A STOCHASTIC MODEL FOR DAILY CLIMATE

by

ANABELA de GUSMÃO BRANDÃO

THESIS

Submitted in fulfilment of the requirements

for the degree of

MASTER OF SCIENCE

In the department of

Mathematical Statistics

University of Cape Town

SUPERVISOR: Professor W Zucchini

March 1987

ACKNOWLEDGEMENTS

I wish to express my immense gratitude to:

(i) My supervisor, Professor Walter Zucchini for bringing this research topic to my attention as well as for his encouragement and guidance throughout this study.

(ii) Mr J Myburgh of the Winter Rainfall Region, Department of Agriculture and Water Affairs for supplying the data.

(iii) Mrs Felicite Train for her immaculate typing of this thesis.

(iv) The CSIR Foundation for Research Development for their financial assistance.

(v) To my parents and sister for their moral and financial assistance.

ABSTRACT

This thesis describes the results of a study to establish whether climate variables could be usefully modelled on a daily basis. Three stochastic models are considered for the description of daily climate sequences, which can then be used to generate artificial sequences. The climate variables under consideration are rainfall, maximum and minimum temperature, evaporation, sunshine duration, windrun and maximum and minimum humidity. A simple Markov chain-Weibull model is proposed to model rainfall. Three multivariate models (one proposed by Richardson (1981), two new) are suggested for modelling the remaining climate variables. The model parameters are allowed to vary seasonally, while the error term is assumed to follow an autoregressive process. The models were validated and their general performance was found to be satisfactory. Some weaknesses were identified and are discussed. The main conclusion of this study is that daily climate sequences can indeed be usefully described by means of stochastic models.

TABLE OF CONTENTS

PAGE

PAGE

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

1. INTRODUCTION

Climate is a critical factor in determining the variety and abundance of vegetation and animal life in a region. It also imposes limits on the range of human activities which can economically take place. It is therefore not surprising that various aspects of climate such as precipitation, temperature, solar radiation, humidity, windspeed and many others are recorded regularly throughout the world.

The purpose of these measurements is to build up our knowledge of the behaviour of climate and thereby to help us determine which activities are feasible and how these can be most efficiently carried out.

The main objective of this thesis is to demonstrate that climate can be usefully described by means of a stochastic model. Such models provide a concise description of the patterns that exist in the various components of climate; they also quantify the variability of deviations from the "typical" patterns. They can be used to generate artificial climate sequences of arbitrary length which can for example be used as "inputs" to crop growth models. These can then be used to determine the distribution of yield, the probability of crop failure due to adverse weather conditions, optimal planting dates and so on. Agriculture is an important but by no means the only activity in which climate plays a vital role.

For many purposes, artificial climate sequences generated by a good stochastic model are more useful than the original historical record.

Firstly, they are free of the typical imperfections, such as incorrect recordings and missing observations, which are especially numerous in historical climate records. Secondly, the historical records available are often quite short and therefore only reflect a small fraction of the different climate sequences which could occur. It is sometimes argued that, as the parameters of a stochastic model have to be estimated from the historical record, the artificial sequences generated by the model are no more than complicated extrapolations of the historical record. This conclusion is false. A model contains more than the information which can be extracted from a single historical record. It contains information in the form of assumptions about climate which are based on our general knowledge about the behaviour of climate derived from observations at other locations and from theory. For example, it is reasonable to assume that certain average properties of climate variables are periodic and vary smoothly with time. Such assumptions give climate models structure which may not be evident in a single short historical record.

Three models for describing climate on a daily basis are investigated in this thesis. The first was proposed by Richardson (1981) and the remaining two are new. It is demonstrated that the models are capable of usefully describing the joint behaviour of seven aspects of climate, namely precipitation, maximum and minimum temperature, evaporation, sunshine duration, windrun and maximum and minimum humidity. The models were fitted to the historical record at Elsenburg and their performance and relative merits were assessed.

The thesis is structured as follows: The preliminary statistical analysis of the data is described in Chapter 2. This includes a description of the data, the types of difficulties encountered in detecting and dealing with faults in the data and the statistics computed to identify the structure present in the climate sequence.

Chapter 3 gives a theoretical description of the three climate models which were investigated and of the methods used to estimate the model parameters. Details on the implementation of the model to the historical record at Elsenburg are given in Chapter 4. The algorithms for implementing the theory are listed in Chapter 5. These include algorithms for generating artificial climate sequences. Extensive tests were performed on the fitted models in order to assess their performance in preserving the important properties of climate sequences. The results of this model validation investigation are summarized in Chapter 6. A brief summary of the study and the main conclusions reached are given in Chapter 7.

CHAPTER 2

2. THE DATA SET AND PRELIMINARY ANALYSIS

It is common knowledge that there is no such thing as a "clean" data set and that the number of "bugs" usually increases in proportion with the size of the data set. The data set considered in this study is no exception to this.

Firstly, a quite high proportion of the observations are missing. Although missing observations are relatively easy to detect, they lead to complications in the analysis. In particular, the multivariate time series models considered here require simultaneous observations of all the variables, and secondly, the serial correlation structure in the series does not allow one to simply discard observations as one could do if the observations were serially independently distributed.

Incorrect readings (or incorrectly recorded readings) are often quite difficult to detect, especially if the values fall within the feasible range of the variable under consideration. This problem which also occurs in the data set considered here is particularly difficult to deal with satisfactorily.

This chapter describes the data set used, some of the problems encountered and the method used to overcome them. At the end some preliminary analyses which were performed on the data set for initial model identification are discussed.

Since this study is intended to investigate the feasibility of modelling climate variables on a daily basis and since the preparation of the data for analysis is an extremely laborious and time consuming task, the climate record at only one station was investigated. The station chosen for this purpose was Elsenburg in the Cape Province (Latitude 33° 51', Longitude 18° 50').

For this station, daily climate readings are available for the years 1940 to 1984. On closer inspection though, the data for years 1950 through to 1957 were found to be missing. Within the years with records available, apart from the occasional missing values, whole months were found to be missing as well.

As one is dealing with a multivariate time series (as will be discussed later), it is required to have simultaneous observations for all variables. Table 2.1 shows the years for which data is available for each variable. The only years for which data is available for all of the variables simultaneously is 1978 to 1984. However, some variables only had observations starting from May in 1978, therefore the model parameters were estimated using the historical observations recorded in the years 1979 to 1984. The only exception was rainfall. Since rainfall was modelled independently of the other variables, the records from 1978 to 1984 were used.

TABLE 2.1: OBSERVATIONS AVAILABLE FOR EACH VARIABLE

Variable	Years Available	
Rainfall	1940 - 1950	1972 - 1984
Maximum Temperature	1940 - 1950	1972 - 1984
Minimum Temperature	1940 - 1950	1972 - 1984
Evaporation	1957 - 1972	1978 - 1984
Sunshine Duration	1972 - 1984	
Windrun	1978 - 1984	
Maximum Humidity	1972 - 1984	
Minimum Humidity	1972 - 1984	

Climate data is available for the following variables:

- Rainfall (mm)
- Maximum temperature (°C)
- Minimum temperature (°C)
- Evaporation (mm)
- Sunshine duration (hours)
- Windrun (km/day)
- Maximum humidity (%)
- Minimum humidity (%).

The unit of measurement for each variable is shown above in brackets following the variable name.

It is important to note that the readings were recorded by multiplying each value by ten, i.e. a record of 10.2 is given as 102. This convention was used throughout the study and so all results are given using this convention. This does not affect the generation of climate sequences, one merely obtains sequences which can be easily converted to the original units by dividing by ten.

There is always a possibility of readings being recorded incorrectly. The type of recording error which can easily be spotted is when the value recorded lies outside the permissible range, for example a recorded value for sunshine duration of 25 hours. When the time series were checked for values lying outside the permissible ranges, the variables maximum and minimum humidity had such recordings. In addition, the variable evaporation had some high values which were probably incorrect but this could not be established with certainty. Preceding and succeeding values of evaporation were examined to see if they could give an indication whether or not these values should be considered as outliers. The observations did not support such a high evaporation reading. Barry and Charley (1968) state that evaporation can be expressed by:

(a) duration of sunshine
(b) mean air temperature
(c) mean air humidity
(d) mean wind speed.

Therefore, the values of other variables observed at the same time were also examined. For example, one would expect to see an increase of evaporation with an increase of sunshine duration. The observations across the variables at these times did not show these expected patterns. Considering that all these high values were exactly the same in all cases and that they resembled the numbers that were found to be outliers in the other two variables and they occurred in the same year as these other outliers, it was decided to take these values as being outliers. These results are shown in Table 2.2.

Outliers were treated as missing values.

Table 2.2: OUTLIERS IN THE TIME SERIES

Variable	Value	Day	Year
Evaporation	488	4th April	1980
	488	9th April	1980
	488	15th April	1980
	488	20th December	1980
Maximum Humidity	9488	8th April	1980
	9488	20th April	1980
	9488	10th November	1980
Minimum Humidity	9488	9th April	1980
	9488	19th April	1980
	9488	20th April	1980
	9488	9th November	1980

2.1 Treatment of leap years.

Whenever it was a leap year, the value observed on the 29th of February was added to the value observed on the 28th of February for the variables

- rainfall and
- evaporation.

For the variables

- Maximum temperature
- Minimum Temperature
- Sunshine duration
- Windrun
- Maximum humidity, and
- Minimum humidity,

the mean of the observed values of the 29th of February and of the 28th of February replaced the observed value of the 28th of February.

If the 28th of February had a missing value then it was replaced by the observed value on the 29th of February.

2.2 Distinctive features of the time series useful for model identification.

2.2.1 Seasonality.

A simple moving average smooth was used to filter the series. This is given by:

$$M_t = 1/(2L + 1) \sum_{\ell=-L}^{L} m_{t + \ell}$$

where m_t is the mean of the time series at time t , i.e.

$$m_t = 1/I \sum_{i=1}^{I} x_{i,t}$$

where $x_{i,t}$ is the observation made at time t of the ith year,

$i=1,2,\ldots,I$, I being the number of years for which data is available,

$t=1,2,\ldots,365$

and L is the lag.

Lags of 10, 25, 50 and 100 days were applied.

Note that in the above equation, because m_t is cyclic one has that

$$m_{366} = m_1$$

$$m_{367} = m_2$$

$$\vdots$$

$$m_o = m_{365}$$

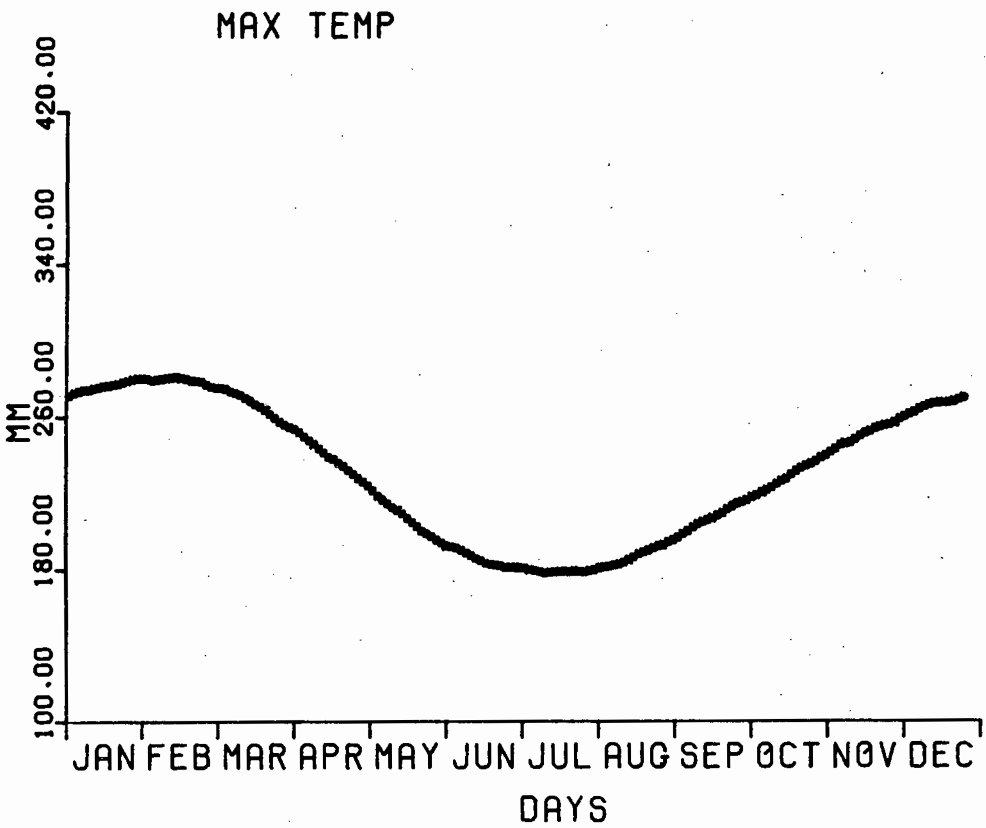
$$m_{-1} = m_{364}$$

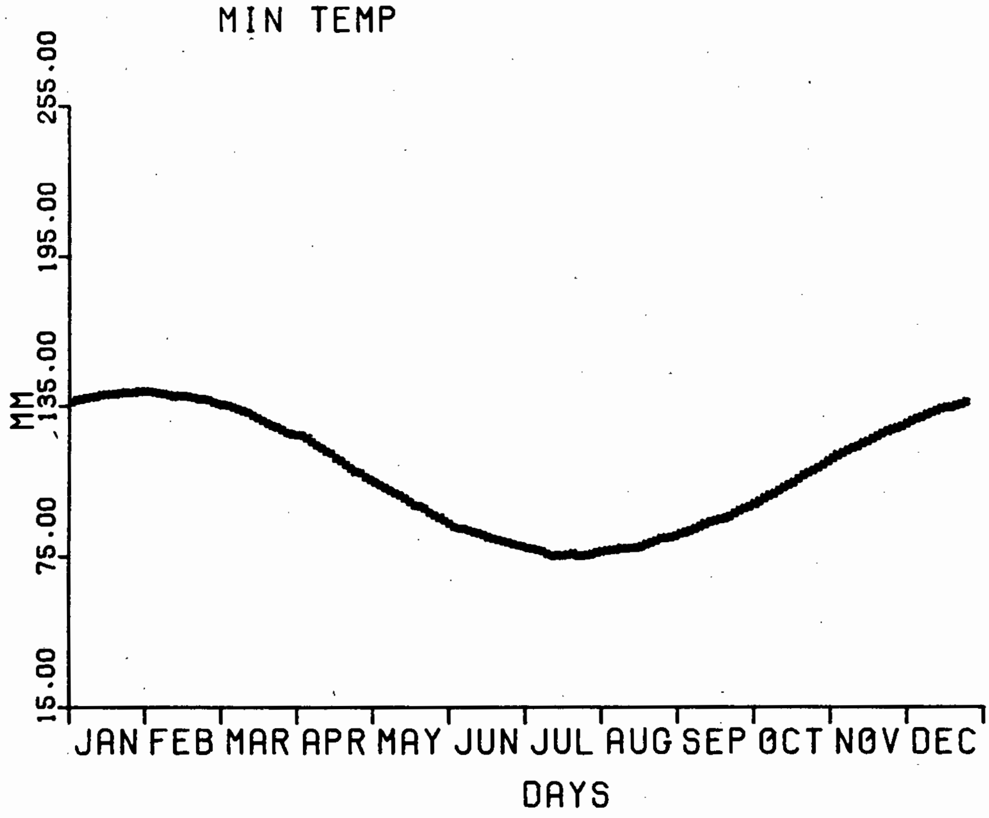
and so on.

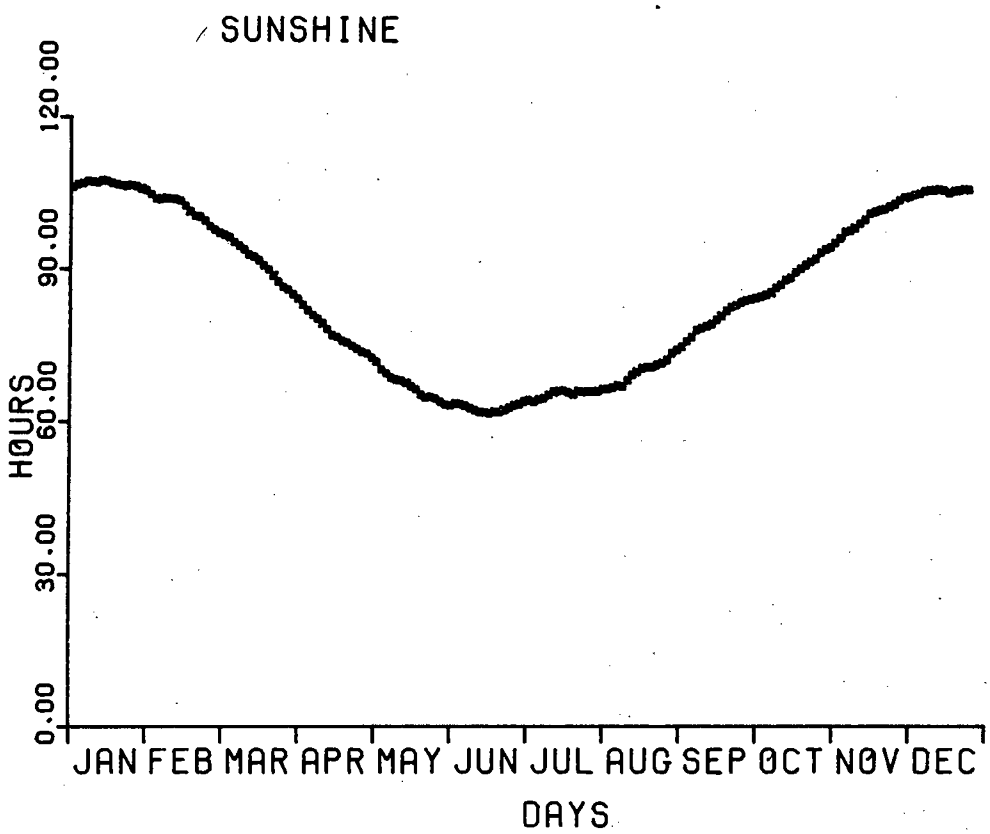
Figure 2.1 shows the smooth plots for the various variables.

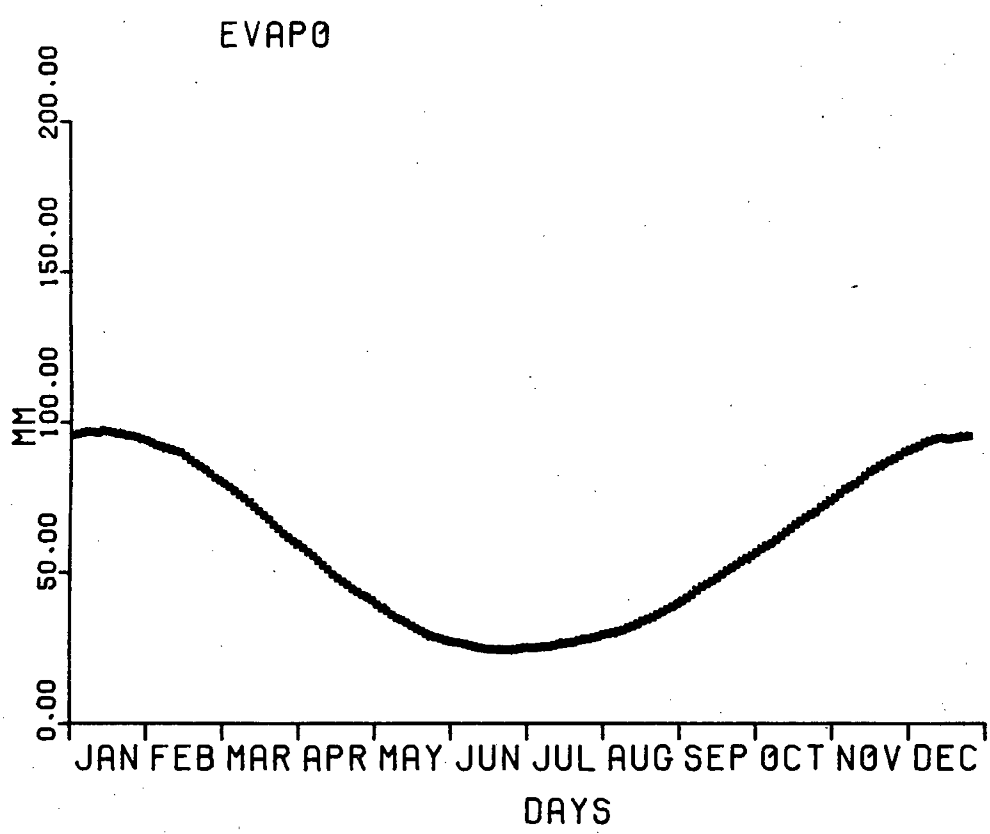
From the smooth plots it can be concluded that each time series of the variables is seasonal, has a cyclic period of one year and has a sinusoidal shape.

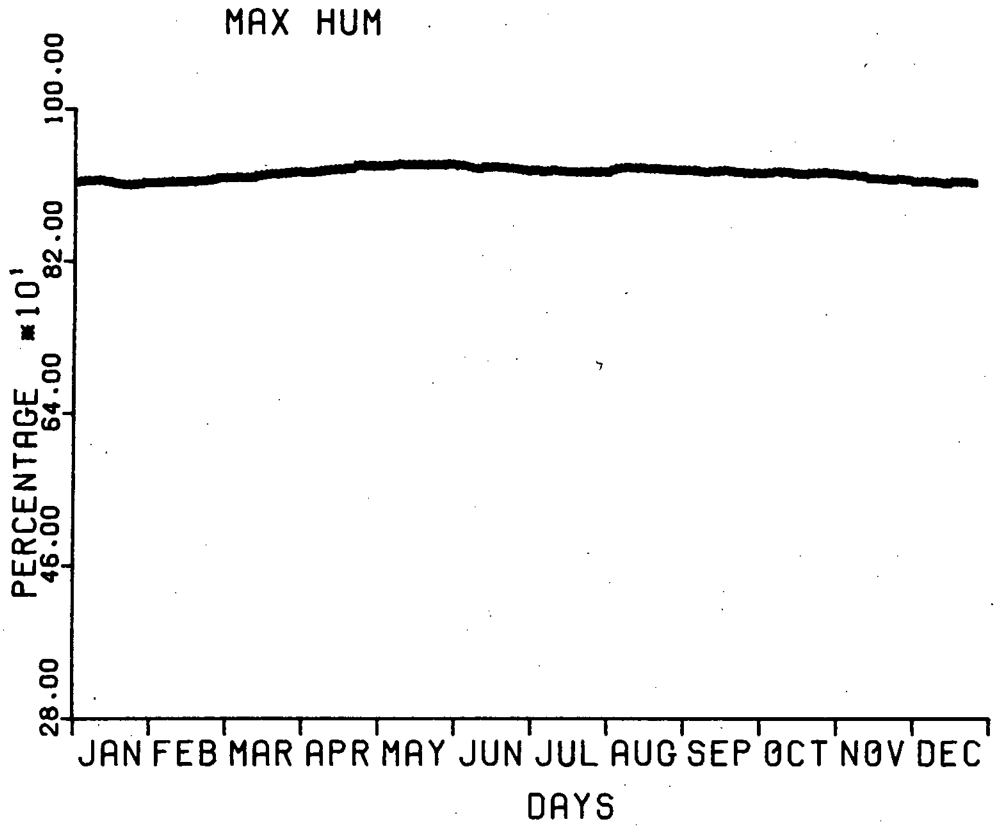
FIGURE 2.1: SIMPLE MOVING AVERAGE SMOOTH (LAG = 50) FOR ALL VARIABLES

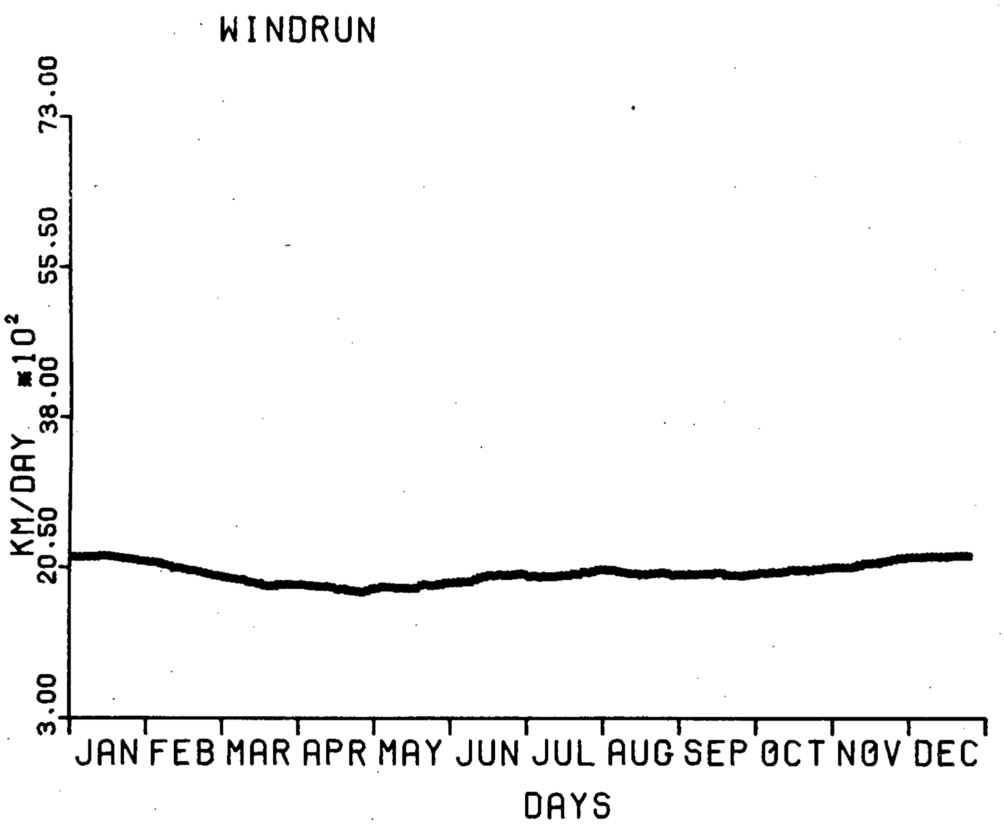


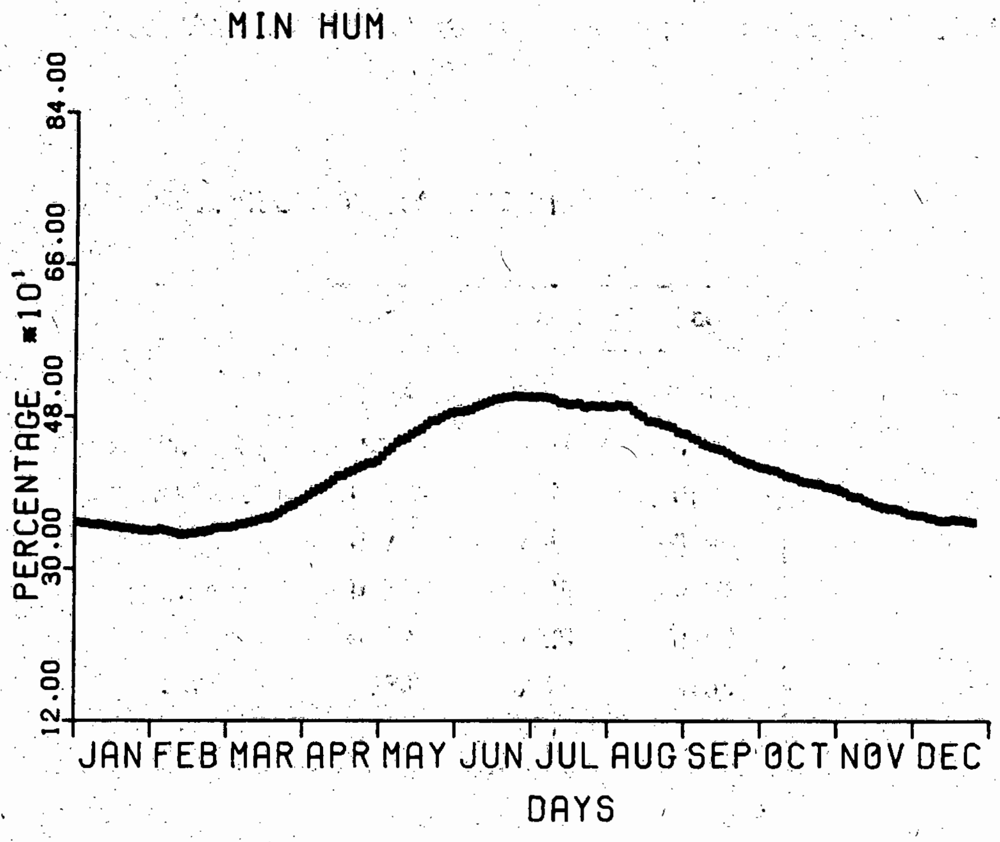



SUNSHINE
HOURS
0.00
30.00
60.00
90.00
120.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


EVAPO
MM
0.00
50.00
100.00
150.00
200.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


MAX HUM
PERCENTAGE *10¹
100.00
82.00
64.00
46.00
28.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


WINDRUN
KM/DAY *10²
73.00
55.50
38.00
20.50
3.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


MIN HUM
PERCENTAGE *10¹
84.00
66.00
48.00
30.00
12.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS

2.2.2 Autocorrelation.

Table 2.3 shows the autocorrelation for each variable up to lags of three.

TABLE 2.3: AUTOCORRELATION COEFFICIENTS

Variable	Lag 1	Lag 2	Lag 3
Rainfall	0.23	0.08	0.05
Maximum Temperature	0.77	0.59	0.51
Minimum Temperature	0.68	0.52	0.45
Evaporation	0.75	0.69	0.65
Sunshine duration	0.50	0.29	0.21
Windrun	0.39	0.09	0.02
Maximum Humidity	0.33	0.14	0.05
Minimum Humidity	0.54	0.33	0.27

From the above table it can be seen that the variables are autocorrelated, i.e. there is a short-term persistence within each variable.

2.2.3 Cross-correlation.

Intuitively, one would expect climate variables to be related to each other in some way, for example one would expect the amount of evaporation to be related to the temperature. In fact as already mentioned, evaporation can be expressed approximately in terms of the other variables. The interdependence among the variables was determined by calculating the lag cross-correlation coefficients of the time series. These cross-correlation coefficients are shown in Table 2.4.

From Table 2.4 it can be seen that the variables are indeed interdependent.

TABLE 2.4: CROSS-CORRELATION COEFFICIENTS BETWEEN VARIABLES

Variables	Lag Cross-Correlation				
	$r_2(j,i)$	$r_1(j,i)$	$r_0(i,j)$	$r_1(i,j)$	$r_2(i.j)$
Max Temp - Min Temp	0.40	0.41	0.51	0.72	0.71
Max Temp - Evapo	0.62	0.72	0.76	0.62	0.54
Max Temp - Sun	0.51	0.63	0.63	0.33	0.23
Max Temp - Wind	-0.08	-0.19	-0.15	0.06	0.10
Max Temp - Max Hum	-0.10	-0.19	-0.27	-0.20	-0.08
Max Temp - Min Hum	-0.41	-0.57	-0.72	-0.39	-0.26
Min Temp - Evapo	0.63	0.60	0.49	0.46	0.46
Min Temp - Sun	0.42	0.24	0.05	0.15	0.22
Min Temp - Wind	-0.03	0.23	0.26	0.13	0.07
Min Temp - Max Hum	-0.17	-0.18	-0.29	-0.06	-0.03
Min Temp - Min Hum	-0.44	-0.33	-0.08	-0.12	-0.19
Evapo - Sun	0.48	0.56	0.72	0.47	0.37
Evapo - Wind	0.08	0.04	0.14	0.11	0.11
Evapo - Max Hum	-0.08	-0.14	-0.27	-0.33	-0.15
Evapo - Min Hum	-0.36	-0.45	-0.60	-0.46	-0.36
Sun - Wind	0.02	-0.19	-0.21	-0.04	0.07
Sun - Max Hum	0.04	-0.02	-0.15	-0.22	-0.14
Sun - Min Hum	-0.15	-0.32	-0.69	-0.49	-0.32
Wind - Max Hum	-0.06	-0.07	-0.13	-0.20	-0.04
Wind - Min Hum	-0.11	-0.05	0.19	0.18	0.01
Max Hum - Min Hum	0.16	0.33	0.32	0.14	0.05

2.2.4 Time series observations differ depending on the wet or dry status of the day.

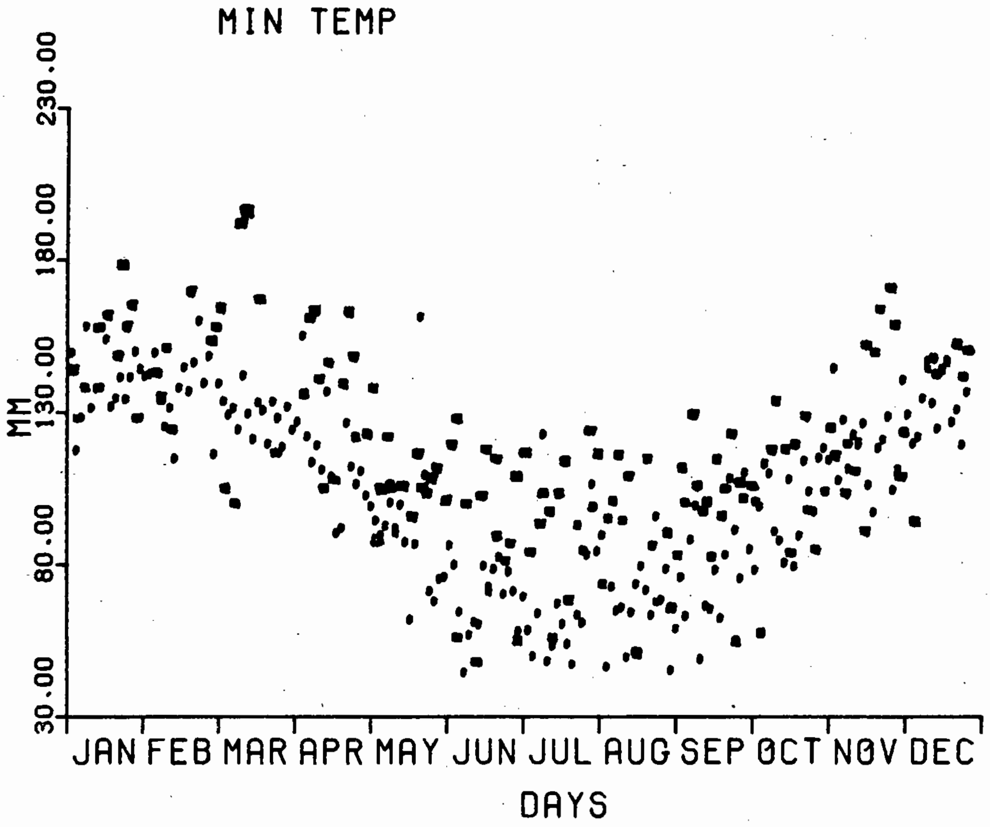
It is known that on days that rain occurs, a marked change also occurs in other climatic variables, for example, temperature and sunshine duration are more likely to be below normal on rainy days than on dry days, humidity, on the other hand will be above average on a rainy day rather than on a dry day. This property of the climate variables was investigated to determine whether the difference was significantly distinct.

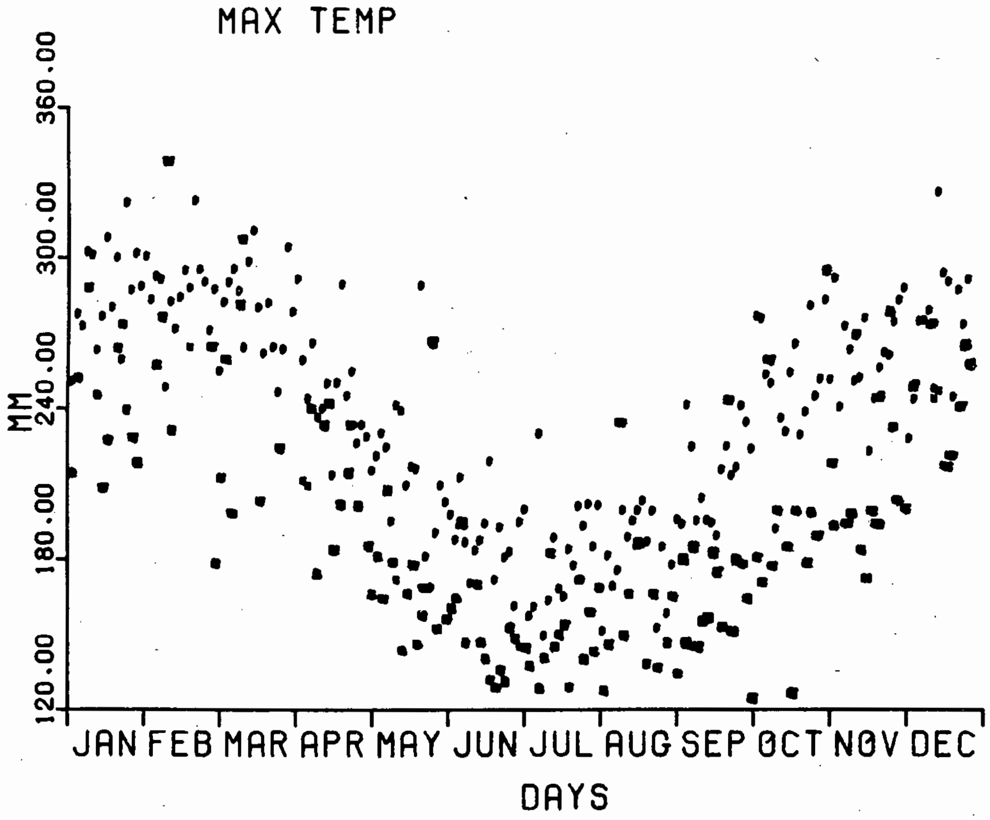
The observations of all variables were found to be significantly different depending on whether rain had or had not occurred in that time period. Figure 2.2 shows the mean time series of each variable conditioned on the wet or dry status of the day. Table 2.5 shows a comparison of the mean for each variable conditioned on the wet or dry status of the day.

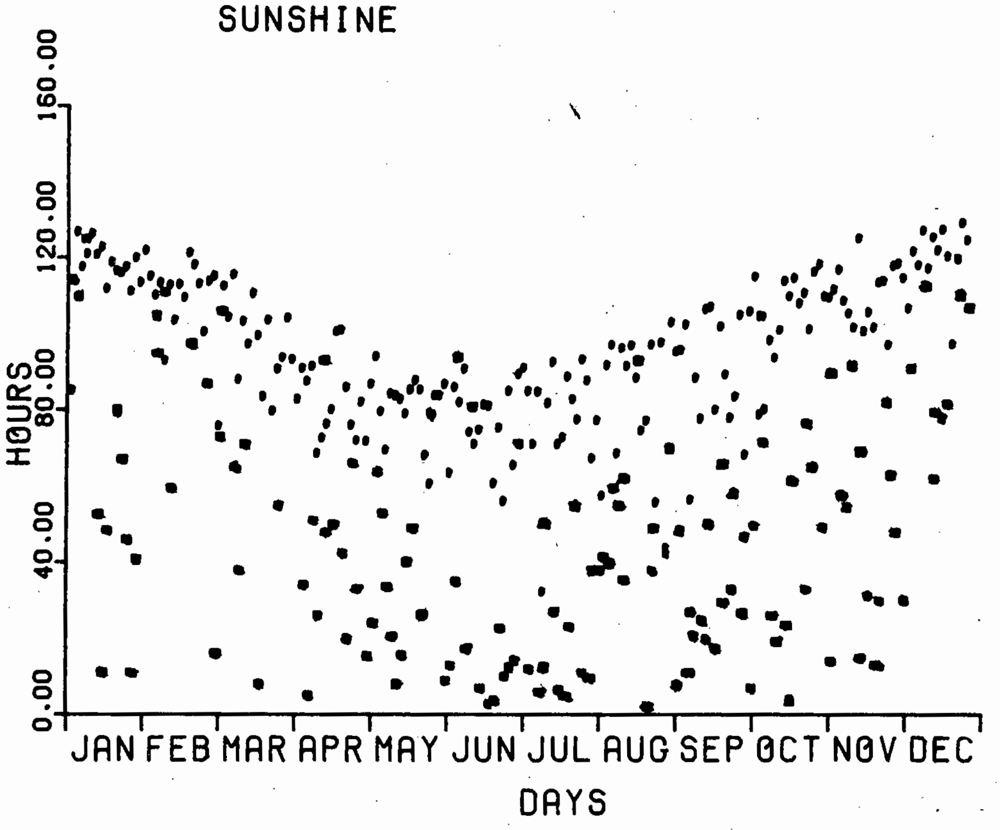
TABLE 2.5: MEAN FOR CONDITIONED TIME SERIES

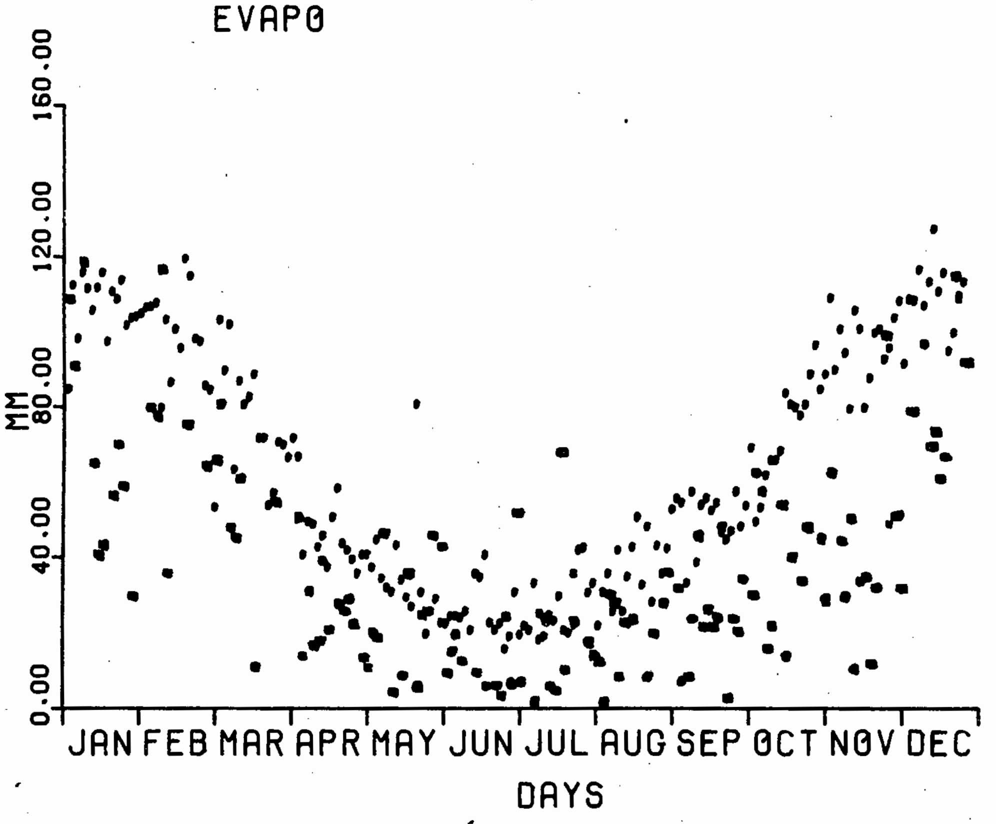
Variable	Dry State	Wet State
Maximum Temperature	238.3	189.9
Minimum Temperature	102.4	114.2
Evaporation	63.39	32.65
Sunshine Duration	94.44	45.24
Windrun	1774	2450
Maximum Humidity	910.7	932.4
Minimum Humidity	374.6	531.0

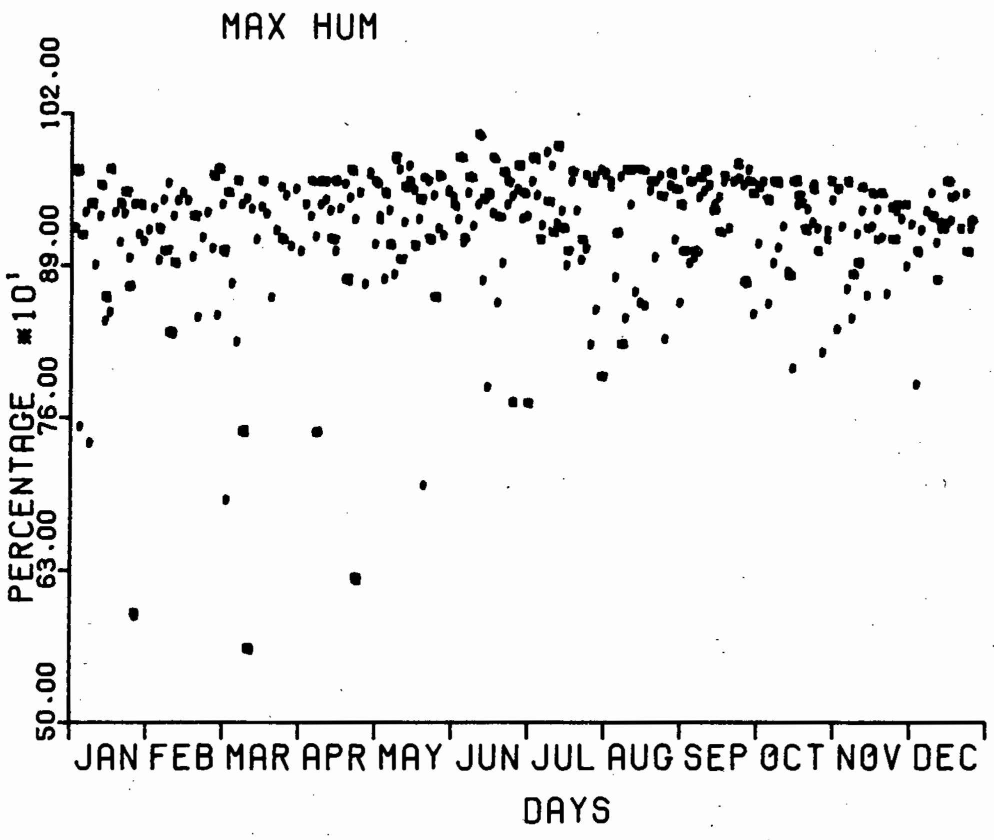
FIGURE 2.2: MEAN TIME SERIES CONDITION ON STATUS OF DAY (D = DRY, ■ = WET) FOR ALL VARIABLES

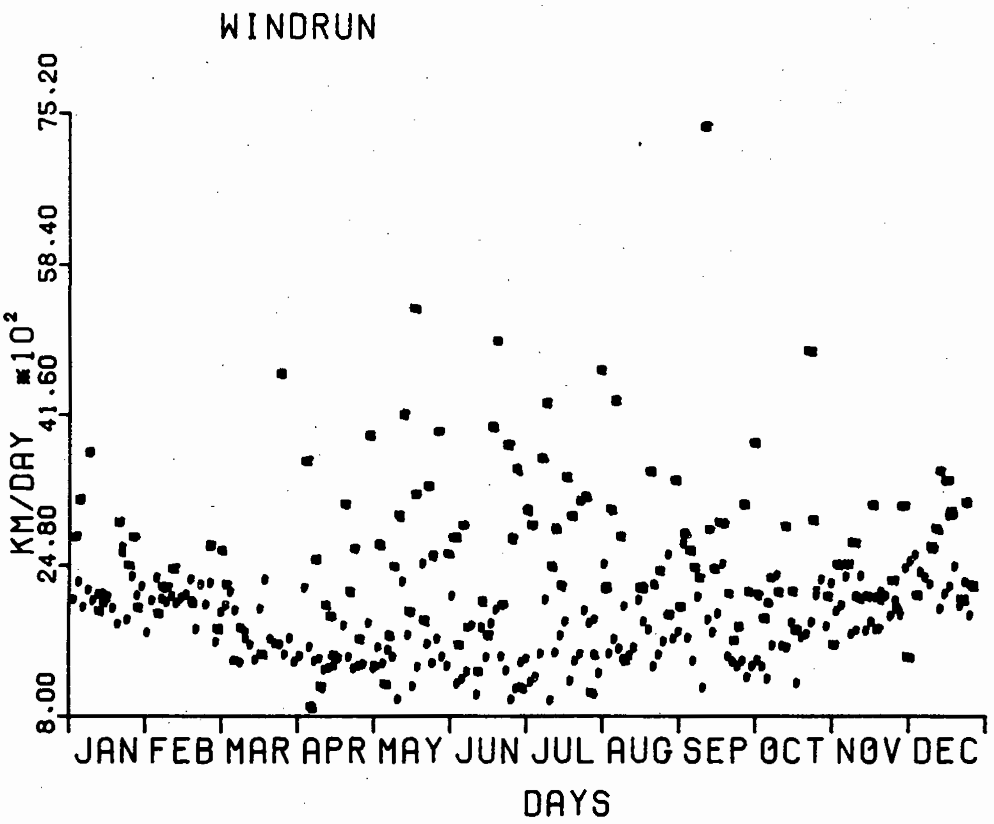


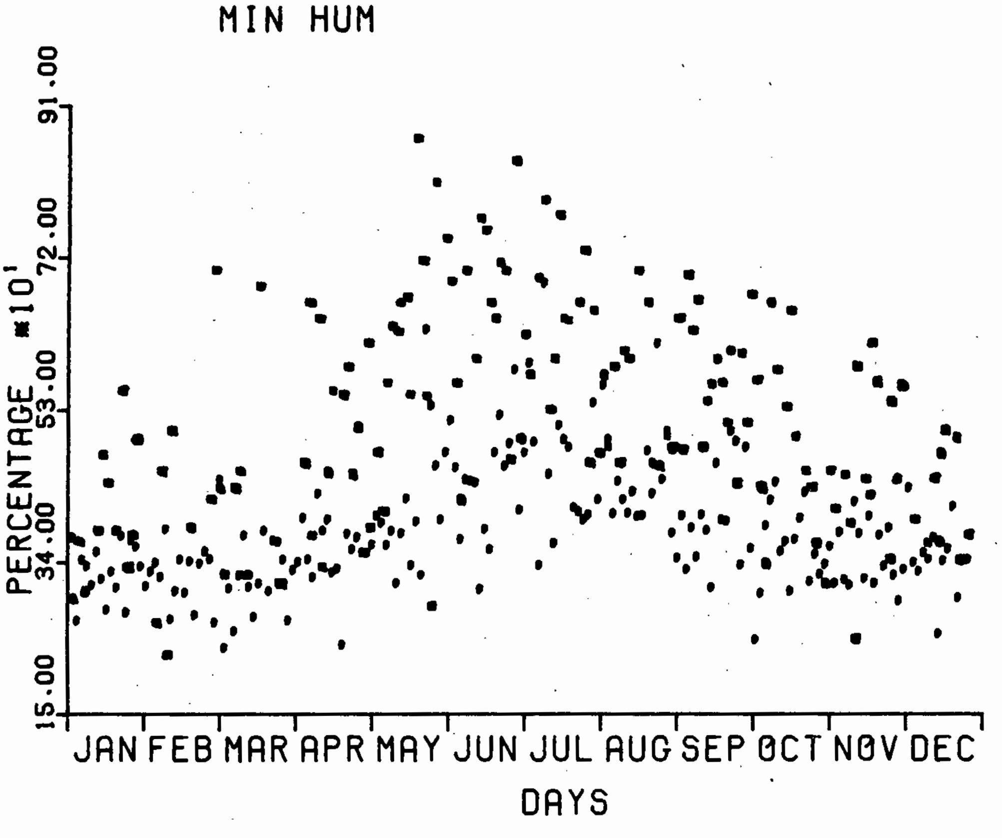



SUNSHINE
HOURS
160.00
120.00
80.00
40.00
0.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


EVAPO
MM
160.00
120.00
80.00
40.00
0.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


MAX HUM
102.00
89.00
76.00
63.00
50.00
PERCENTAGE *10¹
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


WINDRUN
75.20
58.40
41.60
24.80
8.00
KM/DAY *10²
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


MIN HUM
PERCENTAGE *10¹
91.00
72.00
53.00
34.00
15.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS

Having concluded that climatic variables vary depending on whether rain or no rain has occurred, it remains to examine whether the amount of rainfall is related in any way to the observations of the climate variables. Figure 2.3 shows the graphs of rainfall versus each climate variable. From these plots it is concluded that there is no visible pattern to the values of the climate variables in relation to the amount of rainfall.

2.2.5 Rainfall is a "strange" variable.

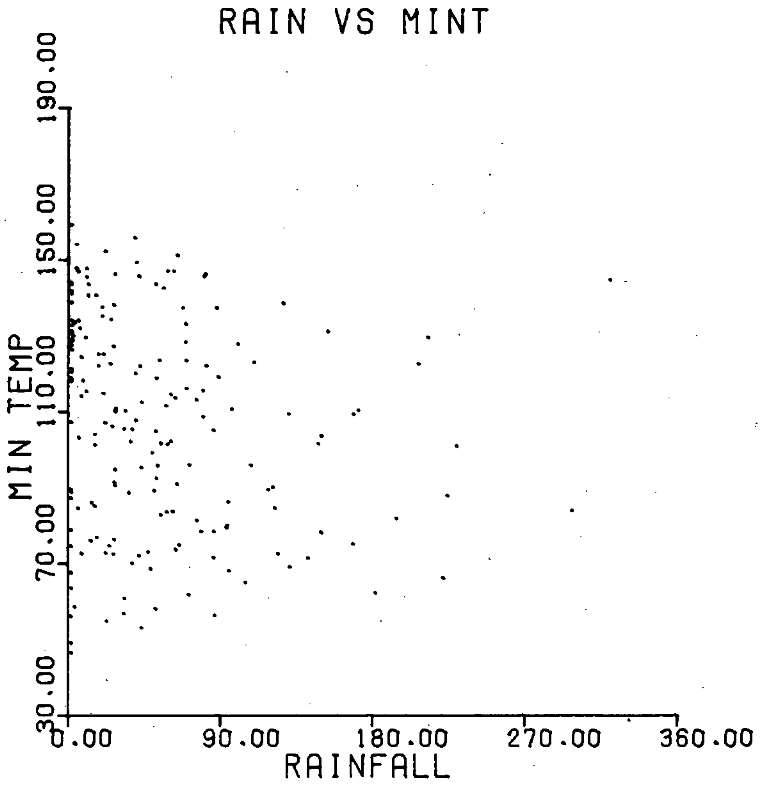
The rainfall variable is somewhat unusual from a statistical point of view in the sense that it exhibits different properties from those of the other climatic variables. The distribution of rainfall is both discrete and continuous. The occurrence or non-occurrence of rainfall is considered as discrete, while on the times that it does rain, the depth of rainfall has a continuous distribution.

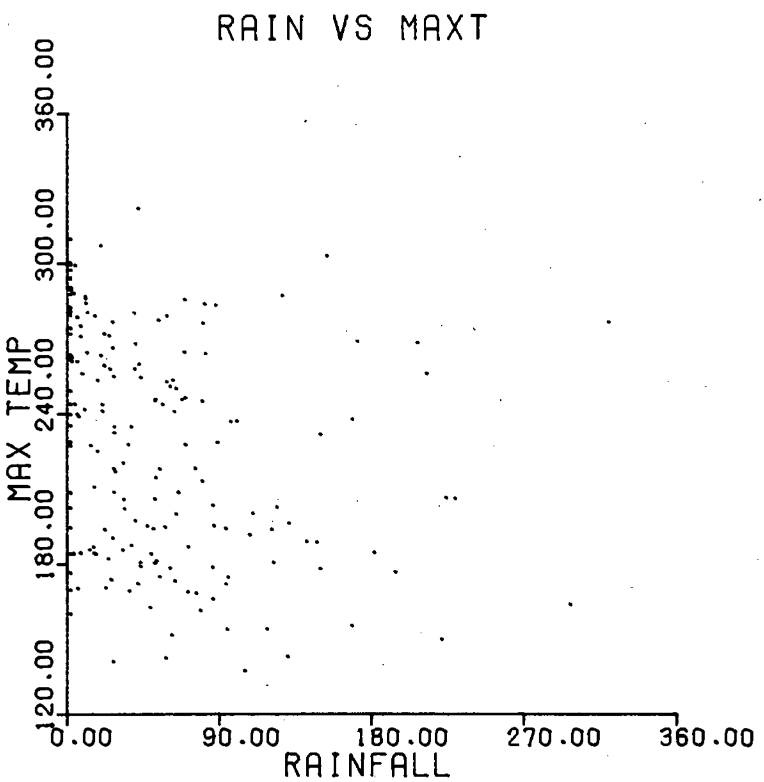
Another distinctive feature of rainfall is that especially in a country like South Africa, the proportion of rainy days is relatively small.

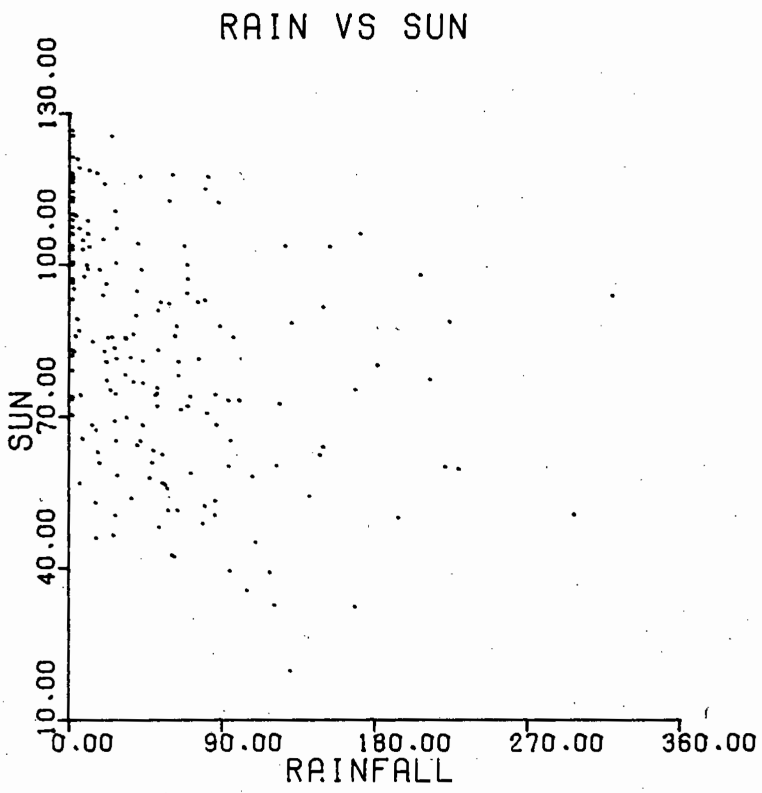
2.3 Concluding remarks.

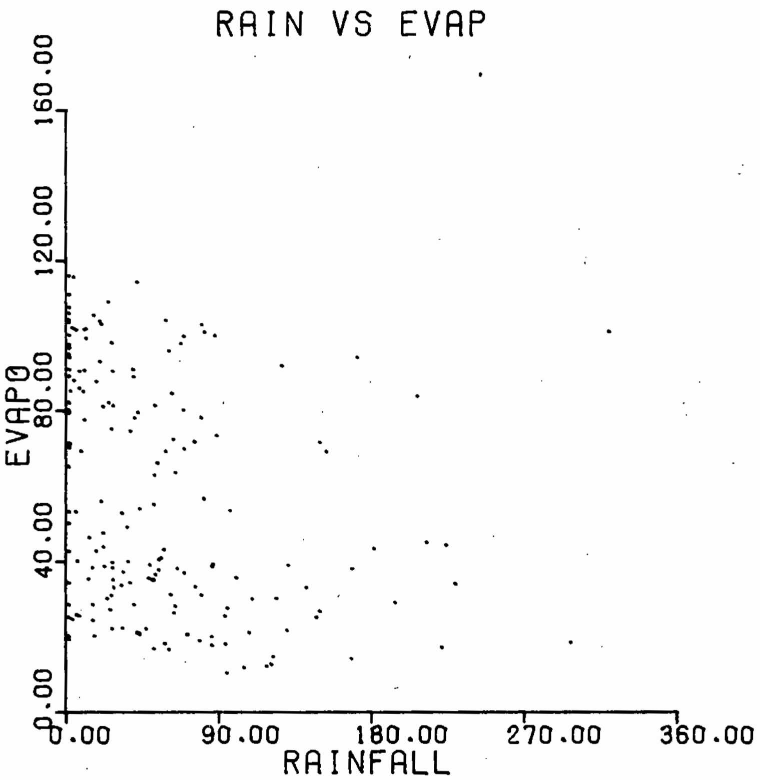
The above preliminary analysis establishes a number of facts. Firstly, the individual climate variables exhibit seasonal fluctuations and these fluctuations appear to follow a sinusoidal pattern. This would suggest that the mean function of each variable could be parsimoniously modelled by means of a truncated Fourier series. Secondly, the individual variables are serially correlated (even after this seasonal fluctuation has been taken into account). In other words, the individual

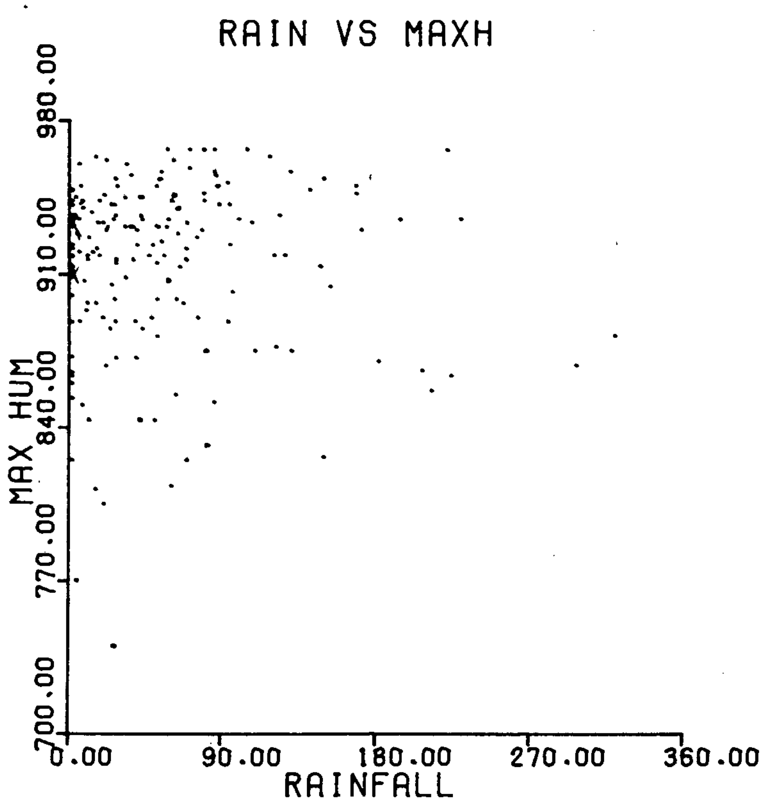
FIGURE 2.3: RAINFALL VERSUS CLIMATE FOR ALL VARIABLES.

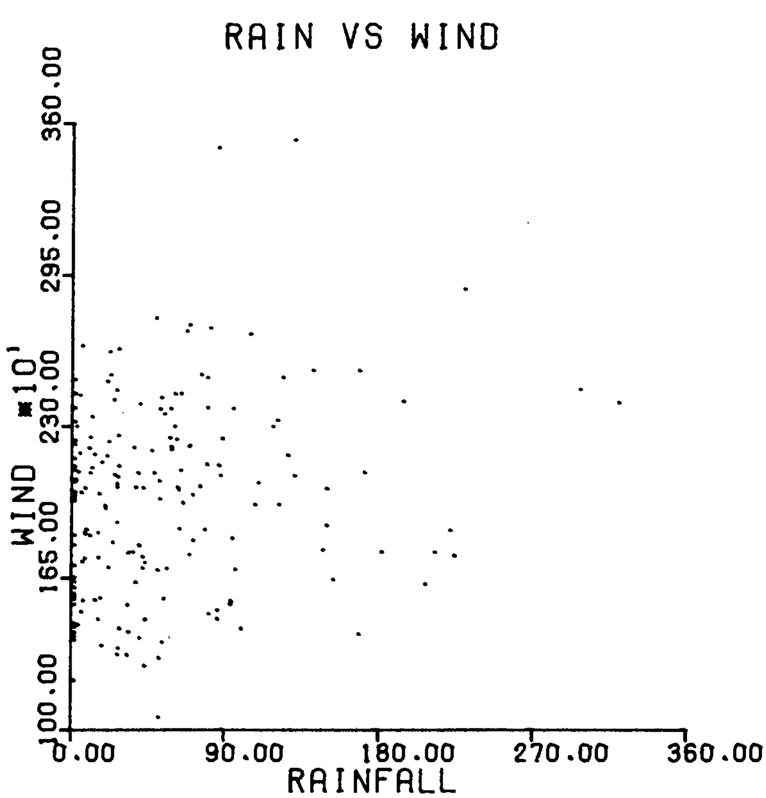


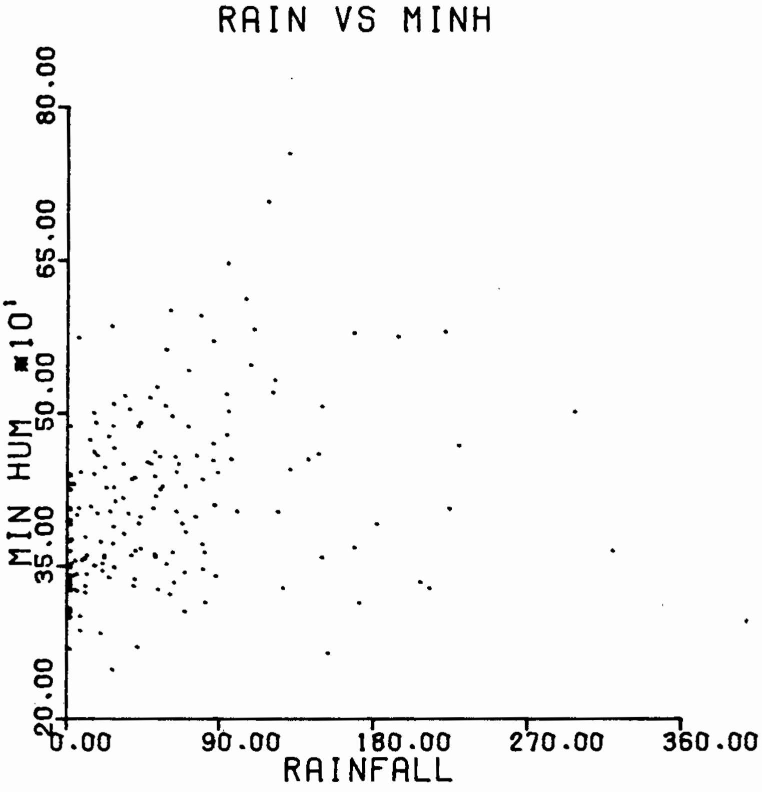



RAIN VS SUN
SUN
130.00
100.00
70.00
40.00
10.00
0.00
90.00
180.00
270.00
360.00
RAINFALL


RAIN VS EVAP
EVAPO
160.00
120.00
80.00
40.00
0.00
0.00
90.00
180.00
270.00
360.00
RAINFALL


RAIN VS MAXH
MAX HUM
980.00
910.00
840.00
770.00
700.00
0.00
90.00
180.00
270.00
360.00
RAINFALL


RAIN VS WIND
WIND *10¹
360.00
295.00
230.00
165.00
100.00
0.00
90.00
180.00
270.00
360.00
RAINFALL


RAIN VS MINH
MIN HUM *10'
80.00
65.00
50.00
35.00
20.00
0.00
90.00
180.00
270.00
360.00
RAINFALL

climate variables constitute time series and have to be modelled as such. This preliminary analysis would suggest that an autoregressive model might be suitable to describe the autocorrelation structure of the variables. Here one has to keep in mind that the number of parameters in the final model must be kept to a minimum in order to avoid the usual statistical problems associated with estimating a large number of parameters. An autoregressive model is ideal in this respect.

Finally, the variables are cross-correlated, that is, they do not vary independently of each other. It follows that it is not possible to model climate by separately modelling its component variables: a multivariate time series model is required.

Seeing that the variable rainfall has some extra properties that have to be taken into consideration and that the remaining climatic variables differ depending on the state of the rainfall variable, it is proposed that the rainfall variable should be determined independently of the other variables and then to condition the other variables for a given day on whether the day was wet or dry.

As no pattern was found between different precipitation amounts and the climate observations, it was decided to consider a non-rainy day as one with a precipitation amount of zero and a rainy day as one with a rainfall depth greater than zero.

CHAPTER 3

3. THE MODELS

The preliminary analysis described in Chapter 2 established that sequences of climate variables exhibit a number of distinctive features. In particular the distribution of each climate variable varies seasonally, the variables are serially correlated, they are interdependent, and finally the distributions of the variables depend on the wet or dry status of the day under consideration. Any useful model for the simultaneous description of climate sequences must of course preserve all these properties.

The models considered in this thesis are constructed in two stages. One begins by constructing a model for the rainfall process. This provides synthetic sequences of wet and dry days. The remaining variables are then modelled according to the wet or dry status of each day. Thus the joint distribution of all the variables other than rainfall changes not only with season but also with changes in wet or dry status.

The rainfall component of the three models to be discussed is common to all three models and is thus described first. The first of the three models is due to Richardson (1981), the remaining two are new.

3.1 The rainfall model.

Several models have been proposed for simulating daily precipitation. (Gabriel and Neumann, 1962; Richardson, 1981; Roldan and Woolhiser, 1982; Stern and Coe, 1984; Zucchini and Adamson, 1984.) Most precipitation models are specified by a discrete occurrence process describing the sequence of wet and dry days, and a continuous distribution function for the amount of precipitation on days with rain. The parameters of the model are allowed to vary seasonally.

3.1.1 A model to describe the occurrence of wet and dry sequences of days.

A first-order Markov chain is used to describe the occurrence of wet and dry days. By this one assumes that the state of day t depends on the state of the previous day, $t-1$. This does not imply that the state at time t is independent of the state on day $t-2$, $t-3$, etc. ..., but rather that the information given by $t-1$ is equivalent to all the information given by $t-1$, $t-2$, etc.... One also assumes that, except for the seasonality, the process is stationary.

A first-order Markov chain has been found to be an adequate model for precipitation occurrence in many different regions. (Gabriel and Neumann, 1962; Caskey, 1963; Weiss, 1964; Hopkins and Robillard, 1964; Haan et al, 1976; Smith and Schreiber, 1973; Woolhiser and Prengram, 1979; Richardson, 1981; Roldan and Woolhiser, 1982; Zucchini and Adamson, 1984.) The order of the Markov chain may of course be increased, but this has to be done at the cost of increasing complexity and the number of parameters in the model. A further problem arises if one attempts to increase the order of the Markov chain in arid areas, namely the estimation of the probability that a rain day follows two or more consecutive rain days. In arid areas there are relatively few runs of three or more consecutive rain days and thus there is hardly any data on which to base estimates of this conditional probability. (Note that this has to be estimated for each day of the year.) Finally, it was demonstrated in Zucchini and Adamson (1984) that a first order Markov chain provides an adequate description of the occurrence of wet and dry sequences of days in the complete range of South African conditions.

(a) Notation and preliminaries.

The day will be used as the time unit. That is, the year is divided into NT (= 365) equal intervals, denoted by t=1,2,...,NT. A day with total rainfall greater than 0mm is considered as a wet day.

The following notation will be used:

R	represents the occurrence of rain (i.e. wet day).
$\bar{R}$	represents the non-occurrence of rain (i.e. dry day).

For t=1,2,...,NT

NR(t)	is the number of times it was wet in period t .
N$\bar{R}$(t)	is the number of times it was dry in period t .
N$\bar{R}$R(t)	is the number of times it was dry in period t-1 and wet in period t .
N$\bar{R}\bar{R}$(t)	is the number of times it was dry in period t-1 and dry in period t .
NRR(t)	is the number of times it was wet in period t-1 and wet in period t .

$ND(t) = N\bar{R}R(t) + N\bar{R}\bar{R}(t)$ is the number of times that it was dry in period $t-1$ and there was an observation (wet or dry) in period t.

$NW(t) = NRR(t) + NR\bar{R}(t)$ is the number of times that it was wet in period $t-1$ and there was an observation (wet or dry) in period t.

$\pi_{R/R}(t)$ the probability that period t is wet given that period $t-1$ is wet.

$\pi_{\bar{R}/R}(t)$ the probability that period t is dry given that period $t-1$ is wet.

$\pi_{R/\bar{R}}(t)$ the probability that period t is wet given that period $t-1$ is dry.

$\pi_{\bar{R}/\bar{R}}(t)$ the probability that period t is dry given that period $t-1$ is dry.

Then
$$\pi_{R/R}(t) + \pi_{\bar{R}/R}(t) = 1$$
$$\pi_{\bar{R}/\bar{R}}(t) + \pi_{R/\bar{R}}(t) = 1.$$

Therefore the transition probabilities are fully defined given $\pi_{R/R}(t)$, $\pi_{R/\bar{R}}(t)$ and the wet or dry state on day $t-1$, and one only needs to estimate these two probabilities.

From elementary probability theory we have

$$NRR(t) \sim B(NW(t), \pi_{R/R}(t))$$
$$N\bar{R}R(t) \sim B(ND(t), \pi_{R/\bar{R}}(t)), \quad t=1,2,\ldots,NT$$

where $B(N,\pi)$ denotes the binomial distribution with parameters N and π.

(b) Estimation.

The functions $\pi_{R/R}(t)$ and $\pi_{R/\bar{R}}(t)$ are estimated using the same method but different data. To simplify the notation in what follows, one makes use of the following generic names:

$$\text{Let} \quad M(t) \sim B(MM(t), \ \pi(t)) , \qquad t=1,2,\ldots,NT.$$

First we note that the binomial distribution belongs to the exponential family. Therefore we have a set of independent random variables $M(t)$, $t=1,2,\ldots,NT$, each with a distribution from the exponential family; each $M(t)$ depends on a single parameter $\pi(t)$ and the distributions of all $M(t)$, $t=1,2,\ldots,NT$, are of the same form (i.e. all binomial). Thus the properties of a generalized linear model are satisfied, and estimates for $\pi(t)$ may be obtained by using the theory for estimation for generalized linear models. (Dobson, 1983.)

The probabilities $\pi(t)$ are assumed to be functions of linear combinations of parameters $\gamma_1,\gamma_2,\ldots,\gamma_L$, $L < NT$. That is

$$g(\pi(t)) = \lambda(t,L)$$

where g is the link function and $\lambda(t,L)$ is a linear combination of the γ_is.

To ensure that the estimated values of $\pi(t)$ are restricted to the interval $[0,1]$, one uses the logit link function, given by

$$g(\pi(t)) = \log(\pi(t)/(1-\pi(t))) = \lambda(t,L) .$$

To obtain the linear combination of the γ_is, $\lambda(t,L)$, we look at some of the properties of $\pi(t)$, namely that it is a smooth, periodic and approximately sinusoidal shaped function. Transforming $\pi(t)$, using the logistic transformation, to a logit $\lambda(t)$ given by

$$\lambda(t) = \log(\pi(t)/(1-\pi(t))) ,$$

one obtains a representation which still has the same properties as $\pi(t)$, and thus we can approximate $\lambda(t)$ by the first few terms of its Fourier representation. This approximation has been used by Stern and Coe (1984) and Zucchini and Adamson (1984).

The exact Fourier representation of $\lambda(t)$ is given by

$$\lambda(t) = \sum_{i=1}^{NT} \gamma_i \, \varphi_i(t) , \qquad t=1,2,\ldots,NT$$

where

$$\varphi_i(t) = \begin{cases} \cos(\omega(t-1)i/2) & i=2,4,\ldots\ldots \\ \sin(\omega(t-1)(i-1)/2) & i=3,5,\ldots\ldots \end{cases}$$

$$\varphi_1(t) = 1 ; \qquad t=1,2,\ldots,NT ,$$

and $\omega = 2\pi/NT$.

Define the function $\lambda(t,L)$ by

$$\lambda(t,L) = \sum_{i=1}^{L} \gamma_i \, \varphi_i(t) \quad , \qquad t=1,2,\ldots,NT \qquad L \leq NT$$

where $\varphi_i(t)$ is defined as before and L is the order of the Fourier series approximation. One is thus making the following approximation:

For some $L < NT$

$$\lambda(t,L) \approx \lambda(t) \; , \qquad t=1,2,\ldots,NT \; .$$

A procedure to choose the order of the Fourier series approximation (i.e. the value of L) will be discussed later. Generally this approximation is accurate for small values of L . The number of parameters, L , is always chosen to be an odd number. The reason for this choice is given in Appendix A. An outline of the properties of the Fourier representation are given in Appendix B.

The log-likelihood function of the observed values as a function of the probabilities $\pi(t)$, is given by

$$\ell(\pi(t);\, M(t)) = \sum_{t=1}^{NT} \left[M(t).\log\left(\frac{\pi(t)}{1-\pi(t)}\right) + MM(t). \log(1-\pi(t)) + \log\binom{MM(t)}{M(t)} \right]$$

Therefore, the log-likelihood function of the observed values as a function of the parameters $\gamma_1, \gamma_2, \ldots, \gamma_L$ is given by

$$\ell(\gamma;M(t)) = \sum_{t=1}^{NT} \left[M(t) \,.\, \lambda(t,L) - MM(t) \,.\, \log(1 + e^{\lambda(t,L)}) + \log\begin{pmatrix} MM(t) \\ M(t) \end{pmatrix} \right] .$$

The score vector U with respect to $\gamma_1, \gamma_2, \ldots, \gamma_L$ has elements given by

$$U_j = \frac{\partial \ell(\gamma;M(t))}{\partial \gamma_j} = \sum_{t=1}^{NT} \left[M(t) - MM(t) \,.\, \frac{e^{\lambda(t,L)}}{1 + e^{\lambda(t,L)}} \right] . \varphi_j(t)$$

$$= \sum_{t=1}^{NT} [M(t) - MM(t) \,.\, \pi(t)] \, \varphi_j(t)$$

since $\mathrm{Var}(M(t)) = MM(t) \,.\, \dfrac{e^{\lambda(t,L)}}{(1 + e^{\lambda(t,L)})^2}$ and

$$E(M(t)) = MM(t) \,.\, e^{\lambda(t,L)}/(1 + e^{\lambda(t,L)}) \quad \text{and so}$$

$$\frac{\partial E(M(t))}{\partial \lambda(t,L)} = \frac{MM(t) \,.\, e^{\lambda(t,L)}}{(1 + e^{\lambda(t,L)})^2} = \mathrm{Var}(M(t)) .$$

Similarly, the information matrix $\mathcal{I}_{L \times L}$ has elements given by

$$\mathcal{I}_{jk} = \sum_{t=1}^{NT} \varphi_j(t) \,.\, \varphi_k(t) \,.\, MM(t) \,.\, \frac{e^{\lambda(t,L)}}{(1 + e^{\lambda(t,L)})^2} .$$

Since $e^{\lambda(t,L)}/(1 + e^{\lambda(t,L)})^2 = \pi(t)(1 - \pi(t))$ it follows that

$$\mathcal{I}_{jk} = \sum_{t=1}^{NT} \varphi_j(t) \, . \, \varphi_k(t) \, . \, MM(t) \, . \, \pi(t)(1 - \pi(t)) \quad .$$

The maximum likelihood estimates for $\gamma_1, \gamma_2, \ldots, \gamma_L$ are then obtained by solving the iterative equation

$$\mathcal{I}^{(m-1)} \hat{\gamma}^{(m)} = \mathcal{I}^{(m-1)} \hat{\gamma}^{(m-1)} + U^{(m-1)}$$

where m indicates the mth approximation and $\hat{\gamma}$ is the vector of estimates.

Computer packages, such as GLIM, can be used to obtain these estimates. They begin by using some initial approximation $\gamma^{(o)}$ to evaluate $\mathcal{I}^{(o)}$ and $U^{(o)}$, then the iterative equation is solved to give $\gamma^{(1)}$ which in turn is used to obtain better approximations for $\mathcal{I}$ and U, and so on until adequate convergence is achieved. When the difference between successive approximations $\gamma^{(m)}$ and $\gamma^{(m-1)}$ is sufficiently small, $\gamma^{(m)}$ is taken as the maximum likelihood estimate vector.

(c) Model selection.

Whenever a model is fitted to observed data, two types of discrepancy arise. The discrepancy due to approximation (the fewer the number of parameters fitted, the higher the value of this discrepancy) and the discrepancy due to estimation (the more parameters fitted, the higher the value of this discrepancy). When choosing the number of parameters to be fitted, one attempts to minimize the combined effect arising from the two discrepancies.

Selection of the number of parameters, L, may be done by using the criterion of the Kullbach-Leibler measure of discrepancy. (Linhart and Zucchini, 1982; Zucchini and Adamson, 1984.)

Under the assumption that for some L_o, $\lambda(t)$ is exactly fitted by L_o parameters, i.e.

$$\lambda(t) = \lambda(t, L_o) , \quad L_o < NT ,$$

the above method leads to the Akaike Information Criterion where

$$AIC = -\ell(\gamma; M(t)) + L$$

where $\ell(\gamma;M(t))$ is the likelihood function given before.

Each value of L leads to a different approximating model. The criterion is computed for $L=1,3,5,\ldots$ and the model which leads to the smallest value of the criterion is selected.

The AIC criterion is much easier to compute than the full Kullbach-Leibler discrepancy and leads to almost identical results if the discrepancy due to approximation is small (which it is in this application).

Another method for the selection of L is based on the log-likelihood ratio statistic given by

$$D = 2[\ell(\gamma_{max}; M(t)) - \ell(\gamma_L; M(t))]$$

called the deviance (Nelder and Wedderburn, 1972; Dobson, 1983), where $\ell(\gamma_{max}; M(t))$ is the log-likelihood function of the maximal model. Consider the hypothesis

$$H_0 : \gamma = [\gamma_1 \dots \gamma_L]^T = \gamma_L$$

and $$H_1 : \gamma = [\gamma_1 \; \gamma_2 \dots \gamma_P]^T = \gamma_P$$

where $$L < P < NT .$$

We test H_0 against H_1 using the difference

$$D = D_L - D_P = 2[\ell(\gamma_P; M(t)) - \ell(\gamma_L; M(t))] .$$

Under H_0 D_L is asymptotically χ^2_{NT-L} and D_P is (asymptotically) χ^2_{NT-P} independently so that

$$D \sim \chi^2_{P-L} .$$

If the observed value of D is greater than the upper tail 100 α% point of the χ^2_{P-L} distribution we reject H_0 and conclude that the model with γ_P parameters provides a significantly better description of the data though this does not of course guarantee that the alternative model gives an adequate description of the data.

Since deviances are computed in the program package (GLIM) used in the estimation of the parameters, model selection was based on the deviance rather than the Akaike criterion.

3.1.2 The distribution of rainfall on days when rain occurs.

Several models have been proposed for the distribution of precipitation amounts given the occurrence of a wet day. These include the exponential (Todorovic and Woolhiser, 1975; Richardson, 1981); gamma (Ison et al, 1981; Buishand, 1977; Stern and Coe, 1984); two-parameter gamma (Buishand, 1978); three-parameter mixed exponential (Woolhiser and Pengram, 1979); kappa (Mielke, 1973); lognormal and Weibull (Zucchini and Adamson, 1984).

Woolhiser and Roldan (1982b) found that out of the exponential, gamma and mixed exponential distributions, the latter fitted the model of precipitation amounts best. Zucchini and Adamson (1984) found that for stations in South Africa, the lognormal distribution did not fit some stations, while the Weibull seemed to provide better fits.

It is known that the distribution of precipitation depths when rain occurs is positively skewed (i.e. smaller amounts occurring more frequently than the larger amounts) and that it exhibits the same seasonal variability as found with the probabilities $\pi(t)$. To account for this seasonality, the simplest solution is to fit a family of distributions and then to allow the parameters to change over the year, where these parameters are expressed in terms of its Fourier series approximation.

The method of modelling precipitation amounts is adopted from Zucchini and Adamson (1984). Here one does not fit any model initially, the first two moment functions of the distribution are fitted instead. These are then used to estimate the parameters (by the method of moments) to any desired two-parameter model. Different families can be fitted to a single record, e.g. one for the rainy season and a second for the dry season.

(a) Notation.

The year is divided into NT equal intervals denoted by $t=1,2,\ldots,NT$.

$M(t)$ represents the number of times that it rained in period t.

$R(i,t)$ represents the rainfall depth on the ith year that it rained in period t, where $i=1,2,\ldots,M(t)$.

C represents the coefficient of variation which we assume to be constant for all t (Zucchini and Adamson, 1984).

$\mu(t)$ represents the mean rainfall per rainy day in period $t=1,2,\ldots,NT$.

(b) Estimating the mean and coefficient of variation.

As observed before $\mu(t)$ can be approximated by its truncated Fourier series representation and thus reducing the number of parameters to be estimated. That is, we make the approximation:

$$\mu(t,L) \approx \mu(t), \qquad t=1,2,\ldots,NT, \quad L < NT$$

where $\mu(t)$ is defined as

$$\mu(t) = \sum_{i=1}^{NT} \mu_i \, \varphi_i(t) \qquad t=1,2,\ldots,NT$$

and

$$\mu(t,L) = \sum_{i=1}^{L} \mu_i \varphi_i(t) \qquad t=1,2,\ldots,NT \quad L \le NT$$

and $\varphi_i(t)$ is defined as before.

Define $m(t)$ to be the observed means for each period, i.e.

$$m(t) = 1/M(t) \sum_{i=1}^{M(t)} R(i,t) , \qquad t=1,2,\ldots,NT \quad i=1,2,\ldots,M(t) \quad M(t) > 0$$

where $m(t)$ is not defined when $M(t) = 0$, i.e. it never rained in period t .

We use the method of least squares on $m(t)$ to estimate $\mu_1,\mu_2,\ldots,\mu_L$, that is, minimize

$$\sum_{t=1}^{NT} (m(t) - \mu(t,L))^2 \tag{1}$$

with respect to the μ_i , $i=1,2,\ldots,L$. The solution, when some of the $M(t) = 0$, something which occurs often in arid regions, is given by

$$\hat{\mu}_i = K(i) \sum_{\substack{t=1 \\ M(t) > 0}}^{NT} m(t) \cdot \varphi_i(t) \tag{2}$$

where $$K(i) = \sum_{\substack{t=1 \\ M(t) > 0}}^{NT} \varphi_i(t)^2 \qquad i=1,2,\ldots,L$$

The $m(t)$ in (1) are given the same weight and so periods which had very little rainfall have a large influence in the estimates of $\mu(t)$. To overcome this difficulty, the following criterion is used instead:

Minimize

$$S(\mu) = \sum_{t=1}^{NT} \sum_{i=1}^{M(t)} (R(i,t) - \mu(t,L))^2 \tag{3}$$

with respect to μ_i, $i=1,2,\ldots,L$.

By adding and subtracting $m(t)$ inside the squared term of (3), $S(\mu)$ can be rewritten as

$$S(\mu) = S + \sum_{t=1}^{NT} M(t)(m(t) - \mu(t,L))^2 \tag{4}$$

where

$$S = \sum_{t=1}^{NT} \sum_{i=1}^{M(t)} (R(i,t) - m(t))^2$$

and $m(t)$ is defined as before if $M(t) \neq 0$ and $m(t) = 0$ if $M(t) = 0$.

To minimize (4) set its partial derivatives equal to zero:

$$\frac{\partial S(\mu)}{\partial \mu_i} = -2 \sum_{t=1}^{NT} M(t)(m(t) - \mu(t,L))\varphi_i(t) ,$$

$$i=1,2,\ldots,L .$$

These L equations can be solved using the Newton-Raphson iteration method. For this, we need the second partial derivatives:

$$\frac{\partial^2 S(\mu)}{\partial\mu_i \partial\mu_j} = 2 \sum_{t=1}^{NT} M(t)\ \varphi_i(t)\ \varphi_j(t)\ , \qquad i,j=1,2,\ldots,L\ .$$

Denote the ith element of the vector $f^{(k)}$ by

$$f_i^{(k)} = \sum_{t=1}^{NT} M(t)(m(t) - \mu^{(k)}(t,L))\ \varphi_i(t)\ , \qquad i=1,2,\ldots,L \tag{5}$$

and the (i,j)th element of the matrix $F^{(k)}$ by

$$F_{ij}^{(k)} = \sum_{t=1}^{NT} M(t)\ \varphi_i(t)\ \varphi_j(t)\ , \qquad i,j=1,2,\ldots,L \tag{6}$$

where k denotes the kth iteration.

Then an algorithm to estimate μ_i , $i=1,2,\ldots,L$ is given by:

Step 1: Obtain initial estimates $\mu_1^{(o)},\ldots,\mu_L^{(o)}$ using (2) and compute $\mu^{(o)}(t,L)$.

Step 2: Compute $f^{(k)}$ using (5) and $F^{(k)}$ using (6).

Step 3: Compute the vector $\delta^{(k)}$ which is the solution to the system of L linear equations given by

$$F^{(k)}\ \delta^{(k)} = f^{(k)}\ .$$

Step 4: Set $\mu^{(k+1)} = \mu^{(k)} - \delta^{(k)}$.

Step 5: Test for convergence, e.g. if the elements of $f^{(k)}$ are sufficiently close to zero. If the convergence criterion is met, stop, otherwise increase k by 1 and go to Step 2.

Note that $F^{(k)}$ is symmetric. This fact can be used to reduce the number of computations performed.

An estimator of C is given by:

$$\hat{C} = \left[\left[\sum_{t=1}^{NT} \sum_{i=1}^{M(t)} (R(i,t) - \hat{\mu}(t))^2\right] / \left[\sum_{t=1}^{NT} M(t)\, \hat{\mu}(t)^2\right]\right]^{\frac{1}{2}}$$

(c) Selecting the number of parameters.

$$\Delta(L) = \sum_{t=1}^{NT} (\mu(t) - E(\hat{\mu}(t,L)))^2 , \quad L=1,3,5,.....$$

would be a suitable discrepancy on which to base the selection, except that some $M(t)$ are zero and so only approximately unbiased estimators are available. The reliability of this criterion is therefore difficult to determine.

If one is prepared to make distributional assumptions, then selection criteria are relatively easy to derive, for example based on the Kullbach-Leibler discrepancy.

A reasonable procedure is to select L for a parametric family of models and then use the same L in the estimation of $\mu(t)$.

(d) Fitting the Weibull family.

Zucchini and Adamson (1984) found the Weibull family to fit the rainfall depth models for stations in South Africa and so this family was used to model the observed rainfall amounts on days that rain was recorded.

Having estimated the mean value function $\mu(t)$ and the coefficient of variation, C, one can apply the method of moments to estimate the parameter functions of the Weibull distribution.

Denote the scale parameter by $\alpha(t)$, t=1,2,...,NT and the shape parameter by β .

Now

$$C = \{\Gamma(1+2/\beta)/\Gamma(1+1/\beta)^2 - 1\}^{\frac{1}{2}} \quad .$$

To obtain β as a function of C a rational function approximation has to be derived as no closed expression of this function is available.

The following approximation has been obtained from Zucchini and Adamson (1984):

$$\hat{\beta} = \frac{339.5410 + 148.4445\hat{C} + 192.7492\hat{C}^2 + 22.4401\hat{C}^3}{1 + 257.1162\hat{C} + 287.8362\hat{C}^2 + 157.2230\hat{C}^3} \quad .$$

Using the relationship

$$\mu(t) = \alpha(t)\Gamma(1 + 1/\beta) \quad , \qquad t=1,2,\ldots,NT$$

we obtain the estimator

$$\hat{\alpha}(t) = \hat{\mu}(t)/\Gamma(1 + 1/\hat{\beta}) \quad , \qquad t=1,2,\ldots,NT.$$

3.2 Model for climate sequences.

Little attention has been given to stochastic modelling of climatic variables such as maximum and minimum temperature, evaporation, sunshine duration, windrun, and maximum and minimum humidity. Recently, though, there have been some models purposed to stochastically simulate possible sequences of maximum and minimum temperature and solar radiation. (Goh and Tan, 1977; Nicks and Harp, 1980; Richardson, 1981; Larsen and Pense, 1982.) Bruhn et al (1980) looked at minimum relative humidity as well.

Variables such as temperature, evaporation, sunshine duration, windrun and humidity are not as difficult to model statistically as precipitation because there is not a high proportion of zero observations and the distributions of these variables are not as skewed as the rainfall distribution.

In the models that follow, because the cross-correlation between the variables is non zero, the variables are considered to be a continuous multivariate stochastic process with the parameters conditioned on the wet or dry status of the day.

3.2.1 Model 1: Multivariate model for climate data proposed by Richardson (1981).

The approach taken here to model the climate variables is the method suggested by Richardson (1981). The weather variables evaporation, windrun, maximum and minimum humidity have been added to the multivariate process and instead of modelling solar radiation, sunshine duration is modelled.

(a) Notation.

Partition the year into NT (= 365) equal intervals, denoted by $t=1,2,\ldots,NT$.

NV	is the number of variables.
NY	is the number of years observed.
R	represents the occurrence of rain.
$\bar{R}$	represents the non-occurrence of rain
$Y_{i,t}$	is the precipitation amount on period t of year i, $i=1,2,\ldots,NY$.
$S_{i,t}$	is the generic name for the observation at period t of the ith year.
$\mu_t^{\bar{R}}$	is the generic name for the mean for a dry day on period t (i.e. $Y_{i,t} = 0$).
μ_t^R	is the generic name for the mean for a wet day on period t (i.e. $Y_{i,t} > 0$).
$\sigma_t^{\bar{R}}$	is the generic name for the standard deviation for a dry day on period t.
σ_t^R	is the generic name for the standard deviation for a wet day on period t.
$X_{i,t}$	is the generic name for the residual component at period t and year i.
$\rho_0(j,k)$	is the lag 0 cross-correlation coefficient between variables j and k.
$\rho_1(j,k)$	is the lag 1 cross-correlation coefficient between variables j and k.

$\rho_1(j)$ is the lag 1 serial correlation for variable j .

(b) The model and assumptions.

Each variable is modelled in the same way. The procedure given below to model $S_{i,t}$ is carried out once for each variable to be included in the multivariate model.

The distribution of $S_{i,t}$ is seasonal and so its parameters, e.g. the mean and standard deviation, are allowed to vary seasonally. As was the case with the parameter functions of the rainfall model, it can be reasonably assumed that the parameter functions of the climate variables are smooth, periodic and sinusoidal in shape. This would again lead one to expect that they can be accurately approximated by the first few terms of their Fourier representation.

The truncated Fourier representations for the daily means and standard deviations for wet days and for dry days are given by:

$$\left.\begin{aligned} \mu_t^R &= \sum_{i=1}^{L} \alpha_i^R \varphi_i(t) \\ \sigma_t^R &= \sum_{i=1}^{L} \xi_i^R \varphi_i(t) \end{aligned}\right\} \quad \text{if} \quad Y_{i,t} > 0$$

$$\left.\begin{aligned} \mu_t^{\bar{R}} &= \sum_{i=1}^{L} \alpha_i^{\bar{R}} \varphi_i(t) \\ \sigma_t^{\bar{R}} &= \sum_{i=1}^{L} \xi_i^{\bar{R}} \varphi_i(t) \end{aligned}\right\} \quad \text{if} \quad Y_{i,t} = 0$$

$$, \qquad t=1,2,\ldots,NT$$

where

$$\varphi_i(t) = \begin{cases} \cos(\omega(t-1)i/2) , & i=2,4,\ldots,L-1 \\ \sin(\omega(t-1)(i-1)/2) , & i=3,5,\ldots,L \end{cases}$$

$$\varphi_1(t) = 1 ,$$

$$\omega = 2\pi/NT , \text{ and}$$

α_i^R , $\alpha_i^{\bar{R}}$, ξ_i^R, $\xi_i^{\bar{R}}$ are the coefficients of the respective Fourier series and, L is the order of the Fourier series approximation, i.e. we assume that for some $L < NT$

$$\mu_t = \sum_{i=1}^{L} \alpha_i \varphi_i(t) \approx \sum_{i=1}^{NT} \alpha_i \varphi_i(t)$$

and $$\xi_t = \sum_{i=1}^{L} \xi_i \varphi_i(t) \approx \sum_{i=1}^{NT} \xi_i \varphi_i(t)$$

where the above two equations hold for both wet and dry days. (Whenever R or $\bar{R}$ is omitted it means that the equation applies for both.)

The number of parameters, L , does not have to be the same in all instances, i.e. the number of parameters for the means of wet days can differ from that for dry days. The same applies for the standard deviations. To avoid complicating the notation, it will be assumed in what follows that L refers to the number of parameters of the particular parameter function under consideration.

The estimation of the Fourier coefficients will be discussed later.

The approach used by Richardson (1981) is to determine the daily means and standard deviations of each variable conditioned on the wet or dry status of each day where Fourier series is used to smooth their seasonality. The time series $S_{i,t}$ is then reduced to a time series of residual elements by removing the periodic means and standard deviations. This residual time series is given by the equations:

$$X_{i,t} = \begin{cases} \dfrac{S_{i,t} - \mu_t^{\bar{R}}}{\sigma_t^{\bar{R}}} & \text{if } Y_{i,t} = 0 \\[2ex] \dfrac{S_{i,t} - \mu_t^{R}}{\sigma_t^{R}} & \text{if } Y_{i,t} > 0 \end{cases}$$

This standardization leads to a residual series for each variable that is stationary in the mean and standard deviation with mean zero and standard deviation of unity.

The serial correlation and cross-correlation coefficients are then calculated to describe the time dependence and the interdependence (respectively) of the residual series.

The model proposed for generating residual series for each variable is the weakly stationary process suggested by Matalas (1967) given by

$$X_{i,t} = A\, X_{i,t-1} + B\, \varepsilon_{i,t} \qquad (7)$$

where $\varepsilon_{i,t}$ is a (NV x 1) matrix of independent random components that are normally distributed with mean zero and a variance of unity, i.e.

$$\varepsilon_{i,t} \sim \text{NID}(0,1) .$$

A and B are (NV x NV) matrices whose elements are defined in such a way that the sequences generated will have the desired serial correlation and cross-correlation coefficients.

This model is based on the assumption that the residuals of the variables are normally distributed and that the serial correlation of each variable may be described by a first-order linear autoregressive model.

(c) Estimation.

Firstly, a method for estimating the matrices A and B will be considered.

From the properties of the distribution of $\varepsilon_{i,t}$ and $X_{i,t}$ we have that

$$E(\varepsilon_{i,t}) = 0 \quad \text{and}$$

$$E(X_{i,t}) = E(X_{i,t-1}) = 0 \ .$$

Postmultiplying (7) through by $X_{i,t-1}^T$, the transpose of $X_{i,t-1}$, and taking expectations we have

$$E[X_{i,t} \, X_{i,t-1}^T] = A.E[X_{i,t-1} \cdot X_{i,t-1}^T] + BE[\varepsilon_{i,t} \, X_{i,t-1}^T] \tag{8}$$

Define

$$M_o = E[X_{i,t-1} \, X_{i,t-1}^T]$$

and

$$M_1 = E[X_{i,t} \; X_{i,t-1}^T] \quad .$$

M_o is an (NV x NV) matrix whose elements are the lag 0 cross-correlation coefficients and M_1 is an (NV x NV) matrix whose elements are the lag 1 cross-correlation coefficients.

The matrices may be written as

$$M_o = \begin{bmatrix} 1 & \rho_o(1,2) & \ldots\ldots\ldots & \rho_o(1,NV) \\ \rho_o(2,1) & 1 & \ldots\ldots\ldots & \rho_o(2,NV) \\ \vdots & & & \vdots \\ \rho_o(NV,1) & \ldots\ldots\ldots & \ldots\ldots\ldots & 1 \end{bmatrix}$$

and

$$M_1 = \begin{bmatrix} \rho_1(1) & \rho_1(1,2) & \ldots\ldots\ldots & \rho_1(1,NV) \\ \rho_1(2,1) & \rho_1(2) & \ldots\ldots\ldots & \rho_1(2,NV) \\ \vdots & & & \vdots \\ \rho_1(NV,1) & \rho_1(NV,2) & \ldots\ldots\ldots & \rho_1(NV) \end{bmatrix}$$

where $\rho_o(j,k)$ is the lag 0 cross-correlation coefficient between variables j and k, $\rho_1(j,k)$ is the cross-correlation coefficient between variables j and k with variable k lagged 1 day in relation to variable j, and $\rho_1(j)$ is the lag 1 serial correlation for variable j.

We can thus rewrite (8) as

$$M_1 = A\,M_0 \qquad \text{since} \quad E[\varepsilon_{i,t} \cdot \varepsilon_{i,t-1}^T] = 0 \ .$$

Since M_0 is a variance covariance matrix, it is non-singular, and therefore its inverse exists.

The matrix A is given by

$$A = M_1\,M_0^{-1} \ .$$

Postmultiplying (7) through by $x_{i,t}^T$ and taking expectations one gets

$$M_0 = A\,M_1^T + BB^T$$

since $E[e_{i,t}\,e_{i,t}^T] = I$, the identity matrix.

Therefore, the matrix B is given by the solution to

$$BB^T = M_0 - M_1\,M_0^{-1}\,M_1^T \ .$$

The Cholesky decomposition (Appendix C) can be used to solve for B .

Now, we will discuss the method to obtain parameter estimates for the coefficients of the truncated Fourier series.

The functions μ_t and σ_t are estimated using the same method but different data sets. The theory will thus be discussed for the mean function μ_t only.

Let $\bar{S}_t$ be the daily mean vector for $S_{i,t}$ and assume that it is given by the linear model

$$\bar{S}_t = x_t^T \beta + e_t \quad , \qquad t=1,2,\ldots,NT$$

with $\quad e_t \sim NID(0,\sigma_t^2) \quad .$

This is a special case of a generalized linear model because the elements $\bar{S}_t$ are independent with distributions $N(\mu_t,\sigma_t^2)$ where

$$\mu_t = x_t^T \beta \quad .$$

Also the normal distribution is a member of the exponential family (provided the σ_t^2 are regarded as known). In this case the link function, g , is the identity function, i.e.

$$g(\mu_t) = \mu_t = \sum_{i=1}^{L} \alpha_i \; \varphi_i(t) = \eta_t$$

where $\sum_{i=1}^{L} \alpha_i \; \varphi_i(t)$ represents the truncated Fourier series of the mean function μ_t , and $\varphi_i(t)$ is defined as before.

The log-likelihood function of the observed mean values as a function of the mean function μ_t is given by:

$$\ell(\mu_t;\, \bar{S}_t) = \left[\sum_{t=1}^{NT} -\bar{S}_t^2/2\sigma_t^2 + \frac{\bar{S}_t\, \mu_t}{\sigma_t^2} - \mu_t^2/2\sigma_t^2 \right.$$

$$\left. - \; 1/2 \log(2\pi\sigma_t^2) \right] \; .$$

Therefore, the log-likelihood function of the observed values as a function of the parameters $\alpha_1,\alpha_2,\ldots,\alpha_L$ is given by

$$\ell(\alpha;\bar{S}_t) = \sum_{t=1}^{NT} \left[-\bar{S}_t^2/2\sigma_t^2 + \frac{\bar{S}_t \cdot \sum_{i=1}^{L} \alpha_i \ \varphi_i(t)}{\sigma_t^2} - \frac{\sum_{i=1}^{L} \alpha_i \ \varphi_i(t)}{2\sigma_t^2} - 1/2 \log(2\pi\sigma_t^2) \right]$$

The score vector U with respect to $\alpha_1,\alpha_2,\ldots,\alpha_L$ has elements given by

$$U_j = \frac{\partial \ell(\alpha;\bar{S}_t)}{\partial \alpha_j} = \sum_{t=1}^{NT} \left[\frac{(\bar{S}_t - \mu_t)}{\sigma_t^2} \cdot \varphi_j(t) \cdot 1 \right]$$

since $E(\bar{S}_t) = \mu_t$,

$Var(\bar{S}_t) = \sigma_t^2$, and

$\partial\mu_t/\partial\eta_t = 1$.

Similarly, the information matrix $\mathscr{I}_{L \times L}$ has elements given by

$$\mathscr{I}_{jk} = \sum_{t=1}^{NT} \frac{\varphi_j(t) \cdot \varphi_k(t)}{\sigma_t^2} \cdot 1^2 .$$

The maximum likelihood estimates for $\alpha_1,\alpha_2,\ldots,\alpha_L$ are then obtained by solving the iterative equation

$$\mathscr{I}^{(m-1)}\,\hat{\alpha}^{(m)} = \mathscr{I}^{(m-1)}\,\hat{\alpha}^{(m-1)} + U^{(m-1)}$$

where m indicates the mth approximation and $\hat{\alpha}$ is the vector of estimates.

Computer packages such as GLIM and GENSTAT can be used to obtain these estimates. When the difference between successive approximations $\hat{\alpha}^{(m)}$ and $\hat{\alpha}^{(m-1)}$ is sufficiently small, $\hat{\alpha}^{(m)}$ is taken as the maximum likelihood estimate vector.

(d) Model selection.

The order of approximation, L , of the mean and standard deviation functions must be selected. Presently, no theoretical procedure is available for this.

The program package (GENSTAT) used for parameter estimation computes the percentage variance accounted by each model fitted, thus the change in percentage variance accounted for was used as an index on which one could subjectively select L . This was regarded as reasonable for the present study as it appears to give reasonable choices for the order of approximation, L .

3.2.2 Model 2: Multivariate model for climate data.

The method for modelling climate data suggested is similar to that of Richardson (1981). The time series of each variable is reduced to a residual time series by removing the seasonal mean where, once again, the seasonality is smoothed by Fourier series representation. The residual series is then assumed to follow an autoregressive process.

(a) Notation.

Partition the year into NT equal intervals, denoted by $t=1,2,\ldots,NT$.

NV	represents the number of variables.
NY	represents the number of years observed.
R	represents the occurrence of rain.
$\bar{R}$	represents the non-occurrence of rain.
$Y_{i,t}$	the precipitation amount on period t of year i , $i=1,2,\ldots,NY$.
$S_{i,t}$	is the generic name for the observation at period t of the ith year.
$\mu_t^{\bar{R}}$	is the generic name for the mean for a dry day on period t (i.e. $Y_{i,t} = 0$).
μ_t^R	is the generic name for the mean for a wet day on period t (i.e. $Y_{i,t} > 0$) .

Since all variables are modelled in the same way, the representation will be given for modelling one variable. The same procedure is then repeated for each of the remaining variables.

(b) Model and assumptions.

The model under consideration is given by:

$$S_{i,t} = \begin{cases} \mu_t^R + u_{i,t} & \text{if } Y_{i,t} > 0 \\ \mu_t^{\bar{R}} + u_{i,t} & \text{if } Y_{i,t} = 0 \end{cases} \quad , \quad \begin{array}{l} i=1,2,\ldots,NY \\ t=1,2,\ldots,NT \end{array}$$

where $u_{i,t}$ is the disturbance term at time period t of year i.

The disturbance term is assumed to be generated by an autoregressive process of order p (AR(p)) defined as

$$u_{i,t} = \Theta_1 u_{i,t-1} + \Theta_2 u_{i,t-2} + \ldots + \Theta_p u_{i,t-p} + e_{i,t}$$

where $\{e_{i,t}\}$ is a set of independent, normally distributed variables with mean zero and constant variance, σ_e^2 , i.e.

$$e_{i,t} \sim NID(0,\sigma_e^2) \quad .$$

That is, $u_{i,t}$ is regressed on past values of $u_{i,t}$ instead of on independent variables as on the classical multiple regression.

The assumption that $u_{i,t}$ is described by an autoregressive process can be substantiated by arguments put forward by Cochrane and Orcutt (1949). The sources for autocorrelation in the error term can be:

1. When modelling climatic variables, errors in modelling arise from faulty descriptions of these variables. Since these variables are themselves autocorrelated, this type of error will be autocorrelated.

2. Error terms may arise from omitting variables from the analysis because these variables are either not available or their importance is not realised or because the influence they have is so small that it is not convenient to insert them. As already indicated these variables are autocorrelated and, therefore, one may expect the resulting error terms to be also autocorrelated.

An autoregressive process of order 1, AR(1), was chosen to describe $u_{i,t}$. The reason for this choice will be discussed later. To simplify the formula from now on we will only show the theory for an AR(1) process. The order of the process can be increased to any order desired, but this has to be done at the cost of increasing both the complexity and the number of parameters to be estimated.

The form of $u_{i,t}$ is thus given by:

$$u_{i,t} = \begin{cases} \Theta^{\bar{R}} u_{i,t-1} + e^{\bar{R}}_{i,t} & \text{if } Y_{i,t} = 0 \\ \Theta^{R} u_{i,t-1} + e^{R}_{i,t} & \text{if } Y_{i,t} > 0 \end{cases}$$

The AR(1) process is sometimes called the Markov process.

This process can also be represented as

$$(1 - \Theta^{\bar{R}}B)u_{i,t} = e^{\bar{R}}_{i,t} \qquad \text{if } Y_{i,t} = 0$$

or $$(1 - \Theta^{R}B)u_{i,t} = e^{R}_{i,t} \qquad \text{if } Y_{i,t} > 0$$

where B is the backward shift operator defined as

$$B^j u_{i,t} = u_{i,t-j} .$$

Then $u_{i,t}$ can be written as a function of the random component $e_{i,t}$, i.e.

$$u_{i,t} = e_{i,t} + \Theta e_{i,t-1} + \Theta^2 e_{i,t-2} + \cdots\cdots$$

where Θ represents either Θ^R or $\Theta^{\bar{R}}$ depending on the status of day t .

(c) Estimation.

The first step is to estimate μ^R_t and $\mu^{\bar{R}}_t$ the seasonal mean of $S_{i,t}$. This is done exactly in the same way as when estimating μ^R_t and $\mu^{\bar{R}}_t$ in the previous model so it will not be discussed again. Fourier series are used to approximate both μ^R_t and $\mu^{\bar{R}}_t$.

Once we have $\hat{\mu}^R_t$ and $\hat{\mu}^{\bar{R}}_t$ the time series $S_{i,t}$ is reduced to a residual time series $u_{i,t}$ by subtracting the seasonal mean, i.e.

$$u_{i,t} = \begin{cases} S_{i,t} - \hat{\mu}^{\bar{R}}_t & \text{if } Y_{i,t} = 0 \\ S_{i,t} - \hat{\mu}^{R}_t & \text{if } Y_{i,t} > 0 \end{cases}$$

Thus the time series $u_{i,t}$ has zero mean and, as stated before, is described by an AR(1) process.

One now needs to estimate the parameters Θ^R and $\Theta^{\bar{R}}$.

Denote

$N(\bar{R})$ to be the set of time periods t such that it did not rain in period t .

$N(R)$ to be the set of time periods t such that it rained on period t , $t=1,2,\ldots,T$, T being the total number of observations.

To estimate Θ^R and $\Theta^{\bar{R}}$ by the method of least squares, the following two equations must be minimized, i.e.

$$G_1 = \sum_{t \in N(\bar{R})} [u_{i,t} - \Theta^{\bar{R}} u_{i,t-1}]^2 \qquad \text{with respect to } \Theta^{\bar{R}}$$

and

$$G_2 = \sum_{t \in N(R)} [u_{i,t} - \Theta^{R} u_{i,t-1}]^2 \qquad \text{with respect to } \Theta^{R}$$

where $u_{i,o}$ is defined to be zero (alternatively consider $u_{i,1}$ to be fixed and the summation only considers $t=2,\ldots,T$.)

