EMPIRICAL DISTRIBUTIONS AND PRODUCTION ANALYSIS: A DOCUMENTATION USING METEOROLOGICAL DATA

Larry W. VanTassell
James W. Richardson
J. Richard Conner

Bulletin 671, September 1989
The University of Tennessee
Agricultural Experiment Station
Knoxville, Tennessee
D. O. Richardson, Dean

LARRY W. VANTASSELL is an Assistant Professor in the Department of Agricultural Economics and Rural Sociology at the University of Tennessee, Knoxville.

JAMES W. RICHARDSON is a Professor in the Department of Agricultural Economics at Texas A&M University, College Station.

J. RICHARD CONNER is a Professor in the Department of Agricultural Economics at Texas A&M University, College Station.

Table of Contents

	Page
Introduction	1
Empirical Probability Distributions	2
Generating Values from Empirical Distributions	3
Temporal Correlation of Empirical Distributions	7
Intertemporal Correlation of Empirical Distributions	9
An Empirical Example	11
Summary and Conclusions	17
References	21
Appendix A	23
Appendix B	30
Description of the Computer Program	31
Main body of the model	31
Subroutine FACTOR	32
Subroutine EMP	33
Subroutine STATS	33
Coding instructions for all input data	34

Abstract

Weather is a primary source of risk and uncertainty in the production of agricultural commodities. Incorporation of meteorological variables in simulation models requires the recreation of the same stochastic relationships which underlie the basic meteorological process. This paper presents a methodology for using Monte Carlo techniques to simulate meteorological values on an aggregated basis (e.g., monthly or quarterly) using empirical distributions. An example for precipitation and temperature variables is developed with endpoints of the empirical functions distributed exponentially, stacked, and standard. The statistical properties observed in the historical series appears to be closely maintained in the simulated series.

EMPIRICAL DISTRIBUTIONS AND PRODUCTION ANALYSIS: A DOCUMENTATION USING METEOROLOGICAL DATA

Introduction

Weather variability induces a substantial amount of uncertainty in the production of agricultural commodities. The development of simulation models for planning, evaluation, and design often requires inclusion of weather uncertainty when risk is to be incorporated into the study.

The relationship between an agricultural process and meteorological variables is complex and often requires elaborate modeling to capture the interactions involved. Studies that track the production process in detail usually simulate a particular system on a daily or more frequent basis. For other studies, it is only necessary to tract the variables under consideration on a monthly or yearly basis to represent the production and variability affecting the system.

Two alternatives are available for representing meteorological conditions when simulating production systems. The first is to use an observed sequence of weather data for the analysis and the second is to use Monte Carlo techniques to simulate alternative weather sequences. The use of an observed sequence of weather data provides an examination of a system under one realization of the weather process and, therefore, does not consider the full range of possible results. The question lingers as to how the system would fair if an equally likely sequence of weather were observed (Richardson, 1981).

Most models that simulate meteorological variables use a Markov chain to determine precipitation occurrence and then utilize various multivariate distributions (e.g., gamma distributions) to generate precipitation amounts, temperatures and other weather variables on a daily basis (Richardson, 1981; Nicks and Harp; Larsen and Pense). When studies only require meteorological data on a monthly or more aggregate basis, simulating data on a daily basis can be unduly time consuming and data demanding.

The objective of this paper is to describe and demonstrate a technique for using Monte Carlo techniques to simulate meteorological values on a monthly or more aggregate basis. The paper is organized in three sections. The following section presents a general procedure for generating random values from multivariate empirical probability distributions. The second section presents the results of an empirical test of the procedure and statistical tests of its performance. The third section summarizes the procedures and the empirical test. An appendix is included to describe the FORTRAN model used to generate multivariate empirical random values for the example.

Empirical Probability Distributions

Several theoretical distributions have been developed and used to describe an underlying distribution for variables that are random or stochastic. Generally a theoretical distribution can be found that accurately describes the stochastic process at hand, but at other times it is difficult to obtain a distribution that provides an adequate fit. Several researchers (Richardson and Condra, 1981; VanTassell; Perry, et al.; Lemieux, et al.) have, therefore, utilized the actual data themselves to define an empirical distribution.

Generating Values from Empirical Distributions

Empirical distributions can be defined when the actual values of the individual observations are available or when only intervals are specified along with the number of observations falling within each interval (Law and Kelton, pp. 176-177). For this study, it is assumed that the actual values of each n observations (X_i's) are available. Given this data, a continuous piecewise linear distribution function F can be specified by sorting the X_i's in the original data set into increasing order. With $X_{(i)}$ representing the i^{th} smallest of the X_i's so that $X_{(1)} \leq X_{(2)} \ldots \leq X_{(n)}$, the distribution can be defined as:

$$[1] \quad F(x) = \begin{cases} 0 & \text{if } x < X_{(1)} \\ \dfrac{i-1}{n-1} + \dfrac{x - X_{(i)}}{(n-1)(X_{(i+1)} - X_{(i)})} & \text{if } X_{(i)} \leq x < X_{(i+1)} \\ & \text{for } i=1,2,\ldots,n-1 \\ 1 & \text{if } X_{(n)} \leq x \end{cases}$$

To generate a continuous random number X that has the distribution function F (where F is continuous and strictly increasing when $0 < F(x) < 1$), the inverse of the function F, defined as F^{-1}, is taken. A uniform random number $U \sim U(0,1)$ is generated and X is set equal to $F^{-1}(U)$ (Law and Kelton, pp. 261-262).

For illustrative purposes, suppose we wish to generate random precipitation levels for the first quarter of the year and have the following sorted first quarter observations (inches of precipitation) for

1979 through 1988: $n = 10$ and $X_{(i)}$'s = {0, 2, 4, 5, 7, 8, 9, 10, 14, 20}. The distribution, F(x), is illustrated in Figure 1. To generate random variables from the continuous empirical distribution function F, a uniform random number U is first generated, where $U \sim U(0,1)$. Next, let $P = (n-1)U$ and let $I = \lfloor P \rfloor + 1$, where $\lfloor P \rfloor$ denotes the greatest integer that is less than or equal to the real number P. The generated random variable X is then equal to $X_{(I)} + (P - I + 1)(X_{(I+1)} - X_{(I)})$. In our example, if $U = .4$, then $P = 3.6$, $I = 4$, and $X_{(4)} = 5$, making $X = 5 + (3.6 - 4 + 1)(7 - 5)$ or 6.2. If a trend is present in the data or if one wishes to establish a trend in the generated random values, the original observations can be subtracted from their mean or trend and the deviates or residuals used as the X_i's in the generation of the distribution. After the random deviates or residuals have been generated, they can be added back to the mean or specified trend to obtain the generated random values.

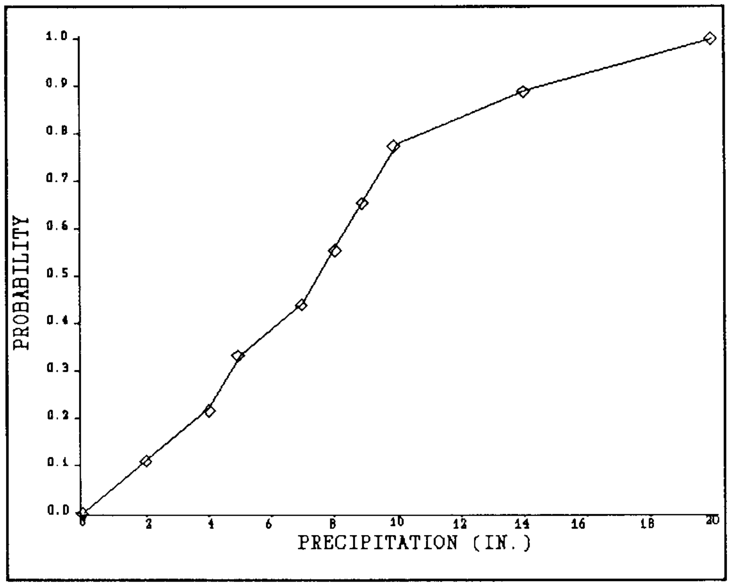

Figure 1. Empirical Distribution Function Developed from the Example Data.

A shortcoming of this formulation (especially for a small data set) is that while the endpoints in our example actually occurred 20 percent of the time in nature, they individually have a zero percent probability of occurrence under this specification. To overcome this criticism, the observation set can be expanded two observations by adding one unit to each of the endpoints. When random values are generated outside the original endpoints, they are accordingly reassigned the values of the original endpoints.

Disadvantages to using empirical distributions to generate the random input variables of interest include the fact that no random variables will be generated outside the observed range of the data set collected. This may not be a limiting problem if the historical data series covers a wide range of occurrences. Also, this anomaly may not be a disadvantage if the generated random variables are used to simulate a production relationship that was estimated from the historical data, since random variables beyond the historical range often cause problems for estimated equations.

When it is desirable to generate values outside the historical range, an "exponential tail" can be put on one or both ends of the empirical distribution (Law and Kelton, pp. 159-160, 177). The exponential distribution function (expo(β)) is defined as:

$$[2] \quad F(x) = \begin{cases} 1 - e^{-x/\beta} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where β is the mean and scale parameter, β^2 is the variance, and the range is $[0,\infty)$. To generate tails on the empirical distribution, the scale parameter β must first be specified. The higher β, the more extended the tail. A distribution function is shown in Figure 2 for a β of 1 and one of 2 (i.e., 1 and 2 inches of precipitation). Once the exponential distribution is called upon to generate a value, there is a 50 percent probability of extending the tail by .7 inches or less of precipitation and a 98 percent probability of extending the original endpoint by 3.9 inches of precipitation with the scale parameter set at 1. With a β of 2, a 50 and 98 percent probability exists of generating an additional 1.4 and 7.8 inches of precipitation, respectively, past the original endpoint.

