YRB 81-20

Stochastic Simulation of Daily Climatic Data for Agronomic Models[1]

G. A. Larsen[2] and R. B. Pense[3]

ABSTRACT

Many agronomic models require the input of daily climatic data. Simulated climatic data may be used when long series of historic data are not available or convenient, or when future data are needed. A stochastic weather simulation model was developed and validated for a wide range of climates. The model produces possible daily sequences of precipitation amount, maximum and minimum air temperature, and total solar radiation at the earth's surface for the entire year.

A first-order, two-state Markov chain is used to simulate the occurrence of wet and dry days. Probabilities are used to simulate the occurrence of trace precipitation amounts on wet days. A two-parameter gamma distribution conditioned by the precipitation status on the previous day is used to generate greater than trace amounts. Two bivariate normal distributions conditioned by the precipitation status on the current day are used to simulate current temperature deviations from long-term average temperature curves. A two-parameter gamma distribution simulates current solar radiation deviations from the calculated maximum clear day radiation on dry days. On wet days, the deviations are simulated with a two-parameter beta distribution.

The model was developed with data from Columbia, Mo. Model validation was done for Columbia, Albuquerque, N.M., Caribou, Maine, Medford, Ore., and Miami, Fla. Various statistical tests were done to detect significant differences in central tendency, dispersion, and distribution. Comparisons were made to the base period used for parameter estimation (ranged from 16 to 20 years) and also to the 80-year period of record available at Columbia. The validation showed that the model produced climatic data which generally did not differ significantly from the base period at any of the locations. At Columbia, it was determined that the 17-year base period was not long enough to adequately represent the 80 years of precipitation data.

Additional index words: Markov chain, Probability distribution, Precipitation, Temperature, Solar radiation.

THERE have been several recent efforts to stochastically simulate possible sequences of daily precipitation occurrence and amount, maximum and minimum air temperature, and total solar radiation received at the earth's surface (5, 12, 13). While the goal of the presently proposed model is the same, it is believed that the methodology differs enough to warrant separate consideration. This study is an expansion of earlier work by Bond (4) in which precipitation and maximum and minimum temperature were simulated for the May through August growing season. The methodology has been refined somewhat for these variables, solar radiation has been added, and the entire model has been expanded to be appropriate for the full year.

Simulated daily weather variates can be used in a variety of settings to replace long series of historic data which may not be available, convenient or appropriate. Simulated data can be used in hydrologic models for watershed planning, evaluation, and design purposes (12). Simulated data can be used in various types of agricultural management models to assess the risk associated with different alternatives (5). In a realtime mode, possible future sequences of data can be used in plant simulation models to make yield forecasts (1). The proposed weather simulation model has been used to estimate probabilities associated with segments of plant model response curves to better judge which input variables realistically produce the greatest change in model output (9).

How closely a stochastic weather simulation model needs to represent the real system depends on the application. While the model can become quite complex, clearly there has to be a balance between complexity and the foreseen uses, or effort may be largely wasted or, at best, simply academic. In view of this, the proposed model is intended to produce simulated data which are statistically comparable to data from the real system in measures of central tendency, dispersion, and distribution while preserving major interrelationships among the variables. The model is also intended to be applicable to a wide range of locations at any time of the year. A rather extensive model validation was done to assess these claims.

The primary purpose of this paper is to present the methods used in the model and a brief set of results to demonstrate the performance. Much of the discussion related to why certain methods were used and complete results of the validation have been omitted for sake of brevity. This information and some suggestions for possible refinement of the model are available on request in an internal staff report (10). The software to run the model is also available.

MATERIALS AND METHODS

Data Base

The data base for model development came from Columbia, Mo. This data consisted of 80 years (1890–1969) of precipitation and temperature values, and 22 years (July 1952–June 1974) of daily solar radiation values. Parameter estimates came from the 17-year period (1953–1969) in which all climatic variables were available. Data for model validation were also obtained for four other locations representing a wide range of latitude, altitude, and precipitation pattern. Twenty years of daily climatic data (1951–1970) were obtained for Albuquerque, N.M., Caribou, Maine, Medford, Ore., and Miami, Fla. Together, the five sites range in latitude from 26° at Miami to 47° at Caribou. Altitudes go from 5 m at Miami to 1620 m at Albuquerque. Average annual precipitation amounts range from less than 20 cm at Albuquerque to about 150 cm at Miami.

Parameter estimates for the additional four sites were made from the entire 20 years of available data at Medford. However, due to missing daily solar radiation values in excess of 20% for some years, the base period of Albuquerque was 19 years, for Caribou, 16 years, and for Miami, 18 years. Since missing solar radiation observations were not likely to be distributed randomly, entire years were left out of the parameter estimation to avoid the possibility of introducing bias.