To minimize G_2 with respect to Θ^R the first derivative is set to zero, i.e.

$$dG_2/d\Theta^R = 0$$

i.e.

$$2 \sum_{t \in N(R)} [u_{i,t} - \Theta^R u_{i,t-1}](-u_{i,t-1}) = 0$$

and so

$$\hat{\Theta}^R = \frac{\sum_{t \in N(R)} u_{i,t} \cdot u_{i,t-1}}{\sum_{t \in N(R)} u^2_{i,t-1}} \quad .$$

The autocovariance function of lag k is defined as

$$\gamma(k) = E[u_{i,t} \cdot u_{i,t+k}]$$

and is estimated by

$$C_k = \sum_{t=1}^{N-k} (u_{i,t})(u_{i,t+k})/N \quad .$$

The autocorrelation function of lag k is defined as

$$\rho(k) = \gamma(k)/\gamma(0)$$

and is estimated by

$$R_k = C_k/C_o \quad .$$

Therefore, by taking the appropriate sets in the summations we get that

$$\hat{\Theta}^R \approx C_1^R/C_o^R = R_1^R$$

This approximate estimator for $\hat{\Theta}^R$ is appealing since R_1^R is an estimator for $\rho^R(1)$ and $\rho^R(1) = \Theta^R$ for a first-order AR process. Furthermore this estimate is particularly easy to compute. As the autocorrelation function is in any case needed for model identification, no additional computation is required to compute $\hat{\Theta}^R$.

Similarly we get

$$\hat{\Theta}^{\bar{R}} \approx C_1^{\bar{R}}/C_o^{\bar{R}} = R_1^{\bar{R}} .$$

The time series $e_{i,t}$ can now be obtained by

$$e_{i,t} = \begin{cases} u_{i,t} - \hat{\Theta}^{\bar{R}} u_{i,t-1} & \text{if } Y_{i,t} = 0 \\ u_{i,t} - \hat{\Theta}^{R} u_{i,t-1} . & \text{if } Y_{i,t} > 0 . \end{cases}$$

An estimate for σ_e^2 can now be obtained using the method of maximum likelihood since

$$e_{i,t} \sim NID(0,\sigma_e^2)$$

where the standard deviation of $e_{i,t}$ is conditioned on the wet and dry status of the day.

An estimate for $\sigma_e^2(R)$ is given by

$$\hat{\sigma}_e^2(R) = 1/C(R) . \sum_{t \in N(R)} (e_{i,t})^2 .$$

Similarly

$$\hat{\sigma}_e^2(\bar{R}) = 1/C(\bar{R}) \sum_{t \in N(\bar{R})} (e_{i,t})^2$$

where

$C(R)$	denotes the number of elements in the set $N(R)$ and
$C(\bar{R})$	denotes the number of elements in the set $N(\bar{R})$.

To generate data, the covariance matrix of the residuals of the different variables is needed to account for the cross-correlation between the variables.

The covariance matrix can be obtained by pre and post multiplying the correlation matrix by a matrix whose diagonal elements are the standard deviation of the variables and whose off-diagonal elements are zeros.

The correlation matrix, $\hat{\Sigma},(NV \times NV)$ has elements given by

$$R_{jk} = \frac{1/T \cdot \sum_{i=1}^{NY} \sum_{t=1}^{NT} e_{i,t}^{(j)} \cdot e_{i,t}^{(k)} - 1/T^2 \cdot \left(\sum_{i=1}^{NY} \sum_{t=1}^{NT} e_{i,t}^{(j)} \right) \left(\sum_{i=1}^{NY} \sum_{t=1}^{NT} e_{i,t}^{(k)} \right)}{\left[1/T \sum_{i=1}^{NY} \sum_{t=1}^{NT} (e_{i,t}^{(j)})^2 - 1/T^2 \left(\sum_{i=1}^{NY} \sum_{t=1}^{NT} e_{i,t}^{(j)} \right)^2 \right]^{\frac{1}{2}} \cdot \left[1/T \sum_{i=1}^{NY} \sum_{t=1}^{NT} (e_{i,t}^{(k)})^2 - 1/T^2 \left(\sum_{i=1}^{NY} \sum_{t=1}^{NT} e_{i,t}^{(k)} \right)^2 \right]^{\frac{1}{2}}}$$

where

R_{jk}	is conditioned on the wet and dry status of the day.
$e_{i,t}^{(j)}$	denotes the residual time series of variable j , $j=1,2,\ldots,NV$.
$e_{i,t}^{(k)}$	denotes the residual time series of variable k , $k=1,2,\ldots,NV$.

(d) Model selection.

The model selection procedure already discussed is used to select the order of the Fourier series transformation and so, we will not discuss it again.

To determine the order of an autoregressive process one can look at the properties of its autocorrelation function, e.g. for an AR(1) process the theoretical autocorrelation function decreases exponentially and the sample function should have a similar shape. When the process is of a higher order though, it becomes difficult to assess the order of the process from the autocorrelation function.

One approach is to look at the partial autocorrelation function (Box and Jenkins, 1970) defined as follows. Denote the last coefficient, Θ_p, when fitting an AR(p) process by π_p. This measures the excess correlation at lag p which is not accounted for by an AR(p-1) process. It is called the pth partial autocorrelation coefficient and when plotted against p gives the partial autocorrelation function. Values of π_p which lie outside the range of $\pm 2/\sqrt{N}$, where N denotes the length of time series, are significantly different from zero at the 5% level. The order of the process is chosen to be that value of p beyond which the sample values of $\{\pi_j\}$ are not significantly different from zero.

Another criterion is given by minimizing Akaike's information criterion (AIC), where

AIC = $-2\ln$ (maximum likelihood) + L (number of independent parameters estimated).

The multivariate time series model described above has a quite large number of parameters all of which have to be estimated from the data. The inclusion of additional parameters not only increases the complexity of the model but also results in an increase in the discrepancy due to estimation. In view of the complexity of the model, it is particularly important to keep the number of parameters down to a minimum. This must be kept in mind when deciding on the order of the autoregressive process imbedded in the model.

3.2.3 Model 3 (ML): Multivariate model for climate data.

The two previous models condition the parameters of the model on the wet or dry status of the day. When generating climate sequences these models represent conditions in which a wet day follows a wet day and a dry day follows a dry day but fail to explain the relationship between conditions such as a wet day following a dry day or a dry day following a wet day.

To generate representative climate sequences, these sequences must be related to the sequences of rain and no-rain days. To achieve this relationship, the parameters of the model must be conditioned on the four possible sequences in the rainfall variable:

1. a dry day follows a dry day,
2. a wet day follows a wet day,
3. a wet day follows a dry day,
4. a dry day follows a wet day.

The method of modelling the climate variables is the same as that of the previous model, only now there are more states to consider.

(a) Notation.

Again partition the year into NT equal intervals, denoted by t=1,2,...,NT.

NV	represents the number of variables.
NY	represents the number of years observed.
W	represents the occurrence of rain.
D	represents the non-occurrence of rain.
$Y_{i,t}$	is the precipitation amount on period t of the ith year (i=1,2,...,NY).
$S_{i,t}$	is the generic name for the observation at period t of the ith year.
μ_t^D	is the generic name for the mean for a dry day on period t (i.e. $Y_{i,t} = 0$).
μ_t^W	is the generic name for the mean for a wet day on period t (i.e. $Y_{i,t} > 0$).
DD	sequence when day t-1 had no rain and day t had no rain.
WW	sequence when day t-1 had rain and day t had rain.
DW	sequence when day t-1 had no rain and day t had rain.
WD	sequence when day t-1 had rain and day t had no rain.

Θ^{DD}	the coefficient of the AR(1) process for $u_{i,t}$ given sequence DD .
Θ^{WW}	the coefficient of the AR(1) process for $u_{i,t}$ given sequence WW .
Θ^{DW}	the coefficient of the AR(1) process for $u_{i,t}$ given sequence DW .
Θ^{WD}	the coefficient of the AR(1) process for $u_{i,t}$ given sequence WD .
T	represents the total number of observations, i.e. NT . NY
N(DD)	is the set of time periods t such that period t was dry and period t-1 was dry, t=1,2,...,T .
N(WW)	is the set of time periods t such that period t was wet and period t-1 was wet.
N(DW)	is the set of time periods t such that period t was wet and period t-1 was dry.
N(WD)	is the set of time periods t such that period t was dry and period t-1 was wet.
C(DD)	is the number of elements in the set N(DD).
C(WW)	is the number of elements in the set N(WW).
C(DW)	is the number of elements in the set N(DW).
C(WD)	is the number of elements in the set N(WD).

Then $T = C(DD) + C(WW) + C(DW) + C(WD).$

The previous convention of representing the non-occurrence of rain by "$\bar{R}$" becomes very cumbersome in the theory that follows, thus to simplify typography, the notation "$\bar{R}$" is replaced by "D" (dry) and R is replaced by "W" (wet).

(b) Model and assumptions.

The model considered is given by:

$$S_{i,t} = \mu_t + u_{i,t}$$

where $u_{i,t} = \Theta u_{i,t-1} + e_{i,t}$

and $e_{i,t} \sim NID(0,\sigma_e^2)$

$i=1,2,\ldots,NY$

$t=1,2,\ldots,NT$.

The above model may be rewritten as

$$S_{i,t} = \mu_t + \Theta(S_{i,t-1} - \mu_{t-1}) + e_{i,t}$$

$i=1,2,\ldots,NY$

$t=1,2,\ldots,NT$.

Therefore, the model that incorporates the different rain sequences is given by:

$$S_{i,t} = \mu_t^D + \Theta^{DD}(S_{i,t-1} - \mu_{t-1}^D) + e_{i,t}$$ if day t-1 was dry and day t was dry

or

$$S_{i,t} = \mu_t^W + \Theta^{WW}(S_{i,t-1} - \mu_{t-1}^W) + e_{i,t}$$ if day t-1 was wet and day t was wet

or

$$S_{i,t} = \mu_t^W + \Theta^{DW}(S_{i,t-1} - \mu_{t-1}^D) + e_{i,t} \quad \text{if day } t-1 \text{ was dry and day } t \text{ was wet}$$

or

$$S_{i,t} = \mu_t^D + \Theta^{WD}(S_{i,t-1} - \mu_{t-1}^W) + e_{i,t} \quad \text{if day } t-1 \text{ was wet and day } t \text{ was dry.}$$

The assumption that the disturbance term $u_{i,t}$ follows an autoregressive process of order 1 is made. (See later for model selection.)

The distribution of the random variable $e_{i,t}$ is given by

$$e_{i,t} \sim NID(0,\sigma_e^2(DD)) \qquad \text{given sequence DD.}$$

$$e_{i,t} \sim NID(0,\sigma_e^2(WW)) \qquad \text{given sequence WW.}$$

$$e_{i,t} \sim NID(0,\sigma_e^2(DW)) \qquad \text{given sequence DW.}$$

$$e_{i,t} \sim NID(0,\sigma_e^2(WD)) \qquad \text{given sequence WD.}$$

In the above equations for $S_{i,t}$ the parameter μ_t was only conditioned on the rain and no-rain status of the day. This was done to simplify the model otherwise for each sequence one would have two equations for $S_{i,t}$ e.g.

if sequence DD has occurred then

$$S_{i,t} = \mu_t^{DD} + \Theta^{DD}(S_{i,t-1} - \mu_{t-1}^{DD}) + e_{i,t}$$

if the sequence DDD was observed

or

$$S_{i,t} = \mu_t^{DD} + \Theta^{DD}(S_{i,t-1} - \mu_{t-1}^{WD}) + e_{i,t}$$

if the sequence WDD was observed.

This not only increases the number of parameters to be estimated but in addition we are no longer assuming that all the information we need of previous values of the model is given by the value of the previous day. The state of the second previous day is also required.

As already discussed, it is reasonable to approximate the mean function μ_t by its truncated Fourier representation, i.e.

$$\mu_t^D = \sum_{i=1}^{L} \alpha_i^D \varphi_i(t) \qquad \text{if } t \text{ dry}$$

and

$$\mu_t^W = \sum_{i=1}^{L} \alpha_i^W \varphi_i(t) \qquad \text{if } t \text{ wet}$$

where $\varphi_i(t)$ is defined as before and L is the order of the Fourier series approximation.

(c) Estimation by maximum likelihood.

The procedure to estimate the parameters α_j^D, α_j^W, Θ^{DD}, Θ^{WW}, Θ^{DW}, Θ^{WD}, $\sigma_e^2(DD)$, $\sigma_e^2(WW)$, $\sigma_e^2(DW)$ and $\sigma_e^2(WD)$; $j=1,2,\ldots,L$ is now discussed.

Maximum likelihood estimates can be obtained by observing that since

$$e_{i,t} \sim NID(0,\sigma_e^2)$$

then

$$f(e_{i,t}) = 1/\sqrt{2\pi}\,\sigma_e \;.\; \exp(-1/2\sigma_e^2 \;.\; (e_{i,t})^2) \;.$$

Therefore the joint likelihood function conditioned on the four different sequences is given by:

$$L(\psi) = L(\alpha_j^D,\ \alpha_j^W,\ \Theta^{DD},\ \Theta^{WW},\ \Theta^{DW},\ \Theta^{WD},\ \sigma_e^2(DD),\ \sigma_e^2(WW),\ \sigma_e^2(DW),$$

$$\sigma_e^2(WD);\ e_{i,t})$$

$$= \prod_{t\in N(DD)} f(e_{i,t}/DD) \;.\; \prod_{t\in N(WW)} f(e_{i,t}/WW) \;.$$

$$\prod_{t\in N(DW)} f(e_{i,t}/DW) \;.\; \prod_{t\in N(WD)} f(e_{i,t}/WD)$$

where $f(e_{i,t}/DD)$ represents the density function of $e_{i,t}$ given that the sequence DD has been observed, i.e.

$$f(e_{i,t}/DD) = 1/\sqrt{2\pi}\,\sigma_e(DD) \;.\; \exp\,(-1/2\sigma_e^2(DD) \;.\; (e_{i,t})^2) \;.$$

where $t\in N(DD)$.

Then

$$L(\psi) = (1/\sqrt{2\pi})^T \;.\; (1/\sigma_e(DD))^{C(DD)} \;.\; (1/\sigma_e(WW))^{C(WW)} \;.$$

$$(1/\sigma_e(DW))^{C(DW)} \;.\; (1/\sigma_e(WD))^{C(WD)} \;.$$

$$\exp\left\{-1/2\left[1/\sigma_e^2(DD) \;.\; \sum_{t\in N(DD)} (e_{i,t})^2 + 1/\sigma_e^2(WW) \;.\right.\right.$$

$$\sum_{t\in N(WW)} (e_{i,t})^2 + 1/\sigma_e^2(DW) \cdot \sum_{t\in N(DW)} (e_{i,t})^2$$

$$\left. + \; 1/\sigma_e^2(WD) \cdot \sum_{t\in N(WD)} (e_{i,t})^2 \right] \Bigg\}$$

One now makes the following transformation:

$$e_{i,t} = S_{i,t} - \mu_t - \Theta(S_{i,t-1} - \mu_{t-1}) \quad .$$

The Jacobian of the transformation is given by

$$\left|\frac{\partial e_{i,t}}{\partial S_{i,p}}\right| = \begin{vmatrix} 1 & 0 & 0 & \cdots\cdots & 0 \\ -\Theta & 1 & 0 & \cdots\cdots & 0 \\ 0 & -\Theta & 1 & \cdots\cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots \; -\Theta & 1 \end{vmatrix} .$$

$$= \prod_{i=1}^{T} 1 = 1$$ since we are dealing with a triangular matrix.

Let

$$\beta(DD) = S_{i,t} - \Theta^{DD} S_{i,t-1} - \mu_t^D + \Theta^{DD} \mu_{t-1}^D$$

$$\beta(WW) = S_{i,t} - \Theta^{WW} S_{i,t-1} - \mu_t^W + \Theta^{WW} \mu_{t-1}^W$$

$$\beta(DW) = S_{i,t} - \Theta^{DW} S_{i,t-1} - \mu_t^W + \Theta^{DW} \mu_{t-1}^D$$

$$\beta(WD) = S_{i,t} - \Theta^{WD} S_{i,t-1} - \mu_t^D + \Theta^{WD} \cdot \mu_{t-1}^W \quad .$$

Then the joint probability density function of $S_{i,t}$ is given by:

$$(1/\sqrt{2\pi})^T \cdot (1/\sigma_e^2(DD))^{C(DD)} \cdot (1/\sigma_e^2(WW))^{C(WW)} \cdot (1/\sigma_e^2(DW))^{C(DW)}$$

$$(1/\sigma_e^2(WD))^{C(WD)} \cdot \exp\left\{-1/2\left[1/\sigma_e^2(DD) \sum_{t\in N(DD)} (\beta(DD))^2\right.\right.$$

$$+ 1/\sigma_e^2(WW) \sum_{t\in N(WW)} (\beta(WW))^2 + 1/\sigma_e^2(DW) \sum_{t\in N(DW)} (\beta(DW))^2$$

$$\left.\left.+ 1/\sigma_e^2(WD) \sum_{t\in N(WD)} (\beta(WD))^2\right]\right\} \quad .$$

Therefore the log likelihood function is given by:

$$\ell(\psi) = -T \log(\sqrt{2\pi}) - C(DD)/2 \log \sigma_e^2(DD) -$$

$$C(WW)/2 \log \sigma_e^2(WW) - C(DW)/2 \log \sigma_e^2(DW) -$$

$$C(WD)/2 \log \sigma_e^2(WD) - 1/2\left[1/\sigma_e^2(DD) \sum_{t\in N(DD)} \beta^2(DD)\right.$$

$$+ \quad 1/\sigma_e^2(WW) \sum_{t\in N(WW)} \beta^2(WW) + 1/\sigma_e^2(DW) \sum_{t\in N(DW)}$$

$$\left.\beta^2(DW) + 1/\sigma_e^2(WD) \sum_{t\in N(WD)} \beta^2(WD)\right]$$

To find the maximum likelihood estimates, differentiate $\ell(\psi)$ with respect to the different parameters and set the derivatives equal to zero. These results are given in Appendix D, here one merely states the parameter estimates. These are given by

$$\hat{\Theta}^{DD} = \frac{\sum_{t\in N(DD)} (\hat{\mu}_t^D - S_{i,t})(\hat{\mu}_{t-1}^D - S_{i,t-1})}{\sum_{t\in N(DD)} (\hat{\mu}_{t-1}^D - S_{i,t-1})^2}$$

$$\hat{\Theta}^{WW} = \frac{\sum_{t\in N(WW)} (\hat{\mu}_t^W - S_{i,t})(\hat{\mu}_{t-1}^W - S_{i,t-1})}{\sum_{t\in N(WW)} (\hat{\mu}_{t-1}^W - S_{i,t-1})^2}$$

$$\hat{\Theta}^{DW} = \frac{\sum_{t\in N(DW)} (\hat{\mu}_t^W - S_{i,t})(\hat{\mu}_{t-1}^D - S_{i,t-1})}{\sum_{t\in N(DW)} (\hat{\mu}_{t-1}^D - S_{i,t-1})^2}$$

$$\hat{\Theta}^{WD} = \frac{\sum_{t\in N(WD)} (\hat{\mu}_t^D - S_{i,t})(\hat{\mu}_{t-1}^W - S_{i,t-1})}{\sum_{t\in N(WD)} (\hat{\mu}_{t-1}^W - S_{i,t-1})^2}$$

$$\hat{\sigma}_e^2(DD) = 1/C(DD) \sum_{t\in N(DD)} \hat{\beta}^2(DD)$$

$$\hat{\sigma}_e^2(WW) = 1/C(WW) \sum_{t\in N(WW)} \hat{\beta}^2(WW)$$

$$\hat{\sigma}_e^2(DW) = 1/C(DW) \sum_{t\in N(DW)} \hat{\beta}^2(DW)$$

$$\hat{\sigma}_e^2(WD) = 1/C(WD) \sum_{t\in N(WD)} \hat{\beta}^2(WD)$$

$$\hat{\alpha}_j^D = \left\{ 1/\hat{\sigma}_e^2(DD) \sum_{t \in N(DD)} \varphi_j(t)^2 - 2\, 1/\hat{\sigma}_e^2(DD)\, \hat{\Theta}^{DD} \,.\right.$$

$$\sum_{t \in N(DD)} \varphi_j(t) \,.\, \varphi_j(t-1) + 1/\hat{\sigma}_e^2(DD)\, (\hat{\Theta}^{DD})^2 \sum_{t \in N(DD)} \varphi_j(t-1)^2$$

$$+ 1/\hat{\sigma}_e^2(DW)\, (\hat{\Theta}^{DW})^2 \sum_{t \in N(DW)} \varphi_j(t-1)^2 + 1/\hat{\sigma}_e^2(WD) \,.$$

$$\left.\sum_{t \in N(WD)} \varphi_j(t)^2 \right\}^{-1} . \left[-A \;\; -M\right]$$

where

$$A = -1/\hat{\sigma}_e^2(DD) \,.\, \sum_{t \in N(DD)} (S_{i,t} - \hat{\Theta}^{DD} \,.\, S_{i,t-1}) \,.\, \varphi_j(t)$$

$$+ 1/\hat{\sigma}_e^2(DD) \,.\, \hat{\Theta}^{DD} \sum_{t \in N(DD)} (S_{i,t} - \hat{\Theta}^{DD} S_{i,t-1})\, \varphi_j(t-1)$$

$$+ 1/\hat{\sigma}_e^2(DW) \,.\, \hat{\Theta}^{DW} \sum_{t \in N(DW)} (S_{i,t} - \hat{\Theta}^{DW} S_{i,t-1})\, \varphi_j(t-1)$$

$$- 1/\hat{\sigma}_e^2(WD) \sum_{t \in N(WD)} (S_{i,t} - \hat{\Theta}^{WD} S_{i,t-1})\, \varphi_j(t) -$$

$$1/\hat{\sigma}_e^2(DW)\, \hat{\Theta}^{DW} \,.\, \sum_{t \in N(DW)} \hat{\mu}_t^W \varphi_j(t-1) -$$

$$1/\hat{\sigma}_e^2(WD)\, \hat{\Theta}^{WD} \sum_{t \in N(WD)} \hat{\mu}_{t-1}^W \varphi_j(t)$$

and

$$M = -1/\hat{\sigma}_e^2(DD)\ \hat{\Theta}^{DD} \sum_{t\in N(DD)} \left(\sum_{\substack{k=1\\k\neq j}}^{L} \hat{\alpha}_k^D\ \varphi_k(t-1)\right) \cdot \varphi_j(t)$$

$$- 1/\hat{\sigma}_e^2(DD)\ \hat{\Theta}^{DD} \cdot \sum_{t\in N(DD)} \left(\sum_{\substack{k=1\\k\neq j}}^{L} \hat{\alpha}_k^D\ \varphi_k(t)\right) \cdot \varphi_j(t-1)$$

$$+ 1/\hat{\sigma}_e^2(DD) \cdot \sum_{t\in N(DD)} \left(\sum_{\substack{k=1\\k\neq j}}^{L} \hat{\alpha}_k^D\ \varphi_k(t)\right) \varphi_j(t) +$$

$$1/\hat{\sigma}_e^2(DD) \cdot (\hat{\Theta}^{DD})^2 \cdot \sum_{t\in N(DD)} \left(\sum_{\substack{k=1\\k\neq j}}^{L} \hat{\alpha}_k^D\ \varphi_k(t-1)\right) \cdot \varphi_j(t-1)$$

$$+ 1/\hat{\sigma}_e^2(DW) \cdot (\hat{\Theta}^{DW})^2 \cdot \sum_{t\in N(DW)} \left(\sum_{\substack{k=1\\k\neq j}}^{L} \hat{\alpha}_k^D\ \varphi_k(t-1)\right) \cdot \varphi_j(t-1) +$$

$$1/\hat{\sigma}_e^2(WD) \sum_{t\in N(WD)} \left(\sum_{\substack{k=1\\k\neq j}}^{L} \hat{\alpha}_k^D\ \varphi_k(t)\right) \varphi_j(t) .$$

The estimate for α_j^W is given by:

$$\hat{\alpha}_j^W = \left\{ 1/\hat{\sigma}_e^2(WW) \sum_{t\in N(WW)} (\varphi_j(t))^2 - 2\ 1/\hat{\sigma}_e^2(WW)\ \hat{\Theta}^{WW} \cdot \right.$$

$$\sum_{t\in N(WW)} \varphi_j(t-1)\ \varphi_j(t) + 1/\hat{\sigma}_e^2(WW) \cdot (\hat{\Theta}^{WW})^2 \cdot$$

$$\sum_{t\in N(WW)} (\varphi_j(t-1))^2 + 1/\hat{\sigma}_e^2(DW) \cdot \sum_{t\in N(DW)} (\varphi_j(t))^2 +$$

$$1/\hat{\sigma}_e^2(WD)\ (\hat{\Theta}^{WD})^2 \sum_{t\in N(WD)} (\varphi_j(t-1))^2\Big\}^{-1} \Big[-A_2 - M_2\Big]$$

where

$$A_2 = -1/\hat{\sigma}_e^2(WW) \cdot \sum_{t\in N(WW)} (S_{i,t} - \hat{\Theta}^{WW} S_{i,t-1})\ \varphi_j(t)$$

$$+ 1/\hat{\sigma}_e^2(WW) \cdot \hat{\Theta}^{WW} \cdot \sum_{t\in N(WW)} (S_{i,t} - \hat{\Theta}^{WW} S_{i,t-1})\ \varphi_j(t-1)$$

$$- 1/\hat{\sigma}_e^2(DW) \sum_{t\in N(DW)} (S_{i,t} - \hat{\Theta}^{DW} S_{i,t-1})\ \varphi_j(t)$$

$$- 1/\hat{\sigma}_e^2(DW)\ \hat{\Theta}^{DW} \cdot \sum_{t\in N(DW)} \hat{\mu}^D_{t-1}\ \varphi_j(t) + 1/\hat{\sigma}_e^2(WD)\ \hat{\Theta}^{WD} \cdot$$

$$\sum_{t\in N(WD)} (S_{i,t} - \hat{\Theta}^{WD} S_{i,t-1})\ \varphi_j(t-1) - 1/\hat{\sigma}_e^2(WD)\ \hat{\Theta}^{WD} \cdot$$

$$\sum_{t\in N(WD)} (\hat{\mu}^D_t)\ \varphi_j(t-1)$$

and

$$M_2 = 1/\hat{\sigma}_e^2(WW) \cdot \sum_{t\in N(WW)} \Big(\sum_{\substack{k=1\\k\neq j}}^{L} \hat{\alpha}^W_k\, \varphi_k(t)\Big)\ \varphi_j(t)\ -$$

$$1/\hat{\sigma}_e^2(WW)\ \hat{\Theta}^{WW} \Bigg[\sum_{t\in N(WW)} \Big(\sum_{\substack{k=1\\k\neq j}}^{L} \hat{\alpha}^W_k\, \varphi_k(t-1)\Big)\ \varphi_j(t)\ +$$

$$\sum_{t\in N(WW)} \Big(\sum_{\substack{k=1\\k\neq j}}^{L} \hat{\alpha}^W_k\, \varphi_k(t)\Big)\ \varphi_j(t-1)\Bigg] + 1/\hat{\sigma}_e^2(WW)\ (\hat{\Theta}^{WW})^2 \ .$$

$$\sum_{t \in N(WW)} \left(\sum_{\substack{k=1 \\ k \neq j}}^{L} \hat{\alpha}_k^W \varphi_k(t-1) \right) \varphi_j(t-1) + 1/\hat{\sigma}_e^2(DW) \; .$$

$$\sum_{t \in N(DW)} \left(\sum_{\substack{k=1 \\ k \neq j}}^{L} \hat{\alpha}_k^W \varphi_k(t) \right) \varphi_j(t) + 1/\hat{\sigma}_e^2(WD) \, (\hat{\Theta}^{WD})^2 \; .$$

$$\sum_{t \in N(WD)} \left(\sum_{\substack{k=1 \\ k \neq j}}^{L} \hat{\alpha}_k^W \varphi_k(t-1) \right) \varphi_j(t-1) \; .$$

These equations cannot be solved explicitly and therefore have to be solved iteratively.

Note that

μ_t is a function of the α_js where the α_js are functions of Θ and σ^2,

Θ is a function of μ_t ,

and σ^2 is a function of μ_t and Θ .

The following algorithm is carried out to estimate the parameters.

Algorithm.

Step 1: Estimate initial $\hat{\mu}_t$ by approximating by its Fourier series transformation and estimating the parameters α_i by the method mentioned in the previous two models.

Step 2: Estimate Θ .

<u>Step 3</u>: Estimate σ^2 .

<u>Step 4</u>: Estimate μ_t .

<u>Step 5</u>: Check for convergence of all parameters, i.e. when the percentage change in parameter estimates is sufficiently small. If convergence is not met, go back to Step 2.

To model the multivariate time series we again need the covariance matrix of the residuals of the different variables.

The autocorrelation matrix is obtained and then the covariance matrix can be computed from it.

The autocorrelation matrix, $\hat{\Sigma}$, has elements given by R_{jk} where R_{jk} is conditioned on the four sequences of possible rain - no-rain outcomes, e.g. for sequence DD .

$$R_{jk}^{DD} = 1/C(DD) \sum_{t \in N(DD)} e_{i,t}^{(j)} \;.\; e_{i,t}^{(k)} - 1/(C(DD))^2 \;.$$

$$\frac{\left(\sum_{t \in N(DD)} e_{i,t}^{(j)}\right)\left(\sum_{t \in N(DD)} e_{i,t}^{(k)}\right)}{\left[1/C(DD) \sum_{t \in N(DD)} (e_{i,t}^{(j)})^2 - 1/(C(DD))^2 \left(\sum_{t \in N(DD)} e_{i,t}^{(j)}\right)^2\right]^{\frac{1}{2}} \;.\; \left[1/C(DD) \sum_{t \in N(DD)} (e_{i,t}^{(k)})^2 - 1/(C(DD))^2 \left(\sum_{t \in N(DD)} e_{i,t}^{(k)}\right)^2\right]^{\frac{1}{2}}}$$

where

$e_{i,t}^{(j)}$ denotes the residual time series of variable j ; $j=1,2,\ldots,NV$.

$e_{i,t}^{(k)}$ denotes the residual time series of variable k ; $k=1,2,\ldots,NV$.

The elements R_{jk}^{WW} , R_{jK}^{DW} and R_{jk}^{WD} are defined similarly.

(d) Model selection.

The order of the autoregressive process for $u_{i,t}$ is chosen in the same way as in the previous model as is the order of the Fourier series approximation.

3.2.4 Model 3 (LS): Multivariate model for climate data.

It was decided, after generating climate sequences using the maximum likelihood estimates of the parameters of the previous model, to fit the same model but to estimate the parameters by the method of ordinary least squares to be able to circumvent the assumption of normality.

(a) Estimation by ordinary least squares.

Given that the model is given by

$$S_{i,t} = \mu_t + \Theta(S_{i,t-1} - \mu_{t-1}) + e_{i,t}$$

then the estimates are obtained by minimizing

$$P = \sum_{i=1}^{NY} \sum_{t=1}^{NT} e_{i,t}^2 = \sum_{i=1}^{NY} \sum_{t=1}^{NT} (S_{i,t} - \mu_t - \Theta(S_{i,t-1} - \mu_{t-1}))^2$$

where the summation is partitioned into the four sequences, i.e.

$$P = \sum_{t\in N(DD)} e_{i,t}^2(DD) + \sum_{t\in N(WW)} e_{i,t}^2(WW) + \sum_{t\in N(DW)} e_{i,t}^2(DW) + \sum_{t\in N(WD)} e_{i,t}^2(WD)$$

where

$e_{i,t}(DD)$ represents the residual series given by the sequence DD , i.e. $e_{i,t}(DD) = S_{i,t} - \mu_t^D - \Theta^{DD}(S_{i,t-1} - \mu_{t-1}^D)$.

Similarly for $e_{i,t}(WW)$, $e_{i,t}(DW)$ and $e_{i,t}(WD)$.

By the same reason as previously, the transformation

$$e_{i,t} = S_{i,t} - \mu_t - \Theta(S_{i,t-1} - \mu_{t-1})$$

can be performed and we obtain

$$P = \sum_{t\in N(DD)} \beta^2(DD) + \sum_{t\in N(WW)} \beta^2(WW) + \sum_{t\in N(DW)} \beta^2(DW) + \sum_{t\in N(WD)} \beta^2(WD) .$$

Now P is minimized with respect to Θ and μ_t by setting the first derivatives with respect to Θ and μ_t equal to zero respectively. This is given in Appendix E.

The estimates obtained for Θ^{DD}, Θ^{WW}, Θ^{DW} and Θ^{WD} using the ordinary least squares are equal to those obtained using the method of maximum likelihood estimation.

The estimates for α_j^D and α_j^W are given by:

$$\hat{\alpha}_j^D = \left\{ \sum_{t\in N(DD)} \varphi_j^2(t) - 2\hat{\Theta}^{DD} \sum_{t\in N(DD)} \varphi_j(t-1)\,\varphi_j(t) + (\hat{\Theta}^{DD})^2 \sum_{t\in N(DD)} \varphi_j^2(t-1) + (\hat{\Theta}^{DW})^2 \sum_{t\in N(DW)} \varphi_j^2(t-1) + \sum_{t\in N(WD)} \varphi_j^2(t) \right\}^{-1} \cdot \left[\sum_{t\in N(DD)} \left((S_{i,t} - \hat{\Theta}^{DD} S_{i,t-1}) - \left(\sum_{\substack{k=1\\k\neq j}}^{L} \hat{\alpha}_k^D \varphi_k(t) - \hat{\theta}^{DD} \left(\sum_{\substack{k=1\\k\neq j}}^{L} \hat{\alpha}_k^D \varphi_k(t-1) \right) \right) \right) \left(\varphi_j(t) - \hat{\Theta}^{DD} \varphi_j(t-1) \right) - \hat{\Theta}^{DW} \sum_{t\in N(DW)} \left[(S_{i,t} - \hat{\mu}_t^W - \hat{\Theta}^{DW} S_{i,t-1}) + \hat{\Theta}^{DW} \left(\sum_{\substack{k=1\\k\neq j}}^{L} \hat{\alpha}_k^D \varphi_k(t-1) \right) \right] \varphi_j(t-1) + \sum_{t\in N(WD)} \left[(S_{i,t} - \hat{\Theta}^{WD} S_{i,t-1} + \hat{\Theta}^{WD} \hat{\mu}_{t-1}^W) - \left(\sum_{\substack{k=1\\k\neq j}}^{L} \hat{\alpha}_k^D \varphi_k(t) \right) \right] \varphi_j(t) \right]$$

and

$$\hat{\alpha}_j^W = \left\{ \sum_{t\in N(WW)} \varphi_j^2(t) - 2\hat{\Theta}^{WW} \sum_{t\in N(WW)} \varphi_j(t-1)\,\varphi_j(t) + \right.$$

$$(\hat{\Theta}^{WW})^2 \sum_{t\in N(WW)} \varphi_j(t-1)^2 + \sum_{t\in N(DW)} \varphi_j^2(t) +$$

$$\left.(\hat{\Theta}^{WD})^2 \sum_{t\in N(WD)} \varphi_j^2(t-1)\right\}^{-1} \cdot \left[\sum_{t\in N(WW)} \left[(S_{i,t} - \hat{\Theta}^{WW} S_{i,t-1}) - \right.\right.$$

$$\left.\left.(\sum_{\substack{k=1\\k\neq j}}^{L} \hat{\alpha}_k^W \varphi_k(t)) + \hat{\Theta}^{WW} \sum_{\substack{k=1\\k\neq j}}^{L} \hat{\alpha}_k^W \varphi_k(t-1)\right]\right[\varphi_j(t) -$$

$$\left.\hat{\Theta}^{WW} \varphi_j(t-1)\right] + \sum_{t\in N(DW)} \left[(S_{i,t} - \hat{\Theta}^{DW} S_{i,t-1} + \hat{\Theta}^{DW} \hat{\mu}_{t-1}^D) - \right.$$

$$\left.\sum_{\substack{k=1\\k\neq j}}^{L} \hat{\alpha}_k^W \varphi_k(t)\right] \varphi_j(t) - \hat{\Theta} \sum_{t\in N(WD)} \left[S_{i,t} - \hat{\Theta}^{WD} \right. .$$

$$\left. S_{i,t-1} - \hat{\mu}_t^D + \hat{\Theta}^{WD} \sum_{\substack{k=1\\k\neq j}}^{L} \hat{\alpha}_k^W \varphi_k(t-1) \right] \varphi_j(t-1) \quad .$$

The estimate for $\sigma_e^2(DD)$ is given by:

$$\hat{\sigma}_e^2(DD) = 1/C(DD) \sum_{t\in N(DD)} (\beta(DD))^2 \; .$$

The estimates for $\sigma_e^2(WW)$, $\sigma_e^2(DW)$ and $\sigma_e^2(WD)$ are given similarly.

The correlation is estimated as in the previous section.

Estimates for the parameters Θ and μ_t are obtained iteratively. The following algorithm gives the procedure for parameter estimation.

Algorithm.

Step 1: Estimate initial $\hat{\mu}_t$ by approximating by its Fourier series transformation and estimating parameters α_i as before.

Step 2: Estimate Θ .

Step 3: Estimate μ_t .

Step 4: Check for convergence of all parameters, i.e. when the percentage change in parameter estimates is sufficiently small. If convergence is not met, go back to Step 2.

The implementation of the models described above is discussed in Chapter 4, where parameter estimates and model selection for a particular data set are given. The procedures to follow in implementing the models as well as those to simulate climate sequences once the parameters have been estimated are described in Chapter 5. A discussion on the validation of the above models is contained in Chapter 6.

CHAPTER 4

4. MODEL IMPLEMENTATION

This chapter gives details of the implementation of the proposed time series models to describe the climate series for the historical record at Elsenburg. In particular, the model selection process is described step by step, the parameter estimates are given and the results of tests to check the model assumptions are discussed.

As already mentioned, the historical record used for model implementation was daily observations for the years 1979 to 1984 and in the case of rainfall, the years 1978 to 1984.

Since rainfall was considered to be the primary variable and all other variables are conditioned on whether a given day was wet or dry, it was modelled independently of all other variables.

4.1 Simple Markov chain to describe the occurrence of wet and dry sequences of days.

The logit transformation of the probabilities $\pi(t)$, $t=1,2,\ldots,NT$ is given by

$$\lambda(t) = \log(\pi(t)/(1 - \pi(t)))$$

where $\lambda(t)$ is represented by a Fourier series approximation, i.e.

$$\lambda(t) = \sum_{i=1}^{L} \gamma_i \, \varphi_i(t) , \qquad t=1,2,\ldots,NT$$

and $\varphi_i(t)$ is defined as in Chapter 3.

The parameters γ_i have to be estimated for the probability that a wet day follows a wet day and for the probability that a wet day follows a dry day. For each of the probabilities the order of the Fourier series approximation, L, has to be selected.

The parameter estimates for approximating models for the probability that a wet day follows a wet day are given in Table 4.1 for $L=1,3,\ldots,13$. The selection of the appropriate approximating model, i.e. the selection of L, was based on the deviance statistic. In particular, the deviances of a number of competing nested models were compared, and a test of hypothesis that the simpler model holds was performed at the 5% level of significance. This procedure led to the selection of $L=3$ (Table 4.2).

Table 4.3 shows the parameter estimates for the probability that a wet day follows a dry day. Again, different numbers of parameters were fitted for the purpose of model selection. The same selection procedure was applied and here too, this led to the selection of $L=3$ (Table 4.4).

TABLE 4.1: PARAMETER ESTIMATES FOR THE FIT OF THE TRUNCATED FOURIER SERIES FOR THE PROBABILITY OF HAVING A WET DAY GIVEN A PRECEDING WET DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\gamma}_0$	-0.06	-0.15	-0.17	-0.18	-0.18	-0.18	-0.18
$\hat{\gamma}_1$		-0.39	-0.42	-0.43	-0.43	-0.43	-0.43
$\hat{\gamma}_2$		-0.08	-0.13	-0.14	-0.14	-0.14	-0.14
$\hat{\gamma}_3$			-0.04	-0.03	-0.04	-0.03	-0.04
$\hat{\gamma}_4$			-0.18	-0.21	-0.22	-0.22	-0.22
$\hat{\gamma}_5$				0.11	0.09	0.09	0.08
$\hat{\gamma}_6$				-0.08	-0.10	-0.10	-0.10
$\hat{\gamma}_7$					0.01	-0.02	-0.02
$\hat{\gamma}_8$					-0.10	-0.11	-0.11
$\hat{\gamma}_9$						-0.08	-0.06
$\hat{\gamma}_{10}$						-0.08	-0.09
$\hat{\gamma}_{11}$							0.09
$\hat{\gamma}_{12}$							-0.01
Devi-ance	428.7	417.6	415.1	413.9	413.2	412.4	411.8

TABLE 4.2: MODEL SELECTION FOR THE PROBABILITY THAT A WET DAY FOLLOWS A WET DAY

$H_0: \gamma = \gamma_L$ vs $H_1: \gamma = \gamma_P$	$D = D_L - D_P$	P-value % based on χ^2_{P-L} distribution	Decision
L = 1, P = 3	11.1	5.991	L = 3
L = 3, P = 5	2.5	5.991	L = 3
L = 3, P = 7	3.7	9.488	L = 3
L = 3, P = 9	4.4	12.592	L = 3
L = 3, P = 11	5.2	15.507	L = 3
L = 3, P = 13	5.8	18.307	L = 3

TABLE 4.3: PARAMETER ESTIMATES FOR THE FIT OF THE TRUNCATED FOURIER SERIES FOR THE PROBABILITY OF HAVING A WET DAY GIVEN A PRECEDING DRY DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\gamma}_0$	-1.57	-1.64	-1.63	-1.65	-1.66	-1.65	-1.65
$\hat{\gamma}_1$		-0.50	-0.49	-0.49	-0.50	-0.49	-0.49
$\hat{\gamma}_2$		-0.29	-0.28	-0.31	-0.32	-0.30	-0.30
$\hat{\gamma}_3$			0.04	0.10	0.10	0.11	0.10
$\hat{\gamma}_4$			0.07	0.07	0.06	0.08	0.08
$\hat{\gamma}_5$				0.15	0.18	0.17	0.15
$\hat{\gamma}_6$				0.16	0.16	0.17	0.17
$\hat{\gamma}_7$					0.07	0.05	0.05
$\hat{\gamma}_8$					0.08	0.04	0.04
$\hat{\gamma}_9$						0.05	0.07
$\hat{\gamma}_{10}$						-0.18	-0.17
$\hat{\gamma}_{11}$							0.06
$\hat{\gamma}_{12}$							0.07
Deviance	373.5	340.8	340.1	335.4	334.3	331.0	330.3

TABLE 4.4: MODEL SELECTION FOR THE PROBABILITY THAT A WET DAY FOLLOWS A DRY DAY

$H_0: \gamma = \gamma_L$ vs $H_1: \gamma = \gamma_P$	$D = D_L - D_P$	P-value % based on χ^2_{P-L} distribution	Decision
L = 1, P = 3	32.7	5.991	L = 3
L = 3, P = 5	0.7	5.991	L = 3
L = 3, P = 7	5.4	9.488	L = 3
L = 3, P = 9	6.5	12.592	L = 3
L = 3, P = 11	9.8	15.507	L = 3
L = 3, P = 13	10.5	18.307	L = 3

4.2 The distribution for rainfall on days when rain occurs.

The mean rainfall per rainy day in period t , $\mu(t)$, can be approximated by its truncated Fourier series representation

$$\mu(t) = \sum_{i=1}^{L} \mu_i \, \varphi_i(t) \quad , \qquad t=1,2,\ldots,NT$$

where $\varphi_i(t)$ is defined as in Chapter 3.

The parameters μ_i need to be estimated and the order of the Fourier series approximation selected.

For this a 3-term Fourier series approximation was chosen (Zucchini and Adamson (1984)). It is sometimes easier to work with the Fourier series coefficients in their polar form. Table 4.5 shows the parameter estimates for mean rainfall, the amplitude and phase representation of the mean rainfall and the estimate for the coefficient of variation. From these, parameters of the corresponding Weibull distribution can be estimated then by the method of moments (see Zucchini and Adamson (1984)).

TABLE 4.5: PARAMETER ESTIMATES FOR THE DISTRIBUTION FOR RAINFALL ON DAYS WHEN RAIN OCCURS

Mean parameters	Polar form	Coeff. of variation
$\hat{\mu}_1$ = 62.13	amplitude (0) = 62.13	1.224
$\hat{\mu}_2$ = -13.03	amplitude (1) = 16.53	
$\hat{\mu}_3$ = 10.16	phase (1) = 144.0	

4.3 Model 1: Multivariate model for climate data proposed by Richardson (1981).

The historical data for each of the climate variables was conditioned on the wet or dry status of the day, thus obtaining a mean function and a standard deviation function for each of the conditioned data sets. The mean and the standard deviation were both approximated by a truncated Fourier series representation. That is

$$\mu_t = \sum_{i=1}^{L} \alpha_i \varphi_i(t) \qquad \text{and}$$

$$\sigma_t = \sum_{i=1}^{L} \xi_i \varphi_i(t) \quad , \qquad t=1,2,\ldots,NT$$

where $\varphi_i(t)$ is defined as in Chapter 3 and where L does not have to be of the same order for both of the mean and the standard deviation function.

Tables 4.6 - 4.12 show the parameter estimates for the mean function given that it rains in period t. For the purposes of model selection the truncation level L, which determines the family of approximating models being fitted, was varied and the fit in each case was examined. To select the appropriate number of parameters the percentage of variance accounted for by each model was examined. The decision on which order of approximation to use was based on a test of hypothesis at the 5% level of significance. Tables 4.13 - 4.19 show the choice of the order of approximation.

TABLE 4.6: PARAMETER ESTIMATES FOR MEAN FUNCTION FITTED BY A TRUNCATED FOURIER SERIES, FOR MAXIMUM TEMPERATURE GIVEN A WET DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\alpha}_0$	188.9	198.1	194.5	194.6	194.5	194.6	194.6
$\hat{\alpha}_1$		45.24	45.48	45.45	45.47	45.46	45.42
$\hat{\alpha}_2$		20.09	21.14	21.31	21.06	21.09	21.21
$\hat{\alpha}_3$			-2.92	-3.30	-3.06	-3.14	-3.19
$\hat{\alpha}_4$			6.67	6.66	6.76	6.82	6.75
$\hat{\alpha}_5$				-1.79	-1.87	-1.99	-1.94
$\hat{\alpha}_6$				-1.96	-1.49	-1.55	-1.71
$\hat{\alpha}_7$					-3.00	-2.83	-2.51
$\hat{\alpha}_8$					2.36	2.09	2.03
$\hat{\alpha}_9$						2.93	3.01
$\hat{\alpha}_{10}$						-1.28	-0.96
$\hat{\alpha}_{11}$							-2.05
$\hat{\alpha}_{12}$							1.47
% Var acc		58.1	59.1	59.0	59.1	59.1	58.9

TABLE 4.7: PARAMETER ESTIMATES FOR MEAN FUNCTION FITTED BY A TRUNCATED FOURIER SERIES, FOR MINIMUM TEMPERATURE GIVEN A WET DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\alpha}_0$	114.0	117.8	118.0	118.0	118.0	118.0	118.1
$\hat{\alpha}_1$		30.63	30.73	30.60	30.59	30.57	30.53
$\hat{\alpha}_2$		17.69	18.19	18.20	18.25	18.29	18.44
$\hat{\alpha}_3$			-1.45	-1.59	-1.63	-1.72	-1.79
$\hat{\alpha}_4$			3.05	2.80	2.78	2.76	2.67
$\hat{\alpha}_5$				1.23	1.25	1.29	1.35
$\hat{\alpha}_6$				-3.09	-3.18	-3.28	-3.49
$\hat{\alpha}_7$					0.63	0.88	1.31
$\hat{\alpha}_8$					-0.41	-0.43	-0.49
$\hat{\alpha}_9$						1.59	1.71
$\hat{\alpha}_{10}$						1.08	1.51
$\hat{\alpha}_{11}$							-2.42
$\hat{\alpha}_{12}$							2.12
% Var acc		58.1	58.3	58.6	58.3	58.2	58.4

TABLE 4.8: PARAMETER ESTIMATES FOR MEAN FUNCTION FITTED BY A TRUNCATED FOURIER SERIES, FOR EVAPORATION GIVEN A WET DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\alpha}_0$	33.14	34.97	34.99	34.93	34.91	34.95	34.93
$\hat{\alpha}_1$		25.07	25.43	25.45	25.46	25.41	25.40
$\hat{\alpha}_2$		0.89	0.87	0.74	0.70	0.71	0.61
$\hat{\alpha}_3$			4.47	4.84	4.79	4.66	4.71
$\hat{\alpha}_4$			2.87	2.88	2.81	2.87	2.91
$\hat{\alpha}_5$				1.18	1.39	1.28	1.26
$\hat{\alpha}_6$				2.32	2.24	2.21	2.34
$\hat{\alpha}_7$					2.10	2.19	1.97
$\hat{\alpha}_8$					1.21	0.89	0.83
$\hat{\alpha}_9$						2.80	2.63
$\hat{\alpha}_{10}$						-2.18	-2.43
$\hat{\alpha}_{11}$							0.45
$\hat{\alpha}_{12}$							-1.59
% Var acc		47.9	49.7	49.8	49.9	50.5	50.3

TABLE 4.9: PARAMETER ESTIMATES FOR MEAN FUNCTION FITTED BY A TRUNCATED FOURIER SERIES, FOR SUNSHINE DURATION GIVEN A WET DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\alpha}_0$	45.52	46.62	46.66	46.61	46.55	46.62	46.58
$\hat{\alpha}_1$		16.66	16.93	16.92	16.97	16.95	16.95
$\hat{\alpha}_2$		0.31	0.30	0.21	0.00	0.04	-0.24
$\hat{\alpha}_3$			2.81	3.01	3.14	3.04	3.15
$\hat{\alpha}_4$			2.43	2.38	2.44	2.49	2.69
$\hat{\alpha}_5$				1.32	1.40	1.29	1.19
$\hat{\alpha}_6$				0.68	0.97	0.95	1.27
$\hat{\alpha}_7$					-1.50	-1.34	-2.11
$\hat{\alpha}_8$					2.30	1.90	1.94
$\hat{\alpha}_9$						4.13	3.65
$\hat{\alpha}_{10}$						-1.49	-2.22
$\hat{\alpha}_{11}$							2.87
$\hat{\alpha}_{12}$							-4.99
% Var acc		14.9	15.1	14.6	14.4	14.9	16.1

TABLE 4.10: PARAMETER ESTIMATES FOR MEAN FUNCTION FITTED BY A TRUNCATED FOURIER SERIES, FOR WINDRUN GIVEN A WET DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\alpha}_0$	2446	2423	2422	2420	2421	2423	2423
$\hat{\alpha}_1$		-158.2	-148.0	-145.8	-146.2	-146.7	-145.7
$\hat{\alpha}_2$		-140.0	-146.7	-151.3	-144.7	-143.6	-142.1
$\hat{\alpha}_3$			138.9	150.8	144.6	142.0	142.1
$\hat{\alpha}_4$			47.8	50.8	49.3	50.6	50.2
$\hat{\alpha}_5$				34.6	34.7	32.4	32.4
$\hat{\alpha}_6$				88.2	76.4	74.2	71.7
$\hat{\alpha}_7$					61.9	68.3	75.8
$\hat{\alpha}_8$					-72.1	-78.8	-77.6
$\hat{\alpha}_9$						83.7	91.5
$\hat{\alpha}_{10}$						-19.3	-11.2
$\hat{\alpha}_{11}$							20.6
$\hat{\alpha}_{12}$							65.6
% Var acc		1.7	2.3	2.1	1.9	1.7	1.2

TABLE 4.11: PARAMETER ESTIMATES FOR MEAN FUNCTION FITTED BY A TRUNCATED FOURIER SERIES, FOR MAXIMUM HUMIDITY GIVEN A WET DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\alpha}_0$	931.1	929.2	928.8	928.6	928.6	928.8	928.7
$\hat{\alpha}_1$		-14.77	-15.39	-15.44	-15.46	-15.52	-15.51
$\hat{\alpha}_2$		-9.71	-10.67	-11.07	-10.63	-10.51	-10.85
$\hat{\alpha}_3$			-1.48	-0.67	-1.08	-1.33	-1.23
$\hat{\alpha}_4$			-9.12	-9.31	-9.44	-9.34	-9.17
$\hat{\alpha}_5$				5.33	5.40	5.25	5.15
$\hat{\alpha}_6$				2.26	1.46	1.23	1.74
$\hat{\alpha}_7$					4.71	5.34	4.15
$\hat{\alpha}_8$					-4.47	-4.99	-4.97
$\hat{\alpha}_9$						7.32	6.61
$\hat{\alpha}_{10}$						-0.93	-2.15
$\hat{\alpha}_{11}$							2.62
$\hat{\alpha}_{12}$							-7.76
% Var acc		3.5	4.1	3.8	3.8	3.9	4.1

TABLE 4.12: PARAMETER ESTIMATED FOR MEAN FUNCTION FITTED BY A TRUNCATED FOURIER SERIES, FOR MINIMUM HUMIDITY GIVEN A WET DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\alpha}_0$	531.4	522.1	521.2	521.0	521.3	521.4	521.4
$\hat{\alpha}_1$		-102.8	-103.2	-103.3	-103.4	-103.4	-103.6
$\hat{\alpha}_2$		-9.25	-11.78	-12.14	-10.45	-10.42	-10.53
$\hat{\alpha}_3$			8.83	9.50	7.96	7.87	7.75
$\hat{\alpha}_4$			-14.62	-14.92	-15.04	-14.79	-14.83
$\hat{\alpha}_5$				5.57	4.93	4.45	4.52
$\hat{\alpha}_6$				0.36	-2.48	-2.48	-2.25
$\hat{\alpha}_7$					10.27	10.39	9.41
$\hat{\alpha}_8$					-21.63	-22.56	-22.90
$\hat{\alpha}_9$						7.14	5.67
$\hat{\alpha}_{10}$						-6.45	-7.55
$\hat{\alpha}_{11}$							-7.79
$\hat{\alpha}_{12}$							-10.89
% Var acc		29.0	29.3	28.9	30.0	29.8	29.8

TABLE 4.13: MODEL SELECTION FOR THE MEAN FUNCTION OF MAXIMUM TEMPERATURE GIVEN THAT A WET DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	58.1	L = 3
L = 3, P = 5	1.0	L = 3
L = 3, P = 7	0.9	L = 3
L = 3, P = 9	1.0	L = 3
L = 3, P = 11	1.0	L = 3
L = 3, P = 13	0.8	L = 3

TABLE 4.14: MODEL SELECTION FOR THE MEAN FUNCTION OF MINIMUM TEMPERATURE GIVEN THAT A WET DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	58.1	L = 3
L = 3, P = 5	0.2	L = 3
L = 3, P = 7	0.5	L = 3
L = 3, P = 9	0.2	L = 3
L = 3, P = 11	0.1	L = 3
L = 3, P = 13	0.3	L = 3

TABLE 4.15: MODEL SELECTION FOR THE MEAN FUNCTION OF EVAPORATION GIVEN THAT A WET DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	47.9	L = 3
L = 3, P = 5	1.8	L = 3
L = 3, P = 7	1.9	L = 3
L = 3, P = 9	2.0	L = 3
L = 3, P = 11	2.6	L = 3
L = 3, P = 13	2.4	L = 3

TABLE 4.16: MODEL SELECTION FOR THE MEAN FUNCTION OF SUNSHINE DURATION GIVEN THAT A WET DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	14.9	L = 3
L = 3, P = 5	0.2	L = 3
L = 3, P = 7	less than 0.1	L = 3
L = 3, P = 9	less than 0.1	L = 3
L = 3, P = 11	less than 0.1	L = 3
L = 3, P = 13	1.2	L = 3

TABLE 4.17: MODEL SELECTION FOR THE MEAN FUNCTION OF WINDRUN GIVEN THAT A WET DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	1.7	L = 1
L = 1, P = 5	2.3	L = 1
L = 1, P = 7	2.1	L = 1
L = 1, P = 9	1.9	L = 1
L = 1, P = 11	1.7	L = 1
L = 1, P = 13	1.2	L = 1

TABLE 4.18: MODEL SELECTION FOR THE MEAN FUNCTION OF MAXIMUM HUMIDITY GIVEN THAT A WET DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	3.5	L = 1
L = 1, P = 5	4.1	L = 1
L = 1, P = 7	3.8	L = 1
L = 1, P = 9	3.8	L = 1
L = 1, P = 11	3.9	L = 1
L = 1, P = 13	4.1	L = 1

TABLE 4.19: MODEL SELECTION FOR THE MEAN FUNCTION OF MINIMUM HUMIDITY GIVEN THAT A WET DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	29.0	L = 3
L = 3, P = 5	0.3	L = 3
L = 3, P = 7	less than 0.1	L = 3
L = 3, P = 9	1.0	L = 3
L = 3, P = 11	0.8	L = 3
L = 3, P = 13	0.8	L = 3

Except for the variables windrun and maximum humidity, a 3-term Fourier series was chosen to approximate the mean function given that rain occurs. For the other two variables a 1-term Fourier series was indicated.

Tables 4.20 - 4.26 show the parameter estimates for the mean function given that it does not rain in period t . For the purpose of model selection, the truncation level was varied and the fit in each case examined.

Tables 4.27 - 4.33 show the choice of order of approximation for the mean function given that a dry day occurs. A 5% level of significance was again used for making the final selection.

TABLE 4.20: PARAMETER ESTIMATES FOR MEAN FUNCTION FITTED BY A TRUNCATED FOURIER SERIES, FOR MAXIMUM TEMPERATURES GIVEN A DRY DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\alpha}_0$	238.2	238.2	238.2	238.2	238.2	238.2	238.2
$\hat{\alpha}_1$		48.41	48.41	48.41	48.41	48.41	48.41
$\hat{\alpha}_2$		18.81	18.81	18.81	18.81	18.81	18.81
$\hat{\alpha}_3$			-7.95	-7.95	-7.95	-7.95	-7.95
$\hat{\alpha}_4$			2.19	2.19	2.19	2.19	2.19
$\hat{\alpha}_5$				-2.20	-2.20	-2.20	-2.20
$\hat{\alpha}_6$				-0.02	-0.02	-0.02	-0.02
$\hat{\alpha}_7$					1.19	1.19	1.19
$\hat{\alpha}_8$					1.16	1.16	1.16
$\hat{\alpha}_9$						2.39	2.39
$\hat{\alpha}_{10}$						-0.04	-0.04
$\hat{\alpha}_{11}$							-0.32
$\hat{\alpha}_{12}$							-0.05
% Var acc		75.6	77.4	77.4	77.3	77.4	77.2

TABLE 4.21: PARAMETER ESTIMATES FOR MEAN FUNCTION FITTED BY A TRUNCATED FOURIER SERIES, FOR MINIMUM TEMPERATURES GIVEN A DRY DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\alpha}_0$	102.5	102.5	102.5	102.5	102.5	102.5	102.5
$\hat{\alpha}_1$		35.00	35.00	35.00	35.00	35.00	35.00
$\hat{\alpha}_2$		16.90	16.90	16.90	16.90	16.90	16.90
$\hat{\alpha}_3$			-1.15	-1.15	-1.15	-1.15	-1.15
$\hat{\alpha}_4$			-0.15	-0.15	-0.15	-0.15	-0.15
$\hat{\alpha}_5$				0.04	0.04	0.04	0.04
$\hat{\alpha}_6$				-1.28	-1.28	-1.28	-1.28
$\hat{\alpha}_7$					-0.47	-0.47	-0.47
$\hat{\alpha}_8$					1.44	1.44	1.44
$\hat{\alpha}_9$						1.06	1.06
$\hat{\alpha}_{10}$						-1.24	-1.24
$\hat{\alpha}_{11}$							-0.92
$\hat{\alpha}_{12}$							0.04
% Var acc		78.5	78.5	78.4	78.4	78.4	78.4

TABLE 4.22: PARAMETER ESTIMATES FOR MEAN FUNCTION FITTED BY A TRUNCATED FOURIER SERIES, FOR EVAPORATION GIVEN A DRY DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\alpha}_0$	63.65	63.65	63.65	63.65	63.65	63.65	63.65
$\hat{\alpha}_1$		41.85	41.85	41.85	41.85	41.85	41.85
$\hat{\alpha}_2$		1.31	1.31	1.31	1.31	1.31	1.31
$\hat{\alpha}_3$			2.49	2.49	2.49	2.49	2.49
$\hat{\alpha}_4$			2.25	2.25	2.25	2.25	2.25
$\hat{\alpha}_5$				-0.87	-0.87	-0.87	-0.87
$\hat{\alpha}_6$				-0.22	-0.22	-0.22	-0.22
$\hat{\alpha}_7$					-1.30	-1.30	-1.30
$\hat{\alpha}_8$					-0.25	-0.25	-0.25
$\hat{\alpha}_9$						1.23	1.23
$\hat{\alpha}_{10}$						1.19	1.19
$\hat{\alpha}_{11}$							1.49
$\hat{\alpha}_{12}$							0.35
% Var acc		90.9	91.4	91.4	91.4	91.5	91.6

TABLE 4.23: PARAMETER ESTIMATES FOR MEAN FUNCTION FITTED BY A TRUNCATED FOURIER SERIES, FOR SUNSHINE DURATION GIVEN A DRY DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\alpha}_0$	94.54	94.54	94.54	94.54	94.54	94.54	94.54
$\hat{\alpha}_1$		20.71	20.71	20.71	20.71	20.71	20.71
$\hat{\alpha}_2$		-2.49	-2.49	-2.49	-2.49	-2.49	-2.49
$\hat{\alpha}_3$			1.24	1.24	1.24	1.24	1.24
$\hat{\alpha}_4$			1.74	1.74	1.74	1.74	1.74
$\hat{\alpha}_5$				-0.91	-0.91	-0.91	-0.91
$\hat{\alpha}_6$				-0.04	-0.04	-0.04	-0.04
$\hat{\alpha}_7$					-0.02	-0.02	-0.02
$\hat{\alpha}_8$					-0.25	-0.25	-0.25
$\hat{\alpha}_9$						1.07	1.07
$\hat{\alpha}_{10}$						0.88	0.88
$\hat{\alpha}_{11}$							1.25
$\hat{\alpha}_{12}$							-1.18
% Var acc		61.2	61.6	61.5	61.3	61.3	61.6

TABLE 4.24: PARAMETER ESTIMATES FOR MEAN FUNCTION FITTED BY A TRUNCATED FOURIER SERIES, FOR WINDRUN GIVEN A DRY DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\alpha}_0$	1764	1764	1764	1764	1764	1764	1764
$\hat{\alpha}_1$		322.5	322.5	322.5	322.5	322.5	322.5
$\hat{\alpha}_2$		-23.50	-23.50	-23.50	-23.50	-23.50	-23.50
$\hat{\alpha}_3$			98.40	98.40	98.40	98.40	98.40
$\hat{\alpha}_4$			44.10	44.10	44.10	44.10	44.10
$\hat{\alpha}_5$				23.30	23.30	23.30	23.30
$\hat{\alpha}_6$				-5.20	-5.20	-5.20	-5.20
$\hat{\alpha}_7$					-70.30	-70.30	-70.30
$\hat{\alpha}_8$					-28.60	-28.60	-28.60
$\hat{\alpha}_9$						7.10	7.10
$\hat{\alpha}_{10}$						39.10	39.10
$\hat{\alpha}_{11}$							28.80
$\hat{\alpha}_{12}$							-4.90
% Var acc		33.1	36.5	36.3	37.8	38.0	37.9

TABLE 4.25: PARAMETER ESTIMATES FOR MEAN FUNCTION FITTED BY A TRUNCATED FOURIER SERIES, FOR MAXIMUM HUMIDITY GIVEN A DRY DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\alpha}_0$	909.6	909.6	909.6	909.6	909.6	909.6	909.6
$\hat{\alpha}_1$		-7.62	-7.62	-7.62	-7.62	-7.62	-7.62
$\hat{\alpha}_2$		1.26	1.26	1.26	1.26	1.26	1.26
$\hat{\alpha}_3$			-6.30	-6.30	-6.30	-6.30	-6.30
$\hat{\alpha}_4$			-3.58	-3.58	-3.58	-3.58	-3.58
$\hat{\alpha}_5$				5.11	5.11	5.11	5.11
$\hat{\alpha}_6$				-1.50	-1.50	-1.50	-1.50
$\hat{\alpha}_7$					9.40	9.40	9.40
$\hat{\alpha}_8$					4.15	4.15	4.15
$\hat{\alpha}_9$						-7.74	-7.74
$\hat{\alpha}_{10}$						-3.47	-3.47
$\hat{\alpha}_{11}$							-6.31
$\hat{\alpha}_{12}$							2.19
% Var acc		0.6	1.1	1.1	2.6	3.5	3.9

TABLE 4.26: PARAMETER ESTIMATES FOR MEAN FUNCTION FITTED BY A TRUNCATED FOURIER SERIES, FOR MINIMUM HUMIDITY GIVEN A DRY DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\alpha}_0$	370.9	370.9	370.9	370.9	370.9	370.9	370.9
$\hat{\alpha}_1$		-56.58	-56.58	-56.58	-56.58	-56.58	-56.58
$\hat{\alpha}_2$		-20.65	-20.65	-20.65	-20.65	-20.65	-20.65
$\hat{\alpha}_3$			19.19	19.19	19.19	19.19	19.19
$\hat{\alpha}_4$			-2.90	-2.90	-2.90	-2.90	-2.90
$\hat{\alpha}_5$				-0.67	-0.67	-0.67	-0.67
$\hat{\alpha}_6$				2.75	2.75	2.75	2.75
$\hat{\alpha}_7$					2.87	2.87	2.87
$\hat{\alpha}_8$					0.79	0.79	0.79
$\hat{\alpha}_9$						-1.47	-1.47
$\hat{\alpha}_{10}$						-1.71	-1.71
$\hat{\alpha}_{11}$							-6.10
$\hat{\alpha}_{12}$							-0.22
% Var acc		34.8	38.1	37.9	37.6	37.3	37.3

TABLE 4.27: MODEL SELECTION FOR THE MEAN FUNCTION OF MAXIMUM TEMPERATURE GIVEN THAT A DRY DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	75.6	L = 3
L = 3, P = 5	1.8	L = 3
L = 3, P = 7	1.8	L = 3
L = 3, P = 9	1.7	L = 3
L = 3, P = 11	1.8	L = 3
L = 3, P = 13	1.6	L = 3

TABLE 4.28: MODEL SELECTION FOR THE MEAN FUNCTION OF MINIMUM TEMPERATURE GIVEN THAT A DRY DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	78.5	L = 3
L = 3, P = 5	less than 0.1	L = 3
L = 3, P = 7	less than 0.1	L = 3
L = 3, P = 9	less than 0.1	L = 3
L = 3, P = 11	less than 0.1	L = 3
L = 3, P = 13	less than 0.1	L = 3

TABLE 4.29 MODEL SELECTION FOR THE MEAN FUNCTION OF EVAPORATION GIVEN THAT A DRY DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	90.9	L = 3
L = 3, P = 5	0.5	L = 3
L = 3, P = 7	0.5	L = 3
L = 3, P = 9	0.5	L = 3
L = 3, P = 11	0.6	L = 3
L = 3, P = 13	0.7	L = 3

TABLE 4.30: MODEL SELECTION FOR THE MEAN FUNCTION OF SUNSHINE DURATION GIVEN THAT A DRY DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	61.2	L = 3
L = 3, P = 5	0.4	L = 3
L = 3, P = 7	0.3	L = 3
L = 3, P = 9	0.1	L = 3
L = 3, P = 11	0.1	L = 3
L = 3, P = 13	0.4	L = 3

TABLE 4.31: MODEL SELECTION FOR THE MEAN FUNCTION OF WINDRUN GIVEN THAT A DRY DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	33.1	L = 3
L = 3, P = 5	3.4	L = 3
L = 3, P = 7	3.2	L = 3
L = 3, P = 9	4.7	L = 3
L = 3, P = 11	4.9	L = 3
L = 3, P = 13	4.8	L = 3

TABLE 4.32: MODEL SELECTION FOR THE MEAN FUNCTION OF MAXIMUM HUMIDITY GIVEN THAT A DRY DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	0.6	L = 1
L = 1, P = 5	1.1	L = 1
L = 1, P = 7	1.1	L = 1
L = 1, P = 9	2.6	L = 1
L = 1, P = 11	3.5	L = 1
L = 1, P = 13	3.9	L = 1

TABLE 4.33: MODEL SELECTION FOR THE MEAN FUNCTION OF MINIMUM HUMIDITY GIVEN THAT A DRY DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	34.8	L = 3
L = 3, P = 5	3.3	L = 3
L = 3, P = 7	3.1	L = 3
L = 3, P = 9	2.8	L = 3
L = 3, P = 11	2.5	L = 3
L = 3, P = 13	2.5	L = 3

For all variables, except maximum humidity, a 3-term Fourier series approximation was chosen to represent the mean function, given a dry day occurs. For maximum humidity, a 1-term Fourier series approximation was indicated.

Thus except perhaps for the case of maximum humidity and also windrun when a wet day occurs, a 3-term Fourier series approximation is estimated to be appropriate. The cases of maximum humidity and windrun (given a wet day has occurred), for which L=1 was indicated as best, were also fitted using L=3 . This was done in order to simplify the implementation and interpretation of the complete (multivariate time series) model.

Tables 4.34 - 4.40 show the parameter estimates for the standard deviation function given that a wet day occurs in period t . For the purpose of model selection the truncation level L was varied and the fit in each case examined.

Tables 4.41 - 4.47 show the choice of the order of the Fourier series approximation. A 5% level of significance was used for making a selection.

TABLE 4.34: PARAMETER ESTIMATES FOR STANDARD DEVIATION FUNCTION FITTED BY A TRUNCATED FOURIER SERIES FOR MAXIMUM TEMPERATURE, GIVEN A WET DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\xi}_0$	1174	1216	1225	1237	1237	1238	1240
$\hat{\xi}_1$		367	353	361	363	363	362
$\hat{\xi}_2$		154	187	208	203	203	211
$\hat{\xi}_3$			-312	-352	-346	-347	-349
$\hat{\xi}_4$			41	63	72	77	73
$\hat{\xi}_5$				-354	-372	-381	-379
$\hat{\xi}_6$				14	28	29	17
$\hat{\xi}_7$					-202	-201	-175
$\hat{\xi}_8$					-36	-53	-54
$\hat{\xi}_9$						116	131
$\hat{\xi}_{10}$						-124	-97
$\hat{\xi}_{11}$							-76
$\hat{\xi}_{12}$							163
% Var acc		1.6	2.3	3.5	3.4	3.2	2.9

TABLE 4.35: PARAMETER ESTIMATES FOR STANDARD DEVIATION FUNCTION FITTED BY A TRUNCATED FOURIER SERIES FOR MINIMUM TEMPERATURE, GIVEN A WET DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\xi}_0$	604.1	608.1	613.0	616.9	617.6	614.8	616.0
$\hat{\xi}_1$		-25.3	-18.5	-16.4	-16.1	-15.2	-16.6
$\hat{\xi}_2$		91.1	103.1	110.1	112.3	110.4	113.7
$\hat{\xi}_3$			8.0	-5.5	-7.1	-3.2	-5.0
$\hat{\xi}_4$			106.4	112.2	114.7	114.4	112.2
$\hat{\xi}_5$				-108.3	-115.3	-115.2	-113.8
$\hat{\xi}_6$				-11.6	-13.7	-9.7	-14.2
$\hat{\xi}_7$					-39.2	-49.8	-41.4
$\hat{\xi}_8$					-58.1	-53.9	-56.0
$\hat{\xi}_9$						-91.2	-90.9
$\hat{\xi}_{10}$						-19.8	-11.7
$\hat{\xi}_{11}$							-72.7
$\hat{\xi}_{12}$							30.2
% Var acc		0.2	0.6	1.0	0.8	0.9	0.8

TABLE 4.36: PARAMETER ESTIMATES FOR STANDARD DEVIATION FUNCTION FITTED BY A TRUNCATED FOURIER SERIES FOR EVAPORATION, GIVEN A WET DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\xi}_0$	481.9	518.2	516.3	516.7	517.2	517.3	515.4
$\hat{\xi}_1$		377.7	383.4	380.7	379.4	378.3	379.6
$\hat{\xi}_2$		18.2	10.5	11.6	15.5	15.4	12.1
$\hat{\xi}_3$			90.7	85.3	80.3	77.9	79.5
$\hat{\xi}_4$			11.5	7.1	3.0	5.4	6.9
$\hat{\xi}_5$				9.7	15.8	9.0	7.4
$\hat{\xi}_6$				-56.1	-65.4	-67.0	-64.5
$\hat{\xi}_7$					80.9	81.7	77.0
$\hat{\xi}_8$					-5.2	-18.0	-16.6
$\hat{\xi}_9$						79.2	82.3
$\hat{\xi}_{10}$						-73.0	-78.6
$\hat{\xi}_{11}$							66.2
$\hat{\xi}_{12}$							2.7
% Var acc		15.9	16.3	16.1	16.2	17.0	16.9

TABLE 4.37: PARAMETER ESTIMATES FOR STANDARD DEVIATION FUNCTION FITTED BY A TRUNCATED FOURIER SERIES FOR SUNSHINE DURATION, GIVEN A WET DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\xi}_0$	1012	1041	1040	1039	1039	1043	1042
$\hat{\xi}_1$		295.1	300.7	299.8	299.6	298.2	298.7
$\hat{\xi}_2$		47.2	43.4	41.7	43.2	46.0	40.0
$\hat{\xi}_3$			76.6	79.6	78.0	72.4	74.6
$\hat{\xi}_4$			25.5	23.3	21.7	21.5	24.7
$\hat{\xi}_5$				30.4	33.4	34.4	32.6
$\hat{\xi}_6$				-5.5	-9.0	-14.9	-6.1
$\hat{\xi}_7$					39.2	54.5	34.6
$\hat{\xi}_8$					-2.1	-6.0	-4.9
$\hat{\xi}_9$						116.6	106.4
$\hat{\xi}_{10}$						44.9	24.6
$\hat{\xi}_{11}$							63.0
$\hat{\xi}_{12}$							-121.2
% Var acc		5.7	5.5	4.9	4.3	4.7	5.3

TABLE 4.38: PARAMETER ESTIMATES FOR STANDARD DEVIATION FUNCTION FITTED BY A TRUNCATED FOURIER SERIES FOR WINDRUN, GIVEN A WET DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\xi}_0$	1315062	1236951	1221674	1211347	1215802	1216517	1216393
$\hat{\xi}_1$		-924535	-927353	-921856	-922090	-921910	-919184
$\hat{\xi}_2$		3002	-40837	-58674	-37723	-37651	-35137
$\hat{\xi}_3$			195360	238715	220034	219293	220067
$\hat{\xi}_4$			-222199	-215782	-214248	-210418	-210739
$\hat{\xi}_5$				161027	145991	138766	138437
$\hat{\xi}_6$				275261	241691	242292	237756
$\hat{\xi}_7$					68100	68193	83415
$\hat{\xi}_8$					-303476	-316756	-313298
$\hat{\xi}_9$						92552	110860
$\hat{\xi}_{10}$						-100429	-83817
$\hat{\xi}_{11}$							70140
$\hat{\xi}_{12}$							145989
% Var acc		7.6	7.7	8.0	8.3	7.8	7.4

TABLE 4.39: PARAMETER ESTIMATES FOR STANDARD DEVIATION FUNCTION FITTED BY A TRUNCATED FOURIER SERIES FOR MAXIMUM HUMIDITY, GIVEN A WET DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\xi}_0$	5156	5447	5546	5608	5580	5565	5596
$\hat{\xi}_1$		1259	1446	1442	1439	1439	1427
$\hat{\xi}_2$		2749	2972	3081	2967	2962	3080
$\hat{\xi}_3$			779	541	637	655	613
$\hat{\xi}_4$			2460	2477	2434	2385	2321
$\hat{\xi}_5$				-1275	-1112	-1022	-985
$\hat{\xi}_6$				-1020	-859	-857	-1031
$\hat{\xi}_7$					318	291	682
$\hat{\xi}_8$					2039	2214	2192
$\hat{\xi}_9$						-1366	-1169
$\hat{\xi}_{10}$						1198	1596
$\hat{\xi}_{11}$							-1279
$\hat{\xi}_{12}$							2359
% Var acc		0.6	0.9	0.6	0.5	0.4	0.7

TABLE 4.40: PARAMETER ESTIMATES FOR STANDARD DEVIATION FUNCTION FITTED BY A TRUNCATED FOURIER SERIES FOR MINIMUM HUMIDITY, GIVEN A WET DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\xi}_0$	18394	18489	18401	18447	18468	18490	18504
$\hat{\xi}_1$		-2546	-2563	-2616	-2613	-2609	-2614
$\hat{\xi}_2$		4634	4383	4460	4546	4550	4602
$\hat{\xi}_3$			1118	905	832	808	789
$\hat{\xi}_4$			-1277	-1362	-1328	-1221	-1249
$\hat{\xi}_5$				-392	-519	-721	-705
$\hat{\xi}_6$				-1855	-1977	-1964	-2041
$\hat{\xi}_7$					-265	-253	-80
$\hat{\xi}_8$					-1571	-1945	-1954
$\hat{\xi}_9$						2662	2752
$\hat{\xi}_{10}$						-2780	-2603
$\hat{\xi}_{11}$							-542
$\hat{\xi}_{12}$							1058
% Var acc		2.6	2.2	1.9	1.5	2.7	2.1

TABLE 4.41: MODEL SELECTION FOR THE STANDARD DEVIATION FUNCTION OF MAXIMUM TEMPERATURE GIVEN THAT A WET DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	1.6	L = 1
L = 1, P = 5	2.3	L = 1
L = 1, P = 7	3.5	L = 1
L = 1, P = 9	3.4	L = 1
L = 1, P = 11	3.2	L = 1
L = 1, P = 13	2.9	L = 1

TABLE 4.42: MODEL SELECTION FOR THE STANDARD DEVIATION FUNCTION OF MINIMUM TEMPERATURE GIVEN THAT A WET DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	0.2	L = 1
L = 1, P = 5	0.6	L = 1
L = 1, P = 7	1.0	L = 1
L = 1, P = 9	0.8	L = 1
L = 1, P = 11	0.9	L = 1
L = 1, P = 13	0.8	L = 1

TABLE 4.43: MODEL SELECTION FOR THE STANDARD DEVIATION FUNCTION OF EVAPORATION GIVEN THAT A WET DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	15.9	L = 3
L = 3, P = 5	0.4	L = 3
L = 3, P = 7	0.2	L = 3
L = 3, P = 9	0.3	L = 3
L = 3, P = 11	1.1	L = 3
L = 3, P = 13	1.0	L = 3

TABLE 4.44: MODEL SELECTION FOR THE STANDARD DEVIATION FUNCTION OF SUNSHINE DURATION GIVEN THAT A WET DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	5.7	L = 3
L = 3, P = 5	less than 0.1	L = 3
L = 3, P = 7	less than 0.1	L = 3
L = 3, P = 9	less than 0.1	L = 3
L = 3, P = 11	less than 0.1	L = 3
L = 3, P = 13	less than 0.1	L = 3

TABLE 4.45: MODEL SELECTION FOR THE STANDARD DEVIATION FUNCTION OF WINDRUN GIVEN THAT A WET DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	7.6	L = 3
L = 3, P = 5	0.1	L = 3
L = 3, P = 7	0.4	L = 3
L = 3, P = 9	0.7	L = 3
L = 3, P = 11	0.2	L = 3
L = 3, P = 13	less than 0.1	L = 3

TABLE 4.46: MODEL SELECTION FOR THE STANDARD DEVIATION FUNCTION FOR MAXIMUM HUMIDITY GIVEN THAT A WET DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	0.6	L = 1
L = 1, P = 5	0.9	L = 1
L = 1, P = 7	0.6	L = 1
L = 1, P = 9	0.5	L = 1
L = 1, P = 11	0.4	L = 1
L = 1, P = 13	0.7	L = 1

TABLE 4.47: MODEL SELECTION FOR THE STANDARD DEVIATION FUNCTION FOR MINIMUM HUMIDITY GIVEN THAT A WET DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	2.6	L = 1
L = 1, P = 5	2.2	L = 1
L = 1, P = 7	1.9	L = 1
L = 1, P = 9	1.5	L = 1
L = 1, P = 11	2.7	L = 1
L = 1, P = 13	2.1	L = 1

In all cases, either a 1-term or a 3-term Fourier series approximation to the standard deviation function, given a wet day occurs, was chosen. To simplify programming by having a common approximation order, L , a 3-term Fourier series approximation was chosen.

Tables 4.48 - 4.54 show the parameter estimates for the standard deviation function, given a dry day occurs. Again the truncated level L was varied for the purpose of model selection, and in each case the fit was examined.

Tables 4.55 - 4.61 show the model selection for the standard deviation function given a dry day occurs. The decision of the appropriate order of approximation was based on a test of hypothesis at the 5% level of significance.

A 3-term Fourier series approximation to the standard deviation function, given a dry day occurs, was chosen. The reason for not choosing a 1-term Fourier series approximation in some of the variables was the same as above.

TABLE 4.48: PARAMETER ESTIMATES FOR STANDARD DEVIATION FUNCTION FITTED BY A TRUNCATED FOURIER SERIES FOR MAXIMUM TEMPERATURE, GIVEN A DRY DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\xi}_0$	1666	1666	1666	1666	1666	1666	1666
$\hat{\xi}_1$		218.9	218.9	218.9	218.9	218.9	218.9
$\hat{\xi}_2$		-23.8	-23.8	-23.8	-23.8	-23.8	-23.8
$\hat{\xi}_3$			-205.8	-205.8	-205.8	-205.8	-205.8
$\hat{\xi}_4$			-62.6	-62.6	-62.6	-62.6	-62.6
$\hat{\xi}_5$				-13.1	-13.1	-13.1	-13.1
$\hat{\xi}_6$				87.2	87.2	87.2	87.2
$\hat{\xi}_7$					64.7	64.7	64.7
$\hat{\xi}_8$					55.0	55.0	55.0
$\hat{\xi}_9$						45.10	45.10
$\hat{\xi}_{10}$						-128.8	-128.8
$\hat{\xi}_{11}$							-104.9
$\hat{\xi}_{12}$							-1.10
% Var acc		1.8	3.5	3.3	3.1	3.5	3.5

TABLE 4.49: PARAMETER ESTIMATES FOR STANDARD DEVIATION FUNCTION FITTED BY A TRUNCATED FOURIER SERIES FOR MINIMUM TEMPERATURE, GIVEN A DRY DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\xi}_0$	816.6	816.6	816.6	816.6	816.6	816.6	816.6
$\hat{\xi}_1$		-161.6	-161.6	-161.6	-161.6	-161.6	-161.6
$\hat{\xi}_2$		-72.5	-72.5	-72.5	-72.5	-72.5	-72.5
$\hat{\xi}_3$			-35.4	-35.4	-35.4	-35.4	-35.4
$\hat{\xi}_4$			-119.5	-119.5	-119.5	-119.5	-119.5
$\hat{\xi}_5$				-36.0	-36.0	-36.0	-36.0
$\hat{\xi}_6$				60.6	60.6	60.6	60.6
$\hat{\xi}_7$					-100.5	-100.5	-100.5
$\hat{\xi}_8$					49.2	49.2	49.2
$\hat{\xi}_9$						59.8	59.8
$\hat{\xi}_{10}$						-20.6	-20.6
$\hat{\xi}_{11}$							-30.5
$\hat{\xi}_{12}$							3.6
% Var acc		2.2	3.1	3.0	3.6	3.4	2.9

TABLE 4.50: PARAMETER ESTIMATES FOR STANDARD DEVIATION FUNCTION FITTED BY A TRUNCATED FOURIER SERIES FOR EVAPORATION, GIVEN A DRY DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\xi}_0$	337.6	337.6	337.6	337.6	337.6	337.6	337.6
$\hat{\xi}_1$		147.7	147.7	147.7	147.7	147.7	147.7
$\hat{\xi}_2$		14.5	14.5	14.5	14.5	14.5	14.5
$\hat{\xi}_3$			-14.6	-14.6	-14.6	-14.6	-14.6
$\hat{\xi}_4$			3.6	3.6	3.6	3.6	3.6
$\hat{\xi}_5$				-21.3	-21.3	-21.3	-21.3
$\hat{\xi}_6$				-21.9	-21.9	-21.9	-21.9
$\hat{\xi}_7$					-10.4	-10.4	-10.4
$\hat{\xi}_8$					-10.2	-10.2	-10.2
$\hat{\xi}_9$						5.6	5.6
$\hat{\xi}_{10}$						-16.0	-16.0
$\hat{\xi}_{11}$							7.1
$\hat{\xi}_{12}$							-40.6
% Var acc		9.2	8.8	8.8	8.3	8.0	8.2

TABLE 4.51: PARAMETER ESTIMATES FOR STANDARD DEVIATION FUNCTION FITTED BY A TRUNCATED FOURIER SERIES FOR SUNSHINE DURATION, GIVEN A DRY DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\xi}_0$	495.2	495.2	495.2	495.2	495.2	495.2	495.2
$\hat{\xi}_1$		-122.9	-122.9	-122.9	-122.9	-122.9	-122.9
$\hat{\xi}_2$		-48.3	-48.3	-48.3	-48.3	-48.3	-48.3
$\hat{\xi}_3$			-57.8	-57.8	-57.8	-57.8	-57.8
$\hat{\xi}_4$			-16.8	-16.8	-16.8	-16.8	-16.8
$\hat{\xi}_5$				54.9	54.9	54.9	54.9
$\hat{\xi}_6$				6.5	6.5	6.5	6.5
$\hat{\xi}_7$					27.4	27.4	27.4
$\hat{\xi}_8$					-16.4	-16.4	-16.4
$\hat{\xi}_9$						-14.7	-14.7
$\hat{\xi}_{10}$						-40.2	-40.2
$\hat{\xi}_{11}$							-32.8
$\hat{\xi}_{12}$							-3.0
% Var acc		1.8	1.8	1.7	1.3	1.0	0.6

TABLE 4.52: PARAMETER ESTIMATES FOR STANDARD DEVIATION FUNCTION FITTED BY A TRUNCATED FOURIER SERIES FOR WINDRUN, GIVEN A DRY DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\xi}_0$	449977	449977	449977	449977	449977	449977	449977
$\hat{\xi}_1$		-94463	-94463	-94463	-94463	-94463	-94463
$\hat{\xi}_2$		-80265	-80265	-80265	-80265	-80265	-80265
$\hat{\xi}_3$			-17775	-17775	-17775	-17775	-17775
$\hat{\xi}_4$			28700	28700	28700	28700	28700
$\hat{\xi}_5$				17174	17174	17174	17174
$\hat{\xi}_6$				52781	52781	52781	52781
$\hat{\xi}_7$					-27471	-27471	-27471
$\hat{\xi}_8$					-46450	-46450	-46450
$\hat{\xi}_9$						25449	25449
$\hat{\xi}_{10}$						6325	6325
$\hat{\xi}_{11}$							5706
$\hat{\xi}_{12}$							18542
% Var acc		1.6	1.2	1.1	1.0	0.5	*

* Residual variance exceeds variance of y-variate.

TABLE 4.53: PARAMETER ESTIMATES FOR STANDARD DEVIATION FUNCTION FITTED BY A TRUNCATED FOURIER SERIES FOR MAXIMUM HUMIDITY, GIVEN A DRY DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\xi}_0$	10168	10168	10168	10168	10168	10168	10168
$\hat{\xi}_1$		98	98	98	98	98	98
$\hat{\xi}_2$		-390	-390	-390	-390	-390	-390
$\hat{\xi}_3$			1378	1378	1378	1378	1378
$\hat{\xi}_4$			1805	1805	1805	1805	1805
$\hat{\xi}_5$				-824	-824	-824	-824
$\hat{\xi}_6$				1193	1193	1193	1193
$\hat{\xi}_7$					-2783	-2783	-2783
$\hat{\xi}_8$					-936	-936	-936
$\hat{\xi}_9$						2871	2871
$\hat{\xi}_{10}$						422	422
$\hat{\xi}_{11}$							571
$\hat{\xi}_{12}$							-811
% Var acc		*	*	*	0.5	1.4	1.0

* Residual variance exceeds variance of y-variate.

TABLE 4.54: PARAMETER ESTIMATES FOR STANDARD DEVIATION FUNCTION FITTED BY A TRUNCATED FOURIER SERIES FOR MINIMUM HUMIDITY, GIVEN A DRY DAY

Param	L = 1	L = 3	L = 5	L = 7	L = 9	L = 11	L = 13
$\hat{\xi}_0$	13454	13454	13454	13454	13454	13454	13454
$\hat{\xi}_1$		-5885	-5885	-5885	-5885	-5885	-5885
$\hat{\xi}_2$		-757	-757	-757	-757	-757	-757
$\hat{\xi}_3$			790	790	790	790	790
$\hat{\xi}_4$			131	131	131	131	131
$\hat{\xi}_5$				-184	-184	-184	-184
$\hat{\xi}_6$				1684	1684	1684	1684
$\hat{\xi}_7$					464	464	464
$\hat{\xi}_8$					-1093	-1093	-1093
$\hat{\xi}_9$						1162	1162
$\hat{\xi}_{10}$						-423	-423
$\hat{\xi}_{11}$							415
$\hat{\xi}_{12}$							173
% Var acc		14.4	14.2	14.9	15.1	15.3	9.0

TABLE 4.55: MODEL SELECTION FOR THE STANDARD DEVIATION FUNCTION FOR MAXIMUM TEMPERATURE GIVEN THAT A DRY DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	1.8	L = 1
L = 1, P = 5	3.5	L = 1
L = 1, P = 7	3.3	L = 1
L = 1, P = 9	3.1	L = 1
L = 1, P = 11	3.5	L = 1
L = 1, P = 13	3.5	L = 1

TABLE 4.56: MODEL SELECTION FOR THE STANDARD DEVIATION FUNCTION FOR MINIMUM TEMPERATURE GIVEN THAT A DRY DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	2.2	L = 1
L = 1, P = 5	3.1	L = 1
L = 1, P = 7	3.0	L = 1
L = 1, P = 9	3.6	L = 1
L = 1, P = 11	3.4	L = 1
L = 1, P = 13	2.9	L = 1

TABLE 4.57: MODEL SELECTION FOR THE STANDARD DEVIATION FUNCTION FOR EVAPORATION GIVEN THAT A DRY DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	9.2	L = 3
L = 3, P = 5	less than 0.1	L = 3
L = 3, P = 7	less than 0.1	L = 3
L = 3, P = 9	less than 0.1	L = 3
L = 3, P = 11	less than 0.1	L = 3
L = 3, P = 13	less than 0.1	L = 3

TABLE 4.58: MODEL SELECTION FOR THE STANDARD DEVIATION FUNCTION FOR SUNSHINE DURATION GIVEN THAT A DRY DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	1.8	L = 1
L = 1, P = 5	1.8	L = 1
L = 1, P = 7	1.7	L = 1
L = 1, P = 9	1.3	L = 1
L = 1, P = 11	1.0	L = 1
L = 1, P = 13	0.6	L = 1

TABLE 4.59: MODEL SELECTION FOR THE STANDARD DEVIATION FUNCTION FOR MINIMUM HUMIDITY GIVEN THAT A DRY DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	14.4	L = 3
L = 3, P = 5	less than 0.1	L = 3
L = 3, P = 7	0.5	L = 3
L = 3, P = 9	0.7	L = 3
L = 3, P = 11	0.9	L = 3
L = 3, P = 13	less than 0.1	L = 3

TABLE 4.60: MODEL SELECTION FOR THE STANDARD DEVIATION FUNCTION FOR MAXIMUM HUMIDITY GIVEN THAT A DRY DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	less than 0.1	L = 1
L = 1, P = 5	less than 0.1	L = 1
L = 1, P = 7	less than 0.1	L = 1
L = 1, P = 9	0.5	L = 1
L = 1, P = 11	1.4	L = 1
L = 1, P = 13	1.0	L = 1

TABLE 4.61: MODEL SELECTION FOR THE STANDARD DEVIATION FUNCTION FOR WINDRUN GIVEN THAT A DRY DAY OCCURS

$\gamma = \gamma_L$ vs $\gamma = \gamma_P$	Change in % of variance accounted for	Decision
L = 1, P = 3	1.6	L = 1
L = 1, P = 5	1.2	L = 1
L = 1, P = 7	1.1	L = 1
L = 1, P = 9	1.0	L = 1
L = 1, P = 11	0.5	L = 1
L = 1, P = 13	less than 0.1	L = 1

The resulting time series obtained by subtracting the fitted mean function and by dividing through by the fitted standard deviation function should be a time series with a mean of zero and a standard deviation of unity. Since the mean value functions and the standard deviation functions which were fitted are based on truncated Fourier series, that is, on approximating models, the means of the residual series would not be exactly zero and the standard deviations would not be exactly one. However, deviations in this respect were found to be quite small (Table 4.62).

Another assumption made by the model is that the residual time series follows an autoregressive process of order 1. If this is true then $p_k = p_1^k$ where p_k is the autocorrelation with lag k. This assumption (or more precisely, this approximation) was checked by comparing $\hat{p}_k$, k=1,2,3,4 with $\hat{p}_1^k$, and was found to be reasonable (Tables 4.63 - 4.65).

TABLE 4.62: MEAN AND STANDARD DEVIATION OF RESIDUAL TIME SERIES OBTAINED BY STANDARDIZING THE DATA

Variable	Mean	Std. Dev.
Max Temp	-0.04	0.99
Min Temp	0.00	1.00
Evapo	-0.01	1.01
Sunshine	0.00	0.99
Windrun	0.01	1.01
Max Hum	-0.01	1.03
Min Hum	0.01	1.00

TABLE 4.63: COMPARISON OF $\hat{\rho}_k$ and $\hat{\rho}_1^k$ FOR THE RESIDUAL TIME TIME SERIES OF MAXIMUM AND MINIMUM TEMPERATURE

k	Max Temp		Min Temp	
	$\hat{\rho}_k$	$\hat{\rho}_1^k$	$\hat{\rho}_k$	$\hat{\rho}_1^k$
1	0.46	0.46	0.40	0.40
2	0.16	0.21	0.17	0.16
3	0.06	0.09	0.05	0.07
4	0.01	0.04	-0.02	0.03

TABLE 4.64: COMPARISON OF $\hat{\rho}_k$ and $\hat{\rho}_1^k$ FOR THE RESIDUAL TIME SERIES OF EVAPORATION, SUNSHINE DURATION AND WINDRUN

k	Evapo		Sunshine		Windrun	
	$\hat{\rho}_k$	$\hat{\rho}_1^k$	$\hat{\rho}_k$	$\hat{\rho}_1^k$	$\hat{\rho}_k$	$\hat{\rho}_1^k$
1	0.22	0.22	0.18	0.18	0.27	0.27
2	0.12	0.05	0.05	0.03	0.03	0.07
3	0.06	0.01	-0.02	0.01	0.01	0.02
4	0.03	0.00	-0.03	0.00	0.01	0.00

TABLE 4.65: COMPARISON OF $\hat{\rho}_k$ and $\hat{\rho}_1^k$ FOR THE RESIDUAL TIME SERIES OF MAXIMUM AND MINIMUM HUMIDITY

	Max Hum		Min Hum	
	$\hat{\rho}_k$	$\hat{\rho}_1^k$	$\hat{\rho}_k$	$\hat{\rho}_1^k$
1	0.28	0.28	0.38	0.38
2	0.12	0.08	0.22	0.14
3	0.05	0.02	0.19	0.05
4	0.05	0.01	0.18	0.02

The results of the above checks would suggest that the residual series do seem to satisfy the required assumptions of the model. It is therefore reasonable to approximate each of the seven series by the sum of a seasonal component and a residual component, to approximate the seasonal component by a 3-term Fourier series and finally to approximate the standard deviation of the residual series by a 3-term Fourier approximation.

Model 1, proposed by Richardson (1981) is given by:

$$\chi_{i,t} = A\chi_{i,t-1} + B\varepsilon_{i,t}$$

Figure 4.1 gives the estimated A matrix and Figure 4.2 gives the estimated B matrix.

$$\begin{bmatrix} 0.44 & 0.53 & -0.03 & -0.07 & 0.05 & 0.17 & 0.00 \\ -0.02 & 0.43 & 0.02 & -0.01 & -0.01 & 0.08 & 0.02 \\ 0.10 & 0.20 & -0.04 & 0.00 & 0.05 & -0.30 & -0.11 \\ 0.23 & -0.09 & 0.01 & 0.05 & 0.02 & -0.03 & -0.03 \\ -0.37 & 0.13 & 0.02 & 0.04 & 0.11 & -0.36 & -0.13 \\ -0.04 & -0.10 & -0.01 & 0.10 & 0.00 & 0.21 & 0.10 \\ 0.00 & -0.21 & 0.07 & 0.13 & -0.07 & 0.05 & 0.50 \end{bmatrix}$$

Figure 4.1: The estimated matrix A.

$$\begin{bmatrix} 0.71 & & & & & & \\ -0.02 & 0.91 & & & & & \\ 0.37 & 0.07 & 0.79 & & & & \\ 0.41 & -0.28 & 0.39 & 0.73 & & & \\ -0.22 & 0.18 & 0.34 & -0.02 & 0.75 & & \\ -0.25 & -0.33 & -0.14 & -0.02 & -0.06 & 0.84 & \\ -0.62 & 0.24 & -0.08 & -0.05 & 0.02 & 0.14 & 0.55 \end{bmatrix}$$

Figure 4.2: The estimated matrix B.

4.4 Model 2: Multivariate model for climate data.

The mean function, μ_t, was modelled in the same way as in the previous section and therefore the parameter estimates for the Fourier series approximation are the same as before.

The residual time series resulting from subtracting the fitted mean function from the historical data is examined to test the assumption that it follows an autoregressive process of order 1. Tables 4.66 - 4.68 show the comparison of $\hat{\rho}_k$ and $\hat{\rho}_1^k$, for the residual time series conditioned on the wet status of the day.

TABLE 4.66: COMPARISON OF $\hat{\rho}_k$ and $\hat{\rho}_1^k$ FOR THE RESIDUAL TIME SERIES OF MAXIMUM AND MINIMUM TEMPERATURE GIVEN A WET DAY

k	Max Temp		Min Temp	
	$\hat{\rho}_k$	$\hat{\rho}_1^k$	$\hat{\rho}_k$	$\hat{\rho}_1^k$
1	0.18	0.18	0.28	0.28
2	0.10	0.03	0.14	0.08
3	-0.05	0.01	0.03	0.02
4	0.00	0.00	-0.15	0.01

TABLE 4.67: COMPARISON OF $\hat{\rho}_k$ and $\hat{\rho}_1^k$ FOR THE RESIDUAL TIME SERIES OF EVAPORATION, SUNSHINE DURATION AND WINDRUN GIVEN A WET DAY

k	Evapo		Sunshine		Windrun	
	$\hat{\rho}_k$	$\hat{\rho}_1^k$	$\hat{\rho}_k$	$\hat{\rho}_1^k$	$\hat{\rho}_k$	$\hat{\rho}_1^k$
1	-0.03	-0.03	-0.02	-0.02	0.37	0.37
2	0.00	0.00	0.02	0.00	0.16	0.14
3	0.02	0.00	-0.05	0.00	0.04	0.05
4	0.04	0.00	0.02	0.00	-0.05	0.02

TABLE 4.68: COMPARISON OF $\hat{\rho}_k$ and $\hat{\rho}_1^k$ FOR THE RESIDUAL TIME SERIES OF MAXIMUM AND MINIMUM HUMIDITY, GIVEN A WET DAY

k	Max Hum $\hat{\rho}_k$	Max Hum $\hat{\rho}_1^k$	Min Hum $\hat{\rho}_k$	Min Hum $\hat{\rho}_1^k$
1	0.03	0.03	0.15	0.15
2	-0.06	0.00	0.12	0.02
3	-0.01	0.00	0.07	0.00
4	0.04	0.00	0.14	0.00

From the above tables it can be seen that the residual series at least approximately conform to those of an AR(1) process.

Tables 4.69 - 4.71 show the comparison of $\hat{\rho}_k$ and $\hat{\rho}_1^k$ for the residual time series, given a dry day occurs.

TABLE 4.69: COMPARISON OF $\hat{\rho}_k$ and $\hat{\rho}_1^k$ FOR THE RESIDUAL TIME SERIES OF MAXIMUM AND MINIMUM TEMPERATURE, GIVEN A DRY DAY

k	Max Temp $\hat{\rho}_k$	Max Temp $\hat{\rho}_1^k$	Min Temp $\hat{\rho}_k$	Min Temp $\hat{\rho}_1^k$
1	0.53	0.53	0.44	0.44
2	0.18	0.28	0.18	0.19
3	0.06	0.14	0.03	0.08
4	0.00	0.08	-0.02	0.04

TABLE 4.70: COMPARISON OF $\hat{\rho}_k$ and $\hat{\rho}_1^k$ FOR THE RESIDUAL TIME SERIES OF EVAPORATION, SUNSHINE DURATION AND WINDRUN, GIVEN A DRY DAY

k	Evapo		Sunshine		Windrun	
	$\hat{\rho}_k$	$\hat{\rho}_1^k$	$\hat{\rho}_k$	$\hat{\rho}_1^k$	$\hat{\rho}_k$	$\hat{\rho}_1^k$
1	0.30	0.30	0.22	0.22	0.25	0.25
2	0.19	0.09	0.02	0.04	0.01	0.06
3	0.09	0.03	-0.03	0.01	0.01	0.02
4	0.08	0.01	-0.03	0.00	0.02	0.00

TABLE 4.71: COMPARISON OF $\hat{\rho}_k$ and $\hat{\rho}_1^k$ FOR THE RESIDUAL TIME SERIES OF MAXIMUM AND MINIMUM HUMIDITY

k	Max Hum		Min Hum	
	$\hat{\rho}_k$	$\hat{\rho}_1^k$	$\hat{\rho}_k$	$\hat{\rho}_1^k$
1	0.39	0.39	0.45	0.45
2	0.18	0.15	0.22	0.20
3	0.07	0.06	0.20	0.09
4	0.08	0.02	0.14	0.04

Again the residual time series would seem to satisfy the assumption that they constitute an AR(1) process.

Table 4.72 gives the estimate for the parameter of the autoregressive process of order 1 for dry and for wet days.

TABLE 4.72: ESTIMATES FOR THE PARAMETER OF THE AR(1) PROCESS.

Variable	Wet Days	Dry Days
Max Temp	0.53	0.18
Min Temp	0.44	0.28
Evapo	0.30	-0.03
Sunshine	0.22	-0.02
Windrun	0.25	0.37
Max Hum	0.39	0.03
Min Hum	0.45	0.15

The estimated vector of variance for wet and dry days is given in Figure 4.3

Wet Days	Dry Days
444.6	1039
446.0	663.4
351.2	317.3
720.1	399.1
1293472	392870
620.8	10406
16242	9992

Figure 4.3: Estimated vectors of variance for wet and dry days.

The estimated correlation matrices for wet and dry days is given in Figure 4.4.

Dry days

1.00	-0.08	0.35	0.35	-0.33	-0.18	-0.75
-0.08	1.00	0.17	-0.22	0.26	-0.45	0.24
0.35	0.17	1.00	0.42	0.38	-0.30	-0.33
0.35	-0.22	0.42	1.00	-0.05	-0.11	-0.39
-0.33	0.26	0.38	-0.05	1.00	-0.12	0.19
-0.18	-0.45	-0.30	-0.11	-0.12	1.00	0.21
-0.75	0.24	-0.33	-0.39	0.19	0.21	1.00

Wet days

1.00	0.26	0.39	0.33	0.02	-0.09	-0.34
0.26	1.00	-0.11	-0.36	0.07	0.00	0.38
0.39	-0.11	1.00	0.61	0.07	-0.03	-0.49
0.33	-0.36	0.61	1.00	-0.19	0.02	-0.62
0.02	0.07	0.07	-0.19	1.00	-0.18	0.13
-0.09	0.00	-0.03	0.02	-0.18	1.00	0.02
-0.34	0.38	-0.49	-0.62	0.13	0.02	1.00

Figure 4.4: Estimated correlation matrices for dry and wet days

4.5 Model 3 (ML): Multivariate model for climate data.

The method of maximum likelihood was used here to estimate the model parameters.

The initial estimates for the mean function are the same as those of the two previous models and will therefore not be repeated. The selection of the order of the Fourier series approximation was also based on the initial estimates of the mean function and so a choice of a 3-term Fourier series was made.

The estimation of the parameters is accomplished by iteration. A particular parameter estimate is deemed to have converged when its value changes by less than 0.01% in successive iterations. The estimation procedure is deemed to have converged when all parameter estimates have converged.

The parameter estimates for the mean function, for dry and wet days are given in Table 4.73.

TABLE 4.73: PARAMETER ESTIMATES FOR THE MEAN FUNCTION, FOR DRY AND WET DAYS

Variable	Day Status	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_3$
Max Temp	Dry	195.3	37.85	11.04
	Wet	178.3	40.56	15.21
Min Temp	Dry	102.6	34.19	17.04
	Wet	118.1	30.93	16.97
Evapo	Dry	62.61	41.14	1.44
	Wet	34.07	25.31	0.65
Sunshine	Dry	94.59	20.18	-2.13
	Wet	45.69	19.40	-1.42
Windrun	Dry	1784	293.7	-30.66
	Wet	2359	-82.90	-117.6
Max Hum	Dry	941.9	-9.60	6.33
	Wet	950.6	-9.12	-0.46
Min Hum	Dry	385.0	-60.21	-17.42
	Wet	515.0	-111.7	-7.95

The parameter estimates for the coefficient of the autoregressive process of order 1 are given in Table 4.74. This is given for the four sequences, $\bar{R}\bar{R}$, RR , $\bar{R}R$, $R\bar{R}$.

TABLE 4.74: PARAMETER ESTIMATES FOR THE COEFFICIENT OF THE AR(1) PROCESS, GIVEN A SEQUENCE HAS OCCURRED

Variable	$\bar{R}\bar{R}$	RR	$\bar{R}R$	$R\bar{R}$
Max Temp	0.88	0.14	0.54	0.17
Min Temp	0.44	0.26	0.41	0.35
Evapo	0.32	0.04	0.12	0.10
Sunshine	0.24	0.00	0.38	-0.02
Windrun	0.29	0.28	0.57	0.22
Max Hum	0.47	0.04	0.27	0.06
Min Hum	0.49	0.15	0.43	0.20

The estimates for the variance for the different rain and no rain sequences are given in Table 4.75

The estimates of the correlation matrix for each of the rain and no rain sequences are given in Figure 4.5

TABLE 4.75: ESTIMATES OF THE VARIANCE FOR THE FOUR SEQUENCES OF RAIN AND NO RAIN

Variable	$\bar{R}\bar{R}$	RR	$\bar{R}R$	$R\bar{R}$
Max Temp	1428	427.9	1057	514.0
Min Temp	663.8	450.8	564.8	674.1
Evapo	321.6	392.2	453.0	321.5
Sunshine	399.3	741.4	1072	728.2
Windrun	391797	1325611	1128595	538518
Max Hum	11026	564.3	9655	1229
Min Hum	10340	17758	17539	11027

<u>$\bar{R}\bar{R}$ sequence</u>

1.00	-0.30	0.28	0.39	-0.35	-0.15	-0.76
-0.30	1.00	0.16	-0.22	0.26	-0.44	0.25
0.28	0.16	1.00	0.43	0.38	-0.28	-0.32
0.39	-0.22	0.43	1.00	-0.05	-0.11	-0.38
-0.35	0.26	0.38	-0.05	1.00	-0.10	0.20
-0.15	-0.44	-0.28	-0.11	-0.10	1.00	0.20
-0.76	0.25	-0.32	-0.38	0.20	0.20	1.00

<u>RR sequence</u>

1.00	0.29	0.29	0.31	0.01	-0.02	-0.32
0.29	1.00	-0.07	-0.35	0.09	0.01	0.37
0.29	-0.07	1.00	0.54	0.11	0.01	-0.41
0.31	-0.35	0.54	1.00	-0.20	0.04	-0.60
0.01	0.09	0.11	-0.20	1.00	-0.18	0.14
-0.02	0.01	0.01	0.04	-0.18	1.00	-0.01
-0.32	-0.41	-0.41	-0.60	0.14	-0.01	1.00

$\bar{R}R$ sequence

$$\begin{bmatrix} 1.00 & -0.12 & 0.52 & 0.62 & -0.03 & -0.26 & -0.75 \\ -0.12 & 1.00 & -0.08 & -0.33 & 0.10 & -0.45 & 0.22 \\ 0.52 & -0.08 & 1.00 & 0.63 & 0.21 & -0.06 & -0.49 \\ 0.62 & -0.33 & 0.63 & 1.00 & -0.01 & -0.03 & -0.63 \\ -0.03 & 0.10 & 0.21 & -0.01 & 1.00 & -0.14 & -0.10 \\ -0.26 & -0.45 & -0.06 & -0.03 & -0.14 & 1.00 & 0.26 \\ -0.75 & 0.22 & -0.49 & -0.63 & -0.10 & 0.26 & 1.00 \end{bmatrix}$$

$R\bar{R}$ sequence

$$\begin{bmatrix} 1.00 & 0.08 & 0.23 & 0.27 & -0.16 & -0.12 & -0.38 \\ 0.08 & 1.00 & 0.01 & -0.37 & 0.14 & 0.01 & 0.44 \\ 0.23 & 0.01 & 1.00 & 0.59 & 0.39 & -0.04 & -0.29 \\ 0.27 & -0.37 & 0.59 & 1.00 & 0.01 & -0.02 & -0.48 \\ -0.16 & 0.14 & 0.39 & 0.01 & 1.00 & -0.09 & 0.05 \\ -0.12 & 0.01 & -0.04 & -0.02 & -0.09 & 1.00 & 0.11 \\ -0.38 & 0.44 & -0.29 & -0.48 & 0.05 & 0.11 & 1.00 \end{bmatrix}$$

Figure 4.5: Estimated correlation matrices for the four sequences of rain and no rain

4.6 Model 3(LS): Multivariate model for climate data.

Ordinary least squares was used here to estimate the model parameters.

Initial estimates for the mean function were obtained in the same way as the previous section when maximum likelihood estimation was performed.

The parameters were estimated iteratively, the iteration process stopping when all parameter estimates converge. The same convergence criterion was used as before.

The parameter estimates for the mean function, given that wet or dry status of the day, are given in Table 4.76

The parameter estimates for the coefficient of the autoregressive process of order 1 are given in Table 4.77 for the rain and no rain sequences.

The estimates for the variance for the different rain and no rain sequences are given in Table 4.78.