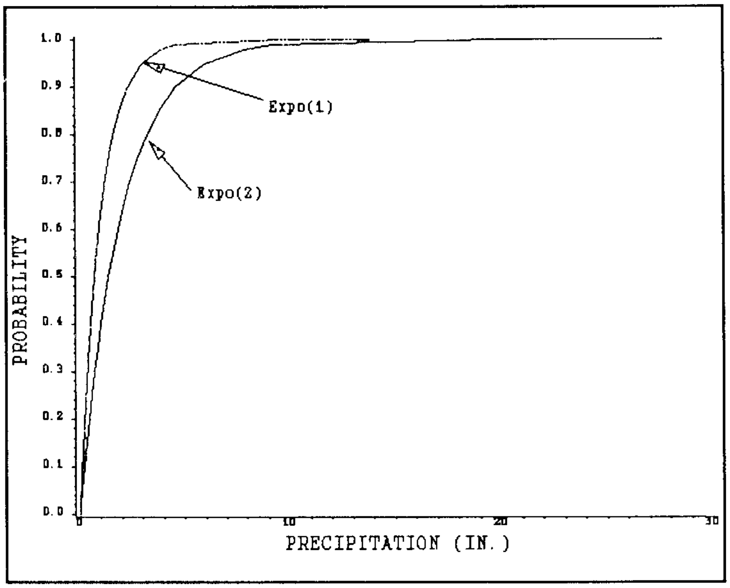

Figure 2. Expo(1) and Expo(2) Distribution Functions.

After a scale parameter has been specified, the probability of events occurring in the "exponential tails" must be specified. This can be accomplished by adding observations to the ends of the sorted empirical distribution function. The probability of a generated value falling between two observed data points on an empirical distribution [1/(n-1)] is dependent upon the number of total observed data points (n) in the distribution. Therefore, the more observations added to the endpoints, the higher the probability of generating observations outside the original data range. The value of the observations added to the endpoints is not important, since it is the number of observations added and not the value itself that determines the probability of occurrence. After the desired number of observations have been added, the empirical distributions are simulated as previously explained, and any values which fall outside the observed range are regenerated via the expo(β) distribution. The exponential distribution may also be truncated at some upper boundary since the tail is bounded by infinity.

__Temporal Correlation of Empirical Distributions__

In the previous section's example, precipitation levels were generated for the first quarter of the year. If minimum and maximum temperatures are also to be generated for the four quarters of the year, values for n = 12 different stochastic variables would need to be generated. If the values for each variable are generated independent of the others, the historical correlation among the variables will not be accounted for. The result could be that more variation is introduced into the system than actually exists. To account for the historical correlations of meteorological variables, the uniform random numbers (U) can be correlated before they are used to generate the random values. To accomplish this, a correlation matrix is developed using as data the deviations from the mean of the meteorological variables under consideration (assuming we are working with deviations from the mean). The correlation matrix (P) is then factored into a unique upper triangular matrix R so that (using matrix notation):

[3] $P = RR'$.

The elements for matrix R can be found by factoring the P matrix by the "square-root" method, i.e., taking the square-root of the matrix (Richardson and Condra, 1978). The R matrix (an n x n upper triangular matrix) is then multiplied by an n x 1 vector of independent standard normal deviates (W) to obtain an n x 1 vector of correlated standard normal deviates (C) as:

[4] $C = RW$.

This procedure is repeated with as many vectors of W as needed to generate the desired number of random values. By doing this, the historical correlations between the variables can be represented in a simulation model. The next step is to convert the correlated standard normal deviates (C) into a vector of uniformly distributed random numbers (U). This is accomplished by integrating the area under the standard normal probability density function from 0 to $c_i/2$ or:

[5] $u_i = 0.5 + \{0.5\ [2/\sqrt{\pi}\ {}_0\int^{(ci/2)} \exp(-t^2)dt]\}$.[1]

Since the transformation of the elements in vector C is a one-to-one transformation, the values in vector U are correlated in the same manner as the values in C. The n values in vector U are used to generate random values from the n respective cumulative distribution functions $F_i(X)$ representing the n meteorological variables.

This procedure can be used to correlate random variables assumed to be distributed beta, triangle or any other distribution which utilizes an inverse function (F^{-1}). Richardson and Condra (1978) have shown the procedure is a generalization of the method for correlating normally distributed random numbers described by Clements, et al. A shortcoming, however, of the procedure is that it does not allow intertemporal correlation of random variables.

[1]The FORTRAN statement for transforming the c_i elements of vector C into U is $u_i = 0.5 + \{0.5 * ERF(c_i/2)\}$, where ERF is an IBM supplied function for integrating the area under the standard normal probability density function.

Intertemporal Correlation of Empirical Distributions

Intertemporal correlation of random variables is of particular importance when simulating meteorological variables. Monthly precipitation and temperatures are not only correlated to one another within a year, but they are also correlated to those in the previous year. For example, assume quarterly precipitation data is to be stochastically generated and then aggregated for a production period beginning with Quarter 3 in the previous year and going through the third quarter in the current year (e.g., $Q^L_3 + Q^L_4 + Q_1 + Q_2 + Q_3$, with L signifying precipitation lagged one year). The problem of correlating these values centers around the fact that while Q_1, Q_2, Q_3, and Q_4 can be properly correlated using the procedure discussed above, Q^L_3 and Q^L_4 are not necessarily correlated to them.[2] One solution would be to include Q^L_3 and Q^L_4 in the variable set to be correlated and generate them as separate random variables. This is not an adequate solution because the precipitation generated in year t for Q_3 and Q_4 would not equal their values generated for the lagged Q_3 and Q_4 in year t+1. This problem of overlapping precipitation variables can be overcome by modifying the method used to obtain correlated empirical deviates.

The procedure to generate intertemporally correlated random variables utilizes the techniques above to correlate the random deviates except that the appropriate U is solved for $Q^L_3 + Q^L_4$ to assure identity in the

[2]Overlooking the correlation problem of lagged variables can be a serious omission because the correlation between years can be drastically different from the correlation existing within years. For example, data from the example used later in this paper reveal the correlation coefficient between February and December in year t is -0.01, while the correlation coefficient between February in year t and December in year t-1 is 0.411.

overlapping data. The 6 x 6 upper triangular correlation matrix of Q_1, Q_2, Q_3, Q_4, Q^L_3, and Q^L_4 would normally be factored into a unique upper right triangular matrix via the "square root method" and multiplied by a vector of random normal deviates d to obtain the vector of correlated random normal deviates c as:

$$[6] \quad \begin{bmatrix} c_{1t} \\ c_{2t} \\ c_{3t} \\ c_{4t} \\ c_{5t} \\ c_{6t} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} & r_{15} & r_{16} \\ & r_{22} & r_{23} & r_{24} & r_{25} & r_{26} \\ & & r_{33} & r_{34} & r_{35} & r_{36} \\ & & & r_{44} & r_{45} & r_{46} \\ & & & & r_{55} & r_{56} \\ & & & & & r_{66} \end{bmatrix} * \begin{bmatrix} d_{1t} \\ d_{2t} \\ d_{3t} \\ d_{4t} \\ d_{5t} \\ d_{6t} \end{bmatrix}$$

where t equals the current year.

The correlated random normal deviate used to calculate Q_{4t}, with t=1, would be calculated as:

$$[7] \quad c_{41} = \sum_{j=1}^{6} (r_{4j} * d_{j1}).$$

In year two, Q^L_{42} would be calculated using c_{62} as:

$$[8] \quad c_{62} = \sum_{j=1}^{6} (r_{6j} * d_{j2}).$$

Because $Q^L_{42} = Q_{41}$, it follows that c_{62} must equal c_{41}. Therefore, substituting c_{41} for c_{62} gives:

$$[9] \quad c_{41} = \sum_{j=1}^{6} (r_{6j} * d_{j2})$$

$$= r_{66} * d_{62}.$$

The random normal deviate d_{62} needed to assure that c_{62} is defined so $Q_{41} = Q^L_{42}$ can be determined by using c_{41} as:

[10] $d_{62} = c_{41}/r_{66}$.

Using this same logic, $Q_{31} = Q^L_{32}$ implies $c_{31} = c_{52}$. Solving for d_{52}, yields:

[11] $$c_{31} = \sum_{j=1}^{6} (r_{5j} * d_{j2})$$

$$= r_{55} * d_{52} + r_{56} * d_{62},$$

with:

[12] $d_{52} = (c_{31} - r_{56} * d_{62}) / r_{55}$.

Therefore, instead of all normal deviates (d) being randomly drawn, d_{5t} and d_{6t} are calculated conditional upon the correlated deviates obtained for c_{3t-1} and c_{4t-1}, thus forcing $Q^L_{42} = Q_{41}$ and $Q^L_{32} = Q_{31}$. At the beginning of each iteration (year 1), initial values must be given to d_{62} and d_{52} to start each iteration at the same point. This procedure can be extended to provide intra/intertemporal correlation of any random variable that can be represented by an inverse function (F^{-1}), an intratemporal correlation matrix, and a fixed intertemporal relationship.

