Simulation of Daily Weather Data Using Theoretical Probability Distributions

J. A. BRUHN AND W. E. FRY

Department of Plant Pathology, Cornell University, Ithaca, NY 13853

G. W. FICK

Department of Agronomy, Cornell University, Ithaca, NY 13853

(Manuscript received 12 November 1979, in final form 4 February 1980)

ABSTRACT

A computer simulation model was constructed to supply daily weather data to a plant disease management model for potato late blight. In the weather model Monte Carlo techniques were employed to generate daily values of precipitation, maximum temperature, minimum temperature, minimum relative humidity and total solar radiation. Each weather variable is described by a known theoretical probability distribution but the values of the parameters describing each distribution are dependent on the occurrence of rainfall. Precipitation occurrence is described by a first-order Markov chain. The amount of rain, given that rain has occurred, is described by a gamma probability distribution. Maximum and minimum temperature are simulated with a trivariate normal probability distribution involving maximum temperature on the previous day, maximum temperature on the current day and minimum temperature on the current day. Parameter values for this distribution are dependent on the occurrence of rain on the previous day. Both minimum relative humidity and total solar radiation are assumed to be normally distributed. The values of the parameters describing the distribution of minimum relative humidity is dependent on rainfall occurrence on the previous day and current day. Parameter values for total solar radiation are dependent on the occurrence of rain on the current day. The assumptions made during model construction were found to be appropriate for actual weather data from Geneva, New York. The performance of the weather model was evaluated by comparing the cumulative frequency distributions of simulated weather data with the distributions of actual weather data from Geneva, New York and Fort Collins, Colorado. For each location, simulated weather data were similar to actual weather data in terms of mean response, variability and autocorrelation. The possible applications of this model when used with models of other components of the agro-ecosystem are discussed.

1. Introduction

Agricultural production requires the management of a complex of interacting physical, biological and economic systems. Uncertainty associated with many of these systems contributes to the complexity of the decision making process and severely reduces the efficiency of agricultural management. Because the need for and effectiveness of management activities depend upon the state of these stochastic components of the agro-ecosystem, effective management policy must achieve a balance between cost effectiveness and risk reduction.

Fluctuations in weather represent a significant portion of the total uncertainty within an agroecosystem. The inclusion of a stochastic weather component with models of other components of an agricultural system should provide a method for measuring the risk due to weather uncertainty that is associated with management alternatives.

The purpose of this report is to describe and evaluate a stochastic weather model that was constructed to provide daily weather data for a plant disease management model of potato late blight (*Solanum tuberosum* and *Phytophthora infestans*). Statistical tests were performed to evaluate the appropriateness of assumptions made during the construction of the weather model and to provide an indication of the performance of the model.

2. Model description

Monte Carlo simulation is used in the model to generate daily values of precipitation, maximum temperature, minimum temperature, minimum relative humidity and total solar radiation. Monte Carlo simulation, which has been used to model weather variables (Wiser, 1966; Jones *et al.*, 1972), employs stochastic and relational information about the components of a system to produce a stochastic representation of the total system's performance. The model developed by Jones *et al.* (1972) generated daily values of precipitation, average temperature and pan evaporation and provided the basis

0021-8952/80/091029-08$06.00

for the construction of our model. Our model differs from that of Jones *et al.* by providing daily values of maximum and minimum temperatures, minimum relative humidity and total solar radiation and by including autocorrelation in the temperature component.

Two processes are common to the generation of values for each weather variable. The first process is the selection of a set of parameters to describe the appropriate probability distribution. The occurrence of rainfall is the foundation for parameter selection (Jones *et al.*, 1972). The second process represents the stochastic framework of the model and is the generation of a random value from the selected probability distribution. This value is assigned as the current day's value of the appropriate weather variable.

Daily rainfall includes the occurrence of rain and the amount of rain. The occurrence of rainfall is described by a first-order Markov chain which means that the probability of rain occurring on given day (parameter selection) is dependent only on whether rain did or did not occur on the previous day. After this probability has been determined, a random number generated from a uniform probability distribution is used to determine if rain occurs on the current day. If the number exceeds the probability of precipitation, daily rainfall is zero, otherwise the amount of rainfall is determined by a random number generated from a gamma probability distribution.