The estimates of the correlation matrix for each of the rain and no rain sequences are given in Figure 4.6

TABLE 4.76: PARAMETER ESTIMATES FOR THE MEAN FUNCTION, FOR DRY AND WET DAYS

Variable	Day Status	$\hat{\alpha}_o$	$\hat{\alpha}_1$	$\hat{\alpha}_3$
Max Temp	Dry	210.2	43.86	15.55
	Wet	187.5	45.81	17.66
Min Temp	Dry	102.4	34.50	16.92
	Wet	117.9	31.53	16.62
Evapo	Dry	62.58	41.12	1.46
	Wet	34.43	25.70	0.57
Sunshine	Dry	93.24	20.32	-2.76
	Wet	47.08	20.03	-0.99
Windrun	Dry	1777	299.7	-32.78
	Wet	2366	-68.90	-119.0
Max Hum	Dry	915.0	-6.79	3.46
	Wet	925.4	-15.26	-8.71
Min Hum	Dry	389.5	-61.59	-16.11
	Wet	520.5	-112.3	-5.29

TABLE 4.77: PARAMETER ESTIMATES FOR THE COEFFICIENT OF THE AR(1) PROCESS, GIVEN A SEQUENCE HAS OCCURRED

Variable	$\bar{R}\bar{R}$	RR	$\bar{R}R$	$R\bar{R}$
Max Temp	0.77	0.13	0.49	0.12
Min Temp	0.44	0.26	0.41	0.35
Evapo	0.32	0.05	0.13	0.10
Sunshine	0.25	0.01	0.42	-0.01
Windrun	0.29	0.28	0.58	0.22
Max Hum	0.40	0.03	0.19	0.20
Min Hum	0.50	0.13	0.43	0.18

TABLE 4.78: ESTIMATES OF THE VARIANCE FOR THE FOUR SEQUENCES OF RAIN AND NO RAIN

Variable	$\bar{R}\bar{R}$	RR	$\bar{R}R$	$R\bar{R}$
Max Temp	1134	445.3	997.9	538.8
Min Temp	662.9	448.2	554.8	666.0
Evapo	316.7	368.9	431.2	273.1
Sunshine	398.0	713.2	1002	669.2
Windrun	391523	1292766	1099653	522136
Max Hum	10327	613.8	8487	1241
Min Hum	10038	16190	17220	9172

$\bar{R}\bar{R}$ sequence

$$\begin{bmatrix} 1.00 & -0.24 & 0.30 & 0.38 & -0.35 & -0.16 & -0.77 \\ -0.24 & 1.00 & 0.16 & -0.22 & 0.26 & -0.44 & 0.25 \\ 0.30 & 0.16 & 1.00 & 0.43 & 0.38 & -0.29 & -0.32 \\ 0.38 & -0.22 & 0.43 & 1.00 & -0.05 & -0.10 & -0.38 \\ -0.35 & 0.26 & 0.38 & -0.05 & 1.00 & -0.10 & 0.20 \\ -0.16 & -0.44 & -0.29 & -0.10 & -0.10 & 1.00 & 0.20 \\ -0.77 & 0.25 & -0.32 & -0.38 & 0.20 & 0.20 & 1.00 \end{bmatrix}$$

RR sequence

$$\begin{bmatrix} 1.00 & 0.31 & 0.32 & 0.31 & 0.03 & -0.08 & -0.32 \\ 0.31 & 1.00 & -0.07 & -0.34 & 0.09 & 0.00 & 0.37 \\ 0.32 & -0.07 & 1.00 & 0.55 & 0.11 & -0.04 & -0.42 \\ 0.31 & -0.34 & 0.55 & 1.00 & -0.20 & 0.01 & -0.60 \\ 0.03 & 0.09 & 0.11 & -0.20 & 1.00 & -0.19 & 0.14 \\ -0.08 & 0.00 & -0.04 & 0.01 & -0.19 & 1.00 & 0.03 \\ -0.32 & 0.37 & -0.42 & -0.60 & 0.14 & 0.03 & 1.00 \end{bmatrix}$$

$\bar{R}R$ sequence

$$\begin{bmatrix} 1.00 & -0.09 & 0.52 & 0.61 & -0.23 & -0.27 & -0.75 \\ -0.09 & 1.00 & -0.08 & -0.33 & 0.10 & -0.46 & 0.23 \\ 0.52 & -0.08 & 1.00 & 0.63 & 0.21 & -0.06 & -0.49 \\ 0.61 & -0.33 & 0.63 & 1.00 & -0.02 & -0.02 & -0.63 \\ -0.23 & 0.10 & 0.21 & -0.02 & 1.00 & -0.14 & -0.10 \\ -0.27 & -0.46 & -0.06 & -0.02 & -0.14 & 1.00 & 0.26 \\ -0.75 & 0.23 & -0.49 & -0.63 & -0.10 & 0.26 & 1.00 \end{bmatrix}$$

RR̄ sequence

1.00	0.12	0.28	0.28	-0.15	-0.11	-0.35
0.12	1.00	0.01	-0.37	0.14	0.07	0.43
0.28	0.01	1.00	0.59	0.38	-0.06	-0.29
0.28	-0.37	0.59	1.00	0.01	-0.07	-0.48
-0.15	0.14	0.38	0.01	1.00	-0.07	0.05
-0.11	0.07	-0.06	-0.07	-0.07	1.00	0.17
-0.35	0.43	-0.29	-0.48	0.05	0.17	1.00

Figure 4.6: Estimated correlation matrices for the four sequences of rain and no rain

The results described in this chapter would suggest that Model 1, Model 2, Model 3(ML) and Model 3(LS) are not inconsistent with the historical record.

The selected models have the following number of parameters:

Model 1: has 86 parameters.

Model 2: has 68 parameters.

Model 3(ML) and (LS): have 98 parameters each.

The model for rainfall occurrences: has 6 parameters.

The model for rainfall depth: has 4 parameters.

Of course the tests described in this chapter cover only some limited aspects of the fit. The issue of model validation is considered more exhaustively in Chapter 6.

CHAPTER 5

5. ALGORITHMS

This chapter describes the various procedures to be followed during model implementation and later during generation of climate sequences.

The preparation of "custom built" computer programs to carry out the preliminary analysis, to fit the models, to validate the models and finally to generate climate sequences would be an enormous task. It is more economical to make use of existing software packages, namely GLIM, GENSTAT AND BMDP. In the algorithms discussed in this chapter it is assumed that these three packages are available. Several FORTRAN programs were written to perform those tasks which could not be carried out using the standard packages. Listings of these programs (referred to in the algorithms below) are given in Appendix F.

The following algorithms are discussed in this chapter:

- Algorithm for fitting the rainfall model.
- Algorithm for generating artificial rainfall sequences.
- Algorithm for fitting Model 1 to climate sequences.
- Algorithm for generating climate sequences using Model 1.
- Algorithm for fitting Model 2 to climate sequences.
- Algorithm for generating climate sequences using Model 2.
- Algorithm for fitting Model 3 (ML) to climate sequences.
- Algorithm for fitting Model 3 (LS) to climate sequences.

- Algorithm for generating climate sequences using Model 3 (ML) or Model 3 (LS).

5.1 Algorithm for implementing the rainfall model.

The following information is required for the parameter estimation programs and must be computed from the historical record:

NT the number of periods in the year (e.g. 365 for daily data).

NY the number of years of data (including the missing values).

For each t=1,2,...,NT

NW(t) the number of times it was wet in period t-1 and there was an observation in period t .

NWW(t) the number of times it was wet in period t-1 and wet in period t .

ND(t) the number of times it was dry in period t-1 and there was an observation in period t .

NDW(t) the number of times it was dry in period t-1 and wet in period t .

R(i,t) the ith non-zero rainfall depth in period t , i=1,2,...,NR(t)

NR(t) the number of times it was wet in period t .

5.1.1 Algorithm for estimating the probabilities of wet and dry sequences.

Step 1: Prepare data sets NW(t) and NWW(t).

- FORTRAN program (Program 1),
- if any of the NW(t) are equal to zero, then delete time period t from data set.

Step 2: Estimate the parameters for the probability that a wet period follows a wet period.

- GLIM package with Error = binomial
  Link = logit (Program 2).

Step 3: Prepare data sets ND(t) and NDW(t) .

- FORTRAN program (Program 1),
- if any of the ND(t) are equal to zero, delete time period t from the data set.

Step 4: Estimate the parameters for the probability that a wet day follows a dry period.

- GLIM package with Error = binomial
  Link = logit (Program 2).

5.1.2 Algorithm to estimate the mean rainfall in wet periods.

Step 1: Prepare the data sets NR(t) and R(i,t).

- FORTRAN program (Program 1).

Step 2: Estimate the parameters of the mean.

Step 3: Estimate the coefficient of variation.

- FORTRAN program to do Step 2 and Step 3. (Program 3.)

5.2 Algorithm for generating artificial rainfall sequences.

Step 1: Set initial state of day to be dry.

Step 2: Generate uniform random number between 0 and 1, inclusive (U(0,1)).

Step 3: If U(0,1) random number is less than the probability of a wet day following a day with the status of the previous time period then

- the status of the present time period is wet.

Otherwise

- the status of the present time period is dry.

Step 4: If present state is wet then determine the rainfall depth.

Step 5: Repeat steps from Step 2 until enough rainfall sequences have been generated.

- FORTRAN program (Program 4).

5.3 Algorithm for implementing Model 1 to climate sequences.

The following information is required for the parameter estimation programs and must be computed from the historical records:

NT the number of periods in the year.

NY the number of years of data.

NV the number of variables in the model.

For each t=1,2,...,NT and for each variable

m(t) the mean of the climate variable at time t .

s(t) the standard deviation of climate variable at time t .

For each variable do:

Step 1: Condition data set according to the wet or dry status of the day. It is easier to create two new data files, one for climate values when rain was observed at a given period and the other with the climate values when no rain was observed at a given period.

- FORTRAN program (Program 5).

For each new data set do:

Step 2: Compute the daily mean vector, m(t)

- FORTRAN program (Program 6).

Step 3: Estimate the parameters of the mean.

- GENSTAT package with Error = normal
  Link = identity. (Program 7.)

Step 4: Obtain residual time series by subtracting from time series the estimated daily mean function.

Step 5: Compute the daily standard deviation vector, s(t).

- FORTRAN program (Program 8).

Step 6: Estimate the parameters of the standard deviation.

- GENSTAT package with Error = normal
  Link = identity. (Program 7.)

Step 7: Obtain new residual series by dividing previous residual series by the estimated daily standard deviation function.

Step 8: Combine the two residual series (i.e. one for rain values and one for no-rain values) into one residual time series.

Step 9: Estimate the lag 1 autocorrelation coefficient.

- GENSTAT package (Program 9).

Step 10: Once the residual time series has been calculated for each variable, estimate the lag 0 and lag 1 cross-correlation coefficients.

- GENSTAT package (Program 10).

Step 11: Using estimates obtained in Step 9 and Step 10 compute the matrices A and B.

- FORTRAN program (Program 11).

5.4 Algorithm for generating artificial climate sequences using Model 1.

Step 1: Generate rainfall sequence (section 5.2).

For each variable do:

Step 2: Generate a N(0,1) random number.

Step 3: Generate residual time series by:

$$\chi_{i,t} = \hat{A}\,\chi_{i,t-1} + \hat{B}\varepsilon_{i,t} \ .$$

- the initial condition of the residual time series is taken to be equal to zero, i.e. $\chi_{1,0} = 0$

Step 4: Generate climate sequence by:

$$S_{i,t} = \begin{cases} X_{i,t}\,\hat{\sigma}_t^R + \hat{\mu}_t^R & \text{if wet} \\ X_{i,t}\,\hat{\sigma}_t^{\bar{R}} + \hat{\mu}_t^{\bar{R}} & \text{if dry.} \end{cases}$$

Step 5: Repeat all of the above steps until the desired amount of climate sequences have been generated.

- FORTRAN program (Program 12).

5.5 Algorithm for implementing Model 2 to climate sequences.

The information required for parameter estimation is the same as that for Model 1.

For each variable do:

Step 1: Perform Step 1 through to Step 4 of the previous parameter estimation algorithm.

Step 2: Estimate the autocorrelation coefficients of the residual time series.

- GENSTAT package (Program 9).

Step 3: Obtain new residual time series by subtracting the autoregressive process from present residual time series, i.e.

$$e_{i,t} = u_{i,t} - \hat{\Theta}\ {}_{i,t-1}$$

<u>Step 4</u>: Estimate the covariance matrix Σ once the above steps have been performed for each variable in the model and for each data set obtained by conditioning the data set on the wet or dry status of the given day.

- GENSTAT package (Program 10)
- BMDP package (Program 13).

5.6 Algorithm for generating artificial climate sequences using Model 2.

<u>Step 1</u>: Generate rainfall sequence.

For each variable do:

<u>Step 2</u>: Generate $N(0,\hat{\Sigma})$ random number where

$$\hat{\Sigma} = \begin{cases} \hat{\Sigma}^{R} & \text{if it is wet} \\ \hat{\Sigma}^{\bar{R}} & \text{if it is dry.} \end{cases}$$

<u>Step 3</u>: Generate residual time series given by

$$u_{i,t} = \begin{cases} (\hat{\Theta}^{R})^{x} \; u_{i,t-1} + e^{R}_{i,t} & \text{if wet} \\ (\hat{\Theta}^{\bar{R}})^{x} \; u_{i,t-1} + e^{\bar{R}}_{i,t} & \text{if dry} \end{cases}$$

where $\hat{\Theta}$ is taken to decrease exponentially in power as the gap between successive dry and wet days increase. "x" denotes the size of the gap. (e.g. The diagram below shows the value x takes for the dry sequence of days

$$
\begin{array}{ccccccc}
\bar{R} & \bar{R} & R & \bar{R} & R & R & \bar{R} \\
 & \uparrow & & \uparrow & & & \uparrow \\
 & x=1 & & x=2 & & & x=3
\end{array} .)
$$

<u>Step 4</u>: Generate daily climate values by:

$$
S_{i,t} = \begin{cases} \hat{\mu}_t^R + u_{i,t} & \text{if it is wet} \\ \hat{\mu}_t^{\bar{R}} + u_{i,t} & \text{if it is dry.} \end{cases}
$$

<u>Step 5</u>: Repeat all of the above steps until the desired amount of climate sequences have been generated.

- FORTRAN program (Program 14).

5.7 <u>Algorithm for implementing Model 3 (ML) to climate sequences</u>.

The following information is required for the parameter estimation programs and must be computed from the historical records:

NT	the number of periods in the year.
NY	the number of years of data.
NV	the number of variables in the model.
T	the total number of observations.
N(DD)	the set of time periods t such that period t was dry and period $t-1$ was dry, $t=1,2,\ldots,T$.
N(WW)	the set of time periods t such that period t was wet and period $t-1$ was wet.

N(DW)	the set of time periods t such that period t was wet and period t-1 was dry.
N(WD)	the set of time periods t such that period t was dry and period t-1 was wet.
C(DD)	Number of elements in the set N(DD) .
C(WW)	Number of elements in the set N(WW) .
C(DW)	Number of elements in the set C(DW)
C(WD)	Number of elements in the set C(WD)

For each variable do:

Step 1: Estimate initial parameters of the mean function by performing Step 1 through to Step 3 of the algorithm for parameter estimation of Model 1.

Step 2: Prepare the data sets of possible sequences, i.e. N(DD), N(WW) , N(DW) and N(WD) .

- FORTRAN program (Program 15).

Step 3: Prepare the data set to ensure that missing values are not taken into the estimation of the parameters. That is, if there is a missing value at any one period of any climate variable for which there is an observation for rainfall, then the missing value is replaced by the nearest non-missing observation of that variable.

For each possible sequence do:

<u>Step 4</u>: Estimate the parameters of the autoregressive process.

<u>Step 5</u>: Estimate the variance.

<u>Step 6</u>: Estimate the parameters of the mean function.

<u>Step 7</u>: Repeat procedure from Step 3 until convergence has been met by all parameters.

- FORTRAN program (Program 16).

<u>Step 8</u>: Obtain residual time series by

$$e_{i,t} = S_{i,t} - \hat{\mu}_t - \hat{\Theta}(S_{i,t-1} - \hat{\mu}_{t-1})$$

where $\hat{\mu}_t$, $\hat{\Theta}$, $\hat{\mu}_{t-1}$ are chosen depending on which sequence the time periods t and $t-1$ satisfy.

<u>Step 9</u>: From the residual series, estimate the off-diagonal coefficients of the covariance matrix Σ .

- FORTRAN program (Program 17).

5.8 <u>Algorithm for implementing Model 3 (LS) to climate sequences.</u>

The information necessary for parameter estimation programs is the same as for Model 3 (ML).

For each variable do:

<u>Step 1</u>: Perform Step 1 through to Step 3 of the algorithm for maximum likelihood estimation.

For each possible sequence do:

<u>Step 2</u>: Estimate the parameters of the autoregressive process.

<u>Step 3</u>: Estimate the parameters of the mean function.

<u>Step 4</u>: Repeat procedure from Step 2 until convergence is achieved by all parameters.

- FORTRAN program (Program 18).

<u>Step 5</u>: Obtain residual time series by:

$$e_{i,t} = S_{i,t} - \hat{\mu}_t - \hat{\Theta}(S_{i,t-1} - \hat{\mu}_{t-1})$$

where $\hat{\mu}_t$, $\hat{\Theta}$, $\hat{\mu}_{t-1}$ are chosen depending on which sequence the time periods t and $t-1$ satisfy.

<u>Step 6</u>: From the residual series, estimate the covariance matrix Σ .

- FORTRAN program (Program 17).

5.9 Algorithm for generating artificial climate sequences using either Model 3 (ML) or Model 3 (LS).

Step 1: Generate rainfall sequence.

For each variable do:

Step 2: Generate $N(0,\hat{\Sigma})$ random number where $\hat{\Sigma}$ is chosen depending on the present sequence.

Step 3: Generate time series given by:

$$u_{i,t} = \hat{\Theta}\, u_{i,t-1} + e_{i,t}$$

where $\hat{\Theta}$ is chosen depending on the present sequence.

Step 4: Generate climate values by:

$$S_{i,t} = \hat{\mu}_t + u_{i,t}$$

where $\hat{\mu}_t$ is chosen depending on the present sequence.

Step 5: Repeat above steps until the desired amount of climate sequences have been generated.

- FORTRAN program (Program 19).

CHAPTER 6

6. MODEL VALIDATION

Once a model has been identified (i.e. given a structure) and the parameters estimated, it remains to decide whether this model is adequate. Model validation is applied with the object of assessing the performance of the model and to uncover any possible lack of fit. In particular one wants to assess whether the model proposed and parameters estimated preserve the properties of the process being examined. This chapter summarizes the results of the checks which were carried out on each of the three models described in Chapter 3.

6.1 Validation of rainfall model.

The rainfall model has been shown to be satisfactory in the various regions of South Africa (Zucchini and Adamson, 1984). They performed extensive checks on the properties of the model, such as:

(a) The annual mean and standard deviation and the distribution of annual totals and sum of k running totals, $k=1,2,...,5$.

(b) The monthly means and variances.

(c) The expected number of wet days at different times of the year.

(d) The distribution of runs of wet and dry days.

(e) The distribution of n-day extreme rainfalls.

The model adopted was found to preserve these properties.

A number of these checks were repeated in this study. For a more complete model validation procedure see Zucchini and Adamson (1984).

Historical data (daily observations for the years 1978 to 1984) was obtained from the weather station at Elsenburg. Its location is given by a latitude of 33° 51 S and a longitude of 18° 50 E.

Twenty years of simulated daily precipitation were compared with the historical data on an annual, monthly and daily basis.

6.1.1 Validation of annual properties.

Table 6.1 shows a comparison of historical and simulated annual mean and standard deviation.

TABLE 6.1 COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MEAN AND STANDARD DEVIATION

Variable	Historical		Simulated	
	Mean	Std. Dev.	Mean	Std. Dev.
Rainfall	63.78	80.27	67.34	85.79

Neither the mean nor the standard deviation of the simulated annual rainfall totals differ significantly from the historical values. It should be noted that only three term Fourier approximations were used in this model thus giving a total of ten parameters for the daily rainfall component of the climate model.

Table 6.2 shows the comparison of historical and simulated annual mean number of wet days. This property has been adequately preserved by the model. This property is an essential one when modelling the remaining climate variables because the latter are conditioned on the wet or dry status of each day.

TABLE 6.2 MEAN NUMBER OF WET DAYS FOR THE YEAR

Historical	Simulated
90.5	93.3

6.1.2 Validation of monthly properties.

The comparison between historical and simulated monthly means for rainfall is given in Figure 6.1. Except for the month of February, the monthly means have been satisfactorily preserved by the rainfall model. For February the mean rainfall depth has been underestimated.

The monthly standard deviations have been satisfactorily preserved by the rainfall model except perhaps for January and February, in which the standard deviations have been somewhat underestimated (Figure 6.2).

The mean number of wet ways for each month has been adequately preserved by the model (Figure 6.3).

As mentioned above, it is especially important that the occurrence of wet days by season be adequately modelled as the generation of the other climate variables is conditioned on the occurrence of wet or

FIGURE 6.1: HISTORICAL (——) AND SIMULATED (---) MONTHLY MEANS FOR RAINFALL

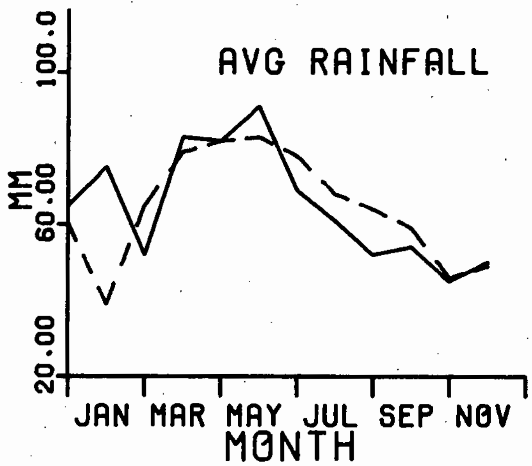


FIGURE 6.2: HISTORICAL (——) AND SIMULATED (---) MONTHLY STANDARD DEVIATIONS FOR RAINFALL

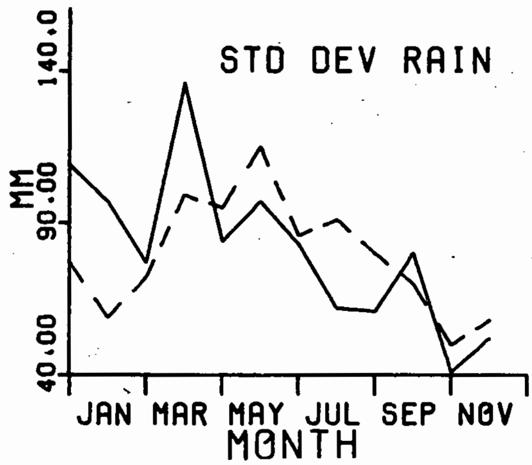
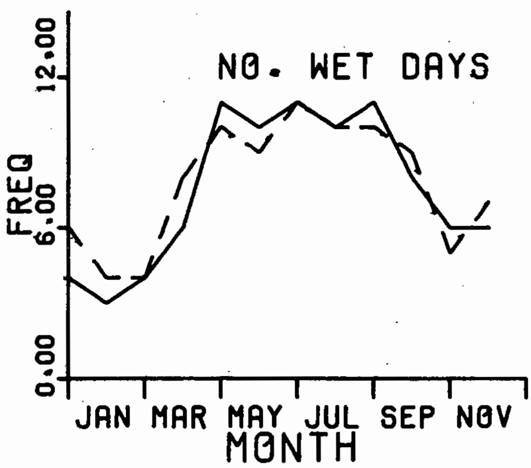


FIGURE 6.3: HISTORICAL (——) AND SIMULATED (---) MEAN NUMBER OF WET DAYS



dry days. The above results indicate that the Markov chain-Weibull model preserves the properties of the rainfall sequence at this location.

6.1.3 Validation of daily properties.

The fits of the truncated Fourier series for the probability of having a wet day given a preceding wet day, and for the probability of having a wet day given the preceding one was dry, are shown in Figure 6.4 and Figure 6.5 respectively. Both fits are generally good.

The truncated Fourier approximation provides a good fit to the mean function of rainfall, but it slightly overestimates the standard deviation (Figure 6.6).

6.2 Validation of Climate Model.

To consider the climate model as adequate in preserving the characteristics of the climate, the multivariate properties of the weather variables must be investigated as well as the univariate characteristics of each individual variable.

The following parameters and parameter functions must be preserved if one is to consider the climate model as satisfactory.

(a) The annual mean and standard deviation for each climate variable.

(b) The annual mean and standard deviation for each variable conditioned on the wet or dry status of the day.

FIGURE 6.4: EMPIRICAL PROBABILITIES AND ESTIMATES BASED ON A 3 PARAMETER MODEL FOR THE PROBABILITY OF A WET DAY GIVEN A PRECEDING WET DAY

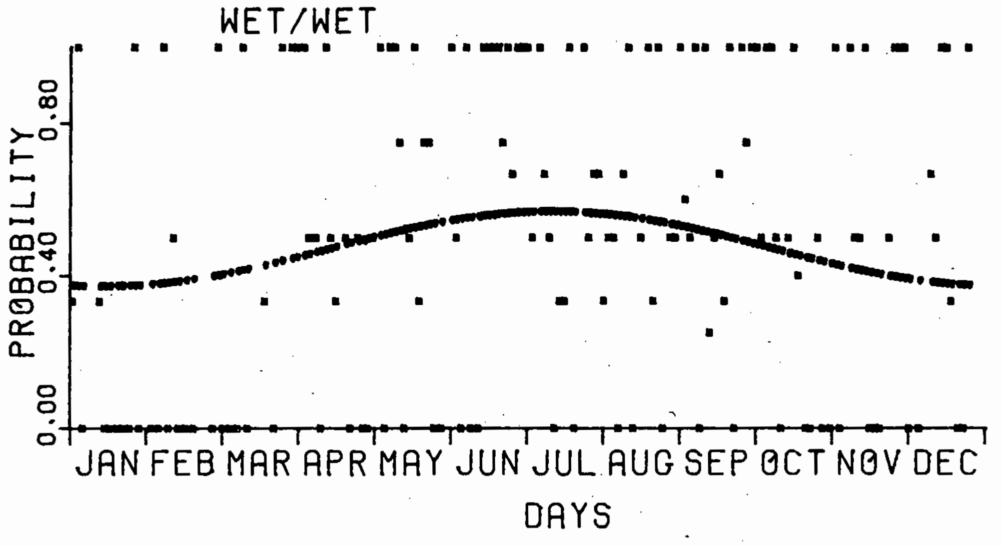


FIGURE 6.5: EMPIRICAL PROBABILITIES AND ESTIMATES BASED ON A 3 PARAMETER MODEL FOR THE PROBABILITY OF A WET DAY GIVEN A PRECEDING DRY DAY

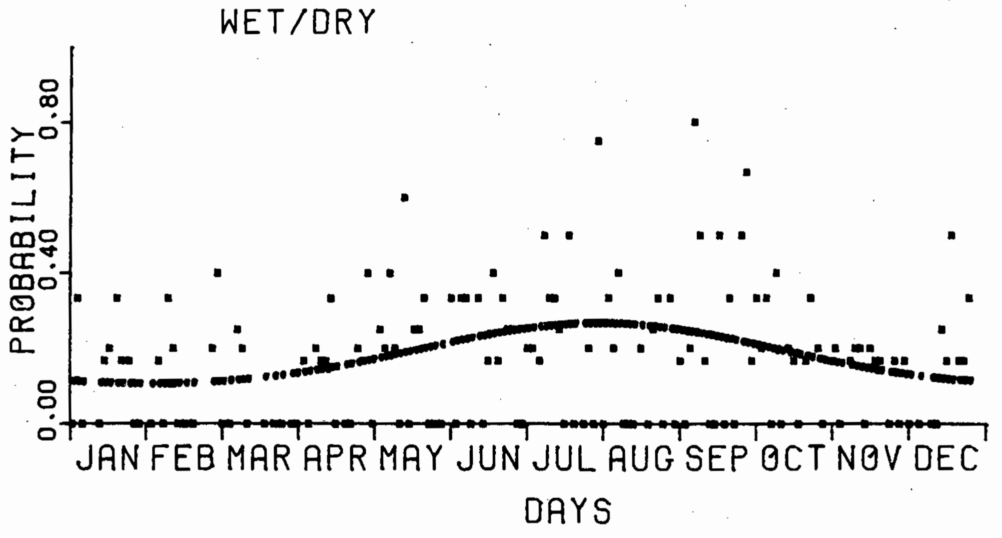
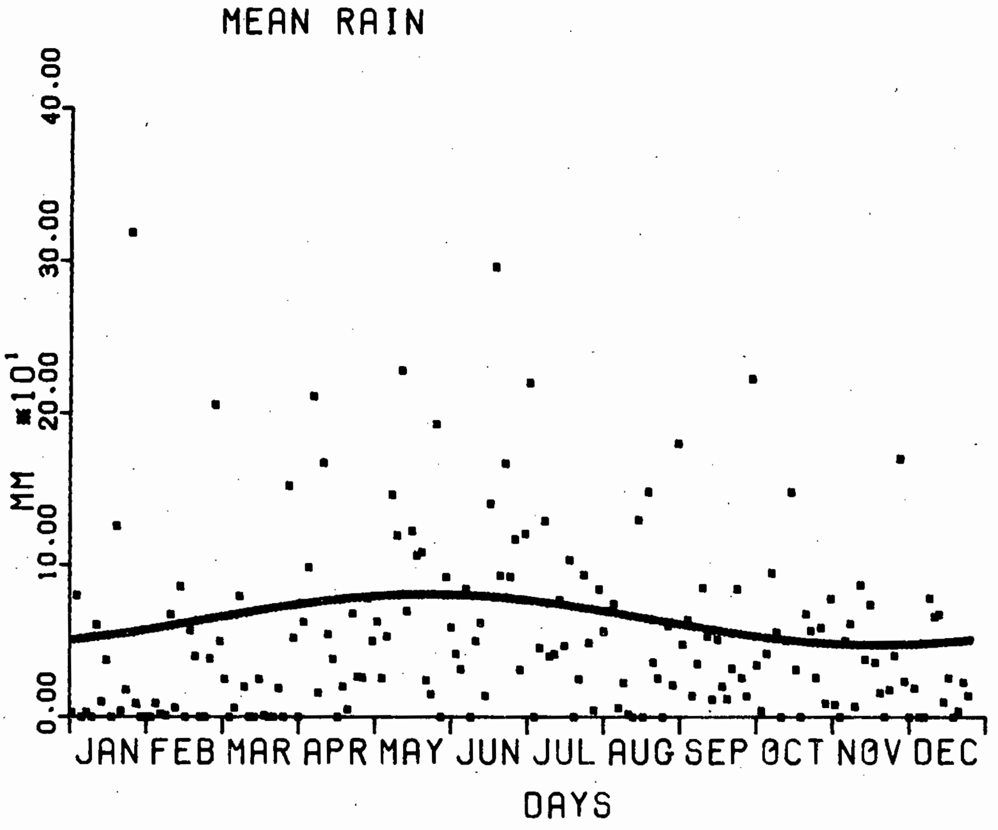
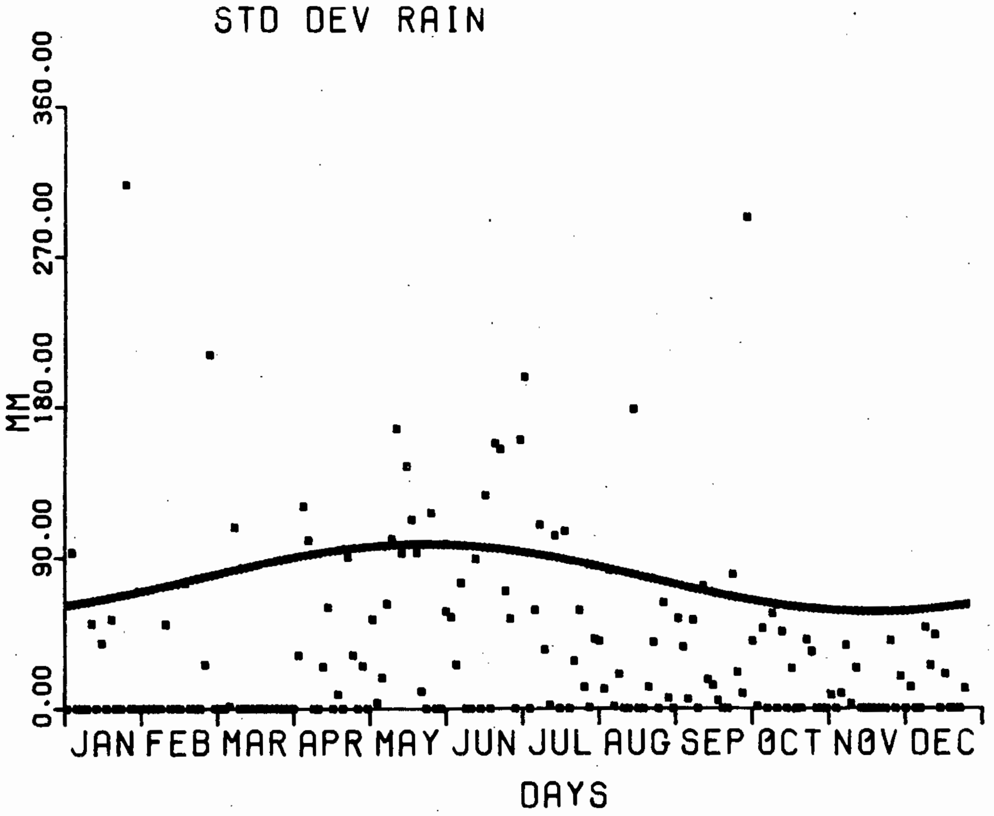


FIGURE 6.6: DAILY AVERAGES AND MEAN FITTED BY A FOURIER SERIES AND DAILY STANDARD DEVIATIONS AND THOSE COMPUTED USING A CONSTANT COEFFICIENT OF VARIATION





(c) The monthly means and standard deviations for each climate variable.

(d) The monthly means and standard deviations for each climate variable conditioned on the wet or dry status of the day.

(e) For the model proposed by Richardson (1981), the standard deviation of each variable as it varies seasonally, conditioned on status of the day.

(f) The expected value of each climate variable as it varies seasonally, conditioned on the status of the day.

(g) The extreme values of each climate variable, i.e. maximum and minimum daily values.

(h) The autocorrelation within each variable.

(i) The autocorrelation within each variable conditioned on the status of the day.

(j) The cross-correlation over all climate variables.

(k) The cross-correlation over all climate variables conditioned on the status of the day.

The checks above test either the multivariate part of the climate model, e.g. the cross-correlation over all variables, or the individual characteristics of each variable, e.g. the monthly means and standard deviations for each variable.

Again twenty years of simulated daily climate sequences were compared with the historical data on an annual, monthly and daily basis.

The following abbreviations for each variable will be adopted:

Max Temp - Maximum Temperature

Min Temp - Minimum Temperature

Evapo - Evaporation

Sunshine - Sunshine Duration

Windrun - Windrun

Max Hum - Maximum Humidity

Min Hum - Minimum Humidity.

6.2.1 Validation of Climate Model 1.

(a) Validation of annual properties.

Table 6.3 shows the comparison of historical and simulated annual mean and standard deviation for each variable.

It can be seen that for every one of the seven variables both the mean and standard deviation have been preserved remarkably well by the model.

TABLE 6.3: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MEAN AND STANDARD DEVIATION

Variable	Historical		Simulated	
	Mean	Std. Dev.	Mean	Std. Dev.
Max Temp	226.5	58.90	228.4	58.45
Min Temp	106.1	38.34	106.9	38.06
Evapo	57.30	37.01	57.80	36.94
Sunshine	82.61	36.74	82.49	36.85
Windrun	1966	911.2	1967	895.9
Max Hum	914.5	96.29	912.1	94.35
Min Hum	412.7	152.8	409.7	151.4

The annual mean and standard deviation are preserved successfully both when the variables are conditioned on a wet day and when they are conditioned on the dry status of the day. These results are shown in Tables 6.4 - 6.5.

TABLE 6.4: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MEAN AND STANDARD DEVIATION GIVEN A WET DAY

Variable	Historical		Simulated	
	Mean	Std. Dev.	Mean	Std. Dev.
Max Temp	182.1	46.73	186.2	47.33
Min Temp	110.1	33.43	109.2	33.77
Evapo	28.49	26.99	29.24	26.22
Sunshine	41.71	33.75	42.01	33.82
Windrun	2500	1194	2483	1185
Max Hum	933.0	72.90	932.1	69.23
Min Hum	548.9	157.7	543.7	154.0

TABLE 6.5: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MEAN AND STANDARD DEVIATION GIVEN A DRY DAY

Variable	Historical		Simulated	
	Mean	Std. Dev.	Mean	Std. Dev.
Max Temp	241.4	54.95	241.9	55.15
Min Temp	104.8	39.77	106.1	39.31
Evapo	66.48	35.00	66.95	35.17
Sunshine	96.16	26.12	95.45	27.08
Windrun	1788	710.4	1801	705.4
Max Hum	908.3	102.3	905.7	100.3
Min Hum	366.7	120.3	366.8	122.7

Table 6.6 shows a comparison of historical and simulated annual maximum and minimum values for each variable.

TABLE 6.6: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MAXIMUM AND MINIMUM VALUES

Variable	Historical		Simulated	
	Maximum	Minimum	Maximum	Minimum
Max Temp	408.0	100.0	441.2	57.94
Min Temp	238.0	13.00	220.1	-21.76
Evapo	185.0	0.00	187.4	-29.28
Sunshine	133.0	0.00	180.1	-61.77
Windrun	7376	293.0	7266	-1671
Max Hum	1000	280.0	1240	511.2
Min Hum	950.0	120.0	1108	-105.6

As can be seen from the above table, the extreme values of the climate variables are not being adequately described by the model. The difficulty that arises with climate data is that it is bounded with values lying outside these boundaries being inadmissible, e.g. having negative sunshine. The historical data showed in many cases that values of variables lied exactly on an upper or lower boundary so that it is expected that simulated sequences will from time to time have values that exceed these boundaries. It remains therefore to see just how many times the simulated values lie outside the limits. To investigate this, a count was taken of those values of the simulated sequence that lied either above the maximum or below the minimum of the historical data. This is shown in Table 6.7. The approximate percentages are given in brackets.

TABLE 6.7: NUMBER OF TIMES SIMULATED VALUES LIE OUTSIDE THE HISTORICAL MAXIMUM AND MINIMUM VALUES.

Variable	Greater than Maximum	Less than Minimum
Max Temp	2 (0%)	652 (9%)
Min Temp	0 (0%)	34 (0.5%)
Evapo	0 (0%)	234 (3%)
Sunshine	83 (1%)	182 (2%)
Windrun	0 (0%)	169 (2%)
Max Hum	489 (7%)	0 (0%)
Min Hum	2 (2%)	48 (0.7%)

From the above table it can be seen that except for maximum humidity, the values of the simulated sequences lie on or below the observed maximum values. Maximum temperature shows a significant number of values lying below the minimum value observed. Suggestions to cope with this weakness are given in the summary at the end of this chapter.

When the data is conditioned on the status of a wet day, the maximum values are preserved to a certain degree by most variables. Maximum and minimum humidity have values which lie outside admissible limits but this can be expected since the observed values lie on or very close to these limits. The same holds when the data is conditioned on the dry status of the day except that now the maximum value for minimum humidity is retained. The minimum values of the simulated sequences on the other hand do not always behave as well for all variables whether conditioned on a wet or dry day. These results are shown in Tables 6.8 - 6.9. This problem remains one of the weakest theoretical aspects of the model. In practice one can of course overcome it by simply truncating the simulated values.

TABLE 6.8: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MAXIMUM AND MINIMUM VALUES GIVEN A WET DAY

Variable	Historical		Simulated	
	Maximum	Minimum	Maximum	Minimum
Max Temp	365.0	100.0	352.3	57.9
Min Temp	238.0	23.0	219.2	16.2
Evapo	122.0	0.00	137.5	-29.28
Sunshine	124.0	0.00	174.1	-61.77
Windrun	73.76	618.0	7266	-1672
Max Hum	1000	340.0	1234	677.1
Min Hum	950.0	170.0	1108	-6.60

TABLE 6.9: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MAXIMUM AND MINIMUM VALUES GIVEN A DRY DAY

Variable	Historical		Simulated	
	Maximum	Minimum	Maximum	Minimum
Max Temp	408.0	120.0	441.2	68.30
Min Temp	234.0	13.00	220.1	-21.80
Evapo	185.0	0.00	187.4	-16.93
Sunshine	133.0	0.00	180.1	-2.44
Windrun	6804	293.0	4073	-1371
Max Hum	1000	280.0	1240	511.2
Min Hum	850.0	120.0	897.3	-105.6

(b) Validation of monthly properties.

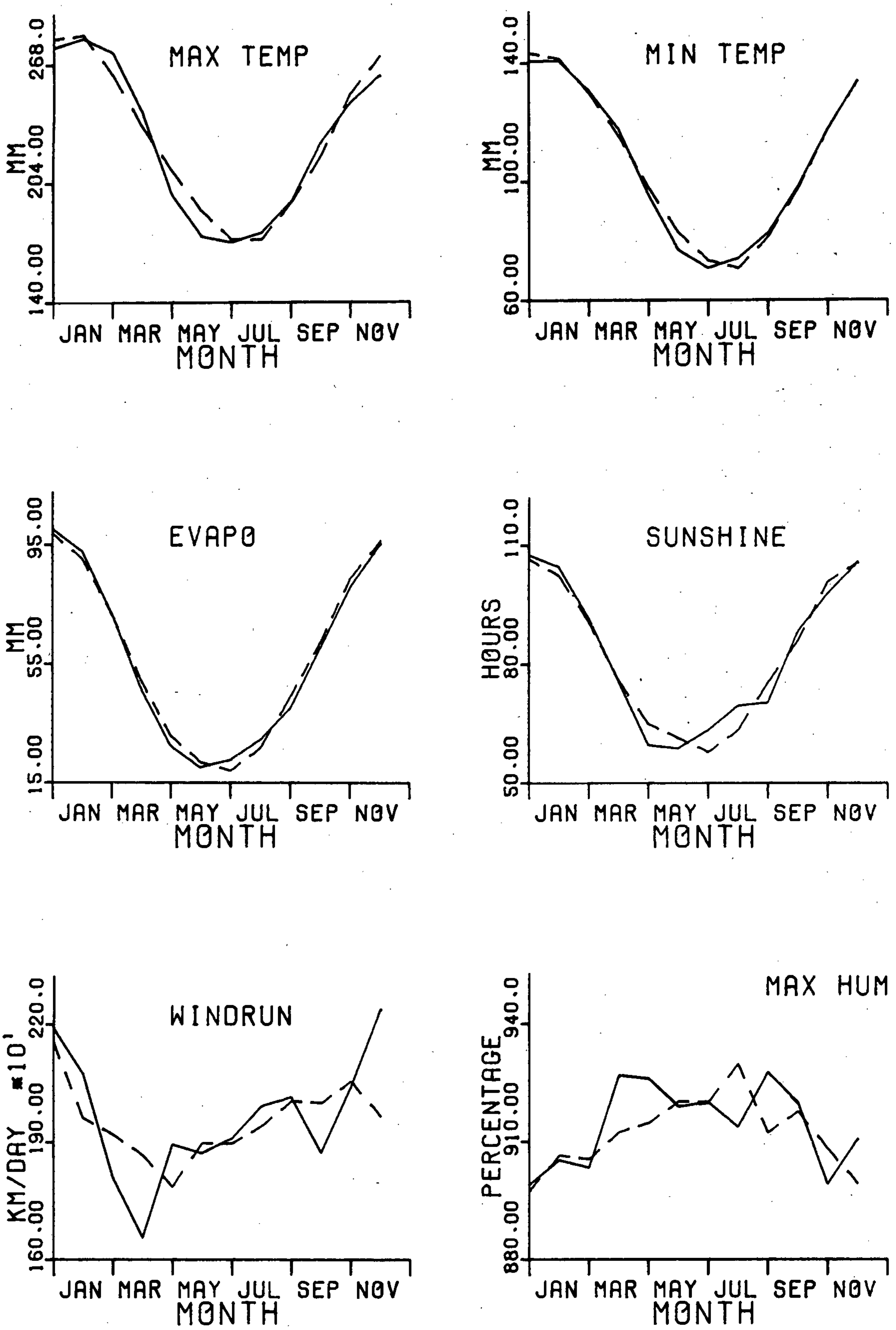
It is important that the monthly characteristics of each climate variable, mainly the mean and standard deviation be adequately described by the model. The monthly means and standard deviations of the simulated sequence were compared to those of the historical record (Figures 6.7 and 6.8).

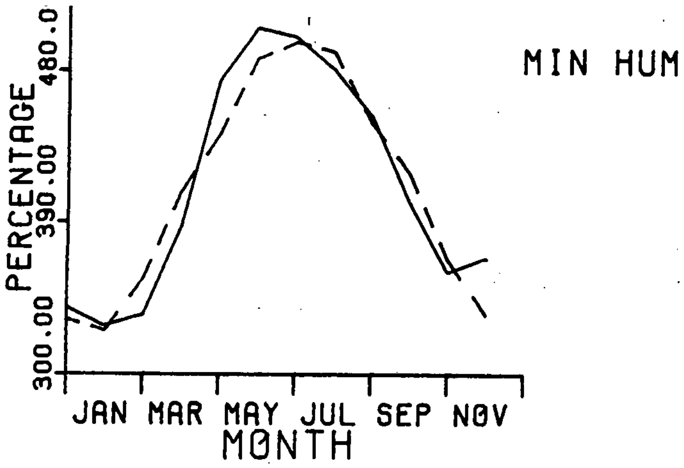
From the figures it can be seen that the monthly means have been preserved very successfully by the model. The variable windrun shows a slight difference between the historical and simulated mean for the month of April, but in fact this represents only a thirteen percent difference between the two values. It can be noted here that the three variables showing a greatest difference between the mean of the observed sequence and that of the generated sequence are maximum and minimum humidity and windrun. Looking at the original sequence for these three variables we see that they do not follow an approximate sinusoidal shape, one of the assumptions made when fitting the mean by a truncated Fourier series, so it seems that these observable differences may be accountable for this. The model is successful in describing the standard deviations of the climate variables except for the months of April and August of maximum humidity and for April and May of windrun.

The monthly means of the conditioned variables on wet and dry days was also examined, the results being shown in Figures 6.9 - 6.10.

It can be concluded that the model fits the monthly means very well both on wet and on dry days. Here again, the variables maximum

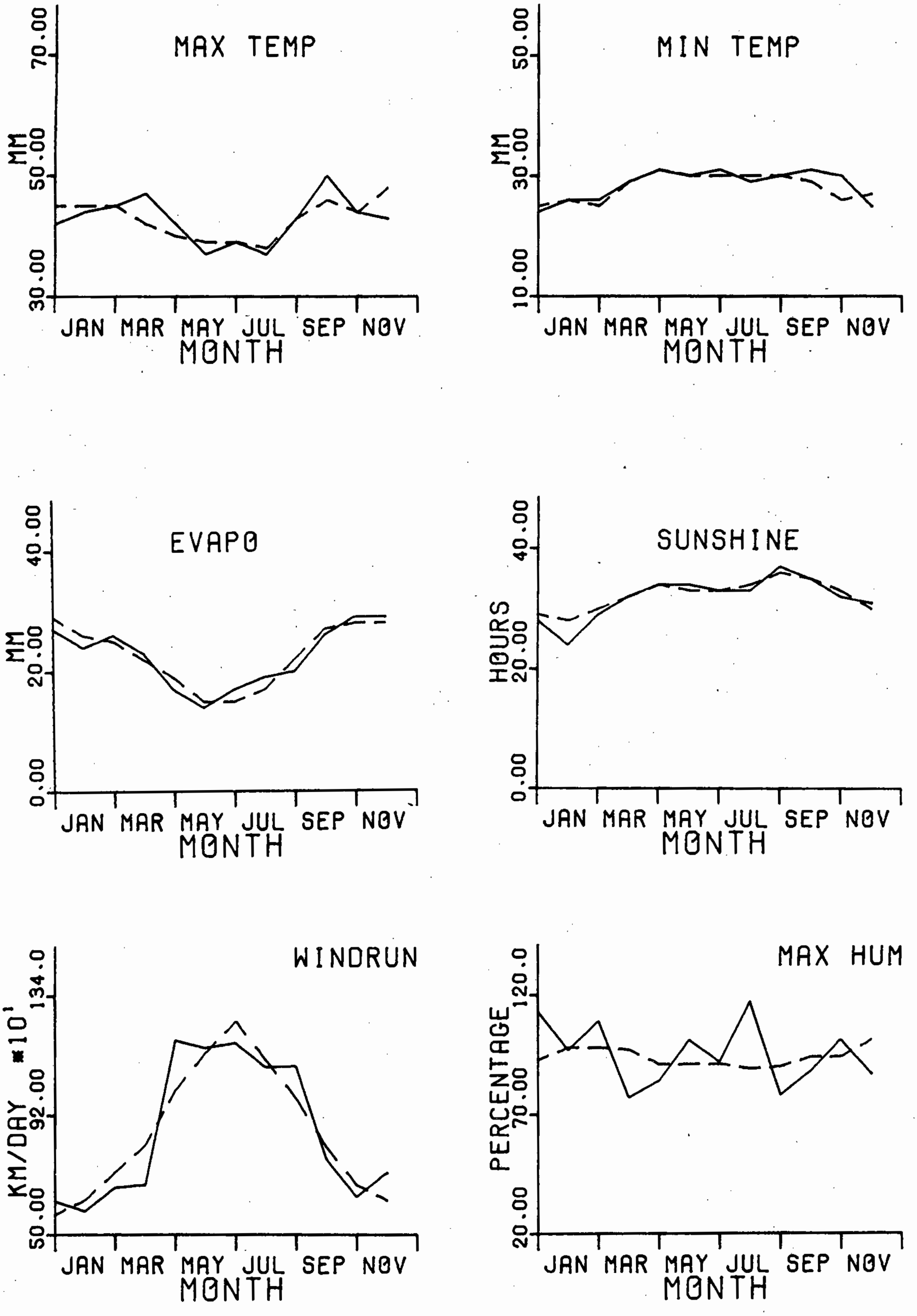
FIGURE 6.7: HISTORICAL (——) AND SIMULATED (---) MONTHLY STANDARD DEVIATIONS FOR ALL VARIABLES





MIN HUM
PERCENTAGE
300.00
390.00
480.0
JAN
MAR
MAY
JUL
SEP
NOV
MONTH

FIGURE 6.8: HISTORICAL (——) AND SIMULATED (---) MONTHLY STANDARD DEVIATIONS FOR ALL VARIABLES



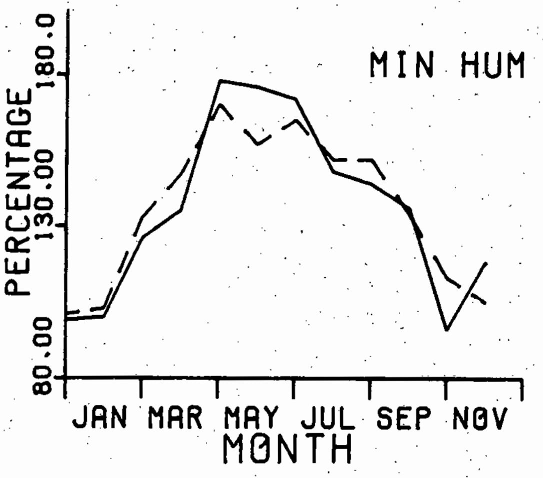
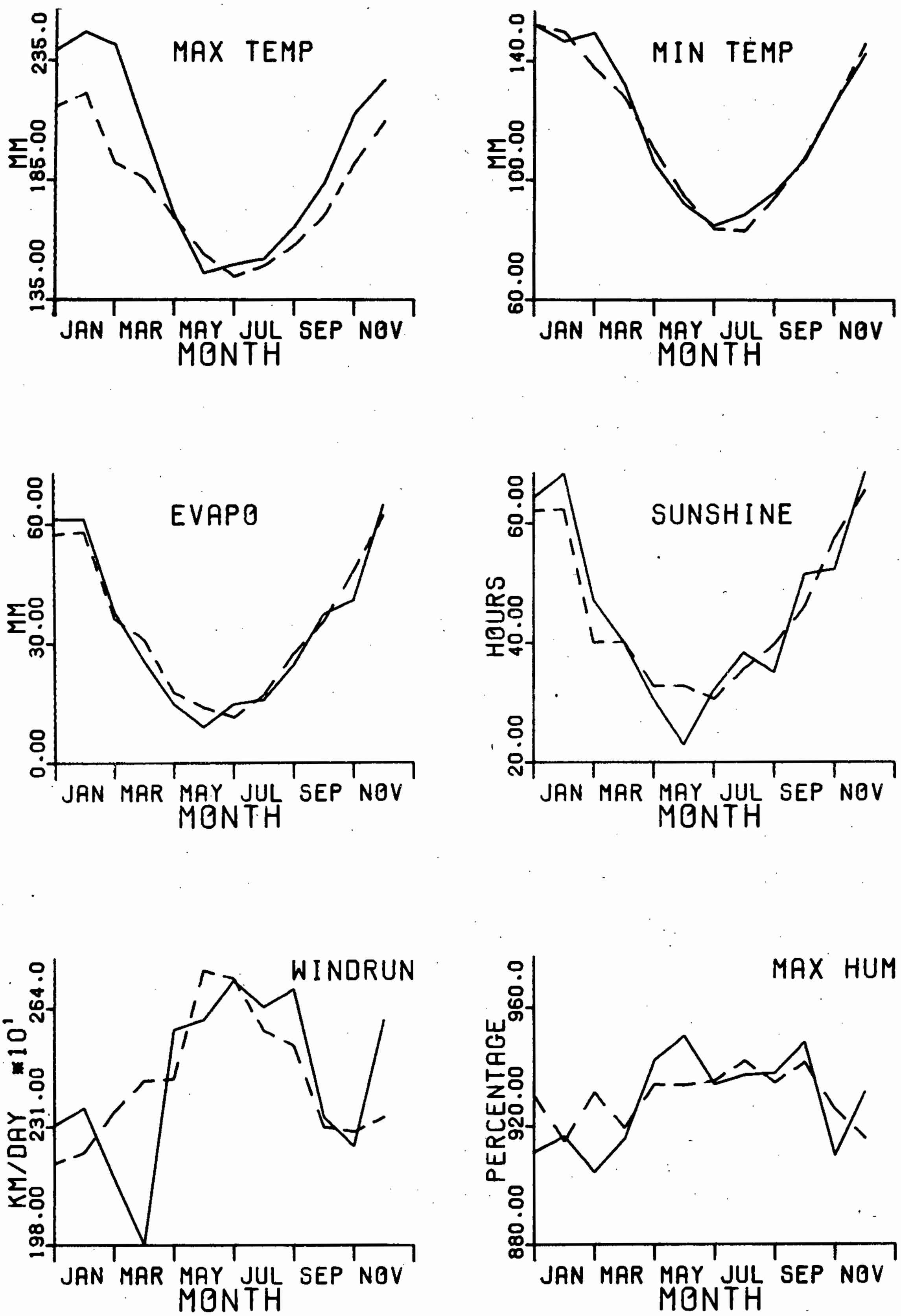

MIN HUM
PERCENTAGE
180.0
130.00
80.00
JAN
MAR
MAY
JUL
SEP
NOV
MONTH

FIGURE 6.9: HISTORICAL (——) AND SIMULATED (---) MONTHLY MEANS GIVEN A WET DAY FOR ALL VARIABLES



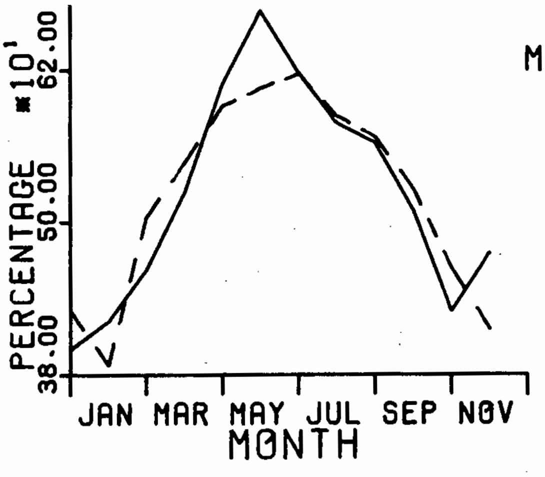
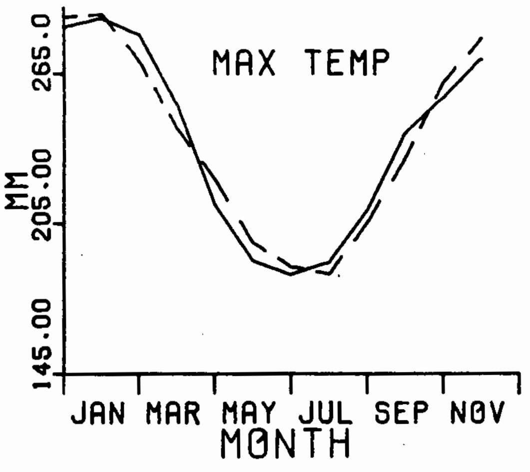
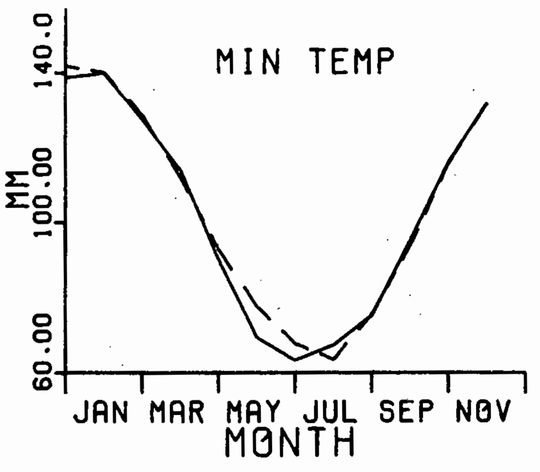
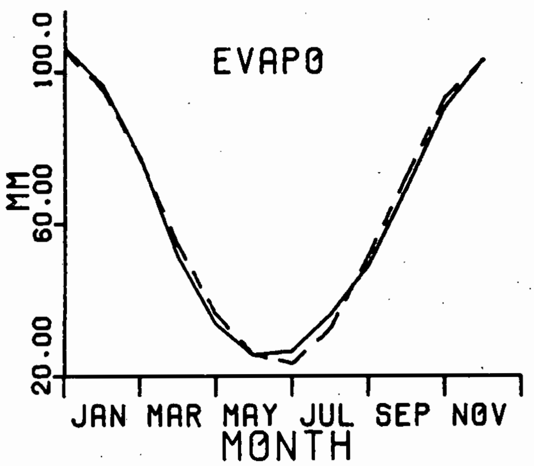
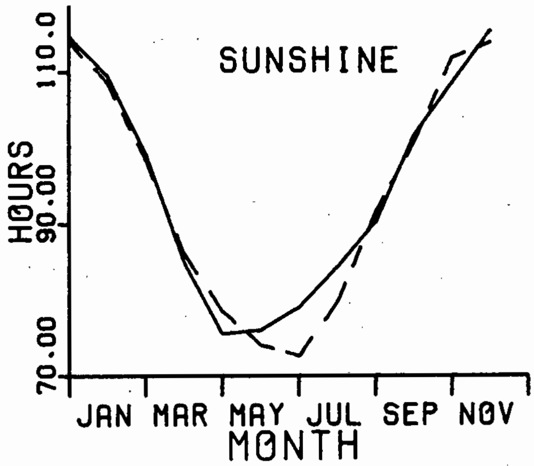


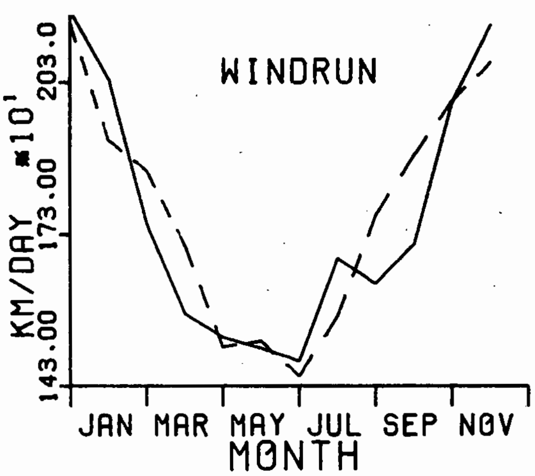
FIGURE 6.10: HISTORICAL (——) AND SIMULATED (---) MONTHLY MEANS GIVEN A DRY DAY FOR ALL VARIABLES

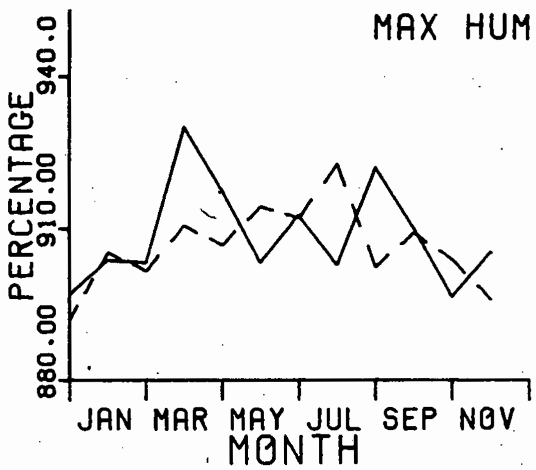


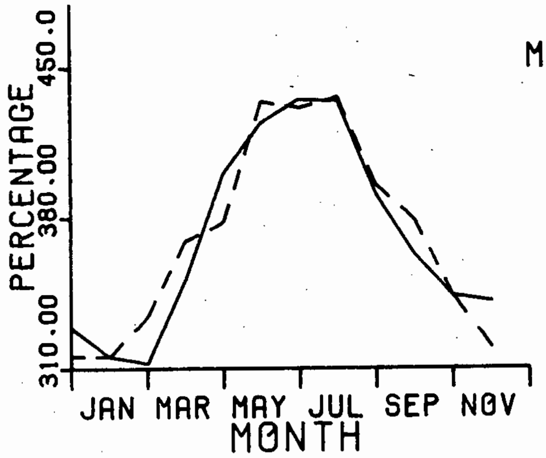



EVAPO
MM
20.00
60.00
100.0
JAN MAR MAY JUL SEP NOV
MONTH


SUNSHINE
HOURS
70.00
90.00
110.0
JAN MAR MAY JUL SEP NOV
MONTH


WINDRUN
KM/DAY *10^1
143.00
173.00
203.0
JAN MAR MAY JUL SEP NOV
MONTH


MAX HUM
PERCENTAGE
880.00
910.00
940.0
JAN MAR MAY JUL SEP NOV
MONTH


MIN HUM
PERCENTAGE
310.00
380.00
450.0
JAN MAR MAY JUL SEP NOV
MONTH

humidity and windrun show the greatest deviations from the historical means for some months.

The monthly standard deviations are preserved for nearly all variables by the model, either when the variables are conditioned on the wet status of the day or on the dry status. The model is more successful in retaining the monthly standard deviations when the variables are conditioned on the dry status of the day. Here only the variables maximum humidity and windrun show quite large differences between historical and simulated sequences. When variables are conditioned on the wet status of the day, the variables minimum humidity and maximum temperature also show differences between historical and simulated standard deviations for some months. These results are shown in Figures 6.11 - 6.12.

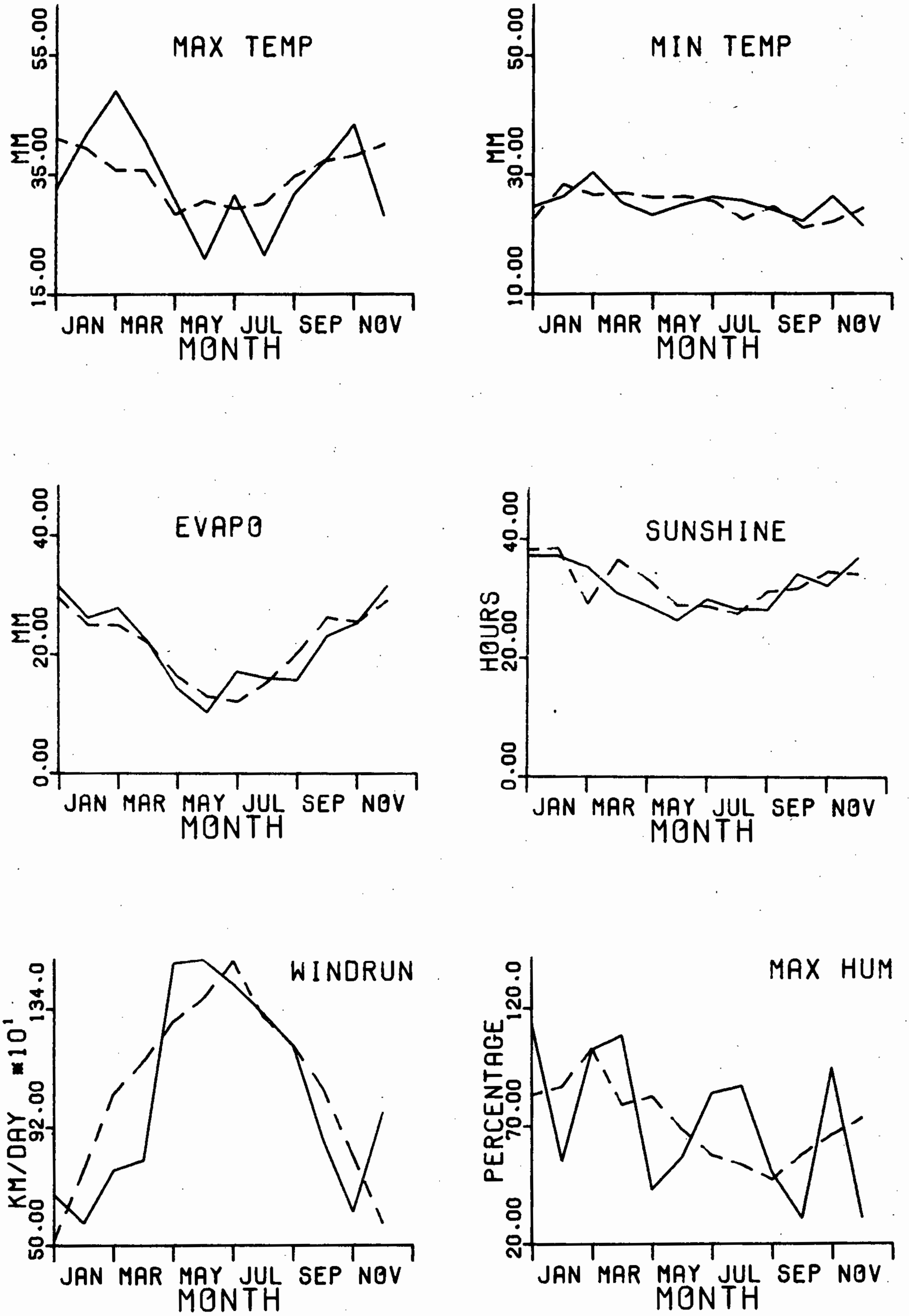
We note that there is a much smaller number of wet days in the year at this station than there are dry days. It would be (statistically) surprising if all the parameter functions associated with wet days fitted the historical record very accurately.

(c) <u>Validation of daily properties.</u>

A visual assessment of the fit of the truncated Fourier series to the mean and standard deviation functions of the various variables can be obtained from Figures 6.13.1 - 6.14.2.

It is clear from these plots that the structure of the means is well maintained by a Fourier series truncated to three terms. This is true for all the climate variables. A truncated Fourier approximation fits

FIGURE 6.11: HISTORICAL (——) AND SIMULATED (---) MONTHLY STANDARD DEVIATION GIVEN A WET DAY FOR ALL VARIABLES



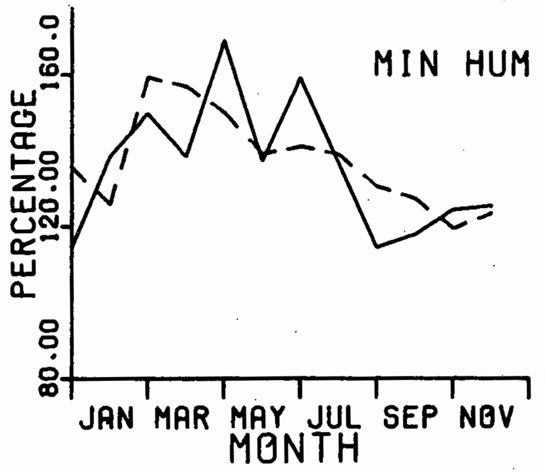
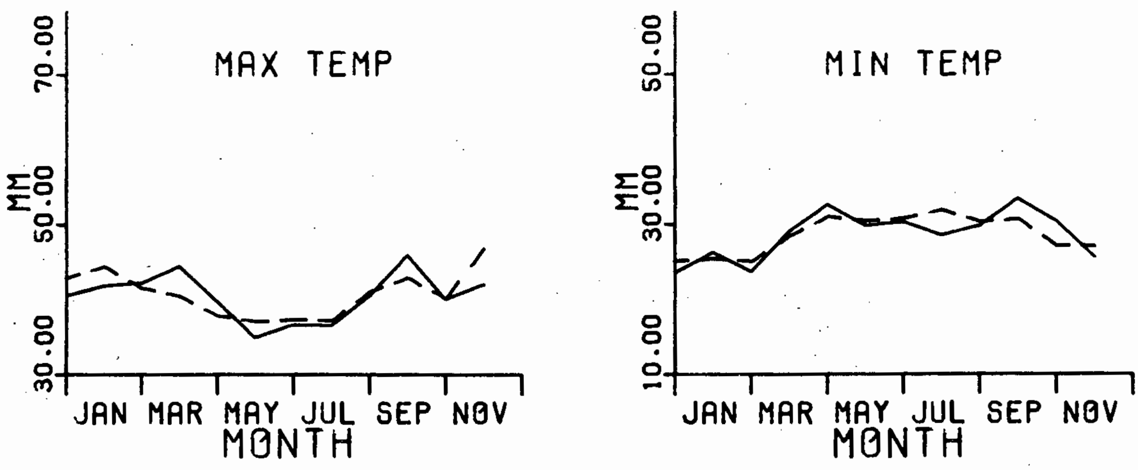


FIGURE 6.12: HISTORICAL (——) AND SIMULATED (---) MONTHLY STANDARD DEVIATIONS GIVEN A DRY DAY FOR ALL VARIABLES



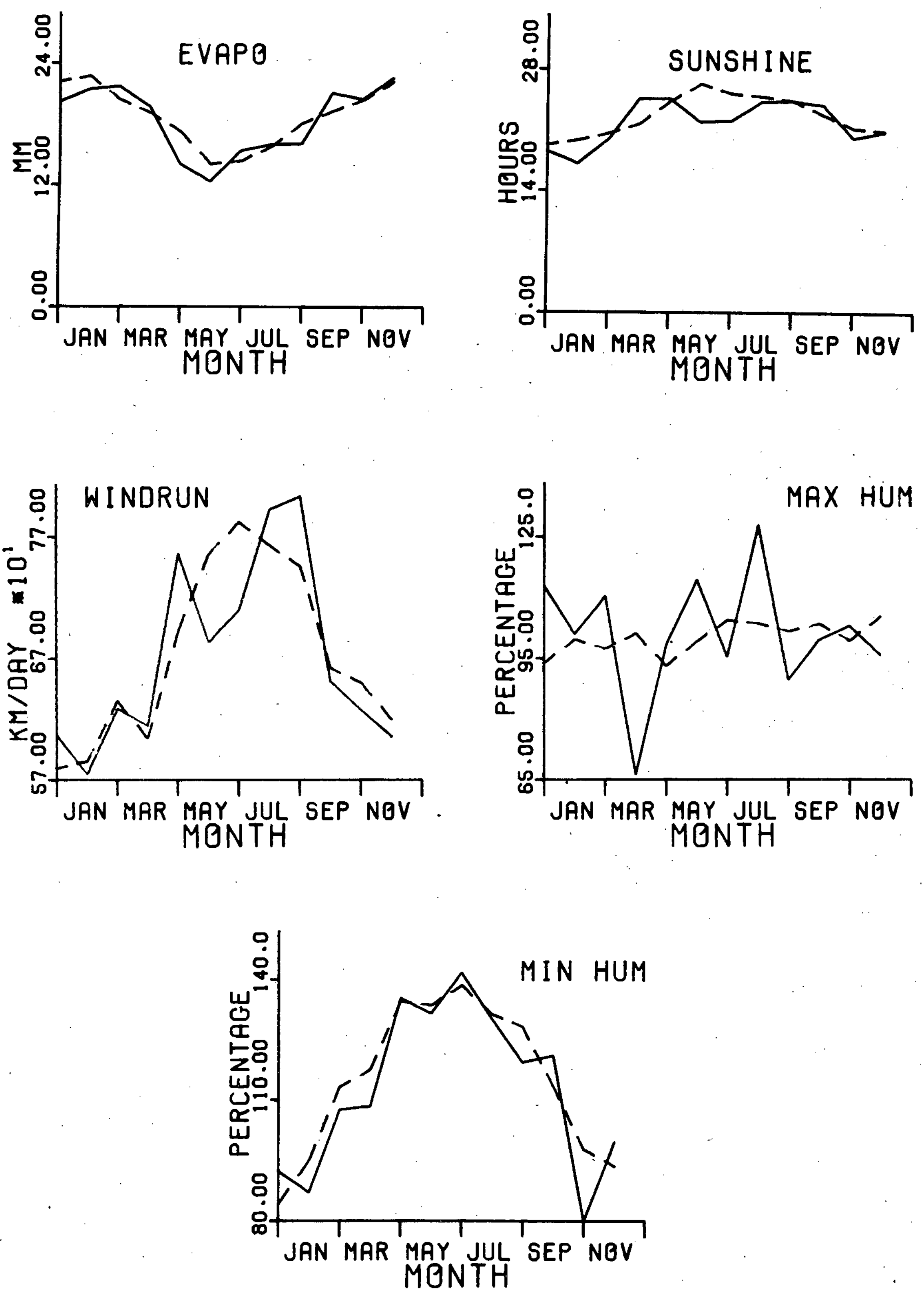
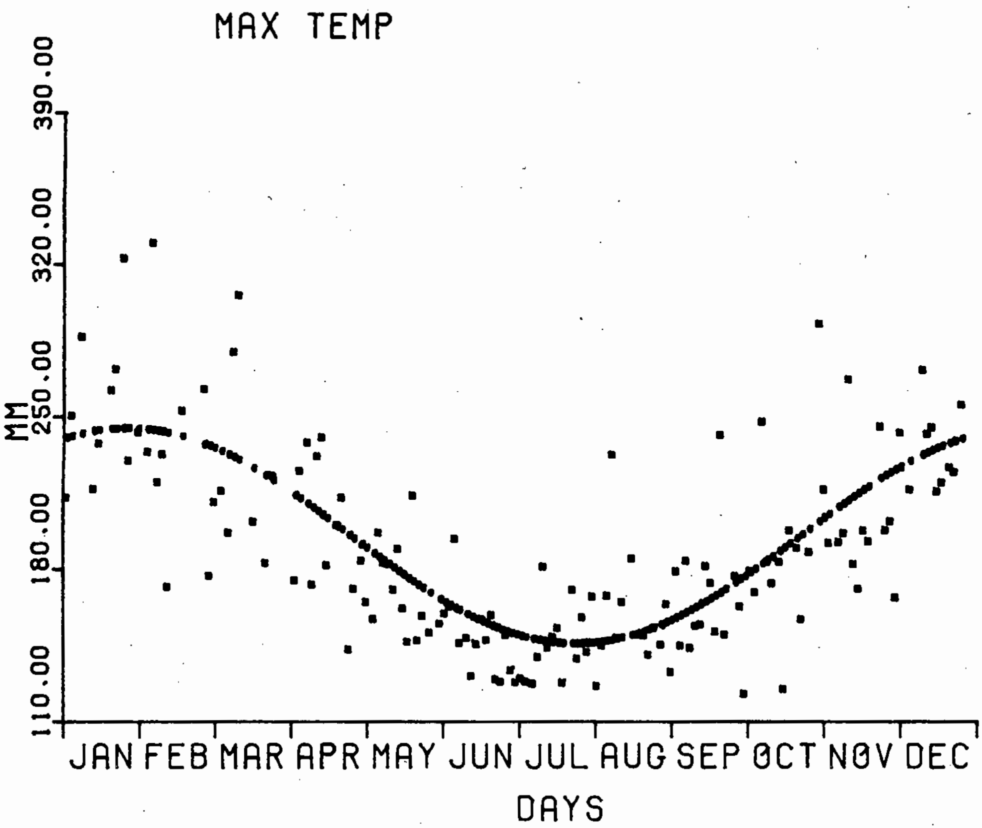
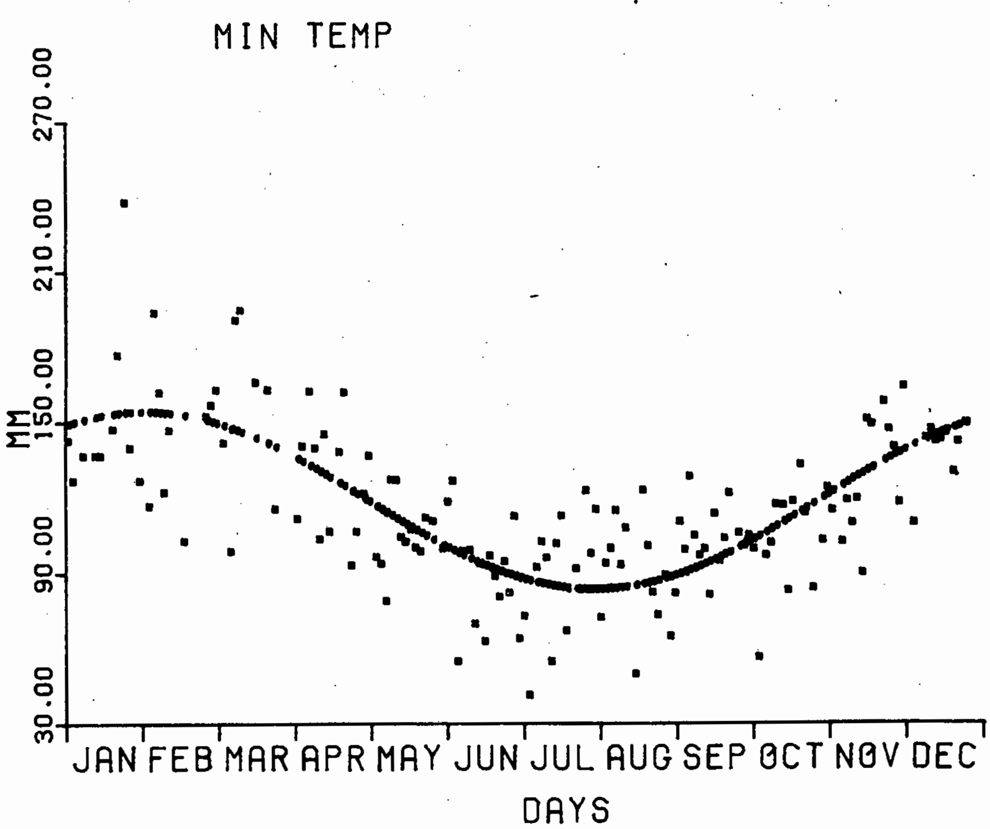
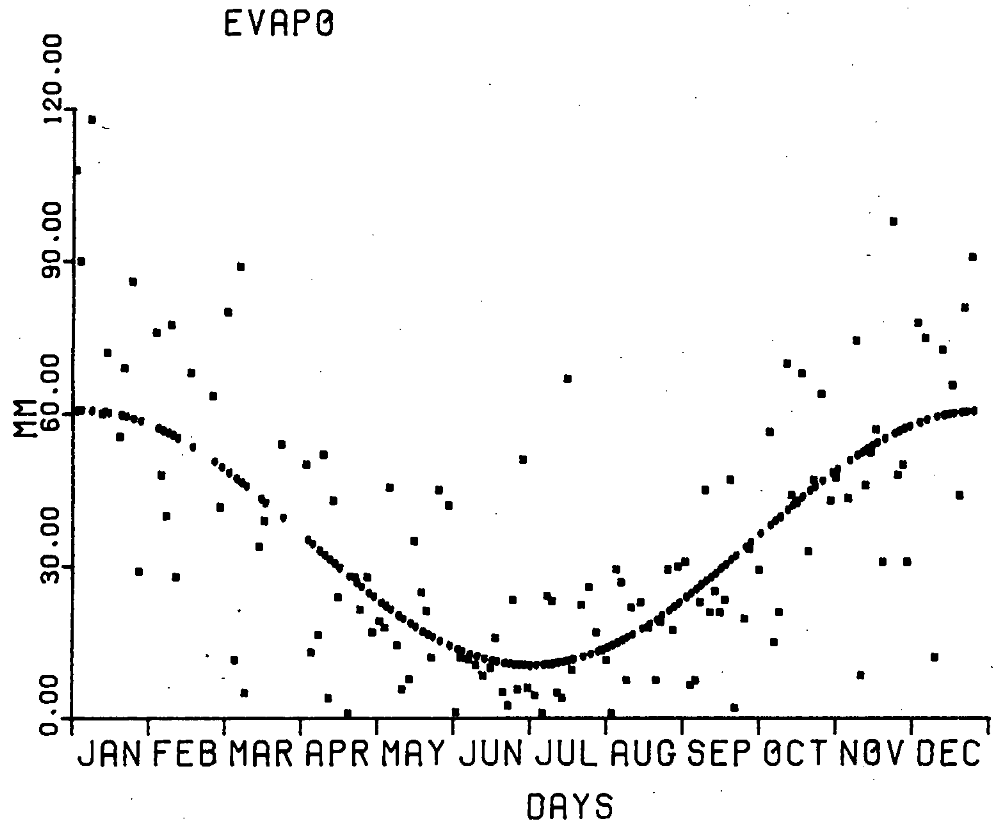
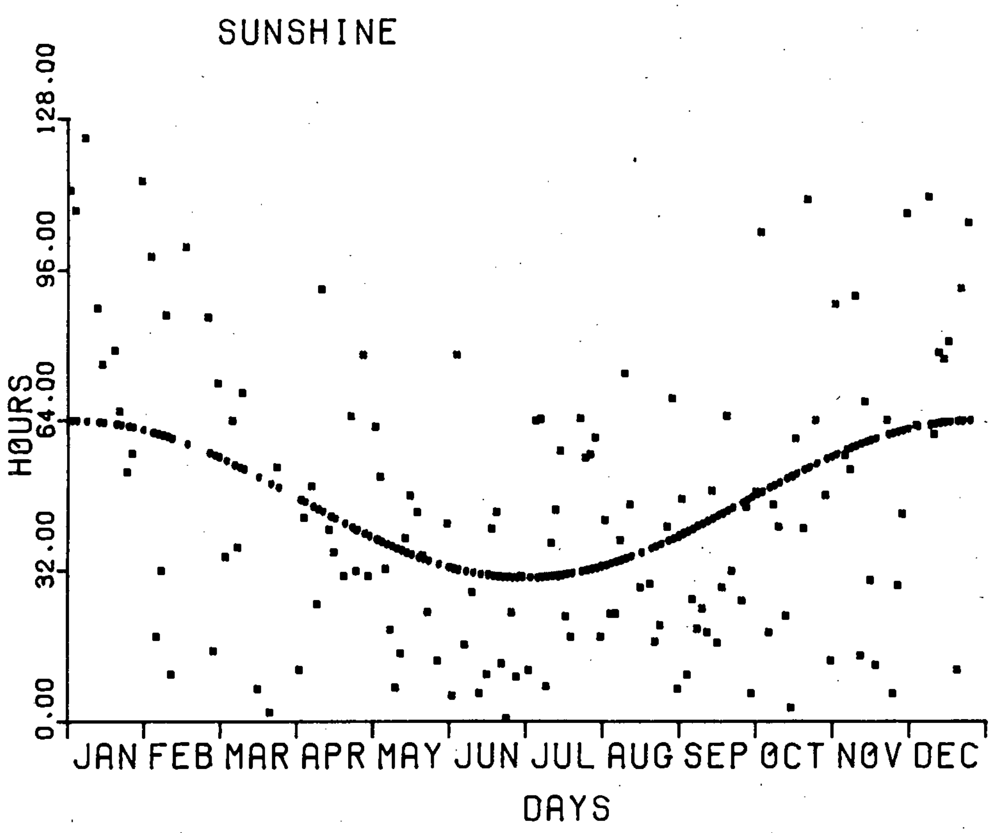

EVAPO
MM
24.00
12.00
0.00
JAN MAR MAY JUL SEP NOV
MONTH
SUNSHINE
HOURS
28.00
14.00
0.00
JAN MAR MAY JUL SEP NOV
MONTH
WINDRUN
KM/DAY *10¹
77.00
67.00
57.00
JAN MAR MAY JUL SEP NOV
MONTH
MAX HUM
PERCENTAGE
125.0
95.00
65.00
JAN MAR MAY JUL SEP NOV
MONTH
MIN HUM
PERCENTAGE
140.0
110.00
80.00
JAN MAR MAY JUL SEP NOV
MONTH

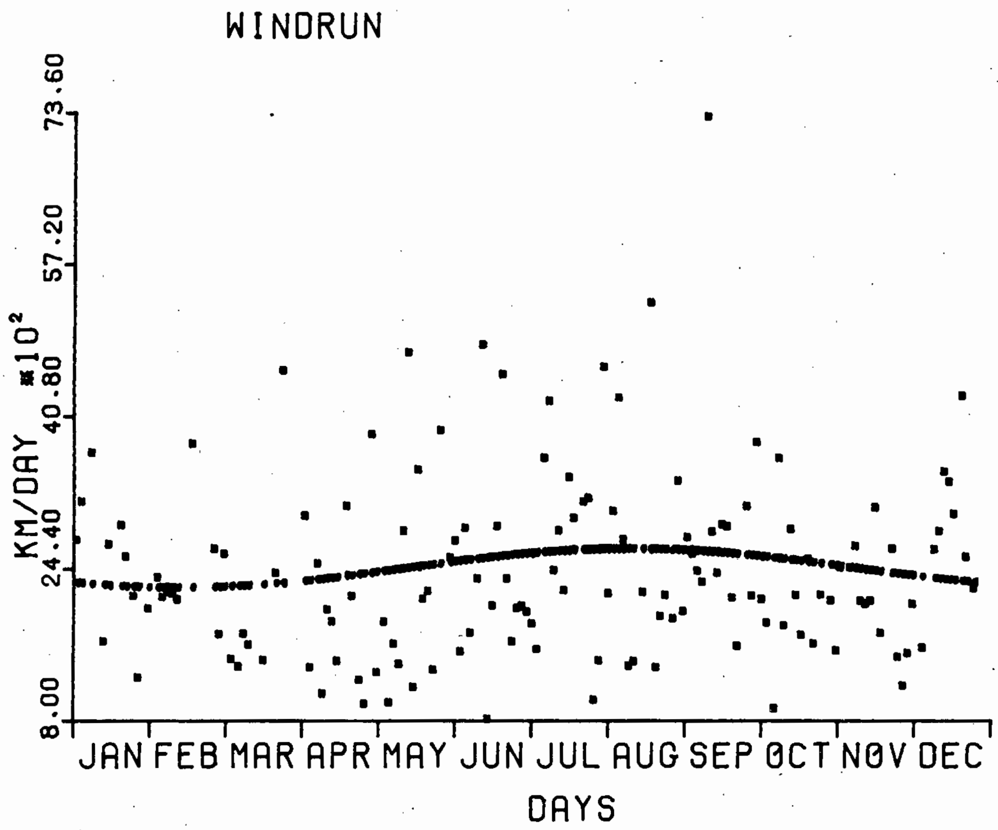
FIGURE 6.13.1: DAILY AVERAGES AND MEAN FITTED BY A FOURIER SERIES GIVEN A WET DAY FOR ALL VARIABLES

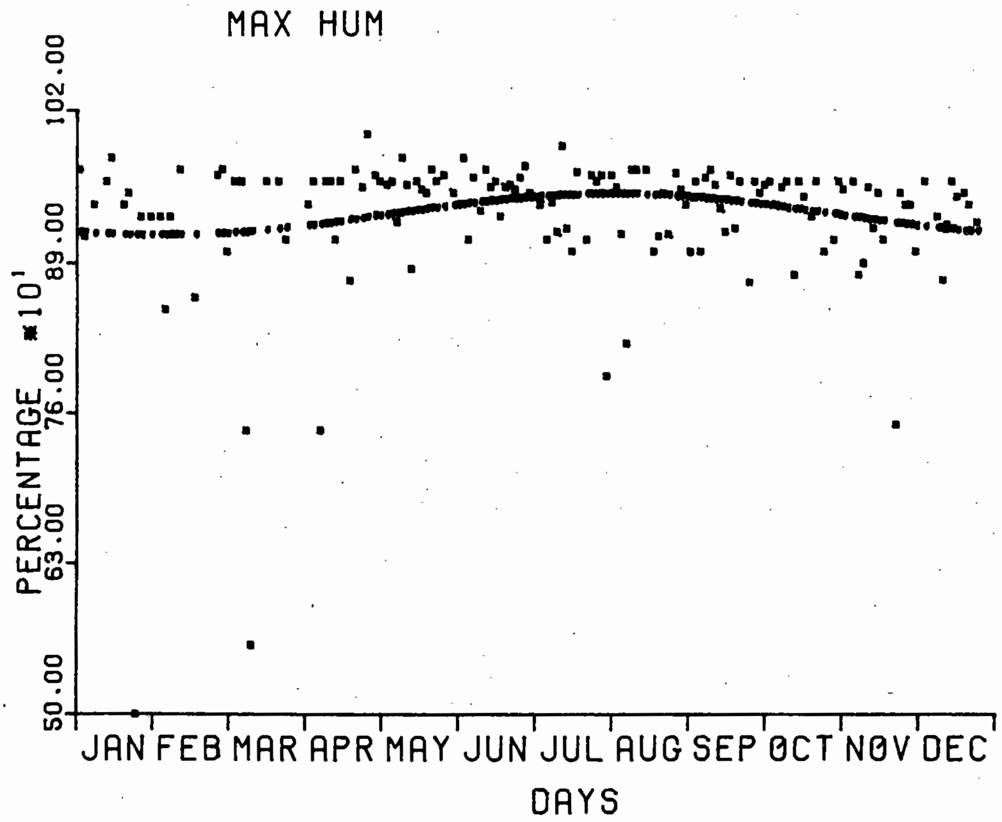





EVAPO
MM
120.00
90.00
60.00
30.00
0.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


SUNSHINE
HOURS
128.00
96.00
64.00
32.00
0.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


WINDRUN
KM/DAY *10²
73.60
57.20
40.80
24.40
8.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


MAX HUM
PERCENTAGE *10¹
102.00
89.00
76.00
63.00
50.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS

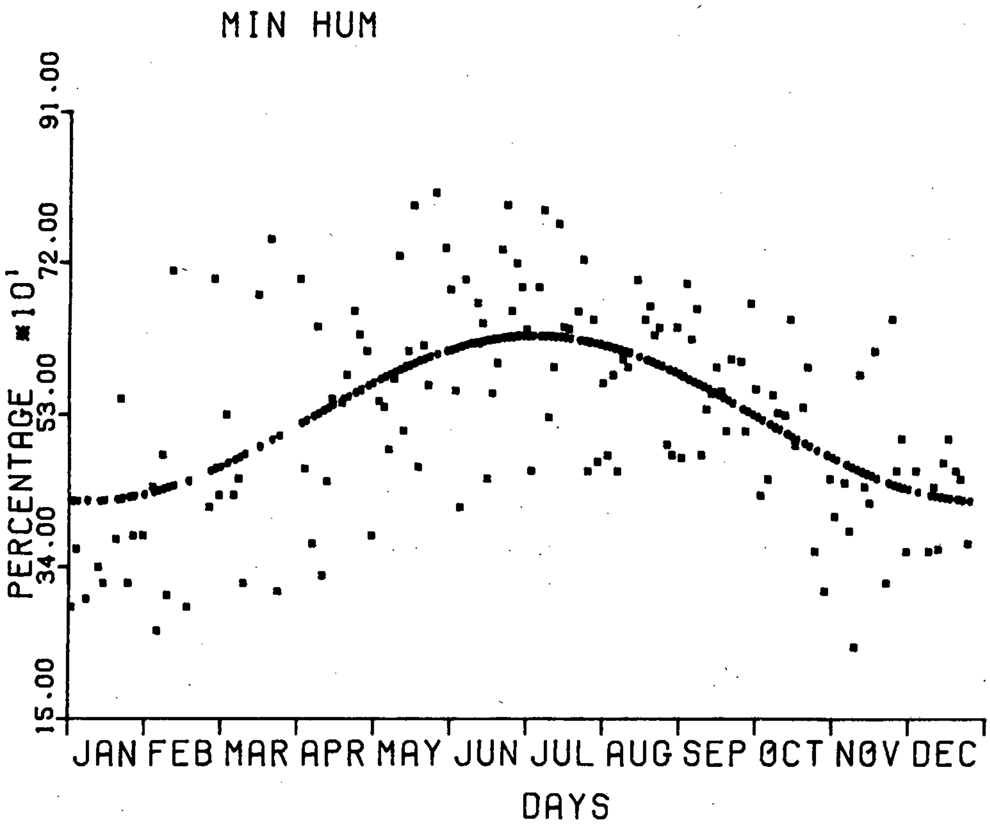
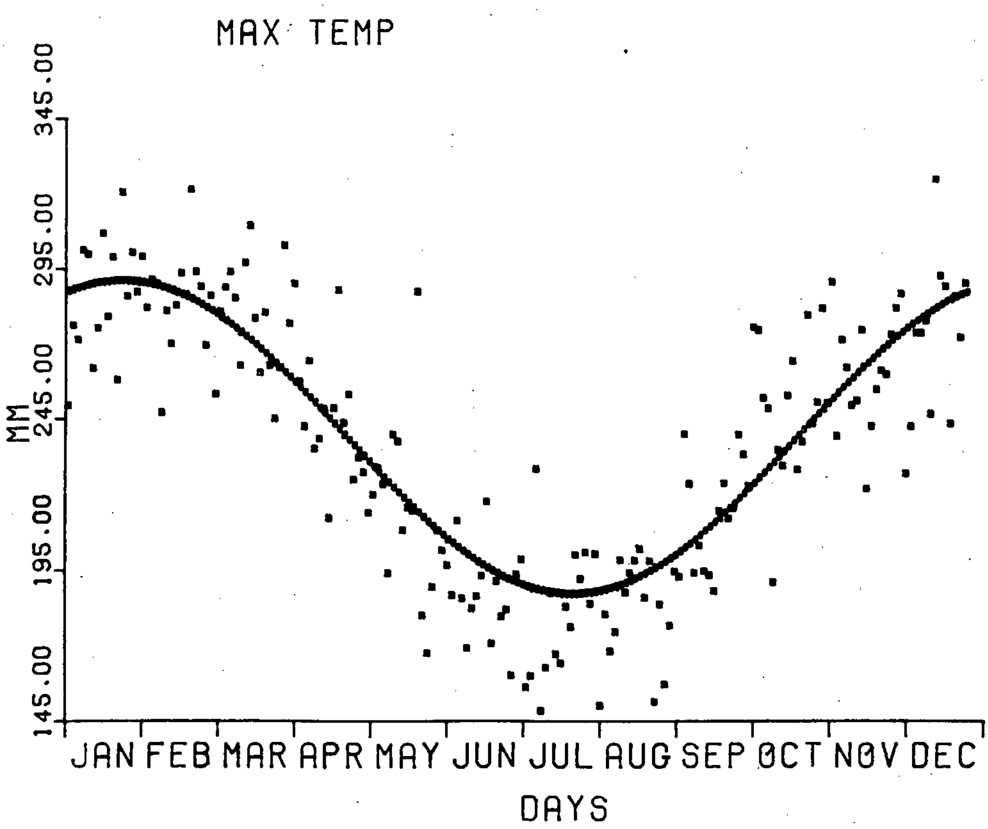
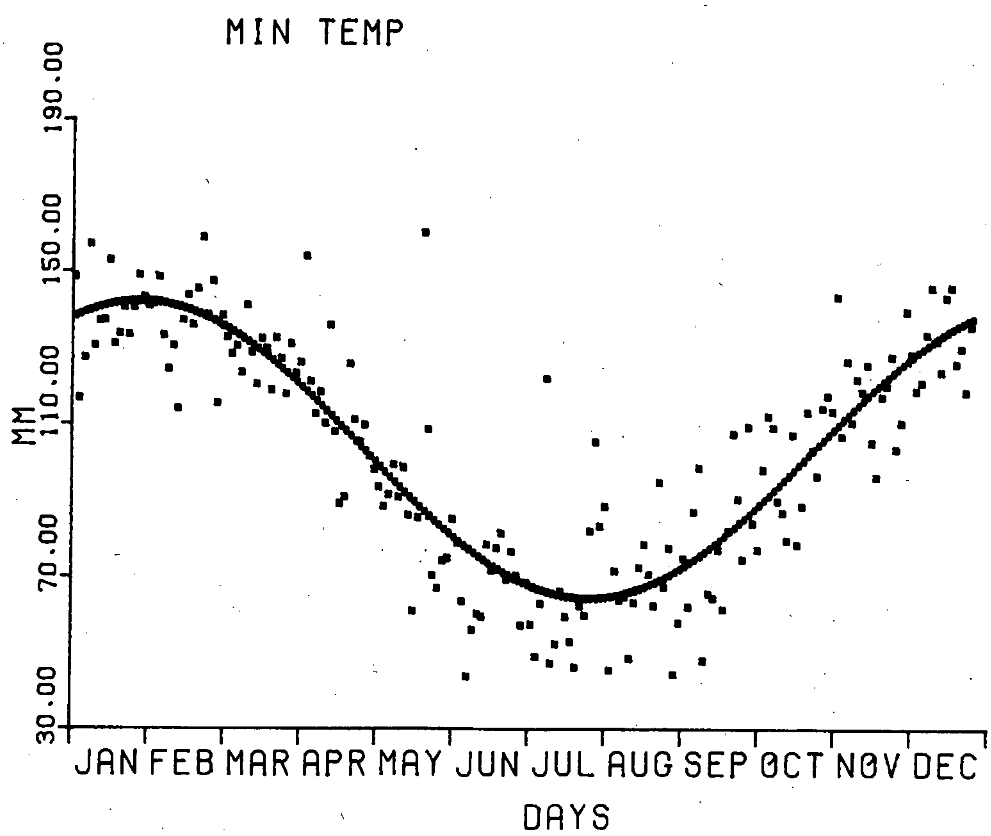
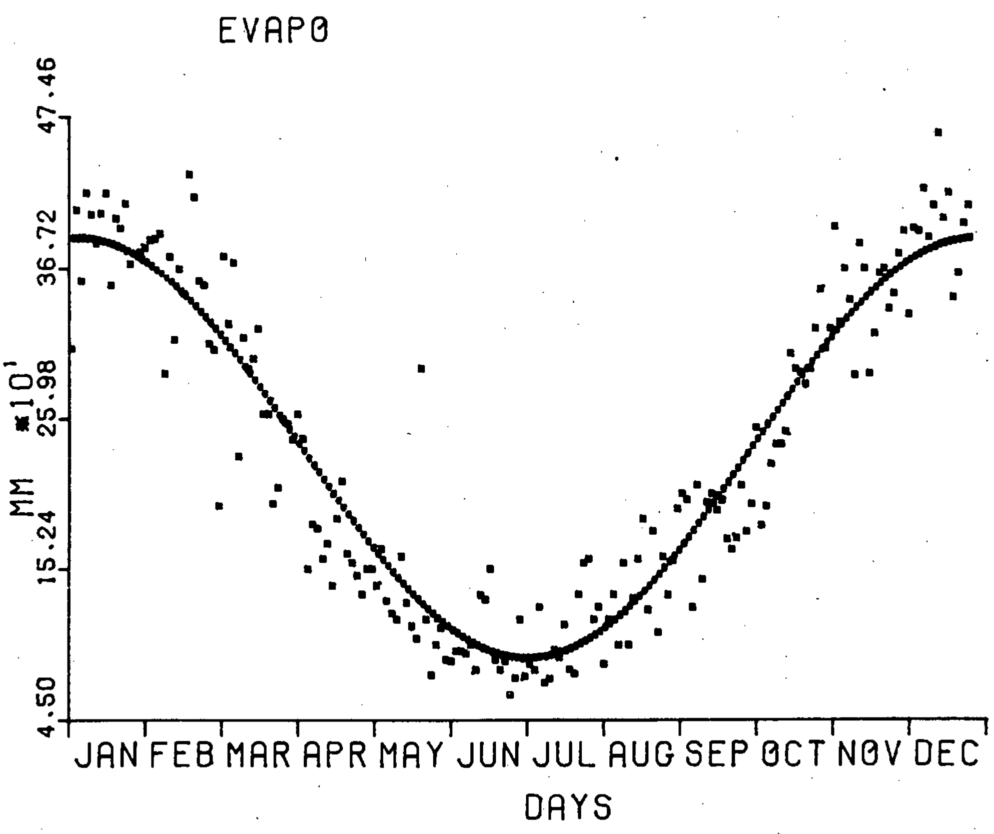
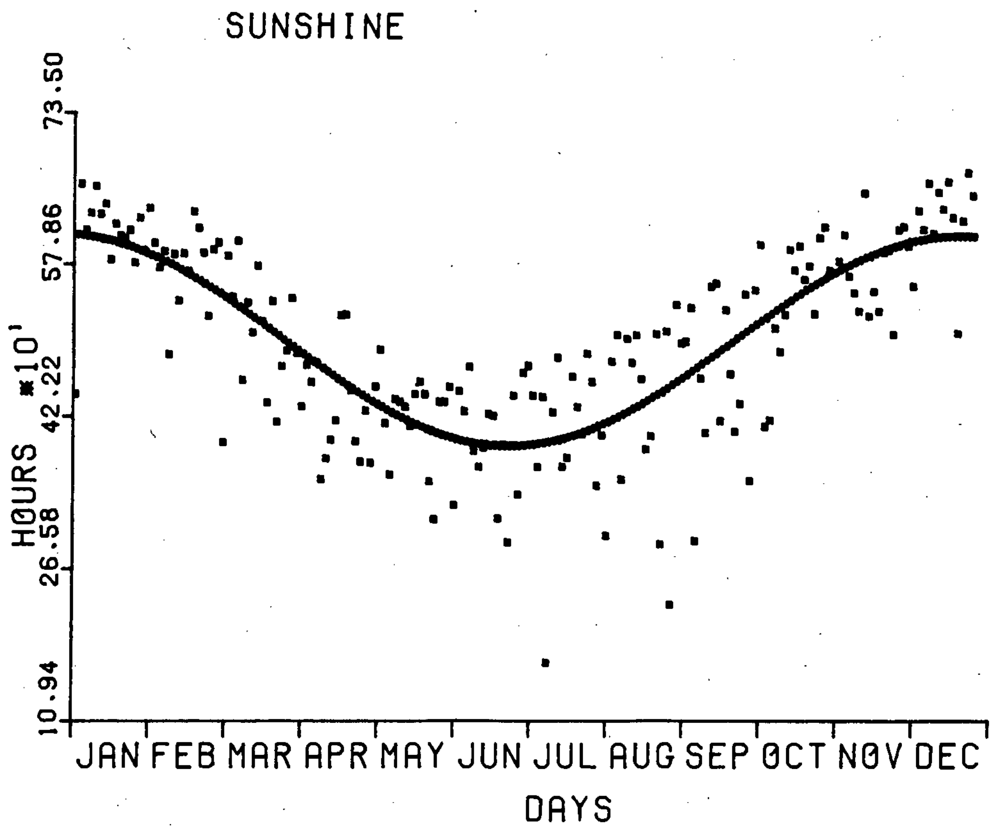


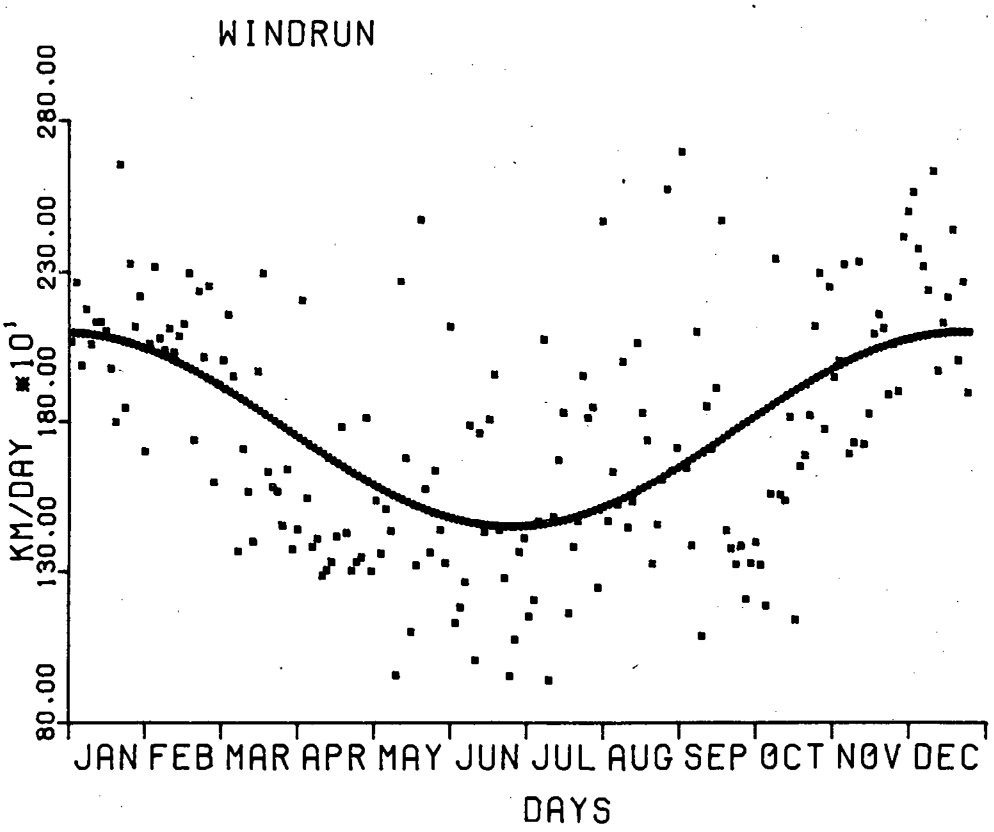
FIGURE 6.13.2: DAILY AVERAGES AND MEAN FITTED BY A FOURIER SERIES GIVEN A DRY DAY FOR ALL VARIABLES

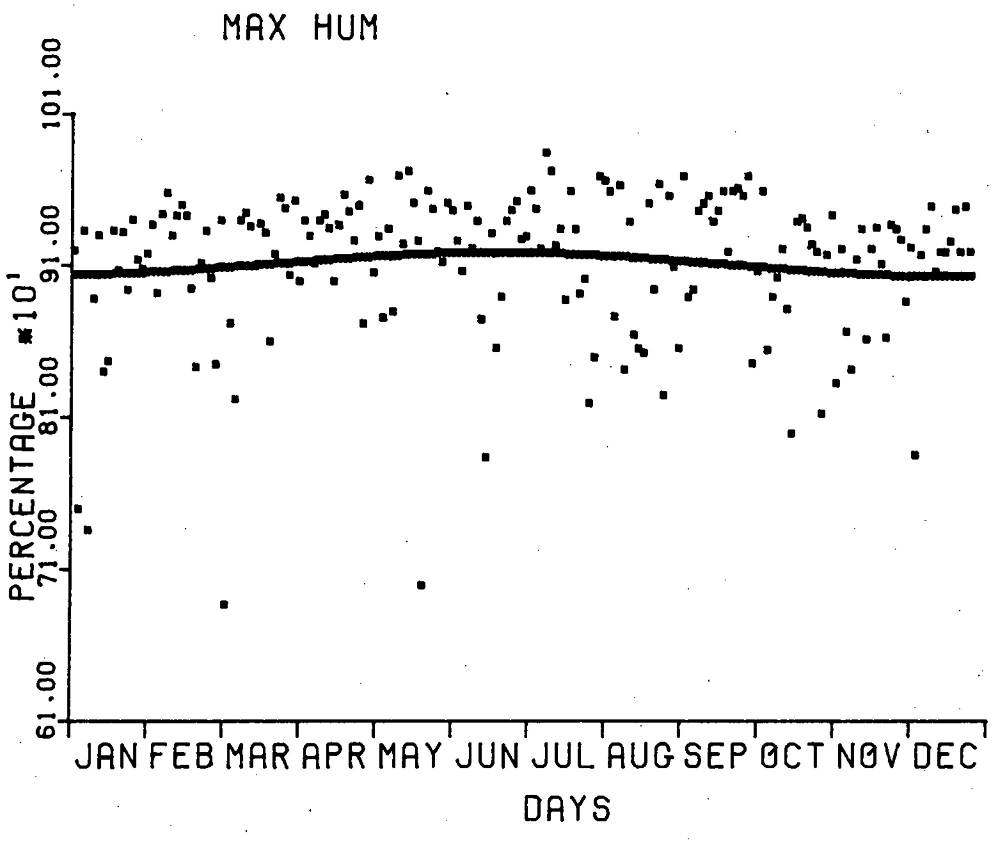



MIN TEMP
190.00
150.00
110.00
70.00
30.00
MM
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


EVAPO
47.46
36.72
25.98
15.24
4.50
MM *10¹
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


SUNSHINE
73.50
57.86
42.22
26.58
10.94
HOURS *10¹
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


WINDRUN
280.00
230.00
180.00
130.00
80.00
KM/DAY *10¹
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


MAX HUM
PERCENTAGE *10[1]
61.00
71.00
81.00
91.00
101.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS

