SIMULATING DAILY WEATHER VARIABLES

Research and Development Branch
Research Division
Statistical Reporting Service
U.S. Department of Agriculture
Washington, D.C.

November 1977

TABLE OF CONTENTS

SIMULATING DAILY WEATHER VARIABLES

Keith Crank

I. Modeling Daily Weather Variables

In recent years attempts have been made to use computers to simulate plant growth, in terms of leaf number and size as well as dry matter accumulation, for different types of farm crops. One such program developed by Professor Gerald Arkin of Texas A&M University, is designed to simulate growth for grain sorghum. This program requires daily inputs of solar radiation, precipitation, and maxiumum and minimum temperature to produce an estimate of crop yield.

The Statistical Reporting Service of the Department of Agriculture would like to use simulation programs to help forecast crop production in the United States. Unfortunately, for models such as Professor Arkin's to work, a knowledge of daily weather data for the entire growing season is required. This knowledge is not available for early season forecasts. However, if daily weather variables satisfy known probability distributions, it would be possible to take random samples from these distributions, input these possible future weather values into the simulation program and thereby make statements about expected crop yields and construct confidence intervals for these expected crop yields. The purpose of this research was to try to find distributions for the weather variables and then to set up a procedure which would produce possible daily values for these weather variables.

The first approach examined was to look at the variables individually. Although this may not seem like a very reasonable idea, since weather variables are related, it would have been beneficial if one variable alone could have been used in determining the others. Unfortunately this was not possible. There was enough between-day correlation of variables to prevent the variables from being treated independently from day to day. This resulted in a new approach in which one day's weather was studied to see how it was affected by the previous day's weather.

Precipitation

Precipitation, which had the highest between-day correlation, was the first variable studied by the new approach. Precipitation was treated as a Markov process with the states determined by whether or not rain occurred. Precipitation was denoted by a zero-one variable: rain by a one and no rain by a zero for each day. Different precipitation values were considered as cutoff points in determining the occurrence or non-occurrence of rain. For example the National Weather Service does not measure amounts less than .01 inch. This level was used as well as levels of .1 inch, .2 inch and .25 inch. By doing a chi-square test on frequency tables for daily data, it could be determined at what value of precipitation one day's knowledge was significant.

Note: This research was done while the author was a trainee for the Statistical Survey Institute, a program sponsored by the American Statistical Association.

Using thirteen years of data from the Des Moines, Iowa airport weather station, the knowledge of the previous day was significantly useful in determining occurrence of rain. This was for a significance level of .05 and occurrence of rain meant at least .01 inch. When this proved successful, the level for occurrence of rain was increased to .1, .2 and .25 inch. However these values were not significant at the .05 level of significance. Using more than one previous day to determine the states in the Markov process was also investigated. For example in a two day process the states were 00, 01, 10, and 11. The states 00 and 11 represented two consecutive days of no rain and two consecutive days of rain, respectively. State 01 was a day of no rain followed by a day of rain, and state 10 was a day of rain followed by a day without rain. Transition probabilities were then determined using the Iowa data (See Figure 1). The level of significance for this varied from month to month; however, for the months of July and August, which are the most important months for simulating weather data for plant growth, using more than one day was not significant.

Therefore the occurrence of rain on a given day was treated as a binomial random variable with the probability dependent upon the state the Markov process was in and the month under consideration. The one day Markov transition matrices were found to converge to the steady state probabilities in approximately four days. (See Figure 2). Unfortunately, this did not give any method for determining amount of rainfall on days that rain occurred. When no method presented itself right away, a completely empirical, and discrete, method was formulated. Instead of breaking the data into two states (occurrence and non-occurrence), it was divided into a state for no rain and a number of states for different levels of rain. (See Table 1). However, this was not completely satisfactory.

Figure 1: Transition Matrix for two day Markov Process.

	00	01	10	11
00	P_1	01 - P_1	0	0
01	0	0	P_2	1 - P_2
10	P_3	1 - P_3	0	0
11	0	0	P_4	1 - P_4

Note: If there have been two consecutive days without rain the process is in State 00. If the following day is dry the process remains in State 00, but if rain occurs the process goes to State 01. It is impossible to go from State 00 to either State 10 or State 11 in one step.