Precipitation Occurrence

A first-order Markov chain was used to simulate the occurrence of precipitation. A first-order Markov chain has

---

[1] Contribution from the USDA, Statistical Reporting Service, Washington, D.C. Received 14 Sept. 1981.

[2] Mathematical statistician, USDA-SRS, P.O. Box E, Fort Collins, CO 80522.

[3] Mathematical statistician, USDA-SRS, Room 4833, South Bldg., Washington, DC 20250.

been used satisfactorily in a number of studies [e.g. see (12) or (13) for a list of references]. In the earlier work by Bond, (4) it was shown that the first-order was appropriate for the months June, July and August but not for May in Columbia, Mo. Bruhn et al. (5) showed that the first-order was appropriate for May, June, July, and September but not for August in Geneva, N.Y.

Two states were used in the Markov chain—wet and dry. A wet day is defined to occur whenever a trace or larger amount of precipitation is recorded. Dry days are days which are not wet. The decision to include trace amounts in the wet category arose primarily from solar radiation simulation considerations.

The assumption underlying the first-order Markov chain is that the probability that the current day is in a particular state (i.e., wet or dry) depends only on the state of the previous day. It has been further assumed that these so-called transition probabilities are independent of the particular day within individual months. The monthly dry day transition probabilities were estimated by counting the number of dry days preceded by a dry day and the number of dry days preceded by a wet day over the base period of record and dividing by the total number of dry previous days and wet previous days, respectively. The corresponding transition probabilities for wet days can be obtained by subtracting the dry day transition probabilities from one.

The wet state was subdivided into occurrences of trace and greater than trace precipitation amounts. The probability that a trace amount occurs on a wet day was computed by month.

Precipitation occurrence is simulated for each day by obtaining a random uniform number between zero and one, inclusive. If the random number exceeds the transition probability of a dry day, then the current day is wet, otherwise it is dry. If the current day is wet, another random uniform number is obtained. If the second random number is smaller than the probability of a trace occurrence, then a trace occurs, otherwise, an amount greater than a trace occurs.

Precipitation Amount

A two-parameter gamma distribution was used to simulate precipitation amounts greater than a trace on wet days. This distribution has been widely used in the past [e.g. (5) and (8)]. The general form of the gamma probability density function has a third parameter, γ, which establishes the lower bound for the random variable X. For precipitation amount we assume $\gamma = 0$ which, indeed, is reasonable since amounts will approach zero but will not be equal to or less than zero. Setting $\gamma = 0$ leaves two parameters, α and β, to be estimated. The gamma distribution has two quite different shapes depending on whether α is less than one or greater than or equal to one. In the first case, the distribution has a reverse "J" shape in the first quadrant where the curve goes asymptotic to both the x and y axes. The second case results in a curve in the first quadrant starting near the origin and then resembling a normal curve with a positive (right) skew eventually going asymptotic to the x-axis. The two-parameter gamma with $0 < \alpha < 1$ is the appropriate distribution for precipitation amounts since this gives relatively high probability to small rainfall amounts and increasingly less probability to larger amounts.

Maximum likelihood (11) estimates are not available when α is less than one and are quite unstable when α is between one and 2.5. Method of moments (11) estimators are even less precise than maximum likelihood and especially so for values of α less than, say, 40. An approximate maximum likelihood parameter estimation procedure suggested by Greenwood and Durand (6) was chosen. The error of this procedure for $\alpha < 1$ is stated by Johnson and Kotz [(7), Vol. 1, p. 189] to not exceed 0.0054%. Using the Greenwood and Durand method, define

$$Y = \log_e\left(\frac{\text{arithmetic mean}}{\text{geometric mean}}\right) = \log_e\left(\frac{\sum_{i=1}^{n} x_i/n}{(\prod_{i=1}^{n} x_i)^{1/n}}\right)$$

then,

$$\hat{\alpha} \simeq \frac{(8.898919 + 9.059950Y + 0.9775373Y^2)}{Y(17.79728 + 11.968477Y + Y^2)}$$

$$\hat{\beta} \simeq \frac{\sum_{i=1}^{n} x_i}{n\hat{\alpha}1}$$

where x_i = precipitation amount on day i

n = number of days in the month

This formula for $\hat{\alpha}$ is appropriate only for $0 < \alpha < 1$. Parameter estimates were made for each month conditioned on the previous day precipitation status. This conditioning is probably preferrable unless the number of observations available for parameter estimation is reduced to the point that precision has to be sacrificed.