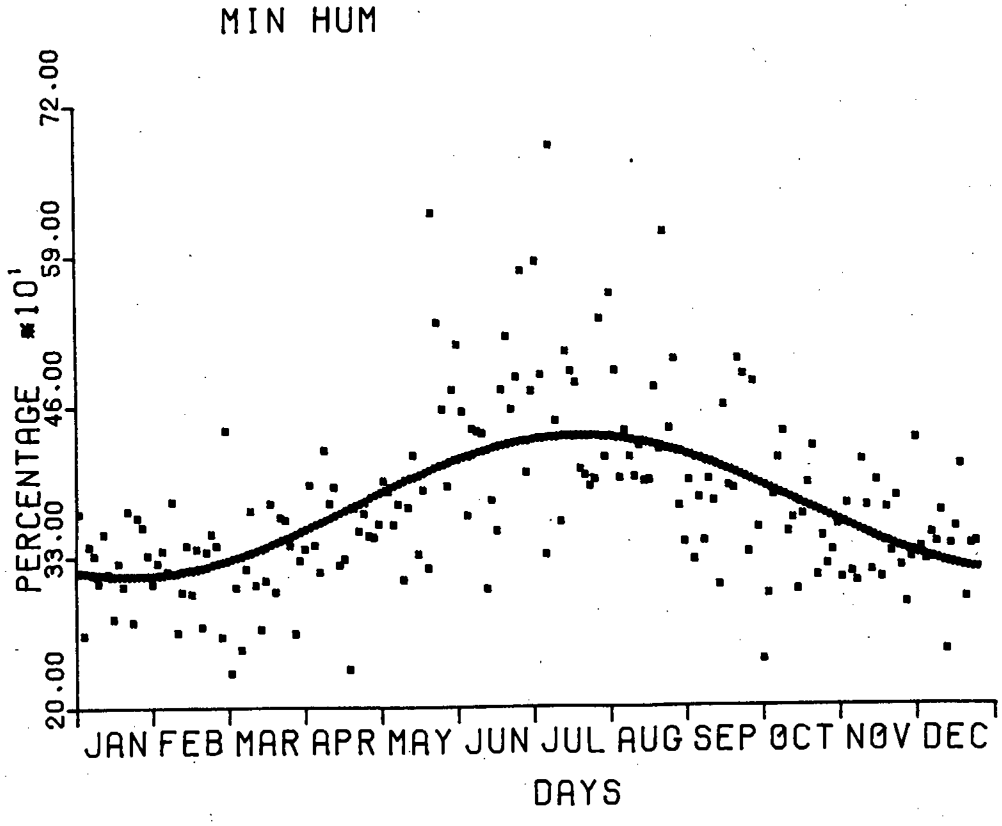
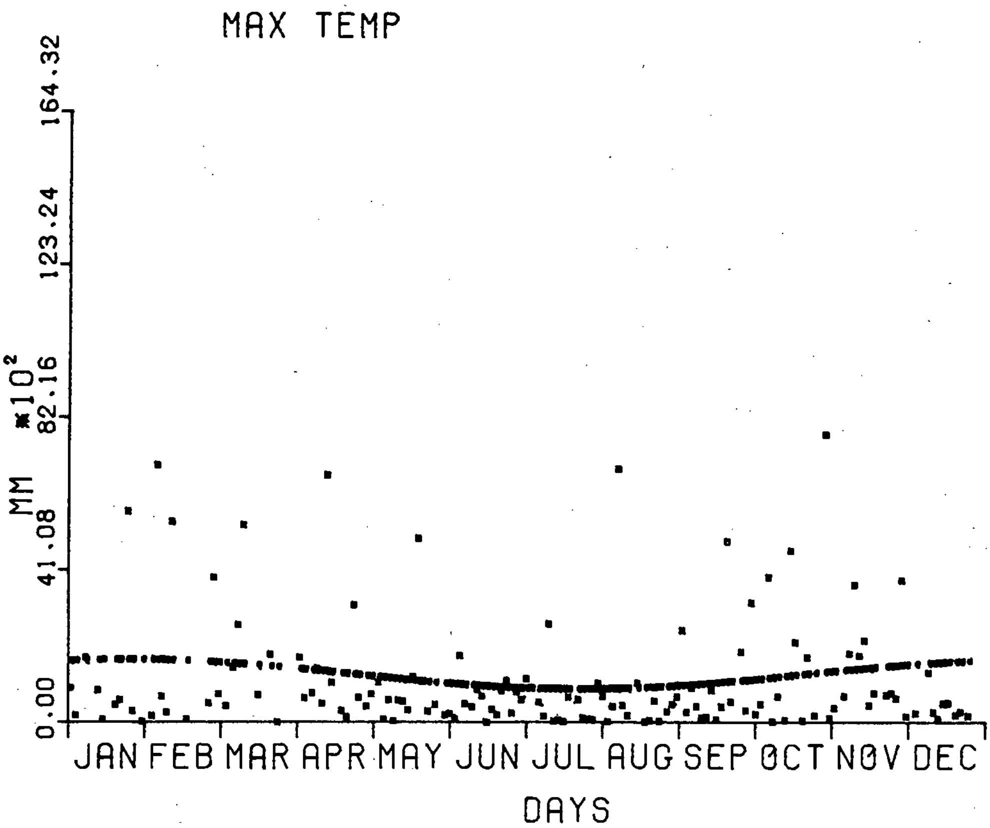
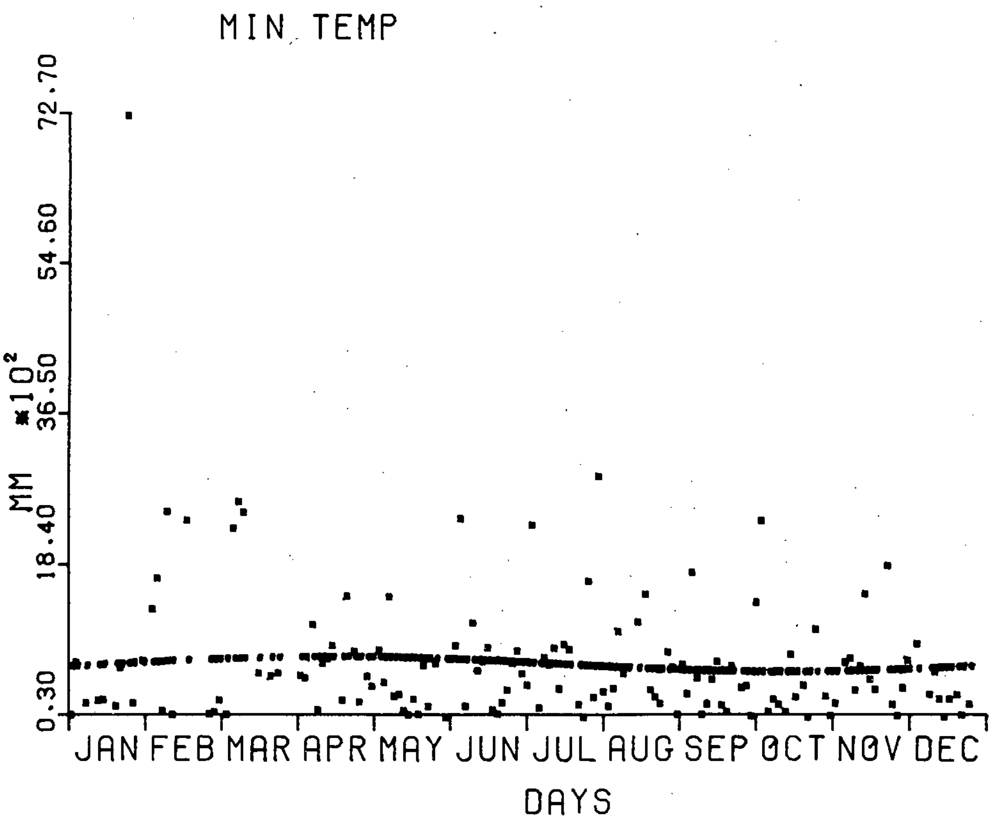
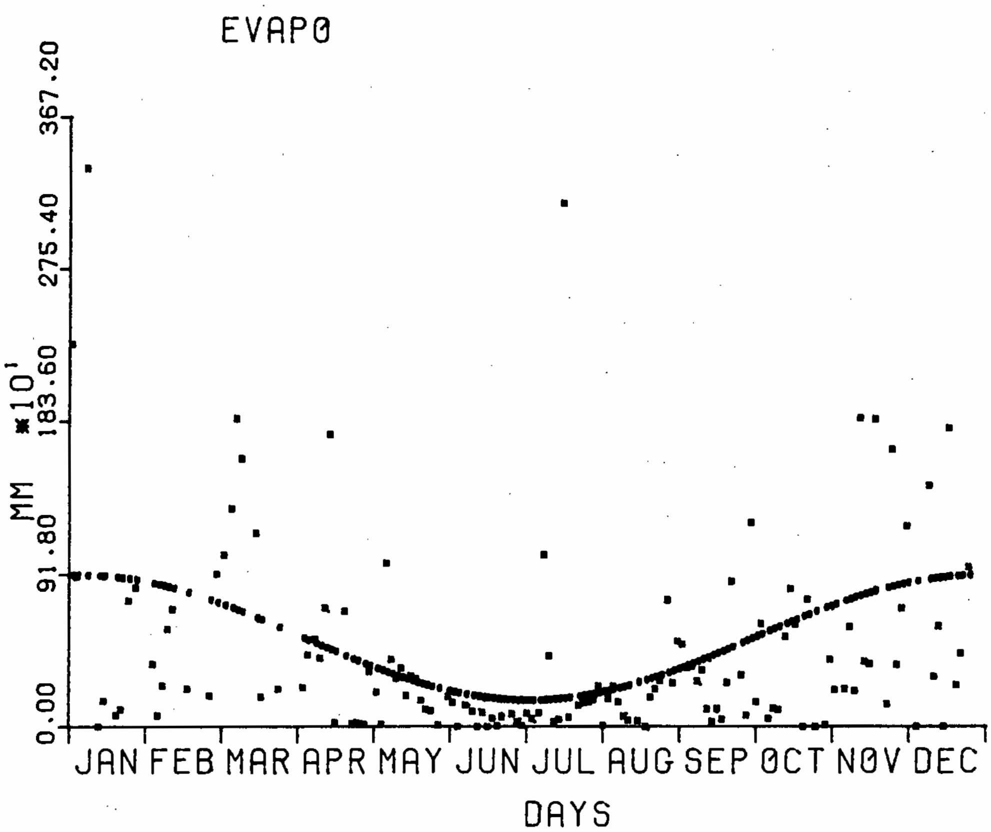
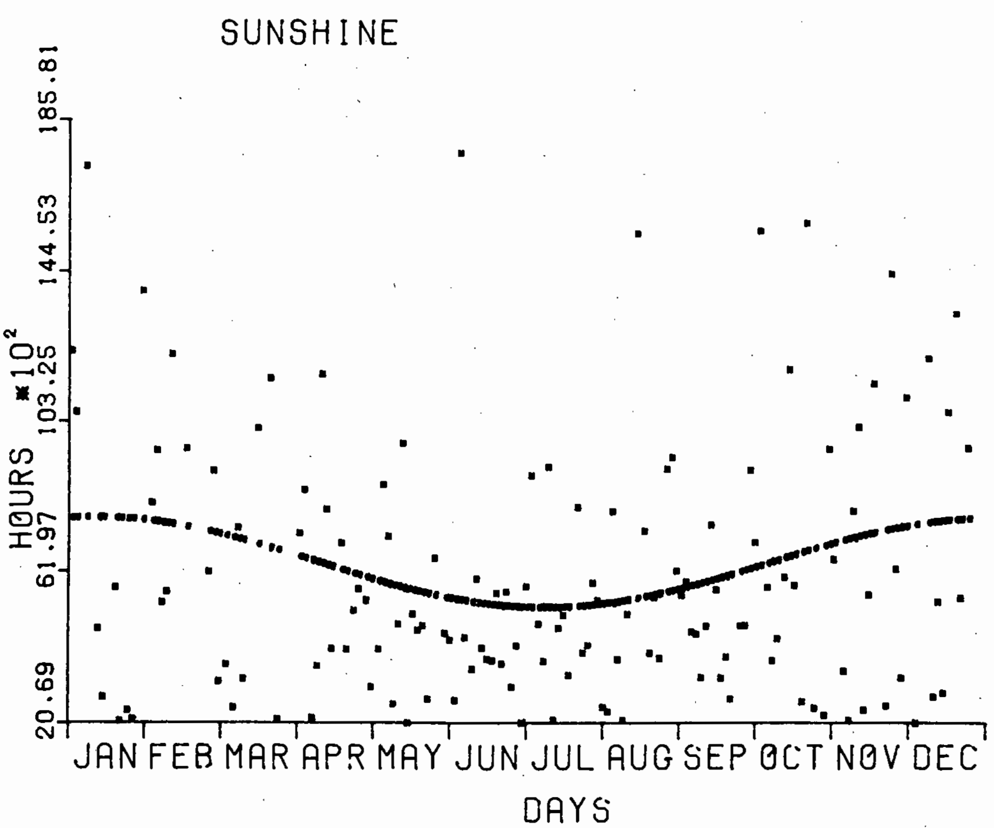

MIN HUM
PERCENTAGE *10[1]
20.00
33.00
46.00
59.00
72.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS

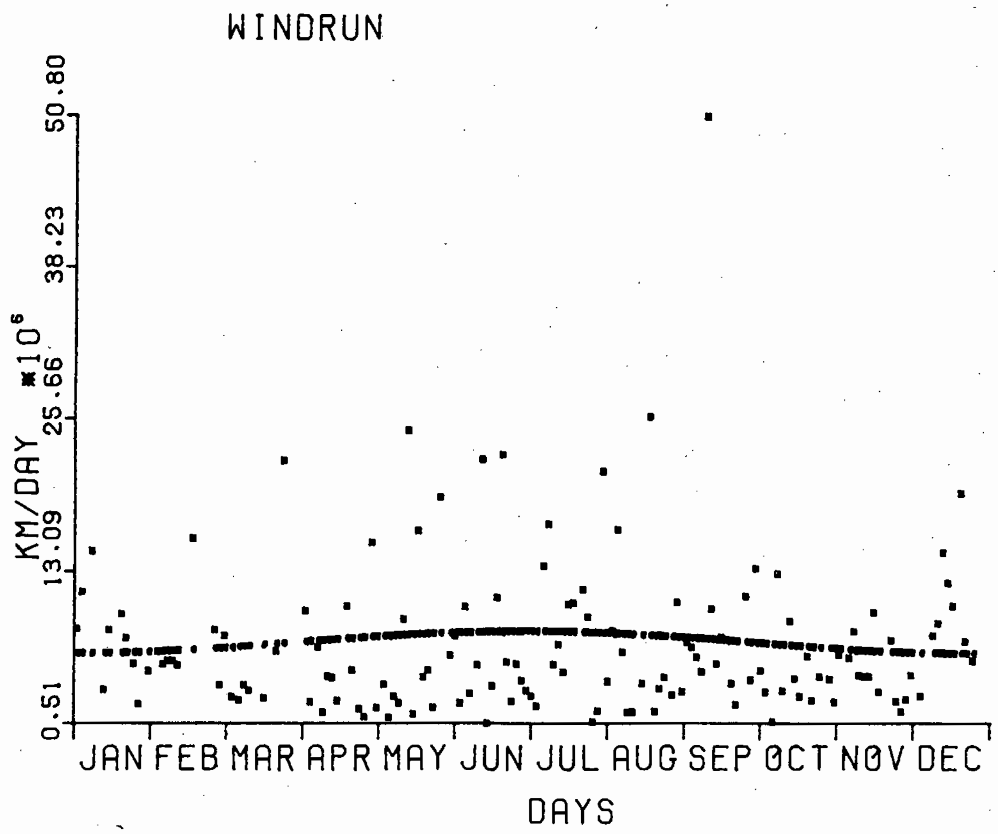
FIGURE 6.14.1: DAILY STANDARD DEVIATIONS AND STANDARD DEVIATION FITTED BY A FOURIER SERIES GIVEN A WET DAY FOR ALL VARIABLES

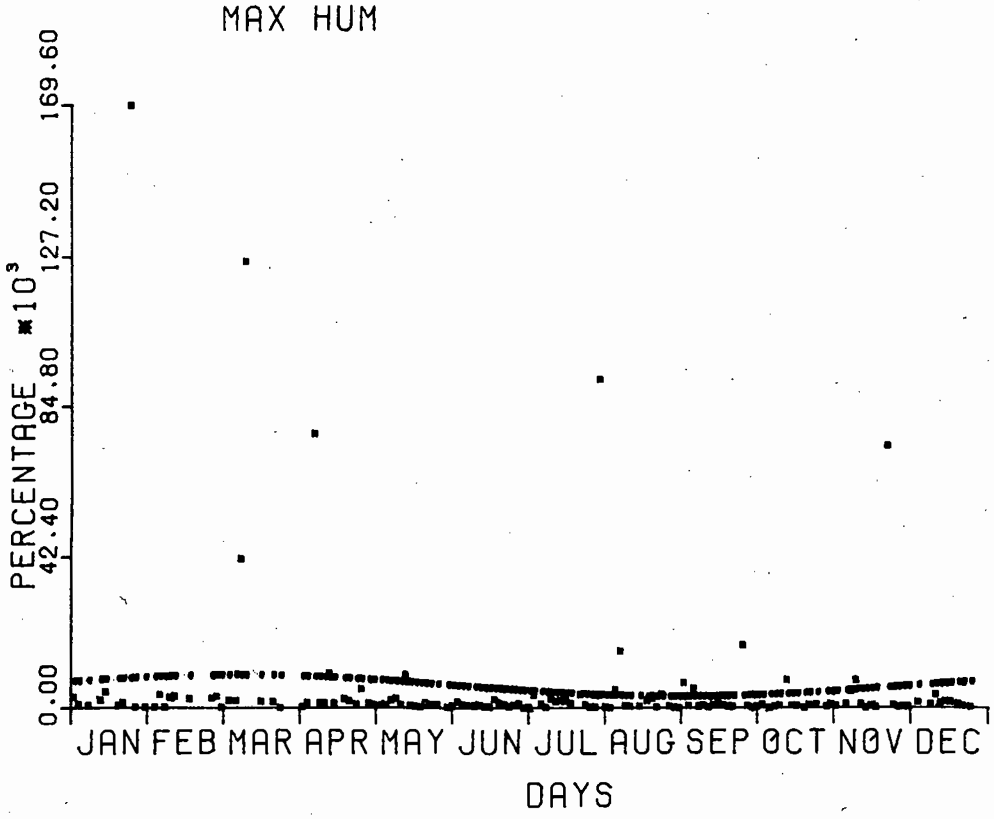





EVAPO
MM *10¹
367.20
275.40
183.60
91.80
0.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


SUNSHINE
HOURS *10²
185.81
144.53
103.25
61.97
20.69
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


WINDRUN
KM/DAY *10^6
50.80
38.23
25.66
13.09
0.51
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


MAX HUM
PERCENTAGE *10^3
169.60
127.20
84.80
42.40
0.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS

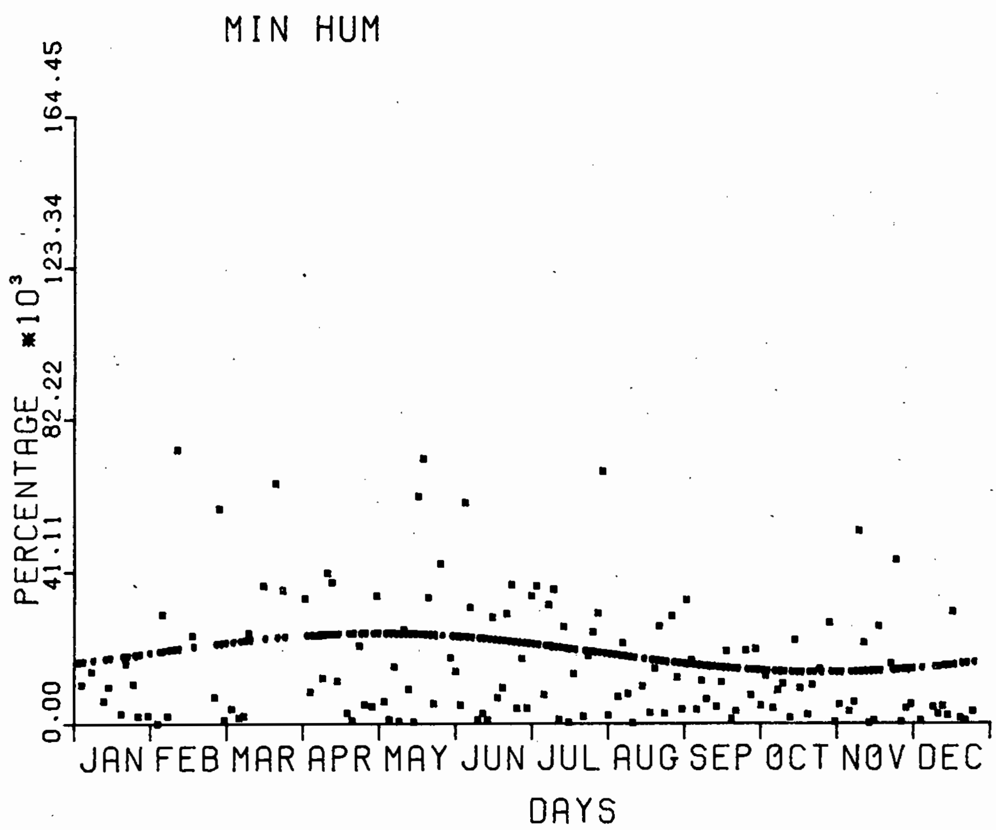
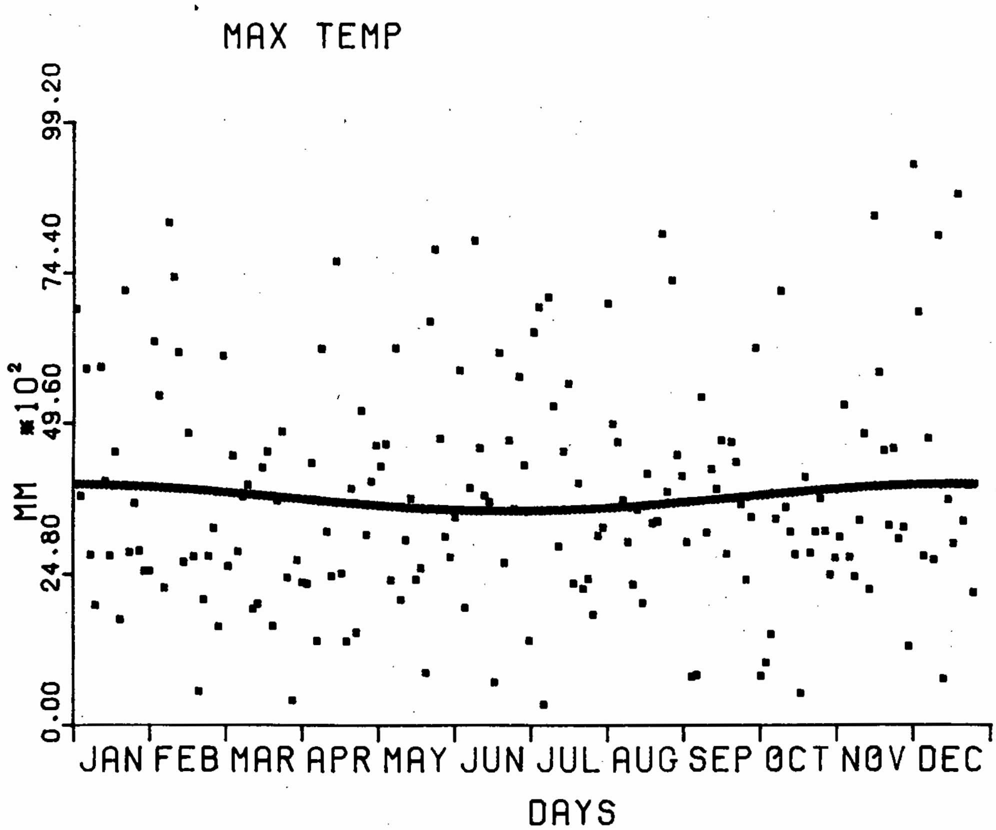
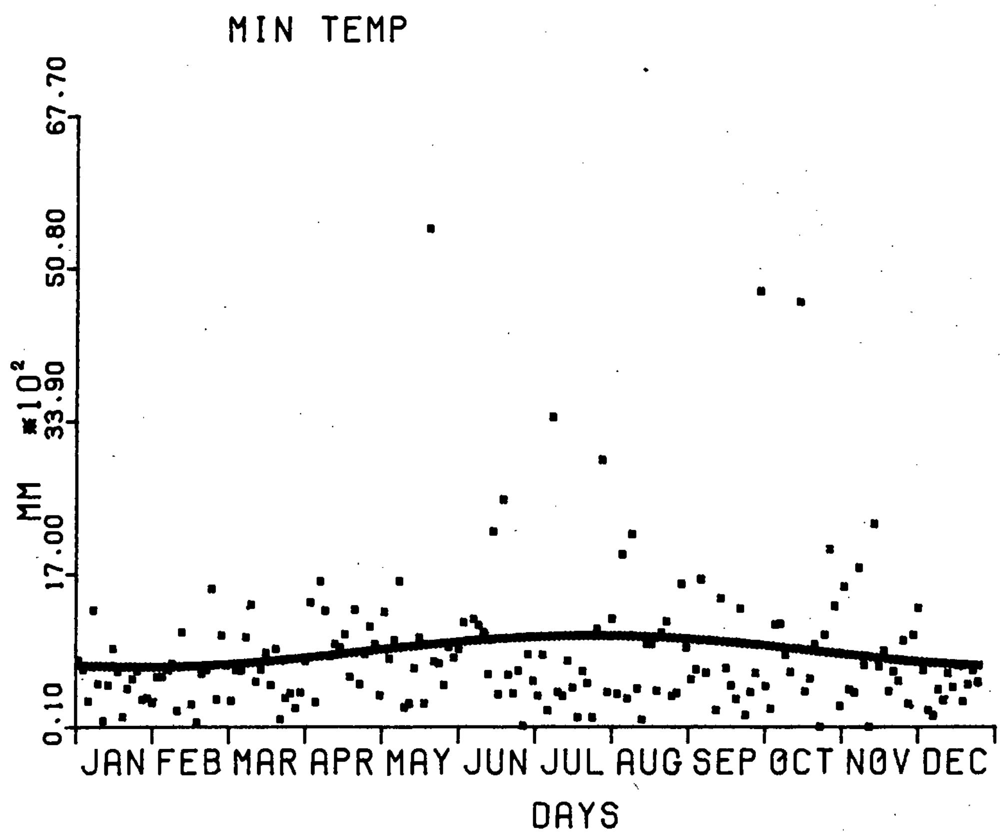
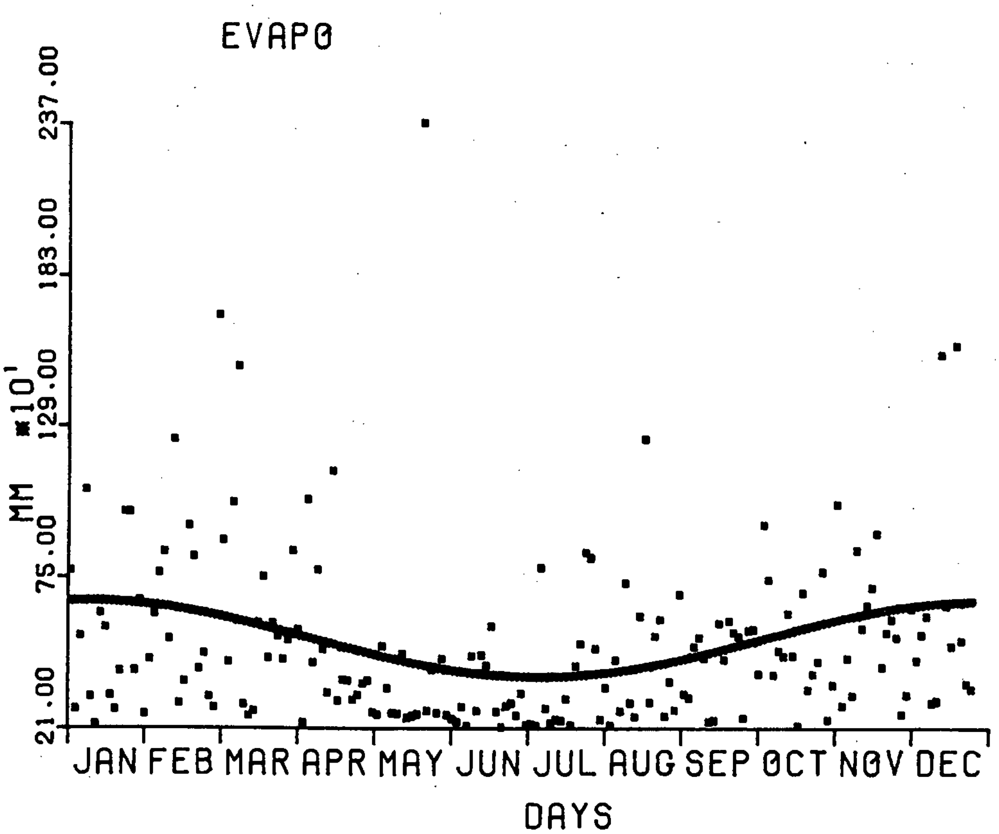
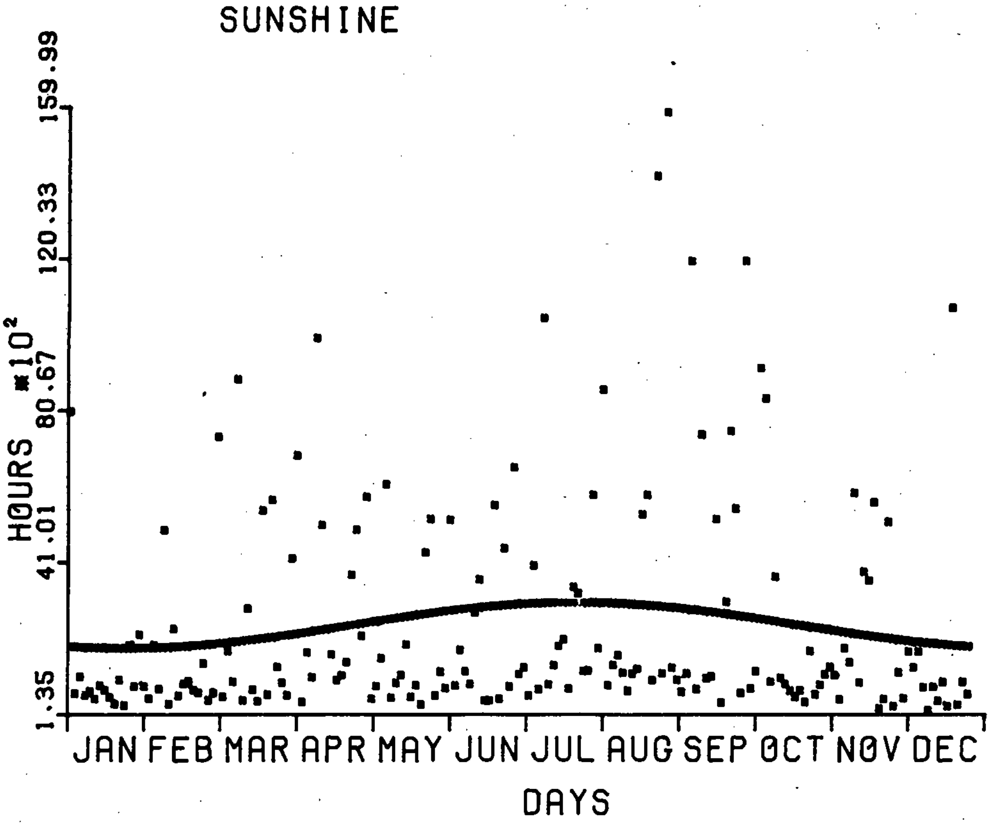


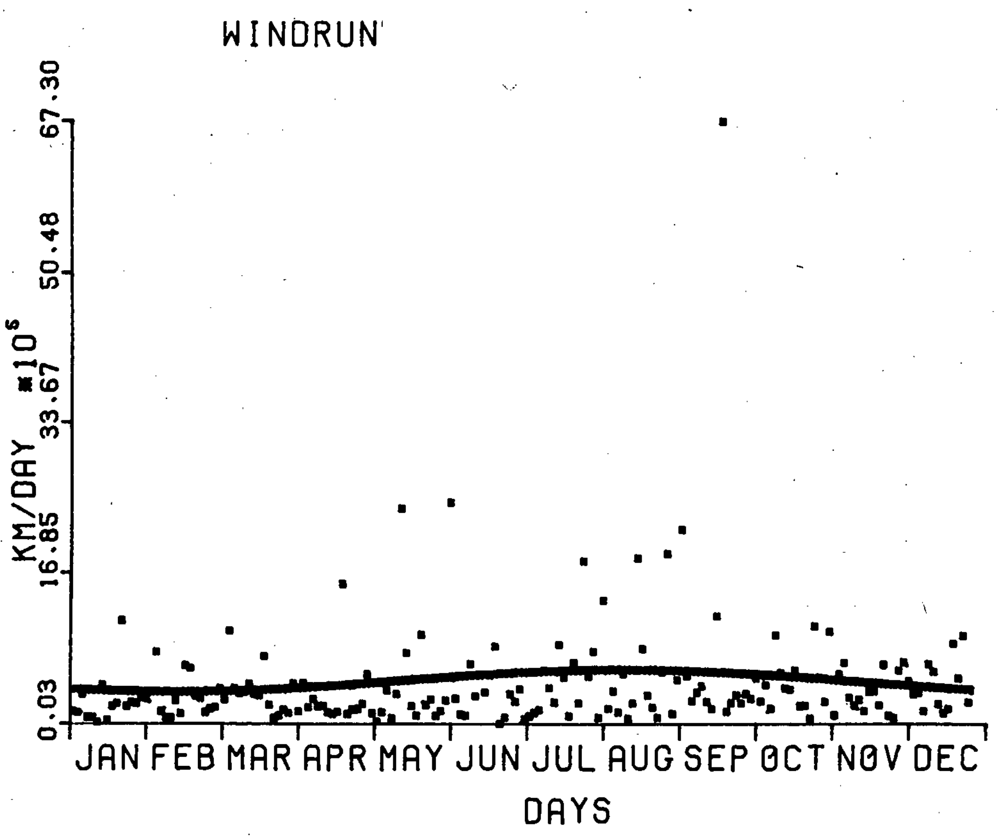
FIGURE 6.14.2: DAILY STANDARD DEVIATIONS AND STANDARD DEVIATIONS FITTED BY A FOURIER SERIES GIVEN A DRY DAY FOR ALL VARIABLES

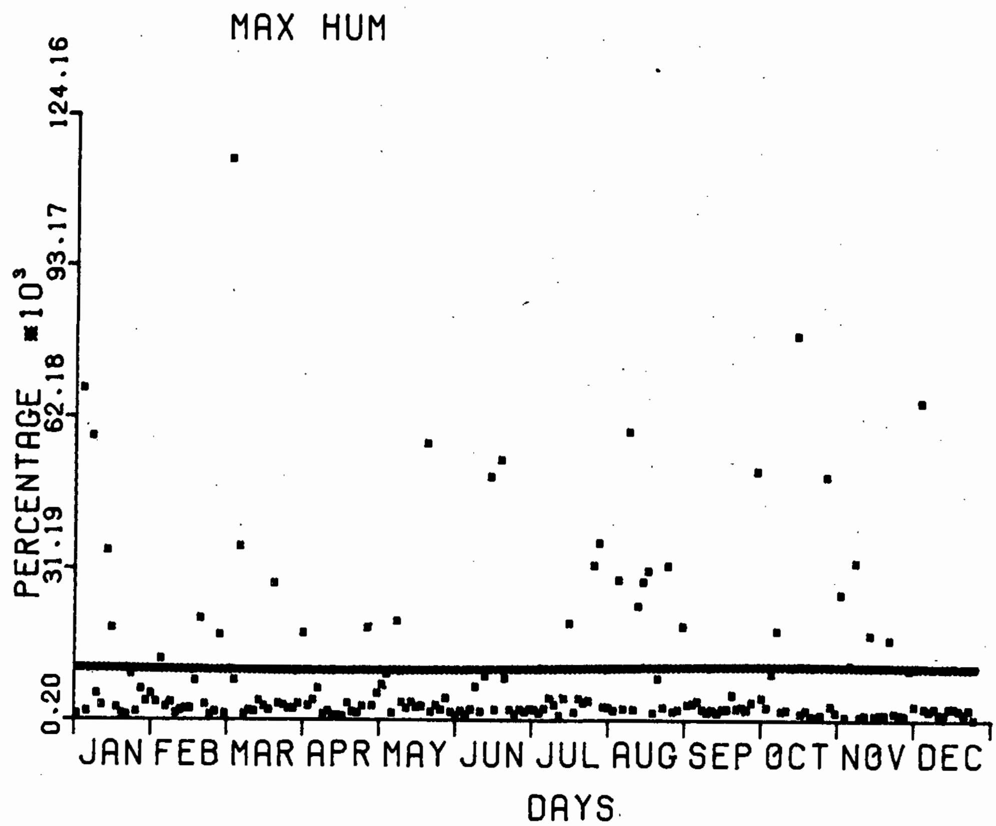


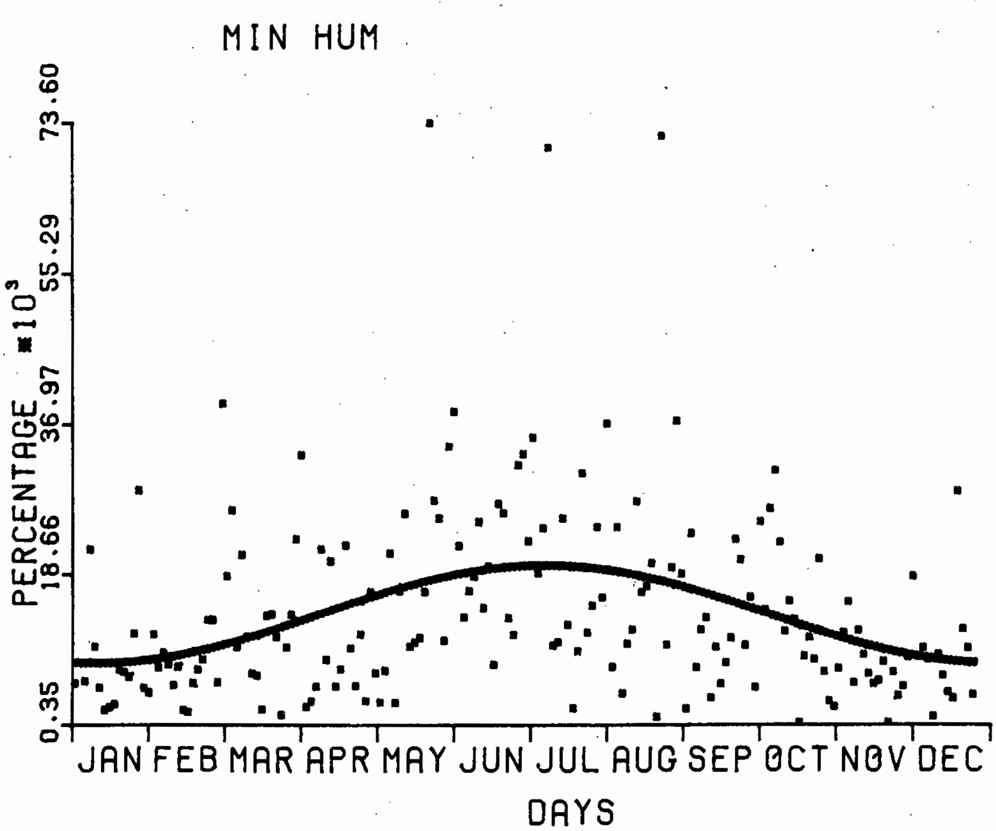

MIN TEMP
MM *10²
67.70
50.80
33.90
17.00
0.10
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


EVAPO
MM *10¹
237.00
183.00
129.00
75.00
21.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


SUNSHINE
HOURS *10²
159.99
120.33
80.67
41.01
1.35
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


WINDRUN
KM/DAY *10⁵
67.30
50.48
33.67
16.85
0.03
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


MAX HUM
PERCENTAGE *10³
124.16
93.17
62.18
31.19
0.20
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


MIN HUM
PERCENTAGE *10³
73.60
55.29
36.97
18.66
0.35
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS

the standard deviations for some variables, while for others, because of the high scatter of the standard deviations, it overestimates the standard deviation, mainly for variables maximum humidity, maximum temperature (given a wet day) and sunshine duration (given a dry day).

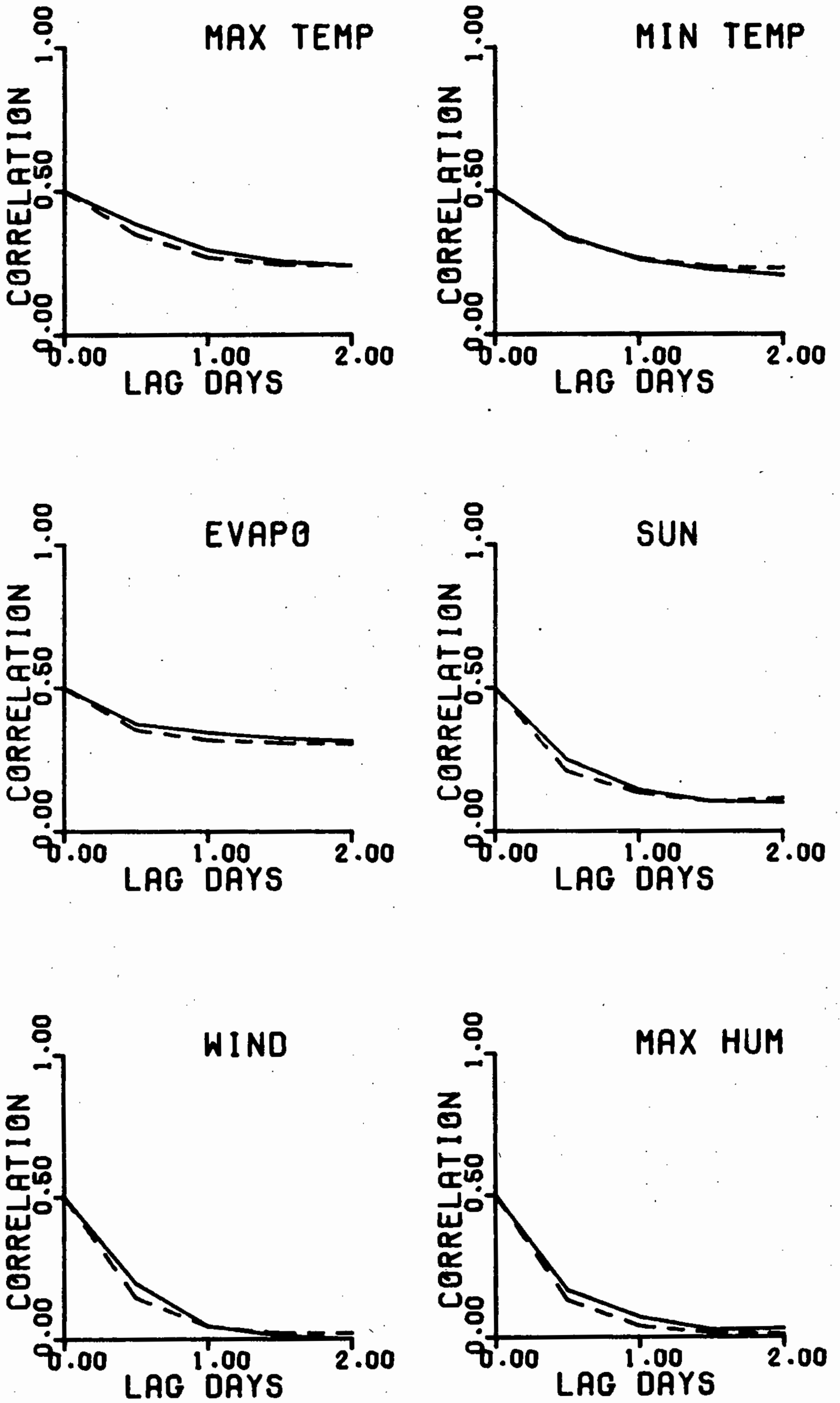
The autocorrelation within each variable in the simulated data was compared to that of the historical record and, as Figure 6.15 shows, the autocorrelation has been retained by the model exceptionally well.

When the climate sequences are conditioned on the wet or dry status of the day, the model still performs very well with respect to the autocorrelation structure of the variables, again the model describes the autocorrelation better given a dry sequence than on a wet sequence. These results are shown in Figures 6.16 - 6.17. Again this is probably because the sample of dry days is much larger than that of wet days.

The cross-correlation coefficients for lags -1, 0, 1 were used in the simulation technique. Figure 6.18 shows that the cross-correlation between most variables has been excellently preserved by the model, with the lag 0 cross-correlation being preserved in all variables very well. For some variables the lag 1 and lage -1 cross-correlation of the generated sequence do differ from those of the historical sequence, but it is apparent that on the whole, the model has retained the cross-correlation between variables.

The cross-correlation coefficients obtained from the historical and from the simulated sequences when the variables are conditioned on the wet or dry status of the day are compared and these comparisons are shown in Figures 6.19 - 6.20.

FIGURE 6.15: HISTORICAL (——) AND SIMULATED (---) AUTOCORRELATION COEFFICIENTS FOR ALL VARIABLES



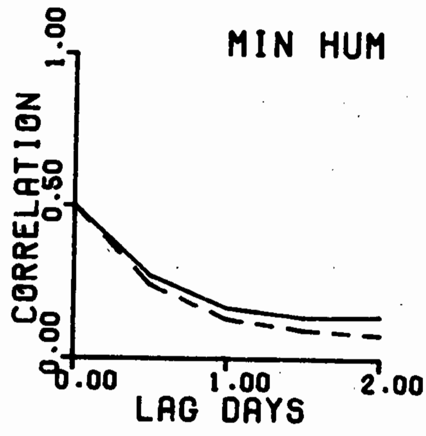
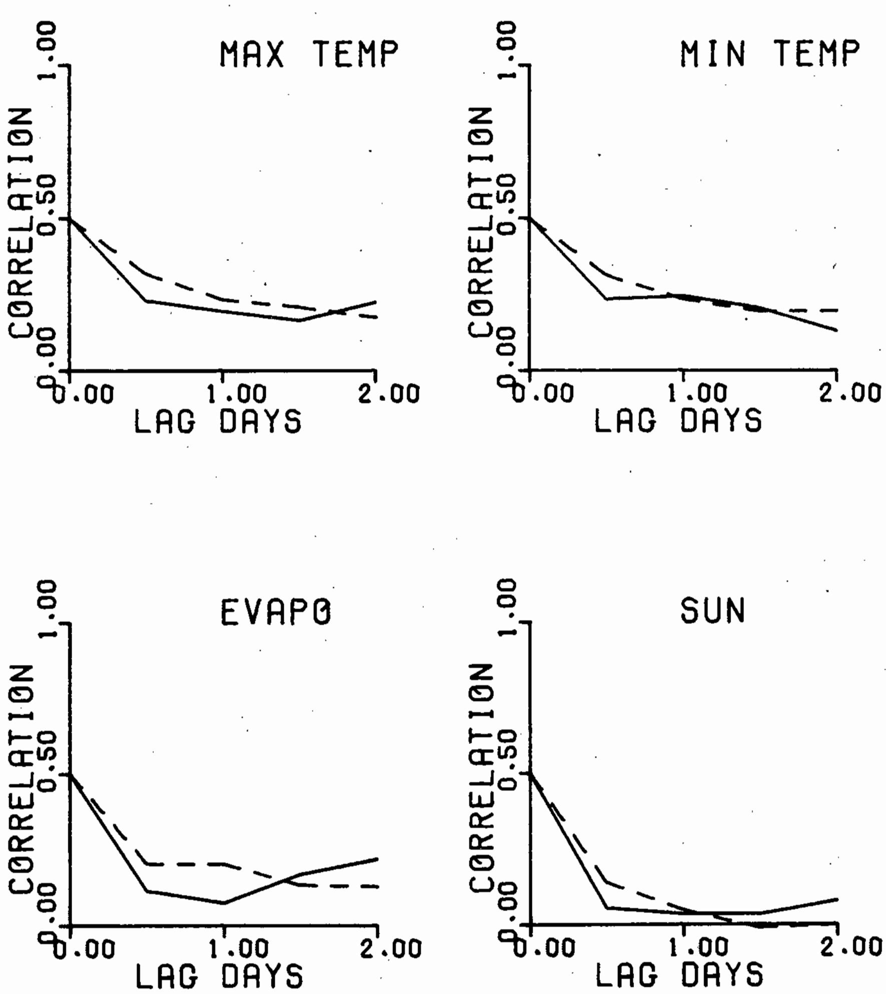


FIGURE 6.16: HISTORICAL (——) AND SIMULATED (---) AUTOCORRELATION COEFFICIENTS GIVEN A WET DAY FOR ALL VARIABLES



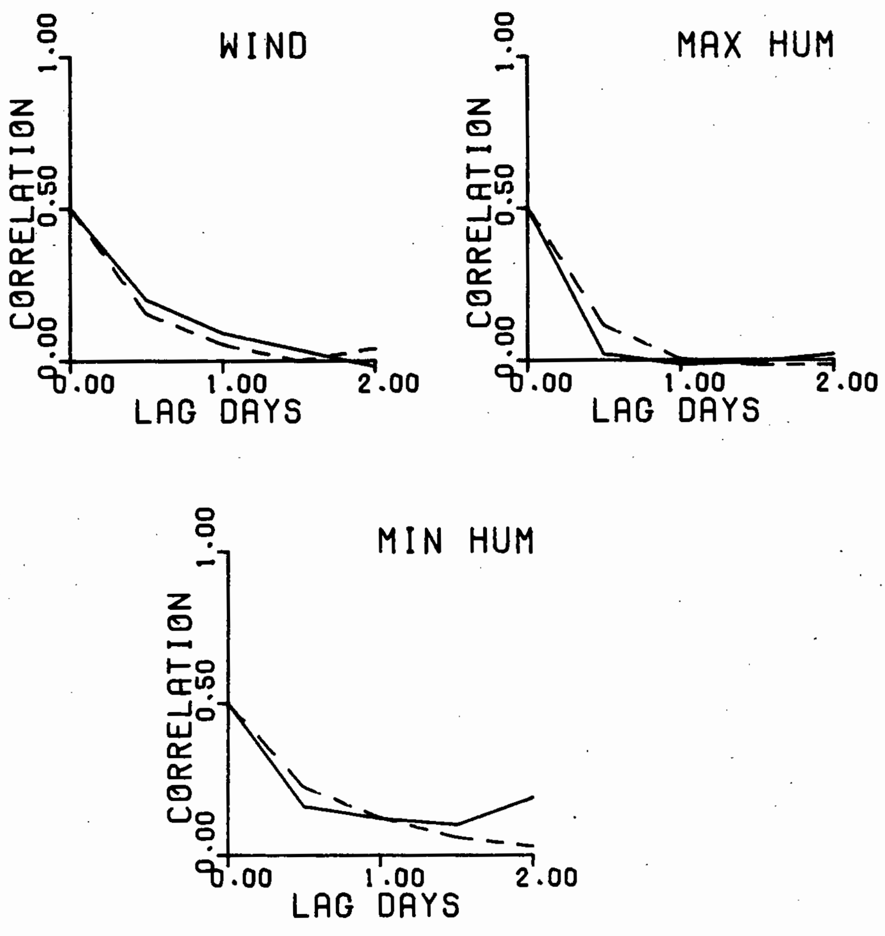
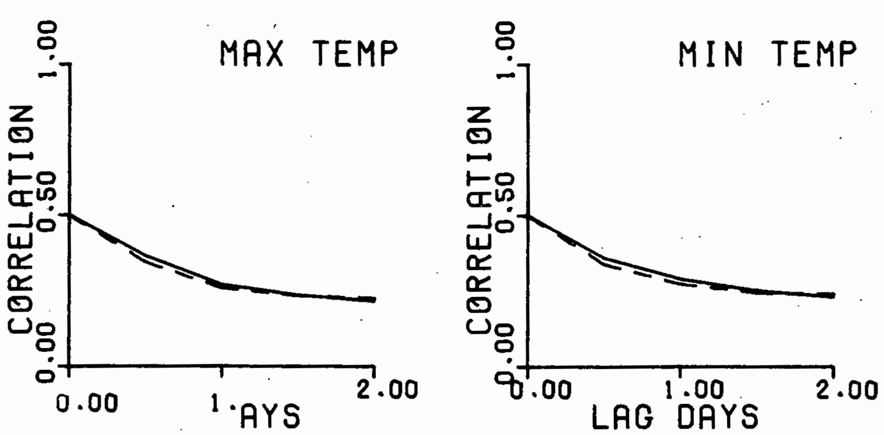


FIGURE 6.17: HISTORICAL (——) AND SIMULATED (---) AUTOCORRELATION COEFFICIENTS GIVEN A DRY DAY FOR ALL VARIABLES



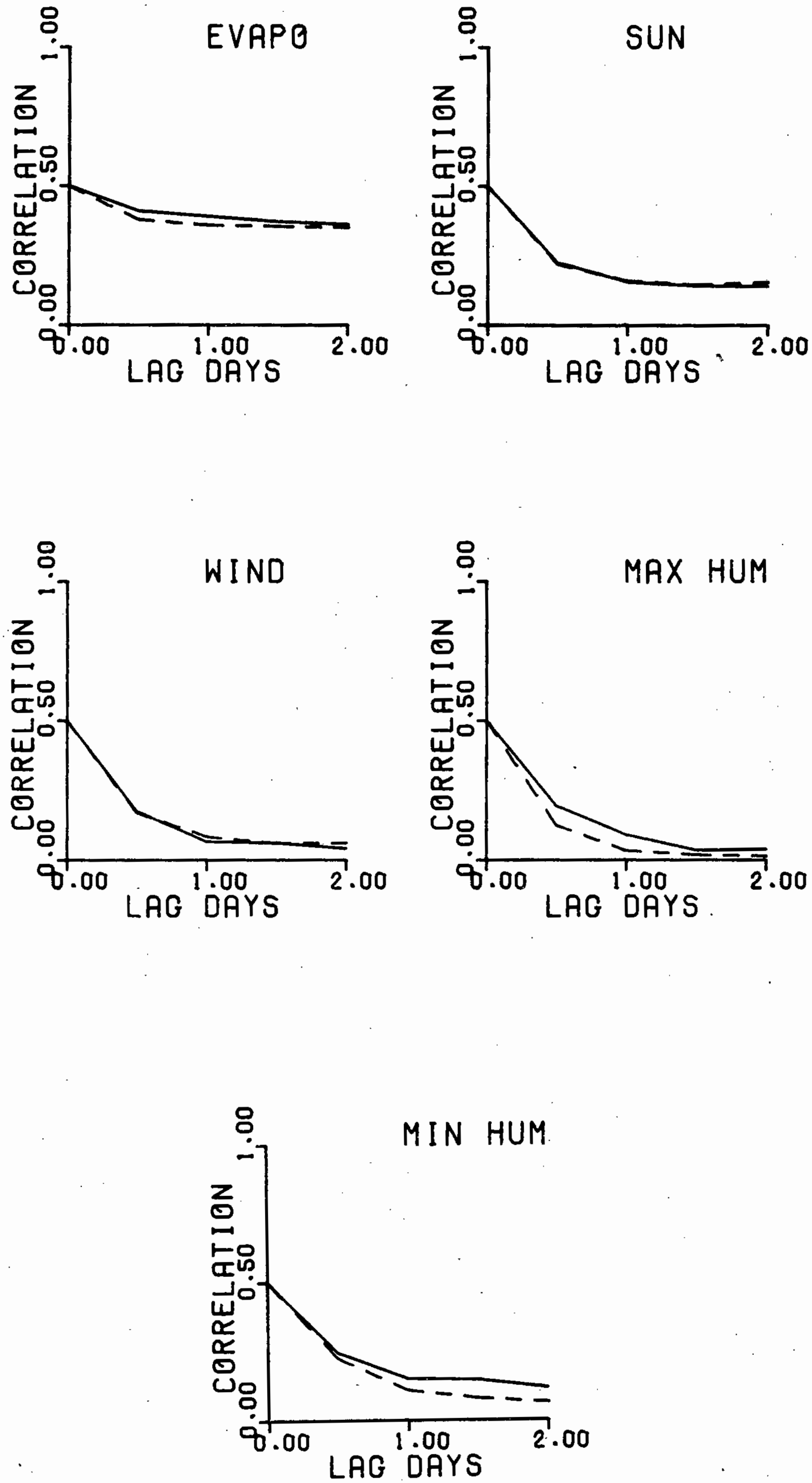
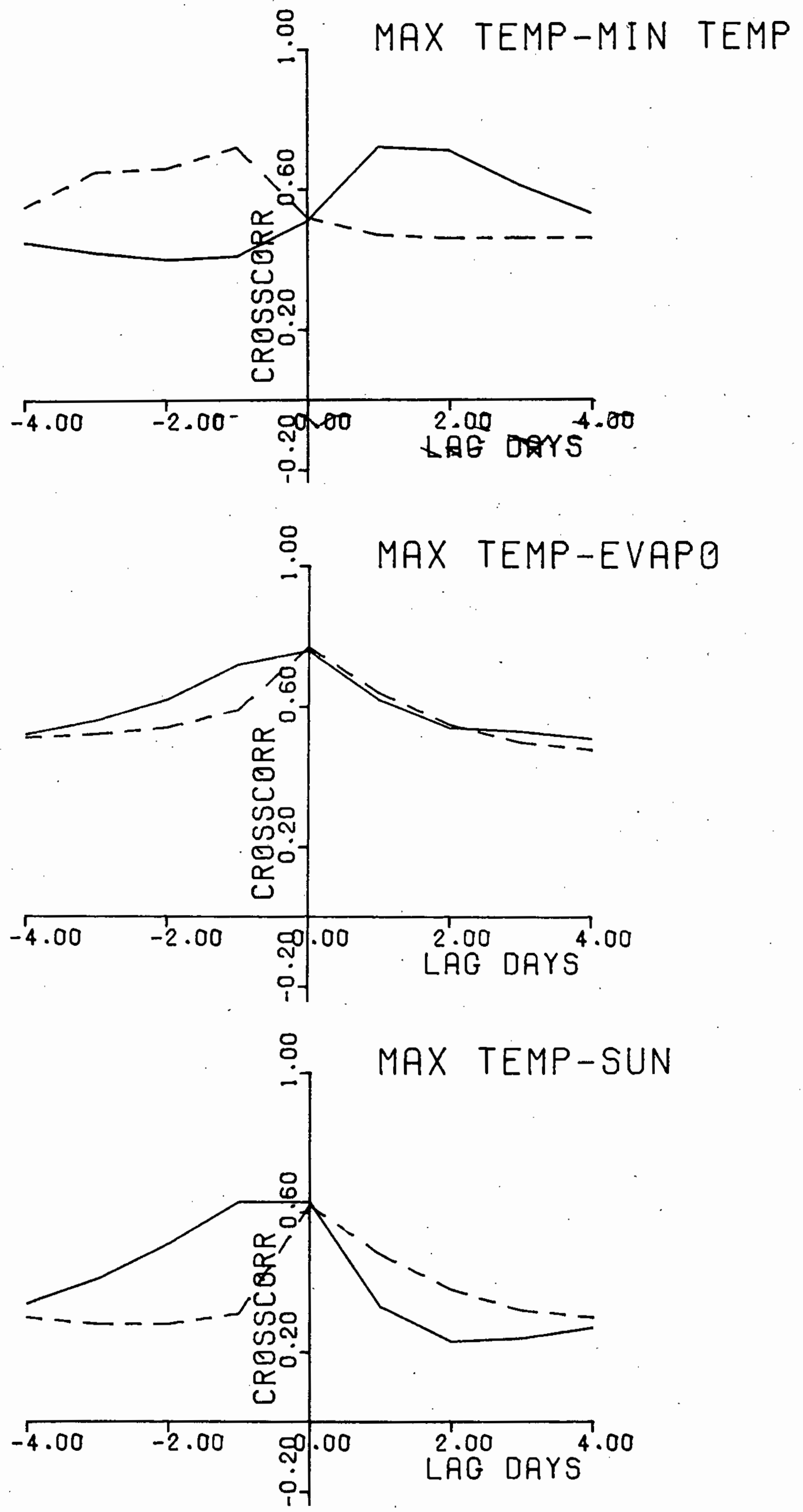
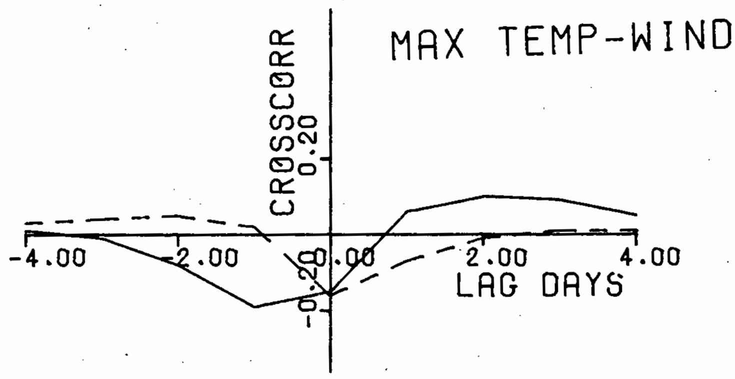
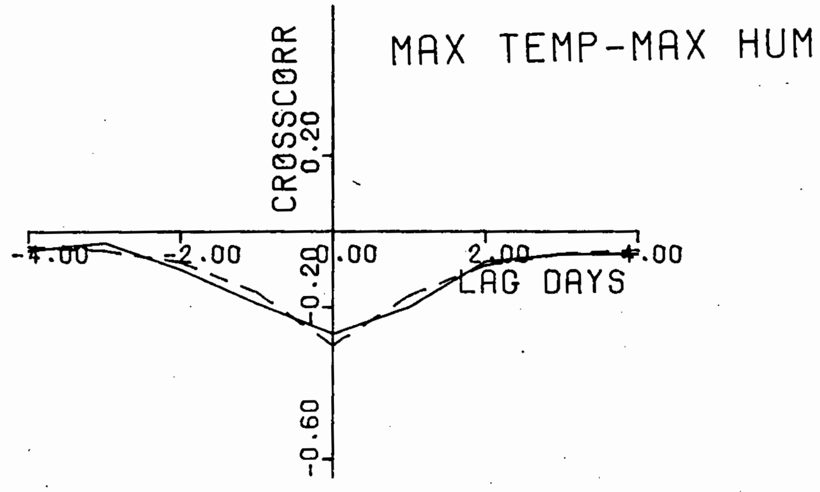
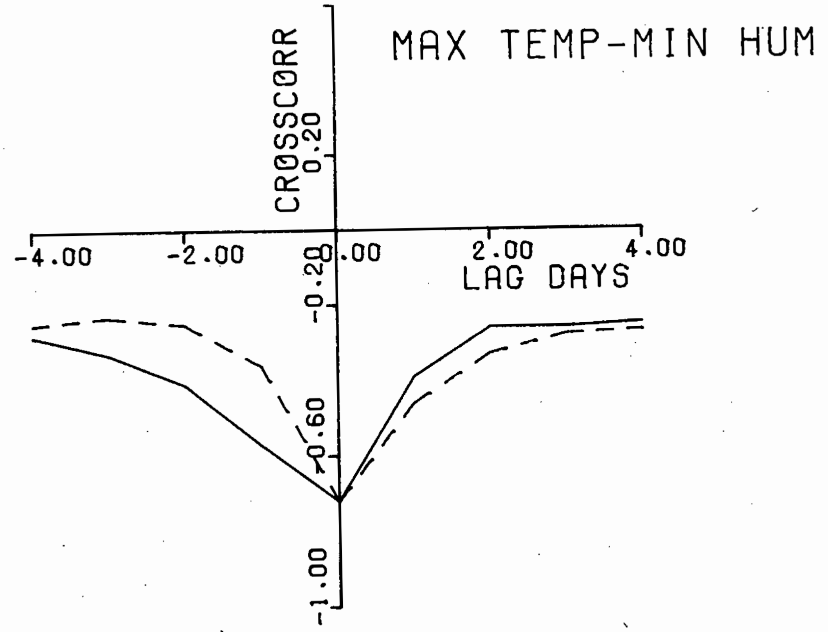

EVAPO
CORRELATION
0.00
0.50
1.00
0.00
1.00
2.00
LAG DAYS
SUN
CORRELATION
0.00
0.50
1.00
0.00
1.00
2.00
LAG DAYS
WIND
CORRELATION
0.00
0.50
1.00
0.00
1.00
2.00
LAG DAYS
MAX HUM
CORRELATION
0.00
0.50
1.00
0.00
1.00
2.00
LAG DAYS
MIN HUM
CORRELATION
0.00
0.50
1.00
0.00
1.00
2.00
LAG DAYS

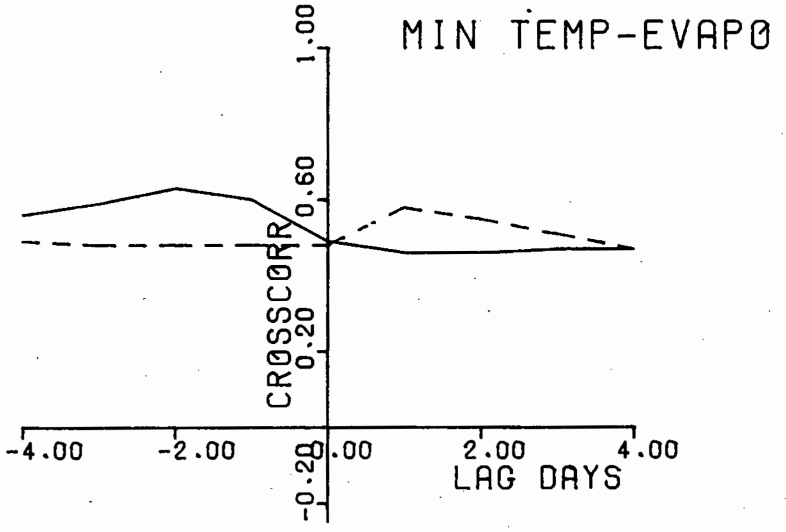
FIGURE 6.18: HISTORICAL (——) AND SIMULATED (---) CROSS-CORRELATION COEFFICIENTS

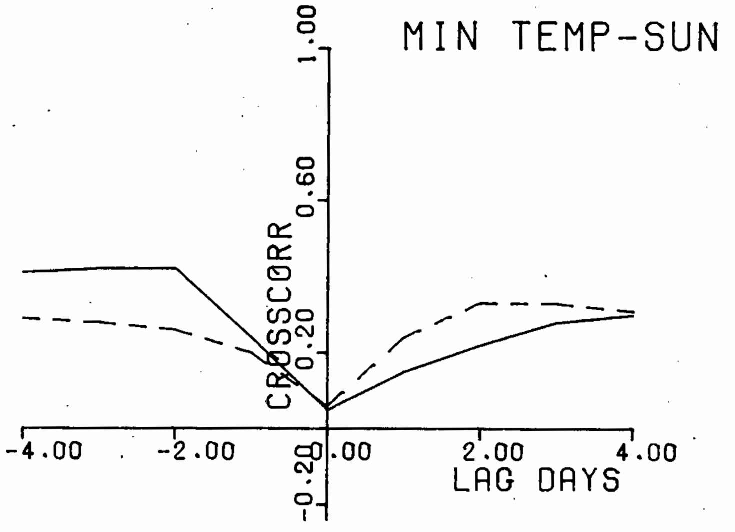


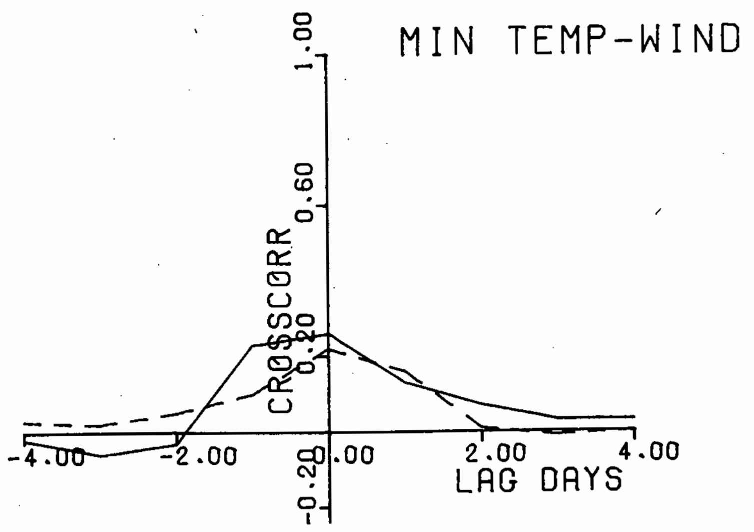

MAX TEMP-WIND
CRØSSCØRR
0.20
-0.20
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

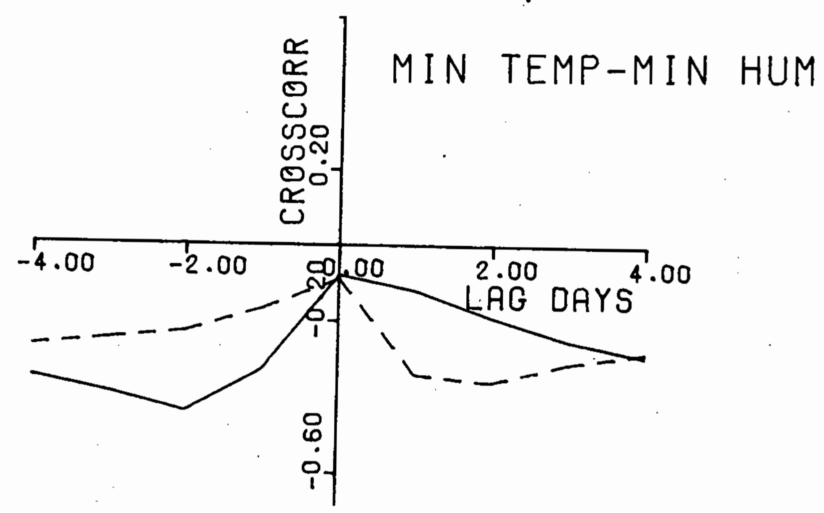

MAX TEMP-MAX HUM
CRØSSCØRR
0.20
-0.20
-0.60
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

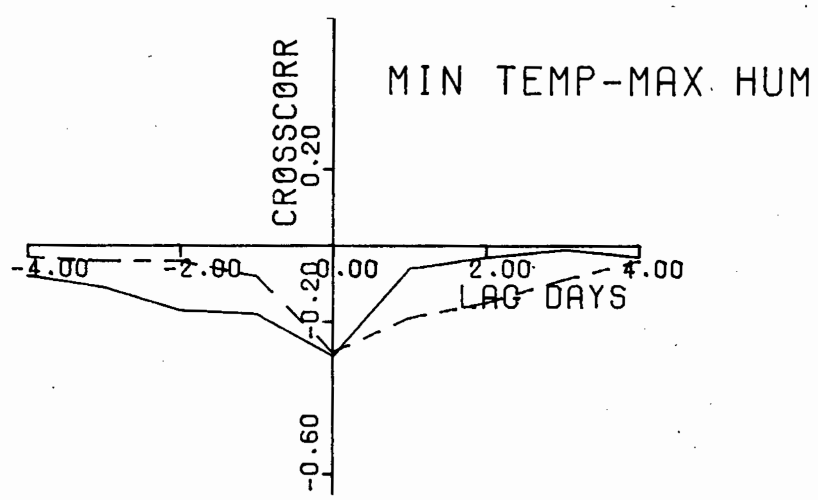

MAX TEMP-MIN HUM
CRØSSCØRR
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

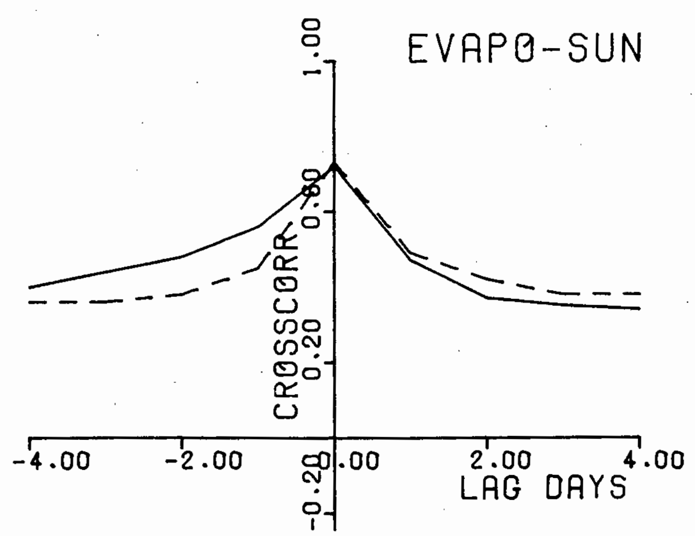

MIN TEMP-EVAPO
CROSSCORR
1.00
0.60
0.20
-0.20
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

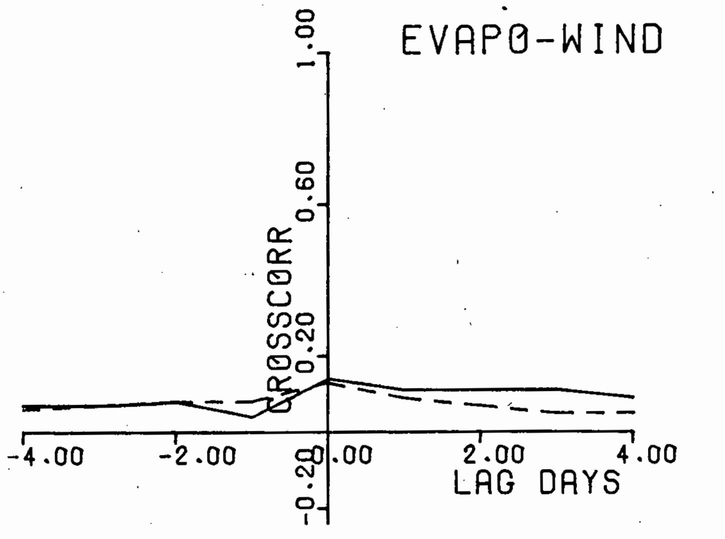

MIN TEMP-SUN
CROSSCORR
1.00
0.60
0.20
-0.20
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

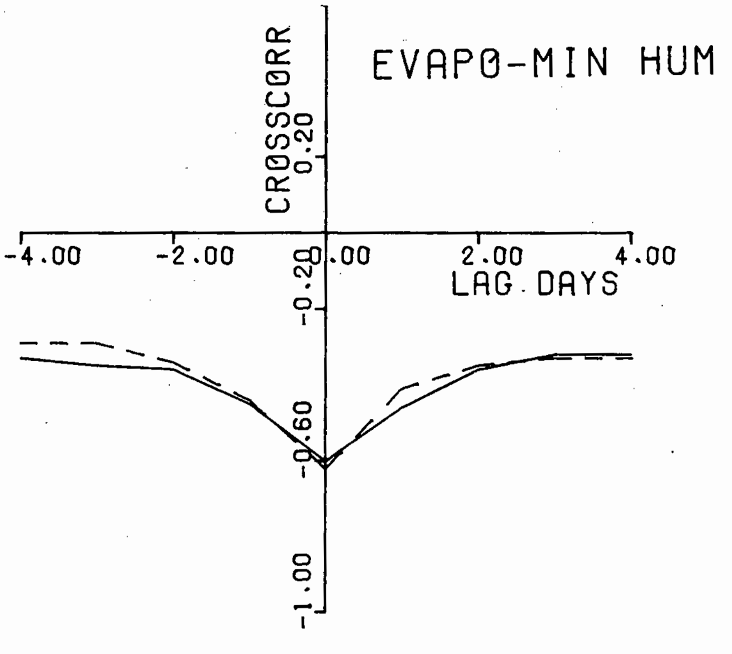

MIN TEMP-WIND
CROSSCORR
1.00
0.60
0.20
-0.20
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

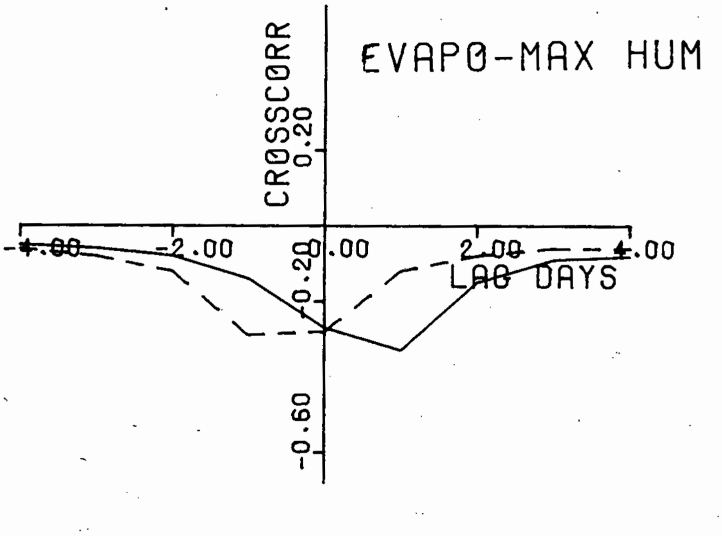

MIN TEMP-MIN HUM
CROSSCORR
0.20
-0.20
-0.60
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

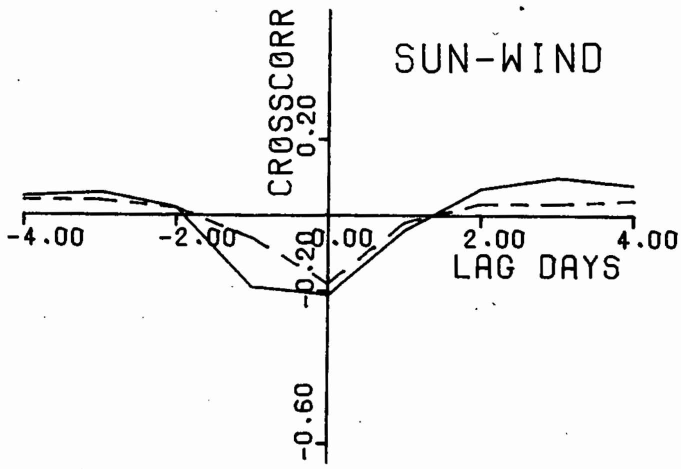

MIN TEMP-MAX HUM
CROSSCORR
0.20
-0.20
-0.60
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

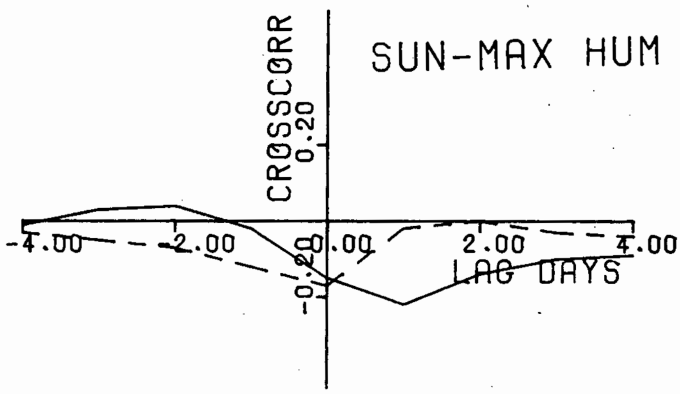

EVAPO-SUN
CROSSCORR
1.00
0.60
0.20
-0.20
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

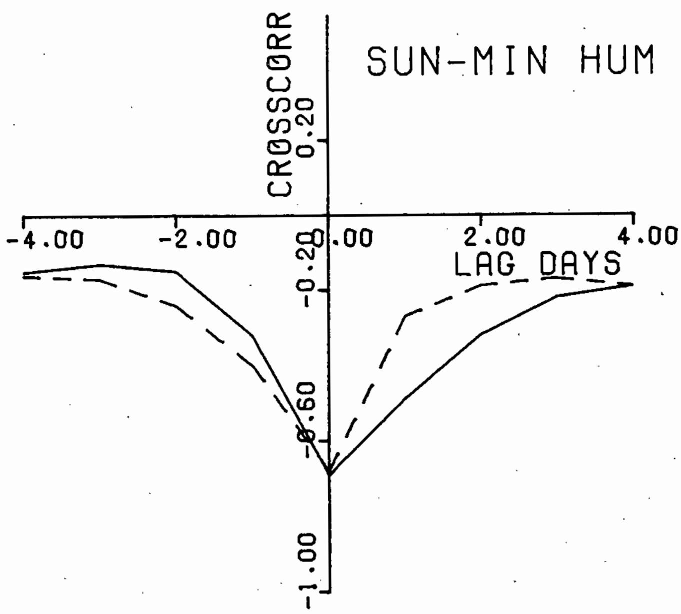

EVAPO-WIND
CROSSCORR
1.00
0.60
0.20
-0.20
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

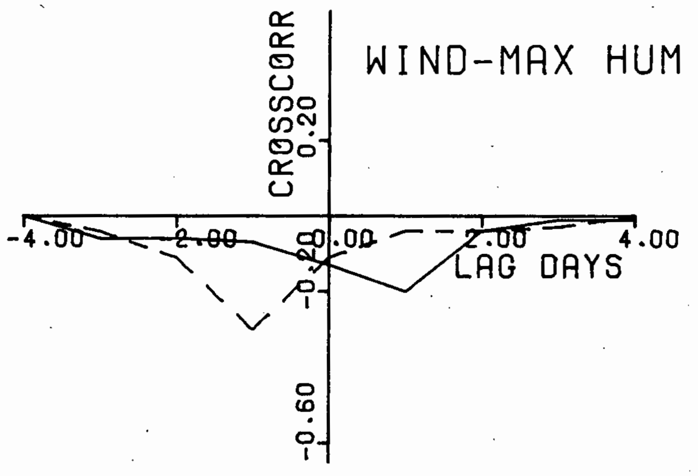

EVAPO-MIN HUM
CROSSCORR
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

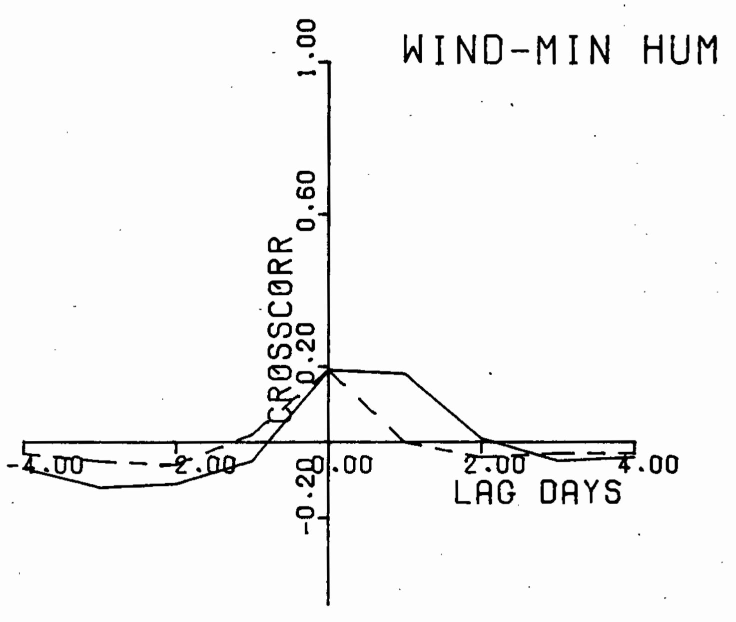

EVAPO-MAX HUM
CROSSCORR
0.20
-0.20
-0.60
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS


SUN-WIND
CROSSCORR
0.20
-0.20
-0.60
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS


SUN-MAX HUM
CROSSCORR
0.20
-0.20
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS


SUN-MIN HUM
CROSSCORR
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS


WIND-MAX HUM
CROSSCORR
0.20
-0.20
-0.60
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS


WIND-MIN HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

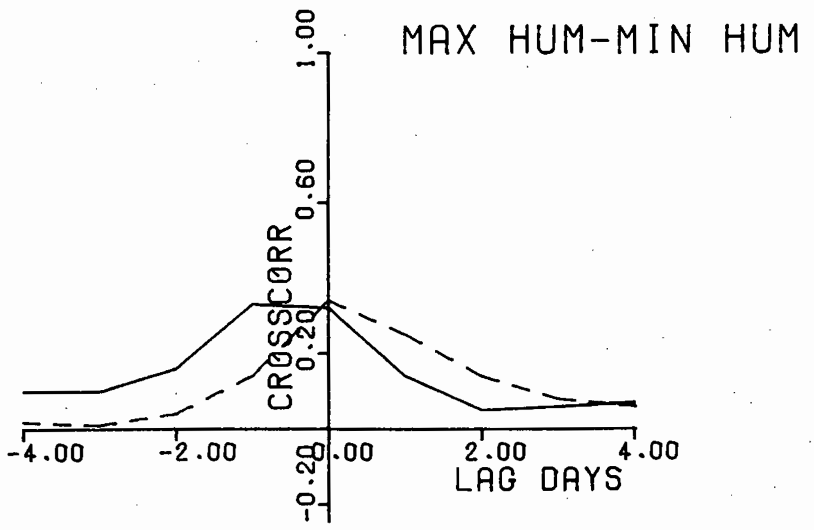
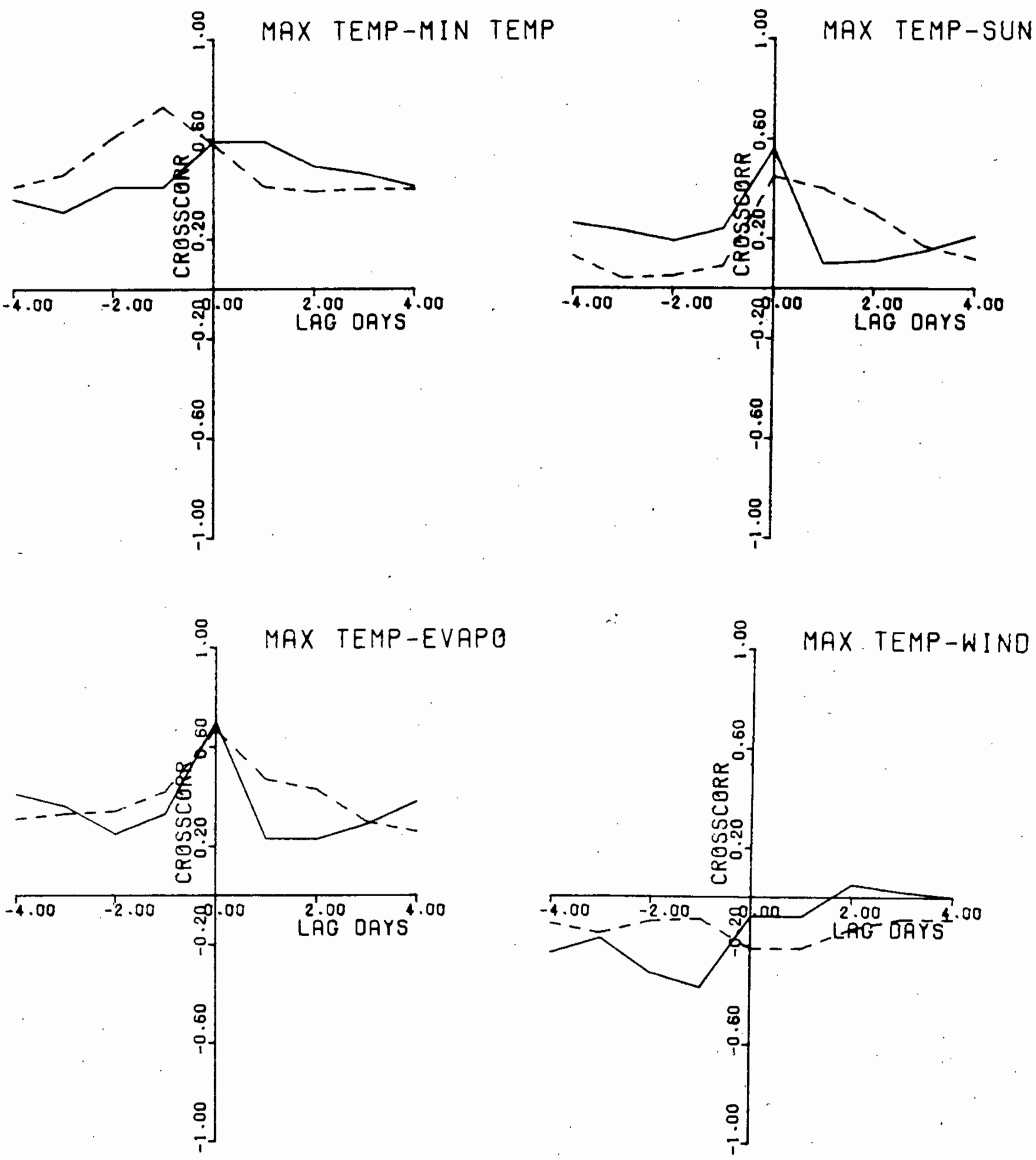
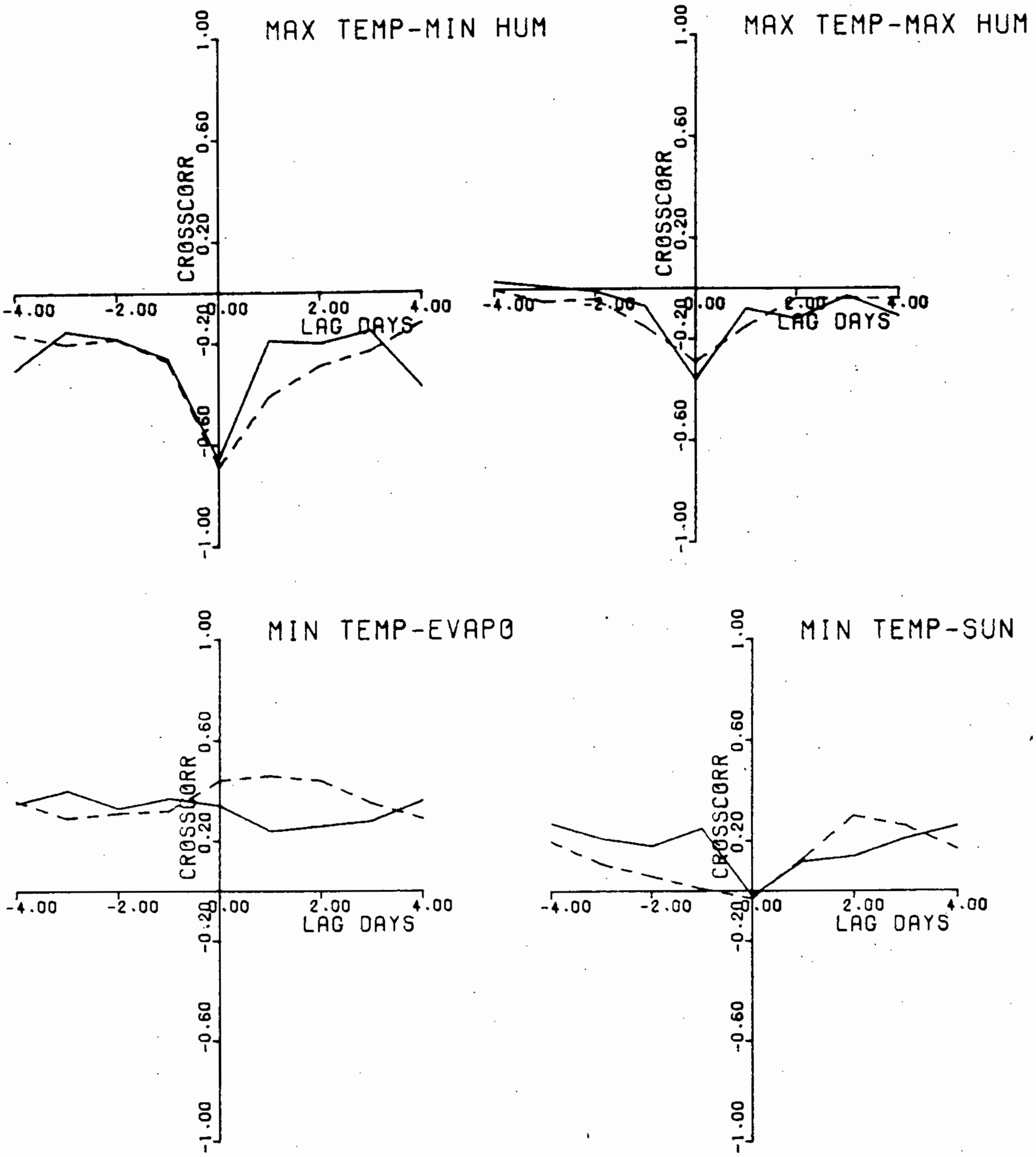
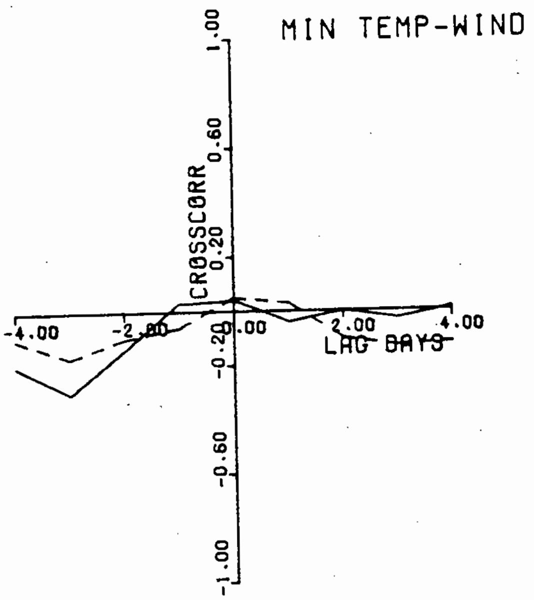
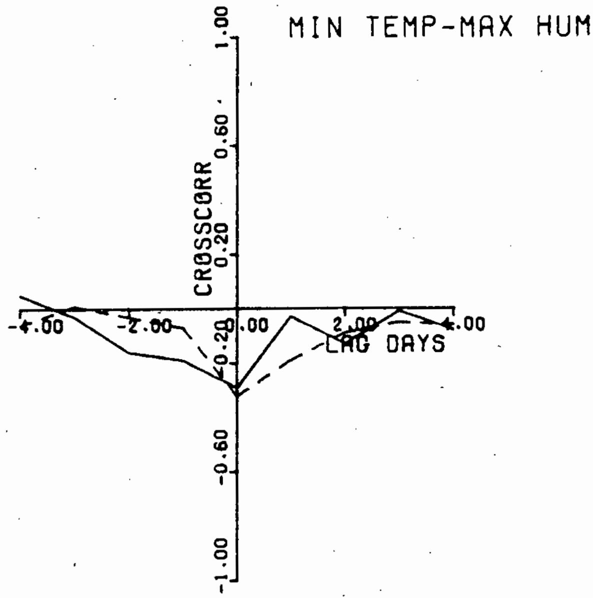

MAX HUM-MIN HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

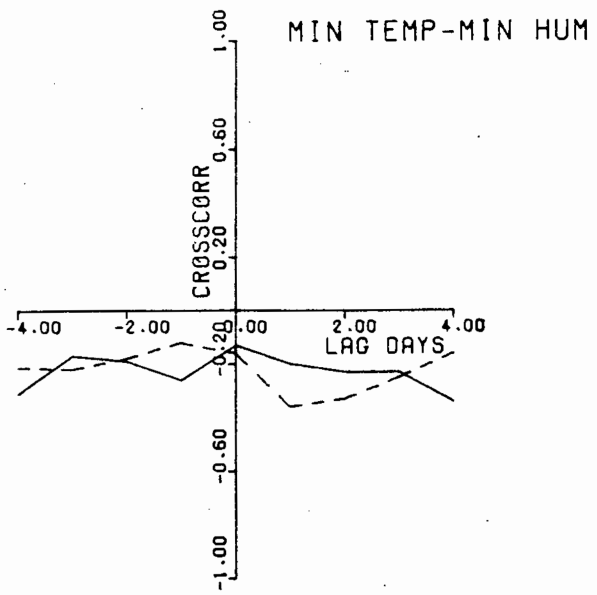
FIGURE 6.19: HISTORICAL (——) AND SIMULATED (---) CROSS-CORRELATION COEFFICIENTS GIVEN A WET DAY

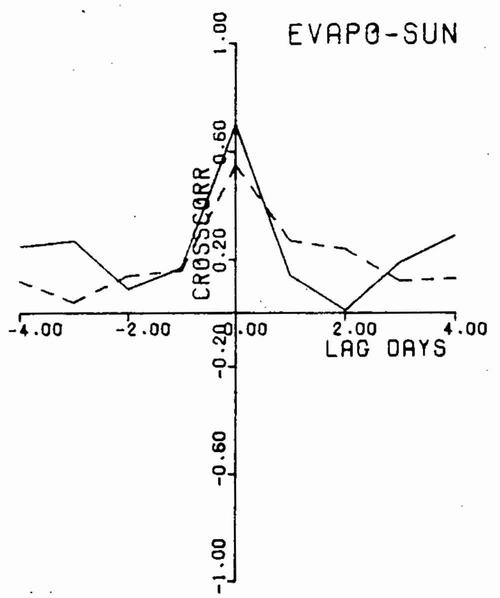


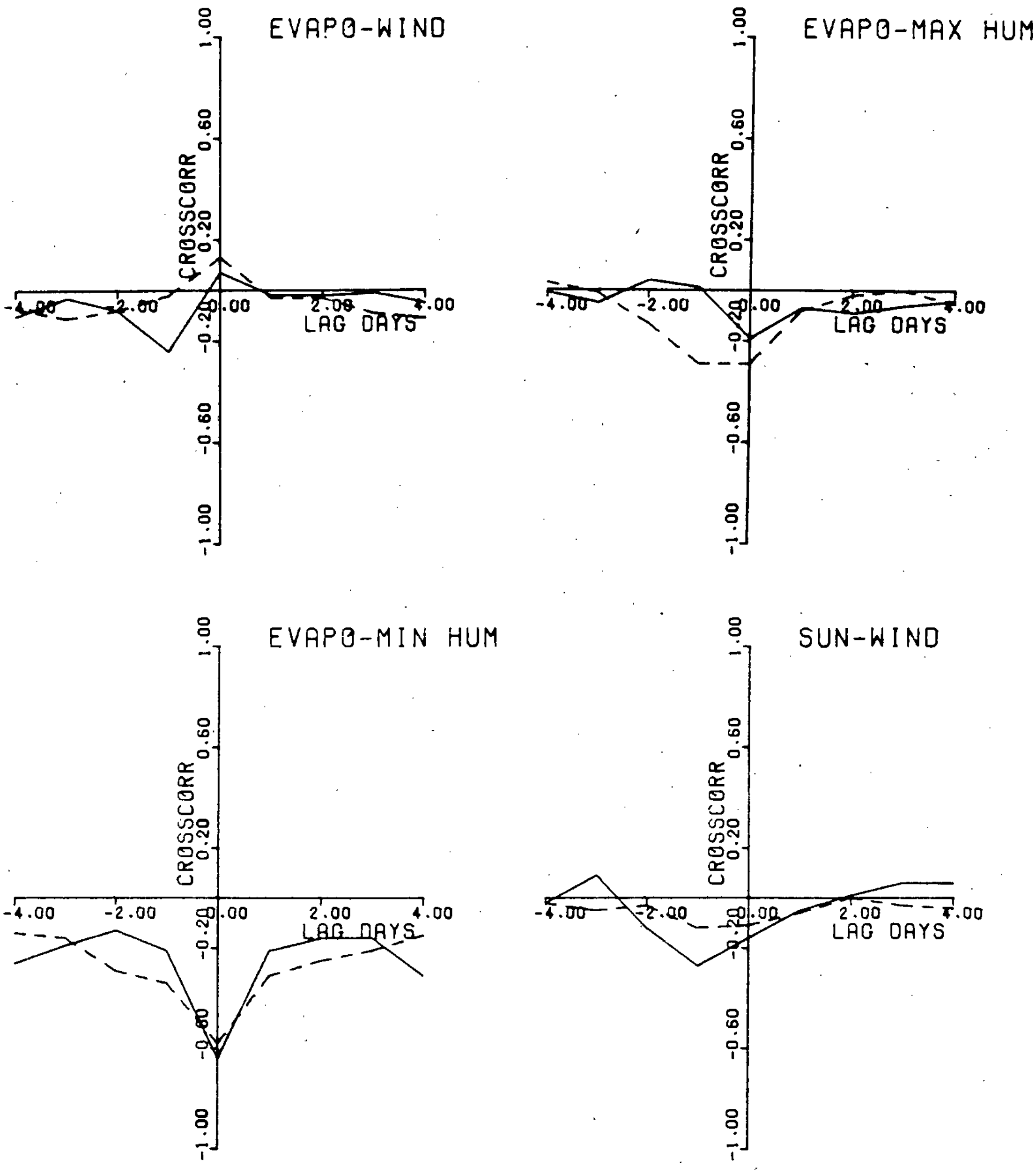

MAX TEMP-MIN HUM
MAX TEMP-MAX HUM
MIN TEMP-EVAPO
MIN TEMP-SUN
CROSSCORR
LAG DAYS

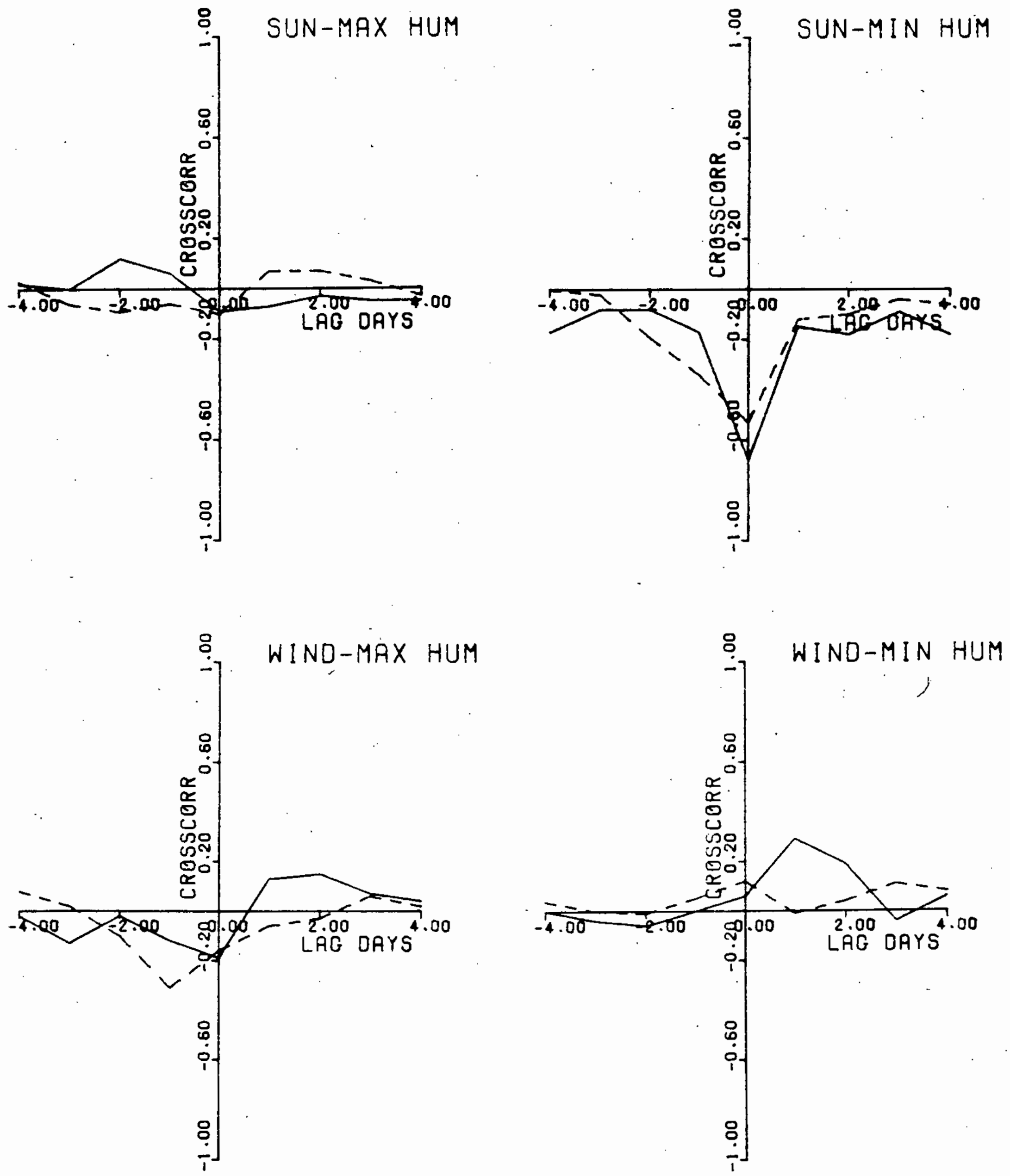

MIN TEMP-WIND
CROSSCORR
LAG DAYS


MIN TEMP-MAX HUM
CROSSCORR
LAG DAYS


MIN TEMP-MIN HUM
CROSSCORR
LAG DAYS


EVAPO-SUN
CROSSCORR
LAG DAYS


EVAPO-WIND
EVAPO-MAX HUM
EVAPO-MIN HUM
SUN-WIND
CROSSCORR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00


SUN-MAX HUM
SUN-MIN HUM
WIND-MAX HUM
WIND-MIN HUM
CROSSCORR
LAG DAYS
-1.00
-0.60
-0.20
0.20
0.60
1.00
-4.00
-2.00
0.00
2.00
4.00

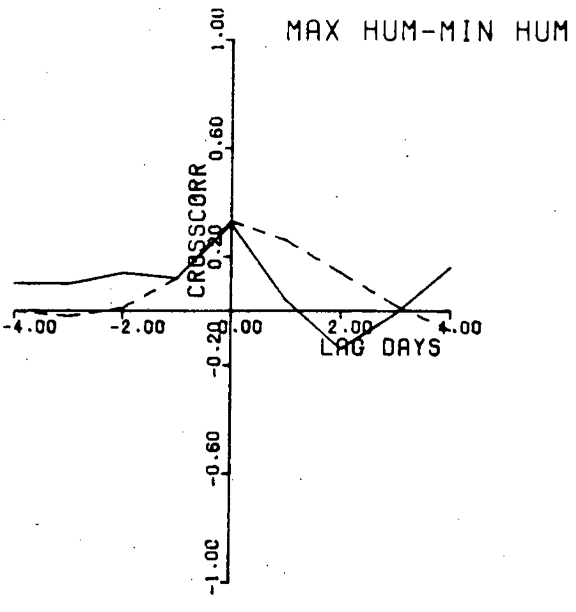
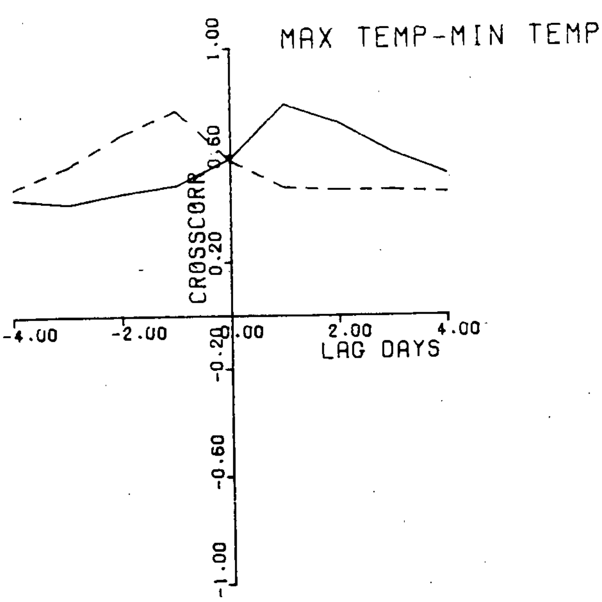
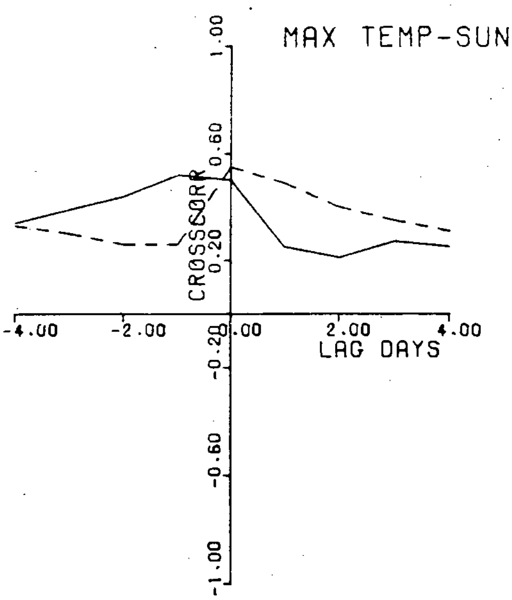
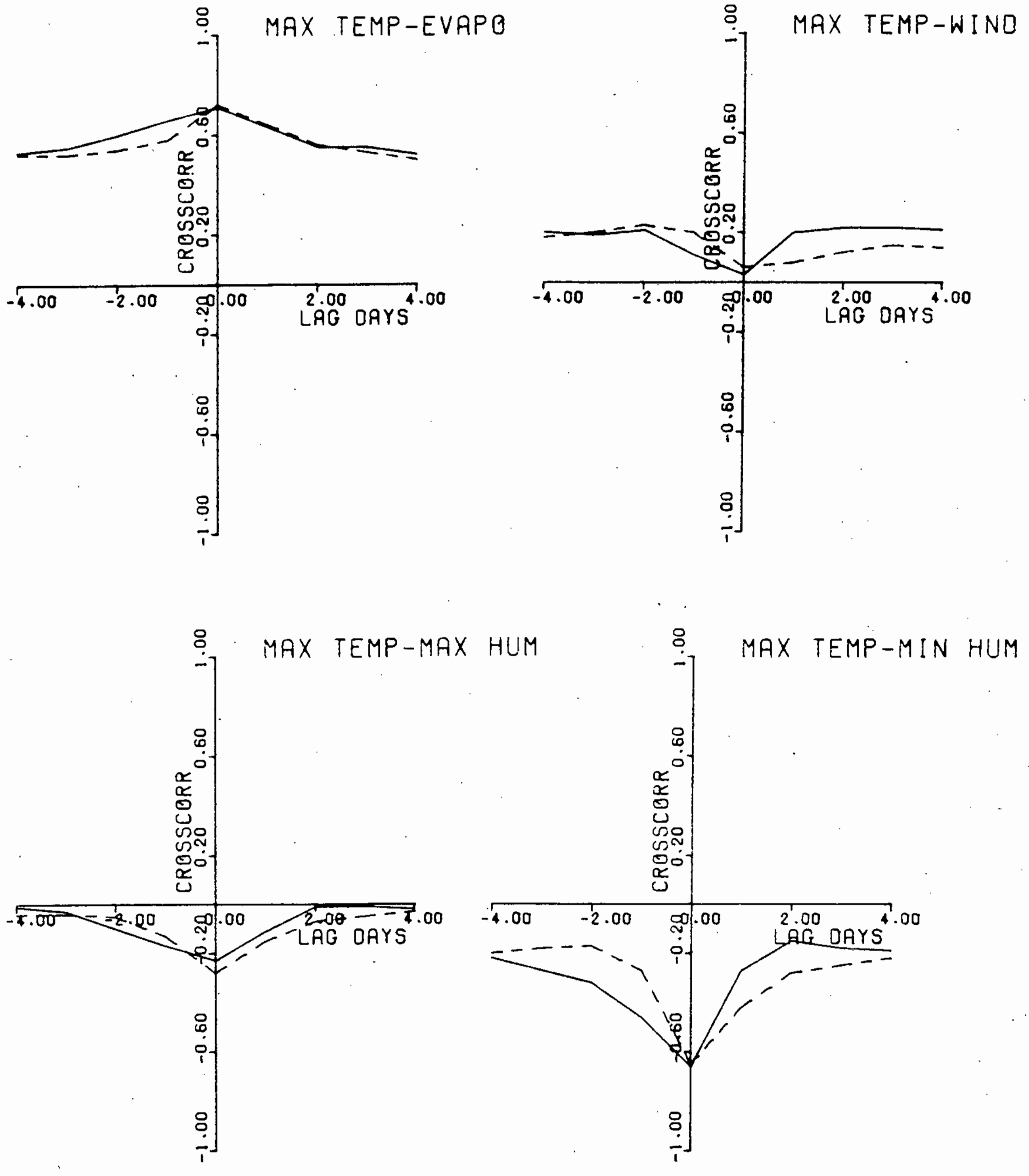
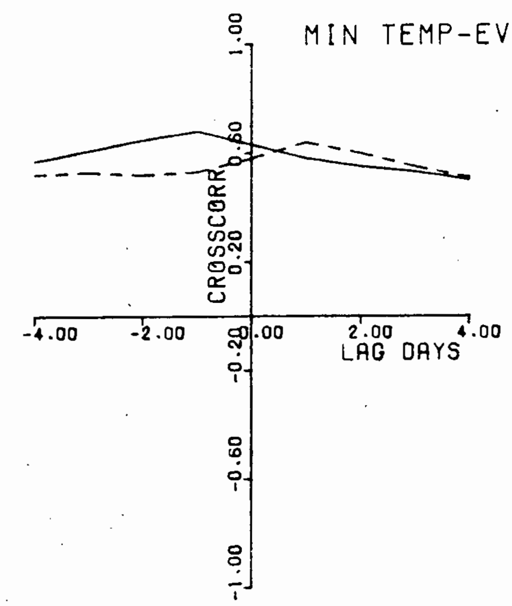


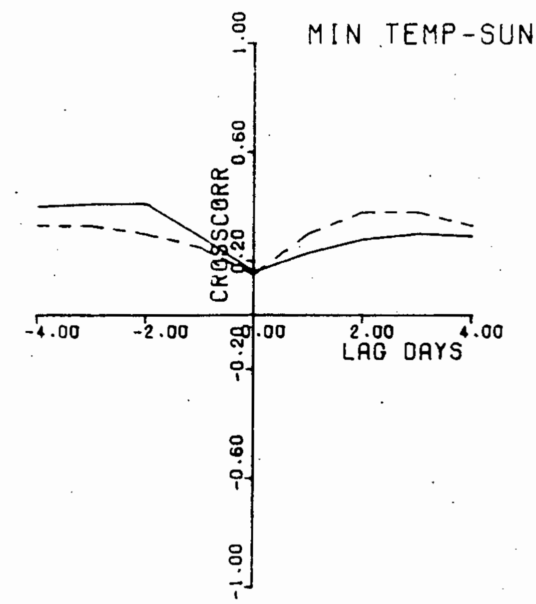
FIGURE 6.20: HISTORICAL (——) AND SIMULATED (---) CROSS-CORRELATION COEFFICIENTS GIVEN A DRY DAY

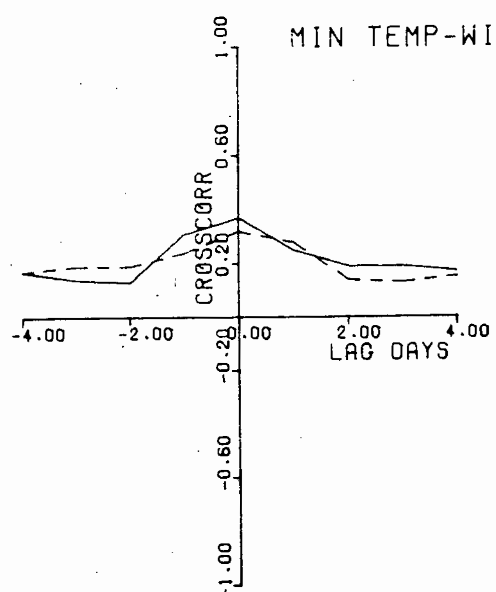


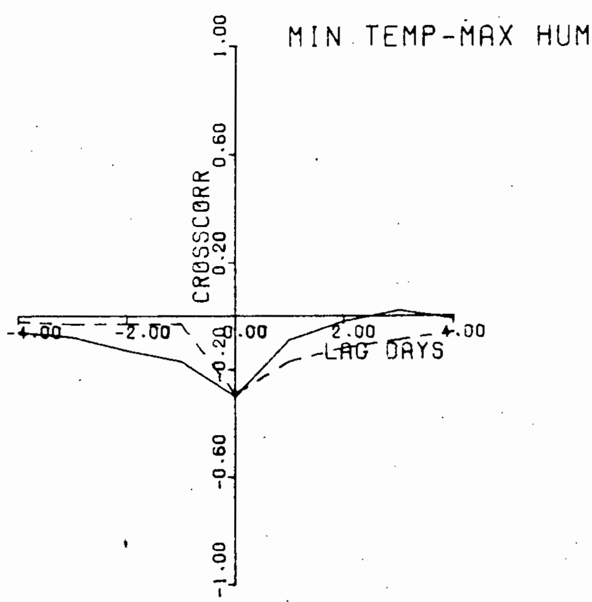


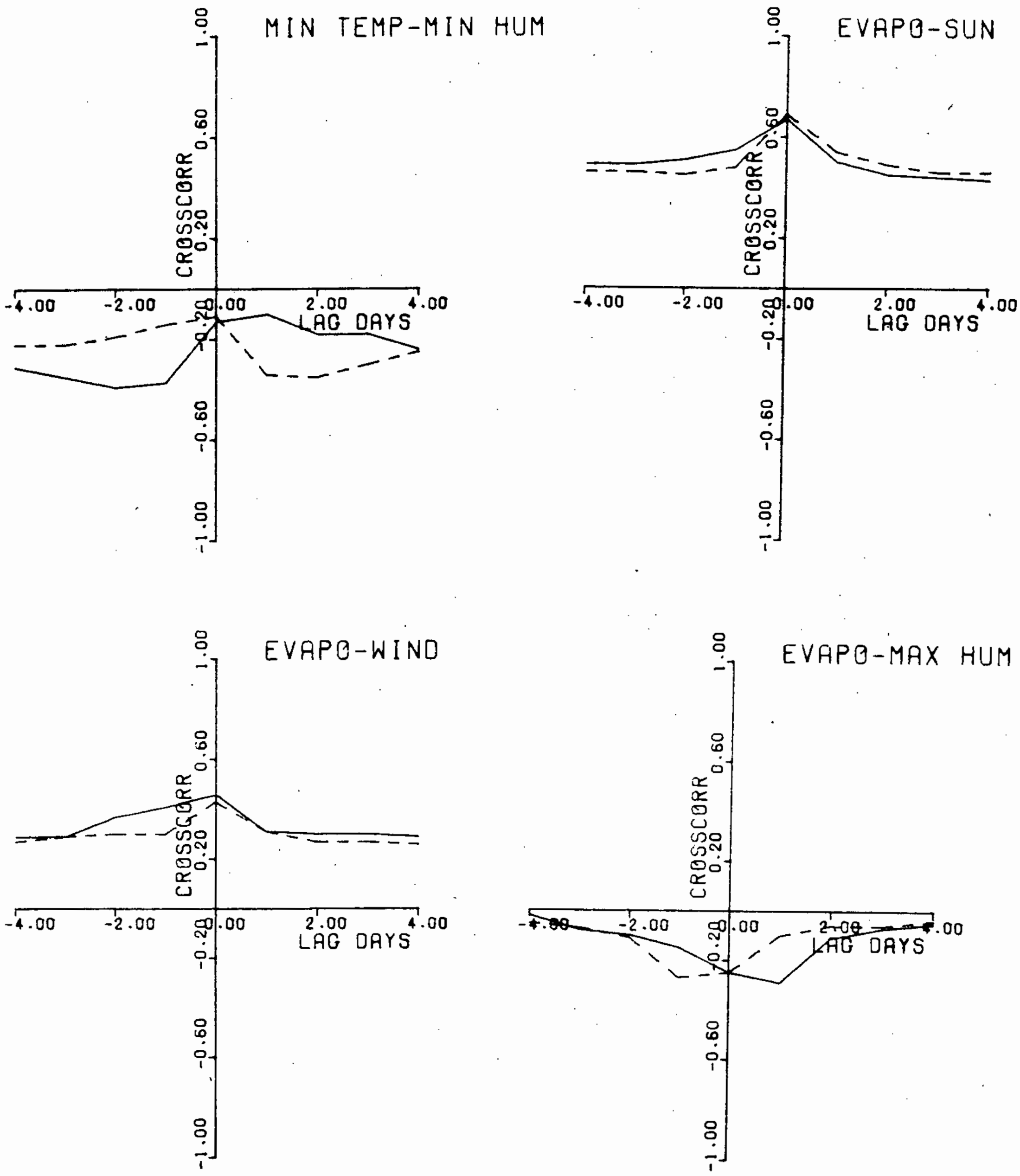

MAX TEMP-EVAPO
MAX TEMP-WIND
MAX TEMP-MAX HUM
MAX TEMP-MIN HUM
CROSSCORR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00

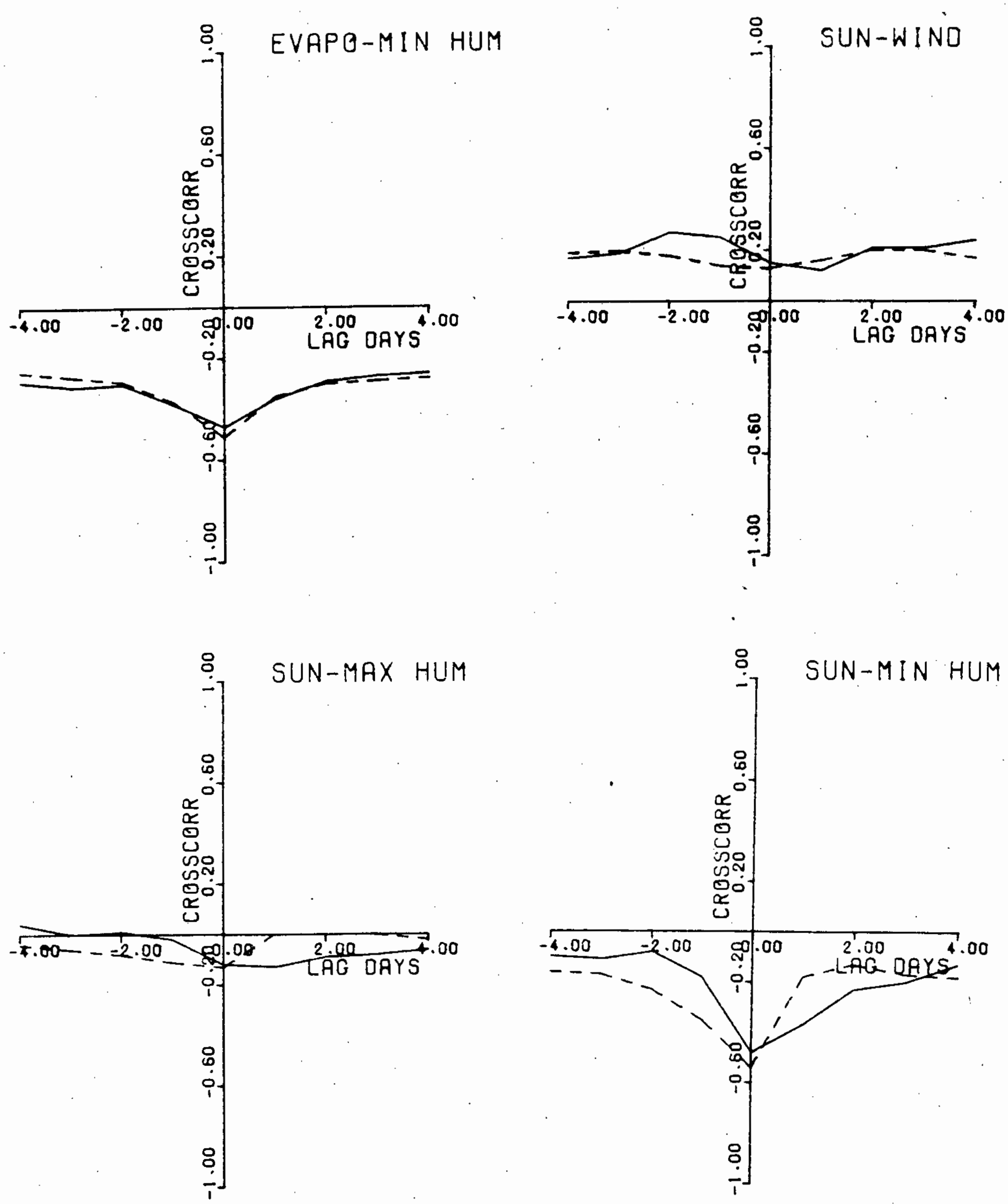

MIN TEMP-EVAPO
CROSSCORR
LAG DAYS

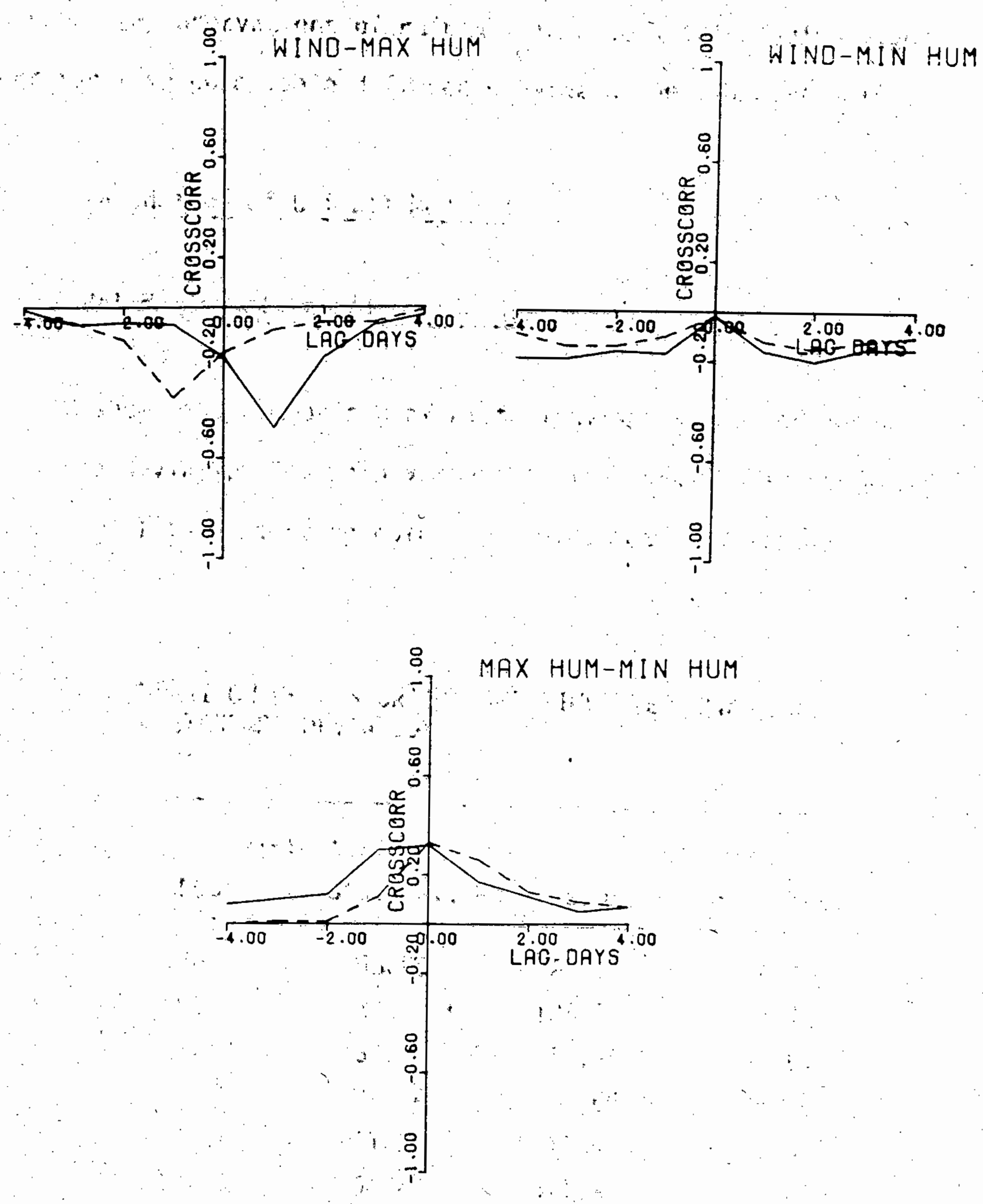

MIN TEMP-SUN
CROSSCORR
LAG DAYS


MIN TEMP-WIND
CROSSCORR
LAG DAYS


MIN TEMP-MAX HUM
CROSSCORR
LAG DAYS


MIN TEMP-MIN HUM
EVAPO-SUN
EVAPO-WIND
EVAPO-MAX HUM
CROSSCORR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00


EVAPO-MIN HUM
CROSSCORR
LAG DAYS
SUN-WIND
CROSSCORR
LAG DAYS
SUN-MAX HUM
CROSSCORR
LAG DAYS
SUN-MIN HUM
CROSSCORR
LAG DAYS


WIND-MAX HUM
CROSSCORR
LAG DAYS
WIND-MIN HUM
CROSSCORR
LAG DAYS
MAX HUM-MIN HUM
CROSSCORR
LAG-DAYS

Generally, the cross-correlations have been preserved by the model for both wet and dry sequences, with the cross-correlations being retained better when a dry sequence occurs. The small number of observations (and a number of missing values present) in a wet sequence due to the relatively few observations of rainfall in a semi-arid region might account for this observable difference between the two data sets.

