated as a complex stochastic process, and a statistical model for observational weather variables is used to obtain statistical model parameters, the statistical model can then be employed to generate time sequences of climate variables. Thus, with this type of statistical model, daily weather sequences can be simulated under certain climate scenarios with using AOGCM outputs (e.g., future local monthly average temperature and precipitation). Simulation experiments performed in this study showed that simulated daily weather sequences associated with the current observed climate conditions (e.g., local monthly average temperature and precipitation) have very high similarities with the observed daily weather sequences. If this similarity persists in the future climate conditions, future climate characteristics based on the daily weather states can be simulated with better precision.

One of the more popular weather generators is based on Markov chain models. Richardson (1981)

*Corresponding author: DING Yuguo, dygnim@126.com

proposed a series of statistical models for precipitation, temperature, and solar radiation to simulate daily weather changes. Racsko et al. (1991) also developed various statistical modeling methods, including the Markov chain method to simulate weather event changes. Bardossy and Plate (1991) proposed the semi-Markov chain model with the use of circulation patterns (Ding and Niu, 1990; Ding, 1994; Ding et al., 2008, 2009). In fact, the Markov chain process has long been one of the major statistical models used in statistical meteorological research and applications. Many publications have examined this subject (Yao and Ding, 1990). Ding and Niu (1990) used a multistate Markov chain model based on the Monte Carlo method to create a statistical model for single-site daily precipitation simulation. Validations showed a high degree of similarity between model-generated climate statistical parameters and the observed ones (Katz, 1974; Richardson, 1981; Racsko et al., 1991; Palutikof et al., 2002). Liao et al. (2004) also developed a first-order two-state Markov chain model based on two parameters as a statistical weather generator for most of the regions in China with reasonable results. Ding et al. (2009) used this weather generator to simulate daily precipitation at six representative stations (over 30 years) with excellent results. The studies show the feasibility of using the Markov chain model as a basis for statistical models for simulating daily meteorological variable changes.

Yet, as a statistical weather generator, current Markov chain models are not without shortcomings. Two major problems have been identified: First, models did not account for the interannual variability in the time sequences of daily weather (e.g., precipitation) under current or future climate conditions, creating obvious inconsistencies with observations. In fact, daily precipitation not only has obvious interannual variability, it also possesses significant seasonal variability. For the latter, a simple treatment is to simulate events in different seasons or months based on different Markov chain models. Secondly, although the two-state Markov chain model described in Bardossy and Plate (1991) was improved, the classification of dry or wet state remained subjective. For example, no finer classification for dry days was used, and dry and wet days were treated as the same state vector with the same transformation characteristics, which is inconsistent with the observations. The objective of the current study was to solve the second problem by constructing a multistate Markov chain model with finer classification of both dry- and wet-day states so that the weather sequences comprised of daily weather states (including precipitation, temperature, humidity, solar radiation, and other variables) could be simulated comprehensively.

2. Improvement of the model

The multistate Markov chain model proposed by Ding et al. (1989) was based on two assumptions: (1) dry- and wet-day sequences follow a time-succession rule and (2) the probability distribution for daily precipitation amount of a wet day takes a certain form. Considering these two aspects, wet-day daily precipitation amounts were classified into many states based on their magnitude, and n states of wet days were assumed to be $s_1, s_2, \ldots, s_n$. Obviously, a more complete precipitation process can be built on terms of $n+1$ states $(s_0, s_1, s_2, \ldots, s_n)$, where the state s_0 denotes a dry day. This classification of states is beneficial to simulating the real daily precipitation process. The results of our study show that classification yielded significant improvement over the two-state Markov chain weather generator. For example, the simulated maximum daily precipitation amount more closely matched the observation data, but the simulation of the dry-day state was very coarse. Dry days were classified into clear, cloudy, and overcast (no rain) based on total cloud amount and were combined with previously classified wet-day states to form a complete daily weather state vector. The matrix of transition probabilities was then calculated. Thus, the improved, mixed, daily weather-state vector was expected to result in better simulation of complete daily weather states.

A continuous-time watershed model (soil water assessment tool, SWAT) developed at the U.S. Department of Agriculture Agricultural Research Service (USDA-ARS) can be used to simulate and predict the long term impact of climate change and human activities on water basin watersheds, sediment, and pollutants. The weather generator of the SWAT model is useful for generating complete meteorological data or for interpolating missing records. But the shortcoming of the weather generator is its oversimplified treatment of clear, cloudy, and rainy days. Also, the empirical parameters and formulas of the SWAT model are all based on observations in some major regions in the United States and are hardly applicable in China. For this reason, this investigators aimed to propose a finer model for daily weather simulation to perform regional climate downscaled numerical modeling, based on the SWAT model weather generator and the aforementioned proposal of classifying dry day into clear, cloudy, and overcast (rain free) three states according to total cloud amount.