Precipitation amounts were simulated by obtaining gamma random variates using the method of Johnk (3). This method uses a rather complicated combination of standard uniform random variates to obtain a random variate appearing to come from a gamma distribution with the desired parameters. Simulated precipitation amounts were rounded to the nearest 0.01 inch. Amounts which were simulated to be smaller than 0.005 inch were not rounded to zero but rather were discarded and another random amount simulated. This procedure was used because zero amounts (i.e., dry) and trace amounts (arbitrarily set equal to 0.001 inch) were previously determined.

Temperature

Two bivariate normal distributions were used to simulate daily maximum and minimum temperature differences conditioned on the current day precipitation status. The temperature differences were obtained by subtracting the observed daily temperature from two fitted three-parameter sine curves representing the mean daily maximum and minimum temperatures. The sine functions are of the form

$$T = \mathrm{SIN}((\mathrm{JDATE} - A) * 0.017214) * B + C$$

where T = daily mean maximum or minimum temperature

JDATE = julian date.

The three parameters A, B, and C were estimated by the non-linear least squares Marquardt method as contained in the Statistical Analysis System computer package (2). The parameter A controls the shift in the horizontal time axis, B establishes the amplitude of the sine curve, and C controls the shift in the vertical temperature axis. Fits using this three-parameter sine function were very good at all locations (R^2 values in excess of 0.93).

Daily temperatures were generated using one bivariate normal to simulate either current maximum or current minimum temperature from previous day maximum temperature. Which current temperature was simulated depended on the higher of the two correlations (i.e., the serial correlation between previous and current maximum temperature and the lag cross-correlation between previous maximum and current minimum temperature). The second bivariate normal was used to simulate the remaining current temperature from the current temperature generated by the first bivariate normal. This procedure takes advantage of the highest correlations. It is noted that in Columbia, Mo., the correlation between previous maximum and current minimum was almost always larger on dry days and sometimes larger on wet days than the correlation between previous maximum and current maximum. Three means, three variances, and three correlations were estimated for each month and precipitation status. These parameters were estimated from the temperature *differences* using the usual formulae. After parameter estimates were made, daily temperature differences were stimulated using the following general equation:

$$T_2 = \hat{\mu}_2 + \hat{\rho}_{12}\hat{\sigma}_2(T_1 - \hat{\mu}_1)/\hat{\sigma}_1 + \hat{\sigma}_2(1 - \hat{\rho}^2_{12})^{1/2}Z$$

where T_1 = difference for either previous day maximum temperature or current temperature
T_2 = current temperature
Z = standard normal random variate.

The simulated temperatures were obtained by adding the appropriate daily values from the fitted sine functions to the values obtained for T_2.

The temperature simulation methodology just described assumes that temperature *difference* parameter estimates are homogeneous within month and precipitation status. This assumption is thought to be much more conservative than the assumption that parameter estimates based on *actual* temperatures are homogeneous within month and precipitation status. The latter assumption is clearly subject to criticism during the spring and fall months when seasonal weather changes are relatively fast.

Solar Radiation

Daily solar radiation differences were simulated from a gamma distribution on dry days and a beta distribution on wet days. The solar radiation differences were obtained by subtracting the observed solar radiation value from the maximum clear day radiation. The latter values were computed from a series of equations which depend only on the latitude and julian date. The equations were obtained from unpublished material with permission from J. T. Ritchie (soil scientist; USDA-ARS; Grassland, Soil and Water Research Laboratory; Temple, Tex.).

Before parameters were estimated for the gamma distribution, a transformation was made to the dry day solar radiation differences. The actual dry day solar radiation values are negatively skewed and, hence, the differences are positively skewed. This fits the general shape of a gamma distribution with $\alpha \geq 1$. Referring back to the general three-parameter gamma distribution discussed in the precipitation amount section, recall that the third parameter, γ, establishes the lower bound. While γ could realistically be assumed to be zero for precipitation, this is generally not a good assumption for solar radiation. Although maximum likelihood estimators exist for all three gamma parameters, the estimates are unstable when α is less than 2.5. Aside from maximum likelihood estimation, a good first approximation for γ is a number slightly less than the observed minimum [(7), Vol. 1, p. 187]. Rather than explicitly estimate γ, the solar radiation differences were transformed by the following equation.

$$TSRD_G = SRD - MINSRD + 3$$

where SRD = solar radiation difference
MINSRD = minimum solar radiation difference within month and precipitation status
$TSRD_G$ = transformed solar radiation difference