Figure 2: Convergence of one day Markov Transition Matrices. P is the Probability Matrix for July.

$$P = \begin{bmatrix} .7396 & .2604 \\ .6174 & .3826 \end{bmatrix} \quad P^4 = \begin{bmatrix} .7034 & .2966 \\ .7032 & .2968 \end{bmatrix} \quad P^6 = \begin{bmatrix} .7034 & .2966 \\ .7034 & .2966 \end{bmatrix}$$

Table 1: Empirical Cumulative Probabilities for Amount of Precipitation based on Previous Day's Precipitations, Des Moines, Iowa; July.

Present Day / Previous Day	<.01"	≥.01"	≥.10"	≥.25"	≥.50"	≥1.0"
<.01"	.7396	.2604	.1667	.1354	.0694	.0312
≥.01"	.6174	.3826	.2000	.1391	.0609	.0348
≥.10"	.6429	.3571	.2000	.1286	.0714	.0429
≥.25"	.6465	.3455	.1818	.1273	.0727	.0545
≥.50"	.6154	.3846	.1154	.1154	.1154	.0769
≥1.0"	.6154	.3846	.1538	.1538	.1538	.1538

Therefore, the data were separated into days of rain and days without rain. For days when rain occurred, the data were sorted by amounts of rain and numbered consecutively from smallest to largest. The graph of the amount of rain against its rank provided the empirical distribution function for each month. A nonlinear regression of amount of rain agains its rank using the equation

$$\text{Rank} = b\ \text{Ln}(a(\text{amount})+1)$$

was performed separately for each month and in each case came up with an R^2 greater than .990. (See Appendix A). In conclusion, the "best" method for determining rainfall was to treat occurrence as a binomial random variable, with the probability of occurrence based on the amount of rain that fell the previous day, i.e. only the first two columns from Table 1 were used. For days when rain was to occur, it was assumed that the amount of rain satisfied the cumulative distribution function

$$F(\text{amount}) = \frac{\hat{b}}{n}\ \text{Ln}(\hat{a}(\text{amount})+1),$$

where $\hat{a}$ and $\hat{b}$ are the parameters estimated in the nonlinear regression, and n is the maximum value for rank in the regression.

Sunshine

Sample correlations were determined for each week between pairs of variables: precipitation, maximum temperature, minimum temperature, and minutes of sunshine. Since the correlation between precipitation and sunshine was highest, this variable was analyzed in a manner similar to that used for precipitation.

First the days were divided into two groups: those when rain occurred and those when no rain occurred. A linear regression was done for each section for each week with minutes of sunshine as the dependent variable and its rank as the independent variable. The model that was used for days that no rain occurred included both a linear and a quadratic term, whereas the model for days tht rain did occur included only a linear term. The model was better for days that had no rain than for days that rain occurred. However, the results obtained seemed sufficiently good enough to use. Unfortunately, when a model was programmed using the parameters estimated by the regressions, it became apparent that temperature also had to be considered in determining minutes of sunshine.

Therefore the final model was based on a three way empirical frequency table of minutes of sunshine conditioned by both precipitation and maximum temperature for that day. (See Table 2). For example on a July day in Des Moines when rainfall is less than .01 inch and the maximum temperature is 87, the empirical probability of having less than 700 minutes of sunshine (from Table 2) is .3587 while the probability of having between 600 and 700 minutes of sunshine is .3587 - .2442 = .1145.

Temperature

The correlation between precipitation and maximum temperature was high enough to indicate that precipitation should be used as a conditioning factor in determining maximum temperature. However, precipitation alone was not enough to give adequate results. The problem that occurred was that a large amount of day to day variation in maximum temperature was present, and this did not seem to be realistic. The result was, as for sunshine, a three way empirical frequency table where maximum temperature was conditioned on amount of precipitation and the previous day's maximum temperature. (See Table 3).