An Empirical Example

Monthly precipitation and average minimum temperatures were determined by VanTassell, et al. (1987), to be crucial in evaluating production relationships for beef forage systems in the Texas Rolling Plains. Their forage production year began in August for year t-1 and extended to either March, June, August, or October in year t, depending upon the decision variable in question. For their production relationships to be used in a range management simulation model, monthly precipitation levels would be needed for August through December in year t-1 and January through December in year t. Average minimum monthly temperatures would also be needed for January through May in year t, bringing the total monthly meteorological variables to 22 (see Table 1 for a list of the variables).

To develop empirical probability distributions, a data series of monthly precipitation and minimum temperatures covering 1936 through 1985 was obtained from the National Oceanic and Atmospheric Administration. The data were first tested for long-run trends, but none were significant. Means for each of the 22 monthly temperature and precipitation variables were then computed for the period 1936 through 1985. From these, deviations from the means were calculated for all observations. A 22 x 22 correlation matrix was then calculated using the deviations from mean from which a factored upper right triangle of the correlation matrix using the square root method was obtained.

A FORTRAN computer model MESS (Meteorological Empirical Stochastic Simulator) was used to generate correlated random deviates and determine an inverse function (F^{-1}) in order to simulate possible 20-year weather patterns for the 22 key variables. MESS used as input the historical series of each variable and the factored correlation matrix (see Appendix B for a documentation of MESS).

Table 1. Meteorological Data Summary Statistics for the Historical Series, Simulated Series A (Distributed Empirically), Simulated Series B (Distributed Empirically with Stacked Endpoints), and Simulated Series C (Distributed Empirically with Exponentially Distributed Endpoints).

VARIABLE[a]	Historical Series[b]				Series A				Series B				Series C			
			Range				Range				Range				Range	
	MEAN	STD DEV	MIN	MAX	MEAN	STD DEV	MIN	MAX	MEAN	STD DEV	MIN	MAX	MEAN	STD DEV	MIN	MAX
JAN-P	1.08	1.15	0.00	5.94	1.07	1.06	0.00	5.92	1.10	1.22*	0.00	5.94	1.19*	1.39*	0.00	10.37
FEB-P	1.33	1.12	0.00	4.37	1.30	1.05*	0.00	4.35	1.34	1.12	0.00	4.37	1.35	1.20*	0.00	7.31
MAR-P	1.39	1.11	0.00	4.78	1.36	1.04*	0.00	4.70	1.41	1.13	0.00	4.78	1.42	1.21*	0.00	8.07
APR-P	2.57	2.22	0.15	9.69	2.51	2.09*	0.15	9.65	2.68*	2.34*	0.15	9.69	2.62	2.37*	0.15	13.68
MAY-P	3.84	2.58	0.15	10.88	3.89	2.46*	0.16	10.88	3.78	2.61	0.15	10.88	3.98*	2.71*	0.15	13.88
JUN-P	2.88	1.94	0.07	9.85	2.84	1.76*	0.07	9.83	2.89	1.99*	0.07	9.85	2.94	2.08*	0.07	14.18
JUL-P	1.92	1.83	0.00	7.38	1.89	1.73*	0.00	7.38	1.96	1.87	0.00	7.38	1.99	1.96*	0.00	10.63
AUG-P	2.14	2.16	0.00	12.07	2.07	1.87*	0.00	11.92	2.20	2.31*	0.00	12.07	2.25*	2.41*	0.00	16.38
SEP-P	3.21	2.95	0.00	10.68	3.12	2.78*	0.00	10.67	3.25	2.94	0.00	10.68	3.22	3.00	0.00	14.44
OCT-P	2.80	2.55	0.00	14.29	2.73	2.22*	0.00	14.01	2.91	2.84*	0.00	14.29	2.94*	2.87*	0.00	16.98
NOV-P	1.31	1.06	0.00	3.75	1.30	1.02*	0.00	3.75	1.31	1.07	0.00	3.75	1.34	1.12*	0.00	8.16
DEC-P	1.14	0.92	0.00	4.03	1.10	0.85*	0.00	3.98	1.15	0.92	0.00	4.03	1.15	1.00*	0.00	7.87
AUG-L	2.17	2.15	0.00	12.07	2.08	1.88*	0.00	11.92	2.18	2.29*	0.00	12.07	2.25	2.39*	0.00	15.72
SEP-L	3.24	2.95	0.00	10.68	3.15	2.79*	0.00	10.67	3.22	2.95	0.00	10.68	3.25	3.01	0.00	14.80
OCT-L	2.80	2.55	0.00	14.29	2.74	2.24*	0.00	14.01	2.91	2.85*	0.00	14.29	2.95*	2.91*	0.00	17.45
NOV-L	1.30	1.06	0.00	3.75	1.29	1.02*	0.00	3.75	1.31	1.06	0.00	3.75	1.34	1.14*	0.00	7.52
DEC-L	1.15	0.91	0.00	4.03	1.10*	0.85*	0.00	3.98	1.15	0.92	0.00	4.03	1.16	1.01*	0.00	8.25
JAN-T	28.49	4.17	21.38	37.71	28.54	3.91*	21.38	37.67	28.42	4.15	21.38	37.71	28.60	4.27	16.69	41.60
FEB-T	32.76	3.73	25.04	42.82	32.63	3.47*	25.05	42.79	32.91	3.79	25.04	42.82	32.67	3.91*	19.30	45.45
MAR-T	40.57	3.81	33.52	48.77	40.52	3.62*	33.53	48.76	40.61	3.87	33.52	48.77	40.55	3.98*	30.17	54.94
APR-T	50.79	2.94	45.43	57.87	50.83	3.77*	45.44	57.87	50.73	2.98	45.43	57.87	50.85	3.13*	39.88	63.42
MAY-T	59.42	2.54	53.77	66.81	59.44	2.31*	53.80	66.77	59.36	2.58	53.77	66.81	59.47	2.77*	46.84	73.05

*Significantly different from historical mean or standard deviation at a 5 percent level of significance.

[a]-P = monthly precipitation; -L = lagged monthly precipitation; -T = average minimum monthly temperatures.

[b]Historical series had 50 observations while the simulated series each had 2,000 observations.

The endpoints of the distributions were handled three ways. First, no additional observations were added to the historical endpoints (Series A). Second, one observation was added to each endpoint, with observations generated outside the historical range stacked on the original endpoints (Series B). Third, one observation was added to each endpoint with the lower tail stacked for the precipitation variables, the upper tail distributed expo(1) for the precipitation variables, and both tails distributed expo(1.5) for the temperature variables. Any observations drawn from the expo distributions that were greater than 4.5 inches of precipitation (0.01 percent probability of occurring) or 7°F (0.01 percent probability of occurring) were redrawn.

Means, standard deviations, and ranges are given in Table 1 for the historical and generated climatic variables simulated over a 20-year planning horizon for 100 iterations. For the most part, mean values and standard deviations are in line with their historical counterparts. Statistical significance of the difference between the historical and simulated means was tested by comparing a calculated t-statistic to tabular values. A Chi Square (χ^2) was also computed and compared to the appropriate value to test the difference between each set of standard deviations (Clements, et al.). We failed to reject the null hypothesis that the sample means were equal to the observed values at a 5 percent level of significance in all cases except for the December precipitation in Series A, April precipitation in Series B, and May, August, October, and October lagged precipitation in Series C. April precipitation (Series A) and August precipitation (Series C) were not significantly different from their historical mean at a 1 percent significance level.[3]

The standard deviations were not as accurately simulated in each series as were the means. This was somewhat expected because of the maneuvering which occurred with the endpoints. All standard deviations for Series A were less than those historically observed. One reason was that the endpoints were not stacked and were, therefore, observed in the simulation less frequently than in nature. While several of the standard deviations in Series A were observed to be quite close to their historical counterparts, the null hypothesis that the sample standard deviations were equal to the observed values was rejected in all cases at a 5 percent level of significance. For Series B, the stacking of endpoints appeared to extend the tails of the distribution sufficiently for several variables since the null hypothesis was rejected for only 7 (6 at a 1 percent level) of the 22 standard deviations. In all cases but one, standard deviations in Series B were either greater than or equal to their historical counterparts. The exponentially distributed variables in Series C overextended the standard deviations in all cases, resulting in the null hypothesis being rejected for 19 of the 22 standard deviations.

For all variables in which the endpoints were stacked and folded back upon their historical endpoint, the minimum and maximum values observed, by definition, were the same. For Series A, where the endpoints were not stacked, the minimum and maximum values obtained were identical to the historical range in most cases. In a few cases, the simulated minimum and

[3]Nicks and Harp, along with Richardson, use a 1 percent level of significance in testing their meteorological data-generating models.

maximum fell slightly below those observed. The ranges for the exponentially distributed endpoints were close, or identical, to the range imposed in the simulation.

Correlation coefficients computed for the historical and simulated meteorological data are reported in Tables A1, A2, and A3 in Appendix 1 for Series A, B, and C, respectively. The upper triangle of the correlation matrices contains the correlation coefficients for the historical data, while the lower triangle contains the correlations for the respective simulated series. Upon visual inspection, the correlation coefficients appear to be close to one another. A statistical procedure for testing the equality of the correlation coefficients described by Steel and Torrie (1960, pp. 188-193) was used to test the null hypothesis that the sample correlation coefficients were equal to the observed values. The tests reveal that at a 5 percent (1 percent) level of significance, the null hypothesis was rejected for 12 percent (6 percent) of the coefficients in Series A, 16 percent (10 percent) in Series B, and 17 percent (10 percent) in Series C. It appears that extending the endpoints may slightly deteriorate the historical correlations.