2.1 *The mixed multistate Markov chain model for daily dry or wet states*

As stated previously, a more complete state describing daily weather processes $(s_0, s_1, s_2, \ldots, s_n)$ is created when mixed dry- and wet-day states are denoted as $n+1$ states: $s_0, s_1, s_2, \ldots, s_n$. The matrix of transition probabilities is similar to that reported in Ding and Zhang (1989). Because wet-day states $s_1, s_2, \ldots, s_n$ have corresponding precipitation states, each state $s_i (i = 1, 2, \ldots, n)$ must correspond to one probability distribution. The thorough study of Ding (1994) showed that daily precipitation amount follows Gamma distribution. Apparently, due to a positive skew of daily precipitation with Gamma distribution, chances of small amounts of daily precipitation are great and chances of extreme daily precipitation are very small. Thus, it is better to classify wet-day states based on the intervals of the daily precipitation amount: smaller intervals for small daily precipitation amounts, and bigger intervals for increased daily precipitation amounts (Ding and Zhang, 1989; Ding et al., 2009). Based on several numerical experiments, this study used wet-day classification intervals of daily precipitation that mimic a geometric series. For the sake of calculation, some adjustments were applied to the limits of the intervals. For dry days, three states [i.e., clear, cloudy, and overcast (rain free)] were assigned based on total cloud amount (i.e., ranked 1–3, 4–6, and 7–10, respectively). Studies have shown that total cloud amount has a U-shaped distribution. As a result, if total cloud amounts are assigned to three groups, then their distribution can be seen as uniform. In other words, dry-day state classification can be adjusted to use s_0, s_1, s_2 to denote clear, cloudy, and overcast (rain-free) states, respectively. Because the probability distribution of wet-day states follows a positively skewed normal distribution with a long tail, wet-day states $s_3 - s_{n-1}$ can be assumed to be uniformly distributed, then s_n can be defined as displaced exponent distribution to represent the characteristic of the long tail of a positively skewed normal distribution. The basis for this assumption is this: Daily precipitation amount, in general, follows gamma distribution, especially in winter, and has an inversed "J" shape, while in summer the distribution bears a single-peak, positively skewed, normal distribution with a long tail. Thus, except for s_n, all other wet-day probability distribution density functions can be approximated as square-shaped uniform distributions, while the probability distribution density function for s_n can be approximated with an exponential function (Ding and Zhang, 1989). In other words, the probability distribution function of states $s_0 - s_2$ is

$$f_i(x) = \frac{1}{c_i - c_{i-1}} \quad (c_{i-1} < x \leqslant c_i\ , \quad i = 0, 1, 2)\ , \quad (1)$$

where c_i and c_{i-1} are the upper and lower limits of the dry-day cloud amount, respectively. Wet-day states $s_3 - s_{n-1}$ also follow uniform distribution approximately:

$$f_i(x) = \frac{1}{m_i - m_{i-1}} \quad (m_{i-1} < x \leqslant m_i,\ i = 3, \ldots, n-1)\ , \quad (2)$$

where m_i denotes the upper limit of wet-day precipitation amount, while the distribution for state s_n is approximately defined by the following exponential function

$$f_n(x) = \lambda e^{-\lambda(x-b)} \quad x \in s_n\ . \quad (3)$$

Here $f_n(x)$ is called the probability density function of displaced exponential distribution, λ is the distribution parameter, and b is the upper limit for state s_{n-1}. The matrix of transition probabilities for $s_0, s_1, s_2, \ldots, s_n$ can be estimated from historical data for the cloud amounts and for precipitation. Then the corresponding probability distributions for each s_i can be estimated using these historical data. Apparently, once the boundaries of the states are known, the probability density functions of the distribution can be determined, while for exponential distribution its probability density function is mainly determined by estimating parameter λ. With all of these, daily precipitation records can be simulated by utilizing simulation methods for discrete random variables (Yao, 1984). Based on the Chapman–Kolmgoroff equation (stationary transition formula) for homogeneous Markov chain, given an initial state and its probability vector $\boldsymbol{p}(0)$, the matrix of transition probabilities at any first k steps can be derived. If an initial vector $\boldsymbol{p}(0) = [\boldsymbol{p}_0(0), \boldsymbol{p}_1(0), \ldots\ldots \boldsymbol{p}_n(0)]$ is known, then the probability vector of daily precipitation process reaching states $s_0, s_1, s_2, \ldots, s_n$ after k steps of transition must be

$$\boldsymbol{p}(k) = \boldsymbol{p}(0)\boldsymbol{p}(0, k) = \boldsymbol{p}(0)\boldsymbol{p}^k \quad (k = 1, 2, \ldots)\ , \quad (4)$$

where $\boldsymbol{p}(0, k)$ represents the transition probability from time 0 to time k. Or it can be written as

$$\boldsymbol{p}_j^{(k)} = \boldsymbol{p}(0)\boldsymbol{p}^k\ , \quad (5)$$

where $\boldsymbol{p}_j^{(k)}$ is a row vector in Eq. (5), and $\boldsymbol{p}(0)$ is initial probability vector, $p_j(0) = 1$, all others are 0, and j can be any one of 0,1,...,n (Green, 1970).

2.2 *The simulation scheme for initial day and its transition process*

To eliminate seasonal effect on the results, in the simulation daily weather changes in a year may be divided into several stages based on the local seasonality within the year. In this study, a scheme based on monthly calculations was adopted. Specifically, for each month, simulation data were obtained using the aforementioned model. The detailed simulation scheme, which is the same as that of Ding and Zhang (1989), is presented here.