As an example of the use of Table 3 suppose that it will not rain today and that yesterday's maximum temperature was in the high 90's. Then the empirical probability that the maximum temperature today will fall below 80 is only .0375, while the probability that the maximum temperature will again be in the 90's is .8750 - .2375 = .6375.

Table 2: Empirical Cumulative Probabilities for Minutes of Sunshine based on Amount of Precipitation and Maximum Temperature; Des Moines, Iowa; July.

Precip < .01"

Sunlight in Mins / Max Temperature	<100	<200	<300	<400	<500	<600	<700	<800	<900
< 80	.1064	.1489	.1702	.1915	.2766	.3404	.4468	.5957	1.000
80 - 89	.0229	.0229	.0534	.1221	.1679	.2442	.3587	.5495	1.000
90 - 99	.0000	.0000	.0000	.0108	.0323	.0968	.2258	.4193	1.000
100 - 104	.0000	.0000	.0000	.0000	.0000	.0000	.0769	.3077	1.000

.01" ≤ Precip < .10"

Sunlight in Mins / Max Temperature	<100	<200	<300	<400	<500	<600	<700	<800	<900
70 - 79	.4286	.5715	.5715	.7144	.7858	1.000	1.000	1.000	1.000
80 - 89	.0400	.1200	.3200	.5200	.6800	.7600	.9200	.9600	1.000
90 - 99	.0000	.0000	.0000	.1111	.2222	.5556	.7778	1.000	1.000

.10" ≤ Precip < .25"

Sunlight in Mins / Max Temperature	<100	<200	<300	<400	<500	<600	<700	<800	<900
70 - 79	.5000	.5000	.5000	.7500	.7500	.7500	.7500	.7500	1.000
80 - 89	.0000	.2500	.2500	.3750	.7500	.7500	.7500	.8750	1.000
90 - 99	.0000	.2500	.2500	.2500	.5000	.5000	1.000	1.000	1.000

Precip ≥ .25"

Sunlight in Mins / Max Temperature	<100	<200	<300	<400	<500	<600	<700	<800	<900
< 80	.5000	.6875	.7500	.8750	.9375	.9375	1.000	1.000	1.000
80 - 89	.0690	.1034	.2414	.3448	.5862	.7586	.9655	.9655	1.000
90 - 104	.0000	.0000	.0000	.1000	.1000	.4000	.6000	1.000	1.000

Table 3: Empirical Cumulative Probabilities for Maximum Temperature based on Precipitation and Previous Day's Maximum Temperature; Des Moines, Iowa; July.

Precip < .01"

Present Day / Previous Day	<70	<80	<90	<100	<104
60 - 69	.0000	1.000	1.000	1.000	1.000
70 - 79	.0159	.3651	1.000	1.000	1.000
80 - 89	.0079	.1496	.7244	.9921	1.000
90 - 99	.0000	.0375	.2375	.8750	1.000
100 - 104	.0000	.0000	.1667	.8333	1.000

.01" ≤ Precip < .10"

Present Day / Previous Day	<80	<90	<100
< 80	.6667	.8889	1.000
80 - 89	.2692	.9615	1.000
90 - 99	.0909	.3636	1.000
100 - 104	.0000	1.000	1.000

.10" ≤ Precip < .25"

Present Day / Previous Day	<80	<90	<100	<104
< 80	.0000	1.000	1.000	1.000
80 - 89	.1429	1.000	1.000	1.000
90 - 99	.3750	.5000	.8750	1.000

Precip ≥ .25"

Present Day / Previous Day	<80	<90	<100	<104
< 80	.4000	1.000	1.000	1.000
80 - 89	.3667	.8667	1.000	1.000
90 - 99	.1500	.7000	.9500	1.000

Once the maximum temperature was determined, minimum temperature became the final random variable to be found. Analysis showed that minimum temperature and maximum temperature were approximately bivariate normal. Under this assumption a certain linear combination of maximum and minimum temperature is a univariate normal random variable. This combination is

$$X_2 - \frac{\sigma_{12}}{\sigma_1^2} X_1,$$

Where X_2 = minimum temperature, X_1 = maximum temperature, σ_1 = standard deviation of X_1, and σ_{12} = covariance of X_1 and X_2. (See Appendix B).