The temperature component is described by a trivariate normal probability distribution. Maximum temperature is correlated with the maximum temperature on the previous day and minimum temperature is correlated with the current day's maximum temperature. The selection of the appropriate values of the parameter set describing this distribution is dependent on the occurrence of rain on the preceding day.

In the model both minimum relative humidity and total solar radiation are normally distributed. The selection of parameter values for total solar radiation is dependent on the occurrence of rainfall on the current day while, for minimum relative humidity, the occurrence of rainfall on both the present day and the previous day determine the appropriate values of the parameter set.

To accommodate changes in the weather as the season progresses, a complete collection of all parameter sets was estimated for each month of the growing season (May–September). The relationships between the weather variables comprising the weather model are summarized in the following functional representation:

rain occurrence $= f(M, R_{t-1}, \mathrm{RU})$,

rain amount $= f(M, \mathrm{RG})$,

maximum temperature $= f(M, R_{t-1}, \mathrm{TMAX}_{t-1}, \mathrm{RN})$,

minimum temperature $= f(M, R_{t-1}, \mathrm{TMAX}_t, \mathrm{RN})$,

minimum relative humidity $= f(M, R_{t-1}, R_t, \mathrm{RN})$,

total solar radiation $= f(M, R_t, \mathrm{RN})$,

where

M	month being simulated,
R_{t-1}	rainfall occurrence on the previous day,
R_t	rainfall occurrence on the current day,
TMAX_{t-1}	maximum temperature on the previous day,
TMAX_t	maximum temperature on the current day,
RU	random variable with a uniform probability distribution,
RN	random variable with a normal probability distribution,
RG	random variable with a gamma probability distribution.

3. Parameter estimation

A large data base was used for parameter estimation. Daily weather data from the period 1946–75 were used to estimate parameters for all weather variables except minimum relative humidity and total solar radiation for Geneva, New York. Parameter estimates for these two weather variables were computed from daily weather data from 1963–75. Because of the large data base available for parameter estimation, maximum likelihood estimators of all population parameters were used to exploit the asymptotic efficiency of these estimators. A more complete discussion of the parameter estimators used in the weather model can be found in Bruhn *et al.* (1979).

4. Model construction

Before model performance can be assessed, the structural and stochastic assumptions associated with each dynamic component of the model require verification. We used information from previous studies, historical data and programming efficiency as criteria for verification. Except where indicated, each analysis of historical data was based on 10 years (1966–76) of daily weather data from Geneva, New York. For brevity, analysis was restricted to data from May and August, the extreme months of the intended simulation period. All historical data used for model verification represented a subset of the data used for parameter estimation.

The assumptions associated with the precipitation occurrence component of the model required extensive verification because of the importance of precipitation occurrence to model structure and because the appropriate order of the Markov chain describing precipitation occurrence varies with geo-

TABLE 1. Determining the proper order of the Markov chain describing the occurrence of rainfall. Results from applying the Akaike Information Criterion (AIC) procedure to rainfall data from Geneva, New York.

Order of the Markov model	May	June	July	August	September	Season
Zero	30.22*	6.77	3.97	−1.58	20.25	40.74
First	−10.85	−6.30	−9.87	−10.80	−6.57	−5.60
Second	−7.51	−5.39	−7.07	−12.33	−4.09	−2.12
Third	−0.98	−3.97	−4.47	−8.34	−2.16	0.08
Fourth	0.00	0.00	0.00	0.00	0.00	0.00
Sample size	972	936	972	972	935	1490
Proper model	first	first	first	second	first	first

* Value of Akaike Information Criterion (AIC) statistic. The minimum value of the AIC statistics identifies the appropriate model.

graphic location and season (Chin, 1977). Chi-squared tests have traditionally been used to determine the appropriate Markov model describing rainfall occurrence (Gabriel and Neumann, 1962; Jones *et al.*, 1972; Weiss, 1964; Wiser, 1966). Recent work (Gates and Tong, 1976) suggests that a new statistic, the Akaike Information Criterion (AIC), may provide a more reliable indication of the appropriate order of a stationary ergodic Markov chain (Tong, 1975). We used the AIC procedure to analyze precipitation data from Geneva.