2.3 *The simulation scheme for daily solar radiation, temperature, humidity*

Studies (Ding and Zhang, 1989, Ding et al., 2009) have shown that better simulation results of daily rainfall amount can be obtained using the Markov chain model, while the Markov chain model simulation of daily temperature and solar radiation have not been as successful. This is because daily temperature and solar radiation are continuous stochastic variables with strong autocorrelation. To obtain statistical characteristics of simulated daily temperature and solar radiation to perform various numerical experiments under different external climate conditions, we used autocorrelated, multivariate regression equations to simulate daily temperature and solar radiation and their statistical characteristics. Here a multivariate linear autoregression model for daily temperature (mean, maximum, and minimum) and solar radiation was constructed. The autoregression model for daily maximum (minimum) temperature and solar radiation is

$$\boldsymbol{\chi}_t(j) = \boldsymbol{A}\boldsymbol{\chi}_{t-1}(j) + \boldsymbol{B}\boldsymbol{\varepsilon}_t(j) \ , \tag{6}$$

where $\boldsymbol{\chi}_t(j)$ is a 3×1 matrix for standardized variables: maximum temperature ($j = 1$), minimum temperature ($j = 2$), and solar radiation ($j = 3$) on the given day t. Also, $\boldsymbol{\chi}_{t-1}(\boldsymbol{j})$ is a 3×1 matrix for the same corresponding variables on the previous day (day $t-1$). Then

$$\chi_t(1) = \frac{T_{\rm mx} - T_{\rm mx,mon}}{\sigma_{\rm mx,mon}} \ , \tag{7}$$

$$\chi_t(2) = \frac{T_{\rm mn} - T_{\rm mn,mon}}{\sigma_{\rm mn,mon}} \ , \tag{8}$$

$$\chi_t(3) = \frac{H_{\rm day} - R_{\rm mon}}{\sigma_{\rm r,mon}} \ , \tag{9}$$

Where $T_{\rm mx}, T_{\rm mn}$, and $H_{\rm day}$ are daily maximum temperature, minimum temperature, and solar radiation, respectively (units of temperature are °C, units of solar radiation are in MJ m^{-2}); $T_{\rm mx,mon}, T_{\rm mn,mon}$, and $R_{\rm mon}$, are monthly average values of daily maximum temperature, daily minimum temperature, and daily solar radiation, respectively; and $\sigma_{\rm mx,mon}$, $\sigma_{\rm mn,mon}$, and $\sigma_{\rm r,mon}$, are the corresponding standard deviations for daily maximum temperature, daily minimum temperature, and daily solar radiation. In Eq. (6), $\boldsymbol{\varepsilon}_i$ is a 3×1 matrix for independent random residuals, and $\boldsymbol{A}$, $\boldsymbol{B}$ are 3×3 matrices whose constant elements are autocorrelation and cross-correlation coefficients of the time series. Matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are defined as

$$\boldsymbol{A} = \boldsymbol{M}_1 \cdot \boldsymbol{M}_0^{-1} \ , \tag{10}$$

$$\boldsymbol{B} \cdot \boldsymbol{B}^{\rm T} = \boldsymbol{M}_0 - \boldsymbol{M}_1 \cdot \boldsymbol{M}_0^{-1} \cdot \boldsymbol{M}_1^{\rm T} \ , \tag{11}$$

where $\boldsymbol{M_0}$ and $\boldsymbol{M_0^{-1}}$ are the cross-correlation matrices for the three variables on the same given day and the inverse matrix, respectively, and $\boldsymbol{M}_1$ and $\boldsymbol{M}_1^{\rm T}$ denote a one-day, lagged, cross-correlation matrix and its transpose, respectively. We have

$$\boldsymbol{M}_0 = \begin{bmatrix} 1 & \rho_0(1,2) & \rho_0(1,3) \\ \rho_0(2,1) & 1 & \rho_0(2,3) \\ \rho_0(3,1) & \rho_0(3,2) & 1 \end{bmatrix} \tag{12}$$

$$\boldsymbol{M}_1 = \begin{bmatrix} \rho_1(1,1) & \rho_1(1,2) & \rho_1(1,3) \\ \rho_1(2,1) & \rho_1(2,2) & \rho_1(2,3) \\ \rho_1(3,1) & \rho_1(3,2) & \rho_1(3,3) \end{bmatrix} \tag{13}$$

where $\rho_0(j,i)$ and $\rho_1(j,i)$ are cross-correlation coefficients between j and i on the same day and one day lagged, respectively, and $j,i = 1,2,3$ denote daily maximum temperature, minimum temperature, and solar radiation, respectively. In general, these matrices can be calculated with observed data for a given location. Based on the calculated values of the standardized variables and using Eqs. (6)–(13) and the corresponding standardization formula, it is not difficult to obtain the simulated daily maximum temperature, daily minimum temperature, and solar radiation. The specific equations are

$$T_{\rm mx} = T_{\rm mx,mon} + \chi_t(1)\sigma_{\rm mx,mon} \ , \tag{14}$$

$$T_{\rm mn} = T_{\rm mn,mon} + \chi_t(2)\sigma_{\rm mn,mon} \ , \tag{15}$$

$$H_{\rm day} = R_{\rm mon} + \chi_t(2)\sigma_{\rm r,mon} \ . \tag{16}$$

Notably, the standard deviation $\sigma_{\rm r},_{\rm mon}$ for daily solar radiation during each month can be estimated as 1/4 of the difference between the maximum and the mean values. The estimation formula (Neitsch et al., 2002, 2005) may be written as