Then $X_2 = \frac{\sigma_{12}}{\sigma_1^2} X_1 + Y$ where Y is a normal random variable. A linear regression of maximum temperature on minimum temperature provides estimates for $\frac{\sigma_{12}}{\sigma_1^2}$ and for E(Y). An estimate of Var(Y) comes from the sample variance of the residuals about the regression line.

Conclusion

The distributions finally used in the simulation program were a combination of empirical frequencies and regressions on both empirical cumulative frequencies and other variables. The empirical frequencies were used in determining maximum temperature, minutes of sunshine, and occurrence of precipitation. The regressions were performed on cumulative frequencies of amounts of rainfall and on maximum temperature to determine amount of rainfall and minimum temperature, respectively.

The methods used in analyzing the available weather data are by no means exhaustive. More types of analysis can and should be done. The program developed in this paper to determine possible sequences of daily summer weather variables is at least adequate for its intended use.

Possible new approaches to the problem could include studying climatic trends such as a supposed decrease in average temperatures, or studying long periods of dry days to see if these can be incorporated into a model. Another idea would be to find out if wet periods produce an approximately constant amount of rainfall, i.e., does a storm that lasts only one day produce the same amount of precipitation as a storm that lasts two or three days? Also of use would be a way to reduce the periods of projection to a weekly or bi-weekly level instead of the present monthly level.

II. Simulation of Daily Weather Variables

The following models were designed to simulate daily weather data for the months of June, July and August. Although it was designed using data from the Des Moines, Iowa airport weather station, it should be possible to modify it for other parts of the country. It should be noted that this model does not attempt to predict or forecast weather, but only to produce possible sequences of daily weather data. The daily weather sequences generated correspond to "normal" weather based on historical records.

In order to begin simulation two input parameters are required. The first, precipitation for the day preceding the first day to be simulated, is used in determining the initial day's precipitation. The second, maximum temperature for the day preceding the first day to be simulated, is one parameter used to determine the maximum temperature.

For each day to be simulated, seven random numbers are generated. Six of these should follow a distribution which is approximately uniform on the interval (0, 1). The seventh should satisfy a standard normal distribution. All of the numbers are considered to be independent both from day to day and within days. How well these properties are satisfied depends, naturally, on the random number generators used.

Using the known value for the previous day's precipitation, the model uses an empirical probability (which differs for each month) and the first uniform random number to determine whether or not there would be rain for that day. (See Table 1). If no rain is to occur that day, the variable "precip" is set at zero and the second uniform random variable is not used. If rain is to occur the second uniform random number is used in an equation to determine the amount of rain. This equation is based on the estimates of the coefficients of a nonlinear regression model based on historic data. This model is fitted to the cumulated frequencies of days by precipitation amounts for days that rain occurred for that month. (See Appendix A).

Table 1: Markov Probabilites of Precipitation based on Knowledge of day before using .01 inch as cutoff point. Des Moines, Iowa; Month of July.

Today / Yesterday	0	≥.01"
0	.7396	.2604
≥.01"	.6174	.3826

Once the amount of rain has been determined, this is used, along with the previous day's maximum temperature, to determine this day's maximum temperature. Using empirical frequency tables and the third uniform random number an initial value is given for maximum temperature. (See Table 2). This value is 60, 70, 80, 90, or 100. This value is changed to integral values of degrees by adding ten times another uniform random number to the initial value at the end of the program except when the initial value is 100. If the value is 100 then it is changed by adding four times the random number to it. This gives a possible temperature range of 60 to 104 with uniform distributions within a ten degree interval and an empirical distribution between intervals.

The minutes of sunlight per day are determined as an alternative to solar radiation in the same way as the maximum temperature, except that the conditioning variables are rainfall and maximum temperature for the same day. (See Table 3). One random number is used to set an initial value for the minutes of sunshine, and a second random number is used to smooth out the possible values this variable can take on. These two random numbers are not the same ones that are used to determine maximum temperature.