The AIC procedure was developed by Akaike (1972, 1974) to overcome inadequacies associated with the application of traditional hypothesis testing procedures to the problem of statistical model identification. Akaike's procedure is an extension of the maximum likelihood principle and considers model identification as an estimation procedure. The advantage of this approach is that model identification is not dependent on the subjectively determined significance levels required by hypothesis testing procedures. Instead of significance levels, the AIC statistic provides the basis for model identification. The magnitude of the AIC statistic is a function of how well the model describes the data and the number of parameters required to describe the model. For selecting an appropriate model from several competing models, the minimum AIC estimate identifies the appropriate model.

We used the AIC procedure of Tong (1975) to analyze rainfall data. Because a large sample size is required (Chin, 1977), 45 years of daily rainfall data were analyzed for each month of the intended period of simulation (May–September). Ten seasons of rainfall data were also analyzed collectively. A first-order Markov chain was appropriate for describing the occurrence of precipitation for all months except August (Table 1), for which a second-order Markov chain was appropriate. A first-order model was appropriate for describing rainfall occurrence when the months of May–September were considered collectively. Therefore, we chose the first-order Markov chain as the appropriate model for describing the occurrence of rainfall for Geneva during the months of May–September.

The amount of precipitation was assumed to be independent of the occurrence of rain on the previous day. We verified this assumption by partitioning observed rainfall amounts into two samples according to rainfall occurrence on the previous day and testing for equality of the two sample distributions with a Kolmogorov-Smirnov two-sample test. The distributions of the two samples were not significantly different ($P = 0.20$) for May or August. The values of the test statistic were 0.083 and 0.107, respectively. From these results we concluded that amounts of daily rainfall were independent of the occurrence of rain on the previous day.

We used a gamma probability distribution to describe the daily amount of rainfall. This distribution has been used similarly by other workers (Barger and Thom, 1949; Dethier and McGuire, 1961; Jones *et al.*, 1972; Raudkivi and Lawgun, 1972). We used the Smirnov test for maximum deviations to determine if a gamma probability distribution could describe the distribution of 10 years of daily rainfall amounts from Geneva. The observed deviations (Schneider and Clickner, 1970) of the sample distribution from the theoretical gamma distribution were not statistically significant ($P = 0.10$) for rainfall amounts observed in May or August. The maximum deviations were 0.068 and 0.040, respectively.

Temperature was assumed to be dependent on the occurrence of rainfall on the previous day because Jones *et al.* (1972) have reported this relationship for average temperature in Mississippi. To verify this, the cumulative frequency distribution of daily temperature on days preceded by rain was compared with the distribution of daily temperature values observed on days not preceded by rain. Maximum and minimum temperature data were analyzed separately. The values of the Kolmogorov-Smirnov statistic indicated significant differences ($P = 0.01$) between the cumulative frequency distributions of

TABLE 2. Testing for bivariate normality of daily maximum and minimum temperature data from Geneva, New York, using Smirnov's maximum deviation test to determine the normality of linear combinations of maximum and minimum temperature.

Linear combinations of maximum and minimum temperature	May		August	
	Wet†	Dry	Wet	Dry
$X + Y$**	0.02‡	0.02	0.04	0.03
$X + 0 \cdot Y$	0.07**	0.02	0.03	0.05
$0 \cdot X + Y$	0.02	0.04	0.04	0.05

* X = maximum temperature, Y = minimum temperature.
** $P = 0.10$.
† Temperature values observed on days preceded by a wet day (precipitation = 0.25 mm/day).
‡ Maximum deviation between the sample cumulative frequency distribution and the theoretical (normal) cumulative probability distribution.

the partitioned temperature data for all comparisons. The maximum deviations obtained for maximum temperature were 0.321 for May and 0.209 for August. The deviations for minimum temperature were 0.230 and 0.156 for May and August, respectively. These results suggest that the distribution of maximum and minimum temperature data obtained from Geneva is dependent on the occurrence of rainfall on the previous day.