Since the transformed solar radiation difference values start at a minimum of three for all months, γ can be assumed to be zero. The addition of three arose from programming considerations to avoid the possibility of roundoff created zero values. The α and β parameters were again estimated using the Greenwood and Durand method. This choice was made because of the limitation in the maximum likelihood estimator for α values less than 2.5. For Columbia, Mo., $\hat{\alpha}$ generally ranged between 2 and 4. The formula for estimating α by the Greenwood and Durand method follows.

$$\hat{\alpha} \simeq \frac{(0.5000876 + 0.1648852Y - 0.0544274Y^2)}{Y}$$

$$\text{where } Y = \log_e\left(\frac{\text{arithmetic mean}}{\text{geometric mean}}\right)$$

This formula for $\hat{\alpha}$ is appropriate for $\alpha \geq 1$. Johnson and Kotz [(7), Vol. 1, p. 189] state that the error of this approximation does not exceed 0.0088%. The β parameter was estimated as before.

The beta distribution was hypothesized for wet day solar radiation differences. The standard form of the beta distribution has two parameters, p and q, and requires the random variable X to be in the interval zero to one, inclusive.

To get the solar radiation differences on the interval [0, 1], the following transformation was made:

$$TSRD_B = \frac{SRD - MINSRD}{MAXSRD - MINSRD}$$

where SRD = solar radiation difference
MINSRD = minimum SRD within month and precipitation status
MAXSRD = maximum SRD within month and precipitation status
$TSRD_B$ = transformed SRD.

Formulas for estimating p and q were obtained using the method of moments (11) and are as follows:

$$\hat{p} = \frac{w - v(1 + w)^2}{v(1 + w)^3}$$

$$\hat{q} = \hat{p}w$$

where v = sample variance
$w = (1 - \bar{x})/\bar{x}$
$\bar{x}$ = sample mean.

After all parameter estimates were made, the appropriate transformed solar radiation differences were simulated by month and precipitation status. In the case of dry days, gamma random variates were simulated using the same procedure as for precipitation. Each random variate was then transformed back to the original scale by adding MINSRD, subtracting 3, and adding the maximum clear day radiation. In the case of wet days, beta random variates were simulated by taking advantage of the following relationship:

$$\beta(p,q) = \frac{\Gamma(p,l)}{\Gamma(p,l) + \Gamma(q,l)}.$$

Thus, a beta random variate was obtained from a combination of two gamma random variates. Daily solar radiation values were then computed in the original scale by reversing the transformations previously indicated.

RESULTS AND DISCUSSION

Extensive model testing was done at Columbia, Mo., because of the availability of a long historic record for precipitation and temperature. Simulated data were compared to the 17-year historic base period and to the entire 80-year length of record. The former tests indicate whether model assumptions are valid and the latter tests show whether the base period is of sufficient length to adequately represent the entire data set.

correlation between previous maximum and current minimum was almost always larger on dry days and sometimes larger on wet days than the correlation between previous maximum and current maximum. Three means, three variances, and three correlations were estimated for each month and precipitation status. These parameters were estimated from the temperature *differences* using the usual formulae. After parameter estimates were made, daily temperature differences were stimulated using the following general equation:

$$T_2 = \hat{\mu}_2 + \hat{\rho}_{12}\hat{\sigma}_2(T_1 - \hat{\mu}_1)/\hat{\sigma}_1 + \hat{\sigma}_2(1 - \hat{\rho}^2_{12})^{1/2}Z$$

where T_1 = difference for either previous day maximum temperature or current temperature
T_2 = current temperature
Z = standard normal random variate.

The simulated temperatures were obtained by adding the appropriate daily values from the fitted sine functions to the values obtained for T_2.

The temperature simulation methodology just described assumes that temperature *difference* parameter estimates are homogeneous within month and precipitation status. This assumption is thought to be much more conservative than the assumption that parameter estimates based on *actual* temperatures are homogeneous within month and precipitation status. The latter assumption is clearly subject to criticism during the spring and fall months when seasonal weather changes are relatively fast.

Solar Radiation

Daily solar radiation differences were simulated from a gamma distribution on dry days and a beta distribution on wet days. The solar radiation differences were obtained by subtracting the observed solar radiation value from the maximum clear day radiation. The latter values were computed from a series of equations which depend only on the latitude and julian date. The equations were obtained from unpublished material with permission from J. T. Ritchie (soil scientist; USDA-ARS; Grassland, Soil and Water Research Laboratory; Temple, Tex.).