$$\sigma_{\rm r,mon} = \frac{S_{\rm xm} - R_{\rm mon}}{4} \ , \tag{17}$$

where $S_{\rm xm}$ represents the maximum values of solar radiation (MJ m^{-2}) that reach the Earth's surface on any given day in the month. In simulating the aforementioned three variables, if the daily weather states are not taken into consideration, then the simulated daily maximum temperature, minimum temperature, and solar radiation sequences are representative of those under clear sky conditions. Thus, a correction has to be made. A common approach is to adjust the simulated sequences based on weather conditions. The revised formula for daily maximum temperature under wet-day conditions is

$$T_{\rm mx,mon,w} = T_{\rm mx,mon,d} - b_T\left(T_{\rm mx,mon} - T_{\rm mn,mon}\right) \ , \tag{18}$$

where there is an implicit relationship:

$$T_{\rm mx,mon}d_{\rm tot} = T_{\rm mx,mon,w}d_{\rm wet} + T_{\rm mx,mon,d}d_{\rm dry} \ , \quad (19)$$

$T_{\rm mx,mon,w}$ denotes the average daily maximum temperature for wet days in the month; $T_{\rm mx,mon,d}$ denotes the average daily maximum temperature for dry days in the month; and, $d_{\rm tot}$, $d_{\rm wet}$, and $d_{\rm dry}$ represent the total number of days, the number of wet days, and the number of dry days in the month, respectively. Moreover, parameter b_T is a scaling factor whose value can be estimated from related data.

Similarly, the revision formula for dry day conditions is

$$T_{\rm mx,mon,d} = T_{\rm mx,mon} + b_T\frac{d_{\rm wet}}{d_{\rm tot}}(T_{\rm mx,mon} - T_{\rm mn,mon}) \ . \quad (20)$$

Thus, the formula for daily maximum temperature in wet days can be written:

$$T_{\rm mx,w} = T_{\rm mx,mon,w} + \chi_t(1)\sigma_{\rm mx,mon} \ . \quad (21)$$

And the formula for daily maximum temperature in dry days is

$$T_{\rm mx,d} = T_{\rm mx,mon,d} + \chi_t(1)\sigma_{\rm mx,mon} \ . \quad (22)$$

Formulas for daily minimum temperature in dry and wet days can also be derived in similar fashion.

The revision formula for daily solar radiation in dry days is

$$R_{\rm mon,d} = b_R R_{\rm day,d} \quad (23)$$

Where an implicit relationship exists:

$$R_{\rm mon}d_{\rm tot} = R_{\rm mon,w}d_{\rm wet} + R_{\rm mon,d}d_{\rm dry} \ . \quad (24)$$

In Eq. (24), $R_{\rm mon,w}$ and $R_{\rm mon,d}$ are the average daily solar radiation in wet and dry days in the month, respectively. Parameter b_R can be estimated from related data. From Eqs. (20) and (21), we have

$$R_{\rm mon,d} = \frac{\mathrm{R}_{\rm mon}d_{\rm tot}}{b_R d_{\rm wet} + d_{\rm dry}} \ . \quad (25)$$

Herein, formula for daily solar radiation in wet days can be written:

$$R_{\rm day,w} = R_{\rm mon,w} + \chi_t(2)\sigma_{\rm r,mon} \ . \quad (26)$$

Formula for daily solar radiation in dry days is

$$R_{\rm day,d} = R_{\rm mon,d} + \chi_t(3)\sigma_{\rm r,mon} \ . \quad (27)$$

Simulation of relative humidity follows similar procedures; details are not included here.

3. Numerical simulations

3.1 *Simulation of precipitation*

To compare simulated results from different simulation schemes, scheme A was used to denote first-order multistate Markov chain model; scheme B was used for a first-order two-state Markov chain model (rainfall follows gamma distribution); scheme C was used for a first-order two-state Markov chain model (rainfall follows skewed distribution); scheme D was used for a first-order two-state Markov chain model (rainfall follows exponential distribution); and scheme E was used for an improved multistate Markov chain model (cloud amount in dry days is considered), respectively. Taking Beijing station as an example, Table 1 lists the simulated results of rainfall characteristics from these five schemes. Based on the observed daily precipitation in 30 years (1961–1990), daily precipitation in each month over 54 years was simulated. Table 1 compares the simulated with observed precipitation data in seven characteristics (i.e., number of dry runs, number of wet runs, monthly rainy days, standard deviation of daily rainfall, daily rainfall, standard deviation of monthly rainfall amount, and maximum of daily rainfall) to assess the five simulation schemes. To obtain stable rainfall simulation, five simulations were created, and the average of these five simulations was used for each scheme. As shown in Table 1, simulations of dry and wet runs and monthly rainy days for all schemes were relatively accurate (relative errors are generally $<5\%$). Simulation of average daily rainfall was also very good. This was especially true for those months in which average daily rainfall is relatively small (January and April in Table 1) with the absolute errors between the simulated rainfall and the observation data of $\leqslant 0.1$ mm. For those months of larger mean daily rainfall (July and October in Table 1), the relative errors between the simulated rainfall and the observations data are also $<5\%$. However, standard deviation of daily rainfall, standard deviation of monthly rainfall, and maximum of daily rainfall, simulations from all the schemes had some errors. Among these the simulations with schemes A and E were closer to the observed distribution and were markedly better than the simulations of the other three schemes. To compare the ultimate results from various simulation schemes, Table 2 presents the comparison of simulated rainfall frequencies with those observed in April at Beijing. As far as the comparison of the χ^2 statistic is concerned, schemes A and E were closer to the observed distribution. This is consistent with the results shown in Table 1.