Cumulative distribution functions are shown in Figures 3 and 4 for September precipitation and January temperatures, respectively. These variables were chosen because of the variation exhibited in their historical series. For both variables, the historical distributions are closely mimicked by each simulated series. The most apparent difference comes at the endpoints, where a higher probability of observing a value at the upper and lower tails of the distribution is exhibited for both Series B and C because of their stacked endpoints. The exponential tails are

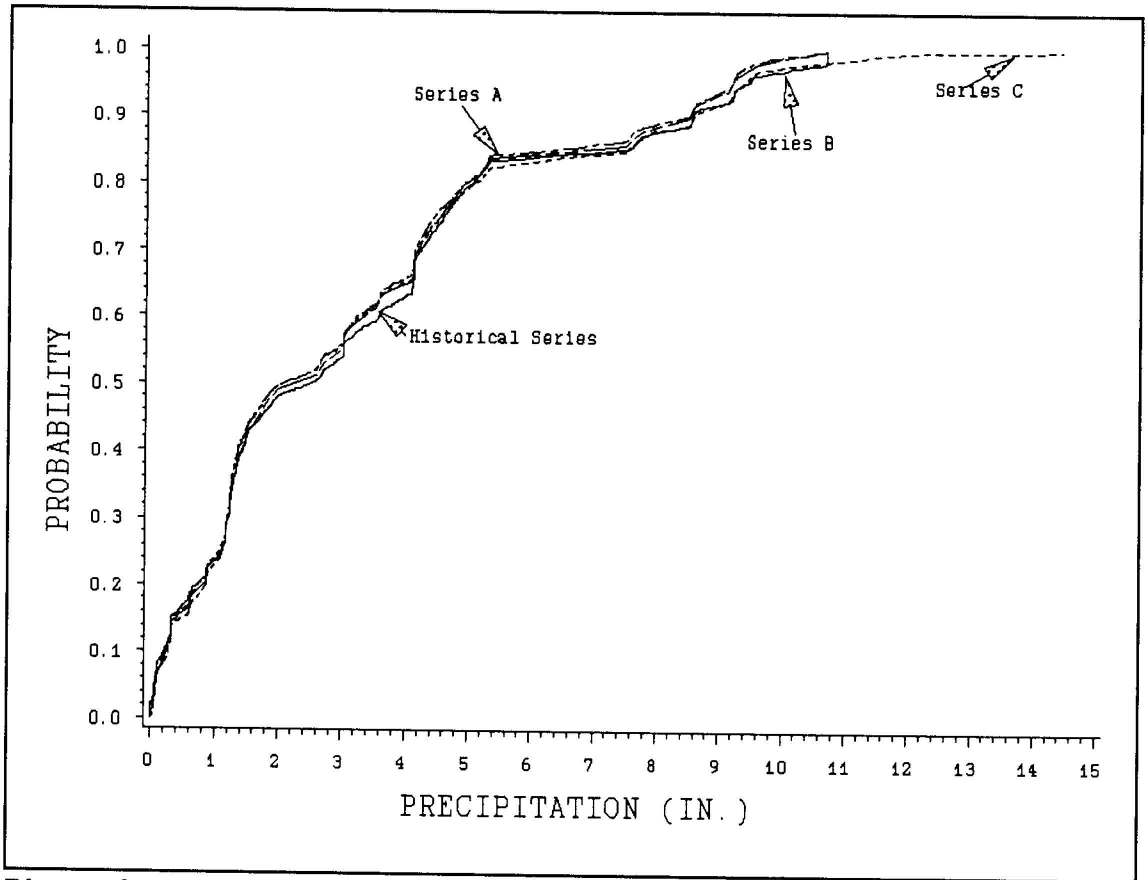

Figure 3. Cumulative Distribution Functions for Precipitation in September.

noticeable for Series C, and the abrupt ending of Series B is caused from stacking the endpoints. While not readily apparent, Series A exhibits a slightly lower probability of observing a value near the endpoints than has historically occurred.

Summary and Conclusions

Weather is a primary source of risk and uncertainty in the production of agricultural commodities. Incorporation of meteorological variables in simulation models requires the recreation of the stochastic relationships which underlie the basic meteorological process. Weather models have been

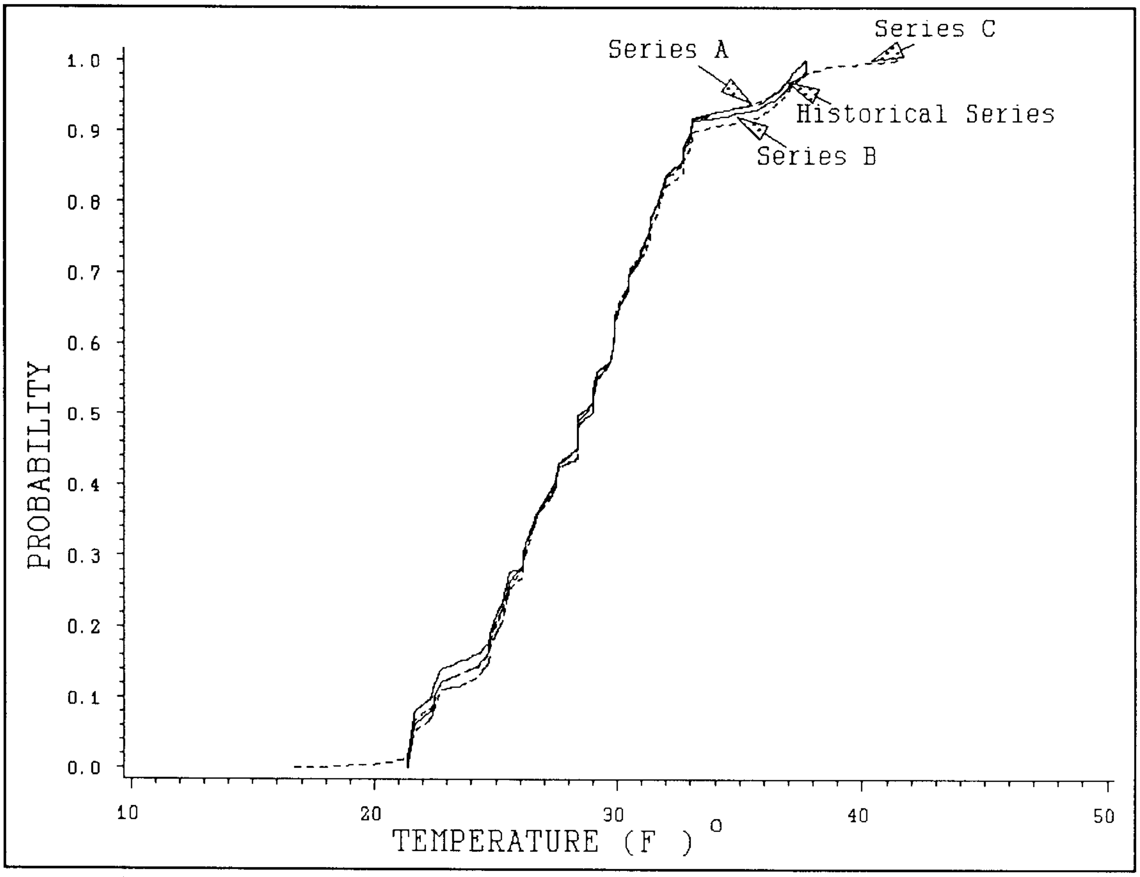

Figure 4. Cumulative Distribution Functions for Average Minimum Temperatures in January.

developed to simulate meteorological variables on a daily or more frequent basis. The methodology described in this paper stochastically simulates meteorological values on a more aggregated basis (e.g., monthly or quarterly) while closely maintaining the statistical properties observed in the historical series.

Empirical probability distributions were developed from monthly precipitation levels and minimum temperatures. To further aggregate precipitation levels into production periods which extend across years, the necessary lagged monthly observations were also generated. The correlation matrix of the meteorological variables in question was factored via the square root method and used to develop random deviates which maintain the historical correlations both within and between years while maintaining the identity between the corresponding lagged and previous years values.

The effect of three different methods for handling the endpoints of empirical distributions was also examined. In Series A, no additional observations were added to the historical endpoints; in Series B one observation was added to each endpoint, with observations generated outside the historical range stacked on the original endpoints; and in Series C, one observation was added to each endpoint with the lower tail stacked for the precipitation variables, the upper tail distributed expo(1) for the precipitation variables, and both tails distributed expo(1.5) for the temperature variables. Observations drawn from the expo distributions which were outside a specified range were truncated.

Statistical tests indicated that, for the most part, the model was capable of representing the statistical characteristics found in the observed data series. The simulated means were generally not significantly different from their historical counterparts for each of the three series. When the endpoints were not extended (Series A), the standard deviations were underestimated; and when the endpoints were extended and distributed exponentially, the standard deviations were overestimated. Extending the endpoints one observation and stacking all observations outside the historical range (Series B) appeared to best represent the historical standard deviation. Another alternative which may work for some data sets is to double the historical observations by imputing each observation twice, then extending the endpoints by one observation. This method would decrease the probability of generating observations outside the historical series while representing the minimum and maximum observations with a higher probability of occurrence than under the standard method where the endpoints are not stacked.

Correlation coefficients of the simulated series were not significantly different from their historical counterparts in the majority of cases. It did appear, though, that the practice of expanding endpoints slightly misrepresented the historical correlations more than when the endpoints were not expanded.

The procedure described is capable of generating a myriad of aggregated meteorological data, including precipitation levels, maximum and minimum temperatures, solar radiation levels, and relative humidity levels. The methods of generating meteorological data shown in this paper show promise for projecting environmental factors required for Monte Carlo simulation models developed to evaluate the performance of management strategies under different climatic conditions.