A trivariate normal probability distribution was used to describe relationships among maximum and minimum temperature on the current day and maximum temperature on the previous day. To verify this assumption we first tested the bivariate normality of maximum and minimum temperature and then assessed the significance of the autocorrelation in maximum temperature data. The normality of average temperature values has been demonstrated (Jones *et al.*, 1972; Miller and Weaver, 1970; Thom, 1973), but we know of no work which demonstrates that a bivariate normal distribution is appropriate for describing maximum and minimum temperature. Because rigorous procedures for testing goodness of fit of bivariate data are unavailable (Kowalski, 1970), we used a "rough" test to determine the bivariate normality of maximum and minimum temperature from Geneva (Table 2). The analysis was performed for temperature data from May and August that was divided into two samples according to the occurrence of rain on the previous day. The test is based on the normality of linear combinations of the component bivariate normal random variables. We used a Smirnov test (Lilliefors, 1967) to evaluate the normality of maximum temperature, minimum temperature and the sum of the two. Since only one of the 12 samples was significantly ($P = 0.20$) different from a theoretical normal distribution, the assumption that a bivariate normal distribution accurately described the distribution of maximum and minimum temperatures values for Geneva was considered verified.

It remained to assess the significance of the autocorrelation of maximum temperature data. This was done by testing for significant deviations from zero of the value of the correlation coefficient for maximum temperature on the current and the previous day. Analysis was performed on subsamples partitioned according to the occurrence of rain on the previous day. All correlation coefficients for May and August were significantly ($P = 0.01$) different from zero indicating that the inclusion of autocorrelation in the generation of values of maximum temperature was justified.

We investigated the dependence of minimum relative humidity upon the occurrence of rainfall on both the preceding and present days. Average relative humidity has been described as a function of temperature and cloudiness (Adem, 1967). Since we had established that temperature was dependent upon the occurrence of precipitation on the previous day and because cloudiness was related to the occurrence of rain on any day, minimum relative humidity was considered dependent upon rainfall occurrence on the preceding day and the present day. To verify this, relative humidity values were separated into four groups on the basis of the occurrence of rain on the previous and present day. The distribution of each group was then compared with that of each of the other groups using a Kolmogorov-Smirnov test. The results of these six comparisons are consistent with our supposition that the distribution of daily values of minimum relative humidity is dependent upon the occurrence of rain on both the preceding and present day (Table 3).

A beta probability distribution has been shown to describe the distribution of relative humidity values (Yao, 1974), but generating random numbers from a beta distribution is prohibitively expensive when using the parameter estimates computed from actual daily weather data. We used a truncated normal distribution to describe minimum relative humidity and found that this was more appropriate than the beta distribution for describing relative humidity

TABLE 3. Dependence of minimum relative humidity on the occurrence of rain on the present day and the preceding day: results of analysis using Smirnov's maximum deviation test.

Comparison	May	August
dry/dry vs wet/wet	0.573**	0.660**
dry/dry vs dry/wet	0.393**	0.401**
dry/dry vs wet/dry†	0.215**	0.300**
wet/wet vs dry/wet	0.284**	0.323**
wet/wet vs wet/dry	0.418**	0.512**
wet/dry vs dry/wet	0.277*	0.217

* $P = 0.10$.
** $P = 0.05$.
† Comparison of relative humidity data occurring on dry days preceded by a dry day versus data from wet days preceded by a dry day. Wet indicates that daily precipitation is not less than 0.25 mm.

data from Geneva that had been partitioned into samples according to the occurrence of rain on the previous day and present day (Table 4).

Although previous reports have indicated that the distribution of solar radiation values is non-normal (Bennett, 1967, 1975), we found that normality was approximated by values of solar radiation from Geneva that had been partitioned on the basis of rainfall occurrence on the present day. The dependence of solar radiation on the current day's rainfall occurrence was introduced to accommodate the effect of cloud cover on incident solar radiation. The distribution of solar radiation values occurring on wet days was significantly different ($P = 0.01$) from the distribution of solar radiation values occurring on dry days for both May and August and suggested that the proposed dependence on the current day's rainfall occurrence was appropriate for solar radiation values from Geneva. The distribution of solar radiation values occurring on wet days did not differ significantly ($P = 0.10$) from a normal distribution for both May and August. The distribution of solar radiation values occurring on dry days, however, was non-normal. The deviations from normality appeared to result from positive skewness in the distribution of sample values. To maintain model simplicity and computing efficiency the normal probability distribution was used to describe all solar radiation values from Geneva although some error was anticipated to result from the skewness of the sample distribution.

TABLE 4. Kolmogorov-Smirnov test results from comparing the distribution of minimum relative humidity with the beta probability distribution and the normal probability distribution.