Simulated average monthly rainy days at Shanghai from schemes A and E are shown in Fig. 1, in which

Table 1. Simulations results (the average of five simulations with each schemes) of rainfall characteristics in January, April, July, and October at Beijing from five simulation schemes.

Month	Model scheme	Number of dry runs	Number of wet runs	Monthly rainy days	Variance of daily rainfall	Daily rainfall	Variance of monthly rainfall	Max of daily rainfall
Jan.	A	24.1	1.5	1.9	0.69	0.1	4.96	14.5
	B	22.4	1.7	2.2	0.66	0.1	4.32	10.6
	C	21.0	1.5	2.1	0.69	0.1	3.91	14.2
	D	21.7	1.6	2.2	0.59	0.1	4.31	9.00
	E	22.7	1.5	2	0.7	0.1	4.49	14.9
	Average	22.4	1.6	2.1	0.7	0.1	4.4	12.6
	Observed	22.7	1.6	2	0.6	0.1	4.5	14.6
	Error(%)	1.4	2.5	4	4.1	0	1.8	13.4
Apr.	A	8.6	1.6	4.6	3.39	0.7	21.96	49.4
	B	8.1	1.6	4.9	2.92	0.8	17.84	36.6
	C	8.0	1.5	4.8	3.01	0.7	19.51	41.7
	D	8.3	1.6	4.8	2.41	0.7	15.01	29.4
	E	8.4	1.6	4.8	3.69	0.8	21.64	48.4
	Average	8.3	1.6	4.8	3.1	0.7	19.2	41.1
	Observed	8	1.5	4.8	3.2	0.7	24.5	51.0
	Error(%)	3.5	5.3	0.4	3.6	5.7	21.5	19.4
Jul.	A	2.7	2.2	13.7	15.4	6.1	99.5	173.8
	B	2.7	2.2	13.8	13.8	5.7	78.4	173.9
	C	2.8	2.2	13.8	15.1	5.4	87.6	192.2
	D	2.6	2.3	14.3	10.9	6	68	96.5
	E	2.7	2.1	13.6	14.9	6	85.0	153.0
	Average	2.7	2.2	13.8	14.0	5.8	83.7	157.9
	Observed	2.7	2.2	14.0	15.8	5.9	98.3	244.2
	Error(%)	0	0	1.1	11.2	1.0	14.9	35.3
Oct.	A	7.5	1.6	5.4	3.8	0.7	22.8	65.1
	B	7.3	1.5	5.4	2.9	0.7	17.7	40.7
	C	7.1	1.6	5.7	3.5	0.7	21.9	62.3
	D	7.2	1.6	5.6	2.4	0.7	14.7	24.7
	E	7.5	1.6	5.4	3.9	0.8	26.4	73.3
	Average	7.3	1.6	5.5	3.3	0.72	20.7	53.2
	Observed	7.3	1.6	5.5	3.5	0.7	22.7	49.4
	Error(%)	0.3	1.3	0.0	6.4	2.9	8.6	7.7

Table 2. Comparison of simulated rainfall frequency distribution from five schemes with the observed in April at Beijing, the units for rainfall is mm d^{-1}.

Rainfall	A	B	C	D	E	Observed
p <0.1	923.2	980.6	940.9	932.4	923.8	919
0.1⩽ p <6.3	413.7	335.4	414.3	446.3	411.2	418
6.3⩽ p <50.7	293.7	320.5	280.3	278.2	291.1	296
p ⩽50.7	43.4	37.5	38.5	17.1	45.9	41
χ^2 Statistic	0.2110	16.7125	1.5398	16.2645	0.8024	

simulation results have a correlation coefficient of 0.97 with the observed distribution for scheme A and 0.98 for scheme E. Therefore, the simulated rainfall features from both schemes A and E have considerable accuracy. Simulations of daily rainfall for all months at other sites showed that simulated results from schemes A and E are all very close to the observed distribution, with correlation coefficients ⩾0.97 (figures not shown). In addition, according to Yao (1984), the matrices of transition probabilities based on 30-year samples are very stable. Thus, roughly speaking, errors due to sampling are very small. Consequently, simulation results of rainfall characteristics from both schemes A and E are better than those from the first-order two-state Markov chain model (gamma distribution) and from the weather generator defined in the original SWAT model. The multistate Markov chain scheme with improved state classification (scheme E)

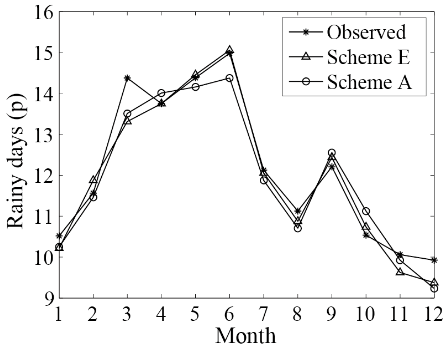

Fig. 1. Comparison of simulated monthly average rainy days at Shanghai.