Before parameters were estimated for the gamma distribution, a transformation was made to the dry day solar radiation differences. The actual dry day solar radiation values are negatively skewed and, hence, the differences are positively skewed. This fits the general shape of a gamma distribution with $\alpha \geq 1$. Referring back to the general three-parameter gamma distribution discussed in the precipitation amount section, recall that the third parameter, γ, establishes the lower bound. While γ could realistically be assumed to be zero for precipitation, this is generally not a good assumption for solar radiation. Although maximum likelihood estimators exist for all three gamma parameters, the estimates are unstable when α is less than 2.5. Aside from maximum likelihood estimation, a good first approximation for γ is a number slightly less than the observed minimum [(7), Vol. 1, p. 187]. Rather than explicitly estimate γ, the solar radiation differences were transformed by the following equation.

$$TSRD_G = SRD - MINSRD + 3$$

where SRD = solar radiation difference
MINSRD = minimum solar radiation difference within month and precipitation status
$TSRD_G$ = transformed solar radiation difference

Since the transformed solar radiation difference values start at a minimum of three for all months, γ can be assumed to be zero. The addition of three arose from programming considerations to avoid the possibility of roundoff created zero values. The α and β parameters were again estimated using the Greenwood and Durand method. This choice was made because of the limitation in the maximum likelihood estimator for α values less than 2.5. For Columbia, Mo., $\hat{\alpha}$ generally ranged between 2 and 4. The formula for estimating α by the Greenwood and Durand method follows.

$$\hat{\alpha} = \frac{(0.5000876 + 0.1648852Y - 0.0544274Y^2)}{Y}$$

$$\text{where } Y = \log_e\left(\frac{\text{arithmetic mean}}{\text{geometric mean}}\right)$$

This formula for $\hat{\alpha}$ is appropriate for $\alpha \geq 1$. Johnson and Kotz [(7), Vol. 1, p. 189] state that the error of this approximation does not exceed 0.0088%. The β parameter was estimated as before.

The beta distribution was hypothesized for wet day solar radiation differences. The standard form of the beta distribution has two parameters, p and q, and requires the random variable X to be in the interval zero to one, inclusive.

To get the solar radiation differences on the interval [0, 1], the following transformation was made:

$$TSRD_B = \frac{SRD - MINSRD}{MAXSRD - MINSRD}$$

where SRD = solar radiation difference
MINSRD = minimum SRD within month and precipitation status
MAXSRD = maximum SRD within month and precipitation status
$TSRD_B$ = transformed SRD.

Formulas for estimating p and q were obtained using the method of moments (11) and are as follows:

$$\hat{p} = \frac{w - v(1 + w)^2}{v(1 + w)^3}$$

$$\hat{q} = \hat{p}w$$

where v = sample variance
$w = (1 - \bar{x})/\bar{x}$
$\bar{x}$ = sample mean.

After all parameter estimates were made, the appropriate transformed solar radiation differences were simulated by month and precipitation status. In the case of dry days, gamma random variates were simulated using the same procedure as for precipitation. Each random variate was then transformed back to the original scale by adding MINSRD, subtracting 3, and adding the maximum clear day radiation. In the case of wet days, beta random variates were simulated by taking advantage of the following relationship:

$$\beta(p,q) = \frac{\Gamma(p,l)}{\Gamma(p,l) + \Gamma(q,l)}.$$

Thus, a beta random variate was obtained from a combination of two gamma random variates. Daily solar radiation values were then computed in the original scale by reversing the transformations previously indicated.

RESULTS AND DISCUSSION

Extensive model testing was done at Columbia, Mo., because of the availability of a long historic record for precipitation and temperature. Simulated data were compared to the 17-year historic base period and to the entire 80-year length of record. The former tests indicate whether model assumptions are valid and the latter tests show whether the base period is of sufficient length to adequately represent the entire data set.

Table 1. Comparison of average historic and simulated climatic values by location and month.