	May		August	
	Beta	Normal	Beta	Normal
dry/dry	0.175**‡	0.090*	0.135**	0.105**
wet/dry†	0.120**	0.081	0.132**	0.132**
dry/wet	0.124**	0.029	0.127**	0.091
wet/wet	0.153**	0.083	0.138**	0.052

* $P = 0.10$.
** $P = 0.05$.
† Wet day preceded by a dry day. Wet indicates that daily precipitation is not less than 0.25 mm.
‡ Maximum deviation between the sample cumulative frequency distribution and the theoretical cumulative probability distribution.

5. Model performance

To evaluate the performance of the weather model, output from the model was compared with weather data from Geneva, New York. To estimate the versatility of the weather model, model parameters were estimated for Fort Collins, Colorado and output was compared with weather data from that location. Five years of simulated data were compared with five randomly selected years of historical data for May–September from each location. For comparisons involving total solar radiation and minimum relative humidity for Geneva the historical weather data used was a subset of the data used for parameter estimation. All other comparisons employed historical weather data that was not used for parameter estimation.

Sequences of wet and dry days were analyzed to evaluate the performance of the Markov model used to describe precipitation occurrence. The relative frequency distribution of sequences of wet days for simulated data was not significantly different from the distribution for actual sequences of wet days for either Geneva or Fort Collins ($\chi^2 = 2.04$, d.f. = 4 for Geneva; $\chi^2 = 7.38$, d.f. = 4 for Fort Collins). The longest sequences of consecutive wet days were 7 (simulated) and 8 (historical) for Geneva and 7 (simulated) and 12 (historical) for Fort Collins. The distribution of sequences of dry days for simulated data was not significantly different from the distribution for actual data for Geneva ($\chi^2 = 5.11$, d.f. = 8). However, significant differences ($P = 0.01$) were observed between the distributions of simulated and actual data for Fort Collins ($\chi^2 = 20.78$, d.f. = 8). This difference was the result of a higher proportion of 8–10 dry-day sequences in the simulated data. The longest sequences of consecutive dry days were 17 (simulated and historical) for Geneva and 19 (simulated) and 21 (historical) for Fort Collins.

Several statistics were used to compare simulated weather data with actual weather data for a given weather variable. Smirnov's maximum deviation statistic was used to measure goodness of fit between the cumulative frequency distribution of actual data and the distribution of simulated data. The first three quartiles (values that exceed 25, 50 and 75% of the sample observations) were compared to detect differences in location between the distributions of simulated and actual data. The interquartile test (Bradley, 1968) and the sample range were used to analyze the variability within the two samples. Finally, the first-order autocorrelation coefficient (Box and Jenkins, 1970) was used to assess the degree of correlation between successive daily values of the weather variable.

The distributions of daily rainfall amounts for simulated and actual data were very similar (Table 5). The sample ranges were similar for simulated and historical data for Geneva and Fort Collins. The cumulative frequency distributions of rainfall amounts for simulated and actual data were not significantly different for either location. The significant difference ($\chi^2 = 6.21$, d.f. = 1, $P = 0.05$) between the first quartile value for simulated and actual rainfall amounts for Fort Collins was the result of irregularities in the form of the distribution of historical rainfall amounts. This difference was very

TABLE 5. Comparison of the cumulative frequency distribution for five seasons (May–September) of simulated weather with the cumulative frequency distribution of five seasons of actual data: daily amount of precipitation (mm).

	Geneva			Fort Collins		
	Actual		Simulated	Actual		Simulated
Autocorrelation	0.05		0.04	0.09		0.03
First quartile	1.25	n.s.	1.00	0.75	*	0.75
Median	3.50	n.s.	3.25	2.00	n.s.	2.25
Third quartile	8.50	n.s.	7.50	6.00	n.s.	5.50
Range	0.25–9.80		0.25–4.10	0.25–8.00		0.25–8.75
Interquartile	$\chi^2 = 0.49$ n.s.			$\chi^2 = 0.82$ n.s.		
Maximum Deviation	$D = 0.06$ n.s.			$D = 0.10$ n.s.		

* $P = 0.05$.
n.s. = not significant ($P = 0.10$).

small (0.25 mm) and was not considered to be of practical significance. No significant correlation was observed between successive values of precipitation for either simulated or actual data from both Geneva and Fort Collins. These results suggest that the rainfall component of the weather model can provide an accurate simulation of precipitation data from both Geneva and Fort Collins.