Table 2: Empirical Cumulative Distribution Table of Maximum Temperature based on Rainfall and Previous Day's Maximum Temperature. Precip <.01", Des Moines, Iowa; July

Today Yesterday	≤69	≤79	≤89	≤99	≤104
60 - 69	.0000	1.000	1.000	1.000	1.000
70 - 79	.0159	.3651	1.000	1.000	1.000
80 - 89	.0079	.1496	.7244	.9921	1.000
90 - 99	.0000	.0375	.2375	.8750	1.000
100 -	.0000	.0000	.1667	.8333	1.000

Example of use of Table 2: If yesterday's maximum temperature was 87 and the uniform random number generated was .5000 the initial value for today's maximum temperature is found by locating the proper row (in this case 3), and going across that row to the first column that has an entry bigger than the random number (in this case 3). The initial value is then found by rounding the number heading that column to the next lower multiple of ten (in this case 80).

Finally minimum temperature is determined. It is assumed that minimum and maximum temperature satisfy a bivariate normal distribution. (See Appendix B). Hence minimum temperature can be determined by summing the following three quantities: 1) A constant (based on the mean of the historical data); 2) A constant multiplied by the maximum temperature (the constant being based on the standard deviation of both maximum and minimum temperature

Table 3: Empirical Cumulative Distribution Table of Minutes of Sunshine based on Amount of Precipitation and Maximum Temperature. Precip $\geq$.25", Des Moines, Iowa; July.

Sunshine Max Temperature	<100	<200	<300	<400	<500	<600	<700	<800	<900
60 - 69	.7273	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
70 - 79	.2632	.4211	.4737	.8421	.8947	.9474	1.000	1.000	1.000
80 - 89	.0000	.0000	.1600	.3200	.6000	.7600	.8800	.8800	1.000
90 -	.0000	.0000	.0000	.0000	.1667	.3333	.8333	.8333	1.000

data and their correlation coefficient); and 3) Another constant multiplied by the normal random number (the constant being based on the standard deviation of minimum temperatures and the correlation coefficient between maximum and minimum temperature).

At this point the day's weather variables are printed out and the amount of precipitation and maximum temperature are carried forward to simulate the same four weather values for the following day. This is done until the daily weather data for the period desired is completely determined.

Comments on Model

Points about the model which should be noted are the following:

1) Empirical frequency distributions for precipitation, maximum temperature, and amount of sunlight are based on data for an entire month.

2) Means, standard deviations, and correlation coefficients for maximum and minimum temperature are based on data separated into weekly intervals.

3) Parameter values used in determining minimum temperature are different for days with rain and days without rain.

This simulation model is based on empirical data for less than 25 years. No climatic trends are incorporated. This is one area where research might be useful if historical records are based on more than 25 years. Also of value would be an attempt to better determine the intricate relations between all of these variables.

III. Fortran Program to Simulate Daily Weather Variables.

Flow Chart of Program:

Start	
Read In Parameters	These are the probability levels used in determining the variables.
Compute Random Numbers	Six uniform (0, 1) random numbers and one standard normal random number.
Compute Amount Of Precipitation	Precipitation is based on amount of precipitation for the day before.
Compute Maximum Temperature	Maximum temperature is based on the amount of rainfall already computed and on the previous day's maximum temperature.
Compute Minutes Of Sunlight	Minutes of sunlight is based on the values of maximum temperature and amount of precipitation that have already been computed.
Compute Minimum Temperature	Minimum temperature is computed from the parameters from a linear regression of minimum temperature on maximum temperature.
Add Control Parameters	These parameters can be used to produce specific weather conditions such as abnormally high average temperatures.
Print Daily Values	
End	