Location	Month	Precipitation occurrence		Precipitation amount		Maximum temperature		Minimum temperature		Solar radiation	
		Hist	Sim	Hist	Sim	Hist	Sim	Hist	Sim	Hist	Sim
				cm		C				ly/day	
Columbia (17-year period)	1	0.488	0.493	3.70	3.79	3.42	3.35	−6.82	−6.71	184.8	193.1
	2	0.442	0.425	4.06	4.1	6.11	6.45	−4.42	−4.15	266.	273.6
	3	0.522	0.503	6.3	6.17	11.13	11.38	−0.18	−0.01	350.	355.7
	4	0.535	0.556	8.9	9.9	19.16	18.84	7.17	7.08	435.	433.
	5	0.488	0.504	10.9	11.6	24.21	24.00	12.42	12.27	530.	516.5
	6	0.445	0.452	10.2	10.3	28.90	28.89	17.34	17.51	571.	572.1
	7	0.431	0.421	10.6	9.3	31.56	31.56	19.84	19.91	580.	582.8
	8	0.349	0.343	5.9	6.8	31.04	31.11	18.71	18.81	530.	527.6
	9	0.386	0.375	9.6	9.8	27.11	27.03	14.25	14.39	427.	432.0
	10	0.338	0.352	8.4	8.9	20.97	20.95	8.42	8.58	317.	316.3
	11	0.363	0.357	3.6	3.67	12.58	12.22	1.40	0.99	212.6	218.8
	12	0.440	0.454	4.3	5.11	5.49	5.31	−4.05	−3.99	162.9	164.1
	Annual	0.436	0.436	86.4	89.7	18.53	18.48	7.06	7.11	380.8	382.6
Columbia (80-year period)	1	0.431	0.493**	4.70	3.85	4.03	3.21**	−6.15	−6.82**	184.8	191.1
	2	0.438	0.446	4.51	4.32	5.74	6.09	−4.68	−4.35	266.	269.8
	3	0.435	0.533**	7.27	6.67	11.84	11.41*	0.45	0.05*	350.	353.4
	4	0.505	0.551**	9.4	9.46	18.67	18.72	6.89	6.95	435.	436.3
	5	0.515	0.499	11.9	11.5	23.94	24.15	12.19	12.35	530.	522.0
	6	0.467	0.456	12.0	10.66	28.82	29.05*	17.37	17.57*	571.	573.1
	7	0.382	0.407	8.9	9.61	31.55	31.77*	19.49	19.98**	581.	581.9
	8	0.385	0.355*	9.1	6.45**	30.89	31.16*	18.61	18.86**	528.	527.7
	9	0.380	0.406	11.0	10.5	26.84	27.08	14.39	14.53	431.	426.0
	10	0.327	0.350	7.5	8.9	20.78	21.12*	8.12	8.66**	322.	316.6
	11	0.363	0.366	5.46	3.82**	12.34	12.29	1.20	1.07	213.8	215.8
	12	0.400	0.457**	4.65	4.87	5.55	5.29	−3.99	−4.04	162.0	162.8
	Annual	0.423	0.443**	96.3	90.7*	18.47	18.50	7.04	7.12	379.9	381.9
Albuquerque	1	0.205	0.195	0.73	0.85	8.56	8.75	−4.71	−4.35	299.5	299.8
	2	0.274	0.283	0.95	0.95	11.18	11.08	−2.72	−2.73	386.5	384.1
	3	0.263	0.284	1.21	1.36	15.13	14.80	0.12	0.06	504.0	504.2
	4	0.197	0.198	1.01	0.91	21.21	20.91	5.07	4.88	629.0	629.3
	5	0.268	0.279	0.97	1.05	26.75	26.58	10.27	10.21	696.4	691.8
	6	0.280	0.286	1.35	1.22	32.21	32.37	15.48	15.48	738.2	732.9
	7	0.576	0.604	3.61	3.82	33.57	33.67	18.56	18.58	681.0	677.2
	8	0.561	0.523	3.6	3.29	32.02	32.20	17.56	17.66	628.5	628.0
	9	0.310	0.313	1.87	1.91	28.70	28.58	13.64	13.57	549.9	551.8
	10	0.219	0.241	2.1	2.65	22.08	22.06	6.90	7.01	442.0	437.4
	11	0.200	0.201	0.82	0.77	13.74	13.73	−0.07	0.01	326.2	326.1
	12	0.218	0.210	1.33	1.25	8.41	8.44	−4.17	−4.19	275.4	274.1
	Annual	0.298	0.302	19.6	20.1	21.17	21.14	6.37	6.39	512.3	511.8
Caribou	1	0.747	0.709	5.43	5.12	−6.47	−7.15*	−16.40	−16.84	143.7	147.3
	2	0.706	0.687	5.61	5.82	−4.62	−4.97	−15.77	−16.27	244.	243.1
	3	0.616	0.596	5.5	5.79	0.61	0.50	−9.49	−9.89	359.	364.7
	4	0.588	0.589	6.4	6.53	7.72	7.95	−1.90	−1.78	419.	433.3
	5	0.603	0.575	7.0	7.03	15.47	15.89	3.88	3.99	463.	474.
	6	0.602	0.609	7.9	8.13	21.40	20.94	9.49	9.27	500.	495.
	7	0.622	0.630	10.0	10.3	24.07	23.92	12.31	12.20	505.	500.4
	8	0.572	0.585	10.0	11.2	22.47	22.13	10.70	10.51	435.	436.6
	9	0.557	0.561	9.2	9.7	17.98	17.82	6.40	6.31	325.	325.4
	10	0.556	0.562	8.3	8.1	11.30	11.53	1.63	1.78	203.6	210.6
	11	0.718	0.733	8.7	9.35	3.59	3.66	−3.94	−3.89	113.0	118.3
	12	0.772	0.771	7.1	6.97	−4.29	−4.45	−12.98	−13.06	109.1	111.2
	Annual	0.638	0.634	91.2	94.0	9.17	9.04	−1.27	−1.41	321.3	322.0
Medford	1	0.732	0.744	9.8	10.5	7.16	6.94	−0.79	−0.79	118.3	122.1
	2	0.563	0.573	5.7	5.95	11.53	11.26	−0.13	−0.16	216.4	219.9
	3	0.552	0.530	4.3	4.15	14.10	14.42	0.83	0.72	349.	351.2
	4	0.440	0.435	2.33	2.29	18.10	17.57*	2.89	2.82	484.	478.9
	5	0.400	0.377	3.3	3.17	22.42	23.25**	6.16	6.29	594.	593.1
	6	0.272	0.262	1.77	1.72	27.12	27.06	9.76	9.66	648.	652.0
	7	0.098	0.098	0.70	0.57	32.51	32.47	12.32	12.23	696.6	698.8
	8	0.137	0.169	0.93	1.13	31.52	31.04*	11.87	11.82	600.7	595.0
	9	0.183	0.205	1.54	1.94	28.28	27.96	8.60	8.64	463.2	459.9
	10	0.374	0.374	4.5	4.55	20.35	20.27	4.28	4.18	294.3	295.1
	11	0.593	0.609	7.6	8.05	11.79	12.00	1.38	1.46	148.4	146.4
	12	0.713	0.715	9.9	9.8	6.95	6.80	−0.38	−0.38	96.3	96.4
	Annual	0.421	0.424	52.3	53.8	19.36	19.29	4.76	4.73	391.1	393.2
Miami	1	0.395	0.404	5.8	6.0	23.81	23.68	14.65	14.64	332.1	333.1
	2	0.382	0.397	5.6	5.92	24.56	24.67	15.36	15.46	410.1	409.8
	3	0.340	0.355	5.2	5.1	26.18	26.18	17.35	17.32	480.	485.0
	4	0.358	0.339	6.7	7.2	28.06	28.06	19.90	19.85	547.	551.3
	5	0.500	0.513	16.5	18.3	29.54	29.31**	21.85	21.86	548.	541.2
	6	0.682	0.673	25.0	25.3	30.85	30.73	23.41	23.57*	515.	513.6
	7	0.694	0.688	15.8	16.6	31.65	31.64	24.37	24.40	546.	548.6
	8	0.713	0.721	16.8	17.6	32.19	32.02**	24.53	24.51	516.	509.0
	9	0.750	0.739	21.9	23.3	31.18	31.18	24.09	24.07	443.	438.4
	10	0.634	0.660	21.3	22.0	29.10	29.12	21.79	21.87	393.1	387.8
	11	0.443	0.420	6.7	7.8	26.51	26.47	18.16	18.18	353.7	352.5
	12	0.350	0.340	4.3	4.60	24.42	24.17	15.37	15.03	323.4	322.2
	Annual	0.521	0.521	152.	159.8	28.19	28.12	20.09	20.08	452.8	449.5