Simulated temperature data compared favorably with historical temperature data for both Geneva and Fort Collins (Table 6). As measured by the Smirnov statistic, there were no significant differences between the distributions of simulated and historical maximum temperature for either Geneva or Fort Collins. A difference of less than 1°C accounted for the significant difference ($P = 0.10$) between the third quartile of the simulated data and the third quartile of the actual maximum temperature data ($\chi^2 = 3.81$, d.f. = 1 for Geneva; $\chi^2 = 3.44$, d.f. = 1 for Fort Collins). Simulated maximum temperature values ranged from 6 to 38°C for Geneva and from 2 to 39°C for Fort Collins while actual maximum temperature data ranged from 5 to 35°C for Geneva and 1 to 37°C for Fort Collins. Significant autocorrelation among maximum temperature values was observed for simulated and actual data for both Geneva and Fort Collins. The distribution of temperature for simulated data was not significantly different from the distribution of actual data for Geneva, but significant differences ($P = 0.05$) were observed between the distributions of simulated and actual minimum temperature data for Fort Collins. Differences in variability occurred among the two samples (Table 6) and 1–2°C differences were observed between the first quartiles and between the medians for simulated and actual minimum temperature data from Fort Collins. The range of minimum temperature values was similar for actual and simulated data for Geneva and Fort Collins. Autocorrelation was significant ($P = 0.01$) for simulated and actual minimum temperature data for both locations, although historical minimum temperature data were more highly autocorrelated than simulated data. Despite the statistically significant differences observed between simulated and actual minimum temperature data, the overall performance of the

TABLE 6. Comparison of the cumulative frequency distribution for five seasons (May–September) of simulated weather with the cumulative frequency distribution of five seasons of actual weather: maximum and minimum temperature (°C).

	Maximum temperature						Minimum temperature					
	Geneva			Fort Collins			Geneva			Fort Collins		
	Actual		Simulated	Actual		Simulated	Actual		Simulated	Actual		Simulated
Autocorrelation	0.73		0.66	0.72		0.68	0.74		0.54	0.78		0.61
First quartile	20	n.s.	20	22	n.s.	22	9	*	8	7	†	6
Median	23	n.s.	23	26	n.s.	26	13	n.s.	13	11	**	10
Third quartile	28	*	28	29	*	29	17	n.s.	16	13	n.s.	13
Range	5–35		6–38	1–37		2–39	−2–22		−2–24	−3–19		−2–19
Interquartile	$\chi^2 = 1.91$ n.s.			$\chi^2 = 4.18$**			$\chi^2 = 3.58$*			$\chi^2 = 18.85$ (0.01)†		
Maximum deviation	0.05 n.s.			0.05 n.s.			0.06 n.s.			0.12**		

* $P = 0.10$.
** $P = 0.05$.
† $P = 0.01$.
n.s. = not significant ($P = 0.10$).

TABLE 7. As in Table 6 except for minimum relative humidity and total solar radiation.

	Minimum relative humidity (%)						Total solar radiation (cal cm^{-2} day^{-1})		
	Geneva			Fort Collins			Geneva		
	Actual		Simulated	Actual		Simulated	Actual		Simulated
Autocorrelation	0.21		0.14	0.35		0.17	0.34		0.11
First quartile	40	n.s.	40	41	n.s.	40	317	†	349
Median	48	*	49	53	n.s.	53	446	n.s.	456
Third quartile	60	n.s.	60	71	†	67	568	n.s.	556
Range	22–100		10–100	0–100		1.2–100	21–796		11–788
Interquartile	χ^2 = 7.99†			χ^2 = 4.49**			χ^2 = 11.49†		
Maximum deviation	0.03 n.s.			0.07 n.s.			0.05 n.s.		

* $P = 0.10$.
** $P = 0.05$.
† $P = 0.01$.
n.s. = not significant ($P = 0.10$).

temperature component appeared adequate for most agricultural applications.

Simulated minimum relative humidity data was similar to actual data in most comparisons (Table 7). The distribution of simulated minimum relative humidity values was not significantly different from the distribution of historical values for either Geneva or Fort Collins. The sample ranges of simulated and actual data were also similar. For both Geneva and Fort Collins interquartile tests did indicate significant differences ($P = 0.01$) in variability between simulated minimum relative humidity data and historical data. The autocorrelation of daily minimum relative humidity values was low for both simulated and actual data at both locations.