References

Clements, A. M., Jr., H. P. Mapp, Jr., and V. R. Eidman. A Procedure for Correlating Events in Farm Firm Simulation Models. Oklahoma State University Agr. Exp. Sta. Bull. No. T-131, Aug. 1971.

Larsen, G. A., and R. B. Pense. Stochastic Simulation of Daily Climate Data. USDA, SRS Staff Report No. AGES810831, 1981.

Law, A. M., and W. D. Kelton. Simulation Modeling and Analysis. New York: McGraw-Hill Book Co. 1982.

Lemieux, C. M., J. W. Richardson, and C. J. Nixon. "Federal Crop Insurance for Texas High Plains Producers: An Application of Whole-Farm Simulation." West. J. of Agr. Econ., 7(1982):141-54.

National Oceanic and Atmospheric Administration. Climatological Data. USDA, Vol 90., 1985.

Nicks, A. D., and J. F. Harp. "Stochastic Generation of Temperature and Solar Radiation Data." Journal of Hydrology, 48(1980):1-17.

Perry, G. M., M. E. Rister, J. W. Richardson, and D. J. Leatham. "The Effects of Equity Position, Credit Policy, and Capital Gains on Farm Survival." Agricultural Finance Review, 45(1985):58-71.

Richardson, C. W. "Stochastic Simulation of Daily Precipitation, Temperature, and Solar Radiation." Water Resources Research, 17,1(1981):182-90.

Richardson, J. W., and G. D. Condra. A General Procedure for Correlating Events in Simulation Models. Texas Agr. Exp. Sta., Department of Agricultural Economics (mimeo), 1978.

Richardson, J. W., and G. D. Condra. "Farm Size Evaluation in the El Paso Valley: A Survival/Success Approach." Amer. J. of Agr. Econ., 63(1981):430-37.

Richardson, J. W., and C. J. Nixon. "Producer's Preference for a Cotton Farmer Owned Reserve: An Application of Simulation and Stochastic Dominance." West. J. of Agr. Econ., (1982):123-32.

Steel, R. G. D., and J. H. Torrie. Principles and Procedures of Statistics. New York: McGraw-Hill, Inc., 1960.

VanTassell, L. W. "Risk Management by Livestock Producers: A Ranch Simulation in the Texas Rolling Plains." Ph.D. dissertation, Texas A&M University, 1987.

VanTassell, L. W., R. K. Heitschmidt, and J. R. Conner. "Modeling Variation in Range Calf Growth Under Conditions of Environmental Uncertainty." Journal of Range Management, 40,4(1987):310-14.

APPENDIX A

Table A1. Correlation Matrix of the Historical Series (Upper Triangle) and Simulated Series A[a] (Lower Triangle).

	JAN-P[b]	FEB-P	MAR-P	APR-P	MAY-P	JUN-P	JUL-P	AUG-P	SEP-P	OCT-P	NOV-P	DEC-P
JAN-P	1.000	0.115	0.159	-0.058	0.005	0.046	0.270	0.004	-0.078	-0.265	0.075	-0.183
FEB-P	0.122	1.000	0.283	0.278	0.128	0.215	0.068	0.080	-0.068	0.167	0.279	-0.011
MAR-P	0.133	0.238**	1.000	0.044	-0.211	0.123	-0.090	-0.027	-0.135	-0.072	0.157	-0.171
APR-P	-0.040	0.262	0.005	1.000	0.087	-0.067	-0.043	-0.018	0.100	0.361	-0.041	-0.205
MAY-P	0.002	0.141	-0.181	0.115	1.000	0.324	-0.083	-0.092	-0.091	0.164	0.178	-0.103
JUN-P	-0.019*	0.172**	0.098	-0.058	0.332	1.000	0.059	-0.040	-0.217	0.151	0.117	0.245
JUL-P	0.236	0.070	-0.052	-0.019	-0.101	0.028	1.000	-0.151	0.178	0.108	-0.141	-0.012
AUG-P	0.045	0.083	-0.033	-0.023	-0.075	-0.045	-0.107**	1.000	0.025	0.047	0.067	0.044
SEP-P	-0.035**	-0.083	-0.135	0.062	-0.117	-0.209	0.192	-0.000	1.000	-0.044	-0.299	-0.115
OCT-P	-0.239	0.131	-0.075	0.316**	0.152	0.123	0.092	0.035	-0.029	1.000	-0.158	0.179
NOV-P	0.079	0.288	0.113**	-0.010	0.233*	0.100	-0.133	0.072	-0.254**	-0.141	1.000	0.295
DEC-P	-0.175	-0.002	-0.178	-0.172	-0.080	0.228	-0.013	0.021	-0.082	0.166	0.234*	1.000
JAN-T	0.131	0.051	-0.023	-0.025	0.261	-0.111	0.009**	-0.127*	-0.192	0.181	-0.026	-0.031
FEB-T	-0.097	0.123	0.089	0.265	0.077	-0.033	0.084**	-0.255**	-0.121	0.196	0.103	-0.001
MAR-T	0.010	0.031	0.218	0.047	-0.173	0.097	0.014*	-0.136	-0.050	0.036	0.196	-0.164
APR-T	-0.196	-0.110	0.066	0.145	0.053	0.071	-0.075	0.039	-0.065	0.158	-0.021	-0.063
MAY-T	-0.086	-0.042	0.069	-0.042	0.025	0.144	-0.268	-0.075	-0.068	-0.022	0.040	0.014
AUG-L	-0.011	-0.033	0.034	0.148	0.107	0.050	0.025	0.097	0.119*	-0.001	-0.112	-0.022
SEP-L	-0.174	-0.268	0.112	0.021	-0.259	-0.275	-0.060	-0.030	0.008	-0.006	-0.168	-0.130
OCT-L	0.007	0.045	-0.014	0.190	-0.071	0.113	0.071	-0.131	-0.026	-0.065	0.044	0.178
NOV-L	-0.135	0.237*	0.299	0.127**	0.213	0.181	-0.110	-0.139	0.016	0.181	-0.004	-0.044
DEC-L	-0.034	0.356*	0.207	0.223**	0.072	-0.119	0.171	-0.224	-0.029	0.150	0.134	-0.084

Table 1A (continued)

	JAN-T	FEB-T	MAR-T	APR-T	MAY-T	AUG-L	SEP-L	OCT-L	NOV-L	DEC-L
JAN-P	0.148	-0.078	0.046	-0.201	-0.078	-0.033	-0.195	0.030	-0.120	-0.060
FEB-P	0.053	0.153	0.060	-0.110	-0.031	-0.074	-0.307	0.044	0.296	0.411
MAR-P	-0.012	0.097	0.223	0.084	0.043	0.076	0.096	-0.005	0.330	0.234
APR-P	-0.000	0.264	0.060	0.156	-0.046	0.120	0.002	0.200	0.177	0.267
MAY-P	0.277	0.063	-0.192	0.018	-0.014	0.116	-0.286	-0.086	0.226	0.081
JUN-P	-0.112	-0.022	0.127	0.046	0.140	0.045	-0.297	0.115	0.188	-0.103
JUL-P	0.059	0.130	0.066	-0.087	-0.254	-0.014	-0.056	0.057	-0.106	0.147
AUG-P	-0.191	-0.301	-0.142	0.065	-0.091	0.128	-0.026	-0.138	-0.157	-0.260
SEP-P	-0.200	-0.118	-0.090	-0.052	-0.039	0.056	0.028	-0.051	0.046	-0.030
OCT-P	0.212	0.209	0.064	0.162	-0.017	-0.008	-0.001	-0.070	0.197	0.144
NOV-P	-0.042	0.083	0.213	-0.061	0.030	-0.104	-0.209	0.032	0.014	0.163
DEC-P	-0.057	-0.023	-0.163	-0.030	0.055	-0.026	-0.127	0.146	-0.071	-0.107
JAN-T	1.000	0.334	0.090	-0.110	-0.011	-0.036	-0.253	-0.245	0.163	0.041
FEB-T	0.289**	1.000	0.217	-0.114	-0.051	-0.168	-0.147	-0.033	-0.048	0.086
MAR-T	0.072	0.188	1.000	0.020	0.009	0.086	0.053	-0.080	-0.101	-0.032
APR-T	-0.135	-0.095	0.057	1.000	0.076	-0.064	0.225	-0.035	0.033	0.052
MAY-T	0.046*	-0.053	0.043	0.098	1.000	-0.402	0.087	-0.201	0.055	-0.189
AUG-L	-0.037	-0.105*	0.094	-0.086	-0.351*	1.000	0.019	0.048	0.074	0.029
SEP-L	-0.217	-0.106	0.096	0.196	0.050	0.012	1.000	-0.043	-0.296	-0.125
OCT-L	-0.267	-0.003	-0.081	-0.055	-0.191	0.035	-0.028	1.000	-0.158	0.182
NOV-L	0.150	-0.051	-0.096	0.037	0.083	0.075	-0.257	-0.146	1.000	0.309
DEC-L	0.027	0.082	-0.013	0.097**	-0.177	0.021	-0.073*	0.165	0.223*	1.000

*,**Significantly different from historical correlation at a 5 percent and 1 percent level of significance, respectively.

[a]Series A was distributed empirically with standard endpoints.

[b]-P = monthly precipitation; -L = lagged monthly precipitation; -T = average minimum monthly temperatures.

Table A2. Correlation Matrix of the Historical Series (Upper Triangle) and Simulated Series B[a] (Lower Triangle).