*,** Significant at the 0.05 and 0.01 levels, respectively.

Model validation at Columbia consisted of several types of tests. T-tests were used to compare the means, and F-tests were used to compare the variances of mean daily precipitation, maximum temperature, minimum temperature, and solar radiation for each month and precipitation status (wet or dry). The ranges of these weather variables were also examined. Chi-square tests were used to compare the frequency of wet days for each month. Two-sample Kolmogorov-Smirnov (K-S) tests were used to compare cumulative distribution functions (CDF's) by month and precipitation status. The means, standard deviations, and ranges of wet spells, dry spells, freezing spells, and hot spells (35 C or above) were computed for each month for the 80 years of historic data and 99 years of simulated data. Finally, the frequency distributions of the wet, dry, freezing, and hot spells as well as the CDF's which were declared significantly different by the K-S tests were graphed for the historic and simulated data.

At the other four locations, comparisons were made using 50 years of simulated data and the entire 20-year historic period. Tests were made for frequency of wet days and means, variances, and CDF's for each climatic variable by month and precipitation status. Ranges were also examined.

Table 1 shows a comparison between the historic and simulated data at each location for frequency of wet days, mean total precipitation, maximum and minimum temperature, and solar radiation by month and annually. Because of space limitations, other results are not presented here but may be obtained from the authors (10).