6.2.2 Validation of Climate Model 2.

(a) Validation of annual properties.

Table 6.10 shows the comparison of historical and simulated annual mean and standard deviation for each variable. This shows that the model is very successful in preserving both the annual mean and standard deviation.

TABLE 6.10: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MEAN AND STANDARD DEVIATION

Variable	Historical		Simulated	
	Mean	Std. Dev.	Mean	Std. Dev.
Max Temp	226.5	58.90	223.1	54.95
Min Temp	106.1	38.34	106.9	37.42
Evapo	57.30	37.01	57.74	36.61
Sunshine	82.61	36.74	82.65	34.99
Windrun	1966	911.2	1954	878.9
Max Hum	914.5	96.29	911.5	95.28
Min Hum	412.7	152.8	411.0	147.9

The same check was performed on the annual means and standard deviations of the conditioned data, on the wet or dry status of the day. The results obtained are illustrated on Tables 6.11 - 6.12.

TABLE 6.11: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MEAN AND STANDARD DEVIATION GIVEN A WET DAY

Variable	Historical		Simulated	
	Mean	Std. Dev.	Mean	Std. Dev.
Max Temp	182.1	46.73	185.6	39.19
Min Temp	110.1	33.43	109.0	31.39
Evapo	28.49	26.99	29.36	25.09
Sunshine	41.71	33.75	42.40	29.54
Windrun	2500	1194	2463	1160
Max Hum	933.0	72.90	932.8	26.99
Min Hum	548.9	157.7	546.9	147.7

From the above table it is clear that the annual means have been well described by the model. Except for the variable maximum humidity the model is successful in maintaining the structure of the annual standard deviation. For maximum humidity the annual standard deviation has been underestimated.

TABLE 6.12: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MEAN AND STANDARD DEVIATION GIVEN A DRY DAY

Variable	Historical		Simulated	
	Mean	Std. Dev.	Mean	Std. Dev.
Max Temp	241.4	54.95	241.7	52.25
Min Temp	104.8	39.77	106.2	39.14
Evapo	66.48	35.00	66.84	35.04
Sunshine	96.16	26.12	95.54	25.53
Windrun	1788	710.4	1792	692.8
Max Hum	908.3	102.3	904.7	107.5
Min Hum	366.7	120.3	367.4	118.7

Both the annual mean and standard deviation of the simulated sequence compare favourably with those of the historical record, when the variables are conditioned on the dry status of the day.

A comparison between the historical and simulated extreme values is shown in Table 6.13. The model performs well with respect to the maximum value for most of the variables but does not perform as well with respect to the minimum value. To assess how the minimum values behave in the simulated sequence with regard to the historical data, Table 6.14 shows the number of times values in the simulated sequence are less than the minimum value observed in the historical data, or when a value exceeds the historical maximum value.

TABLE 6.13: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MAXIMUM AND MINIMUM VALUES

Variable	Historical		Simulated	
	Maximum	Minimum	Maximum	Minimum
Max Temp	408.0	100.0	406.7	76.40
Min Temp	238.0	13.00	232.1	-17.40
Evapo	185.0	0.00	161.3	-38.37
Sunshine	133.0	0.00	177.9	-61.41
Windrun	7376	293.0	6556	-1294
Max Hum	1000	280.0	1255	514.5
Min Hum	950.0	120.0	978.1	-93.40

TABLE 6.14: NUMBER OF TIMES SIMULATED VALUES LIE OUTSIDE A CERTAIN DEVIANCE FROM THE HISTORICAL MAXIMUM AND MINIMUM

Variable	Greater than Maximum	Less than Minimum
Max Temp	0 (0%)	558 (8%)
Min Temp	0 (0%)	17 (0.2%)
Evapo	0 (0%)	337 (5%)
Sunshine	72 (0.1%)	144 (2%)
Windrun	0 (0%)	125 (2%)
Max Hum	477 (7%)	0 (0%)
Min Hum	0 (0%)	63 (0.1%)

These results show that none of the variables have a large percentage of values either below the observed minimum value or above the extreme maximum value observed.

The same test just discussed was repeated on the conditioned sequences. The results are given in Tables 6.15 - 6.16.

TABLE 6.15: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MAXIMUM AND MINIMUM VALUES GIVEN A WET DAY

Variable	Historical		Simulated	
	Maximum	Minimum	Maximum	Minimum
Max Temp	365.0	100.0	300.4	103.3
Min Temp	238.0	23.00	209.1	20.70
Evapo	122.0	0.00	108.5	-38.37
Sunshine	124.0	0.00	134.7	-61.41
Windrun	7376	618.0	6556	-1294
Max Hum	1000	340.0	1017	839.2
Min Hum	950	170.0	978.1	48.50

When the data is conditioned on the wet status of the day, the model retains the maximum value of the sequences. For a dry sequence, the model does not perform as well for some of the variables. For the variables maximum and minimum temperature the minimum values are preserved by the model when the sequences are conditioned on the wet status of the day, otherwise the minimum values are not preserved.

TABLE 6.16: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MAXIMUM AND MINIMUM VALUES GIVEN A DRY DAY

Variable	Historical		Simulated	
	Maximum	Minimum	Maximum	Minimum
Max Temp	408.0	120.0	406.7	76.40
Min Temp	234.0	13.00	232.1	-17.40
Evapo	185.0	0.00	161.3	-30.89
Sunshine	133.0	0.00	177.9	17.58
Windrun	6804	293.0	4042	-989.0
Max Hum	1000	280.0	1255	514.5
Min Hum	850.0	120.0	789.8	-93.40

(b) Validation of monthly properties.

The comparison of historical and simulated monthly means for all variables is illustrated in Figure 6.21.

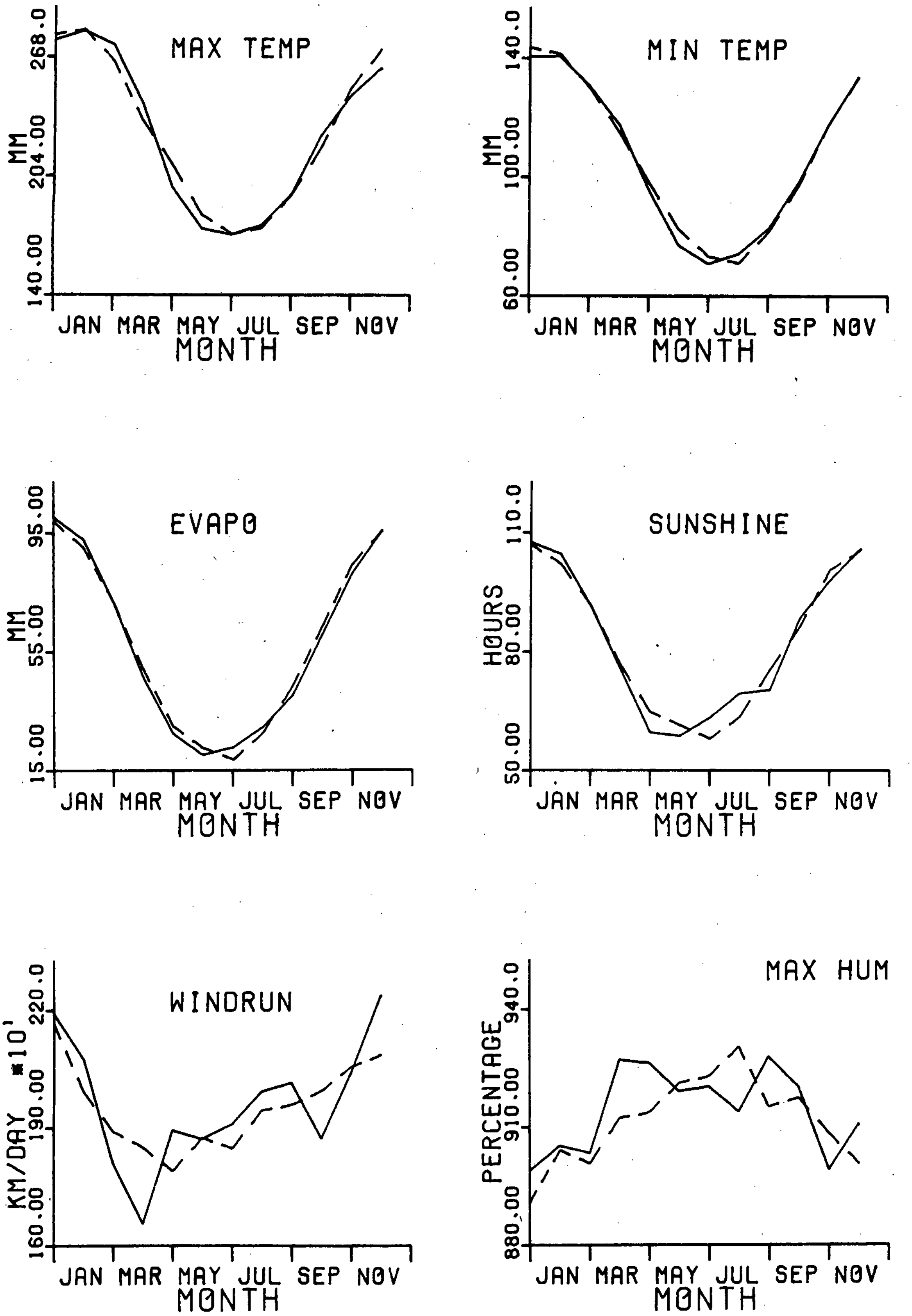
The performance of the model in preserving the monthly means is very satisfactory. For the month of April, in the variable windrun there is a fifteen percent difference between historical and simulated monthly mean. This could easily be an anomaly in the historical record.

Figure 6.22 shows the comparison of historical and simulated monthly standard deviations for all variables.

For maximum temperature, the standard deviations for the months of April and October have been underestimated. In fact for every month the standard deviations have been slightly underestimated. For the variable maximum humidity, there is a difference between the historical and the simulated standard deviations for the months of April and August. The standard deviations for minimum humidity for the months of June and July are underestimated. For the variable windrun, the standard deviation is either overestimated or underestimated for some of the months.

The monthly means, conditioned on the wet or dry status of the day have been preserved by the model. Again the variables maximum humidity and windrun show differences between historical and simulated sequences for some months. These results are contained in Figures 6.23 - 6.24.

FIGURE 6.21: HISTORICAL (——) AND SIMULATED (---) MONTHLY MEANS FOR ALL VARIABLES



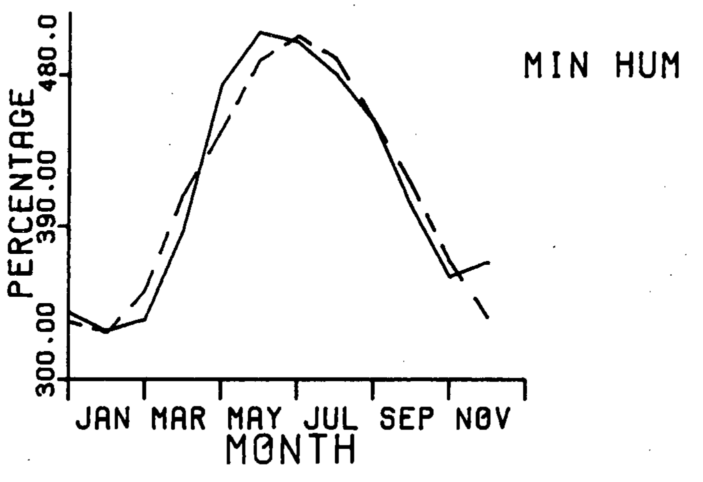
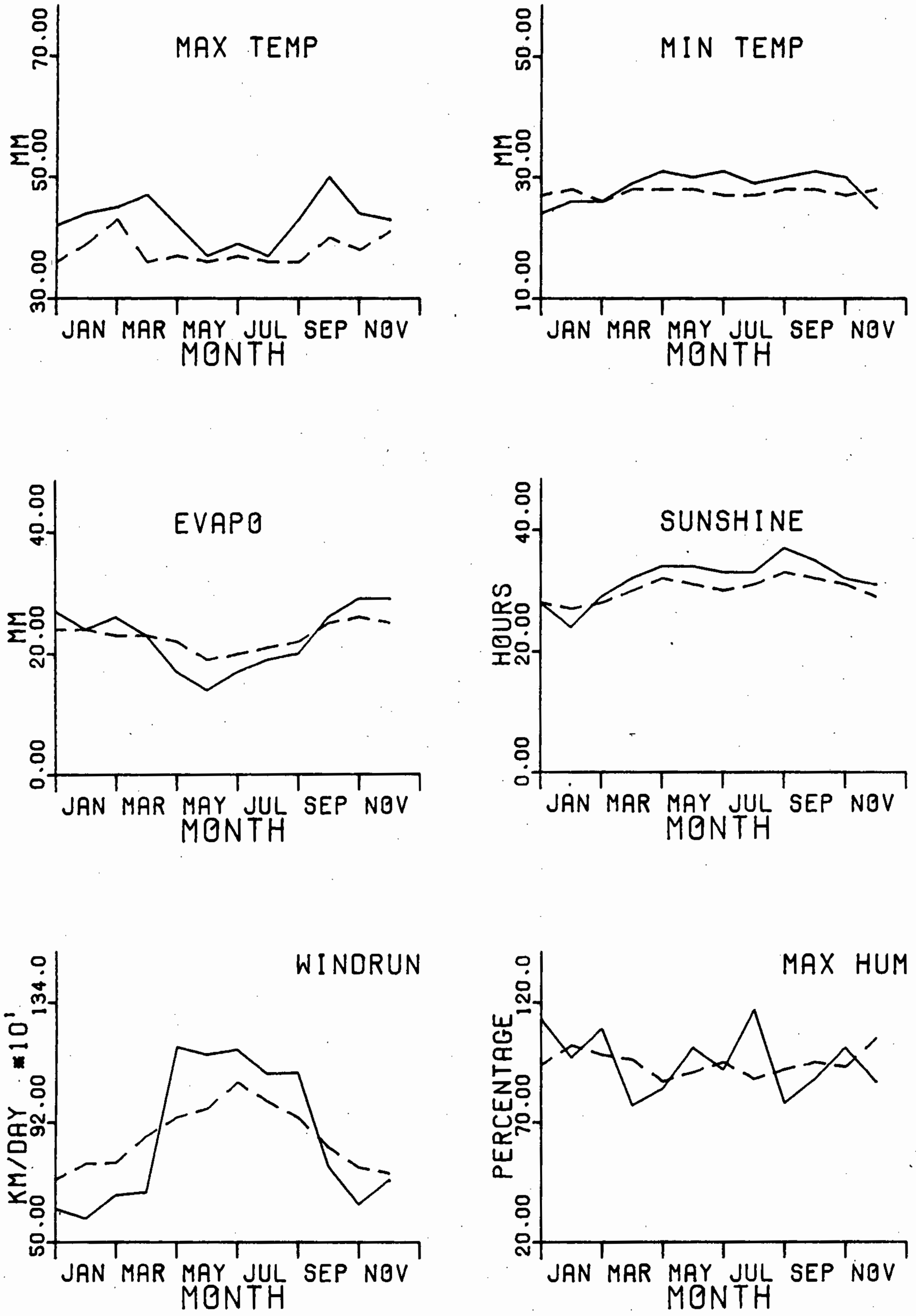

MIN HUM
PERCENTAGE
300.00
390.00
480.0
JAN MAR MAY JUL SEP NOV
MONTH

FIGURE 6.22: HISTORICAL (——) AND SIMULATED (---) MONTHLY STANDARD DEVIATIONS FOR ALL VARIABLES



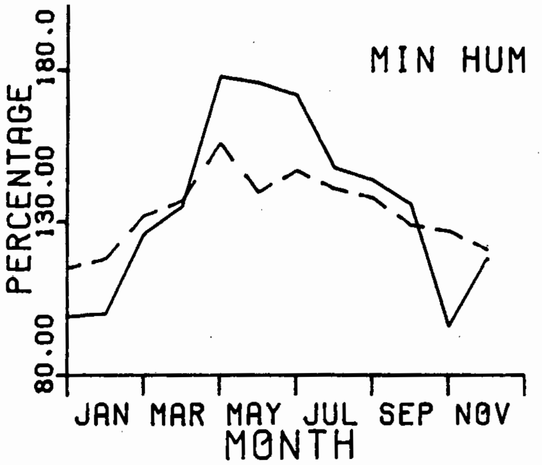
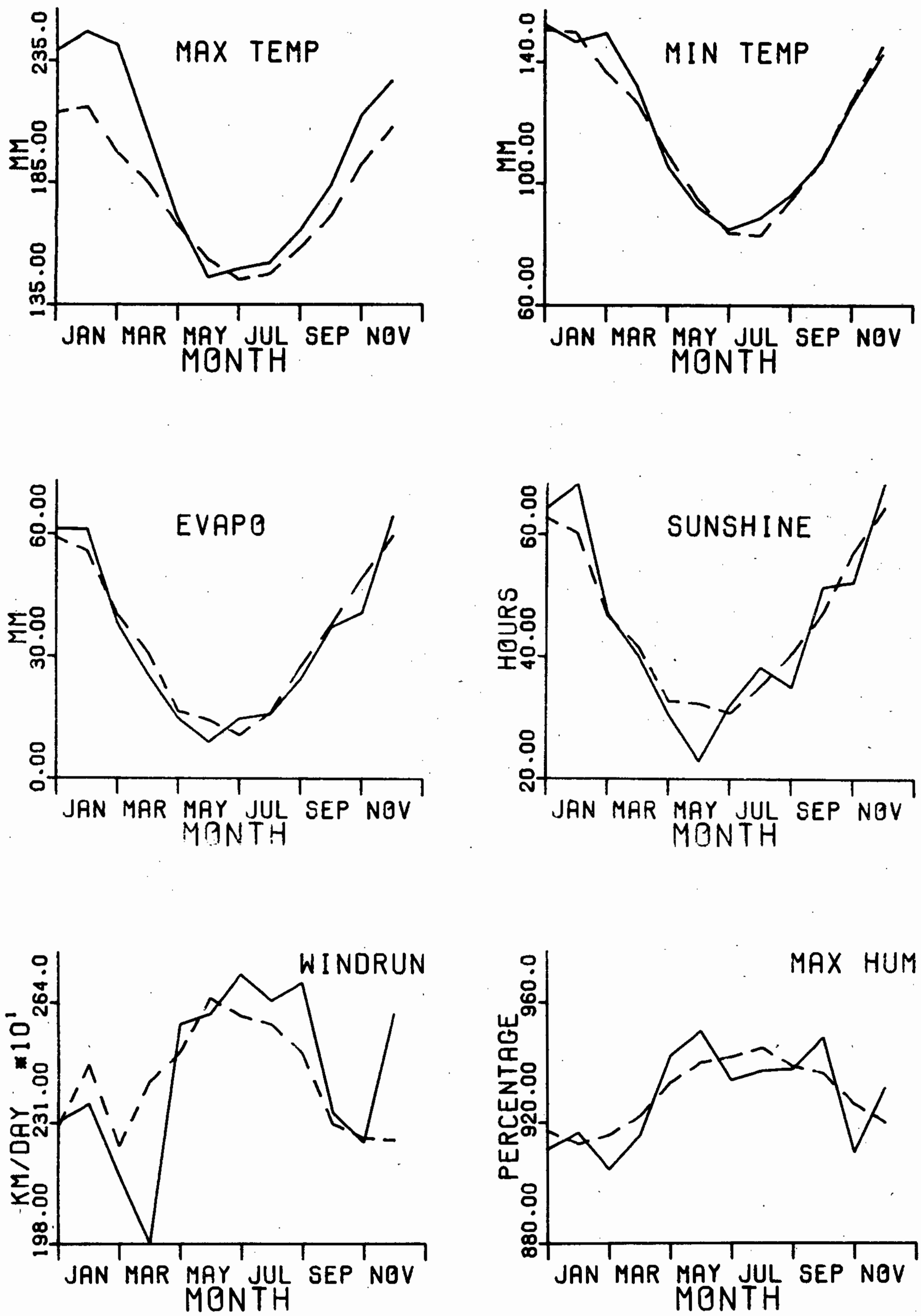

MIN HUM
PERCENTAGE
80.00
130.00
180.0
JAN MAR MAY JUL SEP NOV
MONTH

FIGURE 6.23: HISTORICAL (——) AND SIMULATED (---) MONTHLY MEANS GIVEN A WET DAY FOR ALL VARIABLES



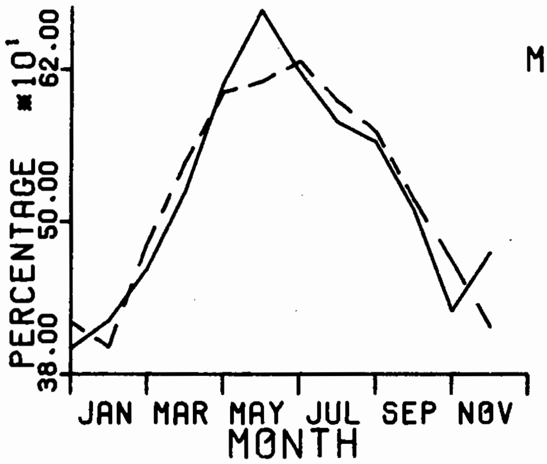
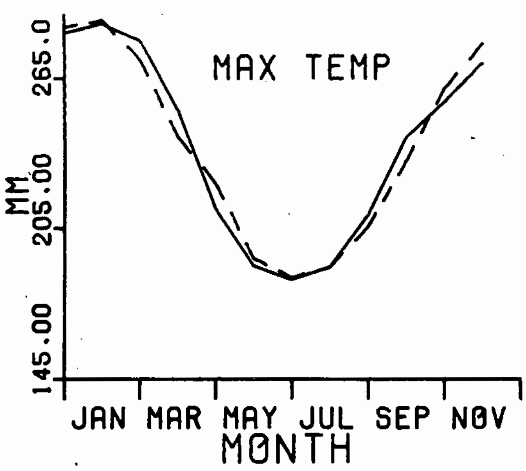
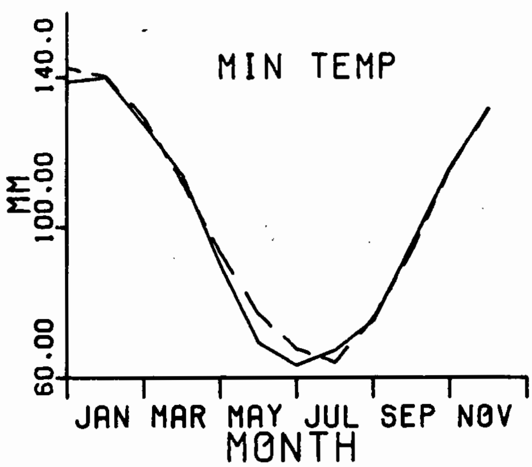
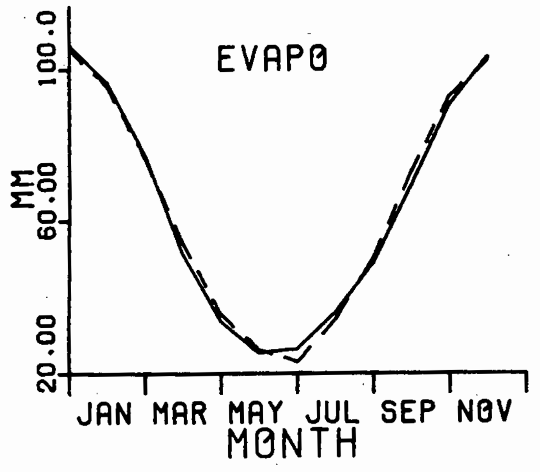
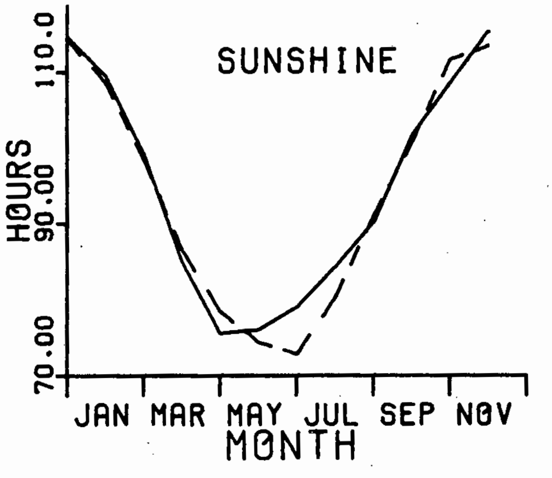


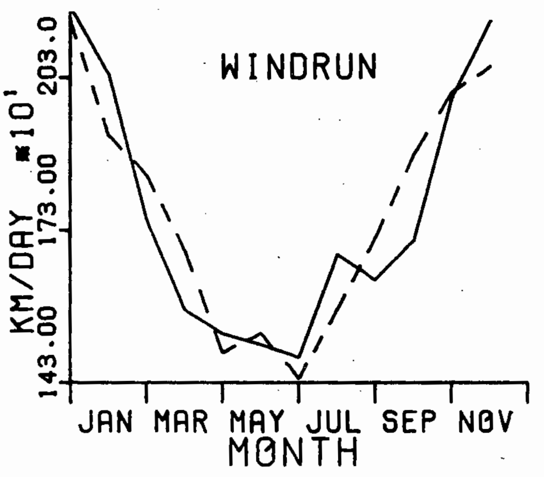
FIGURE 6.24: HISTORICAL (——) AND SIMULATED (---) MONTHLY MEANS GIVEN A DRY DAY FOR ALL VARIABLES

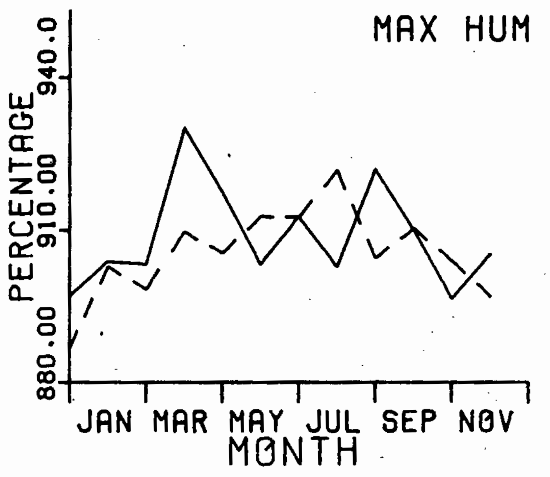


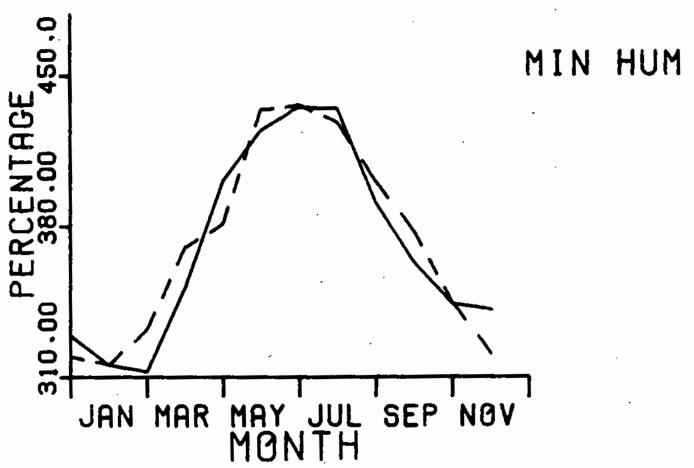



EVAPO
MM
20.00
60.00
100.0
JAN MAR MAY JUL SEP NOV
MONTH


SUNSHINE
HOURS
70.00
90.00
110.0
JAN MAR MAY JUL SEP NOV
MONTH


WINDRUN
KM/DAY *10[1]
143.00
173.00
203.0
JAN MAR MAY JUL SEP NOV
MONTH


MAX HUM
PERCENTAGE
880.00
910.00
940.0
JAN MAR MAY JUL SEP NOV
MONTH


MIN HUM
PERCENTAGE
310.00
380.00
450.0
JAN MAR MAY JUL SEP NOV
MONTH

Figures 6.25 - 6.26 show the comparison of historical and simulated monthly standard deviations for the climate variables conditioned on the status of the day.

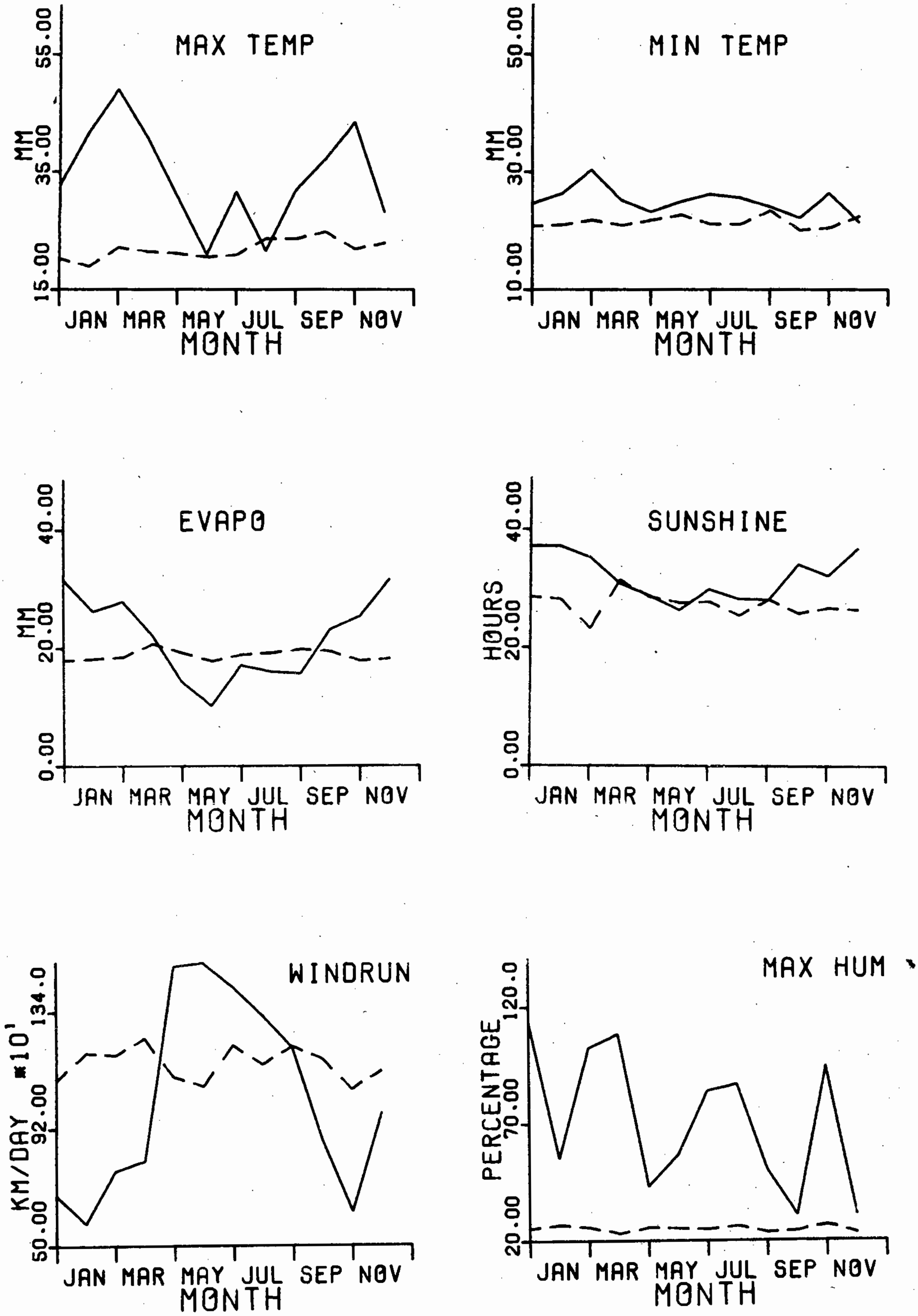
The variable maximum humidity, when conditioned on the dry status of the day, is overestimated for the month of April, while for minimum humidity, the standard deviation for the winter months is underestimated and so are some months when the climate sequences are conditioned on the wet status of the day. The variable windrun has been slightly underestimated in some months. The model has significantly underestimated the monthly standard deviations for the variable maximum temperature when conditioned on the wet status of the day. The same applies for the variable maximum humidity. Again windrun has been overestimated for some months and underestimated for others. All of the above variables for which the standard deviations of the simulated sequence differ from those of the historical record, show a large scatter of standard deviations over the months and this accounts for the differences observed between the simulated and historical standard deviations.

(c) Validation of daily properties.

The Fourier series model fitted to the mean is the same as in the model suggested by Richardson (1981) and the tests to compare the daily averages with the mean fitted by a 3-term Fourier series has already been discussed in the previous section.

It is required that the model proposed should maintain the autocorrelation within each variable. Figure 6.27 shows the autocorrelation

FIGURE 6.25: HISTORICAL (——) AND SIMULATED (---) MONTHLY STANDARD DEVIATIONS GIVEN A WET DAY FOR ALL VARIABLES



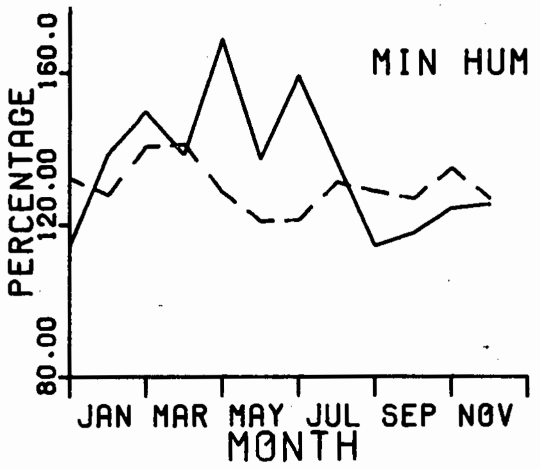
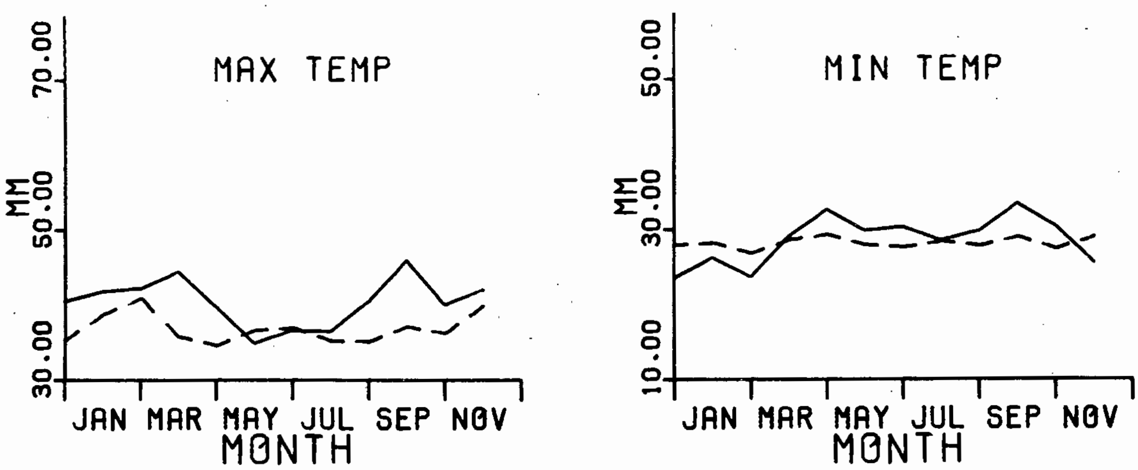
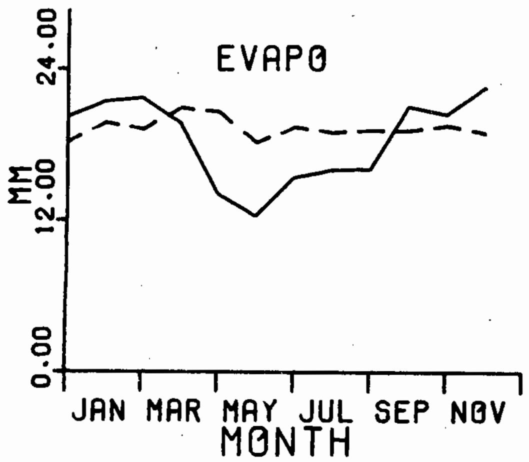
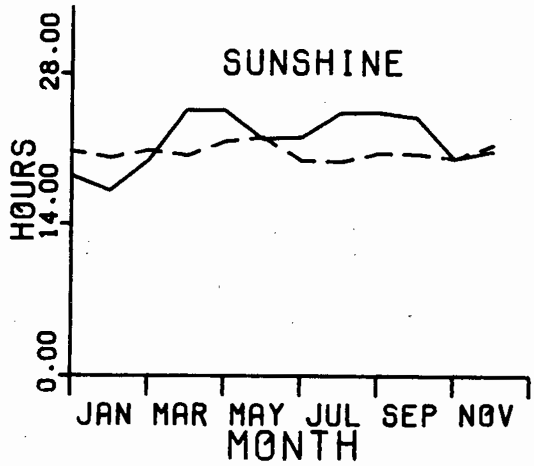
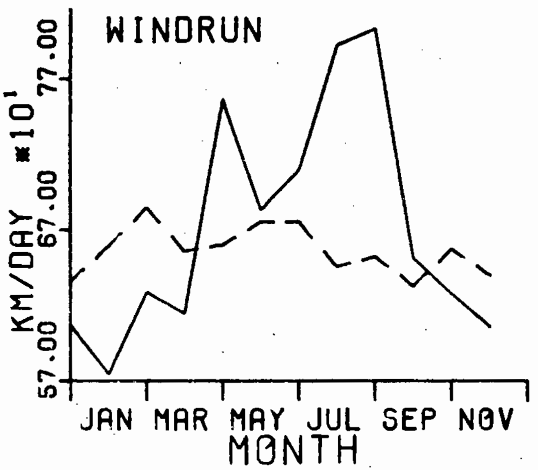


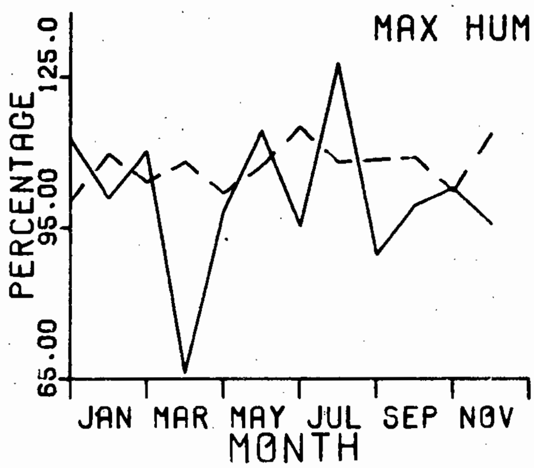
FIGURE 6.26: HISTORICAL (——) AND SIMULATED (---) MONTHLY STANDARD DEVIATIONS GIVEN A DRY DAY FOR ALL VARIABLES




EVAPO
MM
24.00
12.00
0.00
JAN MAR MAY JUL SEP NOV
MONTH


SUNSHINE
HOURS
28.00
14.00
0.00
JAN MAR MAY JUL SEP NOV
MONTH


WINDRUN
KM/DAY *10¹
77.00
67.00
57.00
JAN MAR MAY JUL SEP NOV
MONTH


MAX HUM
PERCENTAGE
125.0
95.00
65.00
JAN MAR MAY JUL SEP NOV
MONTH

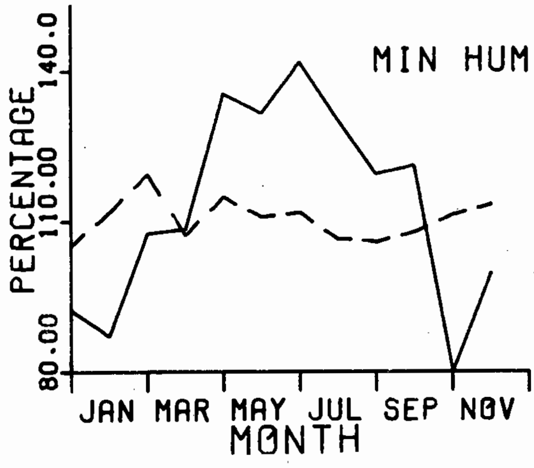
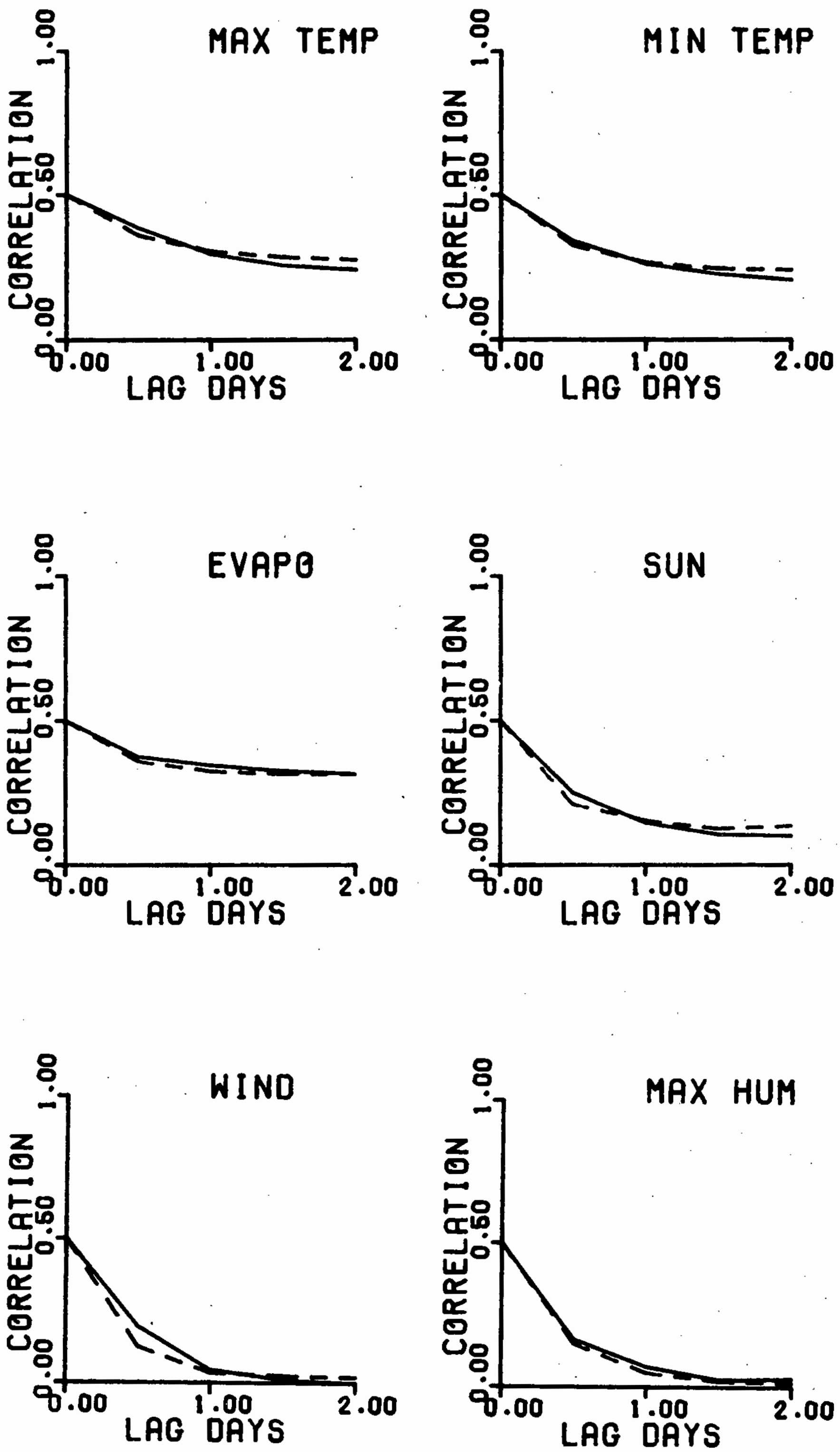

MIN HUM
PERCENTAGE
140.0
110.00
80.00
JAN MAR MAY JUL SEP NOV
MONTH

FIGURE 6.27: HISTORICAL (——) AND SIMULATED (---) AUTOCORRELATION COEFFICIENTS FOR ALL VARIABLES



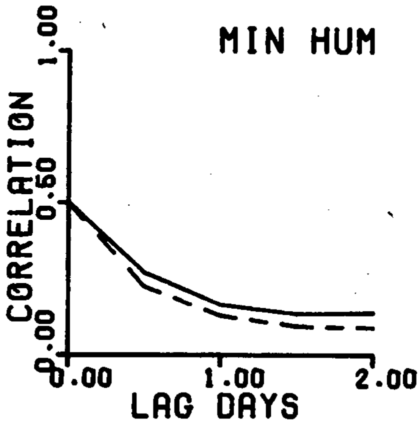
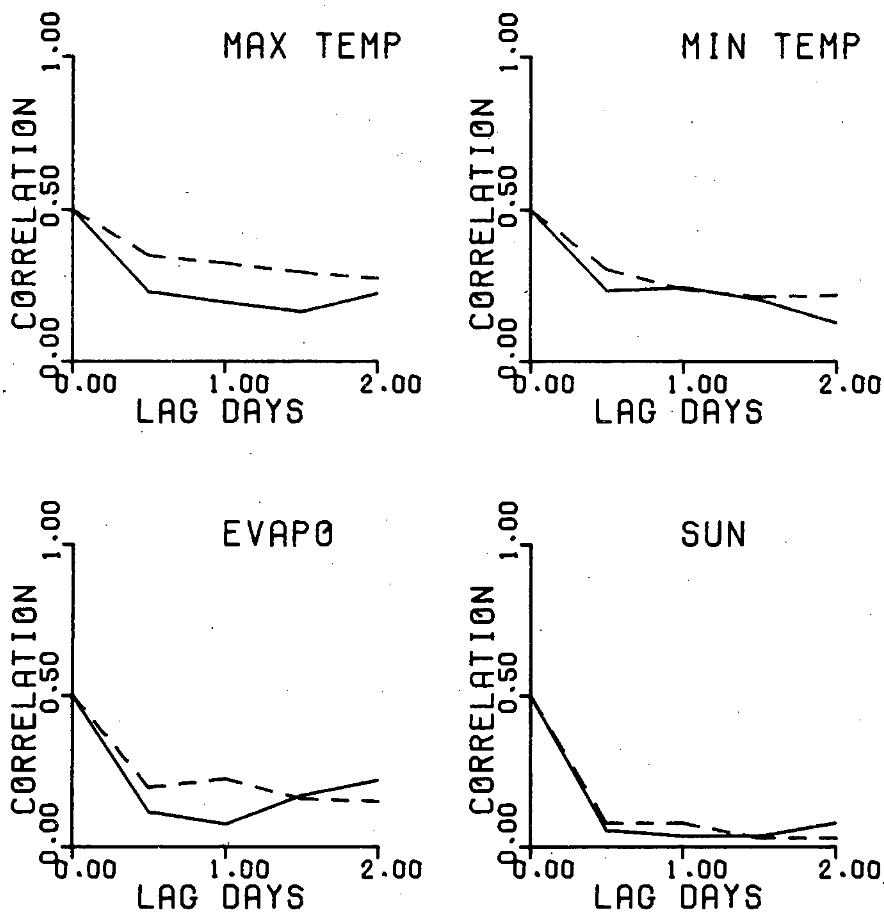


FIGURE 6.28: HISTORICAL (——) AND SIMULATED (---) AUTOCORRELATION COEFFICIENTS GIVEN A WET DAY FOR ALL VARIABLES



coefficients obtained from the historical data and those obtained from the simulated sequences. The plots show that for all variables the autocorrelation structure in the simulated sequences closely resemble that of the observed data.

Figures 6.28 - 6.29 show the comparison between the autocorrelation coefficients of the historical data and simulated sequences conditioned on the wet or dry status of the day.

The model has retained the autocorrelation structure in all variables both for the climate sequences conditioned on a dry day and for the sequence conditioned on the wet status of the day.

The cross-correlation coefficients of the simulated sequence compare favourably with those of the historical sequence. The only exception is for the lag 1 and lag 2 cross-correlation coefficients for maximum temperature lagged with minimum temperature (Figure 6.30).

The cross-correlation coefficients when the sequences are conditioned on the wet or dry status of the day are generally well preserved by the model, with the model performing better with the dry sequences. A possible reason for this has already been discussed in the previous sections. The cross-correlation coefficients for maximum temperature and the other variables (except minimum humidity) appear to show the more marked difference between the simulated and historical sequences. (Figures 6.31 - 6.32.)

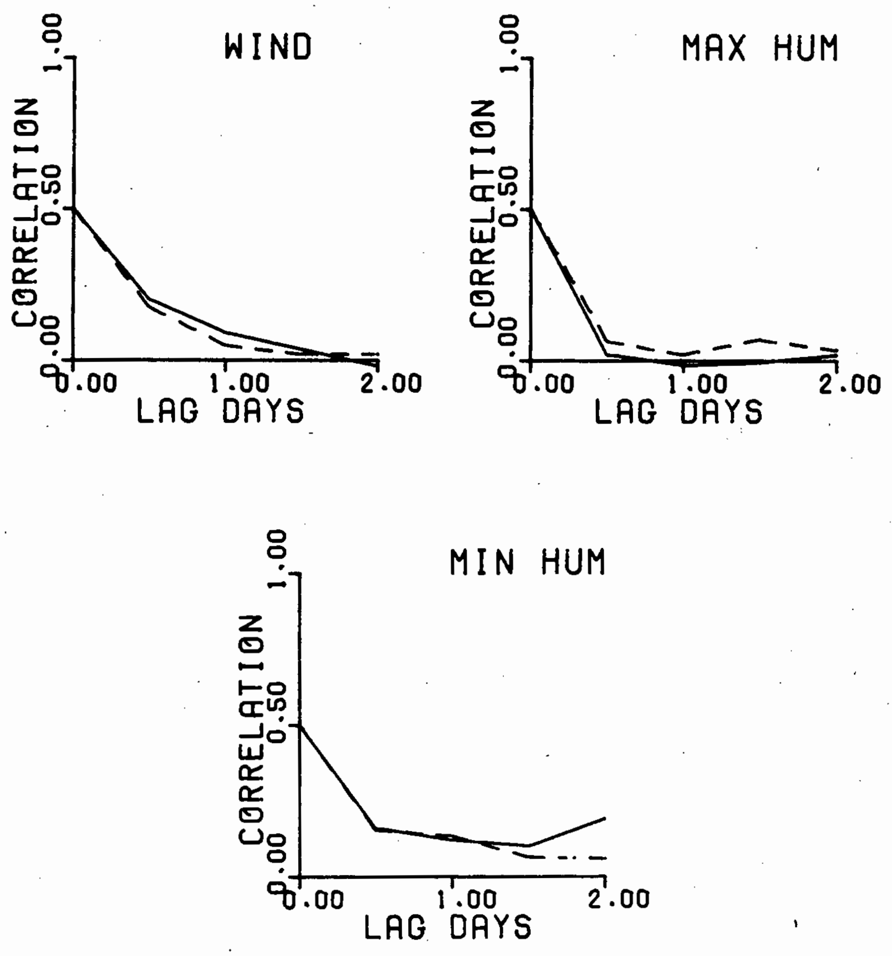
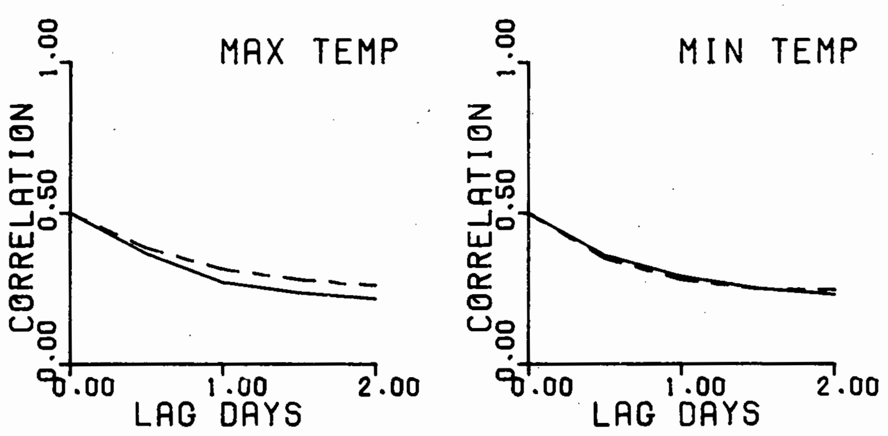


FIGURE 6.29: HISTORICAL (——) AND SIMULATED (---) AUTOCORRELATION COEFFICIENTS GIVEN A DRY DAY FOR ALL VARIABLES



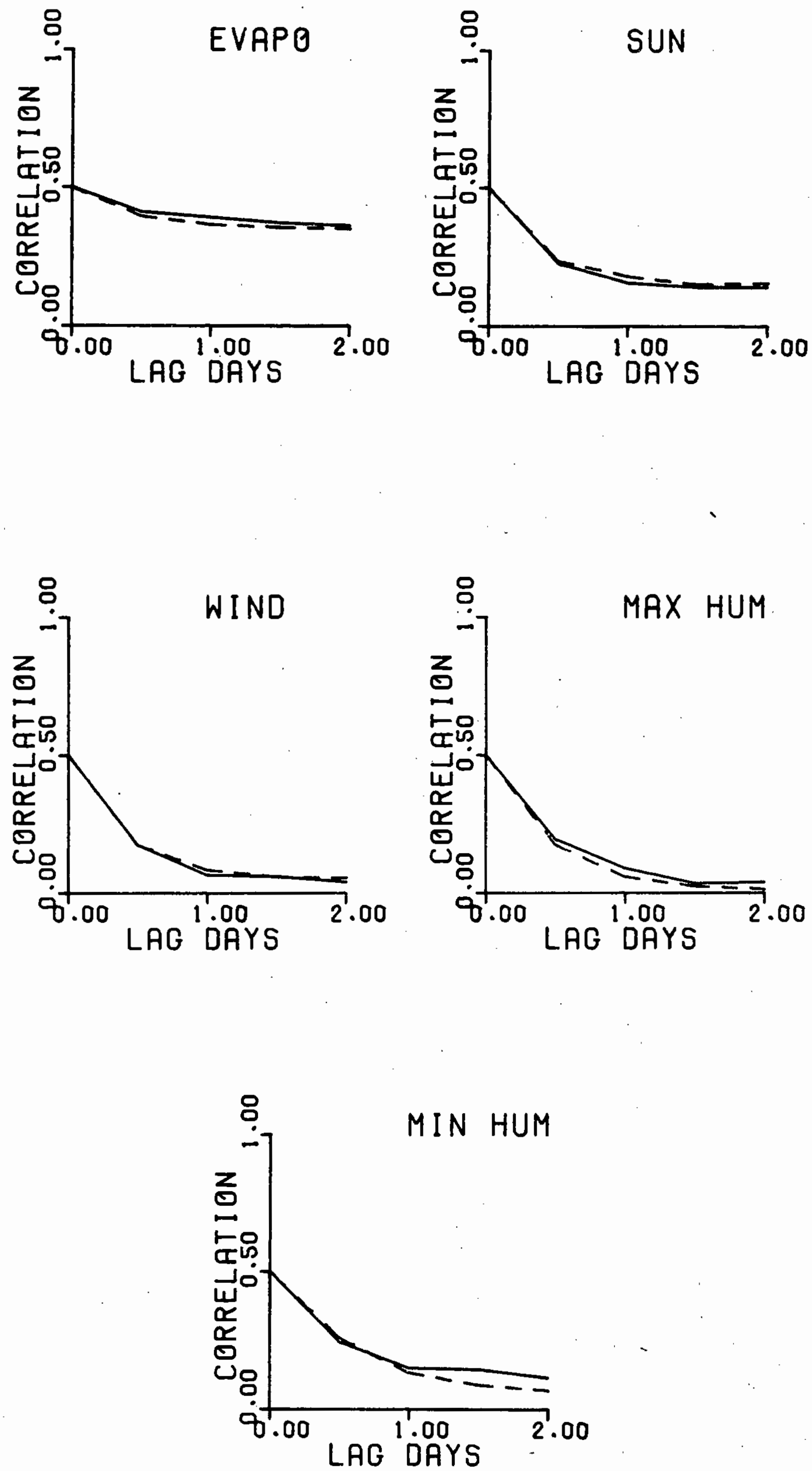
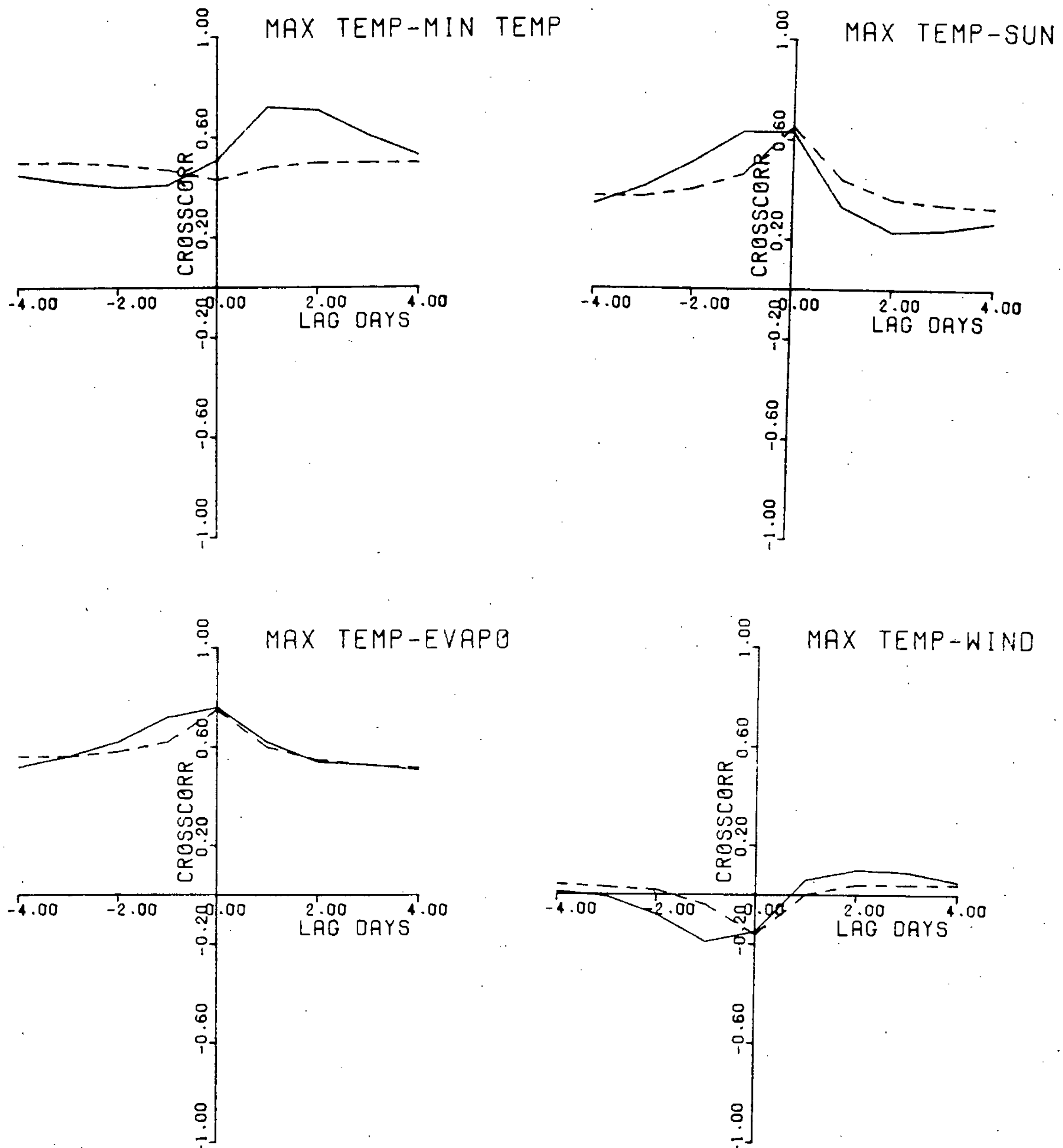
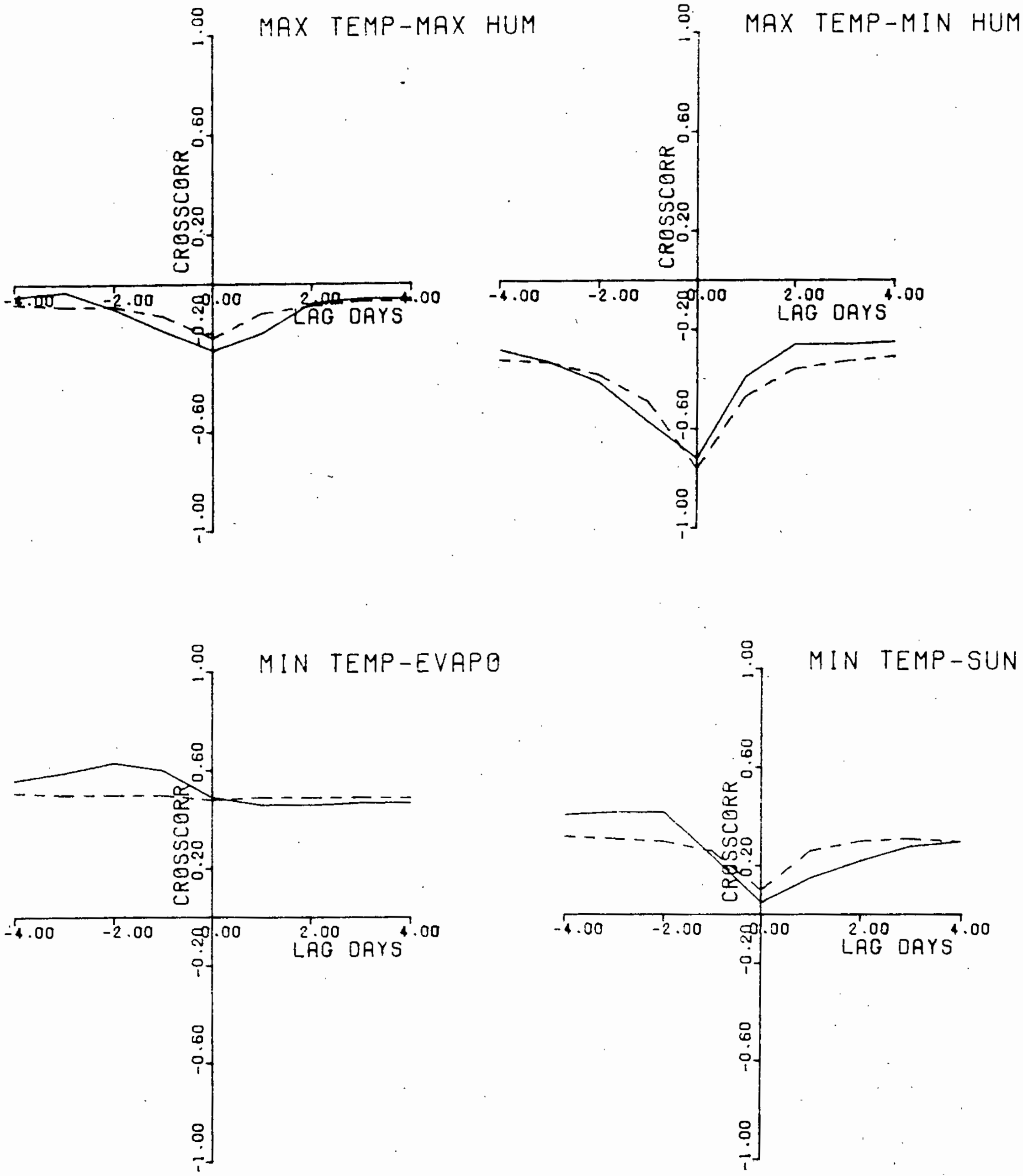
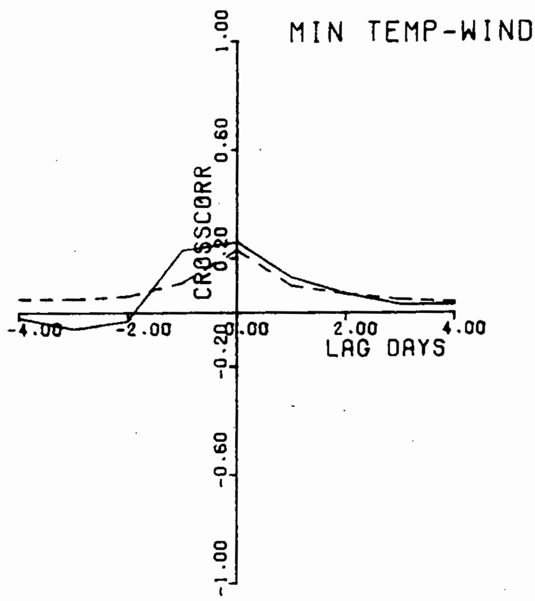
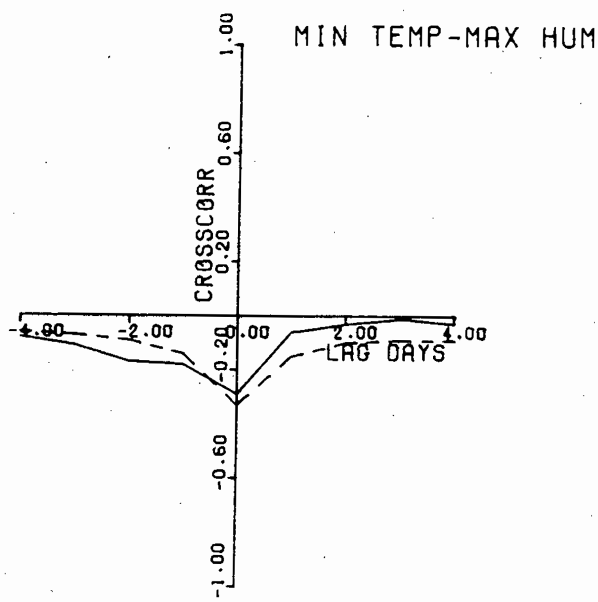

EVAPO
SUN
WIND
MAX HUM
MIN HUM
CORRELATION
0.00
0.50
1.00
LAG DAYS
0.00
1.00
2.00

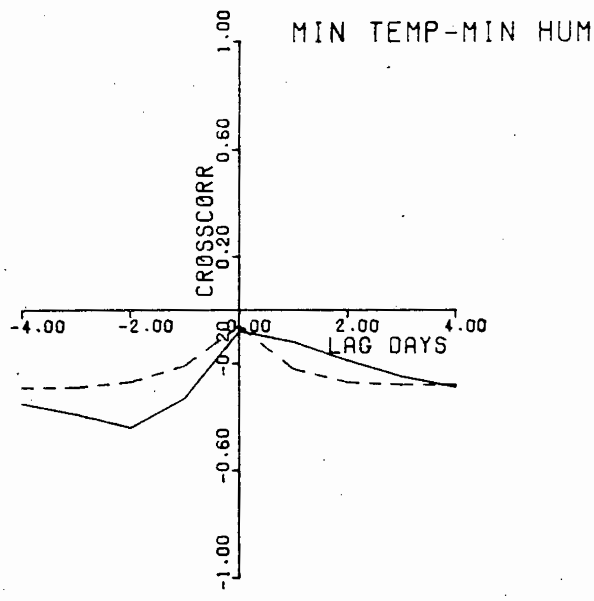
FIGURE 6.30: HISTORICAL (——) AND SIMULATED (---) CROSS-CORRELATION COEFFICIENTS

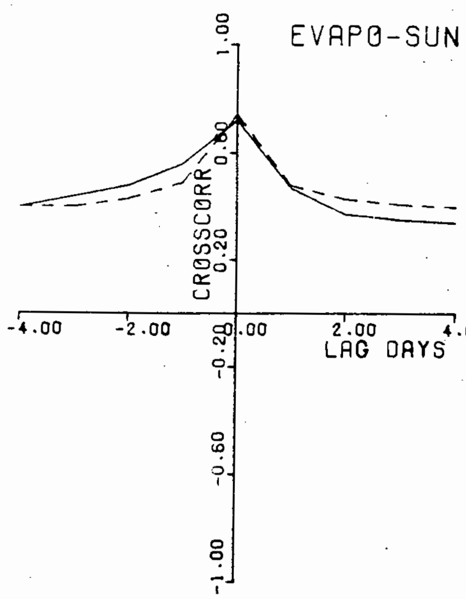


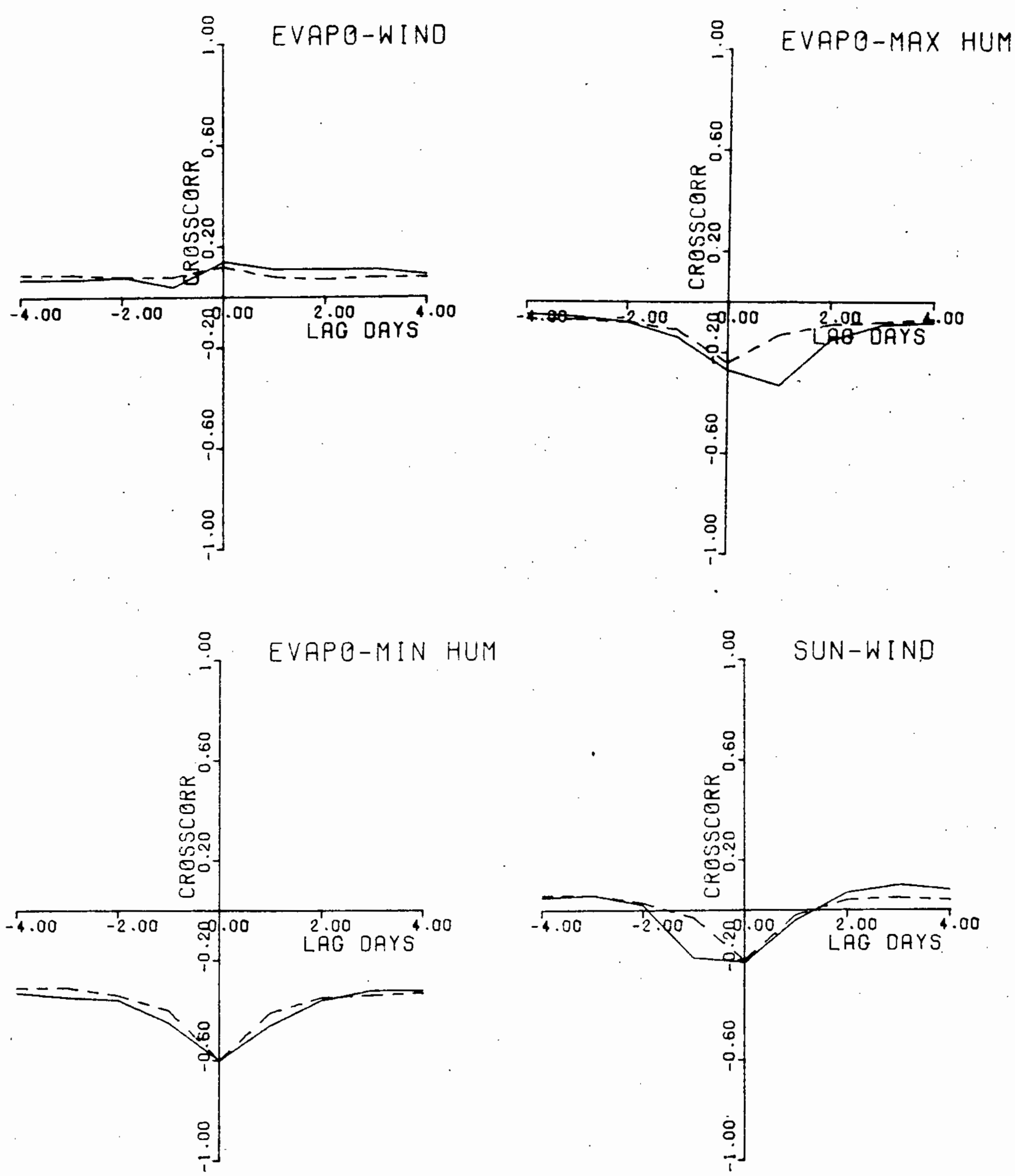

MAX TEMP-MAX HUM
CROSSCORR
LAG DAYS
MAX TEMP-MIN HUM
CROSSCORR
LAG DAYS
MIN TEMP-EVAPO
CROSSCORR
LAG DAYS
MIN TEMP-SUN
CROSSCORR
LAG DAYS

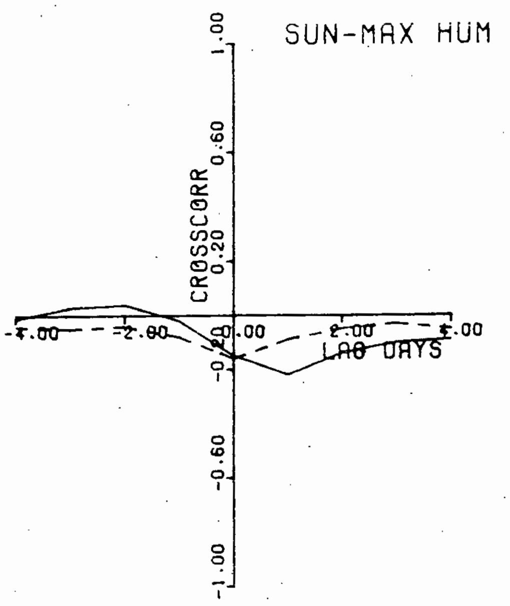

MIN TEMP-WIND
CROSSCORR
LAG DAYS

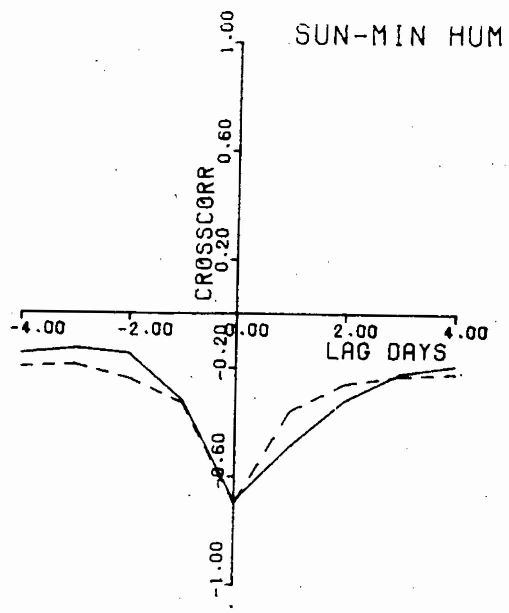

MIN TEMP-MAX HUM
CROSSCORR
LAG DAYS

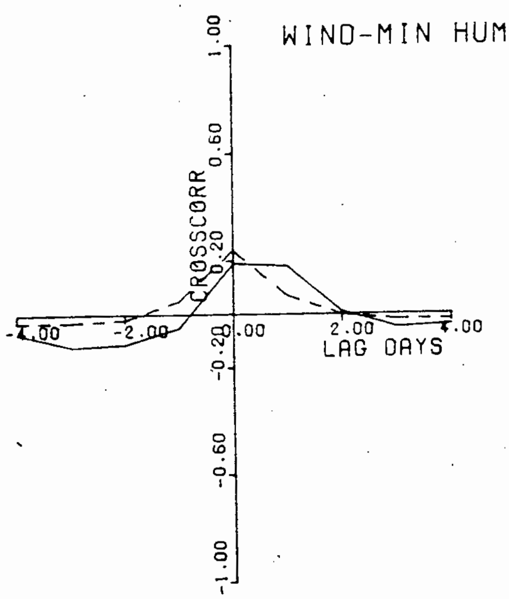

MIN TEMP-MIN HUM
CROSSCORR
LAG DAYS

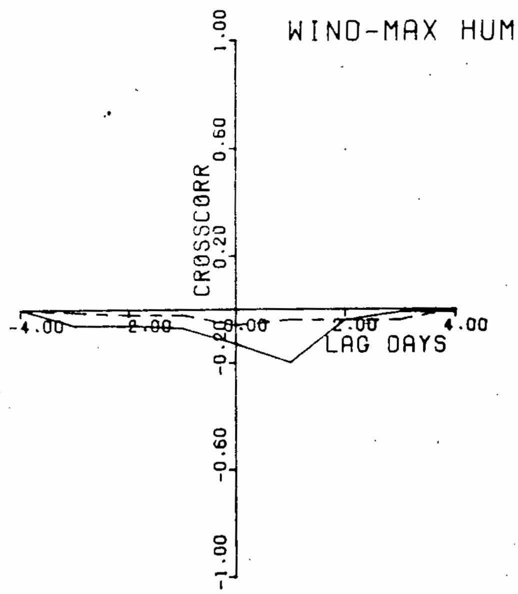

EVAPO-SUN
CROSSCORR
LAG DAYS


EVAPO-WIND
EVAPO-MAX HUM
EVAPO-MIN HUM
SUN-WIND
CROSSCORR
LAG DAYS
-4.00
-2.00
0.00
2.00
4.00
-1.00
-0.60
-0.20
0.20
0.60
1.00


SUN-MAX HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS


SUN-MIN HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS


WIND-MIN HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS


WIND-MAX HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

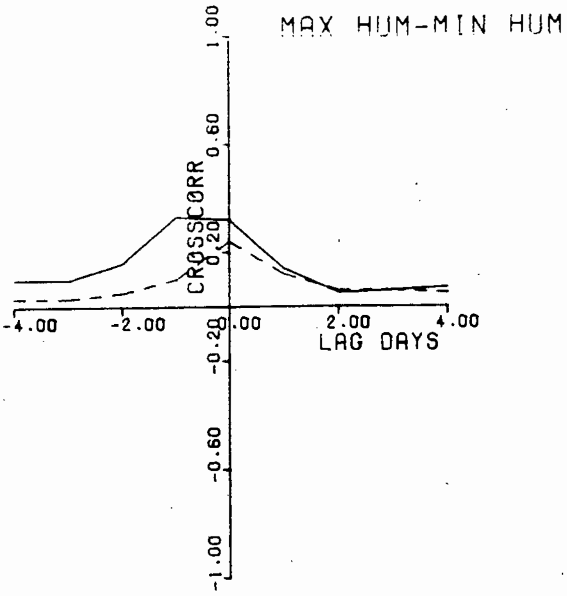
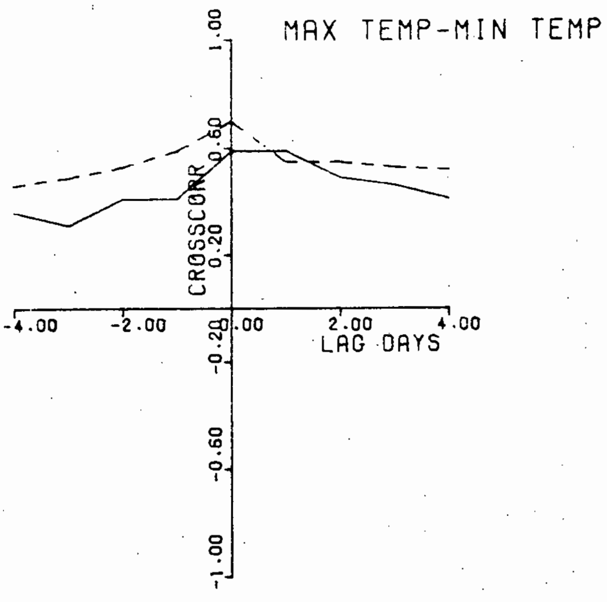
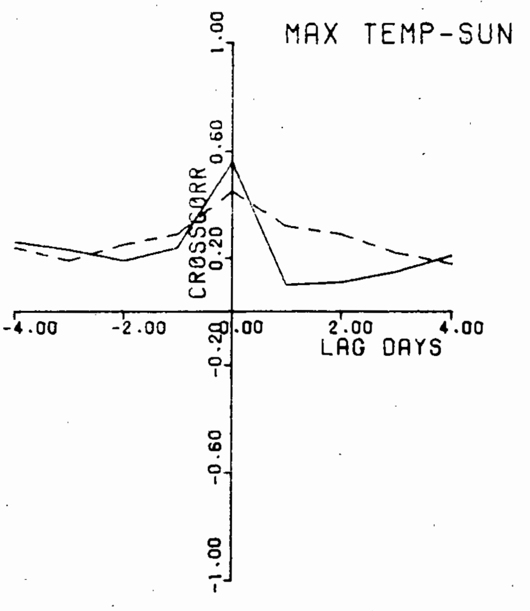
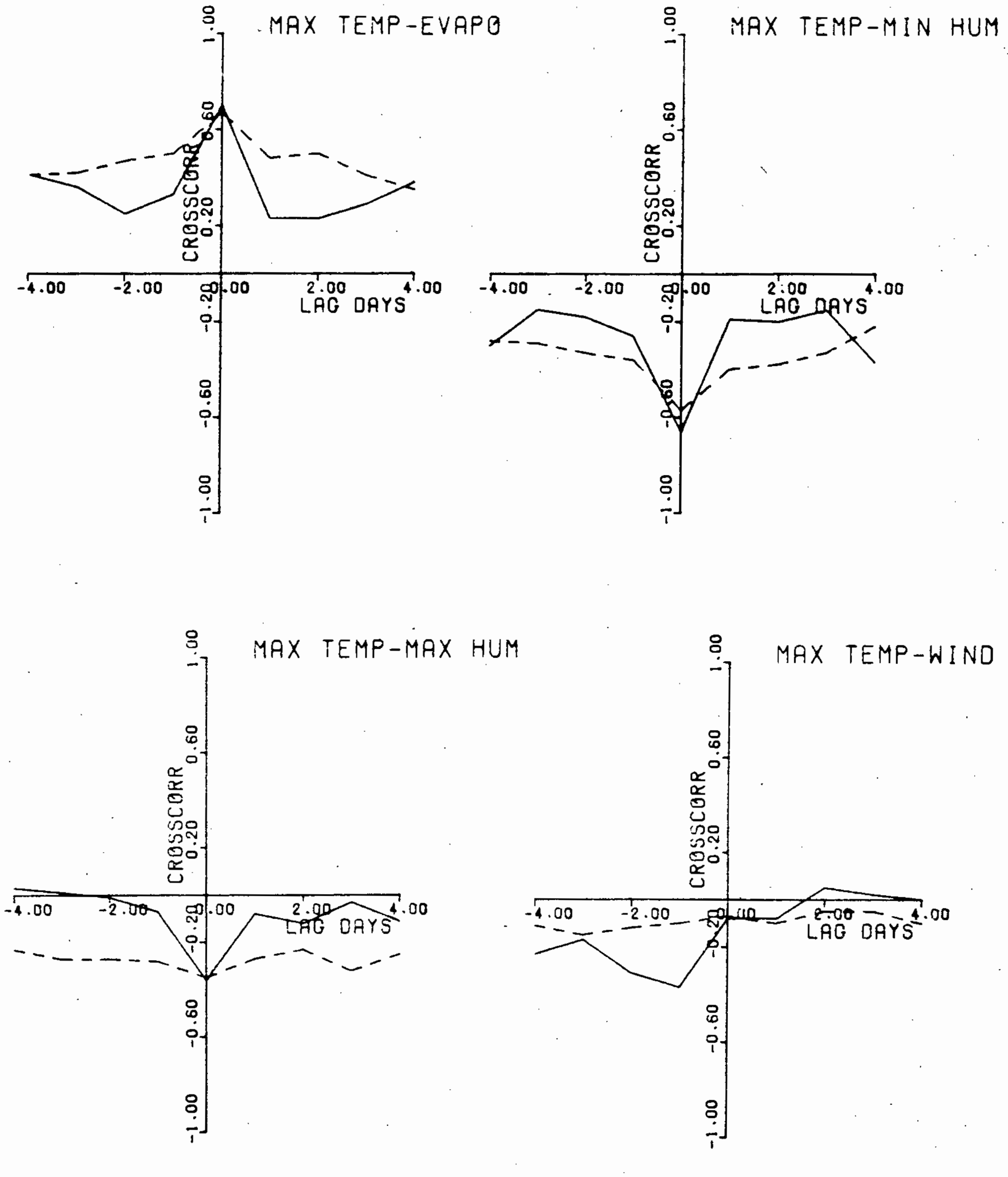
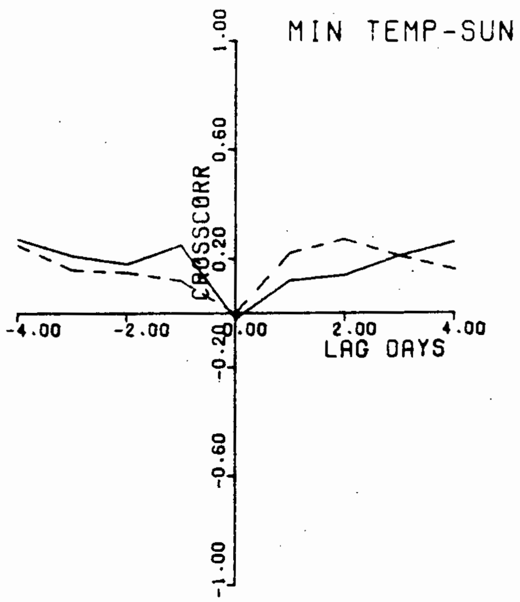


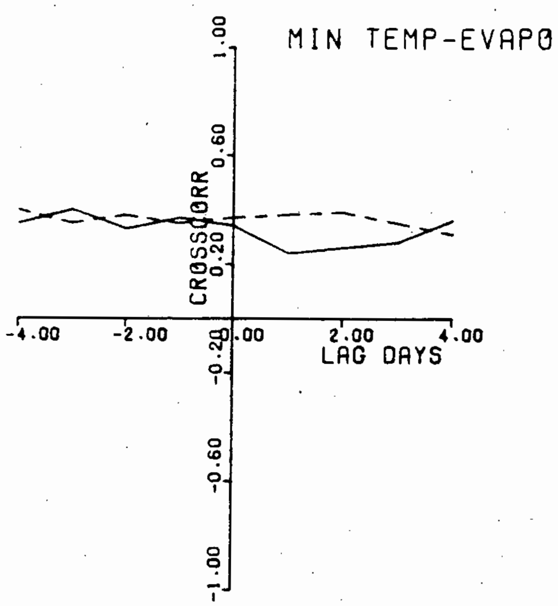
FIGURE 6.31: HISTORICAL (——) AND SIMULATED (---) CROSS-CORRELATION COEFFICIENTS GIVEN A WET DAY

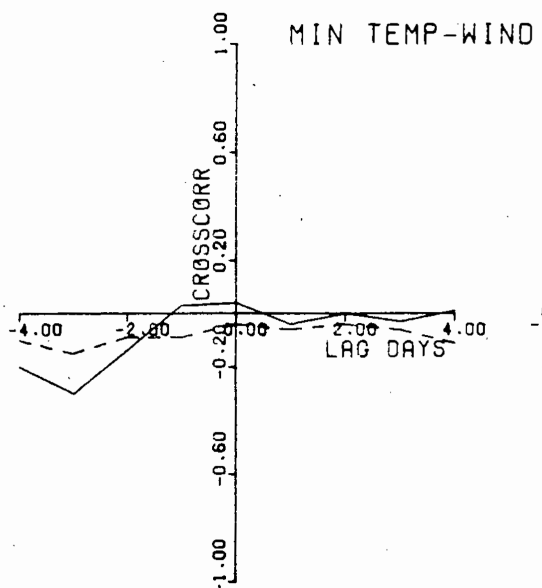


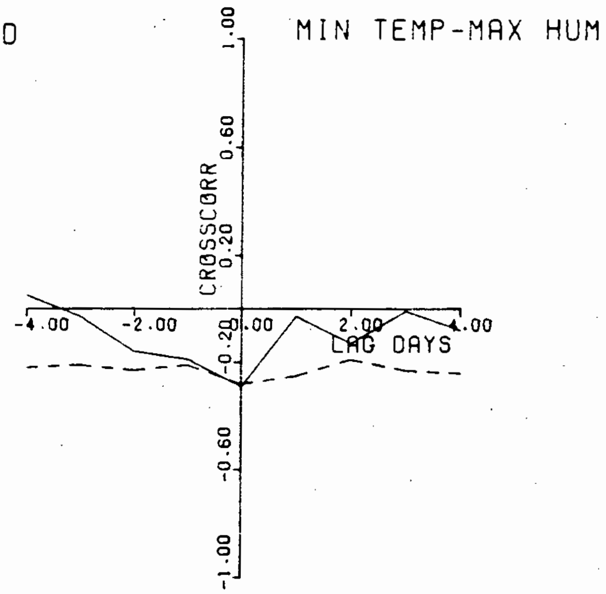


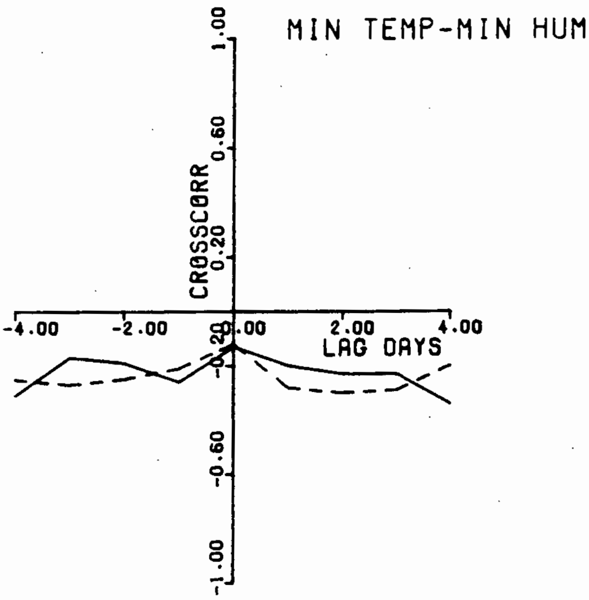

MAX TEMP-EVAPO
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS
MAX TEMP-MIN HUM
MAX TEMP-MAX HUM
MAX TEMP-WIND

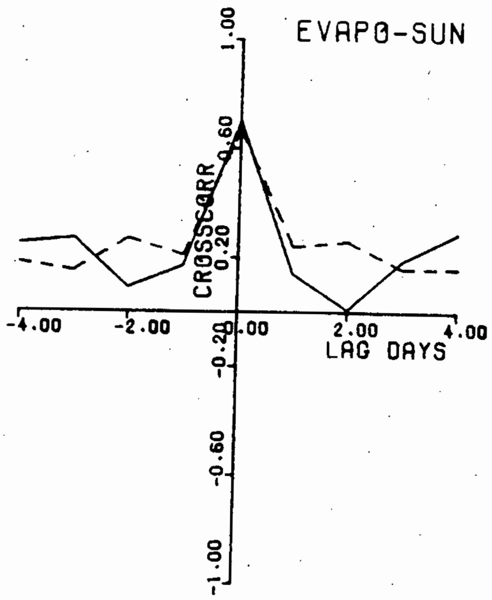

MIN TEMP-SUN
CROSSCORR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00

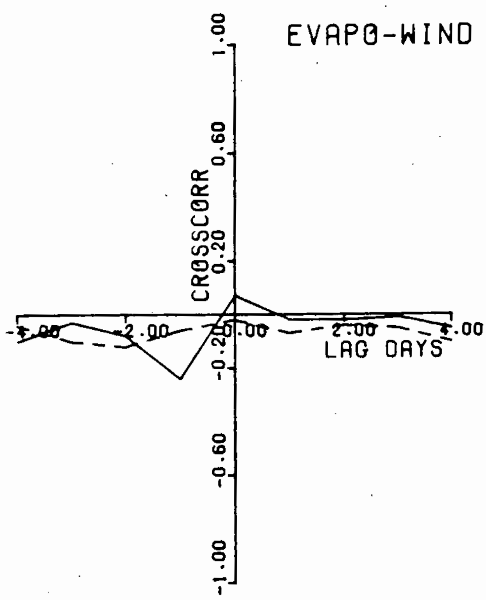

MIN TEMP-EVAPO
CROSSCORR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00

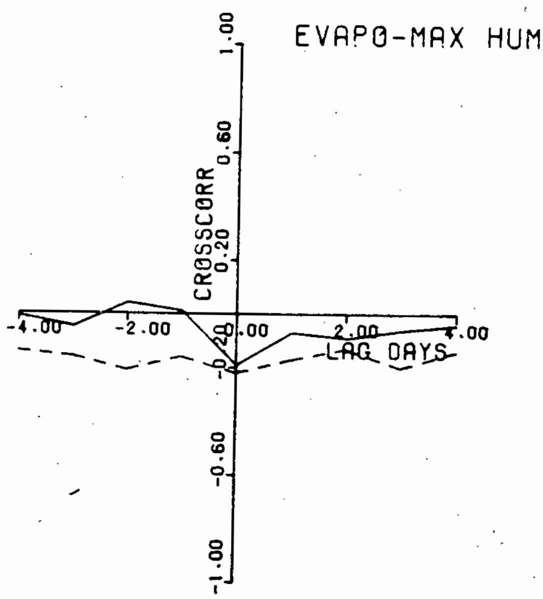

MIN TEMP-WIND
CROSSCORR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00

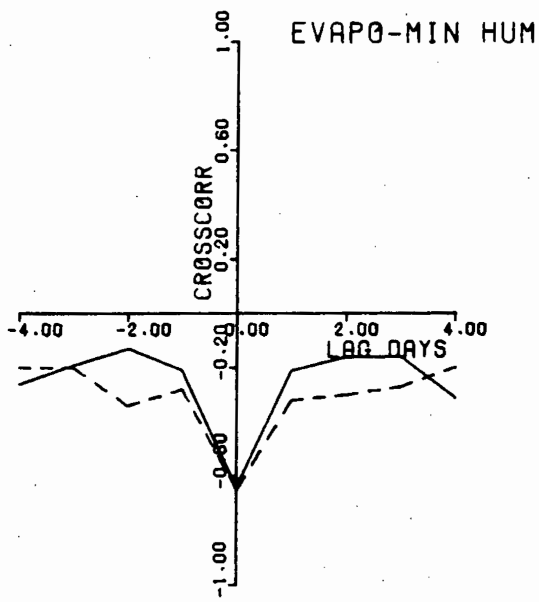

MIN TEMP-MAX HUM
CROSSCORR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00

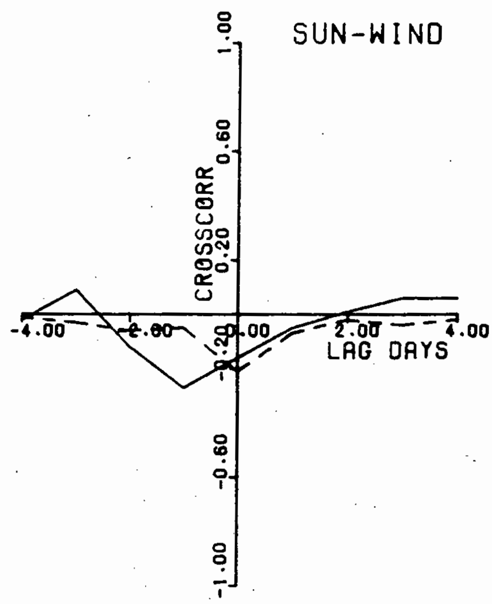

MIN TEMP-MIN HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

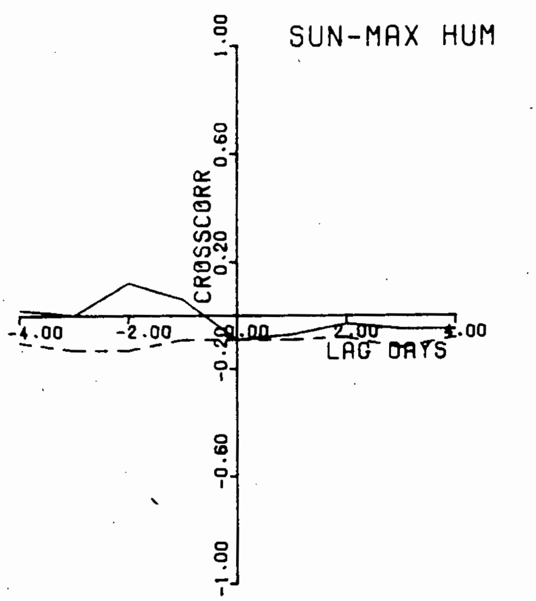

EVAPO-SUN
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

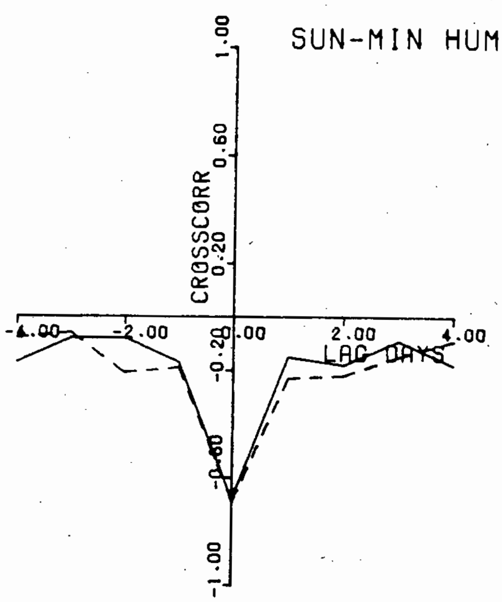

EVAPO-WIND
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-2.00
0.00
2.00
LAG DAYS