Except for small differences in variability, simulated and historical solar radiation values were not significantly different (Table 7). The distributions and ranges of simulated and actual solar radiation were similar. Variability of the simulated data was significantly different ($P = 0.01$) from that of historical data. This difference appeared to result from a higher frequency of solar radiation values below 300 cal cm^{-2} day^{-1} in the actual data. The level of autocorrelation for daily solar radiation values was low for simulated and actual data.

The general performance of the relative humidity and total solar radiation components of the model appeared good although the variability of simulated data was different from that of the historical data. While not directly tested, these deviations in variability appeared to result from asymmetry in the distributions of actual weather data.

As an additional measure of model performance the correlations among the weather variables were estimated for simulated and actual data for Geneva and Fort Collins. In general, the correlations were low for both simulated and actual data but certain significant correlations were consistently observed. Significant correlations were observed between maximum and minimum temperature (simulated = 0.81, actual = 0.85 for Geneva; simulated = 0.82, actual = 0.87 for Fort Collins), solar radiation and precipitation (simulated = −0.23, actual = −0.35 for Geneva), minimum relative humidity and precipitation (simulated = 0.23, actual = 0.36 for Geneva; simulated = 0.21, actual = −0.05 for Fort Collins) and solar radiation and minimum relative humidity (simulated = −0.26, actual = −0.71). Thus the pattern of correlation among the weather variables was similar for simulated and actual weather data.

6. Conclusion

The weather model was developed to fulfill two needs for models of agricultural systems. First, the weather model was to provide realistic daily weather data. Second, the weather model was to provide a stochastic element to the agricultural model through an accurate representation of the variability associated with actual weather data. When using parameters estimated with weather data from Geneva, New York or Fort Collins, Colorado, the weather model provided a useful representation of the mean response, the variability, and the pattern of correlation associated with daily weather data from each location. The simulations of precipitation, minimum temperature and maximum temperature were particularly accurate. This accuracy was very important for the precipitation component because of the primary role of rainfall information to the operation of the weather model. Observed differences between actual and simulated temperature data were usually less than 1°C and never more than 2°C for both Geneva and Fort Collins. The minimum relative humidity and total solar radiation components of the model also performed adequately. The differences in variability observed between simulated and historical relative humidity and solar radiation values appeared to result from skewness in the distributions of historical weather data. A more flexible probability distribution, such as the beta distribution, could be used to describe these weather variables and might reduce these differences. However the generation of random deviates from a beta distribution was prohibitively expensive. The pattern of correlations within and among daily values of the weather variables was also very similar for actual and simulated weather data.

The ability of the weather model to simulate weather data from Fort Collins suggests that the model might be used to simulate weather data from geographic locations other than Geneva. However, because the validity of the stochastic and structural assumptions made during the construction of the model may vary with geographic location (Bennett, 1975; Chin, 1977; Schmidt and Waite, 1962), verification of these assumptions should precede the use of the model for a given location. Seasonal effects also may alter the validity of the relationships that form the weather model (Chin, 1977).

The weather model provides daily weather data that can be used with agricultural models. Because the weather model accurately simulates daily weather data, it eliminates the need for large weather data files within such models. Since the variability of daily weather is also simulated, the weather model provides a stochastic element for models of agricultural systems that can be used to assess the risk associated with management activities when the environment is uncertain. The inclusion of variability in agricultural systems models also implies variability in model predictions which may be used to describe the likelihood of the occurrence of various levels of yield or yield loss. Because the weather model's parameters can be altered easily, it could be used with other agricultural models to evaluate the performance of management strategies under different climatic conditions and thus aid in the identification and construction of optimal management strategies.

Acknowledgments. The authors wish to thank Dr. Robert Seem, Department of Plant Pathology, New York Agricultural Experiment Station, Geneva, New York, for his assistance in obtaining weather data from Geneva, the Department of Seed and Vegetable Sciences, New York Agricultural Experiment Station for making the Geneva weather data available and Mr. Nolan Doesken, Department of Atmospheric Sciences, Colorado State University, Fort Collins, Colorado, for providing the weather data from Fort Collins. This work was supported in part by USDA/SEA CR Grant 701-15-54.