Dictionary For Program

S	An array containing probabilities for levels of sunlight.
T	An array containing probabilities for levels of maximum temperature.
A	An array of random numbers.
C	Control parameters to adjust averages.
R	An array containing probabilities for occurrence of rain.
D	Parameter estimates used to determine amount of rain on days that rain occurs.
E	Parameter estimates used to determine minimum temperature. First three values are used on days without rain. Last three values are used on days that rain occurs.
P	Amount of precipitation on previous day.
YT	Maximum temperature on previous day divided by 10.
IDAY	Day of month
PRECIP	Amount of rainfall
MAXTEMP	Maximum temperature
MINTEMP	Minimum temperature
ISUN	Minutes of sunshine

```
      DIMENSION S(4,5,10),T(4,5,4),A(7),C(4),R(5),D(3),E(6)
      INTEGER IX
      REAL RX
      READ(5,3000)(((S(I,J,K),K=1,10),J=1,5),I=1,4)
      READ(5,3001)(((T(I,J,K),K=1,4),J=1,5),I=1,4)
      READ(5,3000)(R(I),I=1,5),(C(K),K=1,4)
      READ(5,3004)(D(I),I=1,3),(E(K),K=1,6)
      READ(5,3005)P,YT,IX
      WRITE(6,5001)
      WRITE(6,5000)(((S(I,J,K),K=1,10),J=1,5),I=1,4)
      WRITE(6,5002)
      WRITE(6,5003)(((T(I,J,K),K=1,4),J=1,5),I=1,4)
      WRITE(6,5002)
      WRITE(6,5000)(R(I),I=1,5),(C(K),K=1,4)
      WRITE(6,5000)(D(I),I=1,3),(E(K),K=1,6)
      WRITE(6,5005)P,YT,IX
      WRITE(6,2000)
      WRITE(6,2001)
      IX=IX+500
      DO 100 IDAY=1,31
      IYT=YT
      DO 101 K=1,6
      IX=IX*65539
      IF(IX.LT.0) IX=IX+2147483647+1
      RX=IX
101   A(K)=RX*.4656613E-9
      AA=0.0
      DO 50 I=1,12
      IX=IX*65539
      IF(IX.LT.0) IX=IX+2147483647+1
      RX=IX
      RX=RX*.4656613E-9
50    AA=AA+RX
      A(7)=(AA-6.0)
      IF(P.LT..01) I=1
      IF(P.GE..01.AND.P.LT..10) I=2
      IF(P.GE..10.AND.P.LT..25) I=3
      IF(P.GE..25.AND.P.LT..50) I=4
      IF(P.GE..5) I=5
      PRECIP=0
      IF(A(1).LT.R(I)) PRECIP=1
      IF(PRECIP.GT.0) PRECIP=(EXP(D(1)*A(2)/D(2))-1)/D(3)
       IF(PRECIP.LT..01) I=1
      IF(PRECIP.GE..01.AND.PRECIP.LT..10) I=2
      IF(PRECIP.GE..10.AND.PRECIP.LT..25) I=3
      IF(PRECIP.GE..25) I=4
      IF(IYT-6)112,112,113
112   J=1
      GO TO 1000
113   IF(IYT-9)114,114,115
115   J=5
      GO TO 1000
114   J=IYT-5
1000  DO 1001 IK=1,4
      K=5-IK
      IF(A(3).GT.T(I,J,K)) GO TO 1002
1001   CONTINUE
      K=0
 1002 MAXTMP=60+K*10
      J=K+1
      DO 1003 IK=1,10
      K=11-IK
      IF(A(5).GT.S(I,J,K)) GO TO 1004
1003   CONTINUE
      K=0
1004  SUN=100*K+100*A(6)+0.5
      TMPMAX=MAXTMP+10*A(4)+0.5
      IF(TMPMAX.GT.100) TMPMAX=100+4*A(4)+0.5
      MAXTMP=TMPMAX+C(2)
      IF(I.GE.2) I=4
      J=I+1
      K=I+2
     TMPMIN=E(I)+E(J)*MAXTMP+E(K)*A(7)+0.5
      MINTMP=TMPMIN +C(3)
      IP=100.*PRECIP+0.5
      PRECIP=IP/100.+C(1)
      ISUN=SUN+C(4)
      WRITE(6,2002)IDAY,PRECIP,MAXTMP,MINTMP,ISUN,(A(I),I=1,7)
      IF(IDAY.EQ.10) WRITE(6,2004)
      IF(IDAY.EQ.20)WRITE(6,2004)
      YT=MAXTMP/10.
      P=PRECIP
100   CONTINUE
2000  FORMAT(1H1,25X,' JULY')
2001  FORMAT(1H0,5X,'DAY PRECIP MAXTEMP MINTEMP SUNSHINE',///)
2002  FORMAT(1H ,I8,1X,F5.2,I7,I8,1X,I8,7(2X,F7.4))
2004   FORMAT(1H0)
3000  FORMAT(10F6.4)
3001  FORMAT(8F6.4)
3004  FORMAT(9F6.2)
3005   FORMAT(2F6.4,I8)
5000  FORMAT(1H ,10X,10(2X,F9.4))
5001  FORMAT(1H1)
5002  FORMAT(1H0)
5003  FORMAT(1H ,10X,8(2X,F8.4))
5005  FORMAT(1H ,10X,F9.4,F9.4,I8)
      STOP
```