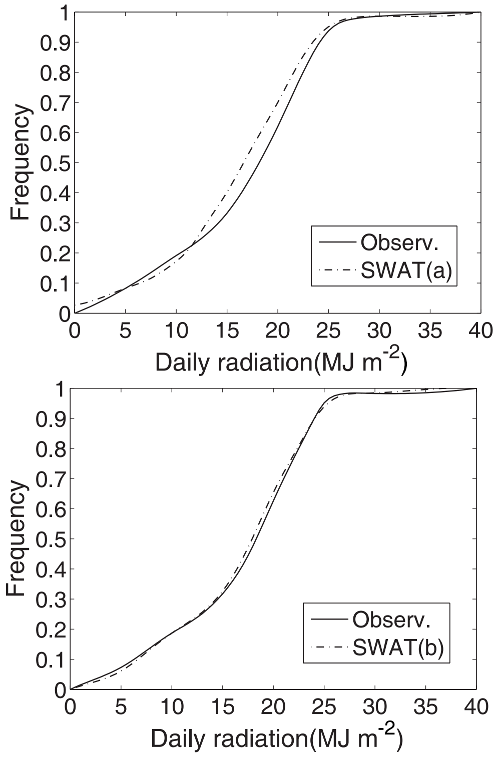

Fig. 2. Comparison of simulated cumulative frequency distribution from SWAT (a) and improved SWAT (b) for daily solar radiation with the observed in April at Ha'erbin.

is more reasonable, in theory, because the cloud amount in dry days is divided into three states: clear, cloudy, and overcast (rain free).

3.2 *Simulation of solar radiation and temperature*

To increase the accuracy in simulating solar radiation and temperature, the weather generator in the original SWAT model had to be improved. To do so, we incorporated cloud amount in a first-order Markov chain model with a gamma distribution simulation method to create a first-order multistate Markov chain model (using gamma distribution simulation for precipitation, whereas joint autoregression simulation was used for temperature and solar radiation). Still using the aforementioned six representative stations as examples, we used 30 years of observation data (1961–1990) to build the models and compared the simulations (46 years at Beijing, 30 years at Shanghai, and 43 years at other stations) with the corresponding observations over the same period of time. The results show that time series of simulated solar radiation and temperature from improved SWAT model with added cloud amount are much better than those from the original SWAT model.

Based on our assumption in the SWAT model, maximum temperature, minimum temperature, and solar radiation under clear and cloudy conditions all had normal distribution. Sample sizes under clear and cloudy conditions, however, were not the same, and thus the stability of the samples is not the same. For instance, both dry-day and wet-day maximum temperature followed normal distribution, but the wet-day sample size was smaller than dry-day sample size. Thus, wet-day normal distribution was less stationary than that of dry days. To make the wet-day normal distribution more stationary, we modified the simple definition in SWAT model by defining those rain-free days with cloud amount >70% as cloudy days to increase the sample size of cloudy days. At the same time, the distribution with the two definitions for dry days were nearly equally stationary, but the modified definition with respect to cloud amount made cloudy day maximum temperature distribution more stationary and led to an apparent decrease in dispersion in sample data. A similar improvement was achieved for daily minimum temperature. As a result of the modified definition, daily solar radiation values improved and became more stationary even more remarkably.

Figures 2a and 2b show the comparison graphs of simulated cumulative probability of solar radiation (43 years) from the original SWAT model and the improved SWAT model with cloud amount included with the observation data. It can be seen from the figure that simulated cumulative probability of daily solar radiation from the improved SWAT model with cloud amount included is remarkably close to the observations (i.e., they almost completely overlap). However, a larger discrepancy was seen between simulated and observed cumulative probability for daily solar radiation from the original SWAT model and the observation data was in the range of 15–25 MJ m^{-2}. This

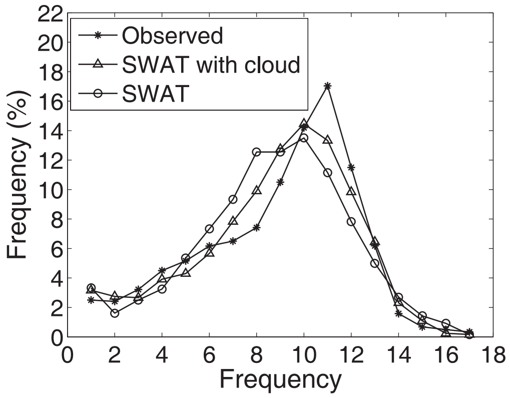

Fig. 3. Comparison of frequency distributions from simulated daily solar radiation with improved SWAT model with cloud amount with the observation in April at Ha'erbin.

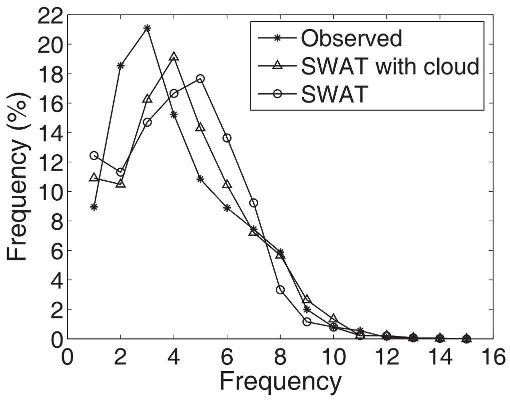

Fig. 4. Comparison of frequency distributions from simulated daily solar radiation with improved SWAT model with cloud amount with the observation in October at Chengdu.

means that the improved model is superior in solar radiation simulation (Figs. 2a and 2b).