REFERENCES

Adem, J., 1967: Parameterization of atmospheric humidity using cloudiness and temperature. *Mon. Wea. Rev.*, **95**, 83–88.

Akaike, H., 1972: Information theory and an extension of the maximum likelihood principle. *Proceedings Second International Symposium on Information Theory*, B. N. Petrov and F. Csaki, Eds. Akademiai Kiado, Budapest, 267–281.

——, 1974: A new look at the statistical model identification. *IEEE Trans. Auto. Control*, **AC-19**, 716–723.

Barger, G. L., and H. C. S. Thom, 1949: Evaluation of drought hazard. *Agron. J.*, **41**, 519–526.

Bennett, I., 1967: Frequency of daily insolation in Anglo North America during June and December. *Solar Energy*, **11**, 41–55.

——, 1975: Variation of daily solar radiation in North America during the extreme months. *Arch. Meteor. Geophys. Bioklim.*, **B22**, 287–299.

Box, G. E. P., and G. M. Jenkins, 1970: *Time Series Analysis: Forecasting and Control*. Holden-Day, Inc., 553 pp.

Bradley, J. V., 1968: *Distribution-Free Statistical Tests*. Prentice-Hall, 388 pp.

Bruhn, J. A., W. E. Fry and G. W. Fick, 1979: WEATHER: A stochastic stimulation model of daily weather. *User's Manual*. Dept. of Plant Pathology, Cornell University, Mimeo 79-1, 47 pp.

Chin, C. H., 1977: Modeling daily precipitation occurrence with Markov chain. *Water Resour. Res.*, **13**, 949–956.

Dethier, B. E., and J. K. McGuire, 1961: Probability of selected weekly precipitation amounts in the northeastern region of the U.S. Cornell University Agric. Exp. Station, Agronomy Mimeo 61-4, 302 pp.

Gabriel, K. R., and J. Neumann, 1962: A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quart. J. Roy. Meteor. Soc.*, **88**, 90–95.

Gates, P., and H. Tong, 1976: On Markov chain modeling to some weather data. *J. Appl. Meteor.*, **15**, 1145–1151.

Jones, J. W., R. F. Colwick and E. D. Threadgill, 1972: A simulated environmental model of temperature, evaporation, rainfall and soil moisture. *Trans. Amer. Soc. Agric. Eng.*, **15**, 366–372.

Kowalski, C. J., 1970: The performance of some rough tests for bivariate normality before and after coordinate transformations to normality. *Technometrics*, **12**, 517–544.

Lilliefors, H. W., 1967: On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *J. Amer. Statist. Assoc.*, **62**, 399–402.

Miller, M. E., and C. R. Weaver, 1970: Extreme monthly and annual temperatures in Ohio. *Ohio Res. Bull.*, No. 1041, 35 pp.

Raudkivi, A. J., and N. Lawgun, 1972: Generation of serially correlated non-normally distributed rainfall durations. *Water Resour. Res.*, **8**, 398–409.

Schmidt, J. L., and P. J. Waite, 1962: Summaries of wet-bulb temperature and wet-bulb depression for grain drier design. *Trans. Amer. Soc. Agric. Eng.*, **5**, 186–189.

Schneider, B. E., and R. P. Clickner, 1970: On the distribution of the Kolmogorov-Smirnov test statistic for the gamma distribution with unknown parameters. Dept. Statistics, Temple University, Mimeo Series 36, 27 pp.

Thom, H. C. S., 1973: Distribution of wet-bulb temperature depression. *Arch. Meteor., Geophys. Biolim.*, **B21**, 43–54.

Tong, H., 1975: Determination of the order of a Markov chain by Akaike's information criterion. *J. Appl. Prob.*, **12**, 488–497.

Weiss, L. L., 1964: Sequences of wet or dry days described by a Markov chain probability model. *Mon. Wea. Rev.*, **92**, 169–176.

Wiser, E. H., 1966: Monte Carlo methods applied to precipitation frequency analyses. *Trans. Amer. Soc. Agric. Eng.*, **9**, 438–540, 542.

Yao, A., 1974: Statistical model for the surface relative humidity. *J. Appl. Meteor.*, **13**, 17–21.