	JAN-P[b]	FEB-P	MAR-P	APR-P	MAY-P	JUN-P	JUL-P	AUG-P	SEP-P	OCT-P	NOV-P	DEC-P
JAN-P	1.000	0.115	0.159	-0.058	0.005	0.046	0.270	0.004	-0.078	-0.265	0.075	-0.183
FEB-P	0.097	1.000	0.283	0.278	0.128	0.215	0.068	0.080	-0.068	0.167	0.279	-0.011
MAR-P	0.157	0.249	1.000	0.044	-0.211	0.123	-0.090	-0.027	-0.135	-0.072	0.157	-0.171
APR-P	-0.082	0.239	0.014	1.000	0.087	-0.067	-0.043	-0.018	0.100	0.361	-0.041	-0.205
MAY-P	-0.027	0.123	-0.178	0.108	1.000	0.324	-0.083	-0.092	-0.091	0.164	0.178	-0.103
JUN-P	0.031	0.161*	0.122	-0.059	0.312	1.000	0.059	-0.040	-0.217	0.151	0.117	0.245
JUL-P	0.222**	0.057	-0.048	-0.017	-0.105	0.029	1.000	-0.151	0.178	0.108	-0.141	-0.012
AUG-P	0.036	0.056	-0.030	-0.029	-0.118	-0.064	-0.076*	1.000	0.025	0.047	0.067	0.044
SEP-P	-0.060	-0.080	-0.116	0.057	-0.113	-0.203	0.188	0.041	1.000	-0.044	-0.299	-0.115
OCT-P	-0.214*	0.103*	-0.059	0.302*	0.133	0.132	0.069	0.024	-0.049	1.000	-0.158	0.179
NOV-P	0.055	0.280	0.132	0.003	0.208	0.089	-0.120	0.041	-0.243*	-0.148	1.000	0.295
DEC-P	-0.171	-0.005	-0.186	-0.137*	-0.084	0.231	0.002	0.024	-0.097	0.230*	0.205*	1.000
JAN-T	0.105	0.060	-0.031	-0.011	0.247	-0.101	0.013**	-0.169	-0.178	0.181	-0.051	-0.035
FEB-T	-0.124**	0.116	0.103	0.278	0.055	-0.017	0.091	-0.274	-0.121	0.190	0.088	0.024**
MAR-T	0.005	0.045	0.220	0.066	-0.174	0.080**	0.041	-0.122	-0.029*	0.024	0.201	-0.167
APR-T	-0.154**	-0.118	0.074	0.119	0.060	0.079	-0.100	0.003*	-0.075	0.135	-0.033	-0.047
MAY-T	-0.069	-0.044	0.067	-0.058	0.036**	0.145	-0.289	-0.086	-0.073	-0.000	0.036	-0.009*
AUG-L	-0.053	-0.059	0.057	0.105	0.093	0.005	0.030**	0.078**	0.082	0.022	-0.096	-0.014
SEP-L	-0.158	-0.266	0.110	0.011	-0.265	-0.243*	-0.053	0.003	0.040	0.002	-0.145*	-0.101
OCT-L	0.047	0.070	0.018	0.159	-0.079	0.108	0.066	-0.121	-0.051	-0.048	0.041	0.145
NOV-L	-0.126	0.231*	0.289**	0.118*	0.219	0.185	-0.143	-0.119	-0.003**	0.161	0.004	-0.045
DEC-L	-0.059	0.341*	0.174*	0.214*	0.078	-0.107	0.140	-0.197*	-0.020	0.104	0.158	-0.073

Table A2. (continued)

	JAN-T[b]	FEB-T	MAR-T	APR-T	MAY-T	AUG-L	SEP-L	OCT-L	NOV-L	DEC-L
JAN-P	0.148	-0.078	0.046	-0.201	-0.078	-0.033	-0.195	0.030	-0.120	-0.060
FEB-P	0.053	0.153	0.060	-0.110	-0.031	-0.074	-0.307	0.044	0.296	0.411
MAR-P	-0.012	0.097	0.223	0.084	0.043	0.076	0.096	-0.005	0.330	0.234
APR-P	-0.000	0.264	0.060	0.156	-0.046	0.120	0.002	0.200	0.177	0.267
MAY-P	0.277	0.063	-0.192	0.018	-0.014	0.116	-0.286	-0.086	0.226	0.081
JUN-P	-0.112	-0.022	0.127	0.046	0.140	0.045	-0.297	0.115	0.188	-0.103
JUL-P	0.059	0.130	0.066	-0.087	-0.254	-0.014	-0.056	0.057	-0.106	0.147
AUG-P	-0.191	-0.301	-0.142	0.065	-0.091	0.128	-0.026	-0.138	-0.157	-0.260
SEP-P	-0.200	-0.118	-0.090	-0.052	-0.039	0.056	0.028	-0.051	0.046	-0.030
OCT-P	0.212	0.209	0.064	0.162	-0.017	-0.008	-0.001	-0.070	0.197	0.144
NOV-P	-0.042	0.083	0.213	-0.061	0.030	-0.104	-0.209	0.032	0.014	0.163
DEC-P	-0.057	-0.023	-0.163	-0.030	0.055	-0.026	-0.127	0.146	-0.071	-0.107
JAN-T	1.000	0.334	0.090	-0.110	-0.011	-0.036	-0.253	-0.245	0.163	0.041
FEB-T	0.290**	1.000	0.217	-0.114	-0.051	-0.168	-0.147	-0.033	-0.048	0.086
MAR-T	0.079	0.190	1.000	0.020	0.009	0.086	0.053	-0.080	-0.101	-0.032
APR-T	-0.129	-0.097	0.050	1.000	0.076	-0.064	0.225	-0.035	0.033	0.052
MAY-T	0.045*	-0.051	0.034	0.105	1.000	-0.402	0.087	-0.201	0.055	-0.189
AUG-L	-0.045	-0.129	0.083	-0.068	-0.337*	1.000	0.019	0.048	0.074	0.029
SEP-L	-0.247	-0.116**	0.086	0.193	0.046	0.048	1.000	-0.043	-0.296	-0.125
OCT-L	-0.241	0.014	-0.066	-0.060	-0.182	0.028	-0.036	1.000	-0.158	0.182
NOV-L	0.140	-0.066	-0.122	0.026	0.092	0.050	-0.254**	-0.147	1.000	0.309
DEC-L	0.001	0.064	-0.036	0.079	-0.189	0.029	-0.090	0.230**	0.197*	1.000

*,**Significantly different from historical correlation at a 5 percent and 1 percent level of significance, respectively.

[a]Series B was distributed empirically with stacked endpoints.

[b]-P = monthly precipitation; -L = lagged monthly precipitation; -T = average minimum monthly temperatures.

Table A3. Correlation Matrix of the Historical Series (Upper Triangle) and Simulated Series C[a] (Lower Triangle).

	JAN-P[b]	FEB-P	MAR-P	APR-P	MAY-P	JUN-P	JUL-P	AUG-P	SEP-P	OCT-P	NOV-P	DEC-P
JAN-P	1.000	0.115	0.159	-0.058	0.005	0.046	0.270	0.004	-0.078	-0.265	0.075	-0.183
FEB-P	0.113	1.000	0.283	0.278	0.128	0.215	0.068	0.080	-0.068	0.167	0.279	-0.011
MAR-P	0.114**	0.217*	1.000	0.044	-0.211	0.123	-0.090	-0.027	-0.135	-0.072	0.157	-0.171
APR-P	-0.030	0.261	-0.002**	1.000	0.087	-0.067	-0.043	-0.018	0.100	0.361	-0.041	-0.205
MAY-P	0.001	0.138	-0.185	0.117	1.000	0.324	-0.083	-0.092	-0.091	0.164	0.178	-0.103
JUN-P	-0.023*	0.172**	0.084	-0.046	0.326	1.000	0.059	-0.040	-0.217	0.151	0.117	0.245
JUL-P	0.227**	0.068	-0.052	-0.016	-0.098	0.021	1.000	-0.151	0.178	0.108	-0.141	-0.012
AUG-P	0.045	0.085	-0.024	-0.020	-0.071	-0.029	-0.106**	1.000	0.025	0.047	0.067	0.044
SEP-P	-0.024*	-0.081**	-0.131	0.061*	-0.115	-0.202	0.187	-0.012	1.000	-0.044	-0.299	-0.115
OCT-P	-0.200*	0.121	-0.071	0.289	0.150	0.116	0.087	0.044	-0.015	1.000	-0.158	0.179
NOV-P	0.071	0.276	0.099*	-0.007	0.234*	0.098	-0.125	0.080	-0.247*	-0.128	1.000	0.295
DEC-P	-0.156	0.006	-0.162	-0.168	-0.076	0.215	-0.005	0.027	-0.075	0.140	0.225*	1.000
JAN-T	0.122	0.042	-0.022	-0.025	0.257	-0.114	0.006*	-0.119*	-0.184	0.171	-0.023	-0.031
FEB-T	-0.083	0.117	0.092	0.255	0.076	-0.022	0.083**	-0.235*	-0.119	0.180	0.104	-0.003
MAR-T	0.019	0.025	0.220	0.044	-0.176	0.097	0.015**	-0.126	-0.045**	0.025	0.191	-0.157
APR-T	-0.179	-0.104	0.064	0.137	0.057	0.073	-0.073	0.028	-0.065	0.145	-0.025	-0.069
MAY-T	-0.080	-0.041	0.062	-0.046	0.021	0.136	-0.268	-0.060	-0.068	-0.019	0.043	0.007**
AUG-L	-0.019	-0.028**	0.027**	0.151	0.104	0.056	0.018	0.085	0.115*	0.005	-0.106	-0.019
SEP-L	-0.155	-0.252*	0.113	0.018	-0.253	-0.265	-0.058	-0.032	0.004	-0.007	-0.168	-0.129
OCT-L	-0.007	0.041	-0.023	0.174	-0.070	0.104	0.072	-0.116	-0.028	-0.060	0.037	0.177
NOV-L	-0.128	0.226*	0.291	0.123*	0.202	0.170	-0.110	-0.123	0.019	0.172	-0.007	-0.039
DEC-L	-0.033	0.350*	0.193	0.229	0.079	-0.105	0.173	-0.207*	-0.037	0.145	0.137	-0.082