Figure 3 displays the simulated solar radiation frequency distribution in April at Ha'erbin. From this figure, the frequency curve from the improved SWAT model can be seen; the cloud amount is closer to that of the observation data. The correlation coefficient of the original SWAT model curve with that of the observation data is only 0.87, whereas that of the improved SWAT model with that of the observation data is as high as 0.96. The increase in accuracy in solar radiation simulation from the improved SWAT model is obvious. Figure 4 shows the comparison curves of the simulations of two models and the observed frequency of solar radiation in October at Chengdu. In terms of the observed frequency, the simulated solar radiation (October) at Chengdu shows an apparent negative skew, whereas the simulated frequency curve from the improved SWAT model with cloud amount is closer to the observation data, manifested by a correlation coefficient of 0.91. The correlation coefficient between the original SWAT model simulated curve and that of the observation data, on the other hand, is merely 0.87. The simulation results of daily solar radiation in various months at the aforementioned representative stations show obvious progress in the improved SWAT model simulations. In fact, the models were built by modifying the definition of clear and cloudy days and are more stationary; the simulated solar radiation is therefore destined to better approximate the observations.

Taking Wuhan station as an example, Table 3 displays several climatic statistics of simulated temperature and solar radiation by two models compared with those of the observation data. The simulated maximum temperature and minimum temperature from both models have high accuracy with a relative error of $<5\%$ (except for January near 0 average minimum temperature results in a larger relative error, but the absolute error is $<0.7°C$). At the same time, the accuracy in simulating the extreme values of daily maximum and minimum temperatures is also relatively high (except for January, for the aforementioned reasons), and the relative error for the improved SWAT model is smaller than the original SWAT model. Although the simulation errors are bigger due to the skewness in the simulated solar radiation from both models, the accuracy of the improved SWAT model with cloud amount, relatively speaking, is higher, considerably so in some months. For example, at Wuhan station the relative error of the simulated daily solar radiation from the improved SWAT model with cloud amount is only 4% in April and is 4.9% in October, whereas the relative errors from original SWAT model are 19.5% and 17.1%, respectively. This demonstrates that improved SWAT model with cloud amount has advantages in simulating solar radiation. Notably, the aforementioned skewedness of daily solar radiation was referred to as resulted skew coefficient of simulated or observed frequency distribution under the assumption that in the SWAT model solar radiation follows normal distribution in theory, which was chosen as one of the indicators to assess the simulation.

4. Conclusions

In this study, a total cloud amount and finer classification of dry states were used, based on a continuous-time watershed model (soil water assessment tool, SWAT) developed at the USDA-ARS. Dry days were classified into three states: clear, cloudy, and overcast (rain-free). A new kind of model with a finer classifica-

Table 3. The climatic statistic of simulated and measured results from two models for temperature and radiation over Wuhan.

Month	Scheme	Mean of max temp	Mean of min temp	Mean of daily radiation	Skewness of daily radiation	Std of monthly radiation
Jan.	A(S)	8.2154	−0.4792	7.6995	0.1660	37.514
	Observed	8.2558	0.1350	7.2311	0.0945	59.568
	Error %	0.49%	454.87%	6.48%	75.54%	37.02%
	B (S)	8.0769	−0.4257	7.6129	0.0422	37.698
	Observed	8.2558	0.1350	7.2311	0.0945	59.568
	Error %	2.16%	415.27%	5.28%	55.41%	36.71%
Apr.	A(S)	21.2042	12.4646	13.2820	0.1551	50.948
	Observed	21.2400	12.7050	12.6990	0.1298	80.796
	Error %	0.17%	1.90%	4.60%	19.52%	36.94%
	B (S)	20.7524	12.1258	12.9010	0.1246	56.412
	Observed	21.2400	12.7050	12.6990	0.1298	80.796
	Error %	2.30%	4.56%	1.60%	4.00%	30.18%
Jul.	A (S)	32.783	25.1014	19.0810	−0.3405	72.672
	Observed	32.884	25.4880	18.3880	−0.5320	122.230
	Error %	0.31%	1.52%	3.77%	35.99%	40.54%
	B (S)	32.7136	25.1094	18.8884	−0.4044	77.194
	Observed	32.884	25.4880	18.3880	−0.5320	122.230
	Error %	0.52%	1.46%	2.72%	23.98%	36.84%
Oct.	A (S)	22.5730	13.4600	11.2700	−0.1231	52.567
	Observed	22.7830	13.8410	10.8190	−0.1485	63.001
	Error %	0.92%	3.48%	4.17%	17.08%	16.56%
	B (S)	22.7574	13.5350	11.2462	−0.1413	53.0278
	Observed	22.7830	13.8410	10.8190	−0.1485	63.001
	Error %	0.11%	2.21%	3.95%	4.87%	15.83%

Note：A(S)：SWAT；B(S)：SWAT with cloud amount.

tion of weather states for daily weather simulation was used to accomplish a comprehensive and downscaled simulation of regional daily climatic states and to improve the SWAT model simulations. Main conclusions are as follows.