The validation showed that the method used to simulate precipitation occurrence worked very well for the base period at all locations. No significant differences were found ($\alpha = 0.05$). At Columbia where 80 years of precipitation and temperature data were available, the simulated precipitation frequencies (based on the most recent 17-year period) did not compare well to the long term averages for 5 of the 12 months.

The simulated precipitation amounts compared favorably to the base period for all locations. No significant differences were found in the means and very few were found in the distributions ($\alpha = 0.05$). A majority of the variances were significantly different at four of the five locations ($\alpha = 0.05$). The annual totals showed a slight positive bias and the associated variances a slight negative bias. Comparison of the simulated data to the entire 80-year period of record at Columbia revealed a negative bias in the average monthly rainfall amounts and variances. The simulated total annual rainfall was significantly lower than the historic.

The algorithm used to generate daily maximum and minimum temperatures worked very well. Few significant differences were found in the means, variances, and distributions ($\alpha = 0.05$). Comparisons to the base period showed a tendency for the monthly means to be low on wet days. The average annual maximum temperatures showed a slight negative bias. The simulated data compared satisfactorily to the 80-year period at Columbia although there was some indication that the 17-year base was slightly warmer. The variances for the average annual maximum temperature were significantly different.

The method used to simulate solar radiation worked extremely well. Very few significant differences were found in the means and distributions ($\alpha = 0.05$). Variances were significantly different less than 25% of the time ($\alpha = 0.05$). No biases were evident in any of the comparisons. There was not a long enough period of record available at Columbia to determine whether the 17-year base period was of sufficient length.

Overall, the validation showed that the simulated values for all climatic variables were very similar to the actual data used for parameter estimation at all locations tested. However, the analysis also indicated that 17 years were not enough to adequately represent the 80 years of recorded precipitation data at Columbia. If adequate representation is desired, a longer base period for parameter estimation would be required. However, many times it is better for the simulated data to represent recent history rather than a long time period. This is particularly true if the simulated data are used to assess the risk of current decisions or represent future weather. It may also be possible that the simulator may function satisfactorily with only a 10 to 15-year base period for temperature and solar radiation. However, too short of a base would cause the simulated data to have unrealistically low variance. The choice of the length of the base period depends in part on the purpose of the simulation.

ACKNOWLEDGMENTS

The authors wish to recognize the earlier work by D. C. Bond, USDA-SRS, which formed a basis for the present model. Thanks are also expressed to P. T. Dyke, USDA-ERS, for providing the climatic data for the four additional validation locations.

LITERATURE CITED

1. Arkin, G. F., S. J. Maas, and C. W. Richardson. 1980. Forecasting grain sorghum yields using simulated weather data and updating techniques. Trans. ASAE 23(3):676–680.
2. Barr, A. J., J. H. Goodnight, and J. P. Sall. 1979. SAS users guide 1979 edition. SAS Institute Inc., SAS Circle, Box 8000, Cary, NC 27511.
3. Berman, M. B. 1971. Generating gamma distributed variates for computer simulation models. Rand Corp. Pub. R-641-PR, Santa Monica, CA.
4. Bond, D. C. 1979. Generating daily weather values by computer simulation techniques for crop yield forecasting models. Research division, SRS, USDA, Washington, DC.
5. Bruhn, J. A., W. E. Fry, and G. W. Fick. 1980. Simulation of daily weather data using probability distributions. J. Appl. Meteorol.
6. Greenwood, J. A., and D. Durand. 1960. Aids for fitting the gamma distribution by maximum likelihood. Technometrics 2:55–65.
7. Johnson, N. L., and S. Kotz. 1970. Distributions in statistics: continuous univariate distributions. Vol. 1, 2. John Wiley and Sons, N.Y.
8. Jones, J. W., R. F. Colwick, and F. D. Threadgill. 1972. A simulated environmental model of temperature, evaporation, rainfall and soil moisture. Trans ASAE 15:366–372.
9. Larsen, G. A. 1981. Progress report on the evaluation of plant growth models. SRS Staff Report AGESS810716. Research Division, SRS, USDA, Washington, DC.
10. ———, and R. B. Pense. 1981. Stochastic simulation of daily climate data. SRS Staff Report AGESS810831. Research Division, SRS, USDA, Washington, DC.
11. Larson, H. J. 1969. Introduction to probability theory and statistical inference. John Wiley and Sons, NY.
12. Nicks, A. D., and J. F. Harp. 1980. Stochastic generation of temperature and solar radiation data. J. Hydrol. 48:1–17.
13. Richardson, C. W. 1981. Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour. Res. 17(1):182–190.