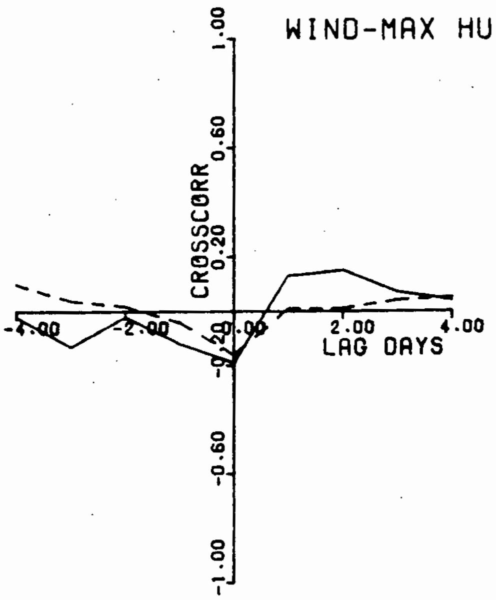

EVAPO-MAX HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
LAG DAYS

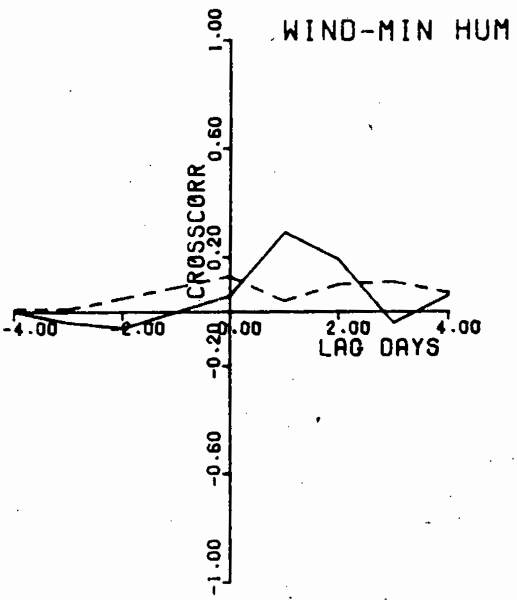

EVAPO-MIN HUM
CROSSCORR
LAG DAYS


SUN-WIND
CROSSCORR
LAG DAYS


SUN-MAX HUM
CROSSCORR
LAG DAYS


SUN-MIN HUM
CROSSCORR
LAG DAYS


WIND-MAX HUM
CRØSSCØRR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS


WIND-MIN HUM
CRØSSCØRR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

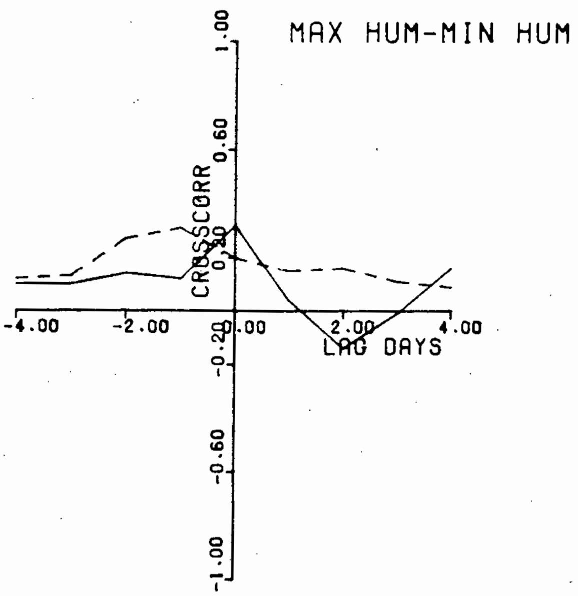
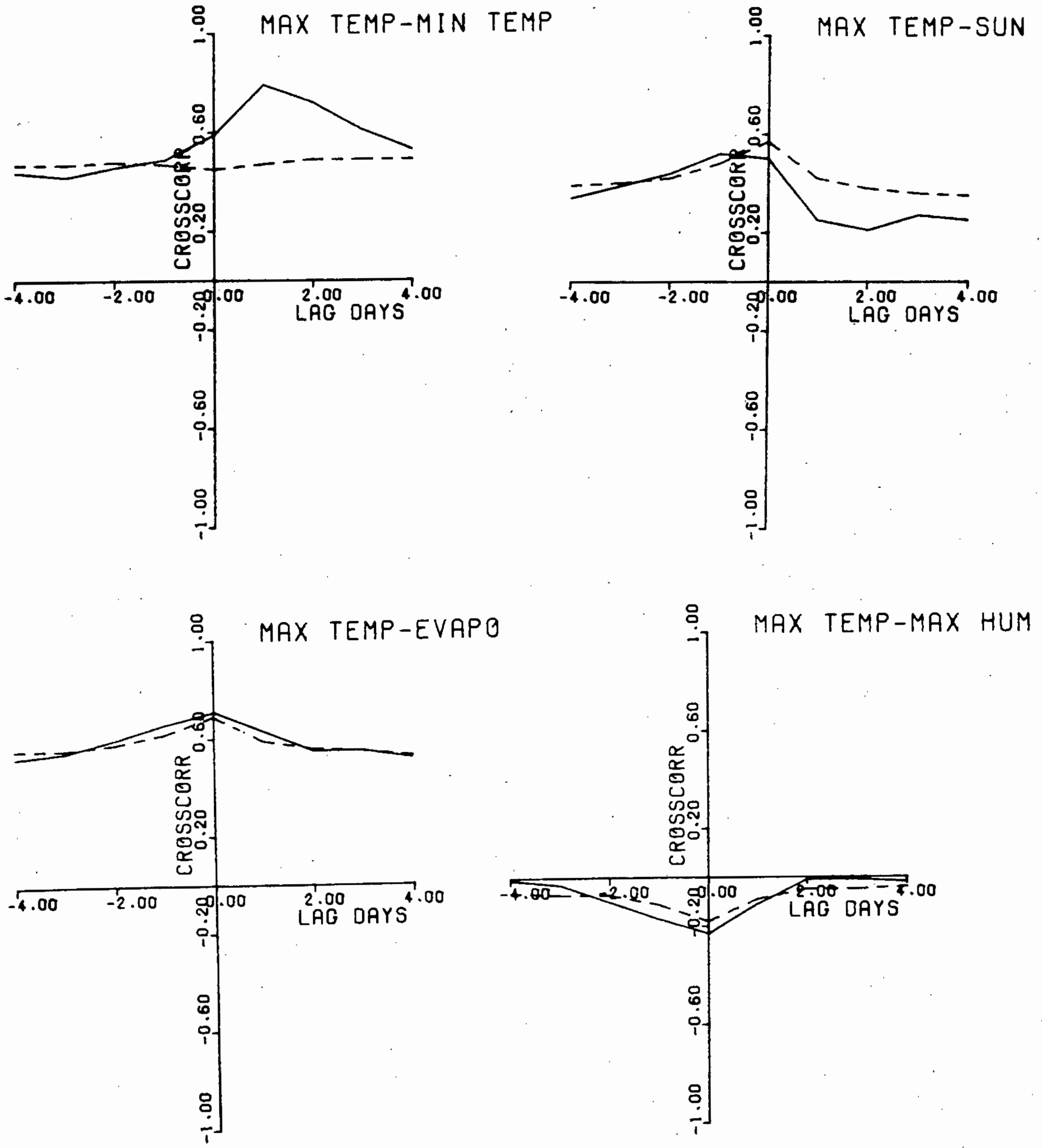
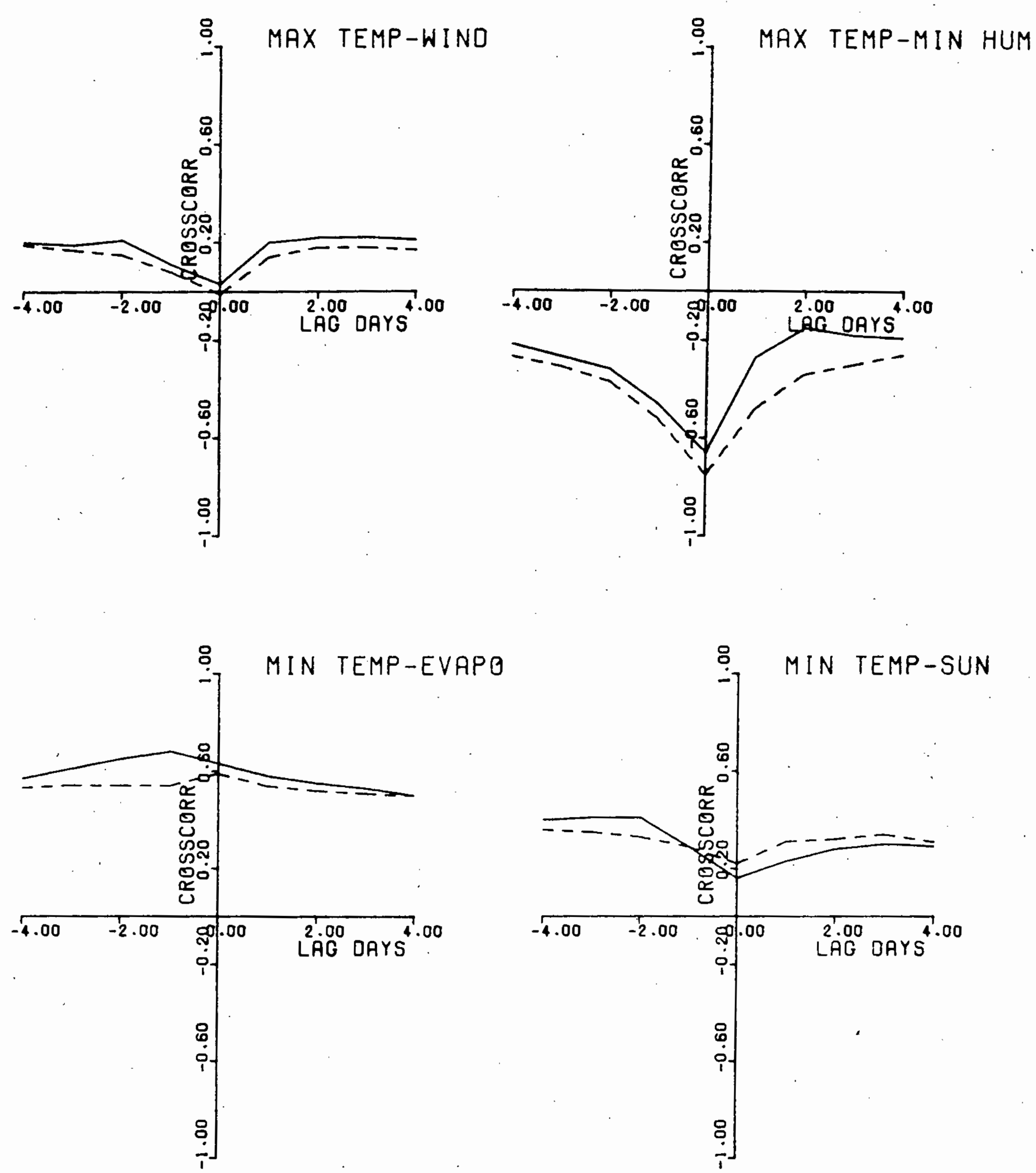
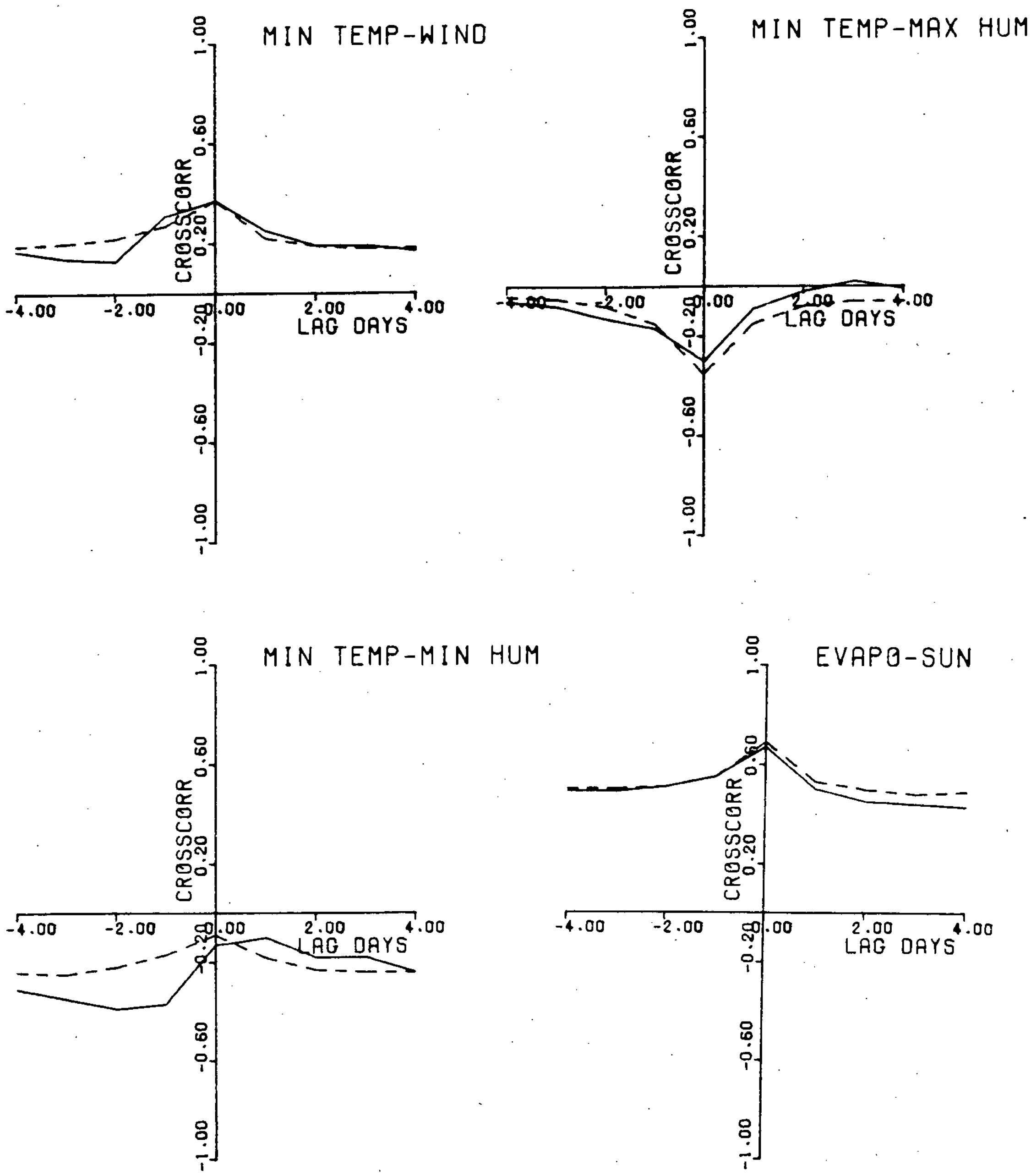
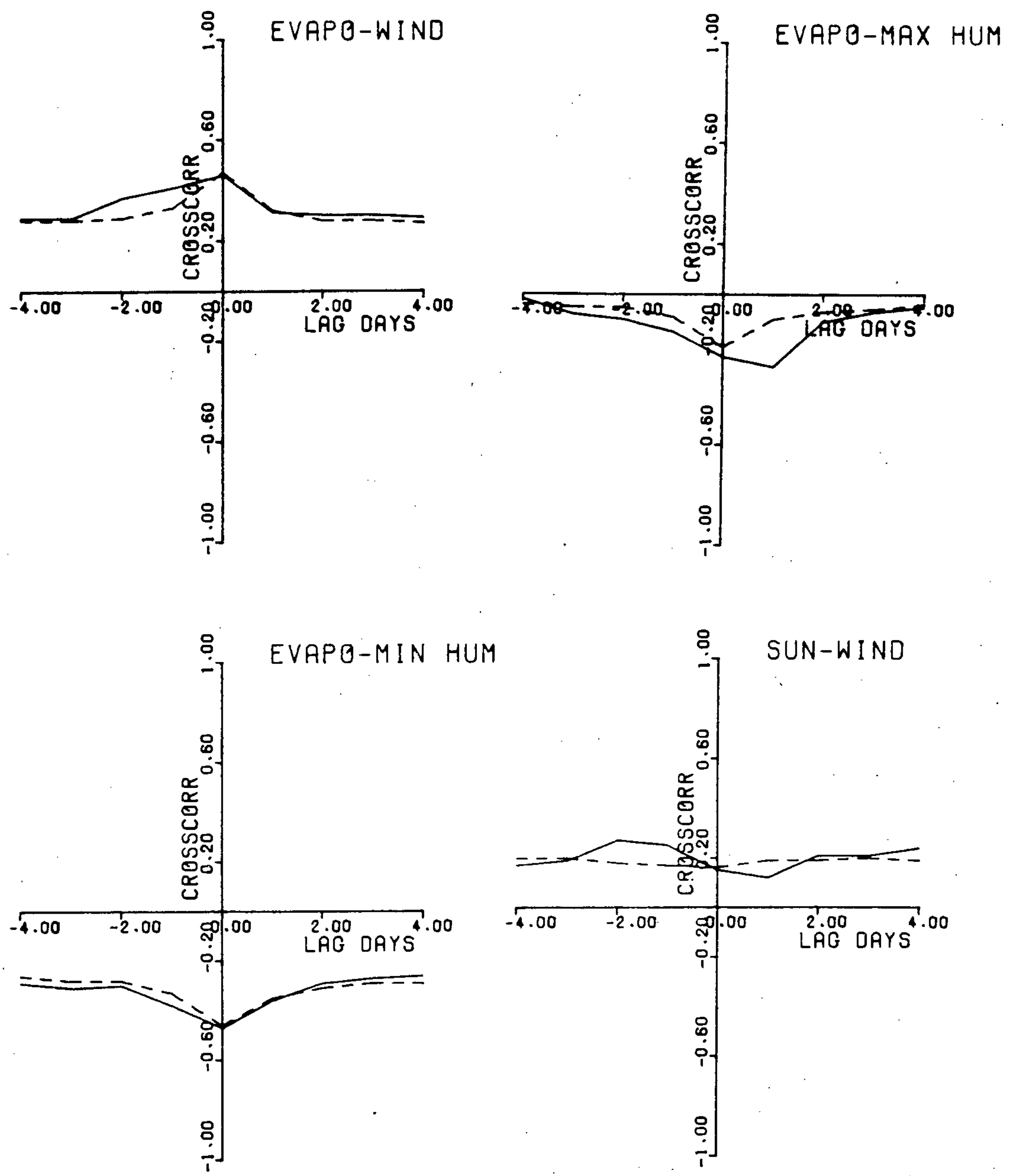

MAX HUM-MIN HUM
CRØSSCØRR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

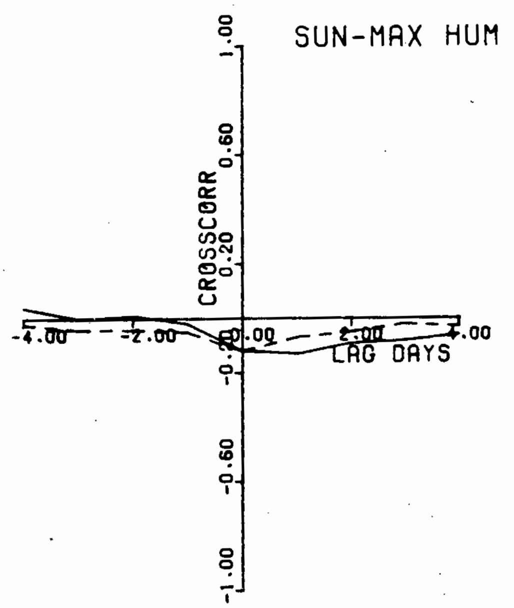
FIGURE 6.32: HISTORICAL (——) AND SIMULATED (---) CROSS-CORRELATION COEFFICIENTS GIVEN A DRY DAY

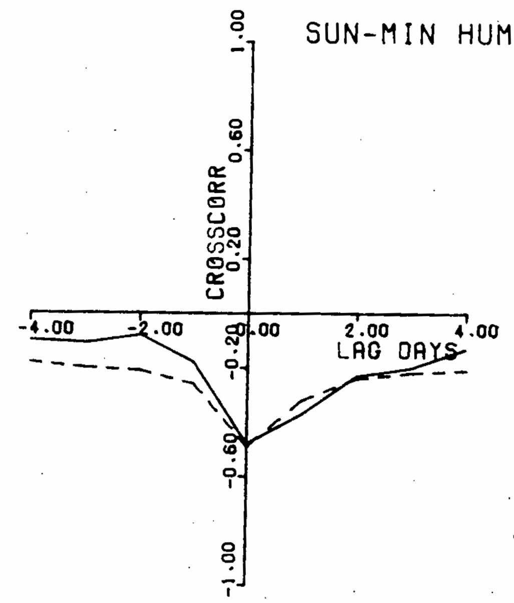


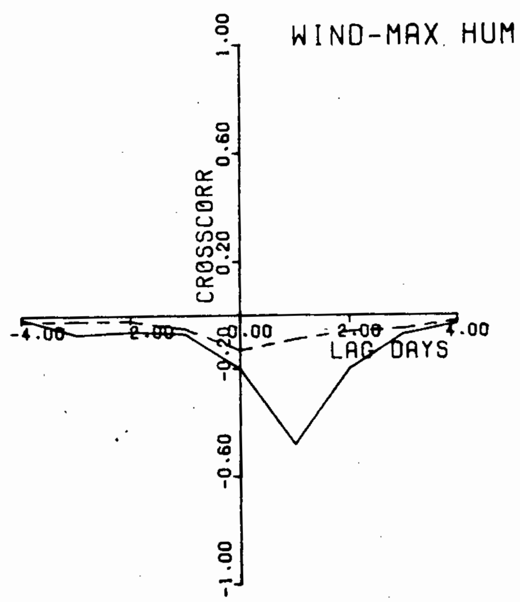

MAX TEMP-WIND
MAX TEMP-MIN HUM
MIN TEMP-EVAPO
MIN TEMP-SUN
CROSSCORR
LAG DAYS
-1.00
-0.60
-0.20
0.20
0.60
1.00
-4.00
-2.00
0.00
2.00
4.00

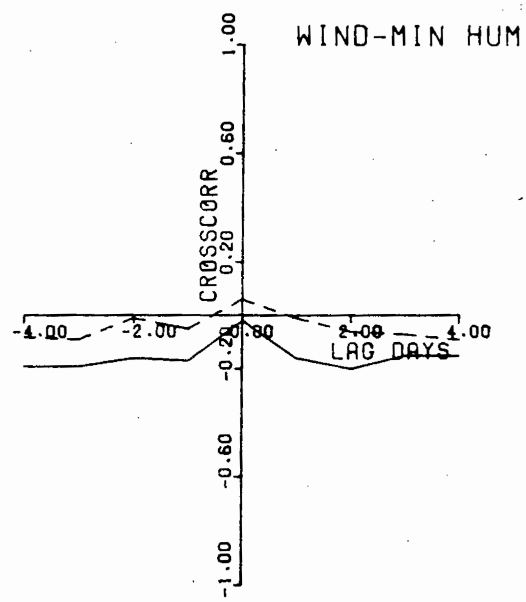

MIN TEMP-WIND
MIN TEMP-MAX HUM
MIN TEMP-MIN HUM
EVAPØ-SUN
CRØSSCØRR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00

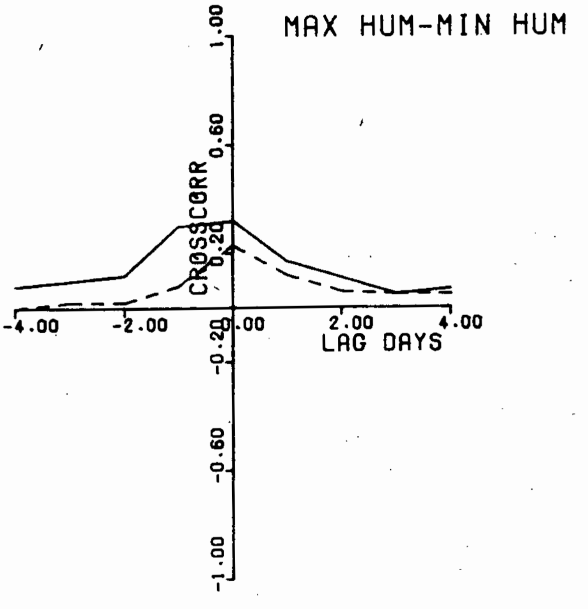

EVAPO-WIND
EVAPO-MAX HUM
EVAPO-MIN HUM
SUN-WIND
CROSSCORR
LAG DAYS


SUN-MAX HUM
CROSSCORR
LAG DAYS


SUN-MIN HUM
CROSSCORR
LAG DAYS


WIND-MAX HUM
CROSSCORR
LAG DAYS


WIND-MIN HUM
CROSSCORR
LAG DAYS


MAX HUM-MIN HUM
1.00
0.60
0.20
-0.20
-0.60
-1.00
CROSSCORR
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

6.2.3 Validation of Climate Model 3 (ML).

The parameters for this model were estimated using the method of maximum likelihood estimation.

(a) Validation of annual properties.

The annual mean and standard deviation for each variable obtained from the historical and simulated sequences are compared in Table 6.17. It is clear that the simulated sequence closely resembles the historical sequence with respect to the annual mean and standard deviation.

The annual mean and standard deviation for the simulated and historical sequences were also compared when the sequences are conditioned on the status of the day. Again, these characteristics of the climate sequence was satisfactorily described by the model. These results are illustrated in Tables 6.18 - 6.19.

TABLE 6.17: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MEANS AND STANDARD DEVIATIONS

Variable	Historical		Simulated	
	Mean	Std. Dev.	Mean	Std. Dev.
Max Temp	226.5	58.90	189.4	65.30
Min Temp	106.1	38.34	107.2	37.82
Evapo	57.30	37.01	56.72	36.69
Sunshine	82.61	36.74	82.32	36.98
Windrun	1966	911.2	1946	876.9
Max Hum	914.5	96.29	940.4	99.53
Min Hum	412.7	152.8	420.5	151.0

TABLE 6.18: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MEAN AND STANDARD DEVIATION GIVEN A WET DAY

Variable	Historical		Simulated	
	Mean	Std. Dev.	Mean	Std. Dev.
Max Temp	182.1	46.73	166.8	45.91
Min Temp	110.1	33.43	109.7	33.75
Evapo	28.49	26.99	28.30	26.71
Sunshine	41.71	33.75	40.72	33.93
Windrun	2500	1194	2383	1143
Max Hum	9330	72.90	952.2	73.91
Min Hum	548.9	157.7	543.7	160.3

TABLE 6.19: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MEAN AND STANDARD DEVIATION GIVEN A DRY DAY

Variable	Historical		Simulated	
	Mean	Std. Dev.	Mean	Std. Dev.
Max Temp	241.4	54.95	196.7	68.83
Min Temp	104.8	39.77	106.4	39.01
Evapo	66.48	35.00	65.82	34.74
Sunshine	96.16	26.12	95.65	26.55
Windrun	1788	710.4	1806	718.7
Max Hum	908.3	102.3	936.7	106.2
Min Hum	366.7	120.3	381.1	124.3

The extreme values of the simulated and historical sequences were compared to check that this characteristic of the climate sequence is described by the model. This comparison is given in Table 6.20

The maximum values are retained by the model for most variables (exceptions are sunshine duration and maximum humidity), however the minimum values are not preserved by the model. The number of times a simulated value is greater than the maximum historical value or less than the minimum value was checked. Table 6.21 contains the results obtained. These results show that there is a large number of values that lie below the minimum value of the historical record observed for maximum temperature, otherwise few or none of the simulated values lie over or below the observed extreme values.

TABLE 6.20: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MAXIMUM AND MINIMUM VALUES

Variable	Historical		Simulated	
	Maximum	Minimum	Maximum	Minimum
Max Temp	408.0	100.0	471.1	-40.11
Min Temp	238.0	13.00	234.6	-18.10
Evapo	185.0	0.00	160.6	-47.14
Sunshine	133.0	0.00	192.5	-76.23
Windrun	7376	293.0	6224	-1806
Max Hum	1000	280.0	1315	562.8
Min Hum	950.0	120.0	1105	-125.4

TABLE 6.21: NUMBER OF TIMES SIMULATED VALUES LIE OUTSIDE THE HISTORICAL MAXIMUM OR MINIMUM VALUES

Variable	Greater than Maximum	Less than Minimum
Max Temp	0 (0%)	1278 (18%)
Min Temp	0 (0%)	22 (0.3%)
Evapo	0 (0%)	385 (5%)
Sunshine	82 (1%)	185 (3%)
Windrun	0 (0%)	160 (2%)
Max Hum	527 (7%)	0 (0%)
Min Hum	2 (0%)	53 (0.7%)

Tables 6.22 - 6.23 show the comparison of historical and simulated maximum and minimum values given the status of the day.

The maximum values of the simulated sequence compare favourably with those of the historical record for most of the variables given a dry day occurs. For a wet sequence, except for maximum humidity, the maximum values are preserved by the model. The same is not true for the minimum values. Except for minimum temperature on a wet day the minimum values are not preserved by the model.

TABLE 6.22: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MAXIMUM MINIMUM VALUES GIVEN A WET DAY

Variable	Historical		Simulated	
	Maximum	Minimum	Maximum	Minimum
Max Temp	365.0	100.0	373.3	-19.40
Min Temp	238.0	23.00	221.6	20.10
Evapo	122.0	0.00	116.9	-47.14
Sunshine	124.0	0.00	136.8	-76.27
Windrun	7376	618.0	6224	-1806
Max Hum	1000	340.0	1260	605.0
Min Hum	950.0	170.0	1104	-125.4

TABLE 6.23: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MAXIMUM AND MINIMUM VALUES GIVEN A DRY DAY

Variable	Historical		Simulated	
	Maximum	Minimum	Maximum	Minimum
Max Temp	408.0	120.0	471.1	-40.10
Min Temp	234.0	13.00	234.6	-18.10
Evapo	185.0	0.00	160.6	-34.15
Sunshine	133.0	0.00	192.5	-5.48
Windrun	6804	293.0	4314	-1327
Max Hum	1000	280.0	1315	562.8
Min Hum	850.0	120.0	826.9	-106.1

(b) Validation of monthly properties.

The monthly means and standard deviations for each variable obtained from the simulated sequence are checked against those obtained from the historical record. This is also performed when the sequences are conditioned on the wet or dry status of the day. These comparisons are given in Figures 6.33 - 6.38.

The monthly means, whether the sequence is conditioned or not and regardless of the state of the day, are underestimated by the model for the variable maximum temperature and is overestimated for the variable maximum humidity. Otherwise the monthly mean structure of the historical sequence has been very well preserved by the model.

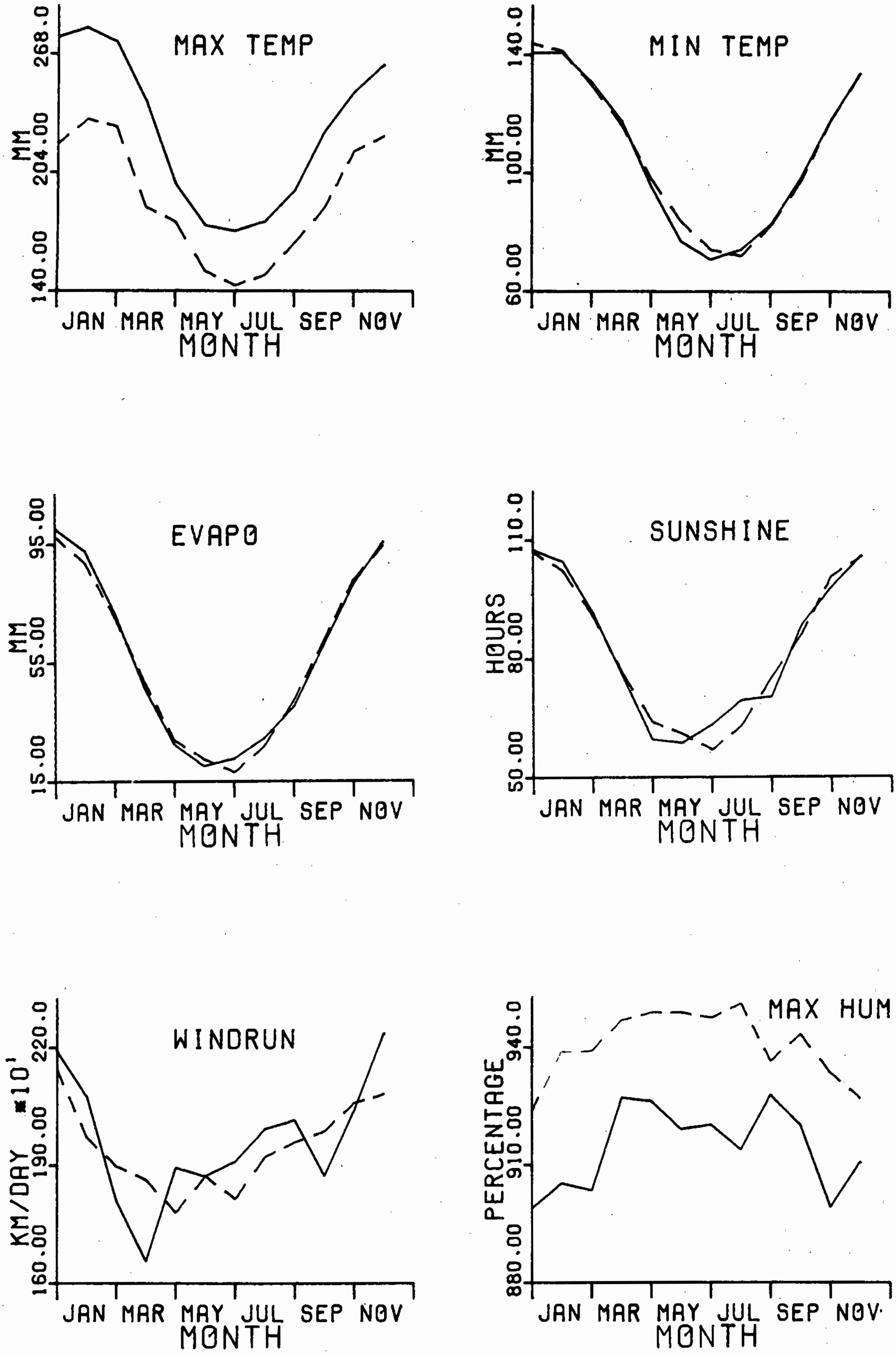
The model performs well with respect to the property of the monthly standard deviations for most variables. The variables maximum temperature, maximum humidity and windrun show slight differences between the standard deviations obtained from the simulated and historical sequences.

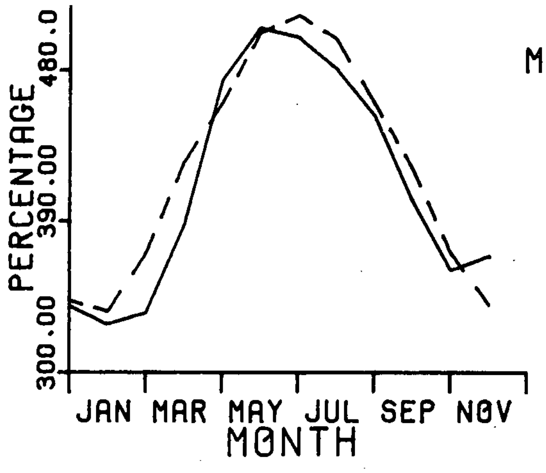
(c) Validation of daily properties.

A good fit by the truncated Fourier approximation is obtained for the mean function, given a dry day, of all climate variables except for maximum temperature, where the mean function is underestimated (Figure 6.40). Given a wet day, the fit is excellent (Figure 6.39).

The simulated sequence was examined to determine whether the model was successful in reproducing the property of autocorrelation that is

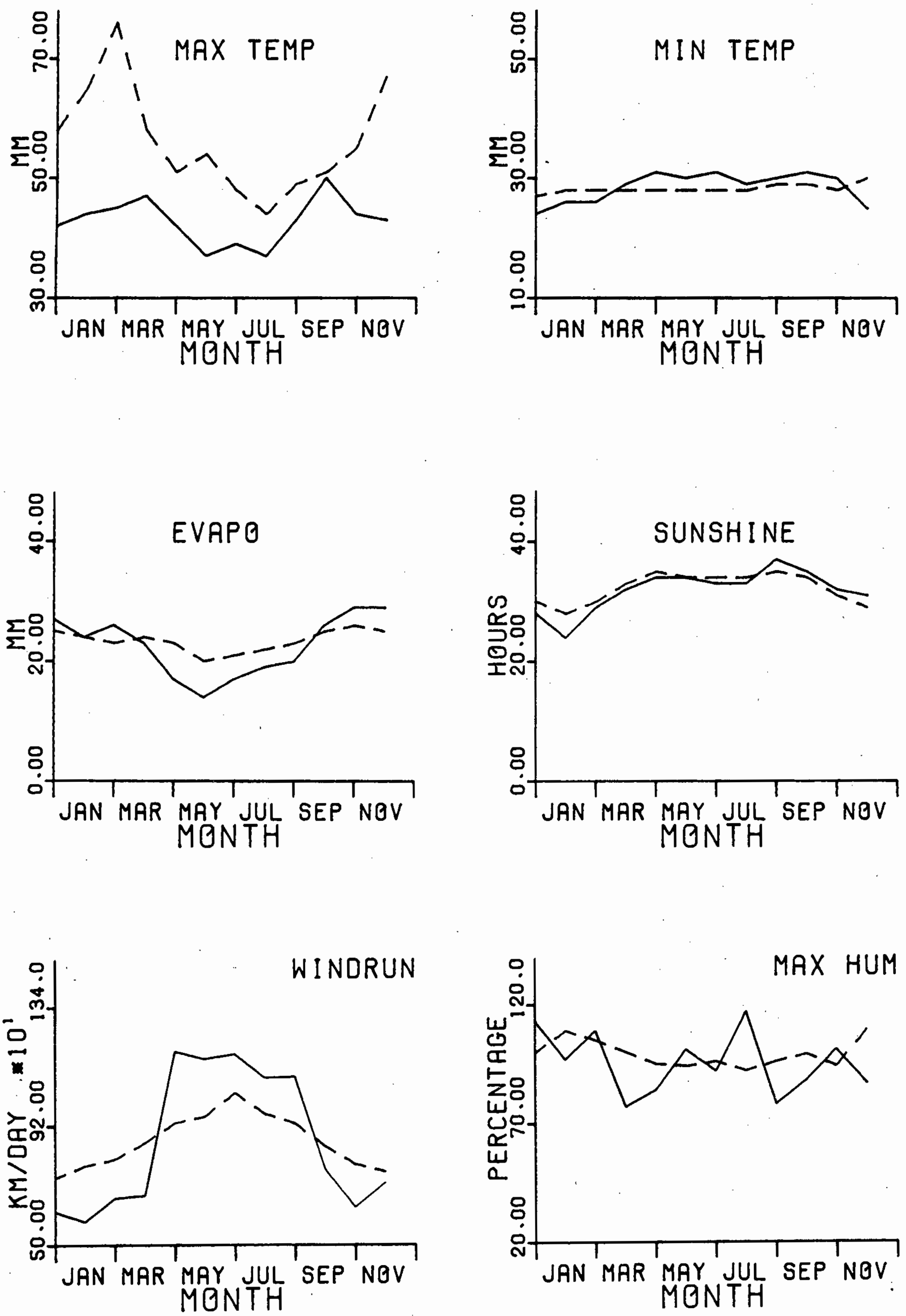
FIGURE 6.33: HISTORICAL (——) AND SIMULATED (---) MONTHLY MEANS FOR ALL VARIABLES





MIN HUM

FIGURE 6.34: HISTORICAL (——) AND SIMULATED (---) MONTHLY STANDARD DEVIATIONS FOR ALL VARIABLES



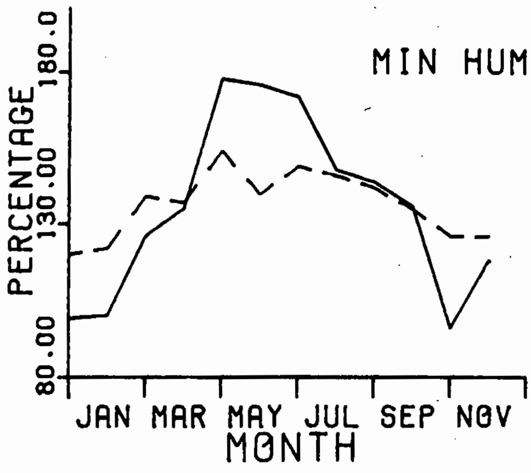
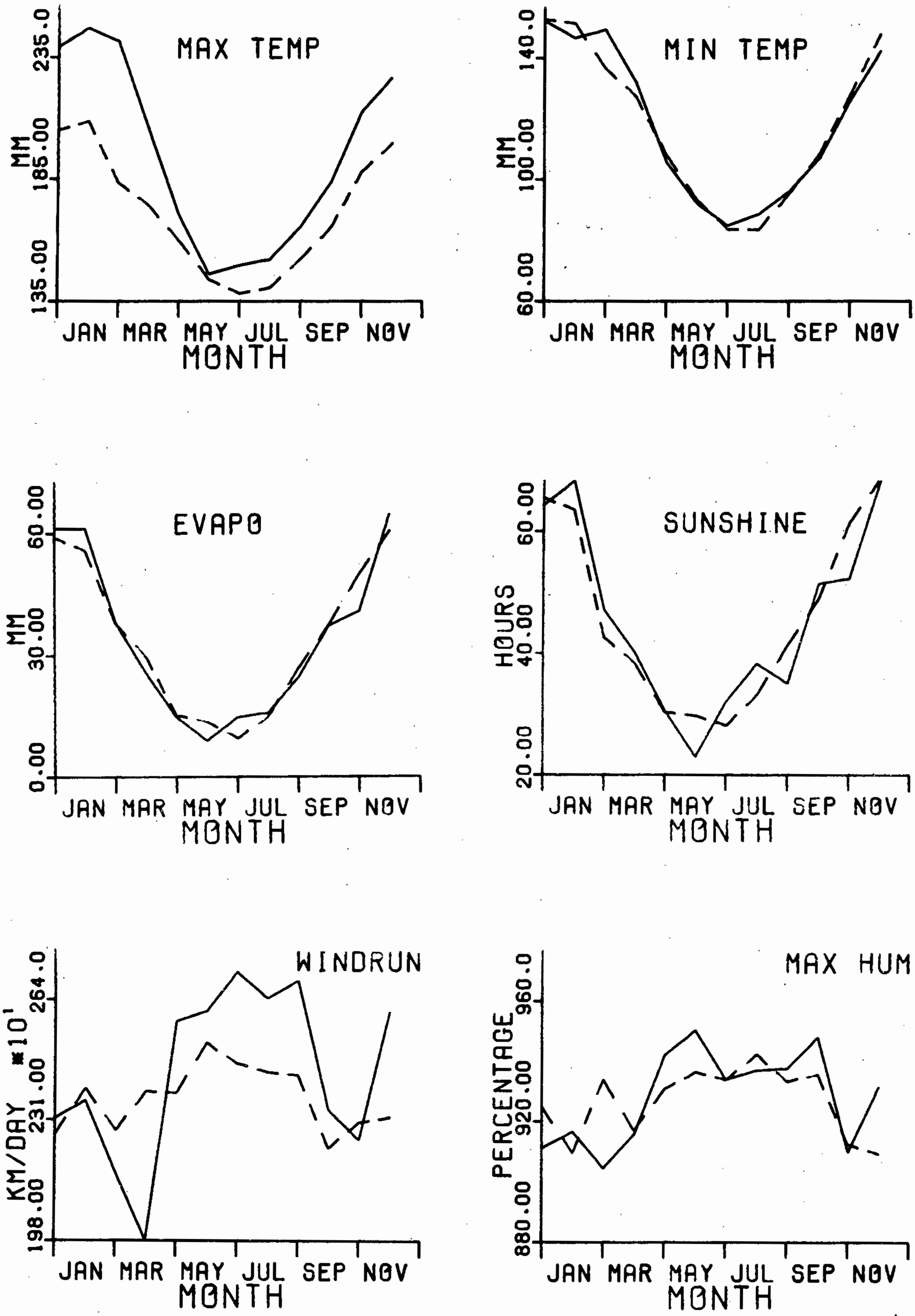

MIN HUM
PERCENTAGE
180.0
130.00
80.00
JAN MAR MAY JUL SEP NOV
MONTH

FIGURE 6.35: HISTORICAL (——) AND SIMULATED (---) MONTHLY MEANS GIVEN A WET DAY FOR ALL VARIABLES



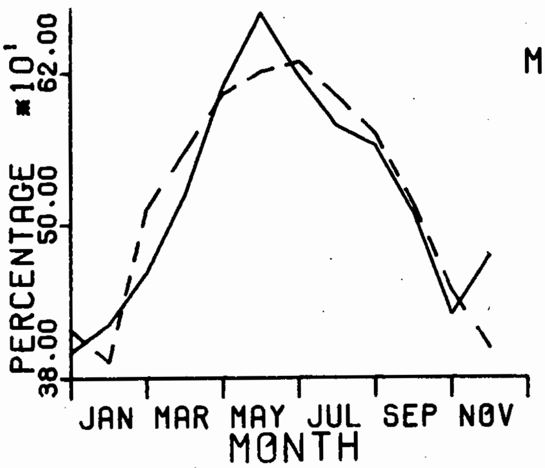
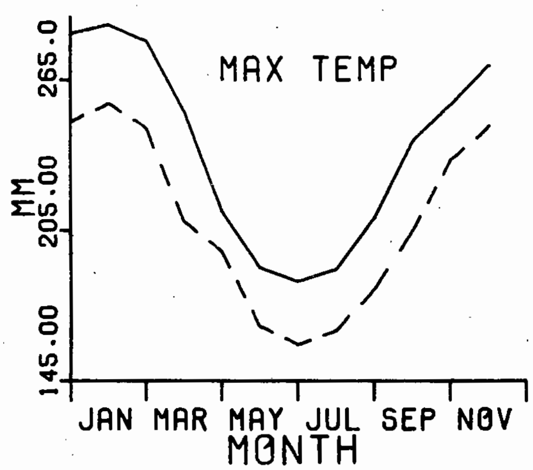
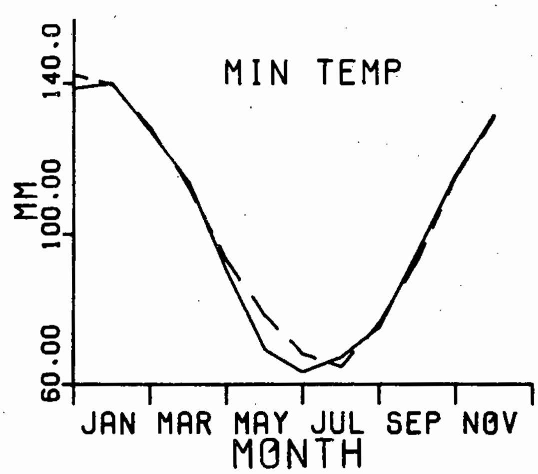


FIGURE 6.36: HISTORICAL (——) AND SIMULATED (---) MONTHLY MEANS GIVEN A DRY DAY FOR ALL VARIABLES





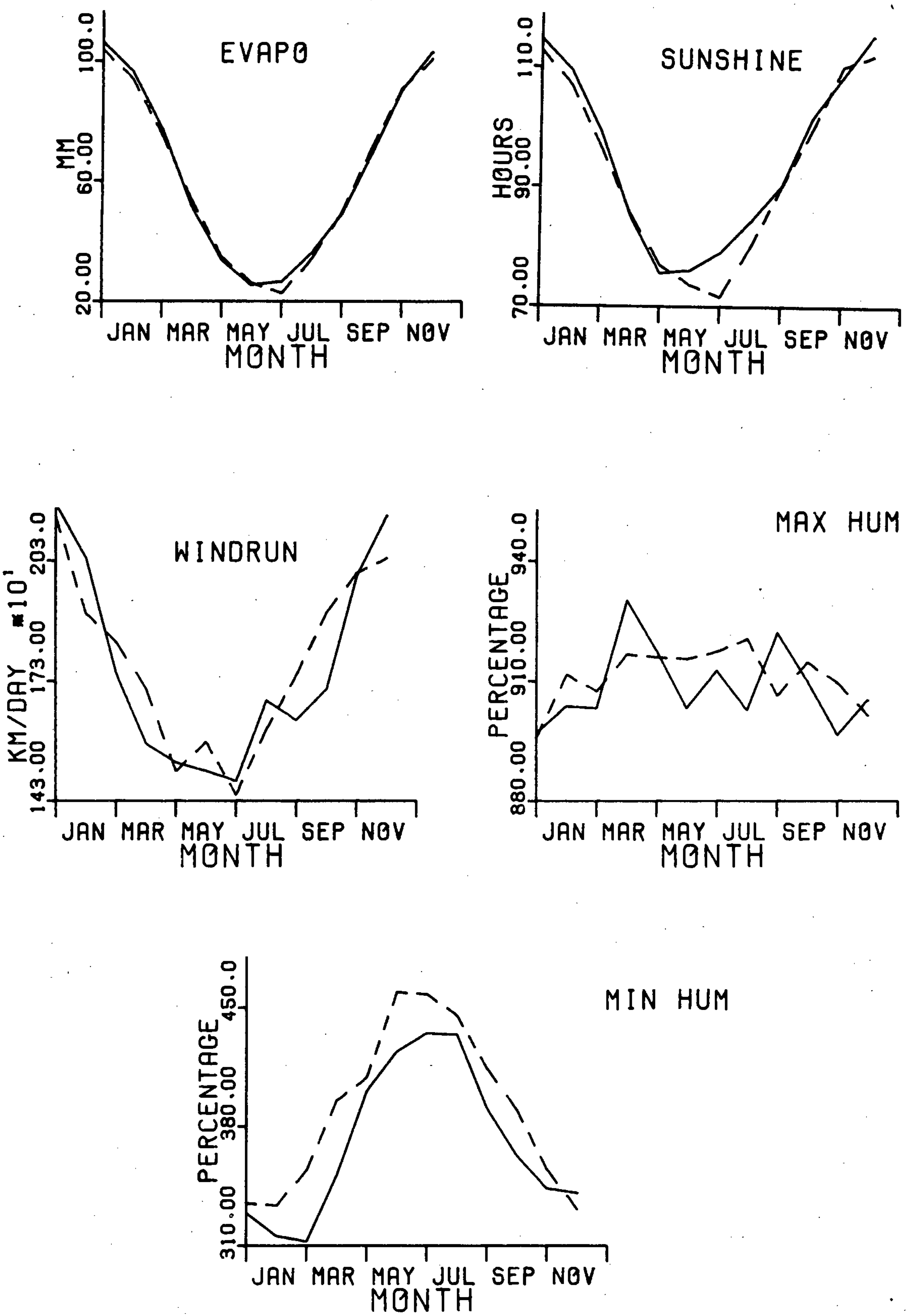
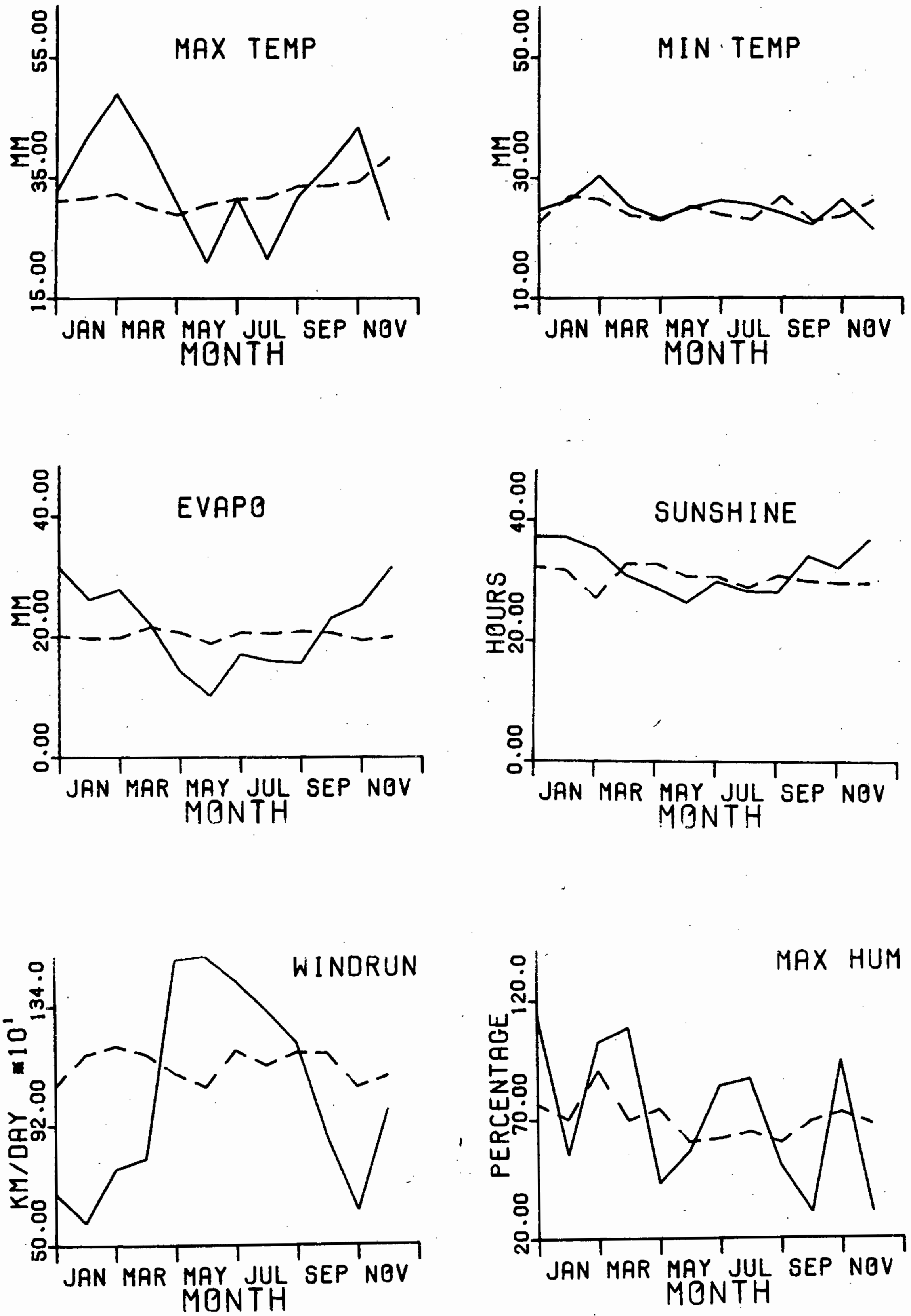

EVAPO
MM
100.0
60.00
20.00
JAN MAR MAY JUL SEP NOV
MONTH
SUNSHINE
HOURS
110.0
90.00
70.00
JAN MAR MAY JUL SEP NOV
MONTH
WINDRUN
KM/DAY *10¹
203.0
173.00
143.00
JAN MAR MAY JUL SEP NOV
MONTH
MAX HUM
PERCENTAGE
940.0
910.00
880.00
JAN MAR MAY JUL SEP NOV
MONTH
MIN HUM
PERCENTAGE
450.0
380.00
310.00
JAN MAR MAY JUL SEP NOV
MONTH

FIGURE 6.37: HISTORICAL (——) AND SIMULATED (---) MONTHLY STANDARD DEVIATIONS GIVEN A WET DAY FOR ALL VARIABLES



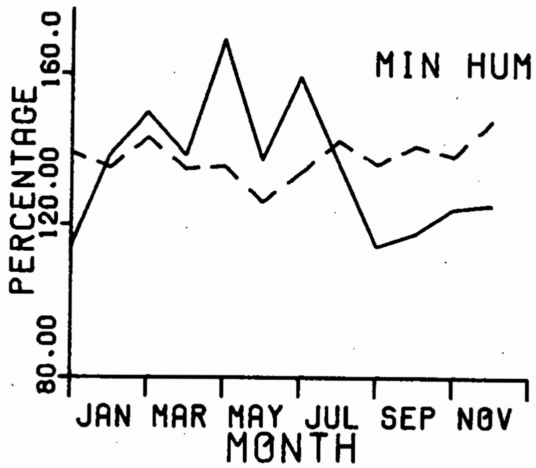
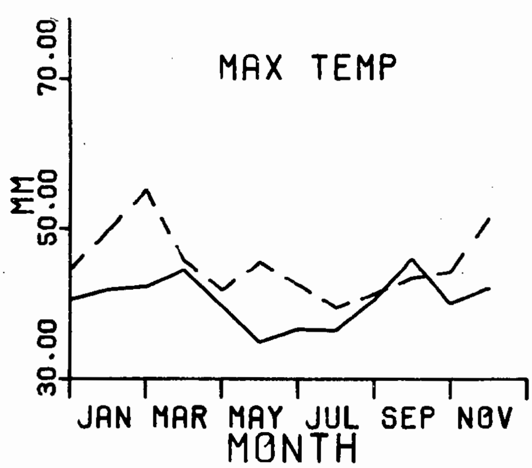
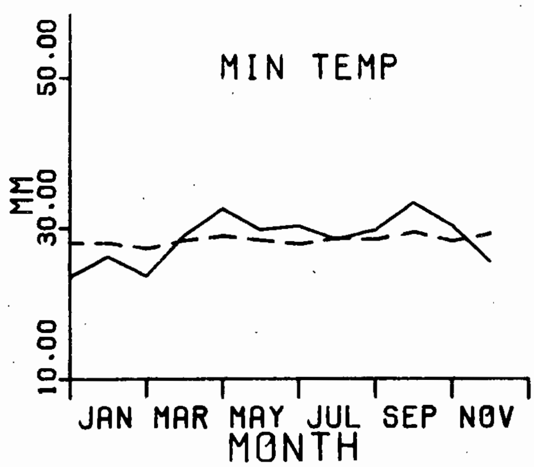
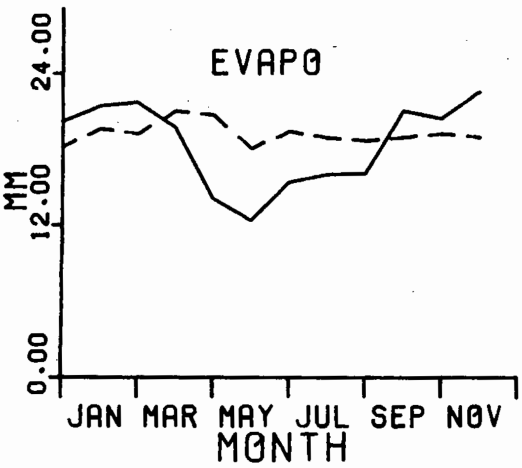
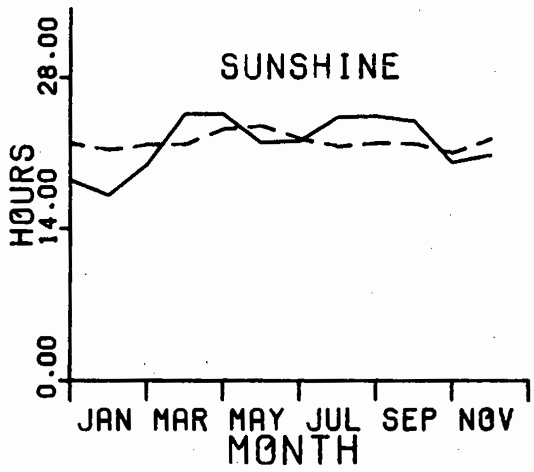


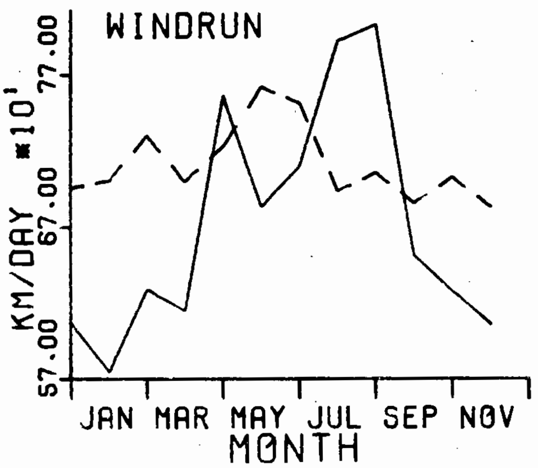
FIGURE 6.38: HISTORICAL (——) AND SIMULATED (---) MONTHLY STANDARD DEVIATIONS GIVEN A DRY DAY FOR ALL VARIABLES

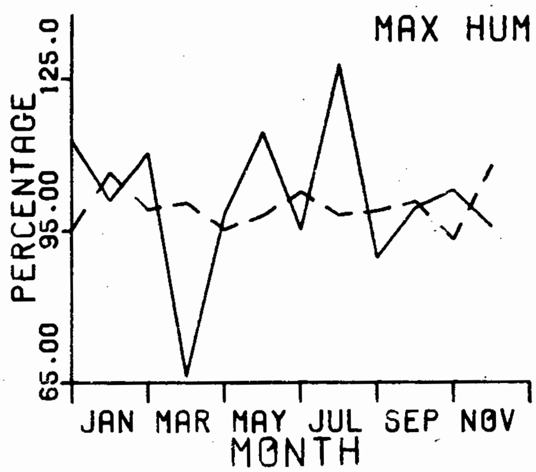





EVAPO
MM
24.00
12.00
0.00
JAN MAR MAY JUL SEP NOV
MONTH


SUNSHINE
HOURS
28.00
14.00
0.00
JAN MAR MAY JUL SEP NOV
MONTH


WINDRUN
KM/DAY *10¹
77.00
67.00
57.00
JAN MAR MAY JUL SEP NOV
MONTH


MAX HUM
PERCENTAGE
125.0
95.00
65.00
JAN MAR MAY JUL SEP NOV
MONTH

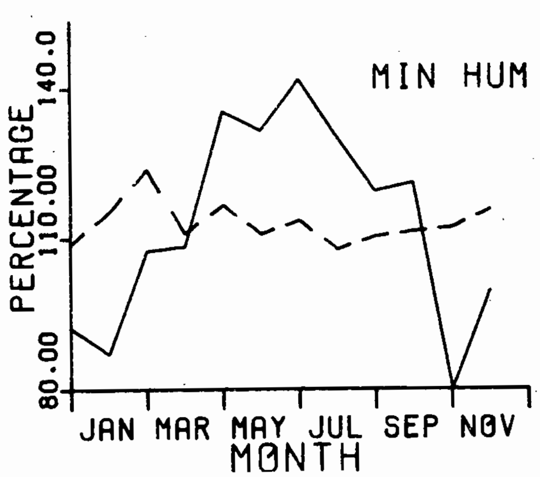
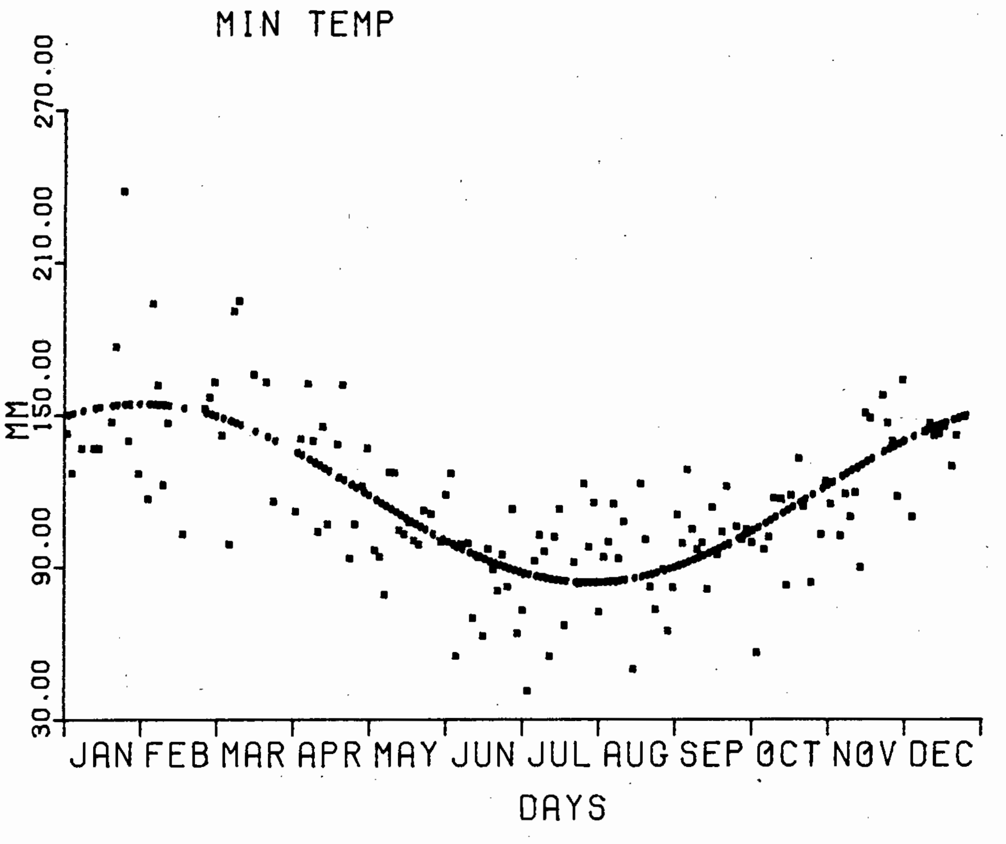
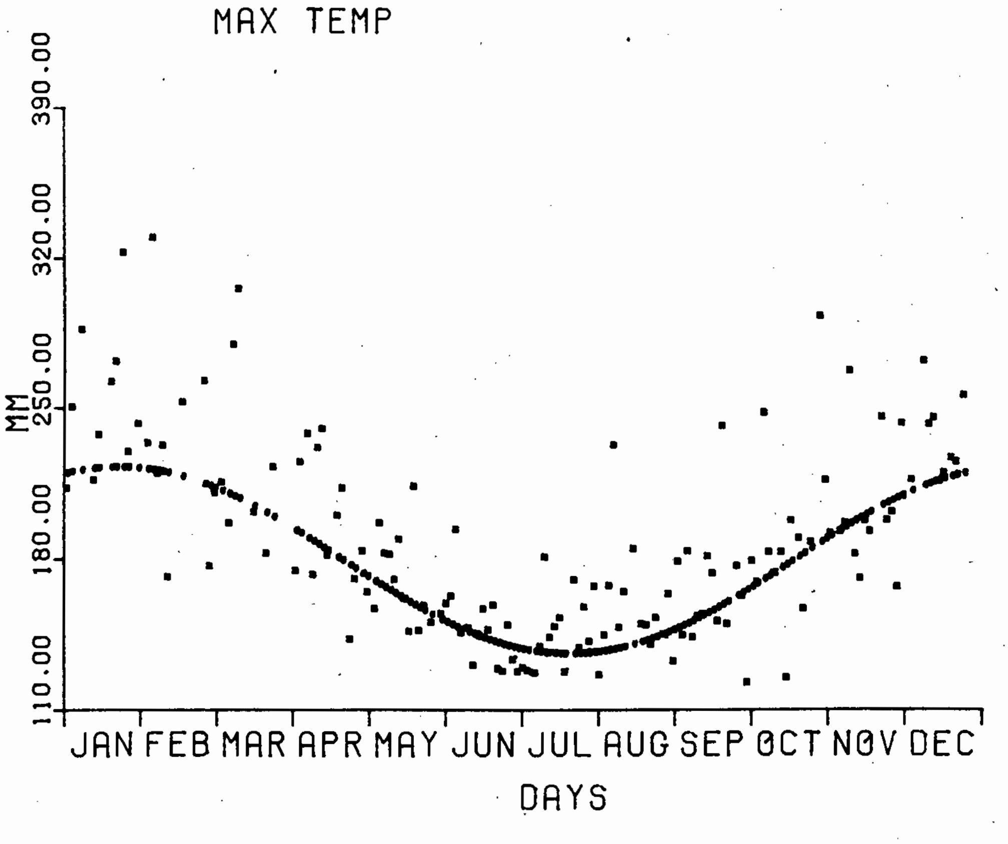
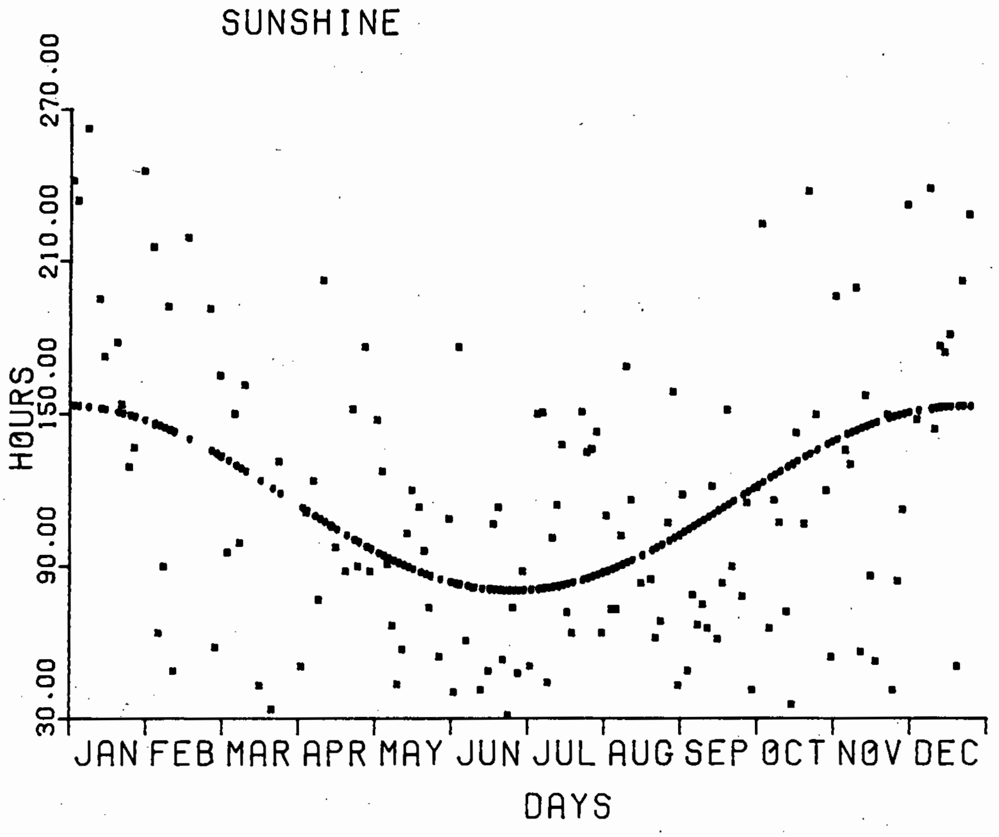
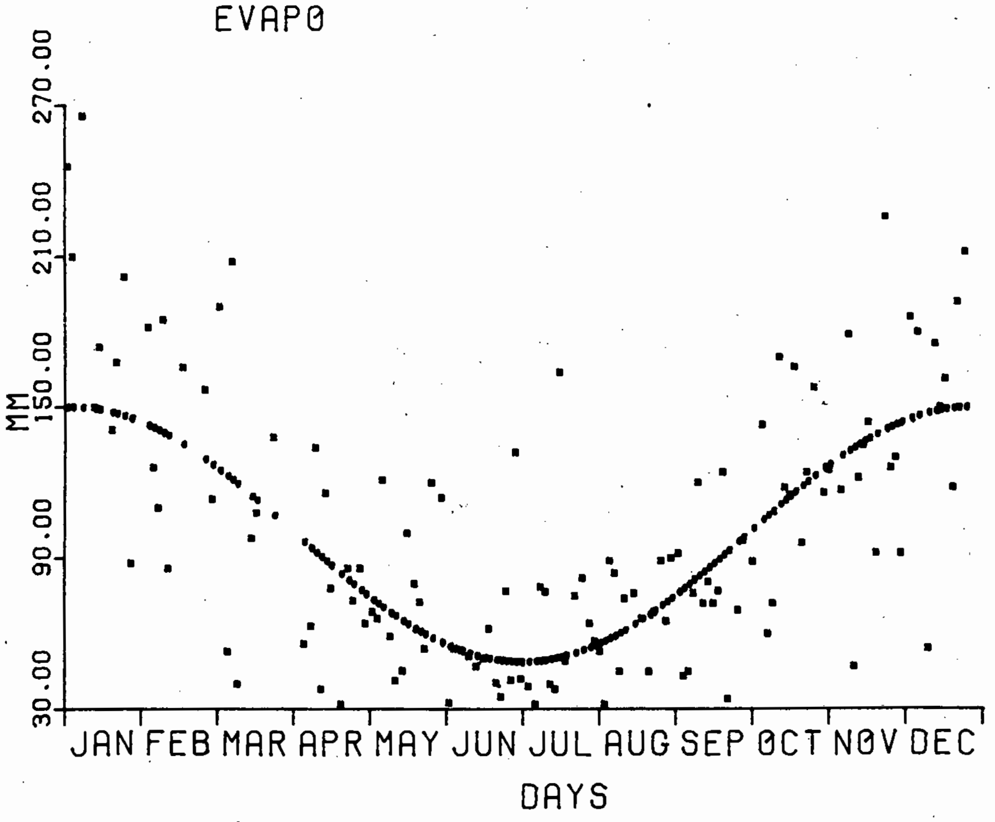

MIN HUM
PERCENTAGE
140.0
110.00
80.00
JAN MAR MAY JUL SEP NOV
MONTH

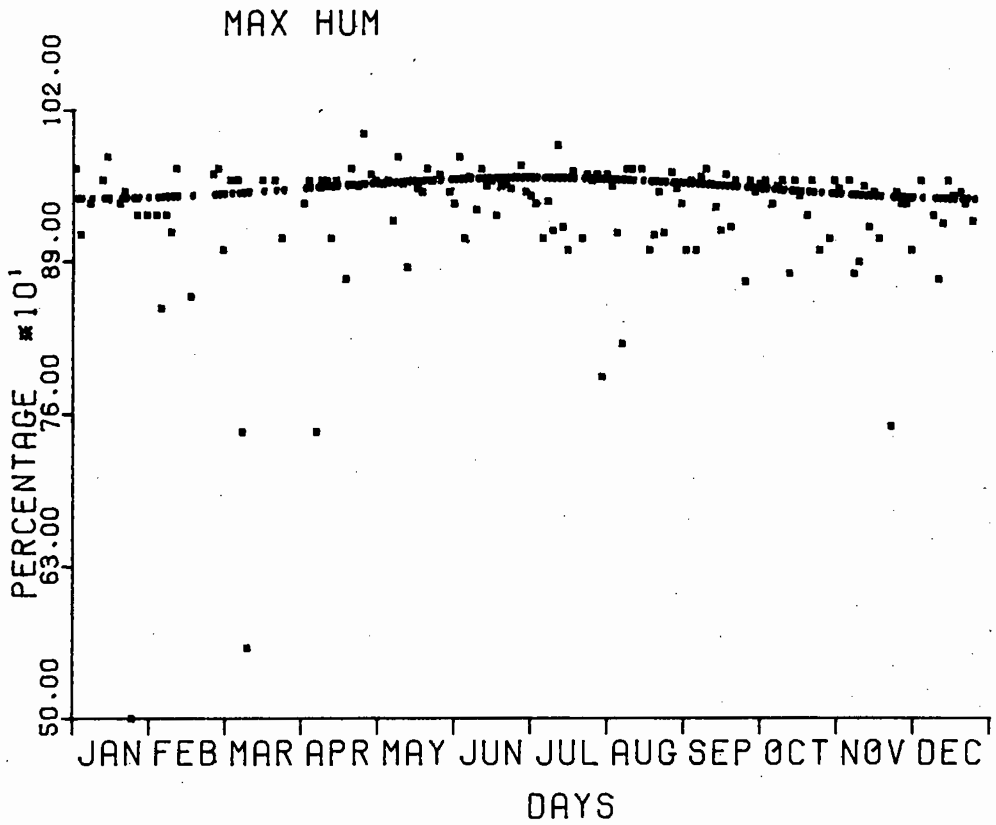
FIGURE 6.39: DAILY AVERAGES AND MEAN FITTED BY A FOURIER SERIES GIVEN A WET DAY FOR ALL VARIABLES

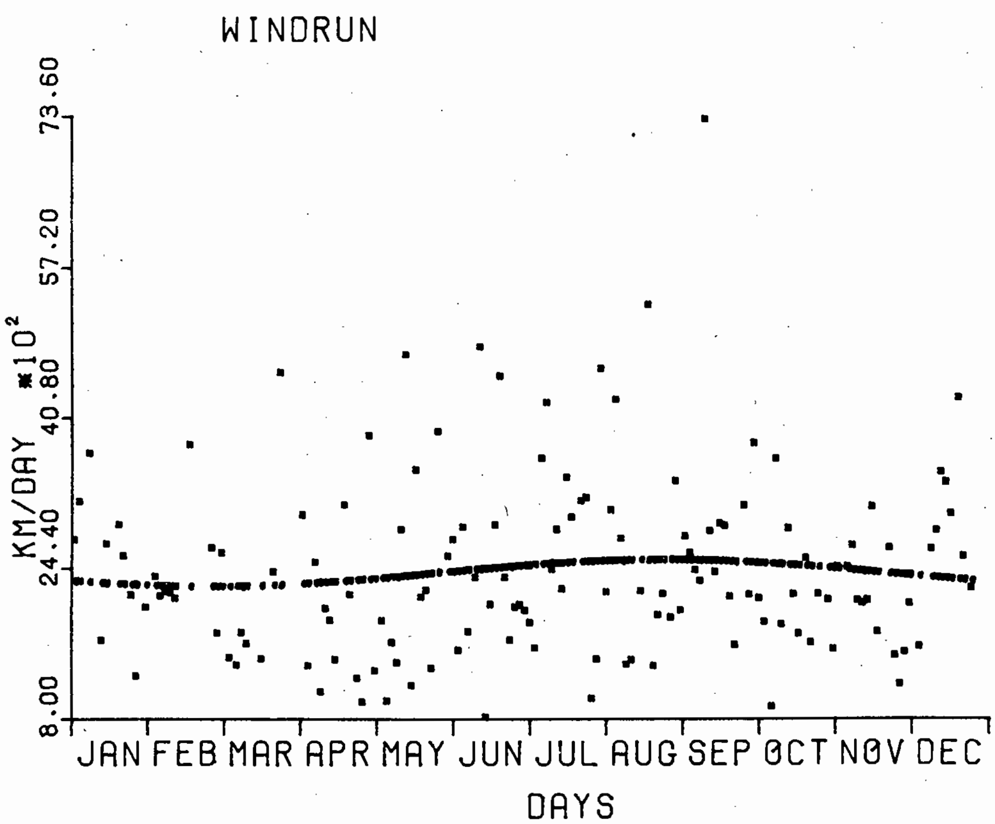





SUNSHINE
HOURS
270.00
210.00
150.00
90.00
30.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


EVAPO
MM
270.00
210.00
150.00
90.00
30.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


MAX HUM
PERCENTAGE *10¹
102.00
89.00
76.00
63.00
50.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


WINDRUN
KM/DAY *10²
73.60
57.20
40.80
24.40
8.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS

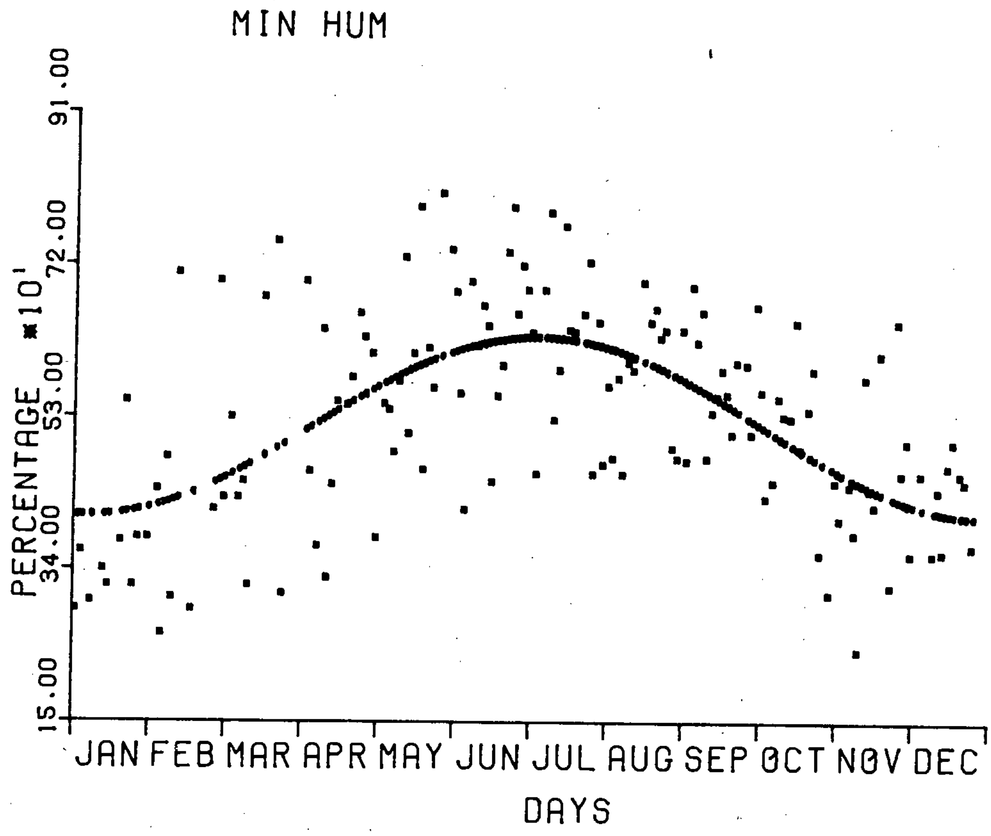
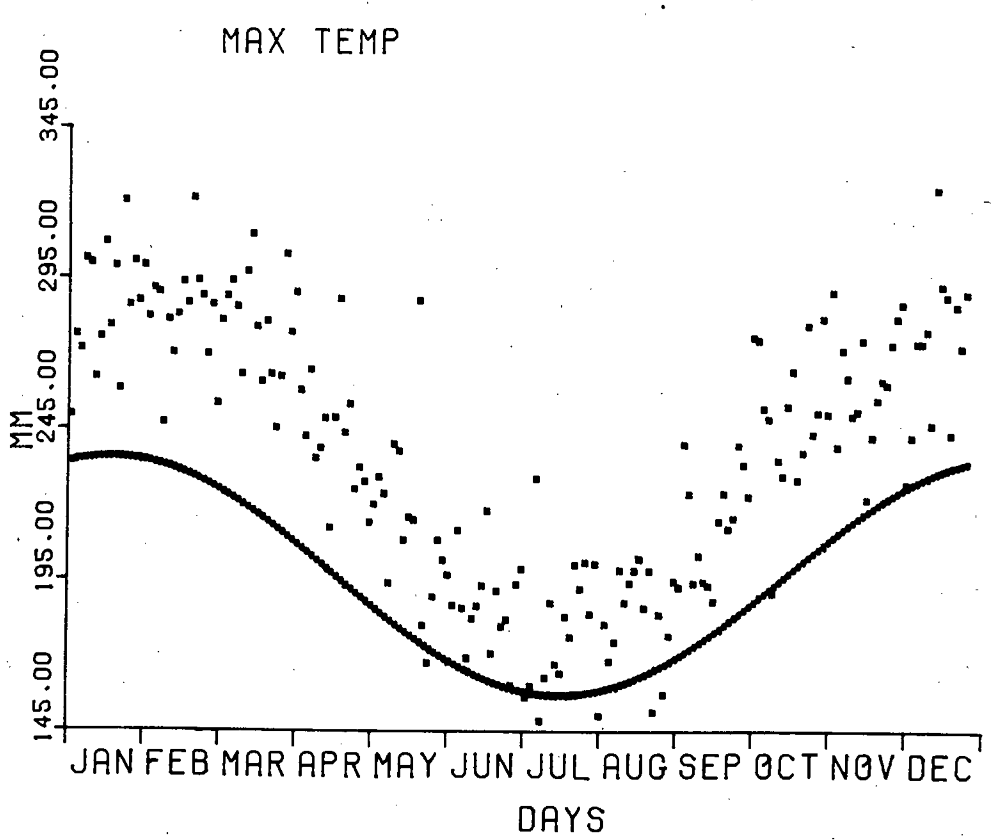
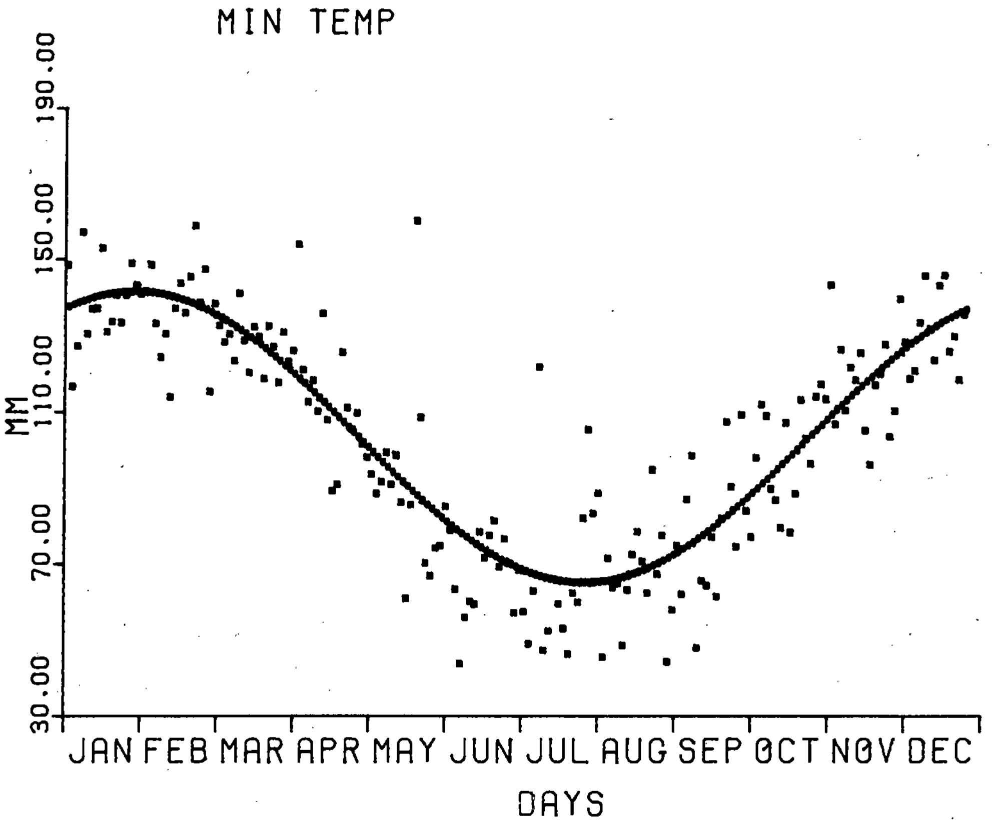
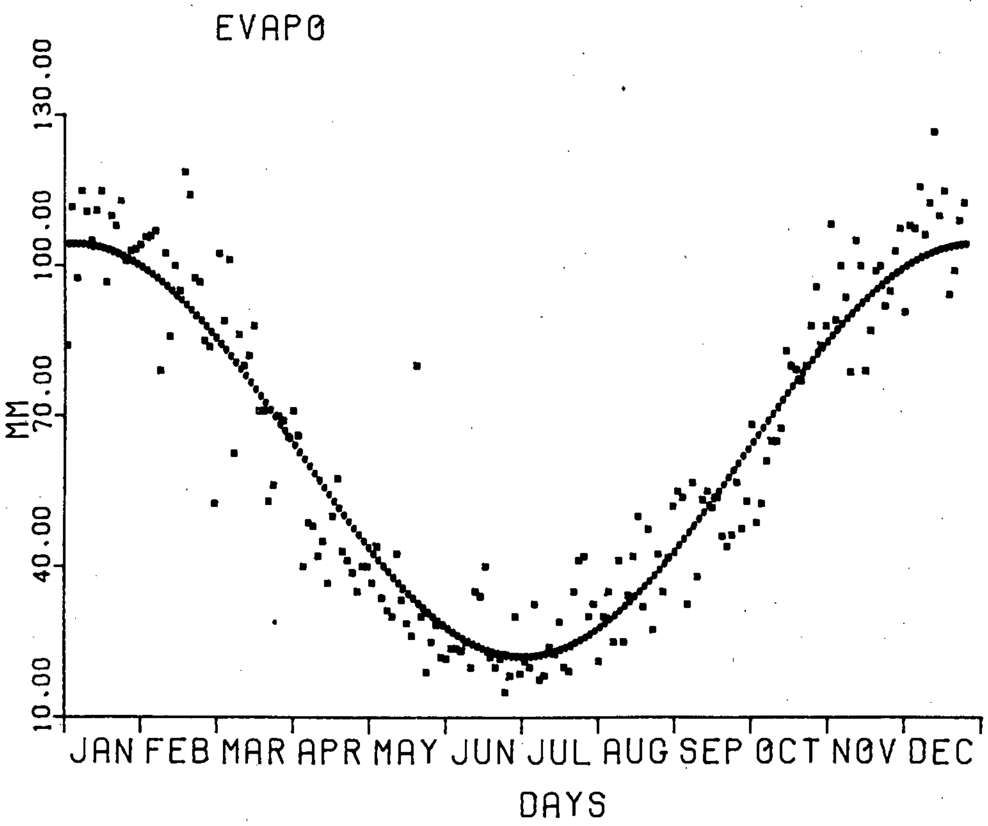
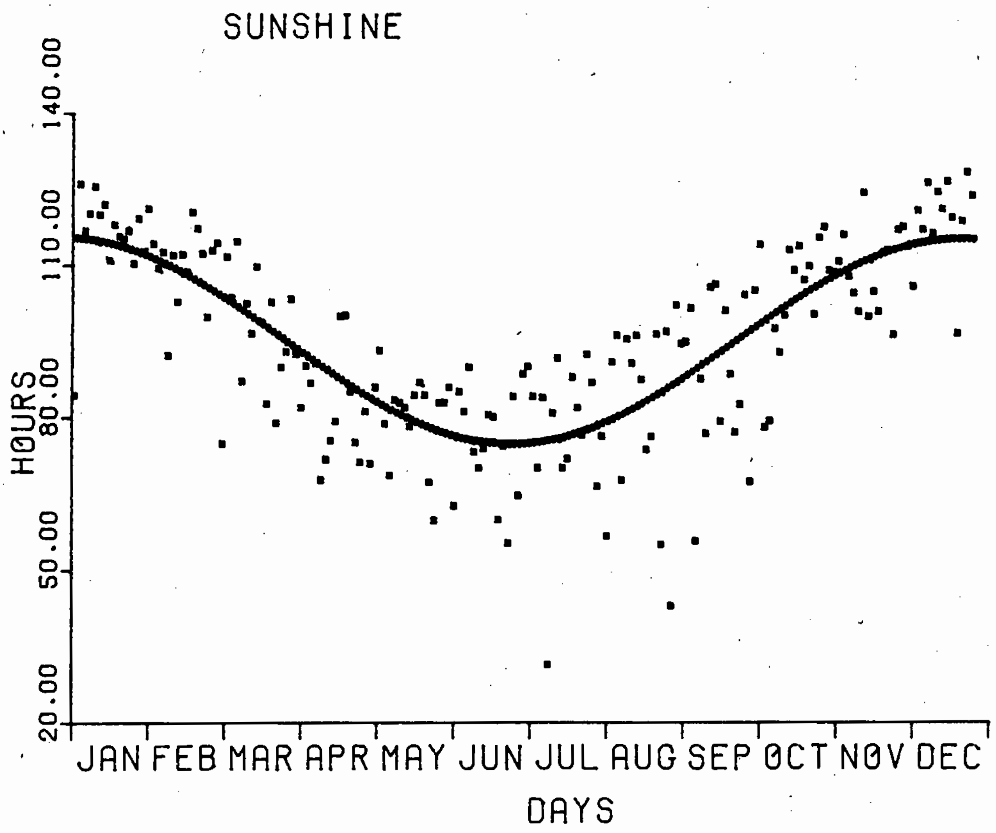


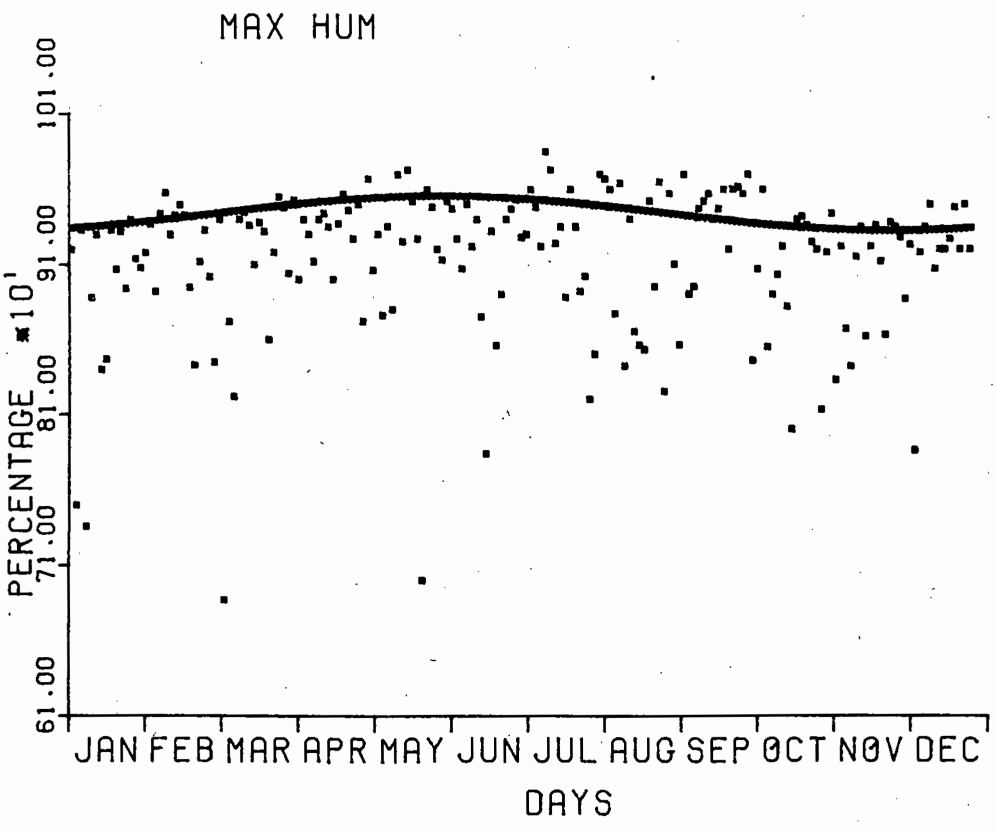
FIGURE 6.40 DAILY AVERAGES AND MEAN FITTED BY A FOURIER SERIES GIVEN A DRY DAY FOR ALL VARIABLES

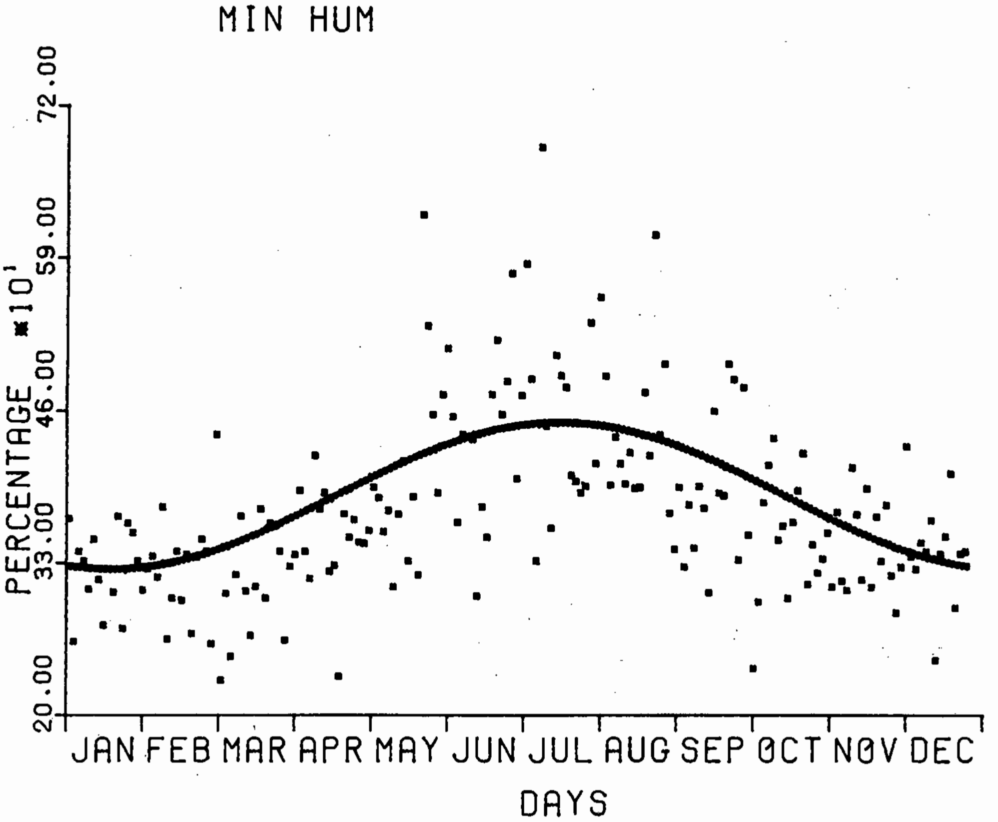


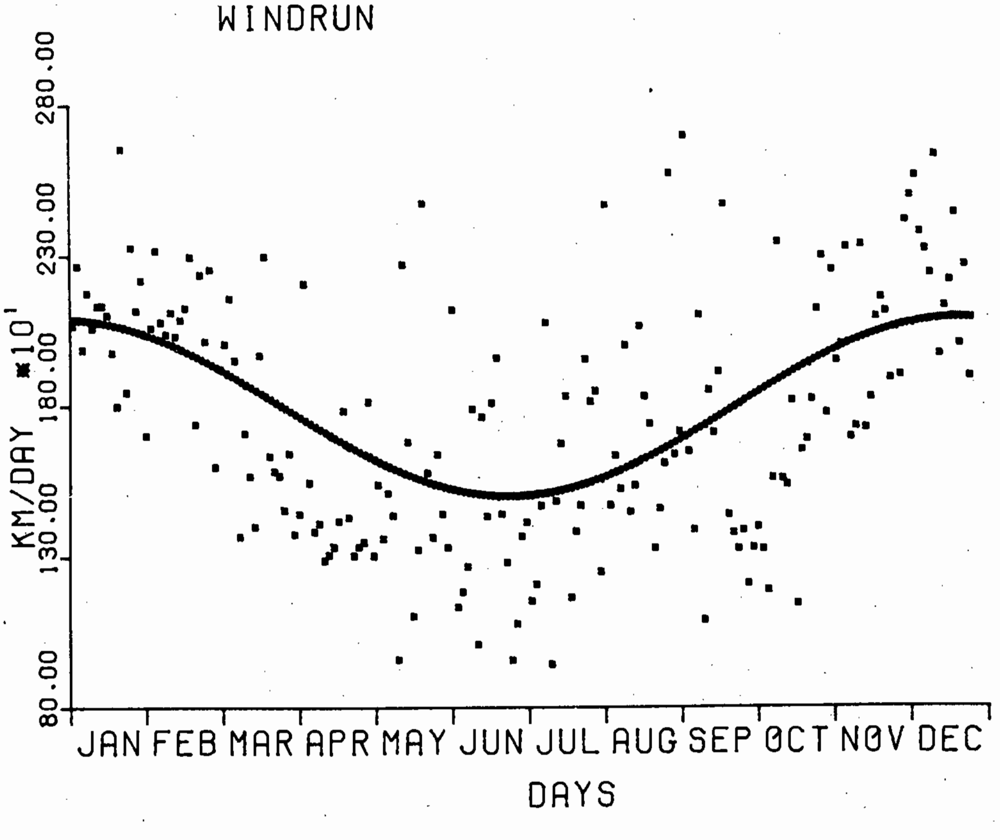

MIN TEMP
190.00
150.00
110.00
70.00
30.00
MM
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


EVAPO
130.00
100.00
70.00
40.00
10.00
MM
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


SUNSHINE
HOURS
140.00
110.00
80.00
50.00
20.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


MAX HUM
PERCENTAGE *10¹
101.00
91.00
81.00
71.00
61.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


MIN HUM
PERCENTAGE *10¹
72.00
59.00
46.00
33.00
20.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


WINDRUN
KM/DAY *10¹
280.00
230.00
180.00
130.00
80.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS

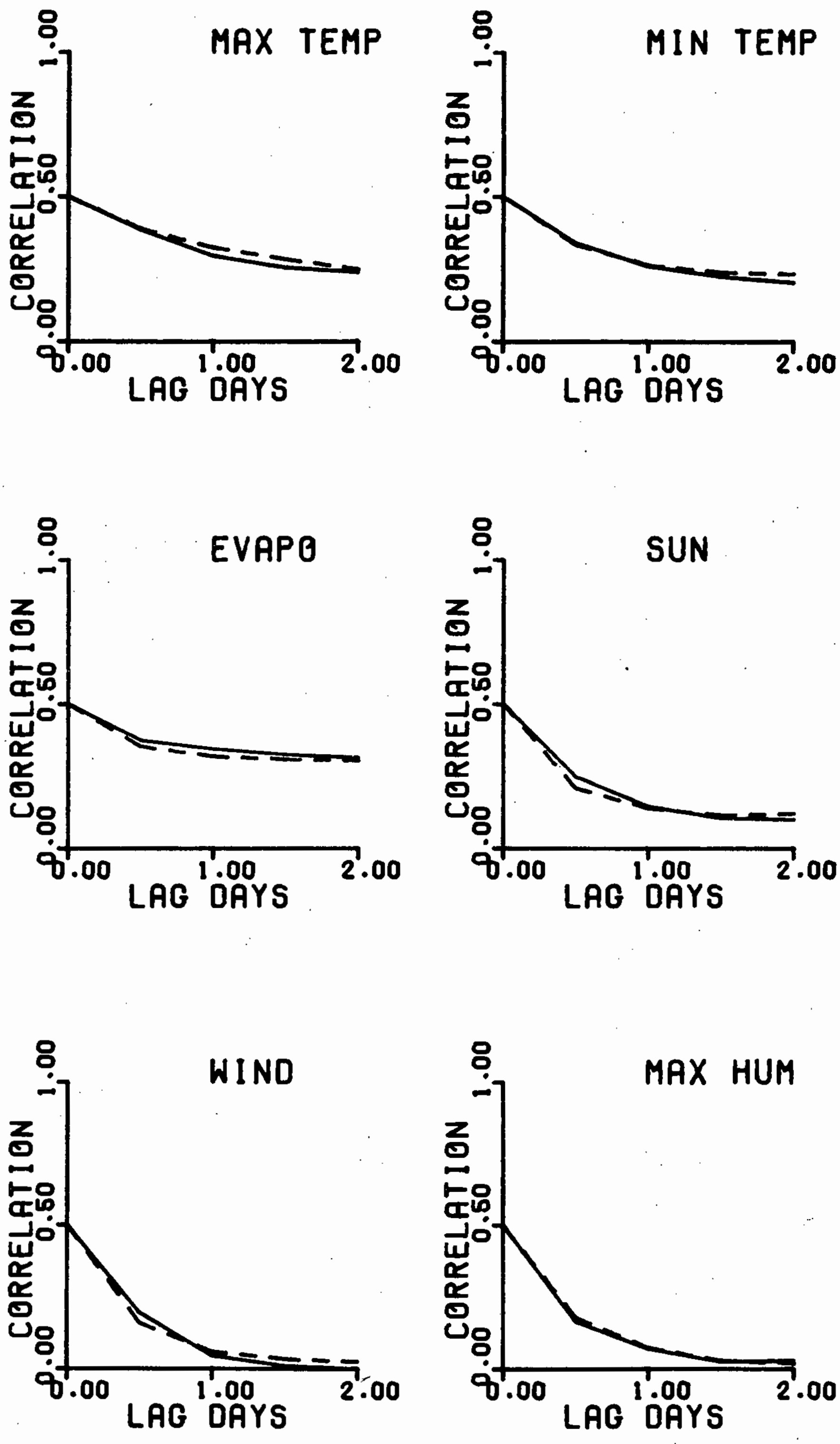
present in the historical record. Figure 6.41 shows the comparison between the autocorrelation of variables in the historical sequence and those of the variables in the simulated sequence. From these comparisons it is concluded that the model has described the autocorrelation property exceptionally well.

The comparison of the autocorrelation coefficients between historical and simulated sequences conditioned on the wet or dry status of the day is illustrated in Figures 6.42 - 6.43. Again the model preserves the autocorrelations very well.

The comparison of the historical and simulated cross-correlation coefficients for all climate variables is illustrated in Figure 6.44. The cross-correlation coefficients have been retained satisfactorily by the model except when maximum temperature is lagged with minimum temperature, evaporation and with sunshine duration. In these cases the cross-correlation coefficients are slightly underestimated (a difference of 0.34 for maximum temperature lagged with minimum temperature for a lag of zero, and a difference of 0.27 in the case of evaporation). For sunshine duration lagged with maximum temperature a difference of 0.2 is observed for a lag of zero and of 0.3 for a lag of one.

The cross-correlation coefficients for all climate variables conditioned on the wet or dry status of the day is well preserved by the model. Again, more marked differences between historical and simulated coefficients are observed for the wet sequence of days. (Figures 6.45 - 6.46.)