APPENDIX A

Figure 1: Plot of amount of rain against its rank on days when rain occurred. Based on this empirical relationship, the ranks are selected with equal probabilities and for a selected rank the precipitation amount is determined.

Graph is attached.

A nonlinear regression was fitted to the data in Figure 1; the regression equation used was

$$1)\quad \text{Rank} = b\ \text{Ln}(a(\text{amount})+1)$$

For the month of June the estimated coefficients were (See Table 4)

$$\hat{a} = 47.66$$

$$\hat{b} = 62.18$$

Using this data set and the estimated parameters, the largest predicted value was approximately 300. Thus by dividing rank by 300 we get values between zero and one.

Letting $U = \frac{\text{rank}}{300}$ Equation (1) becomes

(2) $U = \frac{b}{300}\ \text{Ln}\ (a(\text{amount})+1)$ with $0 \leq u \leq 1$

Solving Equation (2) for amount produces

Amount = $(\text{Exp}(300\ U/\hat{b})-1)/\hat{a}$.

Drawing values of U from a uniform (0, 1) distribution produces values of rain which are distributed approximately the same as the empirical distribution.

APPENDIX B

If X_1, X_2 are random variables which satisfy a bivariate normal distribution where $E(X_1) = \mu_2$ and $\Sigma = \begin{matrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{matrix}$, then there are random variables Y_1, Y_2, which are linear transformations of X_1 and X_2 which are independent, normally distributed random variables.

The transformation which produces Y_1 and Y_2 is $Y_1 = X_1$ and

$$Y_2 = X_2 - \frac{\sigma_{12}}{\sigma_1^2} X_1$$

Then $E(Y_2) = \mu_2 - \frac{\sigma_{12}}{\sigma_1^2} \mu_1$ and $\text{var}\ (Y_2) = \sigma_2^2\ (1 - \rho^2)$ where

$$\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$$

Letting z be a random number drawn from a standard normal distribution, $y = \mu_2 - \frac{\sigma_{12}}{\sigma_1^2} \mu + \sigma_2 \sqrt{(1 - \rho^2)}\ z$ is a random number drawn from the distribution of Y_2. By first determining values for X_1, we can produce values for X_2 by

$$X_2 = Y + \frac{\sigma_{12}}{\sigma_1^2} X_1 = (\mu_2 - \frac{\sigma_{12}}{\sigma_1^2} \mu_1) + \frac{\sigma_{12}}{\sigma_1^2} X_1 + \sigma_2 \sqrt{(1 - \rho^2)}\ z$$

If maximum temperature replaces X_1 and minimum temperature replaces X_2, this equation becomes the one used in the simulation weather model.