Table A3. (continued)

	JAN-T[b]	FEB-T	MAR-T	APR-T	MAY-T	AUG-L	SEP-L	OCT-L	NOV-L	DEC-L
JAN-P	0.148	-0.078	0.046	-0.201	-0.078	-0.033	-0.195	0.030	-0.120	-0.060
FEB-P	0.053	0.153	0.060	-0.110	-0.031	-0.074	-0.307	0.044	0.296	0.411
MAR-P	-0.012	0.097	0.223	0.084	0.043	0.076	0.096	-0.005	0.330	0.234
APR-P	-0.000	0.264	0.060	0.156	-0.046	0.120	0.002	0.200	0.177	0.267
MAY-P	0.277	0.063	-0.192	0.018	-0.014	0.116	-0.286	-0.086	0.226	0.081
JUN-P	-0.112	-0.022	0.127	0.046	0.140	0.045	-0.297	0.115	0.188	-0.103
JUL-P	0.059	0.130	0.066	-0.087	-0.254	-0.014	-0.056	0.057	-0.106	0.147
AUG-P	-0.191	-0.301	-0.142	0.065	-0.091	0.128	-0.026	-0.138	-0.157	-0.260
SEP-P	-0.200	-0.118	-0.090	-0.052	-0.039	0.056	0.028	-0.051	0.046	-0.030
OCT-P	0.212	0.209	0.064	0.162	-0.017	-0.008	-0.001	-0.070	0.197	0.144
NOV-P	-0.042	0.083	0.213	-0.061	0.030	-0.104	-0.209	0.032	0.014	0.163
DEC-P	-0.057	-0.023	-0.163	-0.030	0.055	-0.026	-0.127	0.146	-0.071	-0.107
JAN-T	1.000	0.334	0.090	-0.110	-0.011	-0.036	-0.253	-0.245	0.163	0.041
FEB-T	0.288**	1.000	0.217	-0.114	-0.051	-0.168	-0.147	-0.033	-0.048	0.086
MAR-T	0.076	0.192	1.000	0.020	0.009	0.086	0.053	-0.080	-0.101	-0.032
APR-T	-0.129	-0.089	0.059	1.000	0.076	-0.064	0.225	-0.035	0.033	0.052
MAY-T	0.047*	-0.047	0.039	0.094	1.000	-0.402	0.087	-0.201	0.055	-0.189
AUG-L	-0.034	-0.105*	0.085	-0.081	-0.337*	1.000	0.019	0.048	0.074	0.029
SEP-L	-0.207**	-0.103**	0.095	0.189	0.058	-0.002	1.000	-0.043	-0.296	-0.125
OCT-L	-0.249	0.002	-0.073	-0.052	-0.182	0.042	-0.017	1.000	-0.158	0.182
NOV-L	0.149	-0.056	-0.092	0.023	0.079	0.085	-0.247*	-0.134	1.000	0.309
DEC-L	0.041	0.084	-0.004	0.095	-0.179	0.026	-0.070*	0.137**	0.215*	1.000

*,**Significantly different from historical correlation at a 5 percent and 1 percent level of significance, respectively.

[a]Series C was distributed empirically with exponentially distributed endpoints.

[b]-P = monthly precipitation; -L = lagged monthly precipitation; -T = average minimum monthly temperatures.

APPENDIX B

Description of the Computer Program

The computer program MESS (Meteorological Empirical Stochastic Simulator) consists of approximately 700 lines of source statements. It is a stochastic simulation model written in FORTRAN. The program reads in a historical data series, a factored correlation matrix, and several control parameters. The simulation model generates a stochastic series of meteorological variables which are distributed multivariate empirical. The correlations which historically existed in each series are maintained via generating random numbers utilizing the factored correlation matrix. While illustrated and originally developed for meteorological data, the model is capable of generating empirical distributions from any historical data set.

Main Body of the Model

The main body of the model reads in each historical data series and control parameters, calls various subroutines, accumulates the simulated series into various production periods, and prints output tables. The model can presently accommodate up to 36 variables, made up of any number of historical observations. After reading the factored correlation matrix, the model calls Subroutine FACTOR, where INUM correlated uniform random deviates are created (INUM = number of years in planning horizon * number of iterations). The model then calls Subroutine STATS and obtains the historical average of each variable from which deviations from the mean are created. These deviations, along with the uniform random deviates created in Subroutine FACTOR, are sent to Subroutine EMP, where a series of INUM deviates from the mean are empirically simulated. These deviates are then added back to their respective means before Subroutine STATS is again called to perform a statistical analysis on the simulated data. This simulated series can either be written out to disk (accessed on unit 12) or a hard copy can be obtained. If desired, the computer program will then accumulate the simulated series into a maximum of 12 production periods (maximum of 18 variables summed into one production period), perform a statistical analysis on the accumulated variables, and provide a hard copy and/or disk file (accessed on unit 15) of this data. An option is available which will maintain a specific upper and/or lower bound on each accumulated production series. The main body will also read in variables which have already been simulated and accumulate these into production periods.

Subroutine FACTOR

Subroutine FACTOR utilizes the factored correlation matrix to generate a series of correlated uniform random deviates for each variable. It creates these over two iterations--one for the number of years in the proposed planning horizon and one for the number of iterations or planning horizons to be simulated--to give a total of INUM uniform random deviates. The user may request each planning horizon to start at a different random point or to start with the same random point obtained in iteration 1; or the user may provide specific random points so that each variable will begin each planning horizon at a specified value.

When the appropriate data are provided, Subroutine FACTOR will solve for the necessary random numbers needed to insure identity of lagged variables which are to be intertemporally correlated. For this to be accomplished, the historical data set must be arranged such that the unlagged variables are ordered first, with the lagged variables ordered last. The unlagged variables that correspond to the lagged counterparts must be in the same order (and must be side by side) as their lagged counterparts. For example, if variables A, B, C, and D were to be simulated along with A and B lagged one year (designated A^L and B^L), the historical series would be input in the order of A_t, B_t, C_t, D_t, A^L_t, B^L_t, where $A^L_t = A_{t-1}$ and $B^L_t = B_{t-1}$.

Subroutine EMP

Subroutine EMP creates an empirical distribution for each historical series and utilizes the correlated uniform random deviates from Subroutine FACTOR to generate the simulated series. Each endpoint for each variable can be extended by as many observations as desired. Values generated outside the historical range can then either be stacked on the endpoints or distributed expo(β), where β is the scale parameter and is given by the user. For endpoints distributed expo(β), a truncation point can be designated and all values beyond that point can be stacked or redistributed.

Subroutine STATS

Subroutine STATS provides basic statistics for each variable, consisting of the mean, variance, standard deviation, coefficient of variation, minimum, and maximum. A relative frequency distribution and/or cumulative frequency distribution may also be requested from this subroutine.

Coding Instructions for All Input Data

The first card contains "on/off switches" which control the operations of the model. The analyst must also provide the characteristics of the historical series, the period of time to be simulated, etc. The format followed is (20I1,8I4), with the parameters defined as follows (* implies the parameter can be 0 or 1, with 0 = no and 1 = yes):

Column	Parameter	Description
1	IP(1)*	Read in data that is already generated and accumulate into production periods.
2	IP(2)*	Generate data, but do not accumulate into production periods.
3	IP(3)*	Print historical data.
4	IP(4)*	Print correlation matrix.
5	IP(5)*	Print the uniform random correlated deviates.
6	IP(6)*	Print the generated empirical series.
7	IP(7)*	Print the generated accumulated series.
8	IP(8)*	Write the generated empirical series to unit 12.
9	IP(9)*	Write the generated accumulated series to unit 15.
10	IP(10)*	Print basic statistics for the historical data.
11	IP(11)*	Print relative frequency distributions for the historical series.
12	IP(12)*	Print cumulative frequency distributions for the historical series.
13	IP(13)*	Call Subroutine STATS for the generated empirical series.
14	IP(14)*	Print basic statistics for the generated empirical series.
15	IP(15)*	Print relative frequency distributions for the generated empirical series.
16	IP(16)*	Print cumulative frequency distributions for the generated empirical series.
17	IP(17)*	Call Subroutine STATS for the accumulated series.
18	IP(18)*	Print basic statistics for the accumulated series.
19	IP(19)*	Print relative frequency distributions for the accumulated series.
20	IP(20)*	Print cumulative frequency distributions for the accumulated series.
21-24	NYR	Number of years in the planning horizon.
25-28	ITER	Number of iterations desired.
29-32	NROW	Number of variables to be simulated.
33-36	NCOL	Number of historical observations.
37-40	NQ	Number of accumulated series desired.
41-44	NLAG	Number of variables which are lagged.

Column	Parameter	Description
45-48	NFROM	Row number of the first unlagged variable which will have a corresponding lagged variable.
49-52	NSAME*	Utilize the same random number for the first year of each planning horizon.