FIGURE 6.41: HISTORICAL (——) AND SIMULATED (---) AUTOCORRELATION COEFFICIENTS FOR ALL VARIABLES



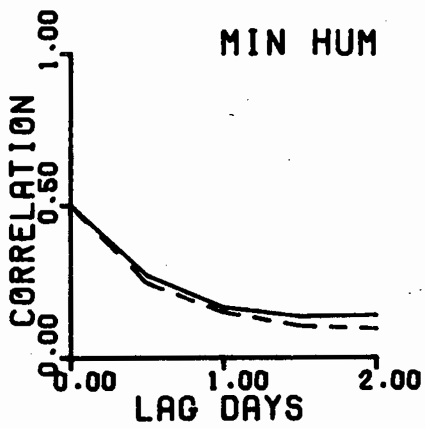
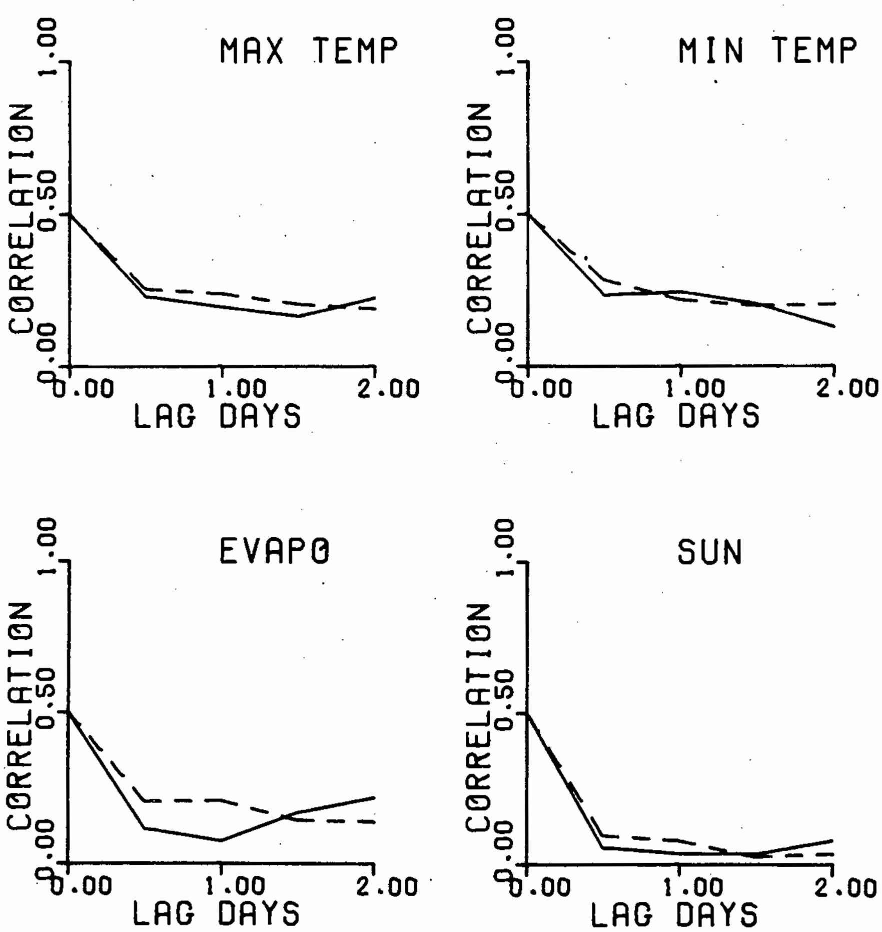


FIGURE 6.42: HISTORICAL (——) AND SIMULATED (---) AUTOCORRELATION COEFFICIENTS GIVEN A WET DAY FOR ALL VARIABLES



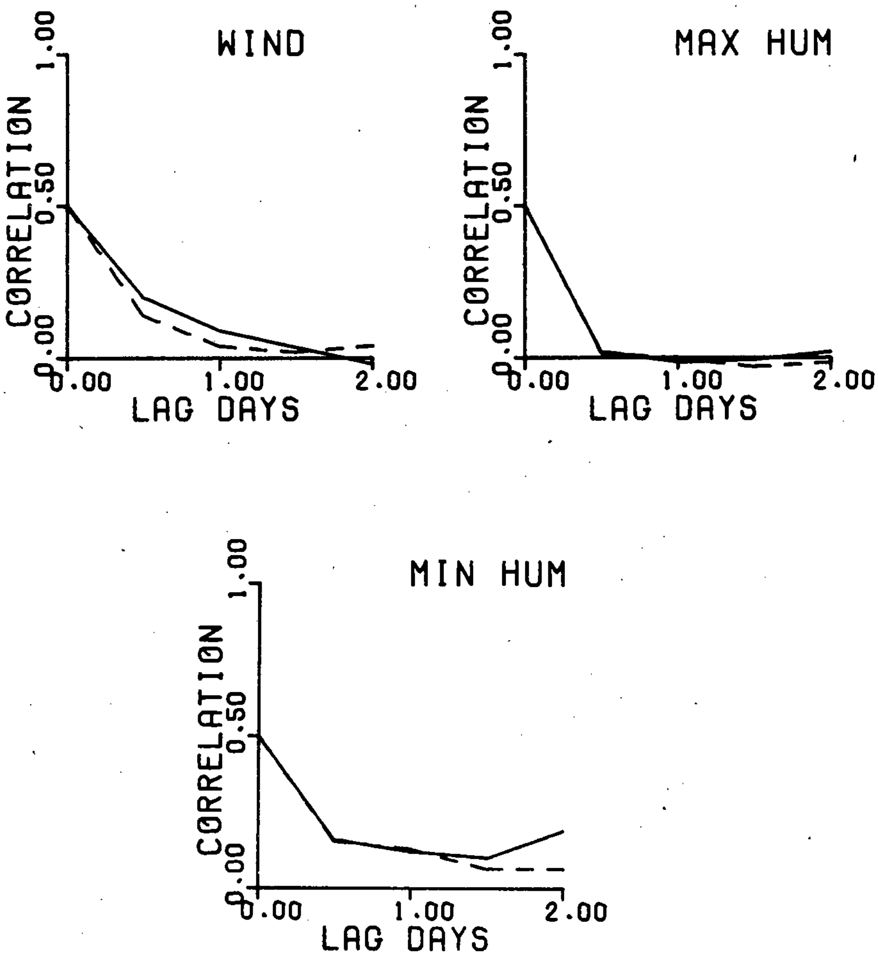
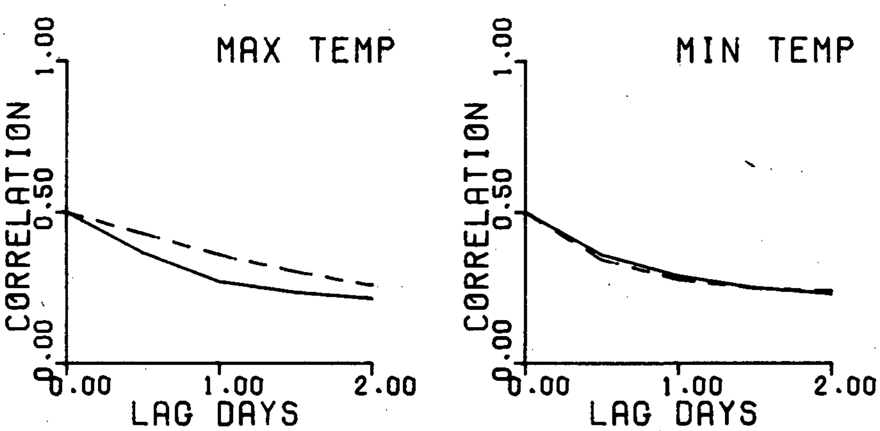


FIGURE 6.43: HISTORICAL (——) AND SIMULATED (---) AUTOCORRELATION COEFFICIENTS GIVEN A DRY DAY FOR ALL VARIABLES



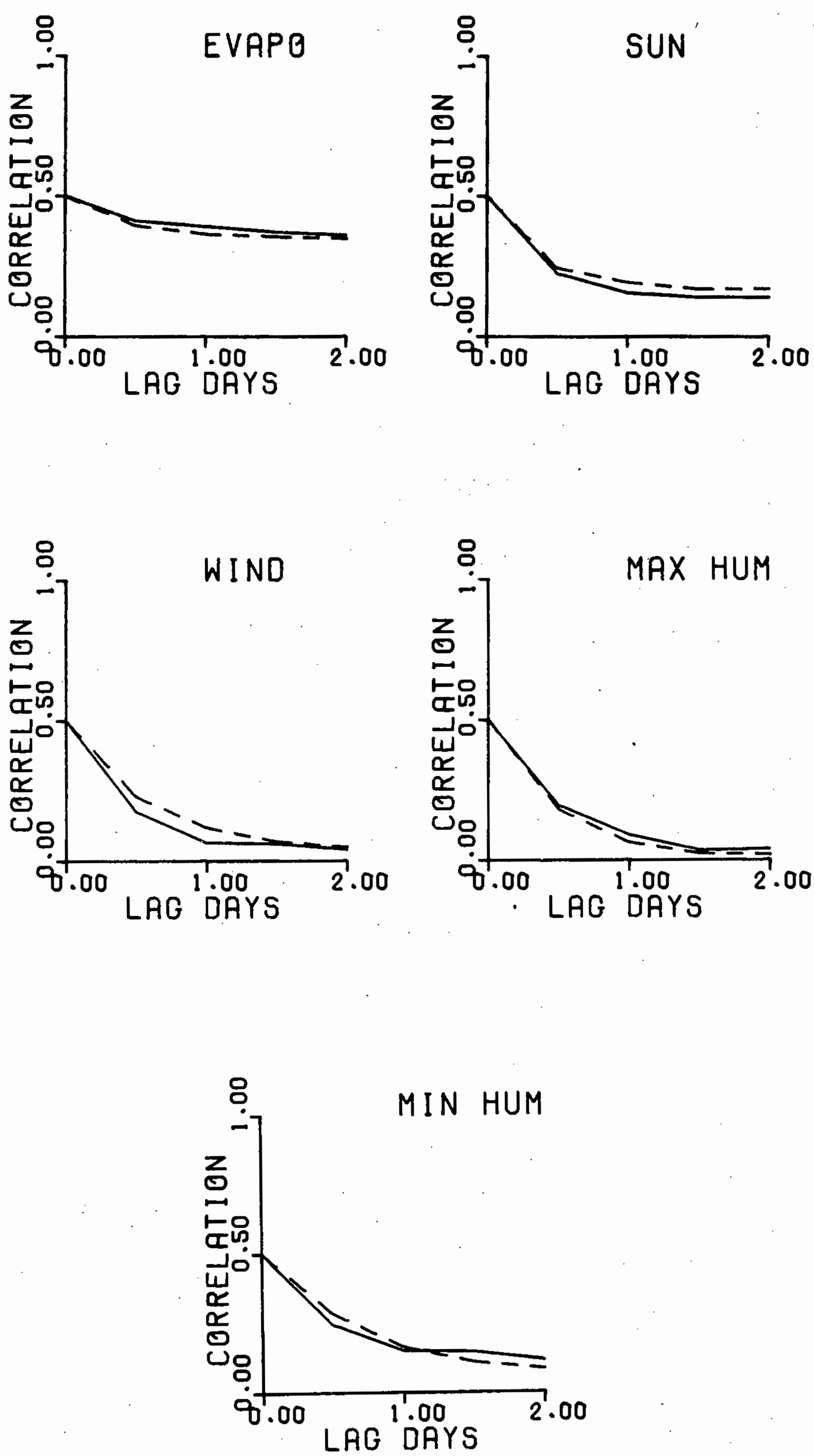
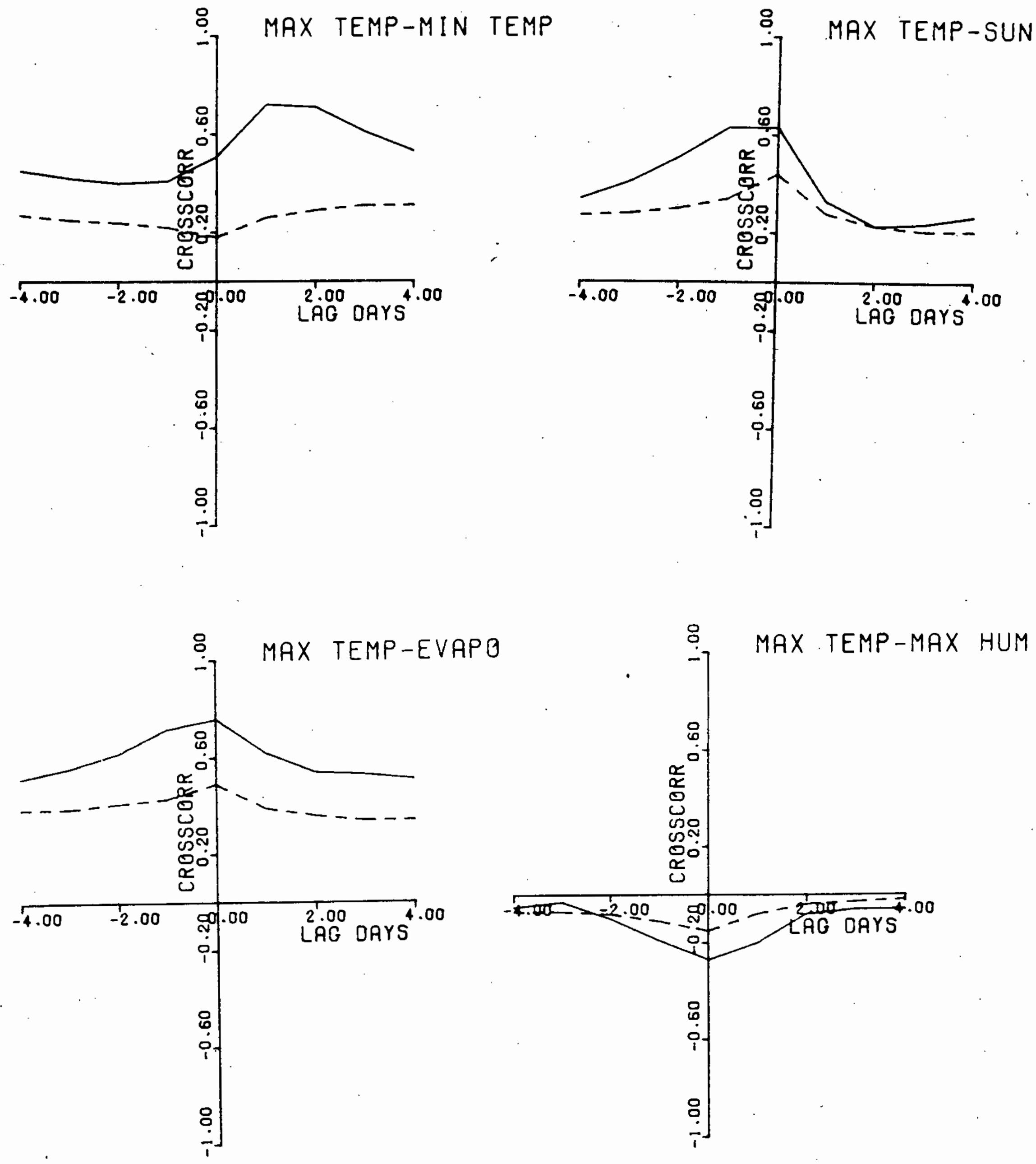
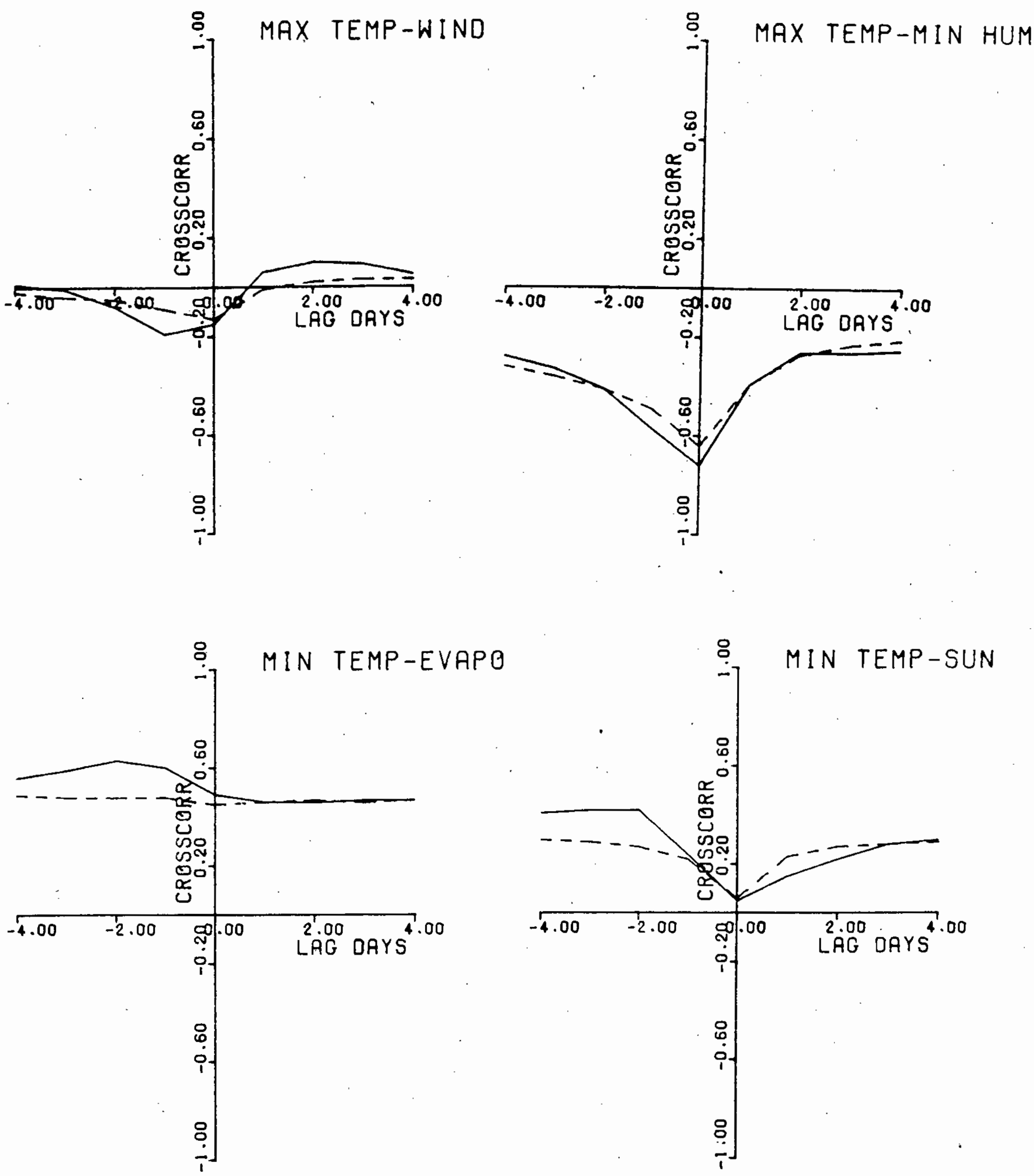
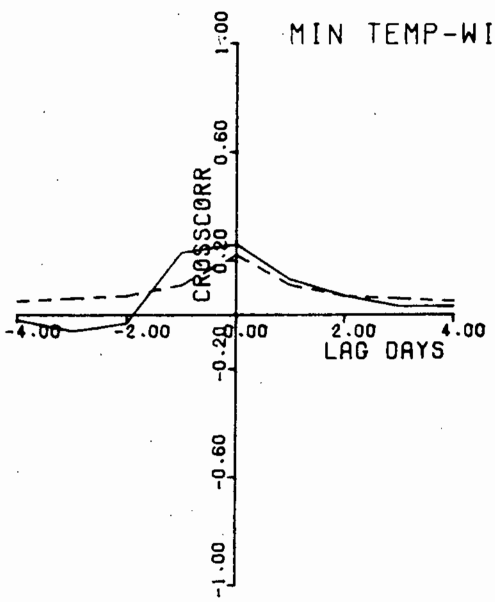
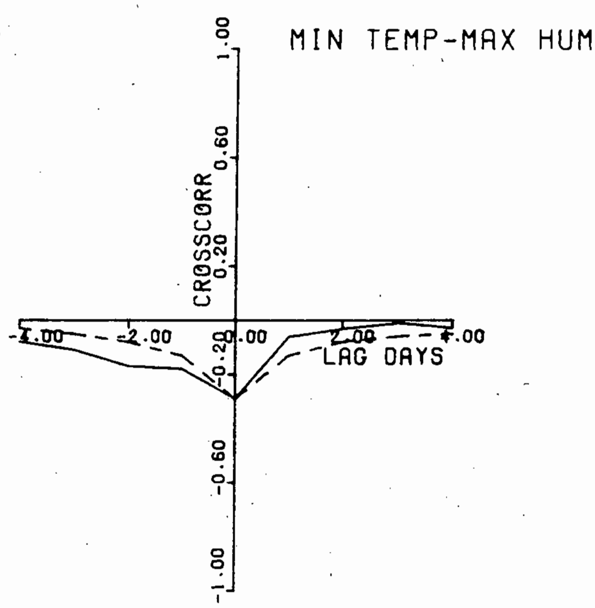

EVAPO
SUN
WIND
MAX HUM
MIN HUM
CORRELATION
0.00
0.50
1.00
LAG DAYS
0.00
1.00
2.00

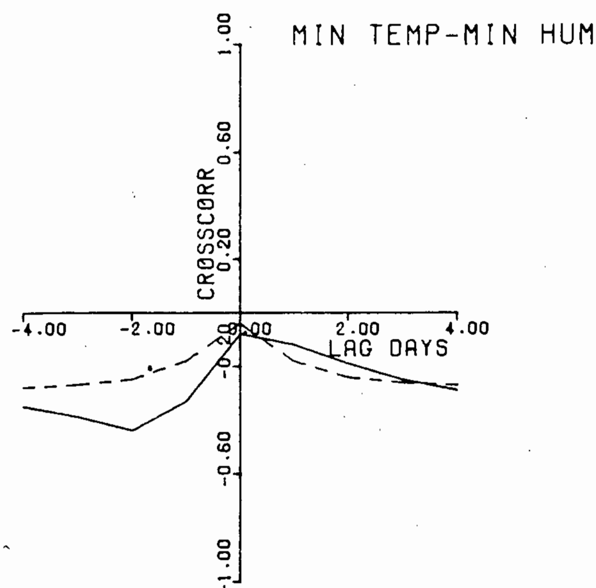
FIGURE 6.44: HISTORICAL (——) AND SIMULATED (---) CROSS-CORRELATION COEFFICIENTS

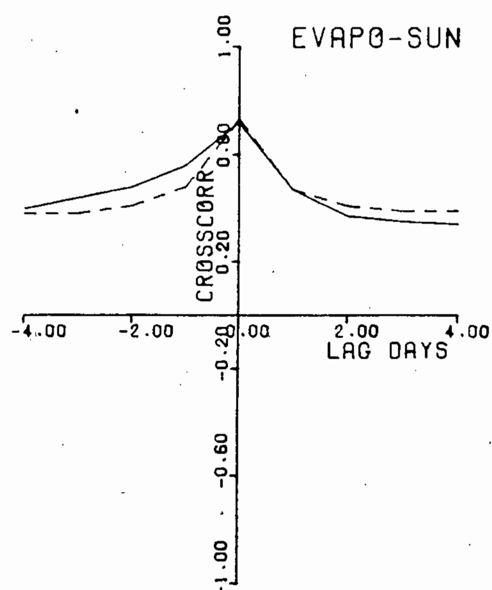


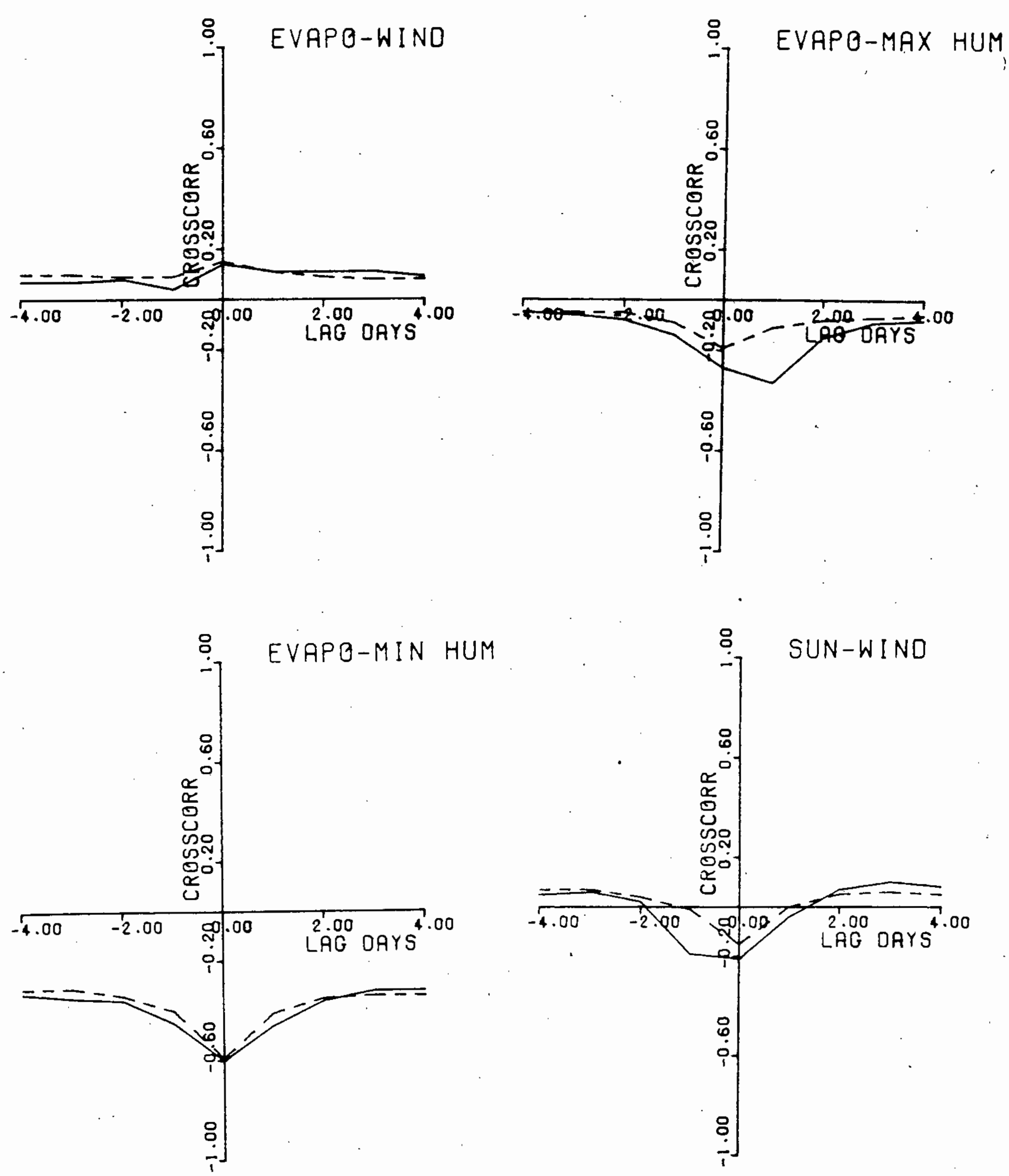

MAX TEMP-WIND
CROSSCORR
LAG DAYS
MAX TEMP-MIN HUM
CROSSCORR
LAG DAYS
MIN TEMP-EVAPO
CROSSCORR
LAG DAYS
MIN TEMP-SUN
CROSSCORR
LAG DAYS

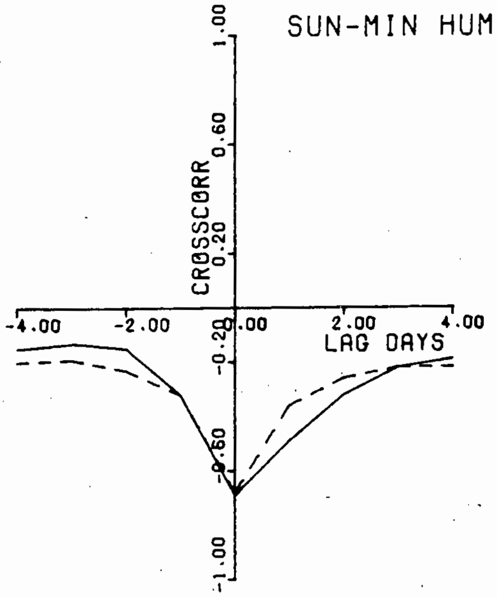

MIN TEMP-WIND
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

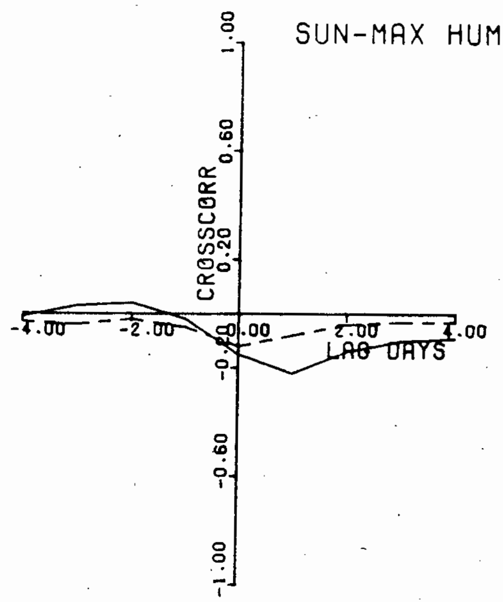

MIN TEMP-MAX HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

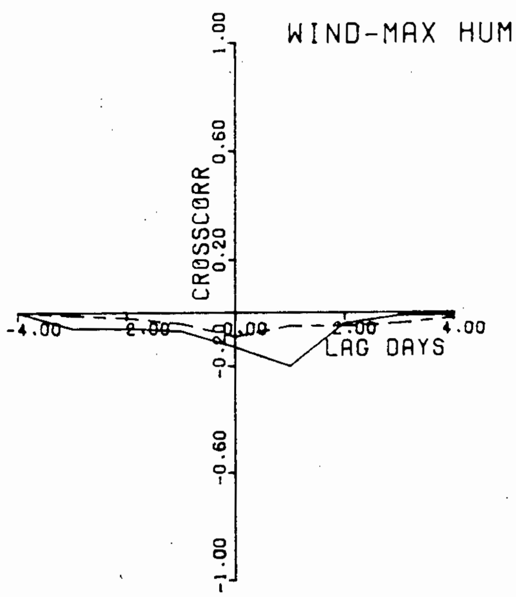

MIN TEMP-MIN HUM
CROSSCORR
1.00
0.60
0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

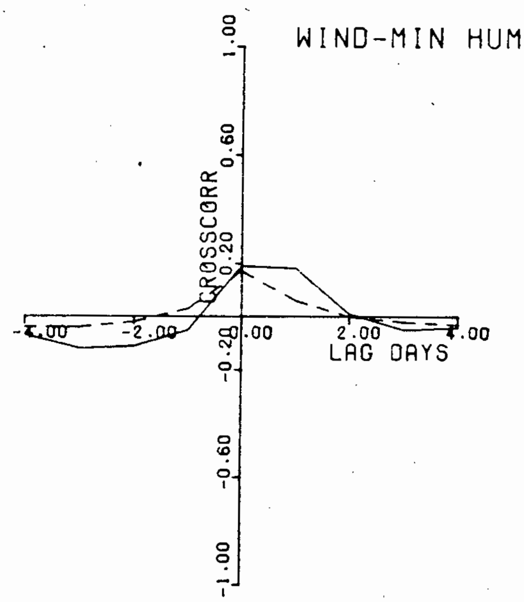

EVAPO-SUN
CROSSCORR
1.00
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS


EVAPO-WIND
EVAPO-MAX HUM
EVAPO-MIN HUM
SUN-WIND
CROSSCORR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00


SUN-MIN HUM
CRØSSCØRR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS


SUN-MAX HUM
CRØSSCØRR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS


WIND-MAX HUM
CRØSSCØRR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS


WIND-MIN HUM
CRØSSCØRR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

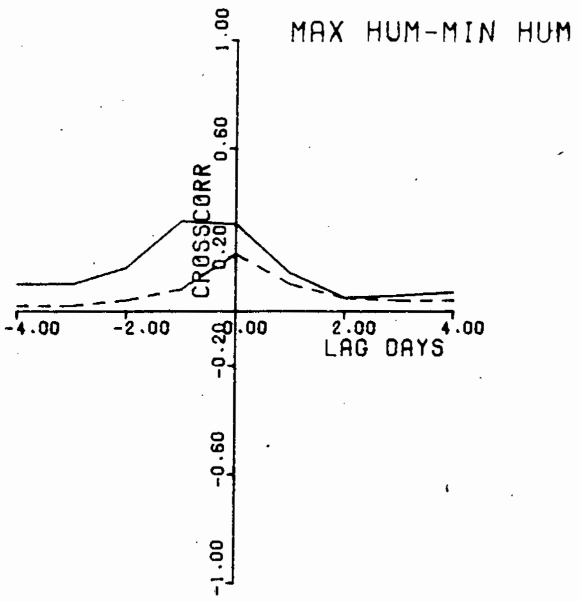
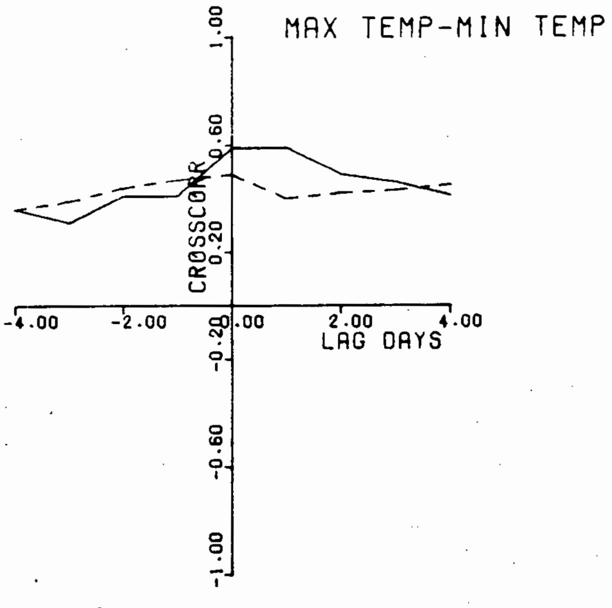
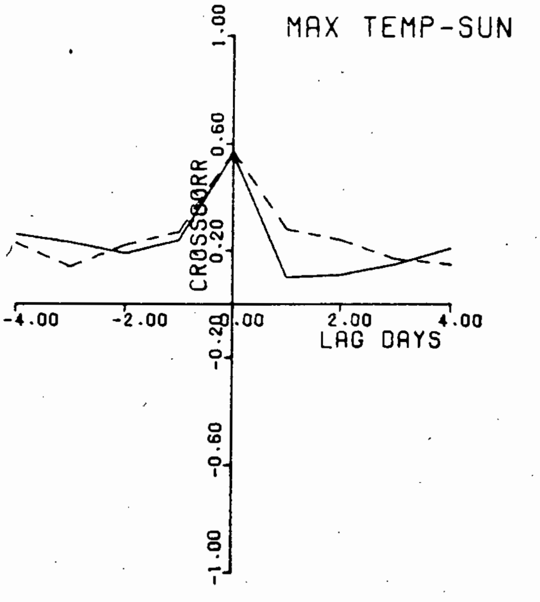
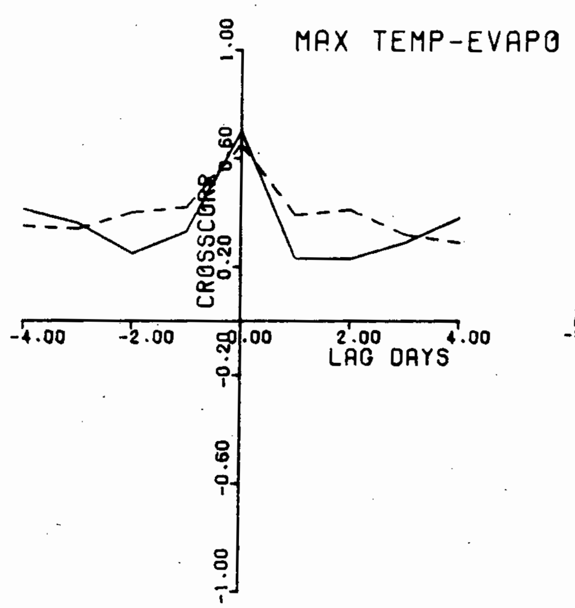
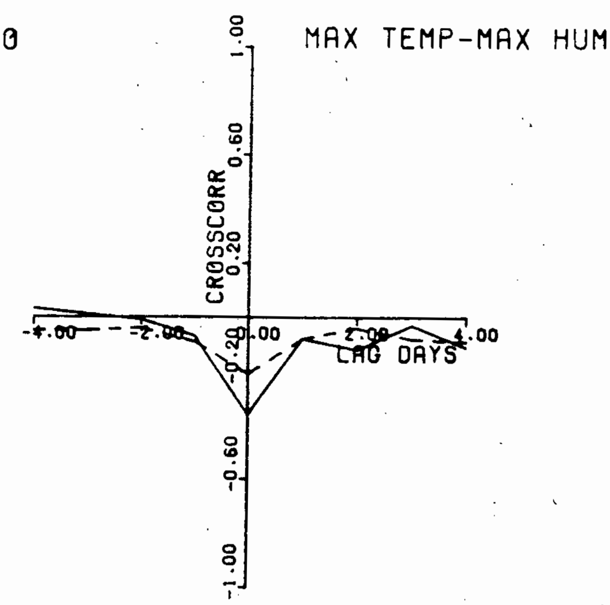


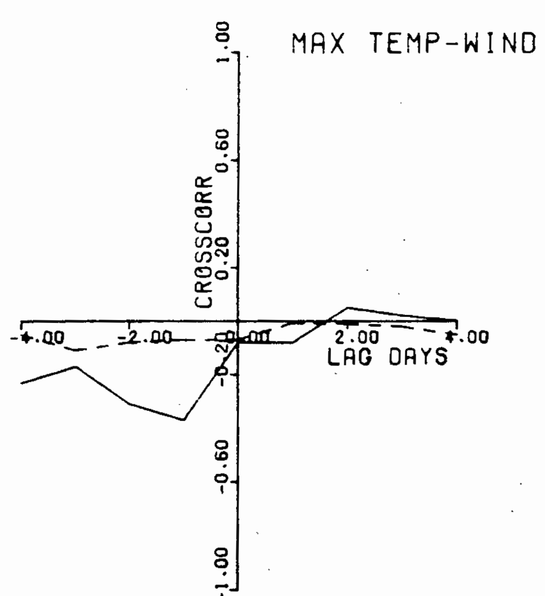
FIGURE 6.45: HISTORICAL (——) AND SIMULATED CROSS-CORRELATION COEFFICIENTS GIVEN A WET DAY

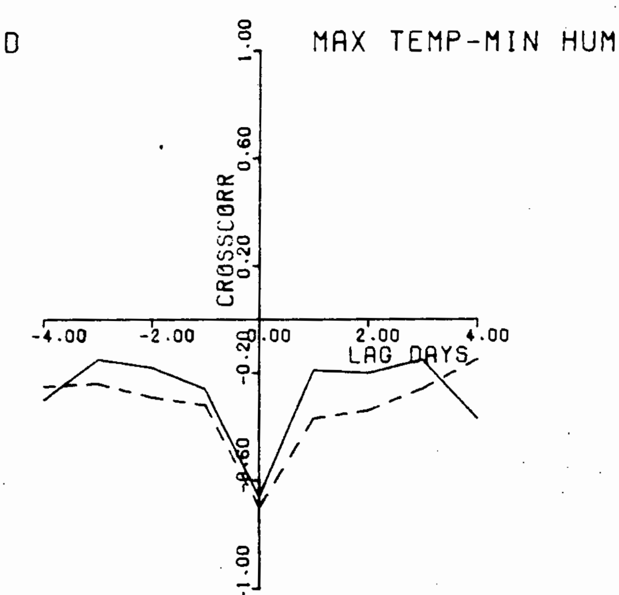


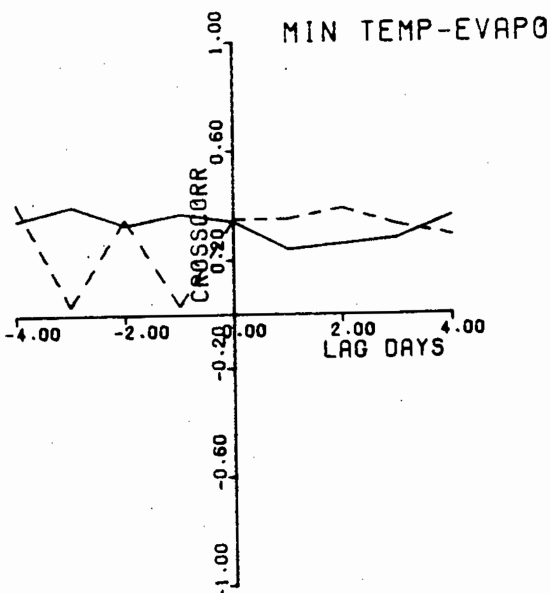


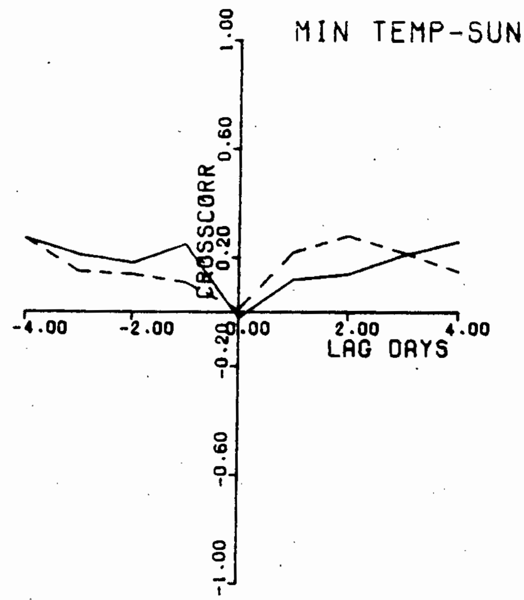

MAX TEMP-EVAPO
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

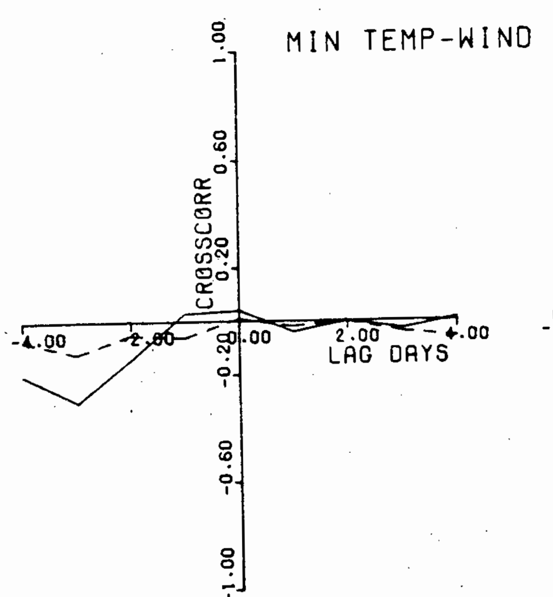

MAX TEMP-MAX HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

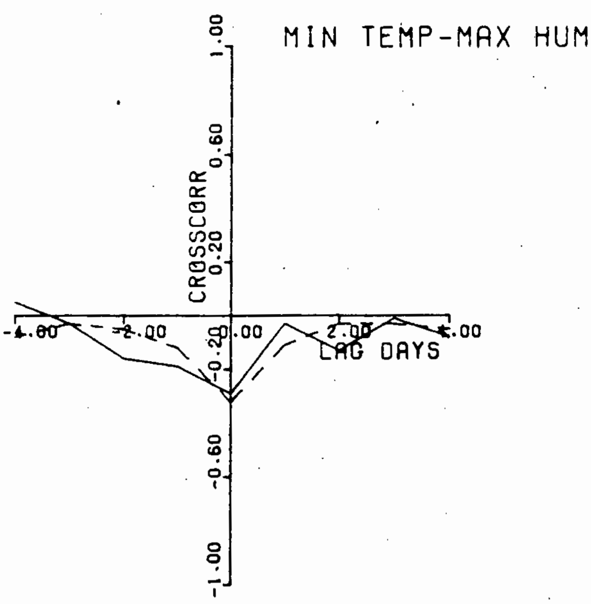

MAX TEMP-WIND
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

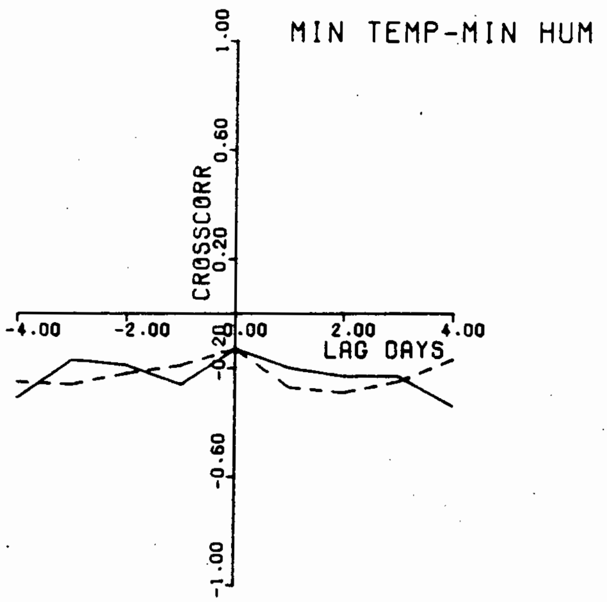

MAX TEMP-MIN HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

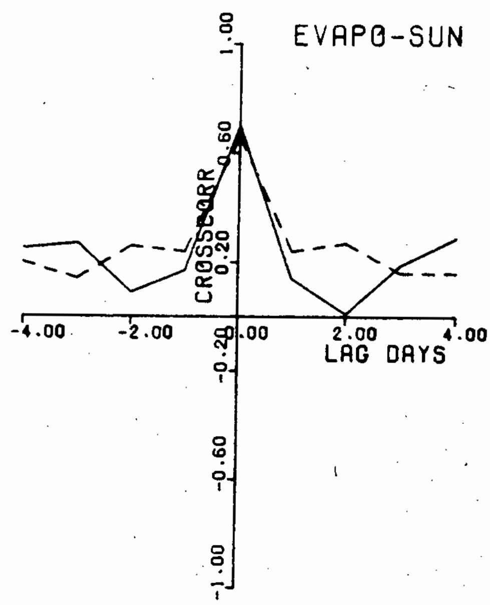

MIN TEMP-EVAPO
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

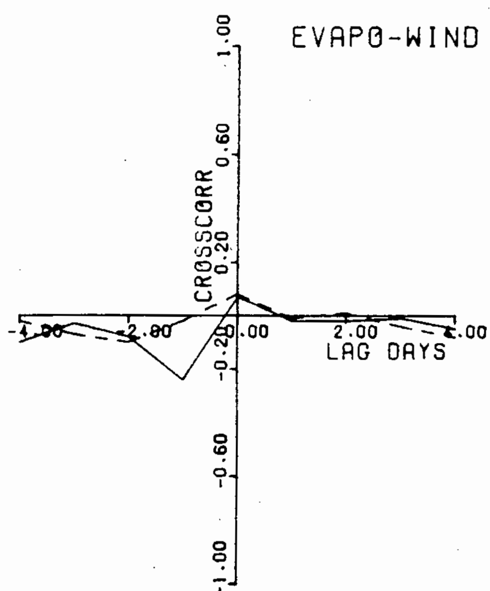

MIN TEMP-SUN
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

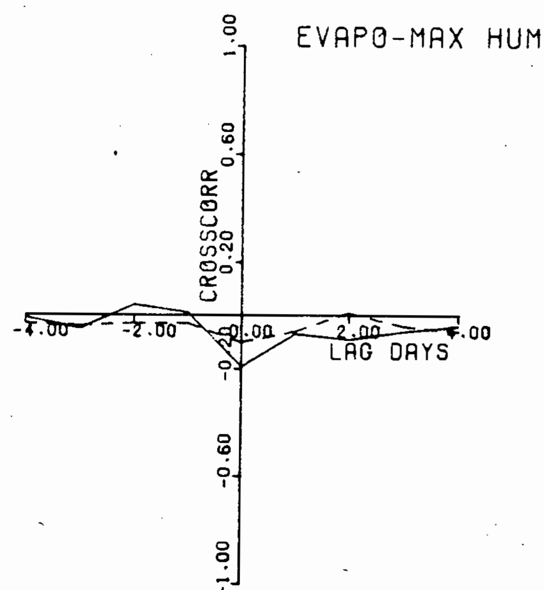

MIN TEMP-WIND
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

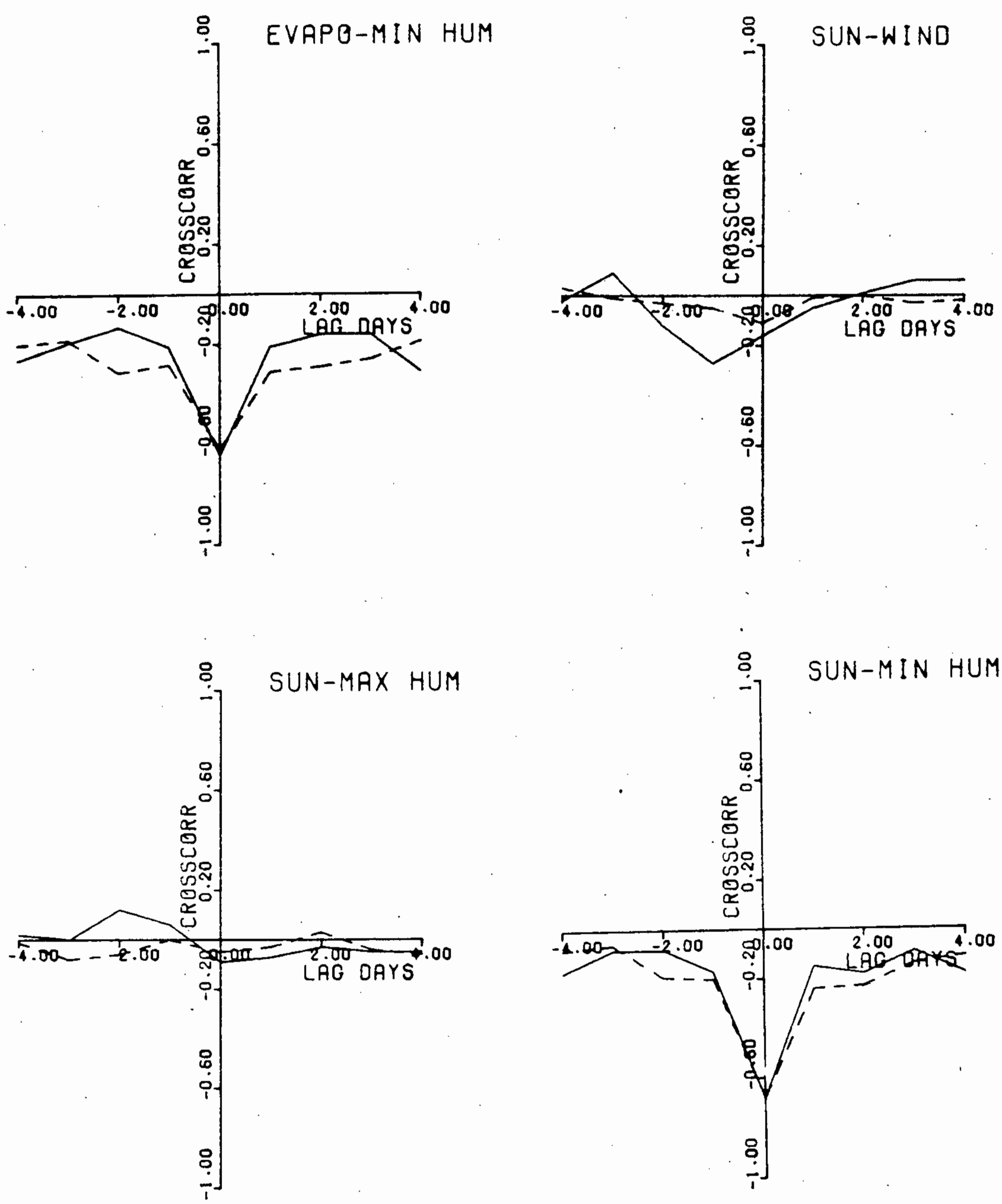

MIN TEMP-MAX HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

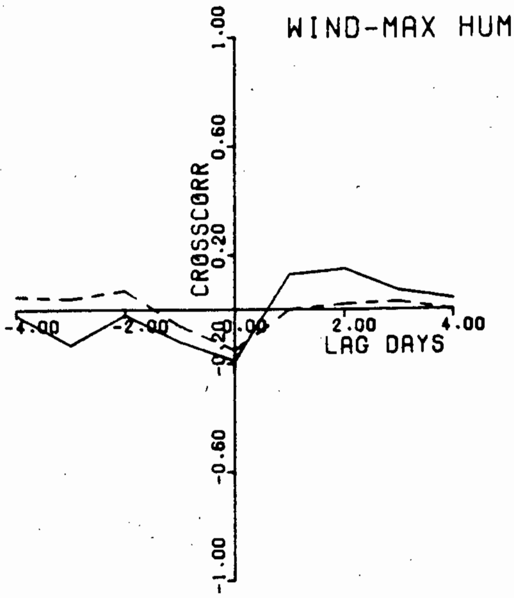

MIN TEMP-MIN HUM
CROSSCORR
LAG DAYS

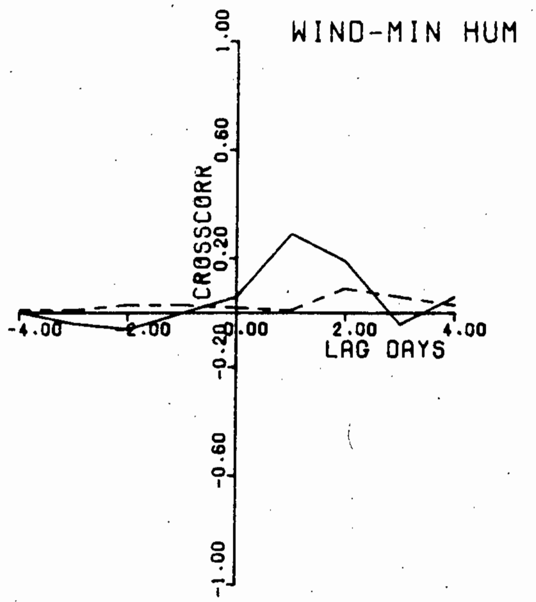

EVAPO-SUN
CROSSCORR
LAG DAYS


EVAPO-WIND
CROSSCORR
LAG DAYS


EVAPO-MAX HUM
CROSSCORR
LAG DAYS


EVAPO-MIN HUM
SUN-WIND
SUN-MAX HUM
SUN-MIN HUM
CROSSCORR
LAG DAYS
-1.00 -0.60 -0.20 0.20 0.60 1.00
-4.00 -2.00 0.00 2.00 4.00


WIND-MAX HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS


WIND-MIN HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

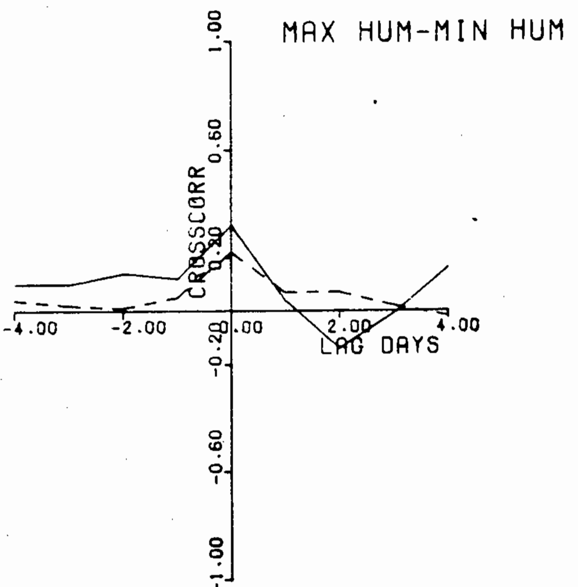
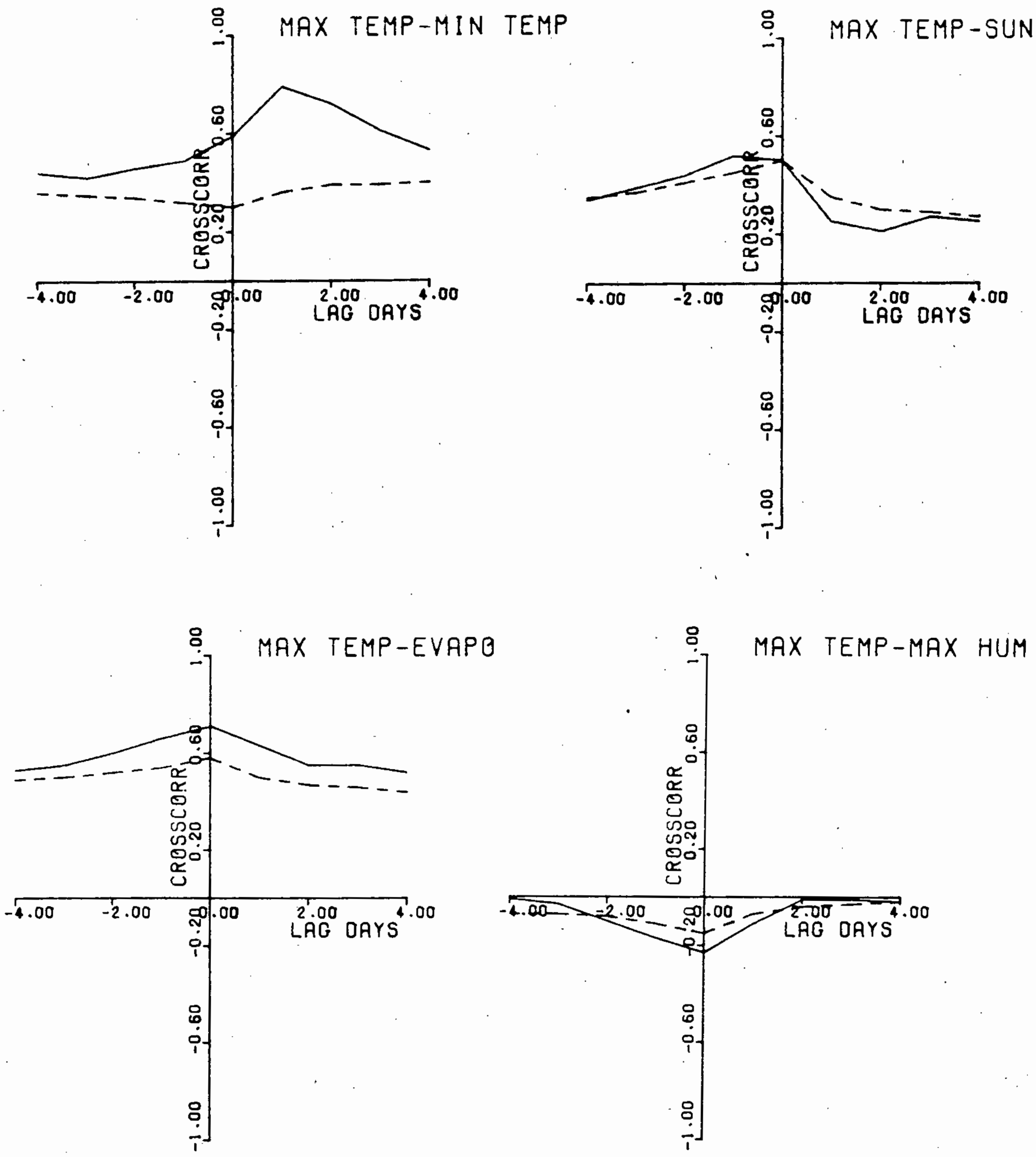
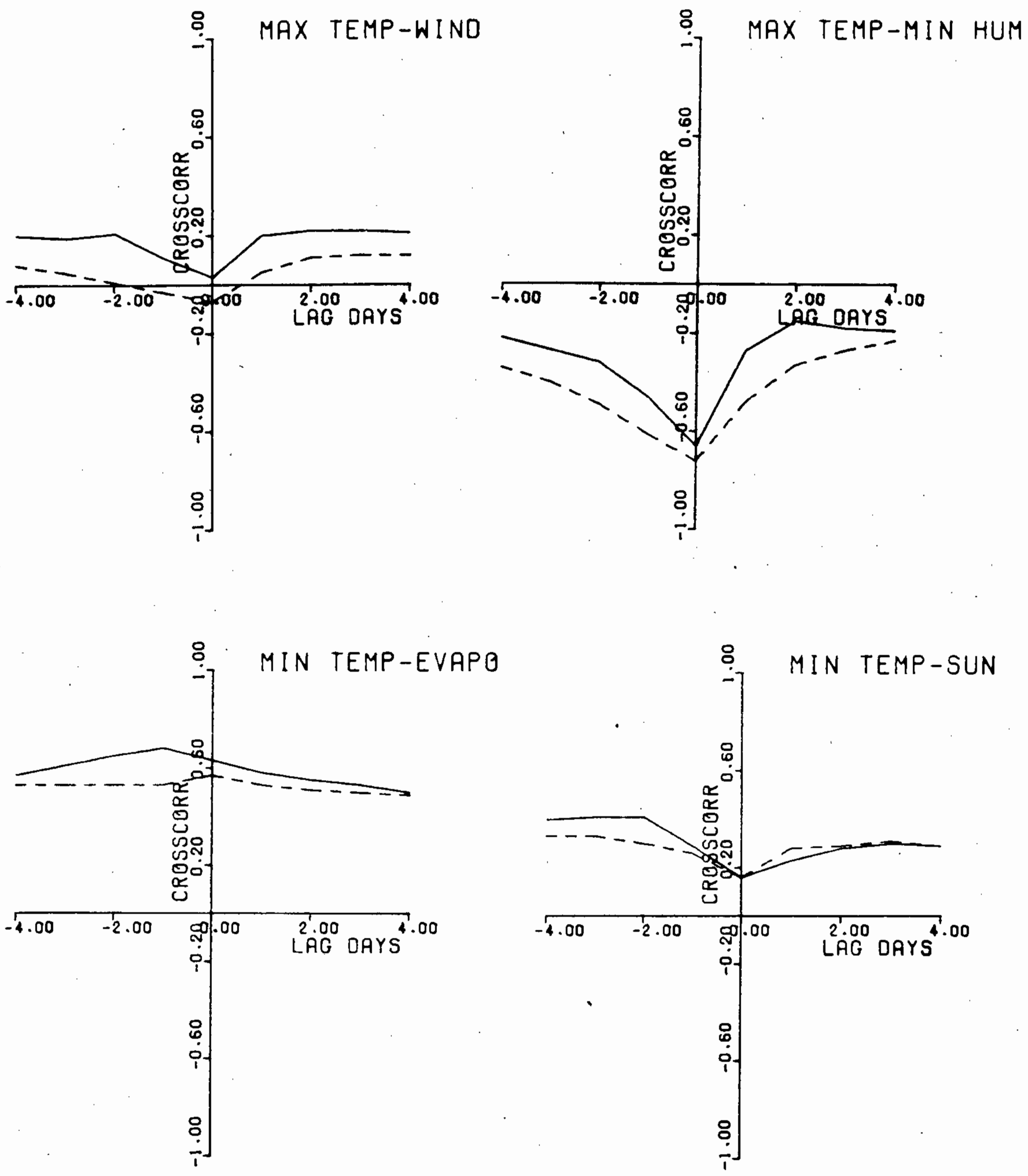
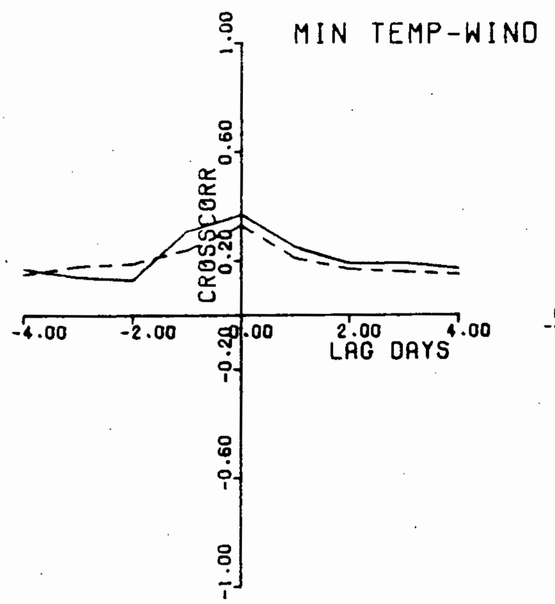
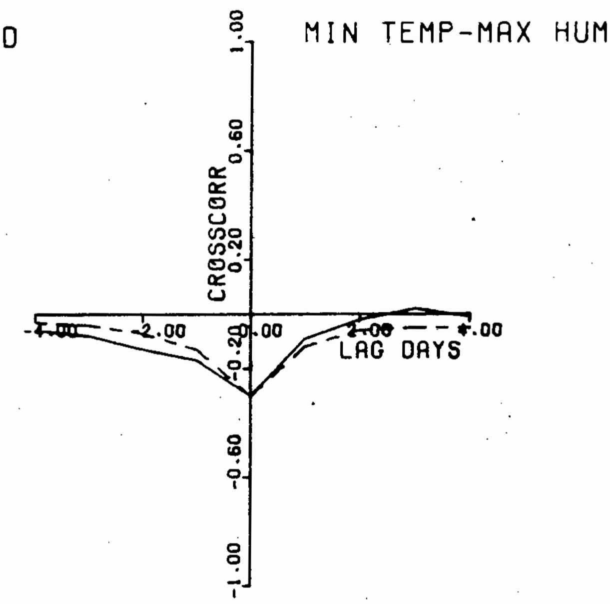

MAX HUM-MIN HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

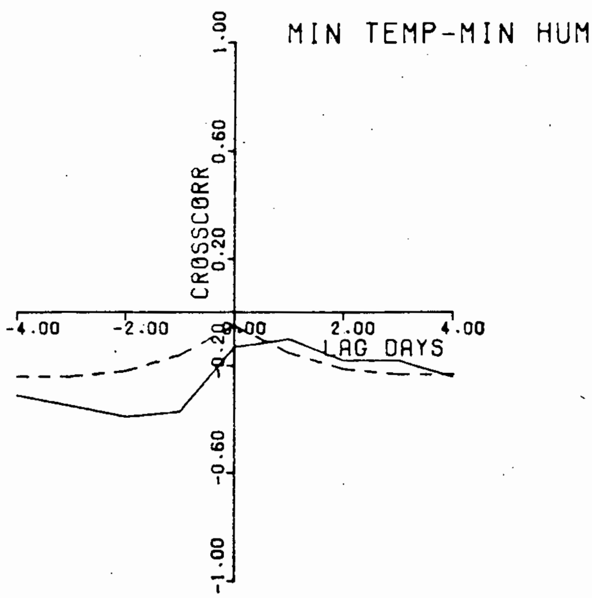
FIGURE 6.46: HISTORICAL (——) AND SIMULATED (---) CROSS-CORRELATION COEFFICIENTS GIVEN A DRY DAY

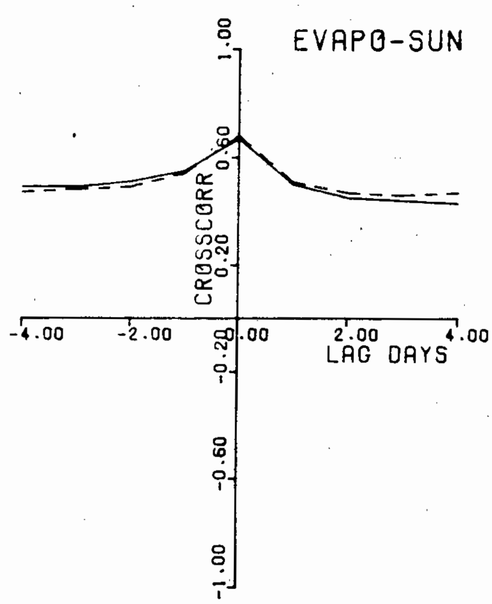


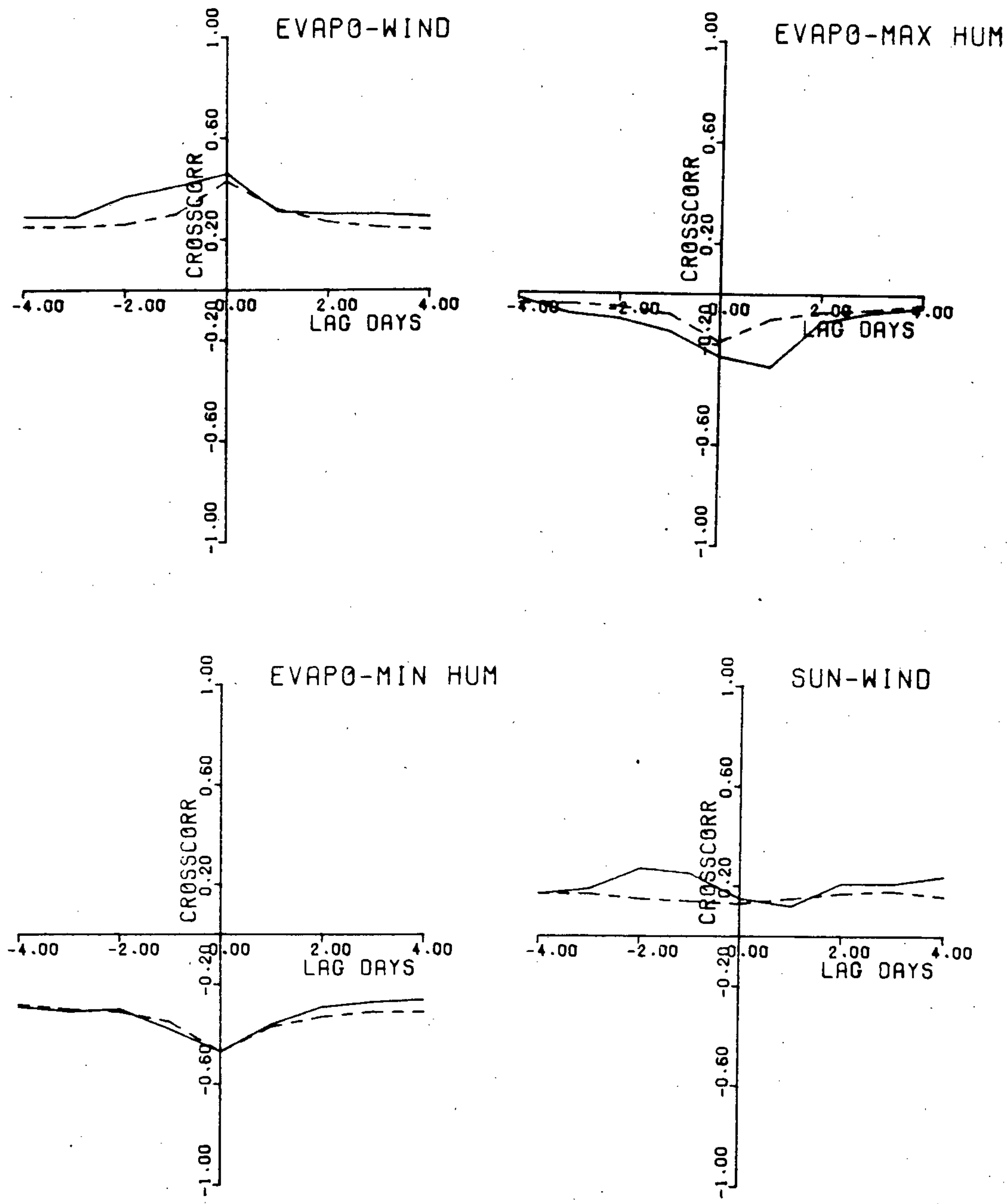

MAX TEMP-WIND
MAX TEMP-MIN HUM
MIN TEMP-EVAPO
MIN TEMP-SUN
CROSSCORR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00

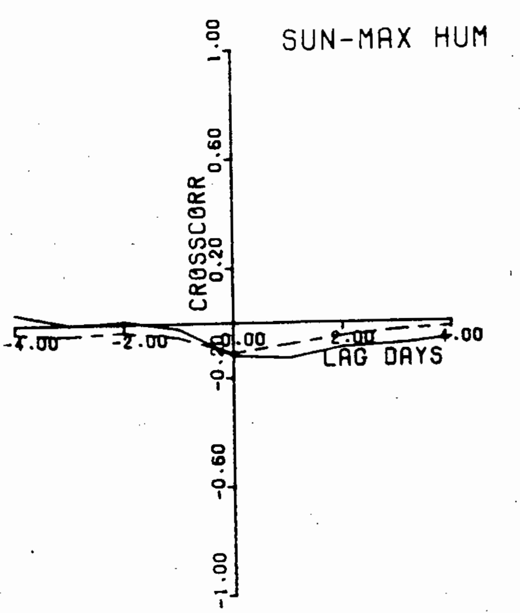

MIN TEMP-WIND
CROSSCORR
LAG DAYS

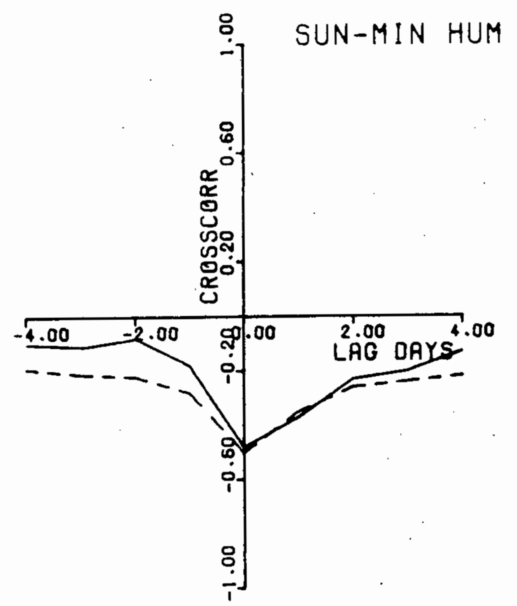

MIN TEMP-MAX HUM
CROSSCORR
LAG DAYS

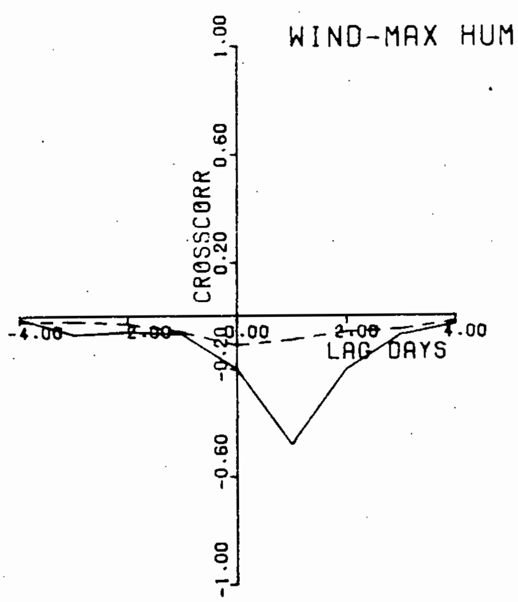

MIN TEMP-MIN HUM
CROSSCORR
LAG DAYS

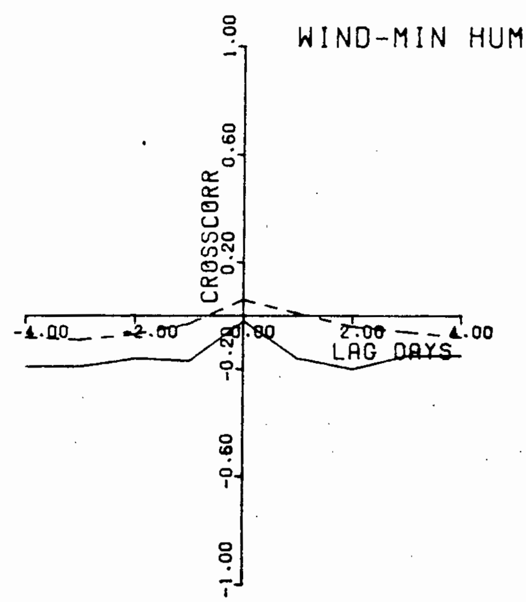

EVAPO-SUN
CROSSCORR
LAG DAYS

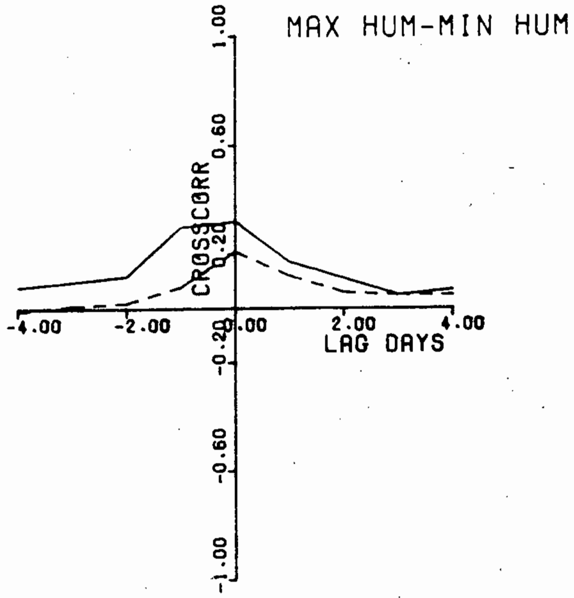

EVAPO-WIND
EVAPO-MAX HUM
EVAPO-MIN HUM
SUN-WIND
CROSSCORR
LAG DAYS
-4.00
-2.00
0.00
2.00
4.00
-1.00
-0.60
-0.20
0.20
0.60
1.00


SUN-MAX HUM
CROSSCORR
LAG DAYS


SUN-MIN HUM
CROSSCORR
LAG DAYS


WIND-MAX HUM
CROSSCORR
LAG DAYS


WIND-MIN HUM
CROSSCORR
LAG DAYS


MAX HUM-MIN HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

6.2.4 Validation of Climate Model 3 (LS).

(a) Validation of annual properties.

The annual mean and standard deviation of the simulated sequence closely resembles those of the historical record. The comparison is given in Table 6.24. The same comparison was carried out on the conditioned sequences. Again, the annual mean and standard deviation are well preserved by the model. (Tables 6.25 - 6.26.)

The comparison of historical and simulated extreme values is shown in Table 6.27. Once more it is found that the maximum values are preserved by the model for most of the variables, but not the minimum values, although only maximum temperature shows a significant number of values lying below the observed minimum value (Table 6.28). Maximum humidity has twelve percent of its values above the maximum observed value, but this is a relatively few number seeing that one is here dealing with a complex model and relatively few historical observations were used in model estimation.

TABLE 6.24: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MEANS AND STANDARD DEVIATION

Variable	Historical		Simulated	
	Mean	Std. Dev.	Mean	Std. Dev.
Max Temp	226.4	58.90	203.6	56.91
Min Temp	106.1	38.34	107.0	37.88
Evapo	57.30	37.01	56.75	36.44
Sunshine	82.61	36.74	81.56	36.13
Windrun	1966	911.2	1943	888.3
Max Hum	914.5	96.29	915.1	93.73
Min Hum	412.7	152.8	425.3	149.0

TABLE 6.25: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MEAN AND STANDARD DEVIATION GIVEN A WET DAY

Variable	Historical		Simulated	
	Mean	Std. Dev.	Mean	Std. Dev.
Max Temp	182.1	46.73	174.8	45.22
Min Temp	110.1	33.43	109.5	33.80
Evapo	28.49	26.99	28.58	26.46
Sunshine	41.71	33.75	41.91	33.43
Windrun	2500	1194	2387	1132
Max Hum	933.0	72.90	929.4	68.82
Min Hum	548.9	157.7	549.0	157.4

TABLE 6.26: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MEAN AND STANDARD DEVIATION GIVEN A DRY DAY

Variable	Historical		Simulated	
	Mean	Std. Dev.	Mean	Std. Dev.
Max Temp	241.4	54.95	212.9	57.17
Min Temp	104.8	39.77	106.3	39.07
Evapo	66.48	35.00	65.77	34.55
Sunshine	96.16	26.12	94.26	26.46
Windrun	1788	710.4	1801	740.4
Max Hum	908.3	102.3	910.5	99.98
Min Hum	366.7	120.3	385.7	122.2

TABLE 6.27: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MAXIMUM AND MINIMUM VALUES

Variable	Historical		Simulated	
	Maximum	Minimum	Maximum	Minimum
Max Temp	408.0	100.0	413.5	23.30
Min Temp	238.0	13.00	234.2	-18.20
Evapo	185.0	0.00	159.8	-45.76
Sunshine	133.0	0.00	188.9	-73.97
Windrun	7376	293.0	6184	-1737
Max Hum	1000	280.0	1265	538.3
Min Hum	950.0	120.0	1100	-123.0

TABLE 6.28: NUMBER OF TIMES SIMULATED VALUES LIE OUTSIDE A DEVIANCE FROM THE HISTORICAL MAXIMUM OR MINIMUM VALUE

Variable	Greater than Maximum	Less than Minimum
Max Temp	14 (0.2%)	1974 (27%)
Min Temp	0 (0%)	20 (0.3%)
Evapo	0 (0%)	399 (5%)
Sunshine	94 (1%)	205 (3%)
Windrun	0 (0%)	152 (2%)
Max Hum	897 (12%)	0 (0%)
Min Hum	1 (0%)	63 (0.9%)

Tables 6.29 - 6.30 show the comparison of historical and simulated annual maximum and minimum values given the status of the day.

TABLE 6.29: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MAXIMUM AND MINIMUM VALUES GIVEN A WET DAY

Variable	Historical		Simulated	
	Maximum	Minimum	Maximum	Minimum
Max Temp	365.0	100.0	349.2	23.30
Min Temp	238.0	23.00	221.2	19.30
Evapo	122.0	0.00	116.6	-45.76
Sunshine	124.0	0.00	138.4	-73.97
Windrun	7376	618.0	6184	-1737
Max Hum	1000	340.0	1206	608.6
Min Hum	950.0	170.0	1101	-123.0

The above table shows that the maximum value for maximum humidity is overestimated. The minimum values, except for minimum temperature, have not been preserved by the model.

The maximum values for sunshine duration and maximum humidity are overestimated while that of windrun is underestimated. The minimum values are not preserved by the model. Suggestions to cope with this weakness of the model are given in the summary at the end of this chapter.

TABLE 6.30: COMPARISON OF HISTORICAL AND SIMULATED ANNUAL MAXIMUM AND MINIMUM VALUES GIVEN A DRY DAY

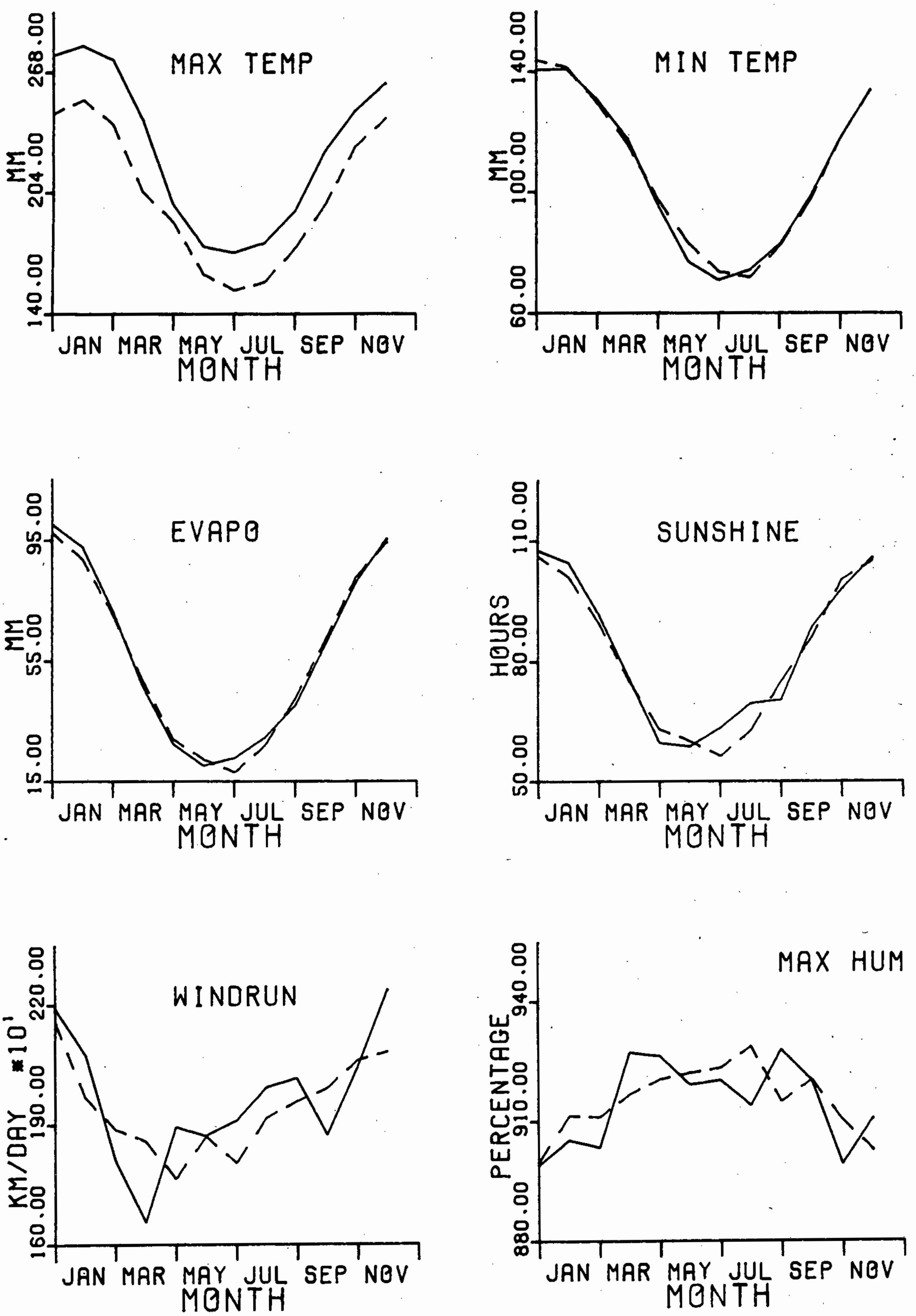
Variable	Historical		Simulated	
	Maximum	Minimum	Maximum	Minimum
Max Temp	408.0	120.0	413.5	24.50
Min Temp	234.0	13.00	234.2	-18.20
Evapo	185.0	0.00	159.8	-33.11
Sunshine	133.0	0.00	188.9	-3.54
Windrun	6804	293.0	4259	-1289
Max Hum	1000	280.0	1265	538.3
Min Hum	850.0	120.0	825.4	-103.0

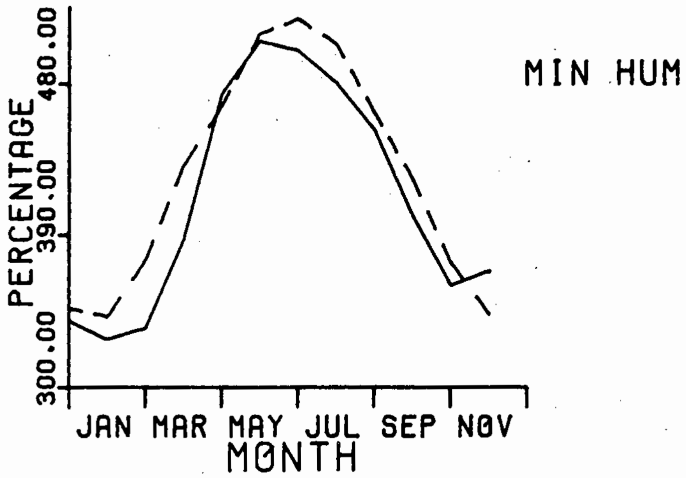
(b) Validation of monthly properties.

The monthly means for each variable of the simulated sequence were compared to those of the historical record (Figure 6.47) and it is concluded that except for maximum temperature, which is slightly underestimated, the model has successfully reproduced the monthly means observed in the historical sequences. The means for March and April of minimum humidity show the largest difference between simulated and historical means but even so this represents only a twelve and sixteen percent difference, respectively. Again a twelve percent difference is observed in the month of April for windrun between the simulated and historical mean. As observed before, this might be indicative of an anomaly in the historical record rather than a weakness in the model.

The comparison of historical and simulated monthly standard deviations

FIGURE 6.47: HISTORICAL (——) AND SIMULATED (---) MONTHLY MEANS FOR ALL VARIABLES




PERCENTAGE
300.00
390.00
480.00
JAN MAR MAY JUL SEP NOV
MONTH
MIN HUM

for each variable is given in Figure 6.48.

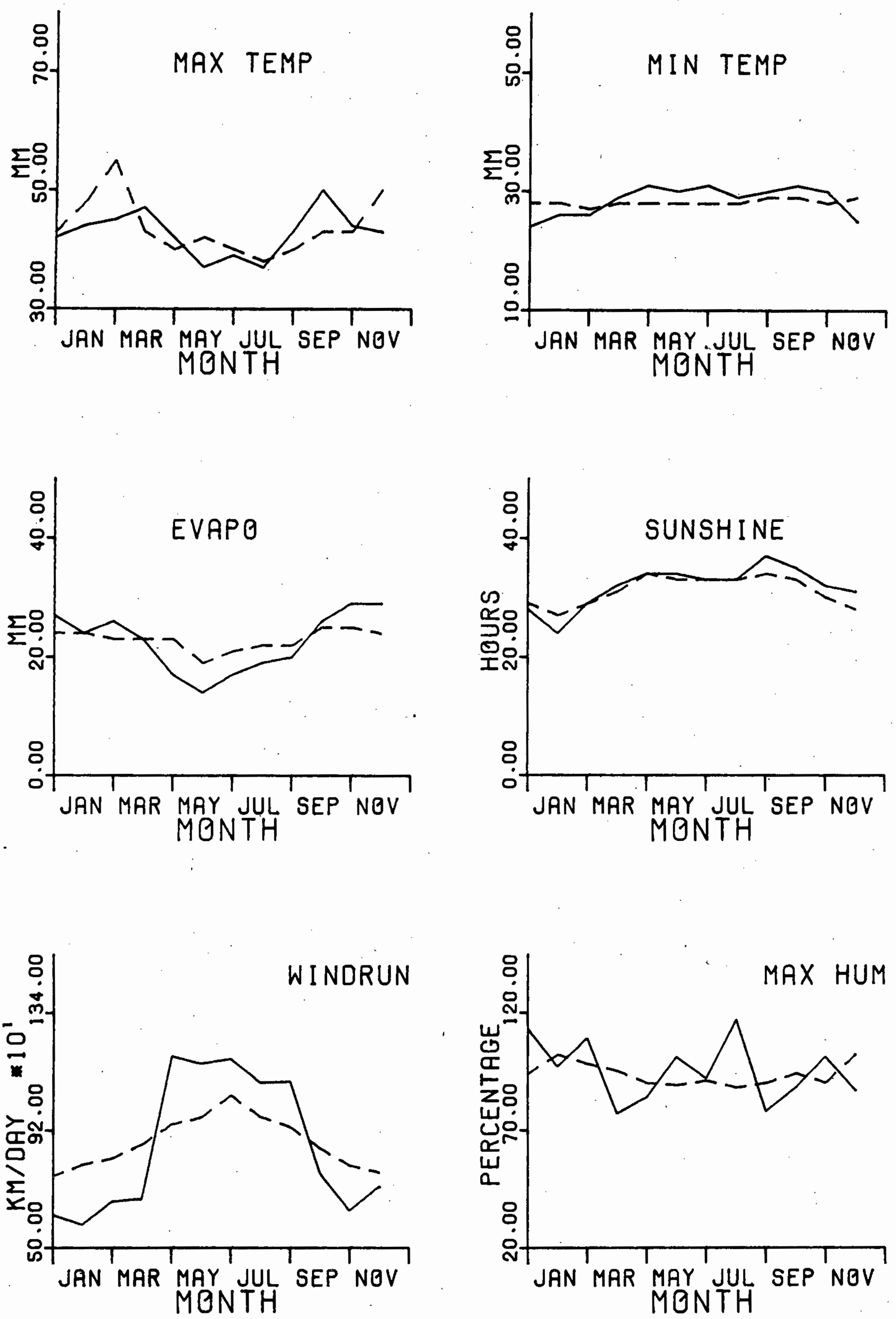
The model is successful in preserving the monthly standard deviations for some variables and for those variables that it does not perform so well it only does so for some months. These variables are evaporation (the months of May to July showing significant differences in the standard deviations), maximum humidity (April and August with significant differences), minimum humidity (July showing a marked difference between the sequences) and windrun, which is either overestimated or underestimated for most months.

Figures 6.49 - 6.50 show the comparison of historical and simulated monthly means for the climate variables conditioned on the wet or dry status of the day.

The model underestimates the monthly means for maximum temperature when the climate sequences are conditioned on the dry status of the day. The monthly means for minimum humidity for some months are overestimated. In some months of windrun the model has not preserved the mean. Otherwise the model has preserved the monthly means very well.

Figures 6.51 - 6.52 illustrate the comparison of historical and simulated monthly standard deviations for the climate variables conditioned on the wet or dry status of the day. Again, the variables maximum and minimum humidity and windrun show significant differences between the standard deviations of the simulated and the historical sequences. As noted before, these variables show large differences between the monthly standard deviations in the historical record and

FIGURE 6.48: HISTORICAL (——) AND SIMULATED (---) MONTHLY STANDARD DIVIATIONS FOR ALL VARIABLES



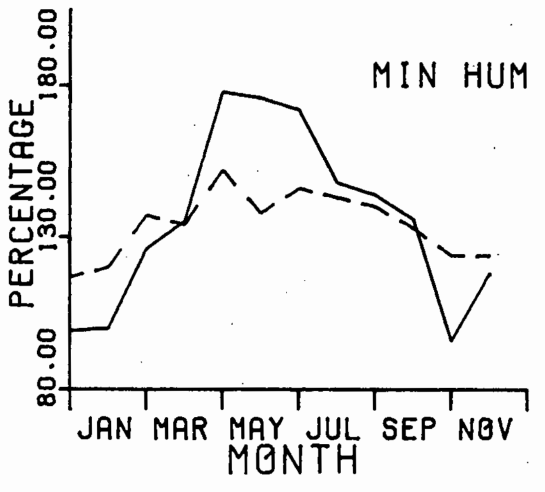
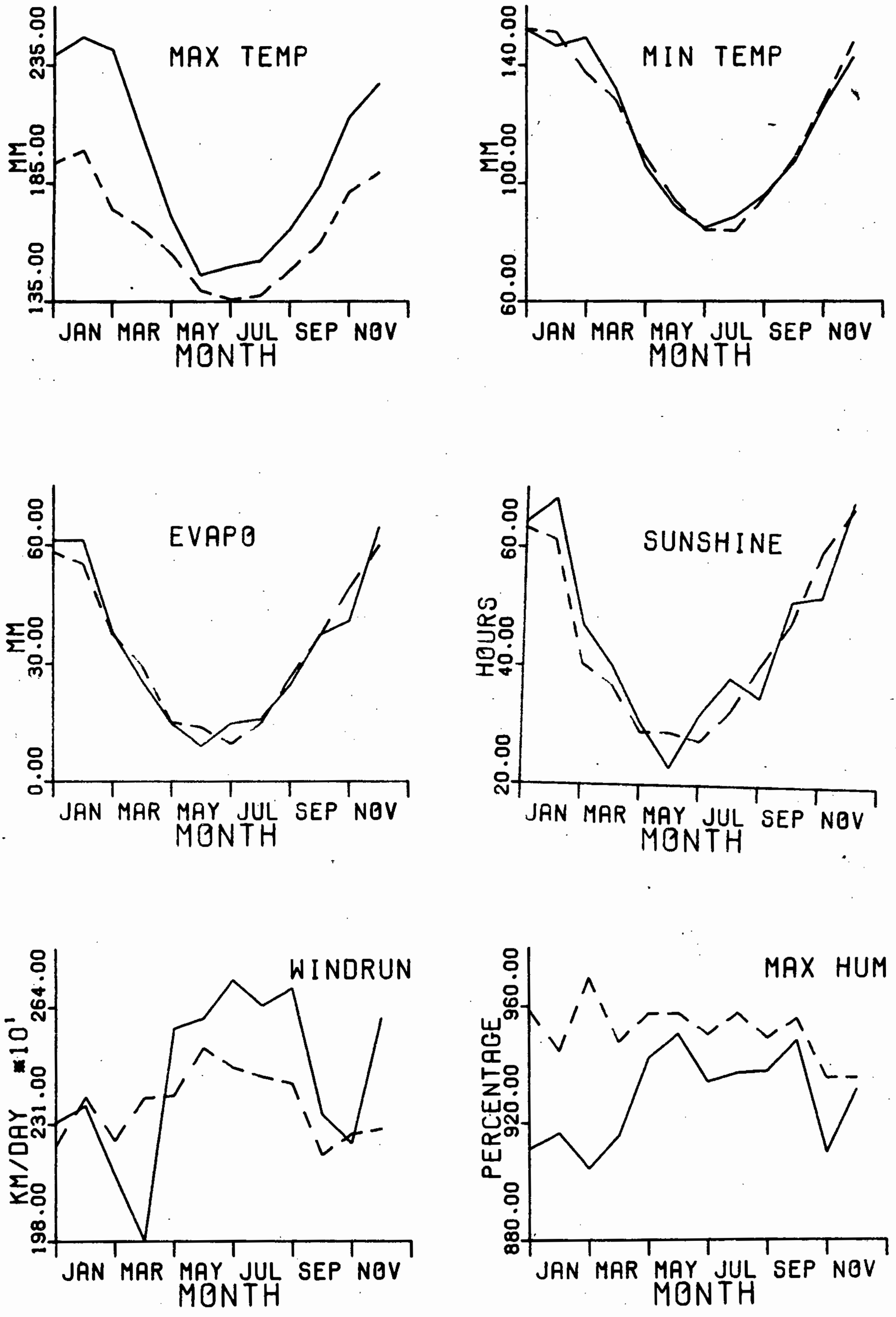

MIN HUM
PERCENTAGE
80.00
130.00
180.00
JAN MAR MAY JUL SEP NOV
MONTH

FIGURE 6.49: HISTORICAL (——) AND SIMULATED (---) MONTHLY MEANS GIVEN A WET DAY FOR ALL VARIABLES



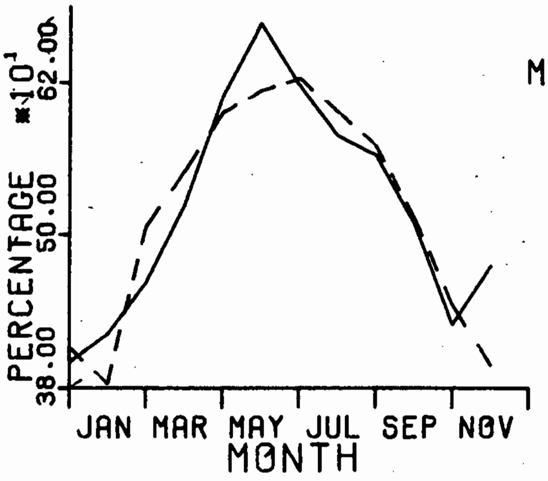
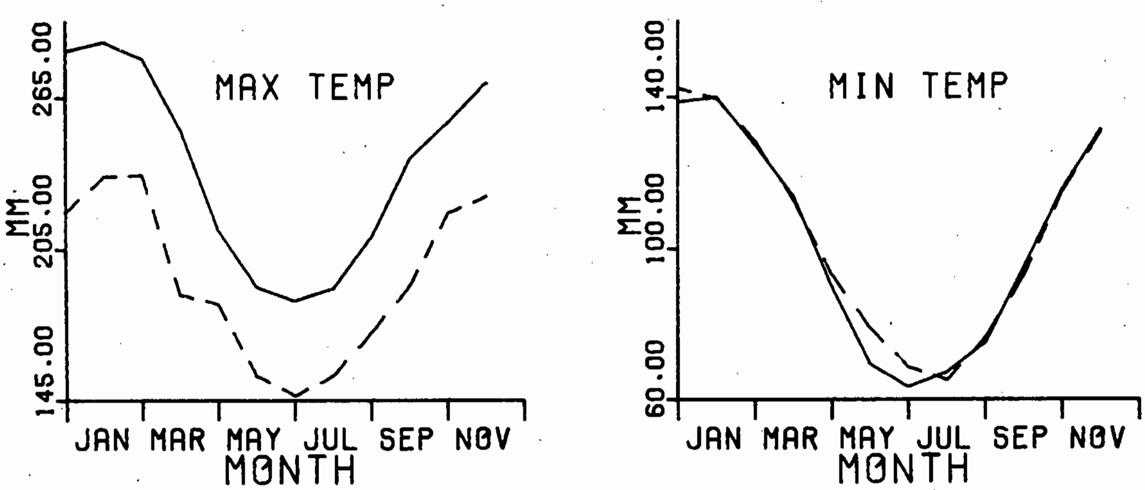
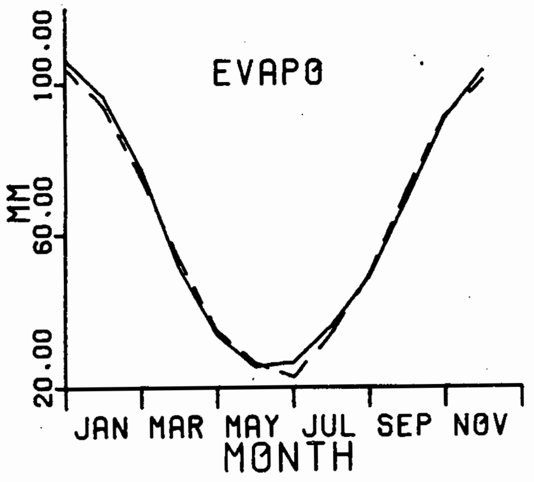
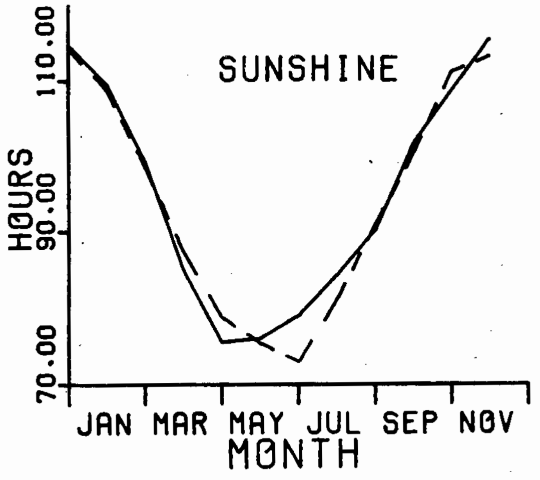
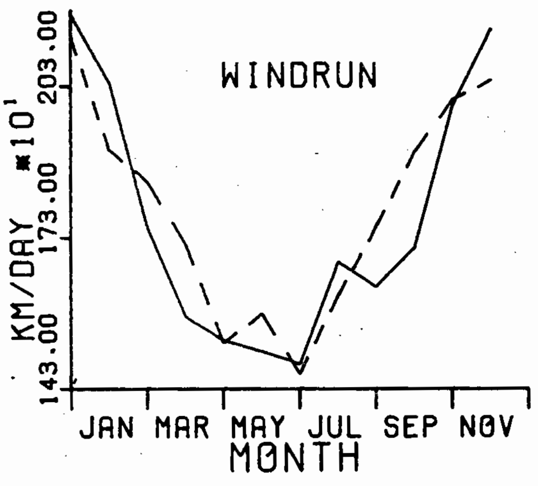


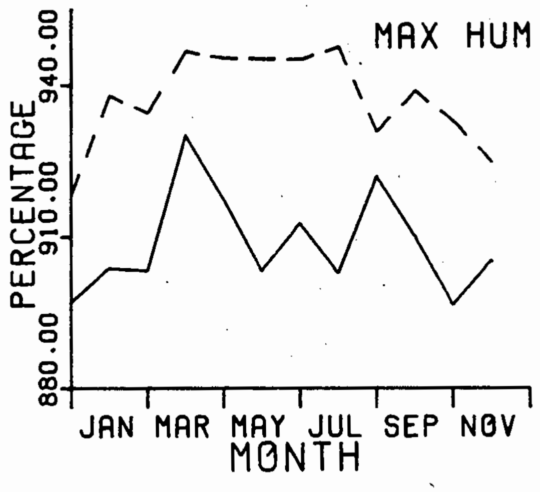
FIGURE 6.50: HISTORICAL (——) AND SIMULATED (---) MONTHLY MEANS GIVEN A DRY DAY FOR ALL VARIABLES




EVAPO
MM
100.00
60.00
20.00
JAN MAR MAY JUL SEP NOV
MONTH


SUNSHINE
HOURS
110.00
90.00
70.00
JAN MAR MAY JUL SEP NOV
MONTH


WINDRUN
KM/DAY *10¹
203.00
173.00
143.00
JAN MAR MAY JUL SEP NOV
MONTH


MAX HUM
PERCENTAGE
940.00
910.00
880.00
JAN MAR MAY JUL SEP NOV
MONTH

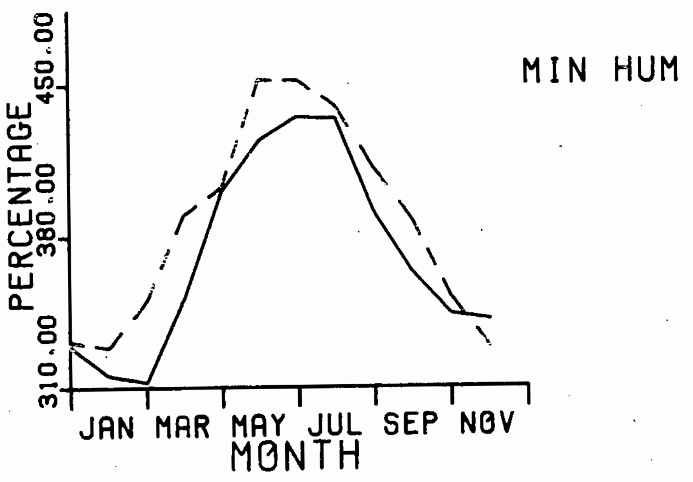
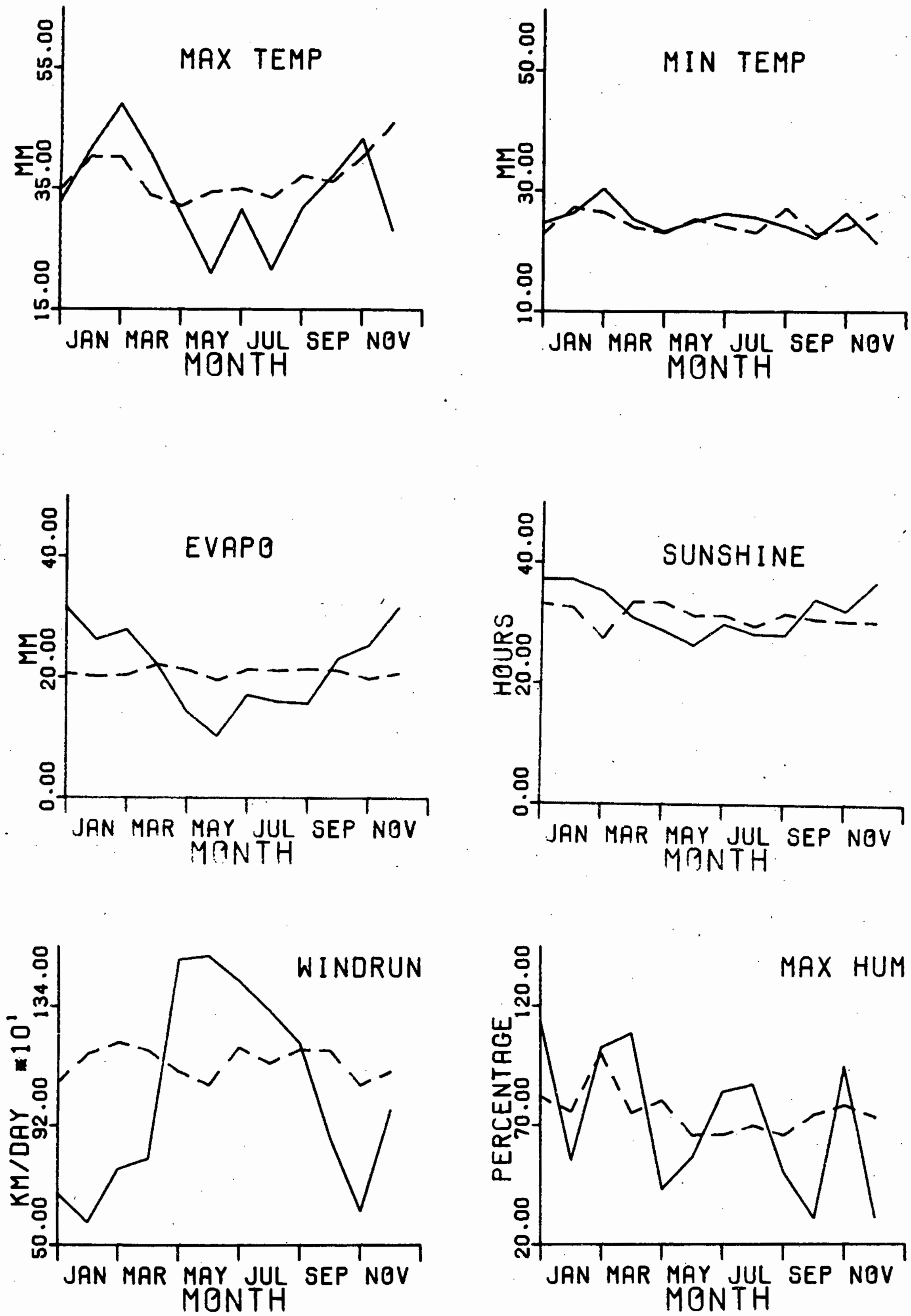

MIN HUM
PERCENTAGE
450.00
380.00
310.00
JAN MAR MAY JUL SEP NOV
MONTH

FIGURE 6.51: HISTORICAL (——) AND SIMULATED (---) MONTHLY STANDARD DEVIATIONS GIVEN A WET DAY FOR ALL VARIABLES



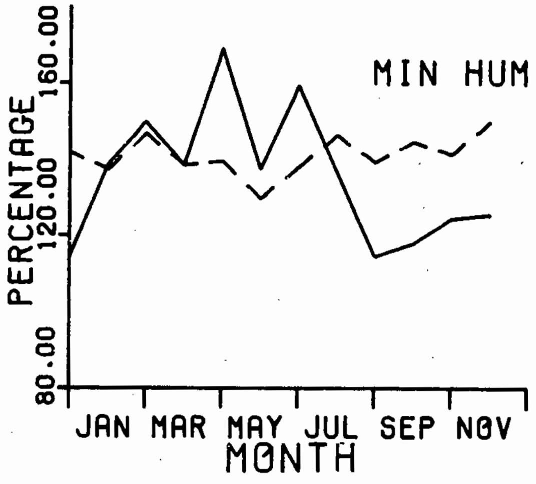
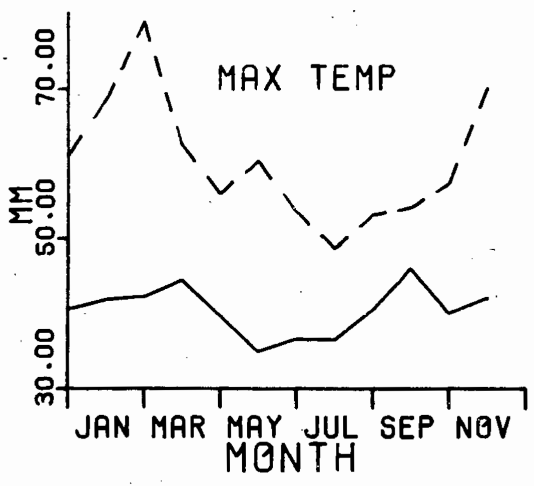
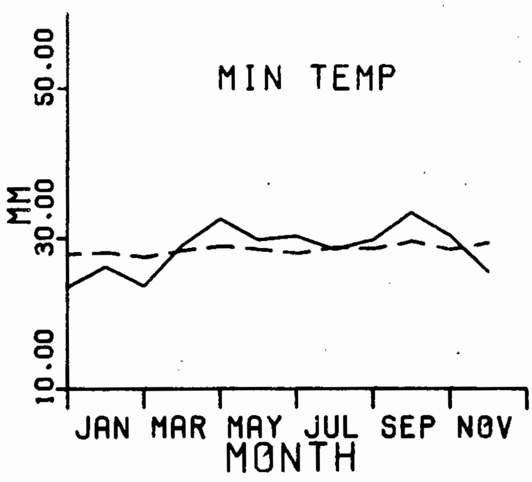
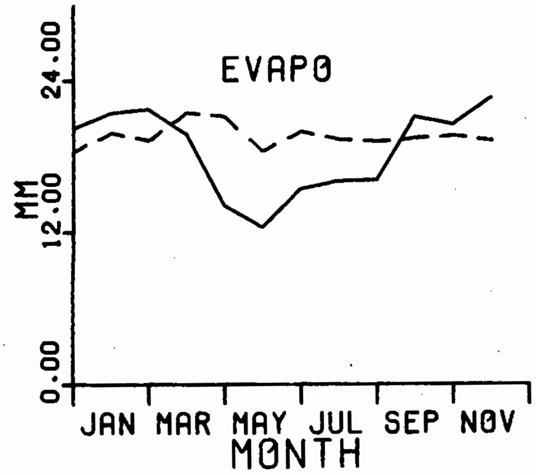
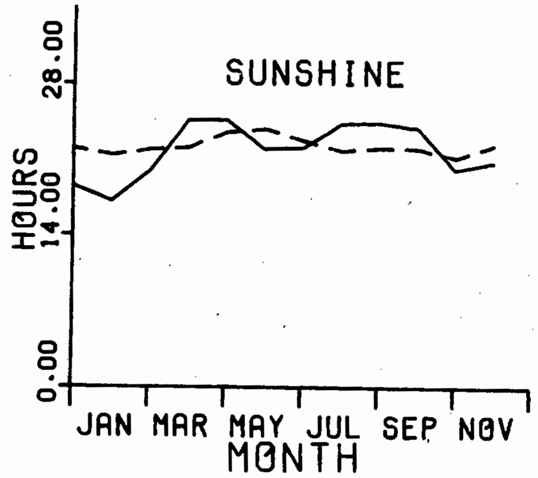


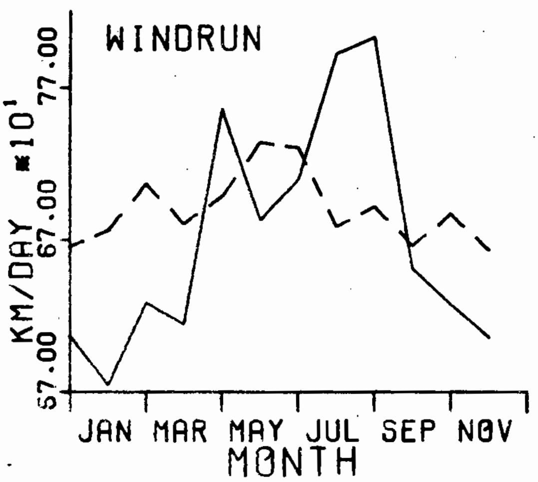
FIGURE 6.52: HISTORICAL (——) AND SIMULATED (---) MONTHLY STANDARD DEVIATIONS GIVEN A DRY DAY FOR ALL VARIABLES

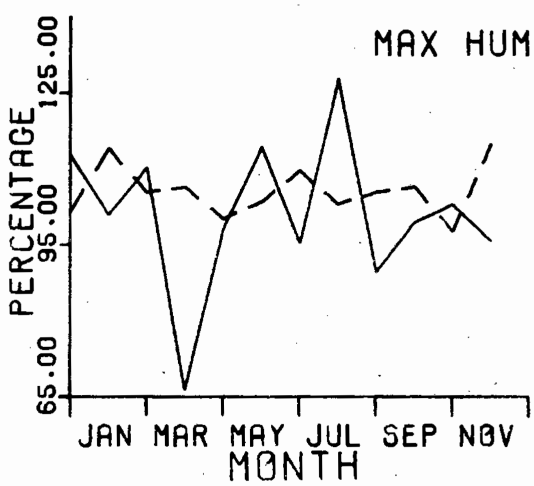


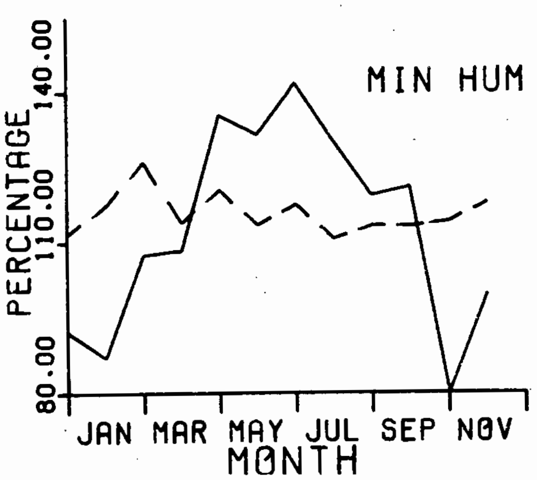



EVAPO
MM
0.00
12.00
24.00
JAN MAR MAY JUL SEP NOV
MONTH


SUNSHINE
HOURS
0.00
14.00
28.00
JAN MAR MAY JUL SEP NOV
MONTH


WINDRUN
KM/DAY *10¹
57.00
67.00
77.00
JAN MAR MAY JUL SEP NOV
MONTH


MAX HUM
PERCENTAGE
65.00
95.00
125.00
JAN MAR MAY JUL SEP NOV
MONTH


MIN HUM
PERCENTAGE
80.00
110.00
140.00
JAN MAR MAY JUL SEP NOV
MONTH

so it is expected that the simulated sequence will show a difference from the historical sequence in standard deviations for some months. That is, the observed differences are due to the scatter in the historical record and not necessarily due to a lack of fit by the model.