PLOT OF PRECIP*N   LEGEND: A = 1 OBS , B = 2 OBS , ETC

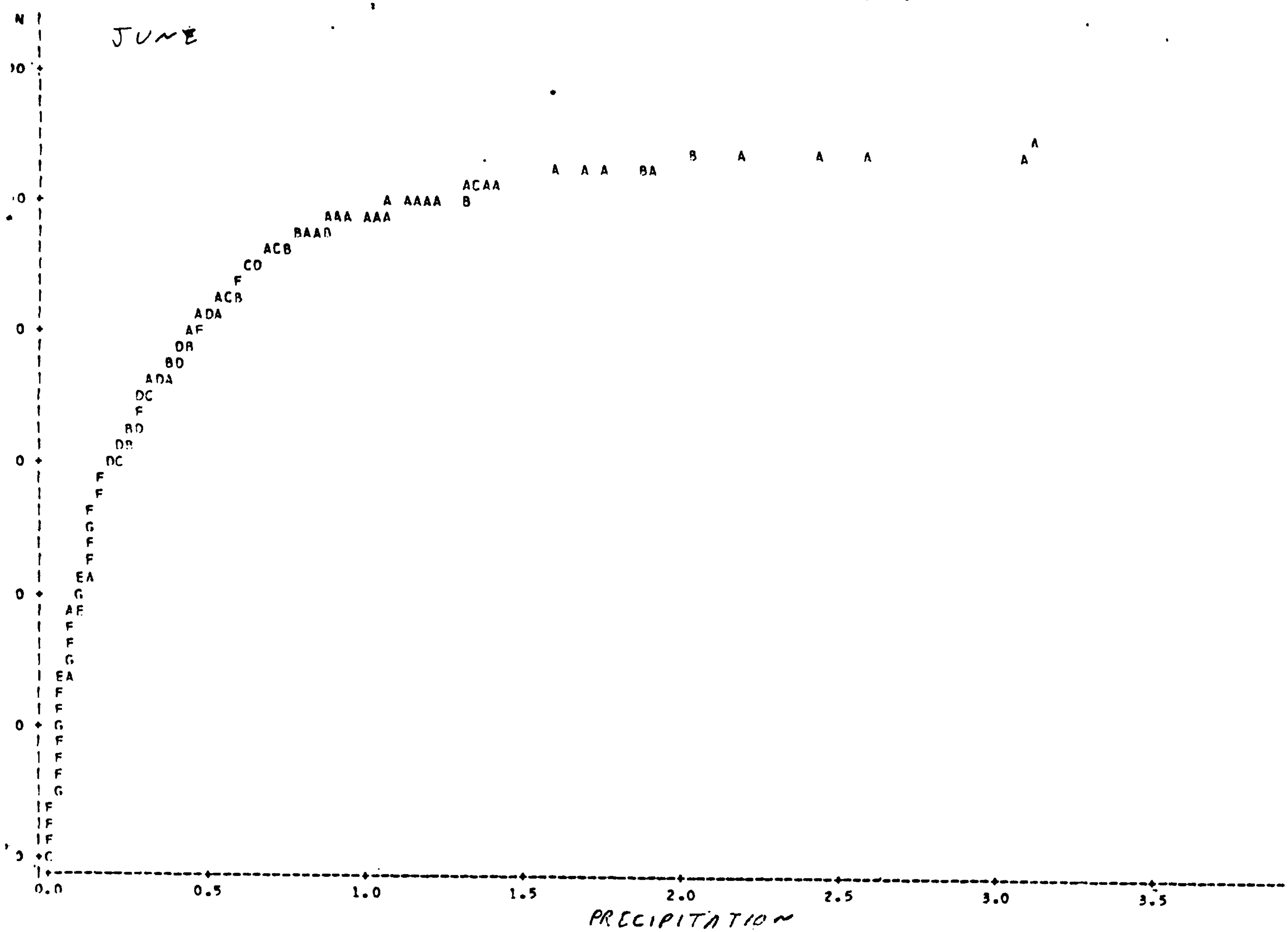

TABLE 4 : NON-LINEAR LEAST SQUARES SUMMARY STATISTICS   DEPENDENT VARIABLE N

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
REGRESSION	2	6725132.71552468	3362566.35776234
RESIDUAL	270	19787.28447532	73.28623880
UNCORRECTED TOTAL	272	6744920.00000000	
(CORRECTED TOTAL)	271	1676948.00000000	

PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR	ASYMPTOTIC 95 % CONFIDENCE INTERVAL LOWER	UPPER
B	62.18267890	0.69193074	60.82039524	63.54496257
A	47.65942586	1.63227729	44.44577372	50.87307799

ASYMPTOTIC CORRELATION MATRIX OF THE PARAMETERS

	B	A
B	1.000000	-0.954981
A	-0.954981	1.000000