The next NROW cards consist of parameters describing the treatment of endpoints for each of the NROW variables to be simulated. The format followed is (A6,7I2,4F6.2), with I = 1,...,NROW variables to be simulated and the parameters defined as follows (* implies the parameter can be 0 or 1, with 0 = no and 1 = yes):

Column	Parameter	Description
1-6	TNAME(I)	Name of the variable to be simulated.
7-8	KE(I,1)*	Expand at least one endpoint.
8-9	KE(I,2)	Number of observations to expand lower endpoint.
10-11	KE(I,3)	Number of observations to expand upper endpoint.
12-13	KE(I,4)	Stack lower outliers on endpoint (0) or distribute exponentially (1).
14-15	KE(I,5)	Stack upper outliers on endpoint (0) or distribute exponentially (1).
16-17	KE(I,6)	Stack lower tail outliers from expo distribution on E(I,3).
18-19	KE(I,7)	Stack upper tail outliers from expo distribution on E(I,4).
20-25	E(I,1)	Scale (β) parameter for lower tail exponential function.
26-31	E(I,2)	Scale (β) parameter for upper tail exponential function.
32-37	E(I,3)	Minimum value to be drawn (truncation value) for the lower exponential tail.
38-43	E(I,4)	Maximum value to be drawn (truncation value) for the upper exponential tail.

The next NQ cards consist of parameters describing the formation of each of the NQ accumulated series. The format followed is (A4,2I2,2F7.0,19I3), with I = 1,...,NQ new series to be created and the parameters defined as follows (* implies the parameter can be 0 or 1, with 0 = no and 1 = yes):

Column	Parameter	Description
1-4	Type(I)	Name of the accumulated series.
5-6	IL(I)*	Put a lower bound on the accumulation.
7-8	IU(I)*	Put an upper bound on the accumulation.
9-15	B1(I)	Lower bound.
16-22	B2(I)	Upper bound.
23-25	L1(I)	Number of simulated variables to be summed.
26-79	L2(I,L1(I))	Row numbers of each variable included in the summation.

The next NCOL cards contain the historical observations (or generated observations if only an accumulation into production periods is to take place) of the NROW variables to be simulated. The format followed is (12A6/12F6.0). Up to 36 variables can be included, but only 12 can be put on one card. After all observations have been included for the first 1 to 12 variables, the next 1 to 12 can be input, and so on, up to 36 variables. The first card for each set of variables should contain the names (12A6) of the variables to follow.

The next set of cards contain the factored correlation matrix (P1(NNROW,NNROW)). The first card will contain the parameters M and NNROW (format = (2I2)) which give the number of stacks and the number of variables to be simulated, respectively. To be read in, the correlation matrix must be broken into M stacks, where one stack consists of up to six columns of the NNROW x NNROW factored correlation matrix. Up to six stacks can be read in, making a total of 36 variables for the simulation. The format for the each stack is (A9,6F11.0), with the first nine columns of each card giving the name of that particular row (IDEN(NNROW)).

The following is an example of all input data previously described:

```
01110001110111011101  20 100  22  50   9   5   8   0   0  40
  JANR 1 1 1 0 1 0 1 01.00 01.00 00.00 04.50
  FEBR 1 1 1 0 1 0 1 01.00 01.00 00.00 04.50
  MARR 1 1 1 0 1 0 1 01.00 01.00 00.00 04.50
  APRR 1 1 1 0 1 0 1 01.00 01.00 00.00 04.50
  MAYR 1 1 1 0 1 0 1 01.00 01.00 00.00 04.50
  JUNR 1 1 1 0 1 0 1 01.00 01.00 00.00 04.50
  JULR 1 1 1 0 1 0 1 01.00 01.00 00.00 04.50
  AUGR 1 1 1 0 1 0 1 01.00 01.00 00.00 04.50
  SEPR 1 1 1 0 1 0 1 01.00 01.00 00.00 04.50
  OCTR 1 1 1 0 1 0 1 01.00 01.00 00.00 04.50
  NOVR 1 1 1 0 1 0 1 01.00 01.00 00.00 04.50
  DECR 1 1 1 0 1 0 1 01.00 01.00 00.00 04.50
  JANN 1 1 1 1 1 1 1 01.55 01.55 07.00 07.00
  FEBN 1 1 1 1 1 1 1 01.55 01.55 07.00 07.00
  MARN 1 1 1 1 1 1 1 01.55 01.55 07.00 07.00
  APRN 1 1 1 1 1 1 1 01.55 01.55 07.00 07.00
  MAYN 1 1 1 1 1 1 1 01.55 01.55 07.00 07.00
  AUGL 1 1 1 0 1 0 1 01.00 01.00 00.00 04.50
  SEPL 1 1 1 0 1 0 1 01.00 01.00 00.00 04.50
  OCTL 1 1 1 0 1 0 1 01.00 01.00 00.00 04.50
  NOVL 1 1 1 0 1 0 1 01.00 01.00 00.00 04.50
  DECL 1 1 1 0 1 0 1 01.00 01.00 00.00 04.50
  A8 0 0 999.99 999.99  8 18 19 20 21 22  1  2  3
 A11 0 0 999.99 999.99 11 18 19 20 21 22  1  2  3  4  5  6
 A13 0 0 999.99 999.99 13 18 19 20 21 22  1  2  3  4  5  6  7  8
 A15 0 0 999.99 999.99 15 18 19 20 21 22  1  2  3  4  5  6  7  8  9 10
  F7 0 0 999.99 999.99  7  2  3  4  5  6  7  8
  J4 0 0 999.99 999.99  4 13 14 15 16
  J5 0 0 999.99 999.99  5 13 14 15 16 17
SEP5 0 0 999.99 999.99  5 19 20 21 22  1
APR3 0 0 999.99 999.99  3  4  5  6
 JANR  FEBR  MARR  APRR  MAYR  JUNR  JULR  AUGR  SEPR  OCTR  NOVR  DECR
 1.08  1.33  0.45  2.06  7.34  0.36  3.29  0.00  9.24  2.96  0.52  0.55
 0.37  0.00  3.37  0.54  0.86  3.15  0.51  2.23  1.52  2.98  0.72  1.52
 2.11  1.96  4.78  0.59  2.30  2.27  1.79  0.00  0.32  0.05  1.35  0.07
   .     .     .     .     .     .     .     .     .     .     .     .
   .     .     .     .     .     .     .     .     .     .     .     .
   .     .     .     .     .     .     .     .     .     .     .     .
 1.35  1.17  1.91  1.30  6.33  4.13  1.66  1.33  0.07  3.96  1.65  0.57
 0.40  1.81  0.19  0.35  0.48  3.99  2.50  3.97  1.45  3.83  3.10  4.03
 0.45  4.37  3.13  8.16  1.40  4.87  1.75  0.56  1.98  2.93  1.80  0.23
 JANN  FEBN  MARN  APRN  MAYN  AUGL  SEPL  OCTL  NOVL  DECL
28.39 32.82 44.90 48.20 59.87  2.14  3.21  2.80  1.31  1.14
26.32 33.54 38.65 50.33 61.45  0.00  9.24  2.96  0.52  0.55
32.69 40.32 48.77 51.20 59.39  2.23  1.52  2.98  0.72  1.52

   .     .     .     .     .     .     .     .     .     .
   .     .     .     .     .     .     .     .     .     .
   .     .     .     .     .     .     .     .     .     .
```

```
 29.68 33.07 40.58 46.03 55.26  0.97  1.16  1.11  2.04  1.62
 24.98 32.18 39.74 46.94 59.72  1.33  0.07  3.96  1.65  0.57
 24.26 30.56 45.50 53.24 60.44  3.97  1.45  3.83  3.10  4.03
4 22
ROW1      0.755757  0.0466075   0.268501   0.138765 -0.0979854  0.0575522
ROW2             0   0.741855   0.129029   0.153954 -0.0467894   0.172063
ROW3             0          0   0.707013  -0.121044  -0.331974   0.130966
 .               .          .          .          .          .          .
 .               .          .          .          .          .          .
 .               .          .          .          .          .          .
ROW20            0          0          0          0          0          0
ROW21            0          0          0          0          0          0
ROW22            0          0          0          0          0          0
ROW1      0.290542 -0.0333912 -0.0579434  -0.130632   0.118505  -0.221032
ROW2     0.0652246   0.211237 -0.0109694   0.150657   0.142511  0.0203264
ROW3     0.0160715   0.186943  -0.109731  -0.151985   0.161662 -0.0791264
 .               .          .          .          .          .          .
 .               .          .          .          .          .          .
 .               .          .          .          .          .          .
ROW20            0          0          0          0          0          0
ROW21            0          0          0          0          0          0
ROW22            0          0          0          0          0          0
ROW1      0.160057  -0.165677  0.0478675  -0.139797 -0.0667243 -0.0131439
ROW2    -0.0831344   0.053537   0.106374 -0.0910455  0.0118219 -0.0883585
ROW3    -0.0637581  0.0969564   0.251678  0.0178241     0.0631  0.0395569
 .               .          .          .          .          .          .
 .               .          .          .          .          .          .
 .               .          .          .          .          .          .
ROW20            0          0          0          0          0          0
ROW21            0          0          0          0          0          0
ROW22            0          0          0          0          0          0
ROW1     -0.242134  0.0177096  -0.106996 -0.0602449
ROW2     -0.216974 0.00971292   0.177473   0.411208
ROW3      0.210377  0.0140054   0.271007   0.234116
 .               .          .          .          .          .          .
 .               .          .          .          .          .          .
 .               .          .          .          .          .          .
ROW20            0   0.957179  -0.225439    0.18162
ROW21            0          0   0.951167   0.308675
ROW22            0          0          0          1
```