(1) An improved scheme for classification of dry- and wet-day states in the SWAT model, based on a first-order multistate Markov chain model has been proposed. Specifically, a complete daily weather state vector was formed by combining three states (i.e., clear, cloudy, and overcast) for dry days classified based on total cloud amount and a wet-day state classified in the original SWAT model to calculate the matrix of transition probabilities. The improved daily weather-state vector was more effective in the comprehensive simulation of the complete structure of daily weather states (including precipitation, temperature, solar radiation, and other characteristics).

(2) Numerical simulation results of representative stations demonstrate that simulated daily precipitation with scheme A (first-order multistate Markov chain model) and scheme E (improved multistate Markov chain model with total cloud amount) are very close to the observation data. For temperature and solar radiation, the improved SWAT model with total cloud amount is more effective. Regarding the improved SWAT model, the weather generator in the original SWAT model was modified to first become a first-order multistate Markov chain weather generator with which daily weather states (i.e., clear, cloudy, overcast, and rainy) were determined. Then rainfall was simulated with gamma distribution, and temperature and solar radiation were simulated using a joint autoregression model.

(3) The improved, finer classification, downscaling model can be used to predict future climatic states, but it can be used to interpolate missing temporal–spatial data and to extrapolate data as well.

Acknowledgements. This work was supported jointly by the National Natural Science Foundation of China (Grant No. 40875058), the Natural Science Key Research of Jiangsu Province High Education (Grant No. 07KJA17020), the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), the National Key Technologies Research and Development Program (Grant No. 2008BAK50B02-04-01), and the CMA Meteorological Special Science Foundation (Grant

No. GYHY200706030).

REFERENCES

Bardossy, A., and E. J. Plate, 1991: Modeling daily rainfall using a semi-Markov representation of circulation pattern occurrence. *J. Hydrol.* **122**, 33–47.

Bouraoui, F., L. Galbiati, and G. Bidoglio, 2002: Climate change impacts on nutrient loads in the Yorkshire Ouse catchment (UK). *Hydrology and Earth System Sciences*, **6**(2), 197–209.

Chin, E. H., 1977: Modeling daily precipitation occurrence process with Markov chain. *Water Resources Res.*, **13**(6), 949–956.

Ding, Y. G., 1994: Research of universality of Gamma distribution model of precipitation. *Scientia Atmospherica Sinica*, **18**(5), 552–560. (in Chinese)

Ding, Y. G., and Y. C. Zhang, 1989: A stochastic simulation test for climatological features of precipitation. *Journal of Nanjing Institute of Meteorology*, **12**(2), 146–154. (in Chinese)

Ding, Y. G., and T. Niu, 1990: A Markov chain simulation for dry and wet month runs. *Journal of Nanjing Institute of Meteorology*, **13**(3), 286–296. (in Chinese)

Ding, Y. G., B. Y. Cheng, and Z. H. Jiang, 2008: A newly-discovered GPD-GEV relationship together with comparing their modelings of extreme precipitation in summer. *Adv. Atmos. Sci.*, **25**(3), 507–516, doi: 10.1007/s00376-008-0507-5.

Ding, Y. G., J. L. Zhang, and Z. H. Jiang, 2009: Simulation experiments of extreme precipitation based on multi-status Markov Chain Model. *Acta Meteorologica Sinica*, **67**(1), 20–27. (in Chinese)

Green, J. R., 1970: A generalized probability model for sequences of wet and dry days. *Mon. Wea. Rev.*, **93**, 338–341

Liao, Y. M., Q. Zhang, and D. L. Chen, 2004: The precipitation simulation using a stochastic weather generator over China region. *Acta Geographica Sinica*, **59**(5), 689–698. (in Chinese)

Neitsch, S. L., J. G. Arnold, J. R. Kinlry, J. R. Williams, and K. W. King, 2002: Soil and water assessment tool theoretical documentation: Version 2000. TWRI Report TR-191, Texas Water Resources Institute, College Station, Texas, 483pp.

Neitsch, S. L., J. G. Arnold, J. R. Kiniry, and J. R. Williams, 2005: Soil and water assessment tool theoretical documentation: Version 2005. Texas Water Resources Institute, College Station, Texas, 476pp

Katz, R. W., 1974: Computing probabilities associated with Markov chain model for precipitation. *J. Appl Meteor*, **15**, 953–954.

Palutikof, J. P., C. M. Goodess, S. J. Watkins, and T. Holt, 2002: Generating rainfall and temperature scenarios at multiple sites: examples from the Mediterranean. *J. Climate*, **15**, 3529–3548.

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

Racsko, P., L. Szeidl, and M. Semenov, 1991: A serial approach to local stochastic weather models. *Ecological Modelling*, **57**, 27–41.

Yao, Z. S., 1984: *Basis of Climatic Statistics*. Science Press, Beijing, China, 594pp. (in Chinese)

Yao, Z. S., and Y. G. Ding, 1990: *Climate Statistics*. Chinese Meteorological Press, Beijing, 698–721. (in Chinese)

Zhang, Y. C., 1990: A stochastic distribution model for N daily rainfall. *Journal of Nanjing Institute of Meteorology*, **13**(1), 23–30. (in Chinese)

Zhang, Y. C., and Y. G. Ding, 1990: Statistical characteristics of daily rainfall series at five representative stations in Eastern China. *Journal of Nanjing Institute of Meteorology*, **13**(2), 194–203. (in Chinese)