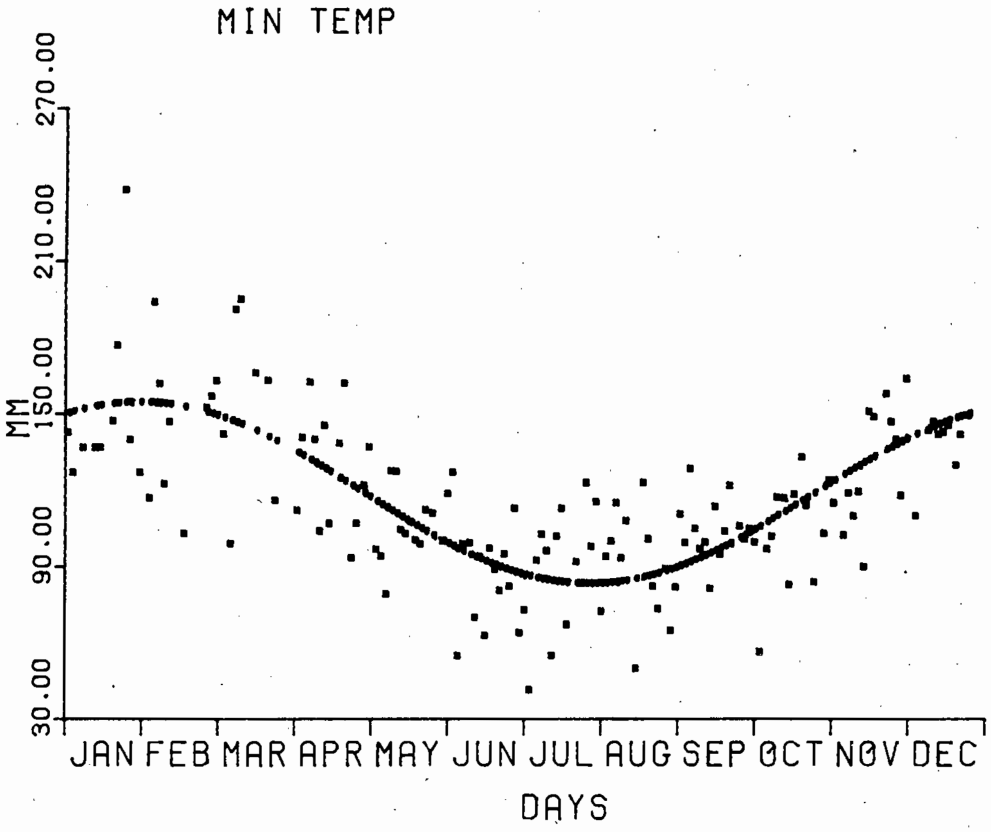
(c) Validation of daily properties.

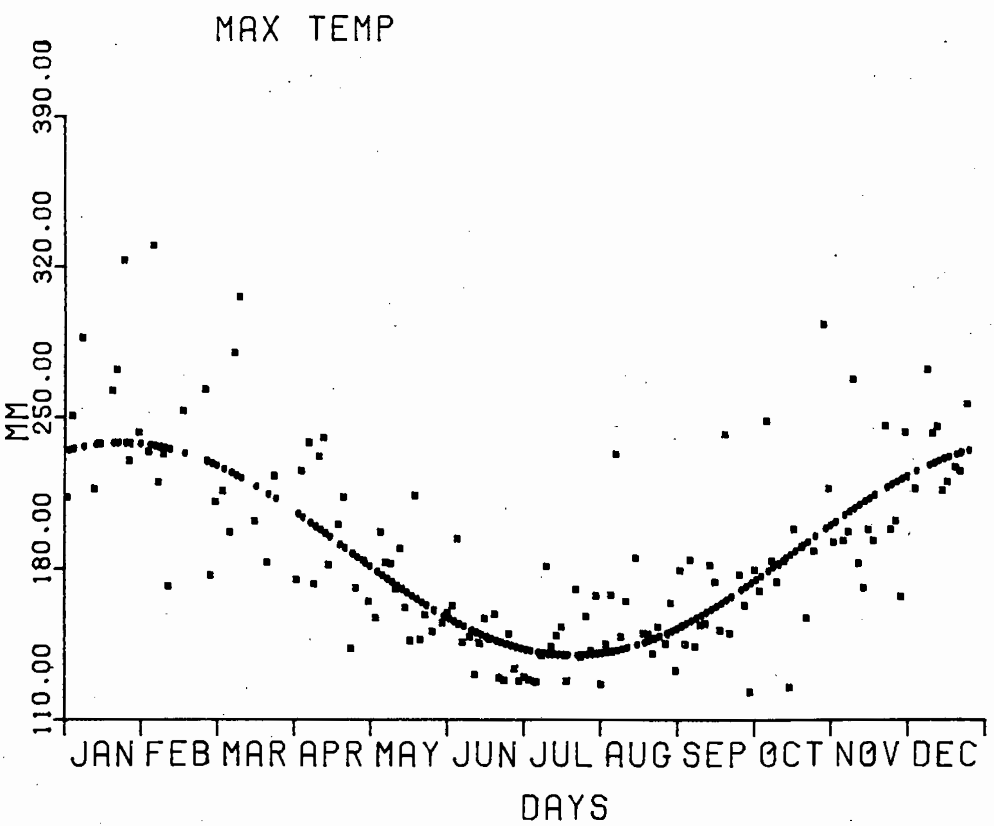
The truncated Fourier series provides a good fit to the mean function of the model. This is true for all climate variables whether the status of the day is wet or dry. The only exception is maximum temperature on a dry day, which is somewhat underestimated. (Figures 6.53 - 6.54.)

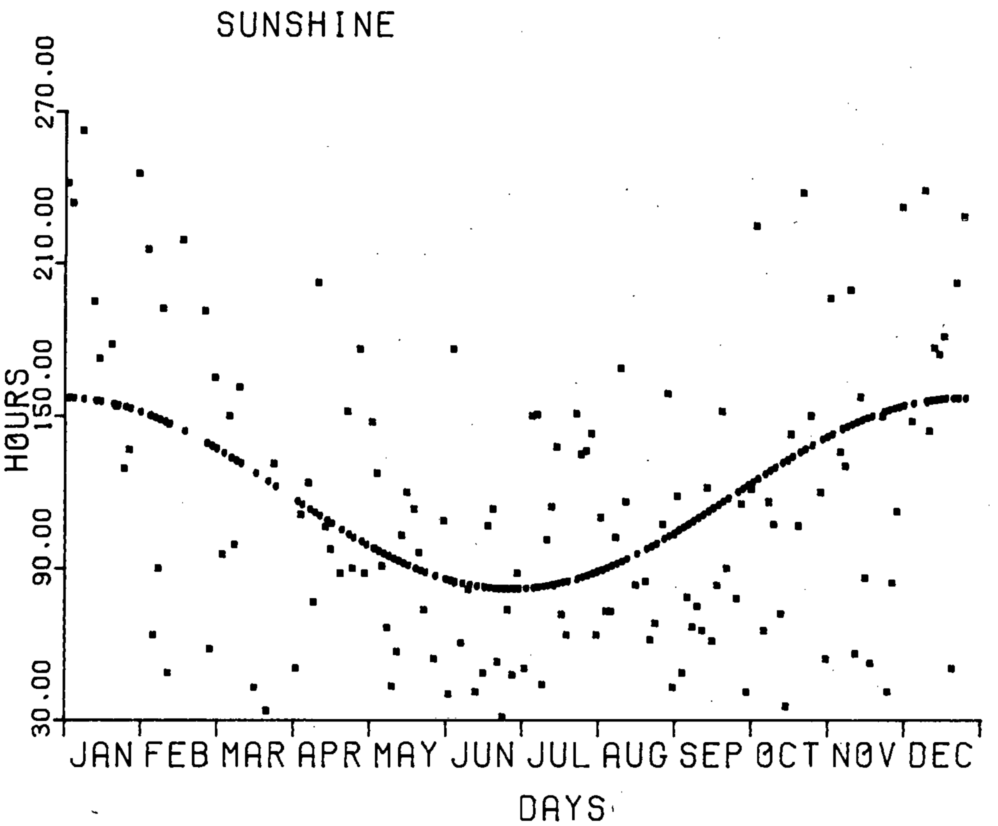
Figure 6.55 shows the comparison between the autocorrelation of variables in the historical data and those of the variables in the simulated data. Figures 6.56 - 6.57 illustrate the comparison between the autocorrelation coefficients of the simulated and historical sequences conditioned on the wet or dry status of the day. In all cases the model has successfully maintained the autocorrelation coefficients.

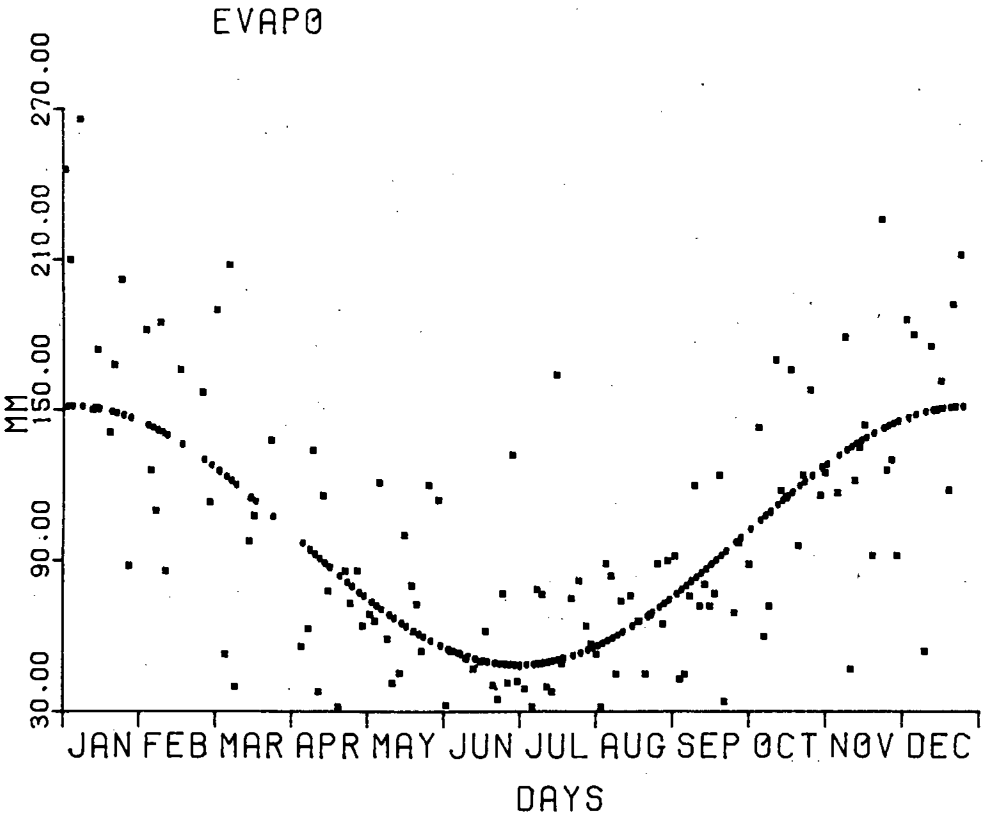
Figure 6.58 shows the comparison of the historical and simulated cross-correlation coefficients for all climate variables. Generally the model has successfully preserved the cross-correlation coefficients. Maximum temperature lagged with minimum temperature is underestimated for a lag of one and two (approximately a difference of 0.3). Maximum temperature lagged with evaporation is slightly underestimated with the largest difference between historical and simulated coefficients being approximately 0.2.

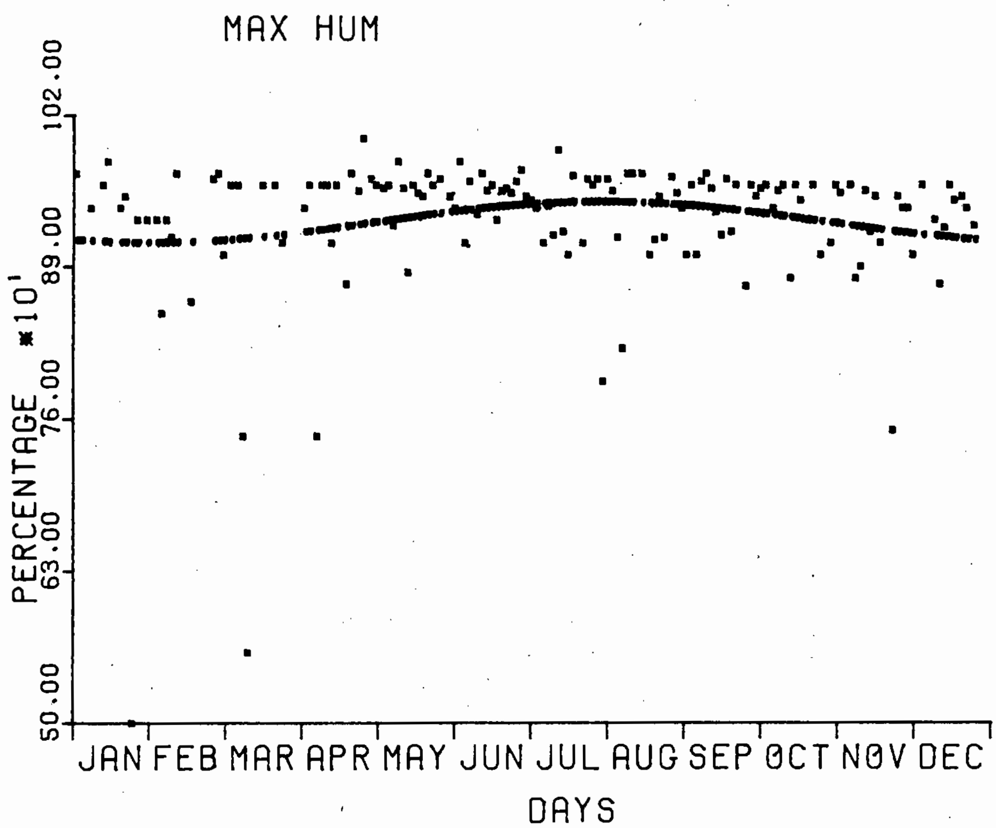
FIGURE 6.53: DAILY AVERAGES AND MEAN FITTED BY A FOURIER SERIES GIVEN A WET DAY FOR ALL VARIABLES

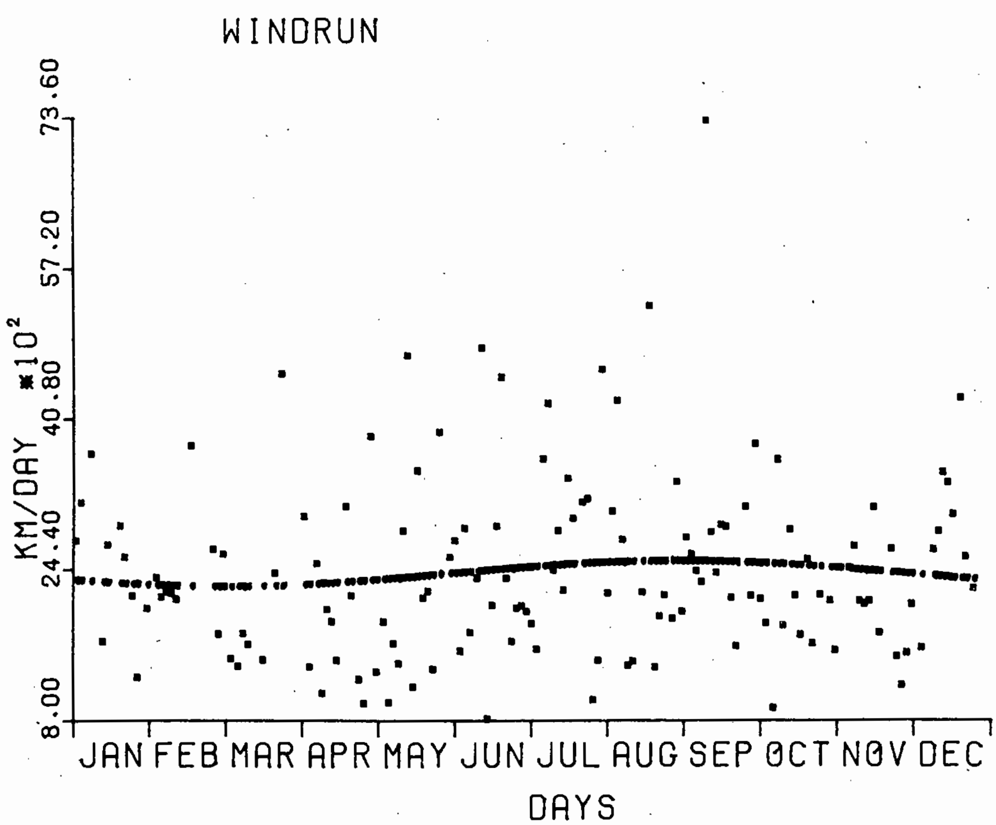





SUNSHINE
HOURS
270.00
210.00
150.00
90.00
30.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


EVAPO
MM
270.00
210.00
150.00
90.00
30.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


MAX HUM
PERCENTAGE *10^1
50.00
63.00
76.00
89.00
102.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


WINDRUN
KM/DAY *10^2
8.00
24.40
40.80
57.20
73.60
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS

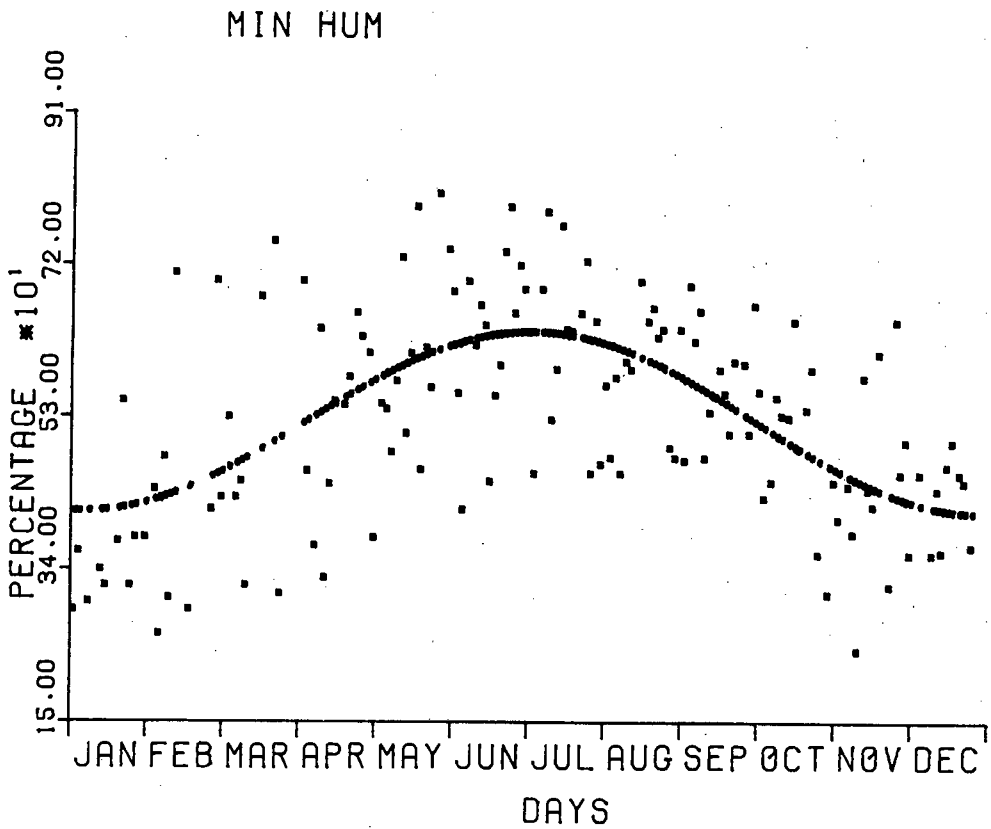
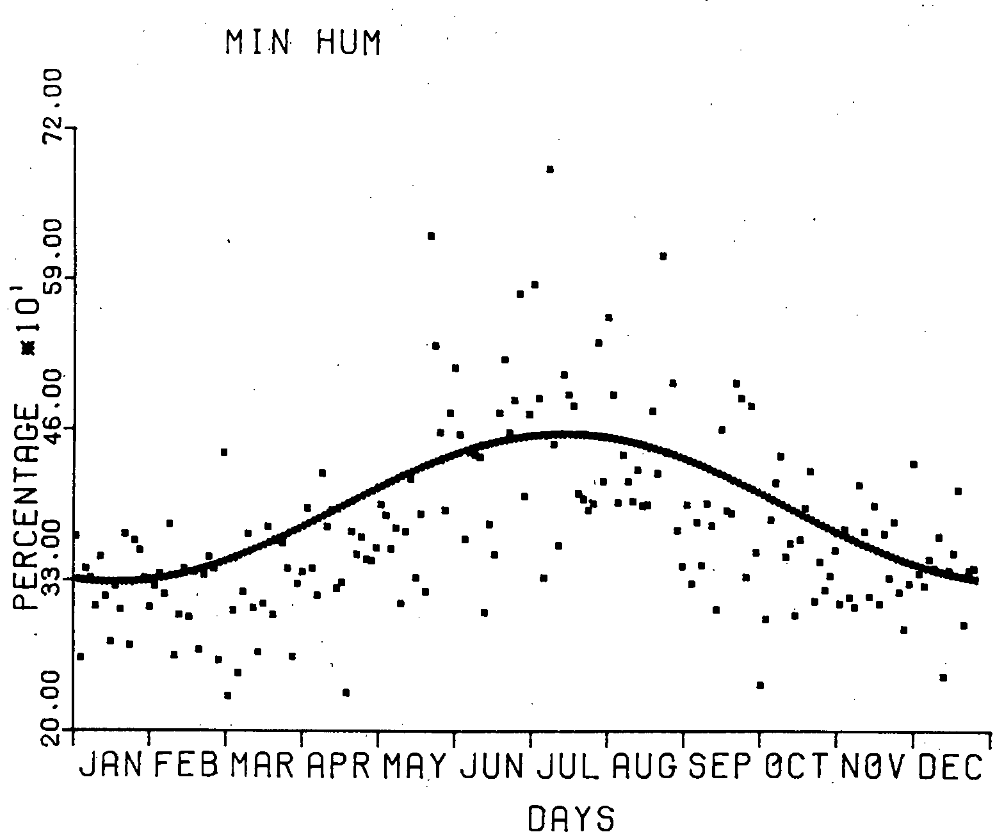
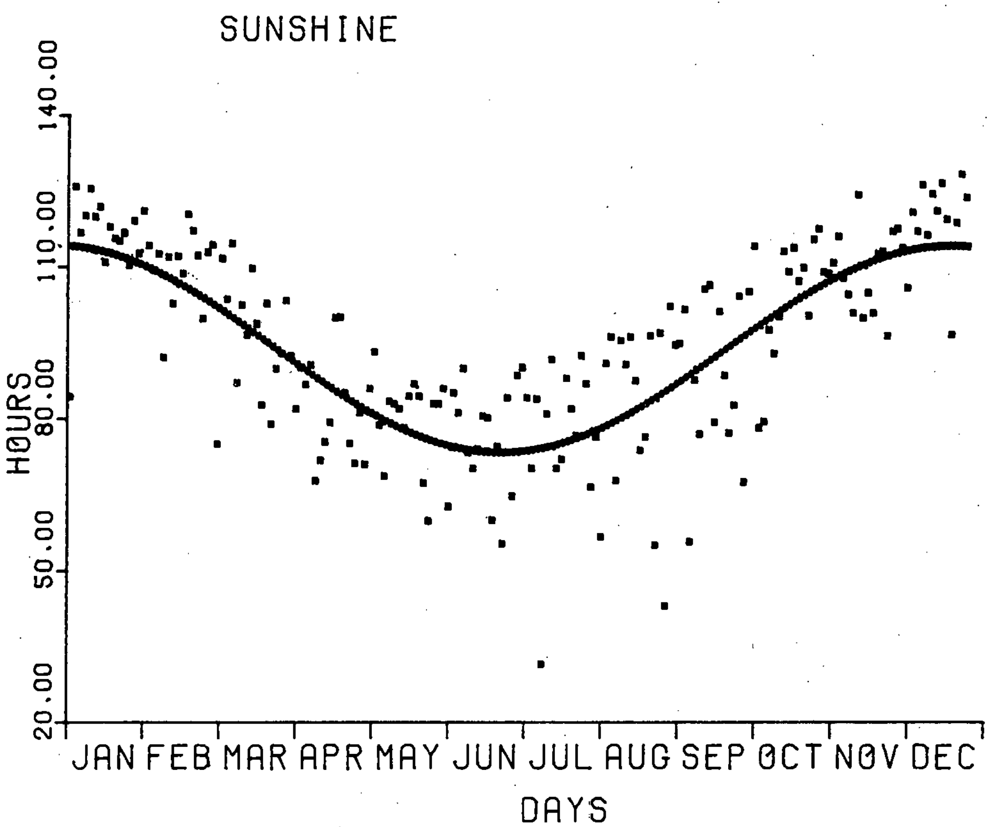
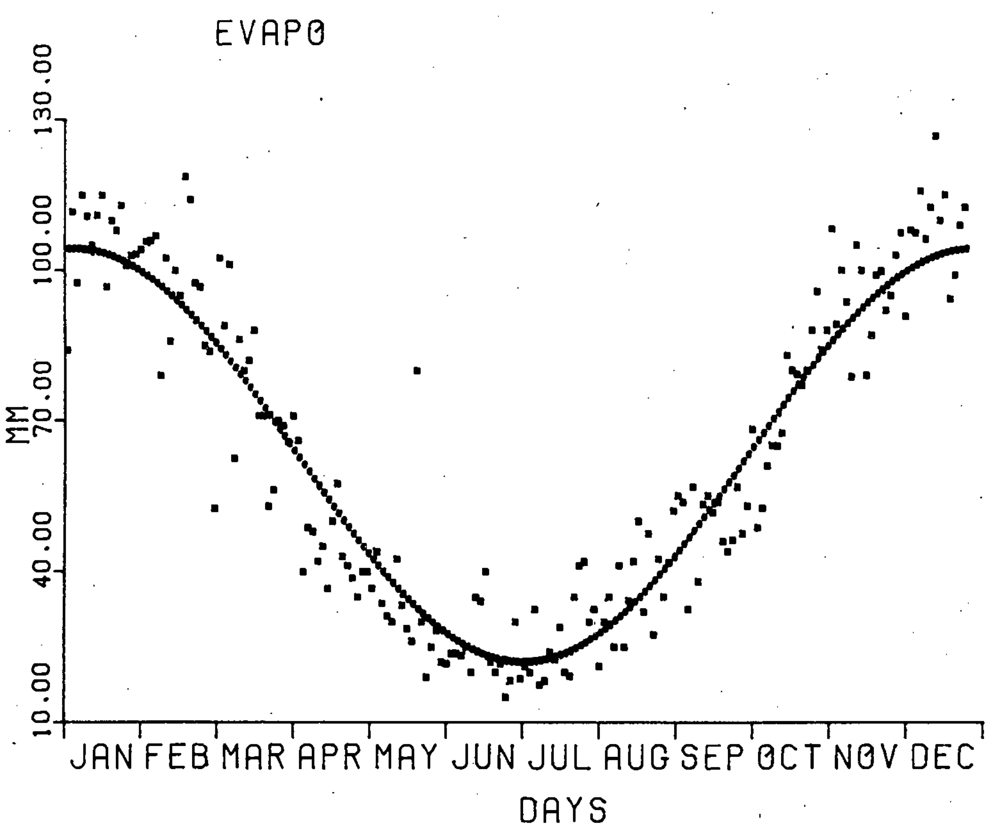
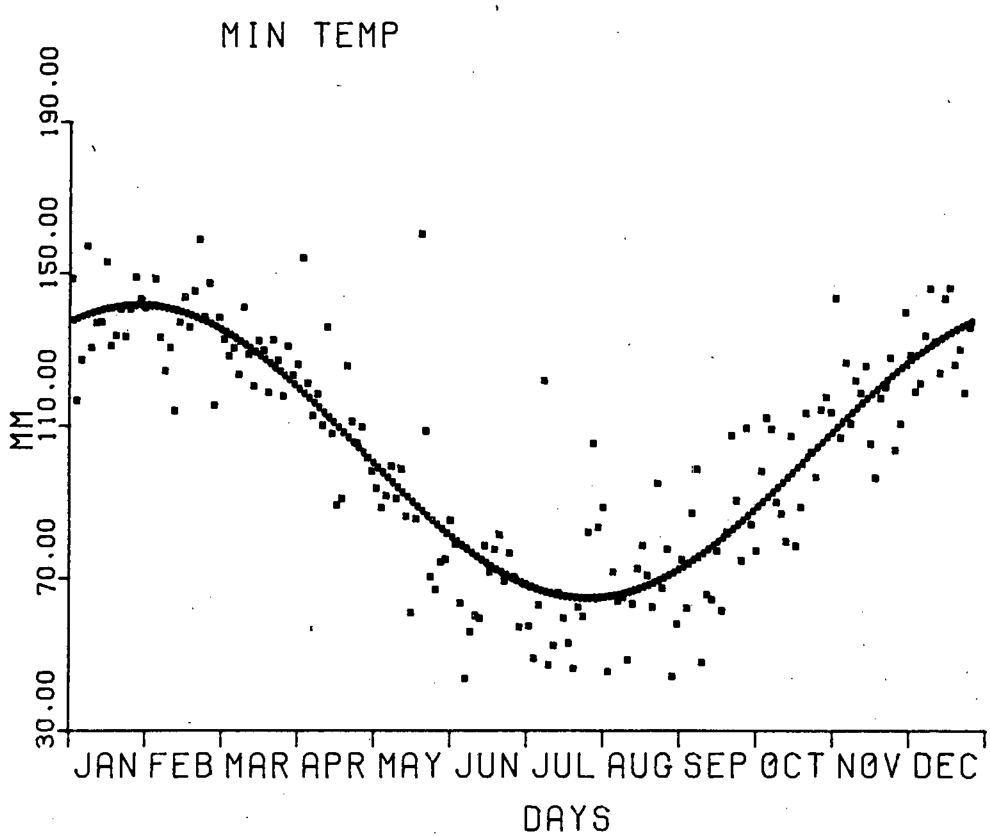


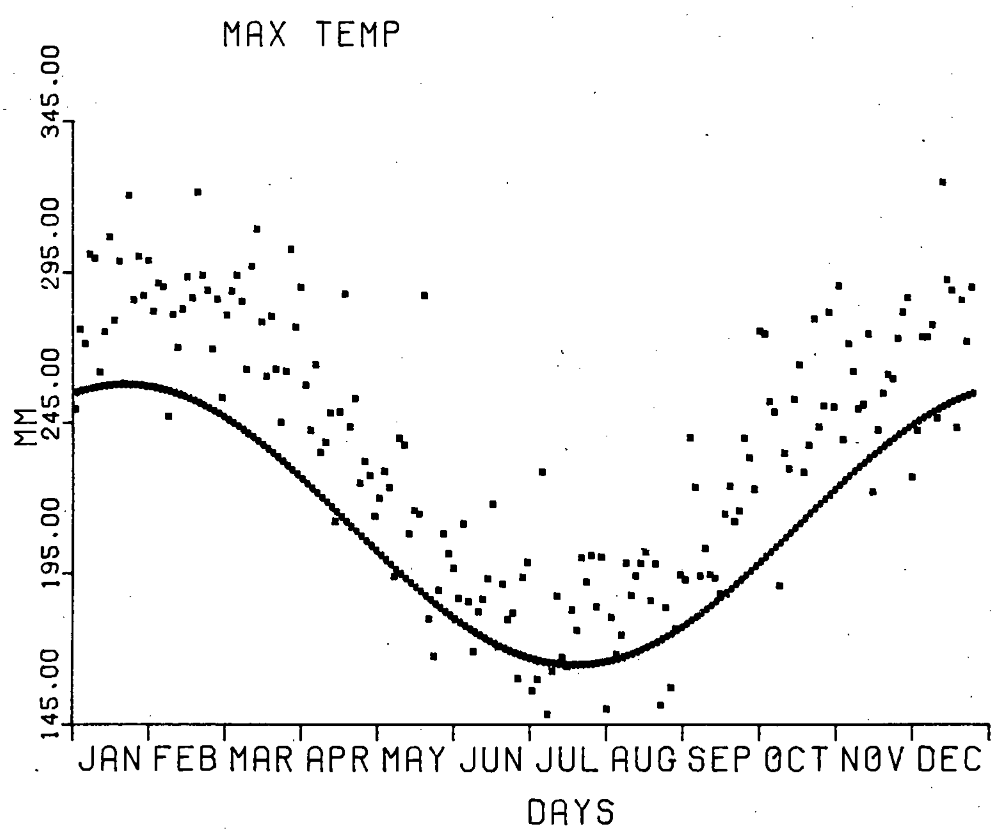
FIGURE 6.54: DAILY AVERAGES AND MEAN FITTED BY A FOURIER SERIES GIVEN A DRY DAY FOR ALL VARIABLES

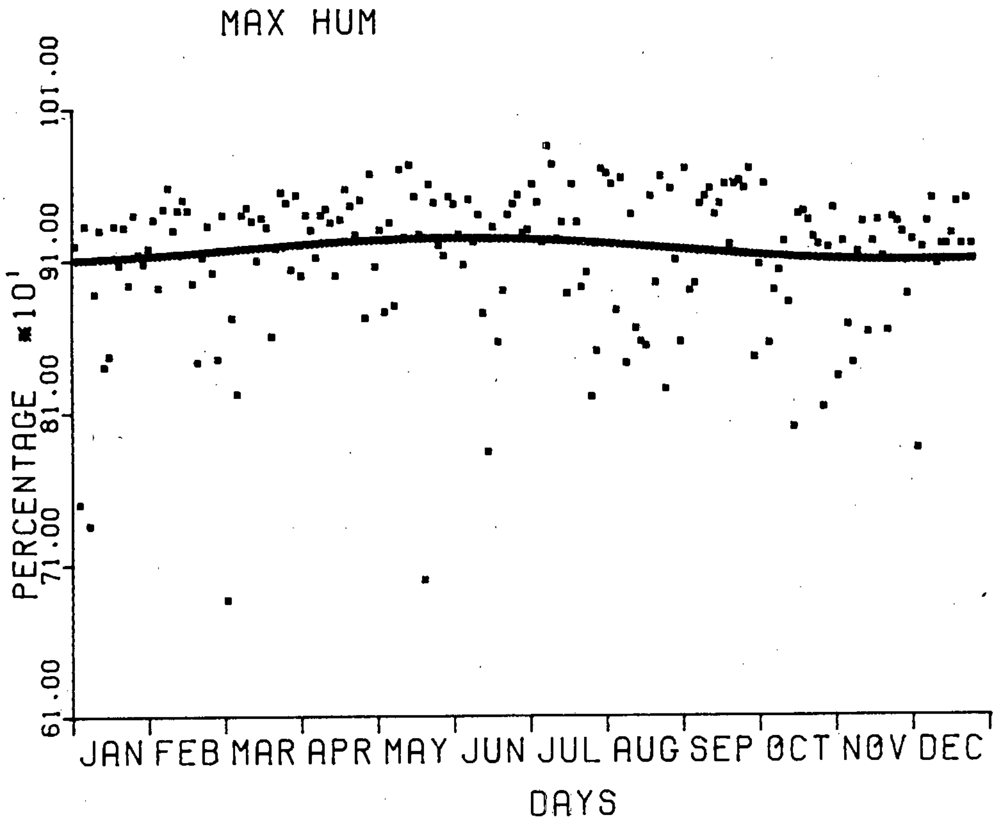



SUNSHINE
HOURS
140.00
110.00
80.00
50.00
20.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


EVAPO
MM
130.00
100.00
70.00
40.00
10.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


MIN TEMP
190.00
150.00
110.00
70.00
30.00
MM
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


MAX TEMP
345.00
295.00
245.00
195.00
145.00
MM
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS


MAX HUM
PERCENTAGE *10¹
101.00
91.00
81.00
71.00
61.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS

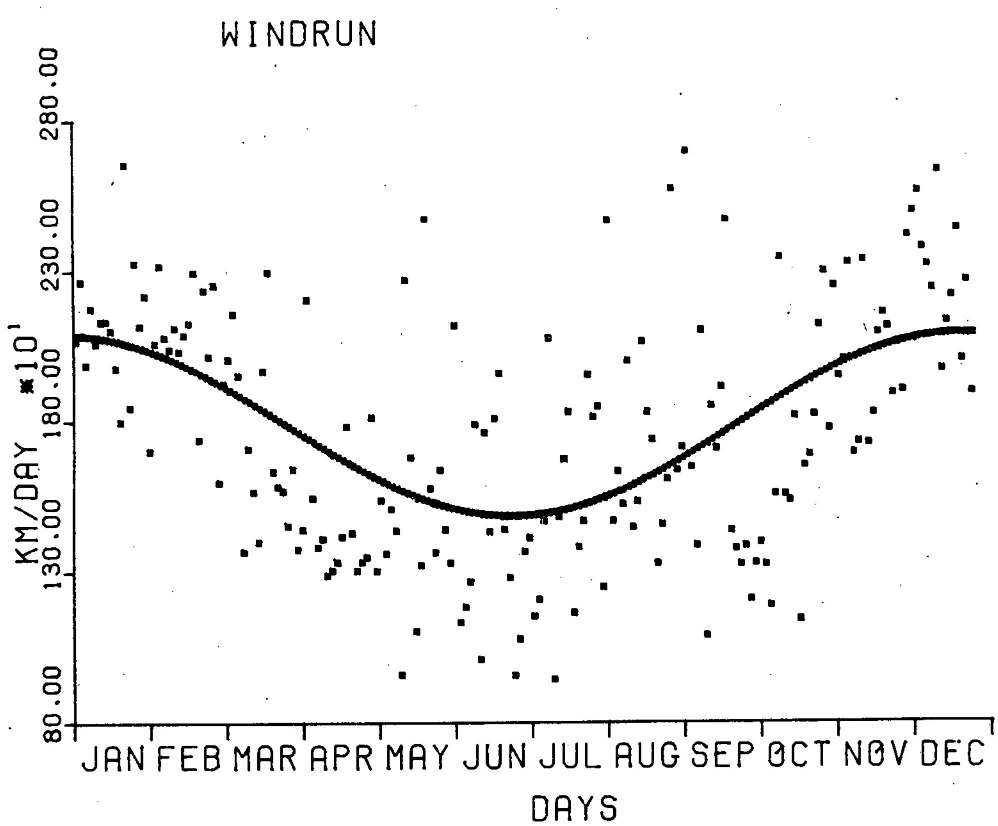
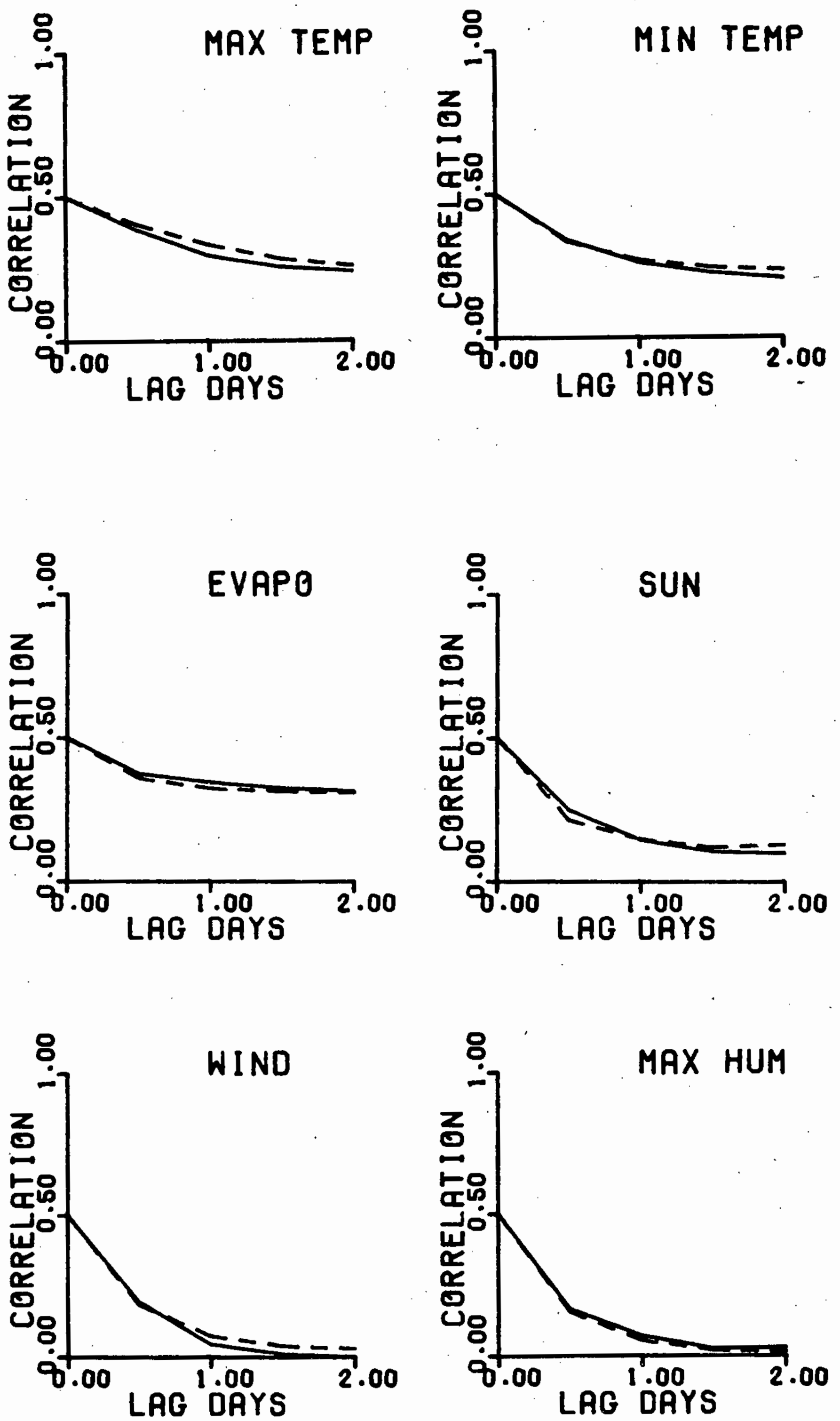

WINDRUN
KM/DAY *10¹
280.00
230.00
180.00
130.00
80.00
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DAYS

FIGURE 6.55: HISTORICAL (——) AND SIMULATED (---) AUTOCORRELATION COEFFICIENTS FOR ALL VARIABLES



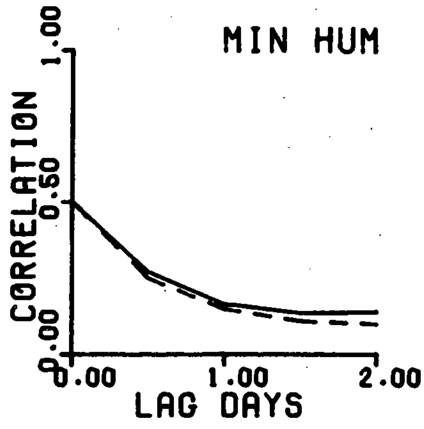
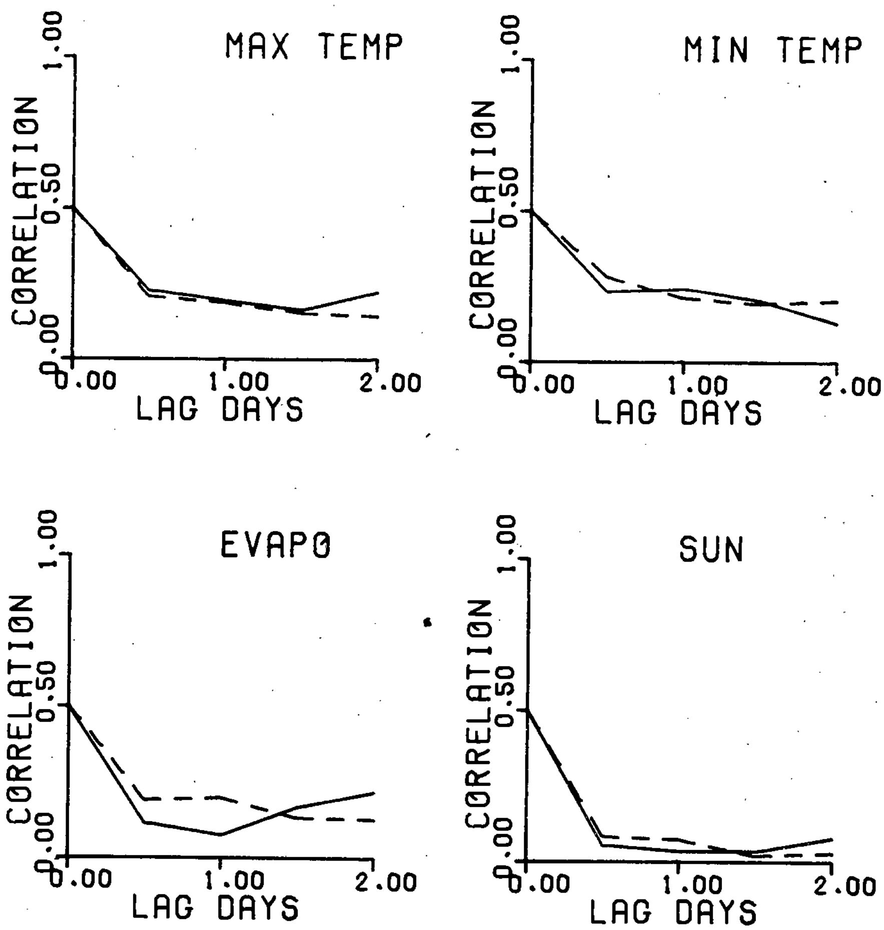


FIGURE 6.56: HISTORICAL (——) AND SIMULATED (---) AUTOCORRELATION COEFFICIENTS GIVEN A WET DAY FOR ALL VARIABLES



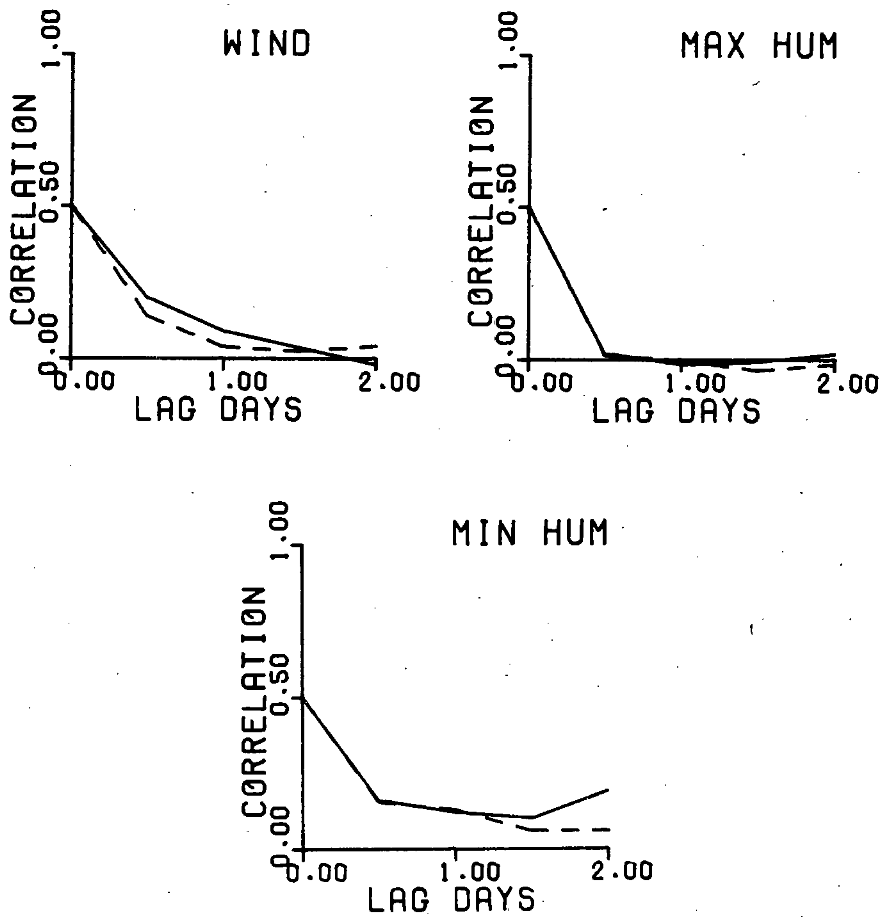
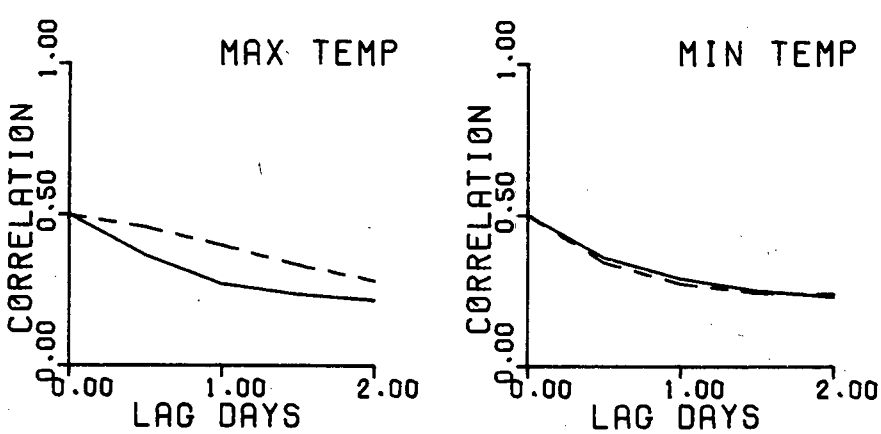
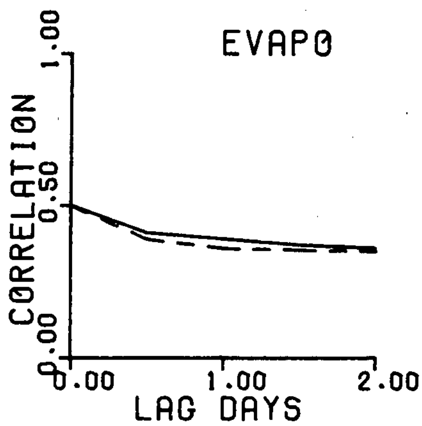
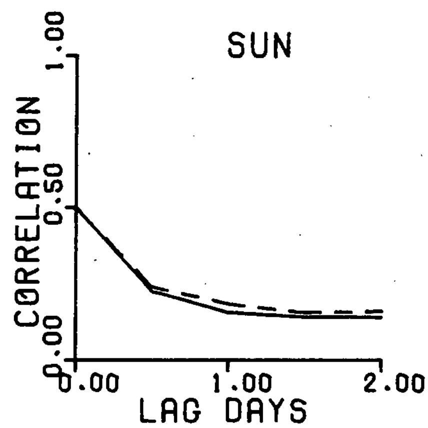
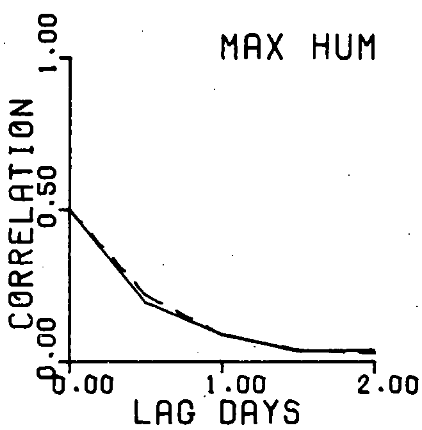


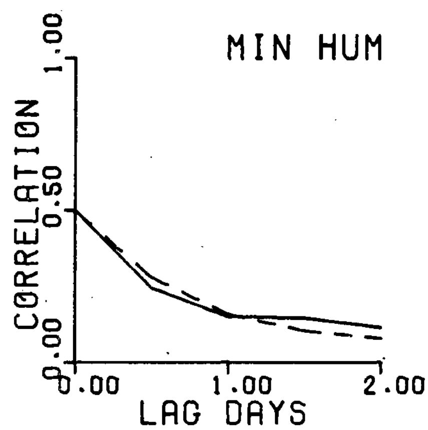
FIGURE 6.57: HISTORICAL (——) AND SIMULATED (---) AUTOCORRELATION COEFFICIENTS GIVEN A DRY DAY FOR ALL VARIABLES




EVAPO
CORRELATION
0.00
0.50
1.00
0.00
1.00
2.00
LAG DAYS


SUN
CORRELATION
0.00
0.50
1.00
0.00
1.00
2.00
LAG DAYS


MAX HUM
CORRELATION
0.00
0.50
1.00
0.00
1.00
2.00
LAG DAYS


MIN HUM
CORRELATION
0.00
0.50
1.00
0.00
1.00
2.00
LAG DAYS

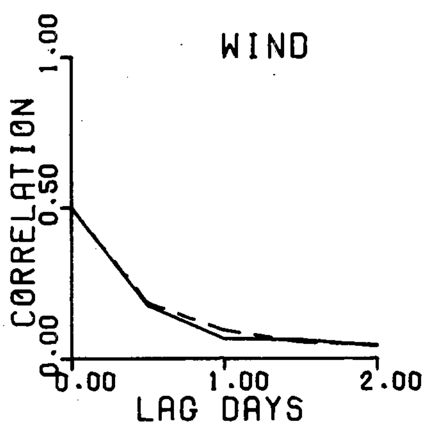
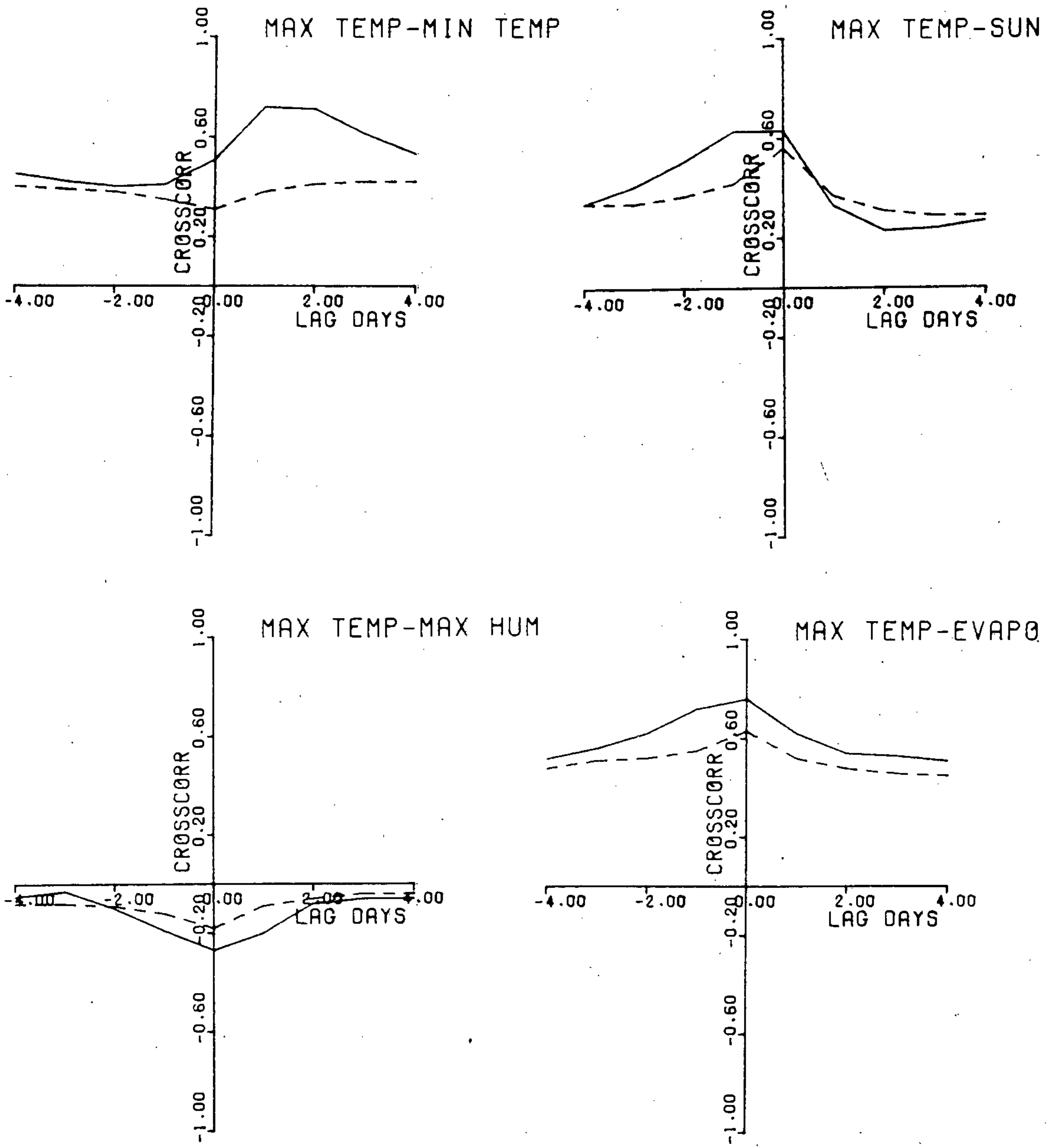
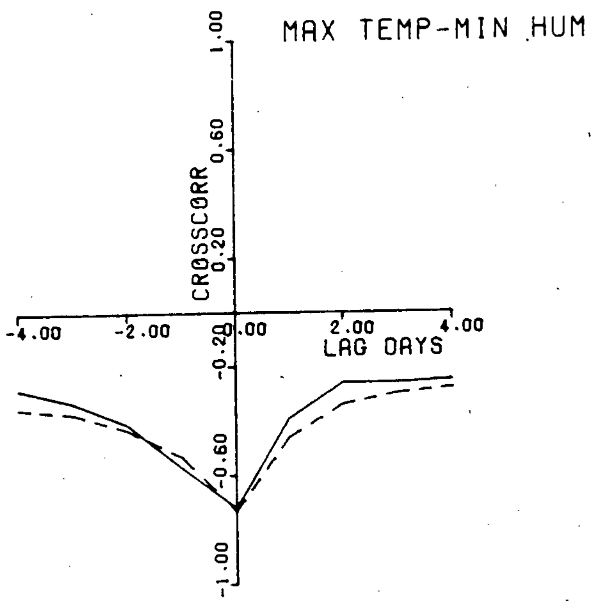
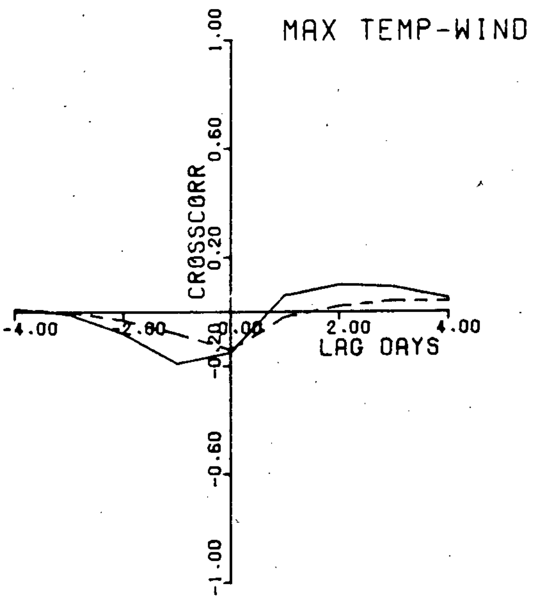
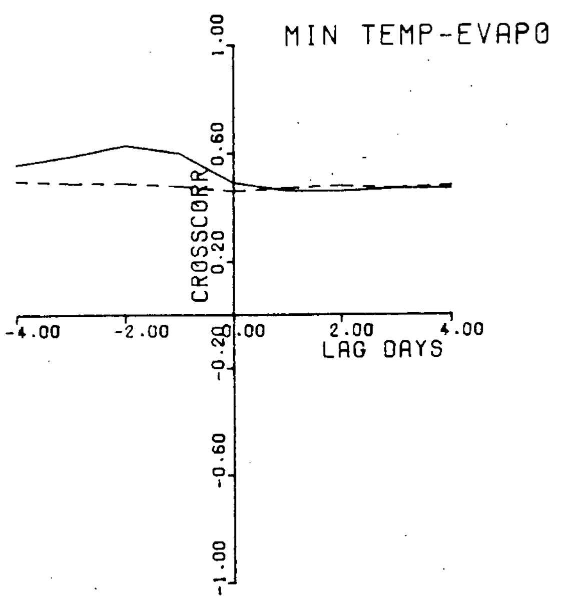

WIND
CORRELATION
0.00
0.50
1.00
0.00
1.00
2.00
LAG DAYS

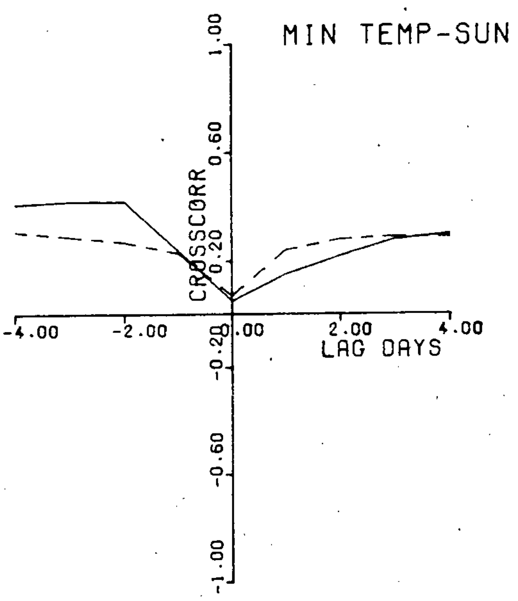
FIGURE 6.58: HISTORICAL (——) AND SIMULATED (---) CROSS-CORRELATION COEFFICIENTS

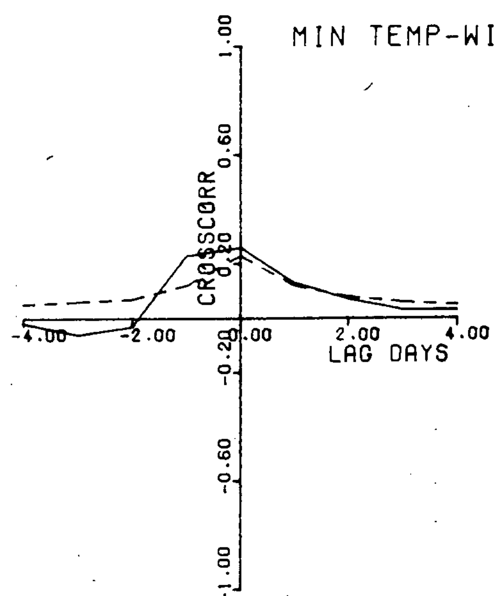


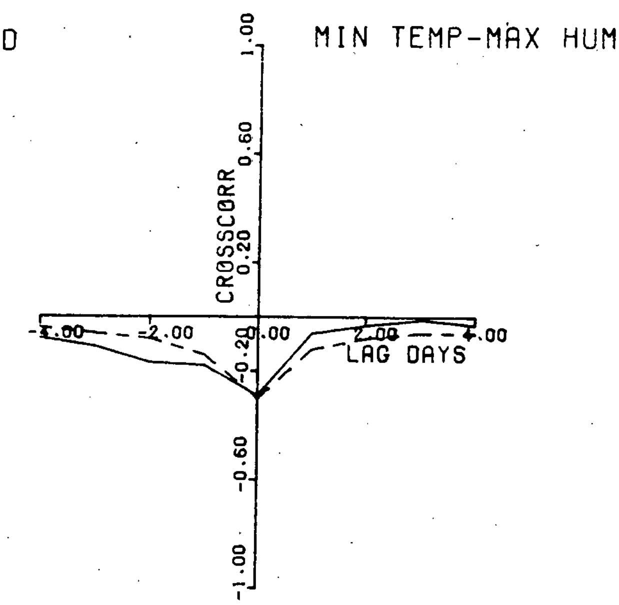

MAX TEMP-MIN HUM
CROSSCORR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00

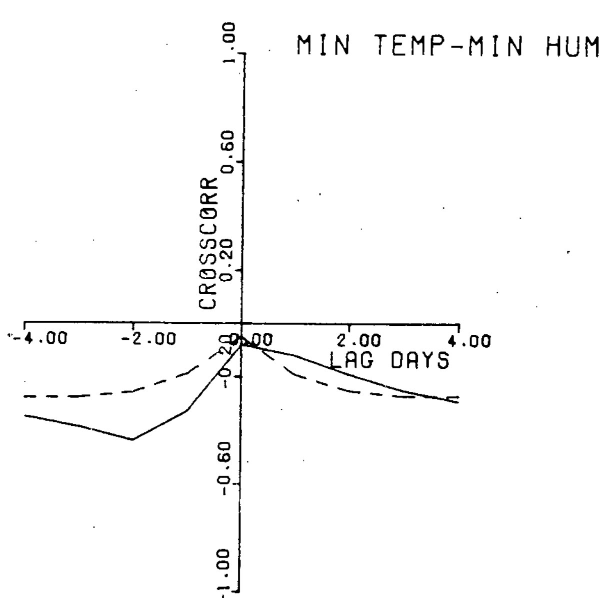

MAX TEMP-WIND
CROSSCORR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00

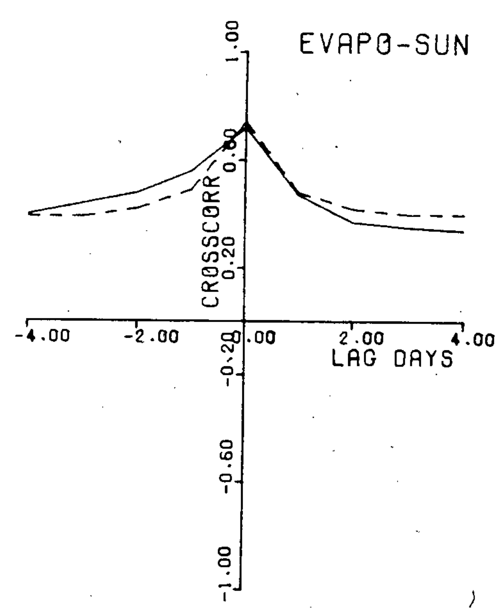

MIN TEMP-EVAPO
CROSSCORR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00

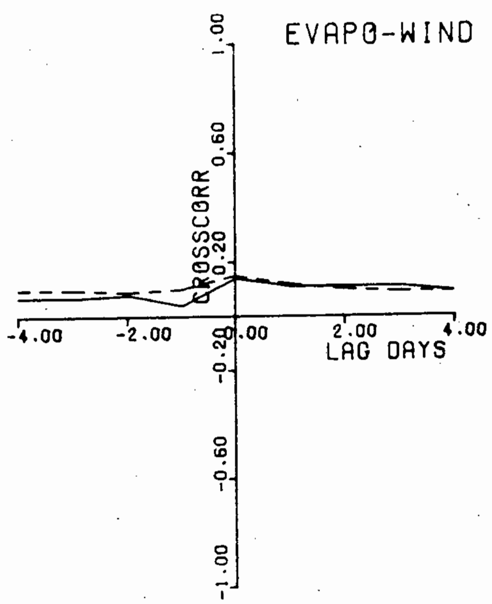

MIN TEMP-SUN
CROSSCORR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00

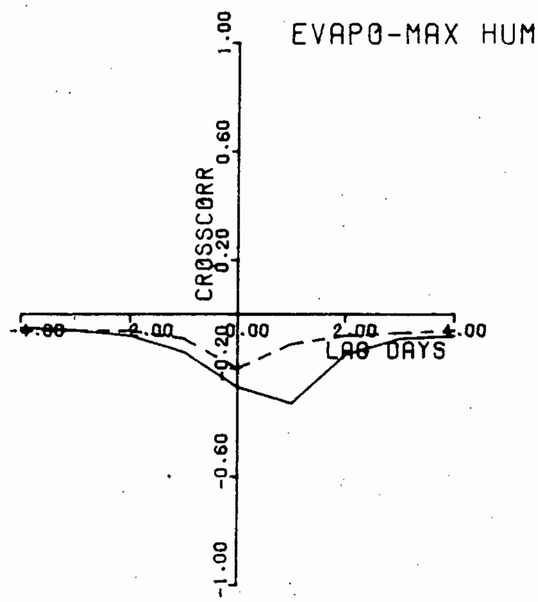

MIN TEMP-WIND
CROSSCORR
LAG DAYS

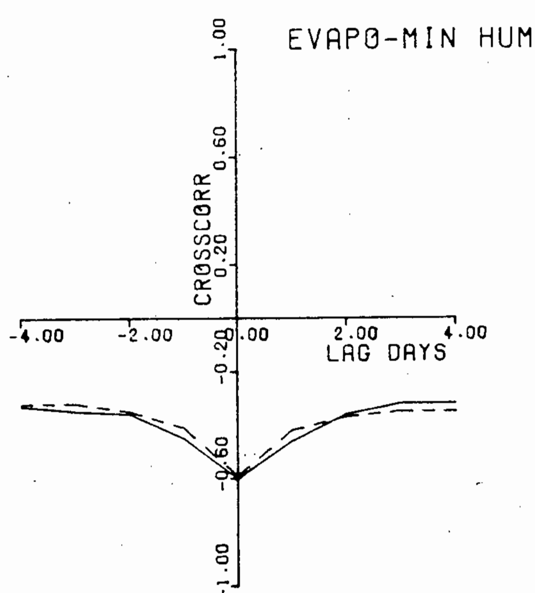

MIN TEMP-MAX HUM
CROSSCORR
LAG DAYS

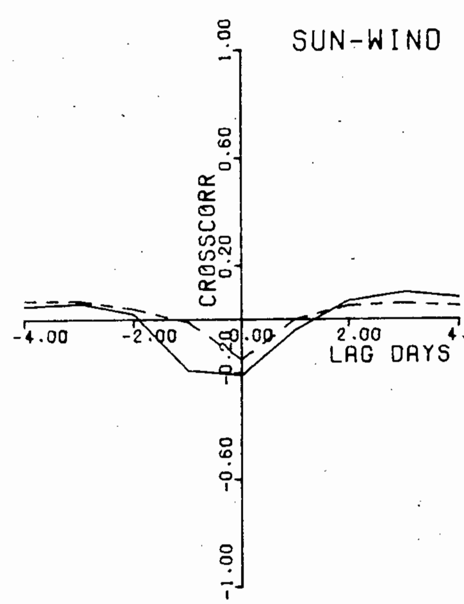

MIN TEMP-MIN HUM
CROSSCORR
LAG DAYS

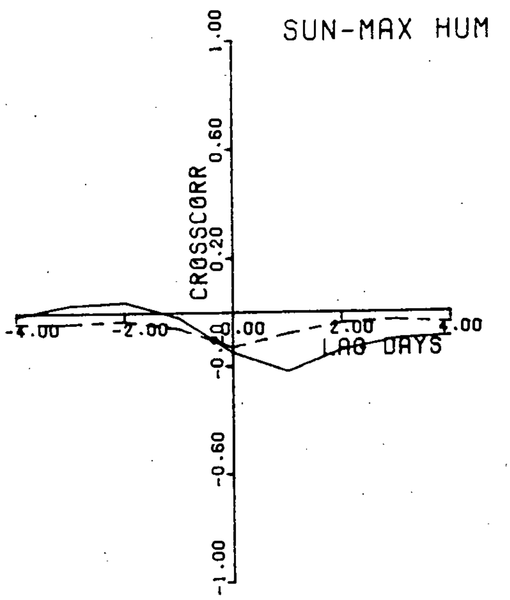

EVAPO-SUN
CROSSCORR
LAG DAYS

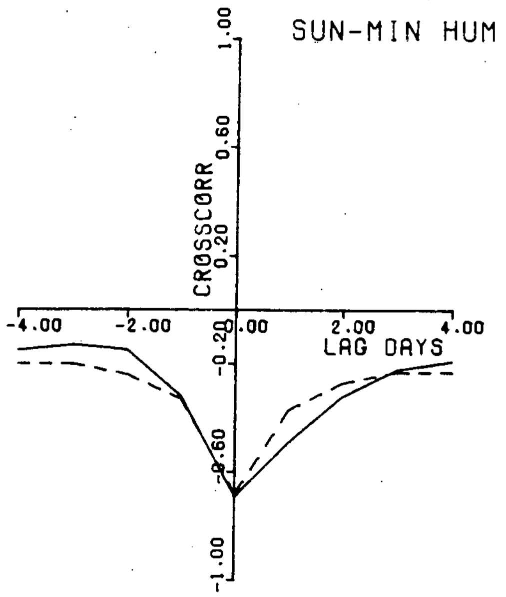

EVAPO-WIND
CROSSCORR
LAG DAYS

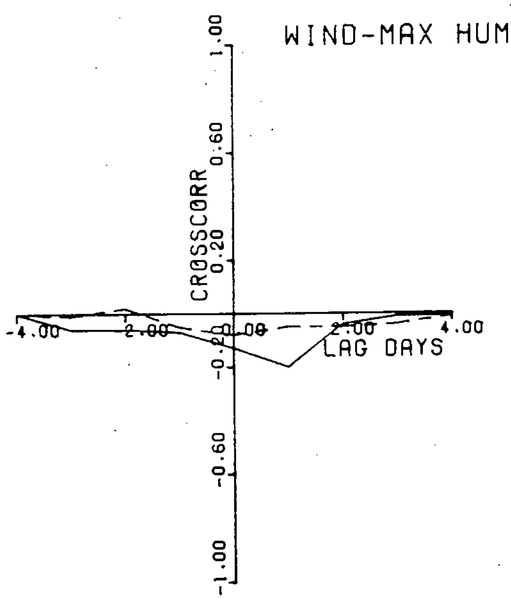

EVAPO-MAX HUM
CROSSCORR
LAG DAYS

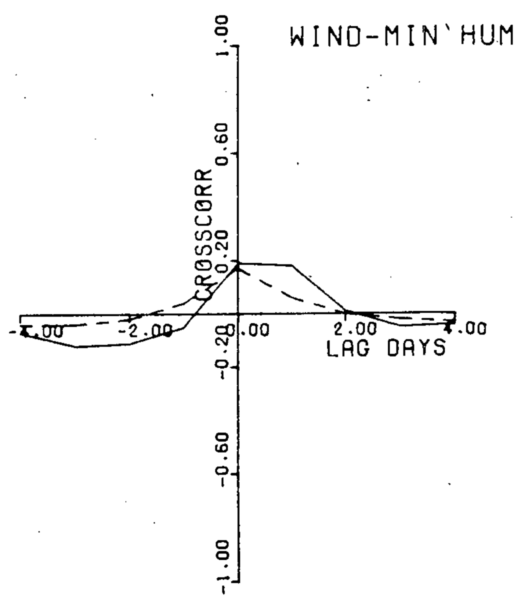

EVAPO-MIN HUM
CROSSCORR
LAG DAYS


SUN-WIND
CROSSCORR
LAG DAYS


SUN-MAX HUM
CRØSSCØRR
LAG DAYS


SUN-MIN HUM
CRØSSCØRR
LAG DAYS


WIND-MAX HUM
CRØSSCØRR
LAG DAYS


WIND-MIN'HUM
CRØSSCØRR
LAG DAYS

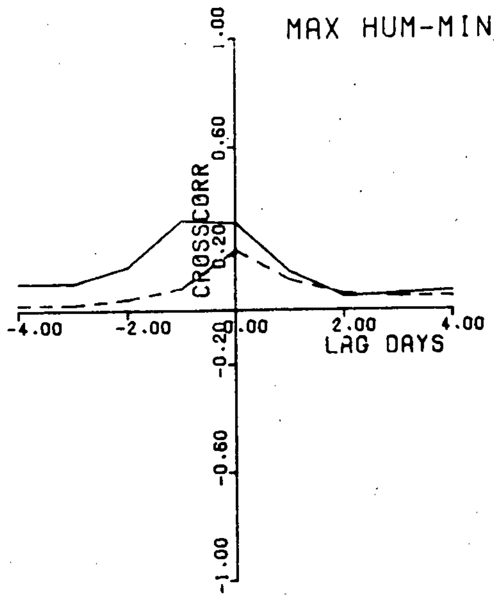
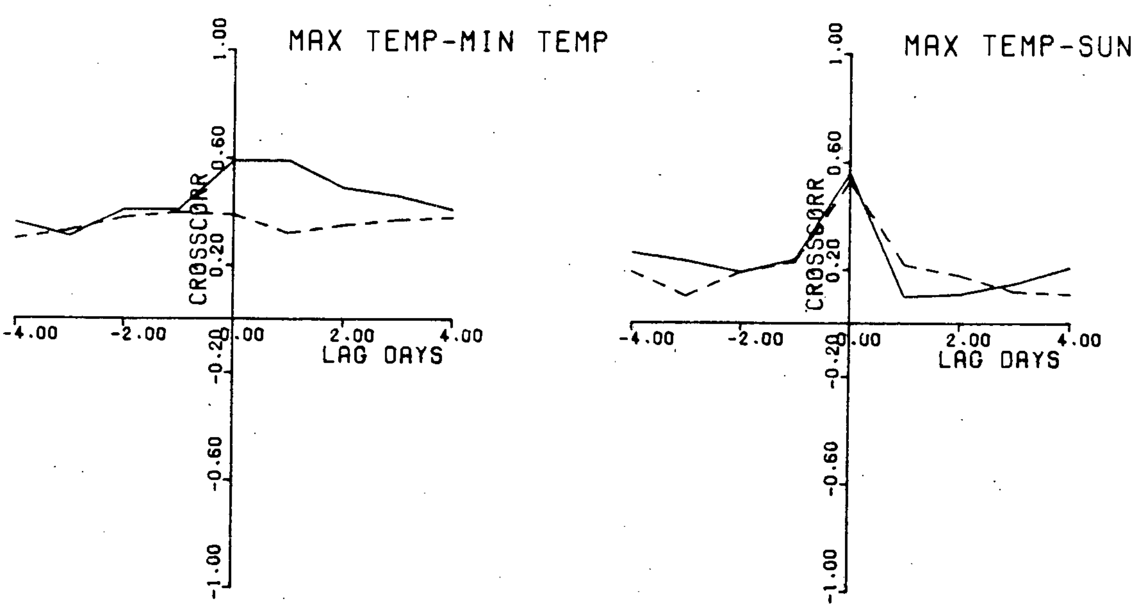


FIGURE 6.59: HISTORICAL (——) AND SIMULATED (---) CROSS-CORRELATION COEFFICIENTS GIVEN A WET DAY



The model describes the cross-correlation coefficients of all climate variables very well whether a dry sequence is observed or a wet one (Figures 6.59 - 6.60).

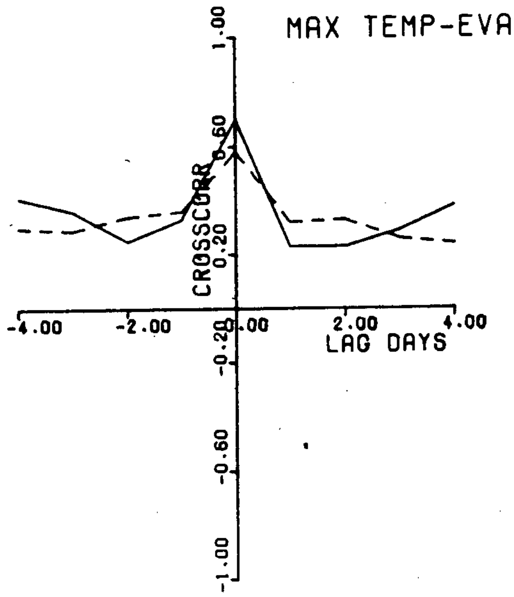
6.3 SUMMARY

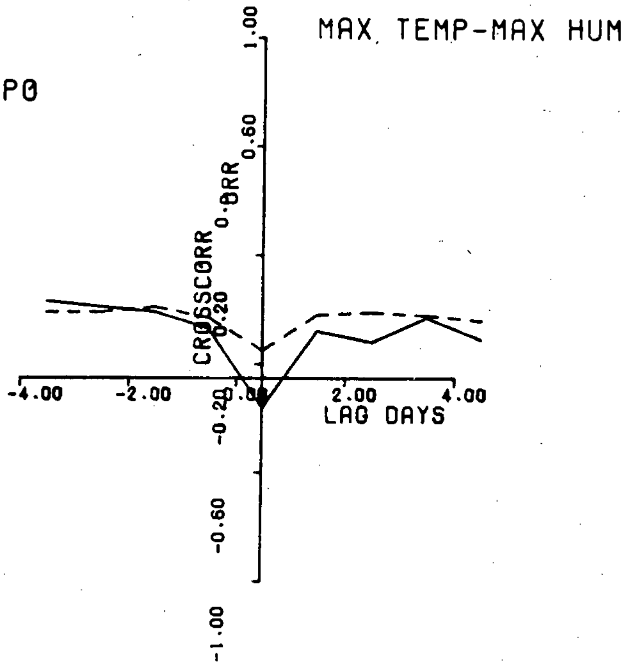
If one reflects on the complexity of the climate process and in particular on the large number of properties which the models are required to preserve it can be reasonably concluded that each of the three models performs remarkably well. All but few of the relevant parameter functions and cross-correlations are preserved surprisingly faithfully by the models. There are of course a number of weaknesses. The most important of these are:

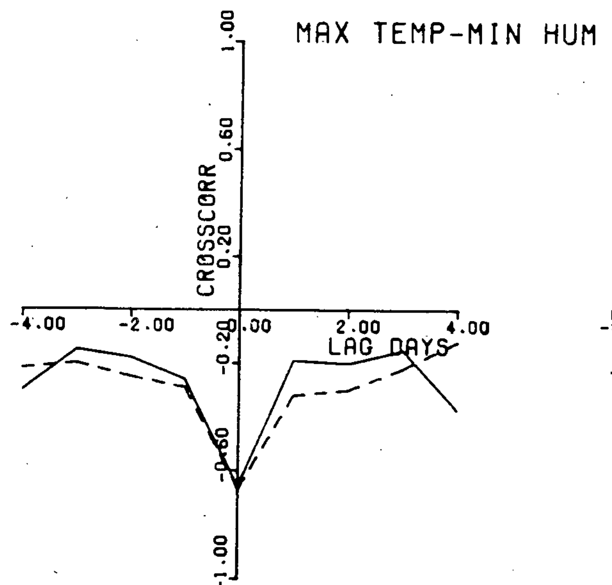
(i) The extreme values of the variables are not preserved by the model, especially the minimum values. As a result of this some of the simulated values fall outside the permissible range of the variables.

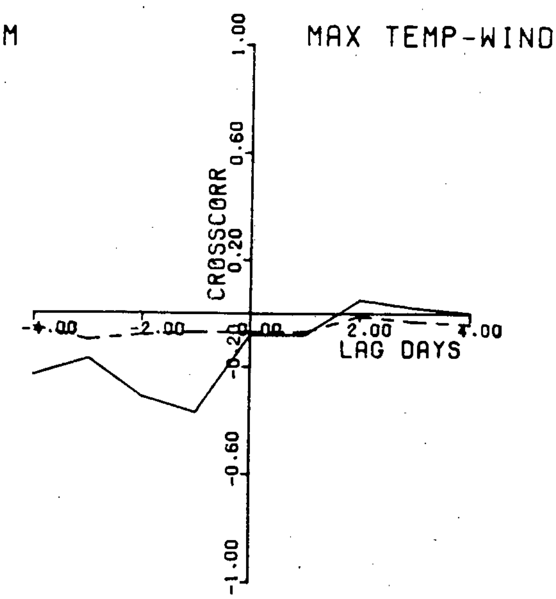
(ii) Model 3, when parameters are estimated by the method of maximum likelihood estimation, does not retain the property of the monthly means in the variable maximum temperature (they are underestimated).

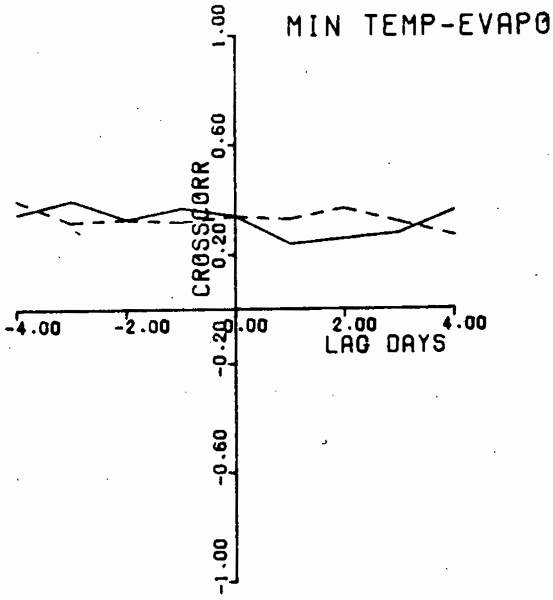
(iii) The models show a weakness in maintaining the cross-correlation coefficients of maximum temperature and minimum temperature for lags of one or more days. For Model 3 (ML) this weakness is also present when maximum temperature is

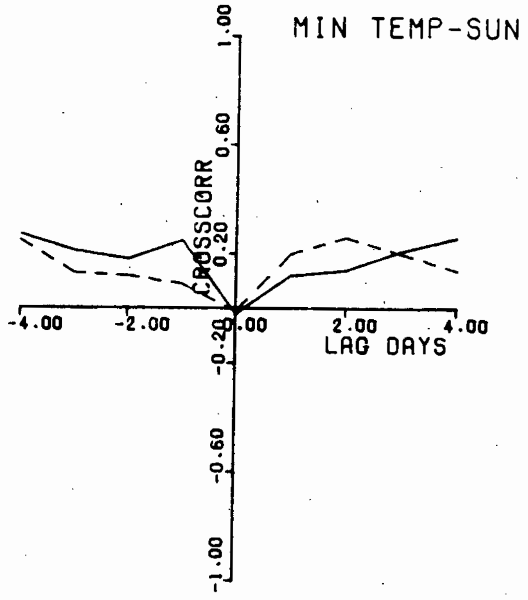

MAX TEMP-EVAPO
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

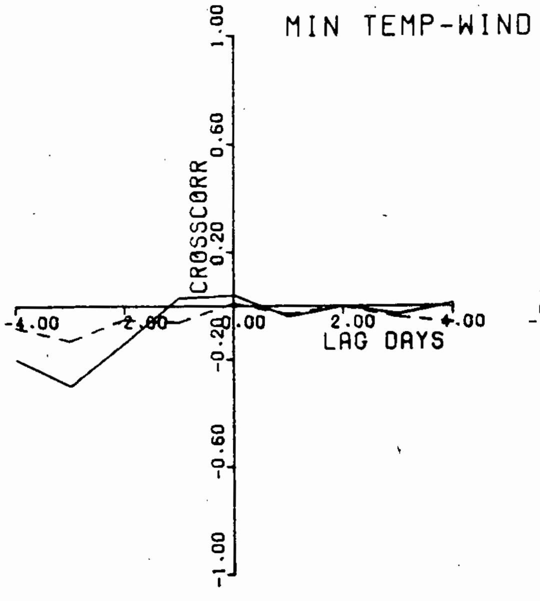

MAX TEMP-MAX HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
2.00
4.00
LAG DAYS

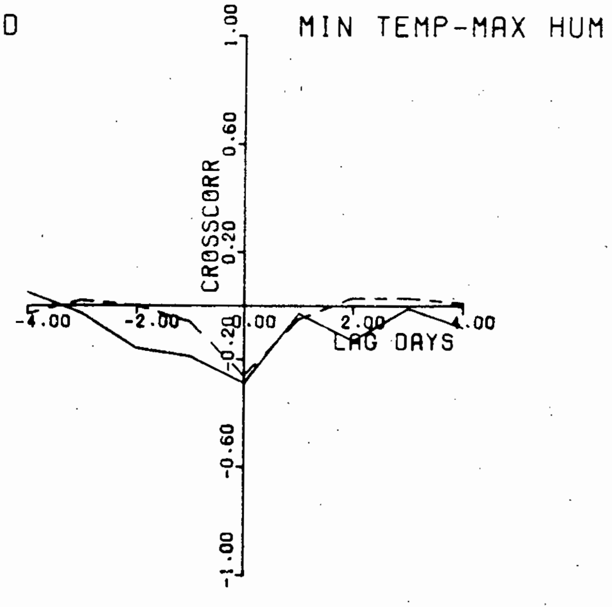

MAX TEMP-MIN HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

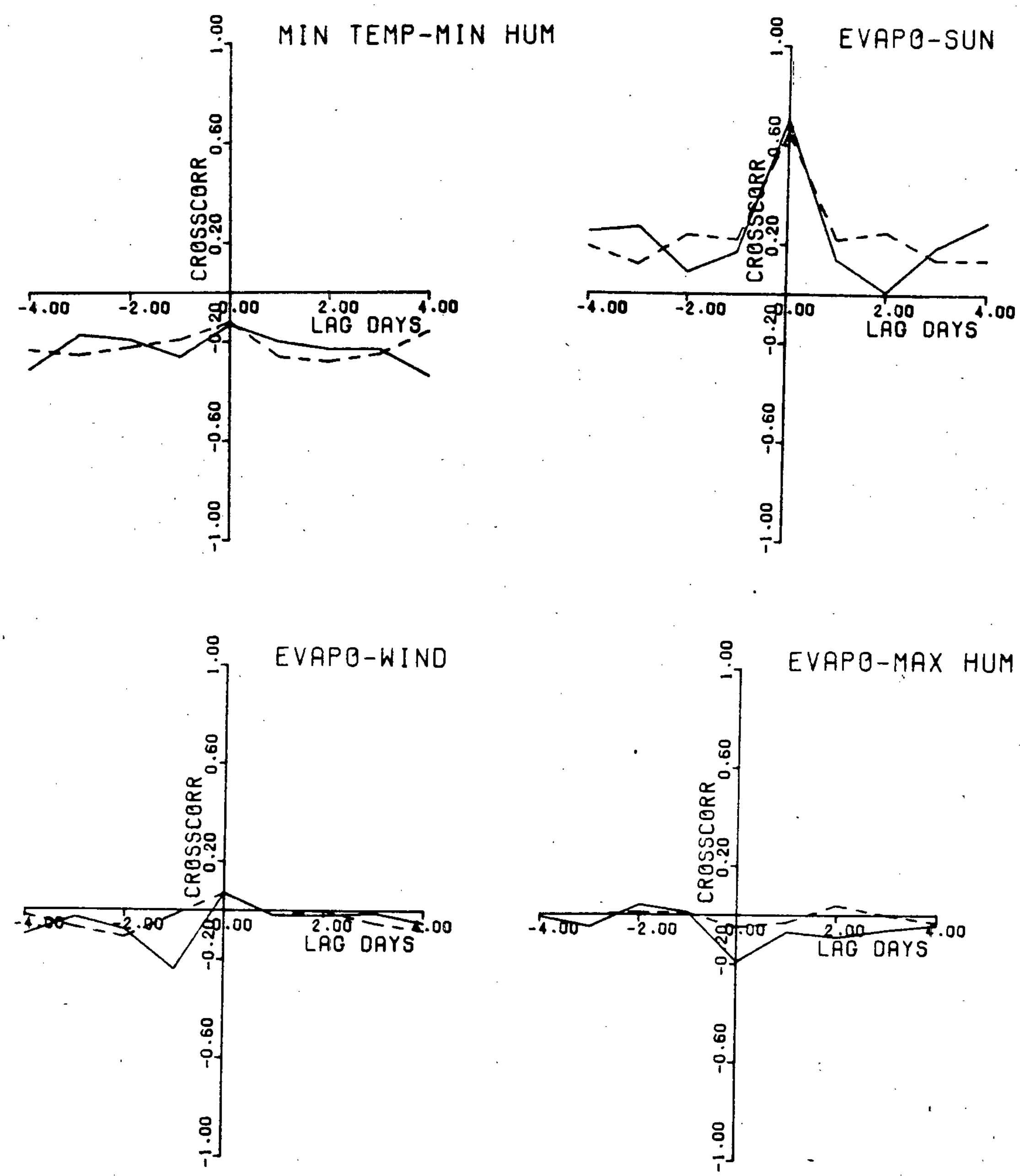

MAX TEMP-WIND
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-2.00
2.00
LAG DAYS

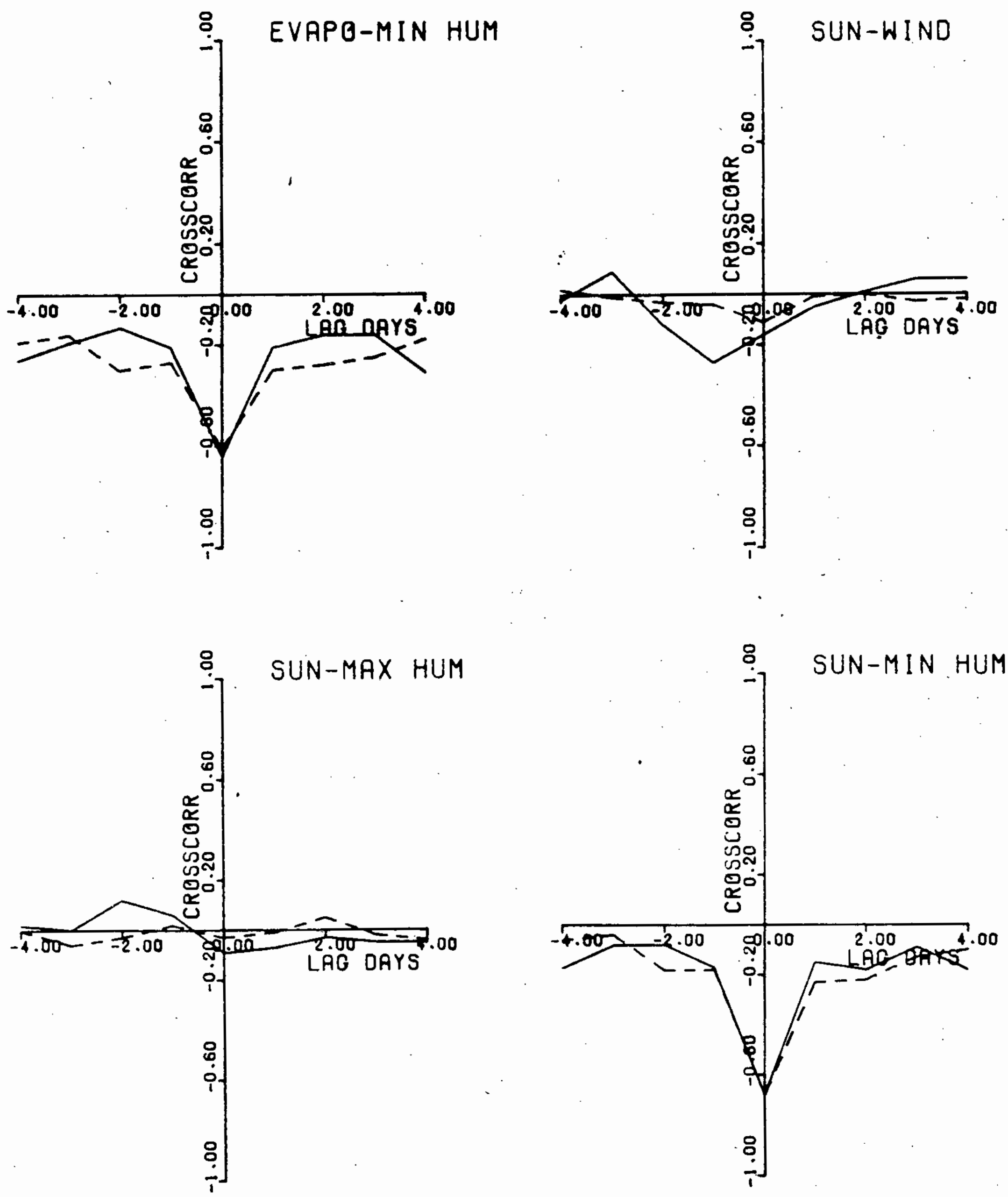

MIN TEMP-EVAPO
CROSSCORR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00

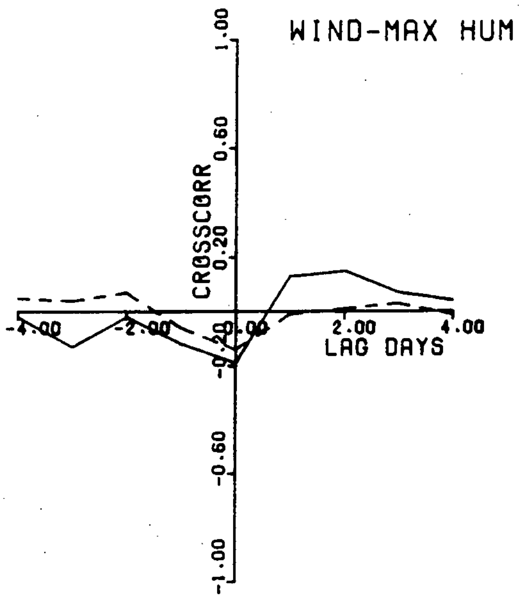

MIN TEMP-SUN
CROSSCORR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00

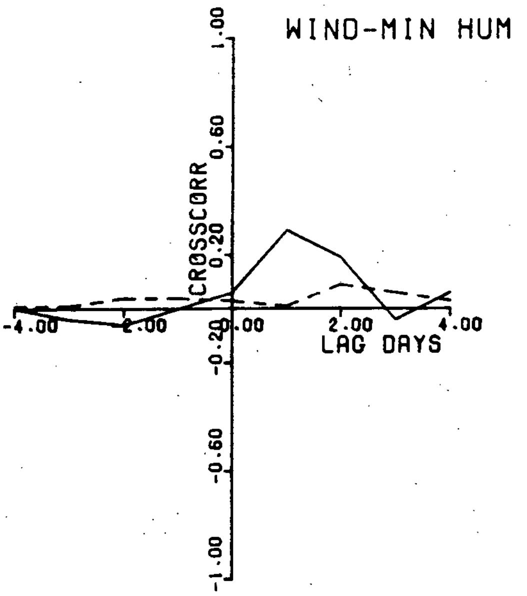

MIN TEMP-WIND
CROSSCORR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00


MIN TEMP-MAX HUM
CROSSCORR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00


MIN TEMP-MIN HUM
EVAPO-SUN
EVAPO-WIND
EVAPO-MAX HUM
CROSSCORR
LAG DAYS
-4.00
-2.00
0.00
2.00
4.00
-1.00
-0.60
-0.20
0.20
0.60
1.00


EVAPO-MIN HUM
SUN-WIND
SUN-MAX HUM
SUN-MIN HUM
CROSSCORR
LAG DAYS
-4.00
-2.00
0.00
2.00
4.00
1.00
0.60
0.20
-0.20
-0.60
-1.00


WIND-MAX HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS


WIND-MIN HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

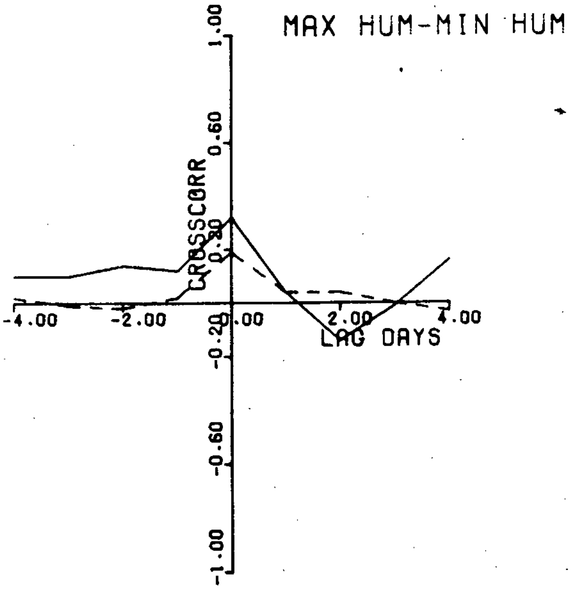
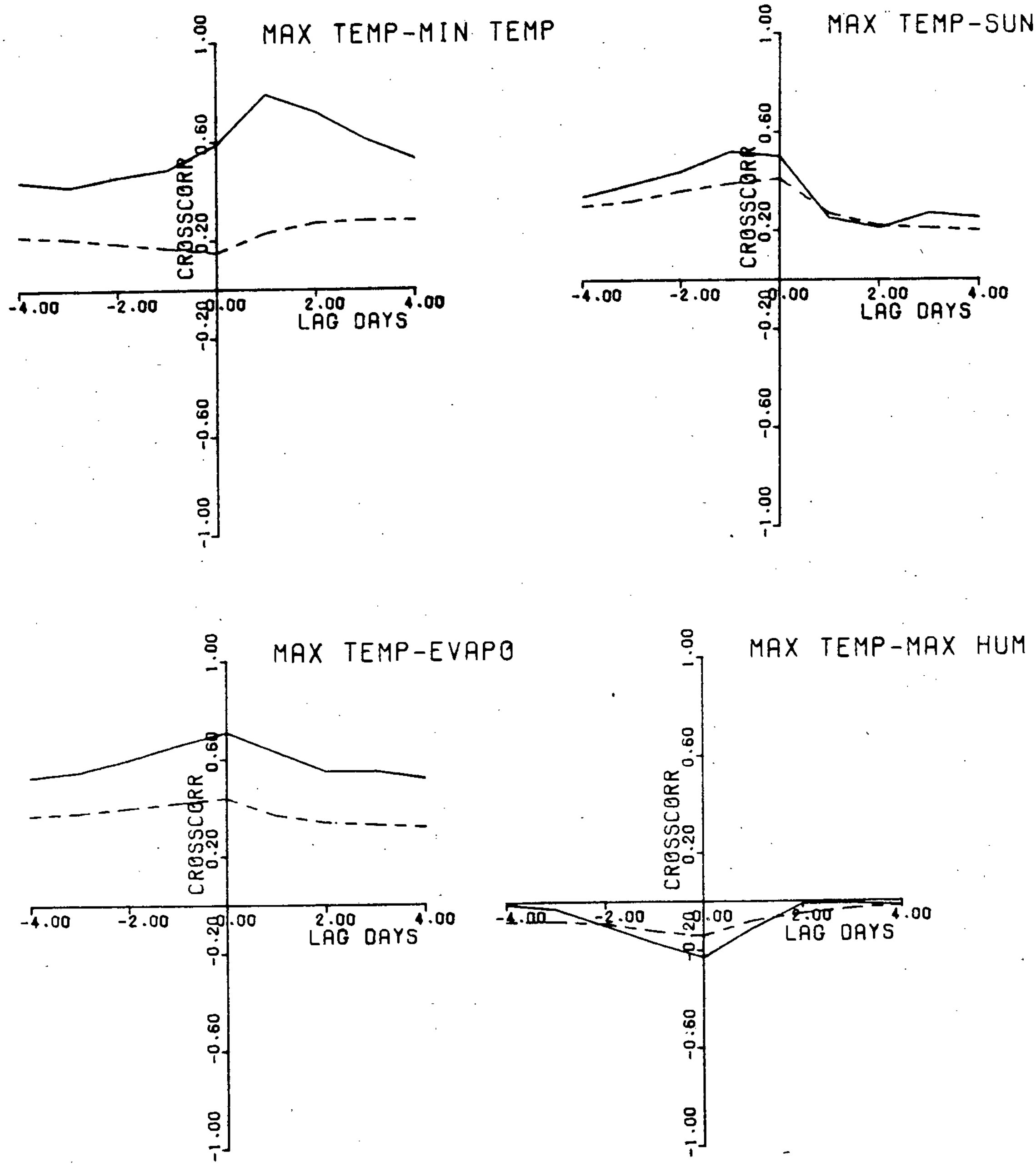
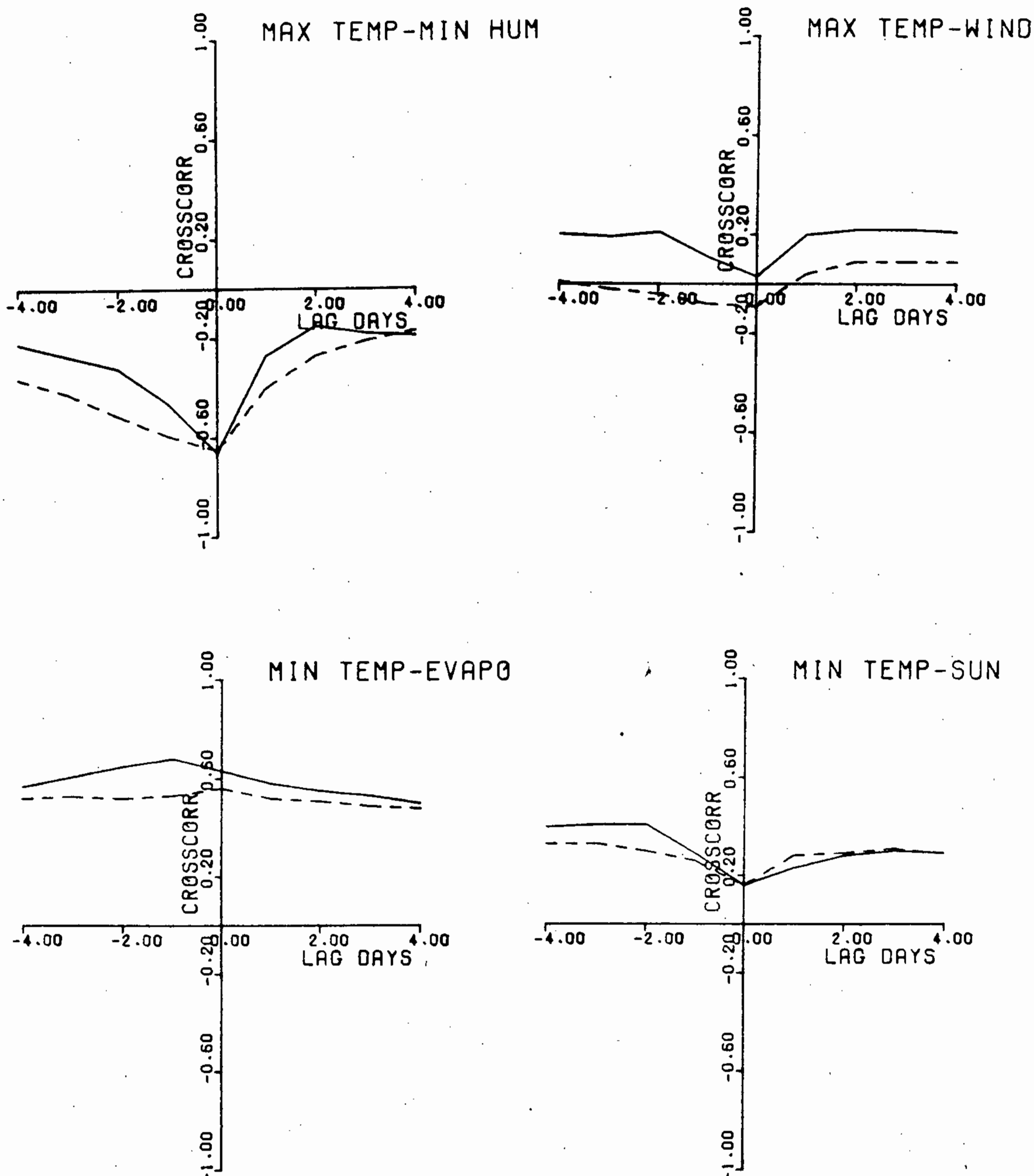
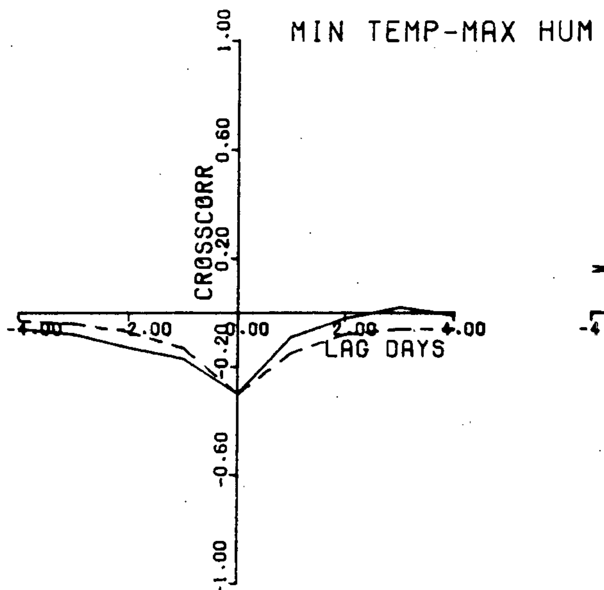
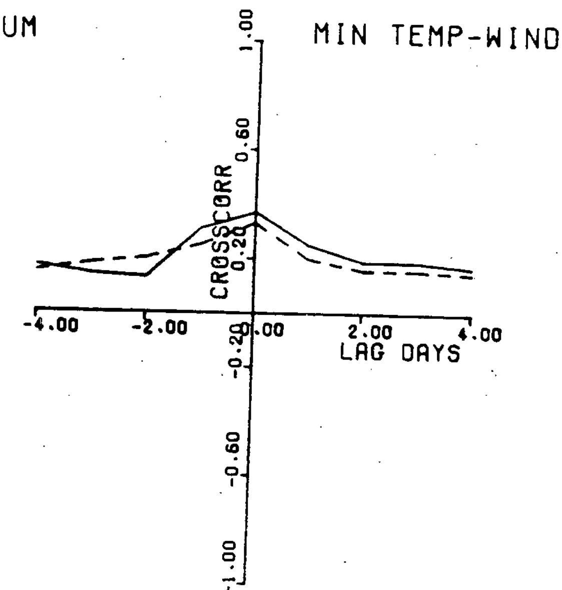

MAX HUM-MIN HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

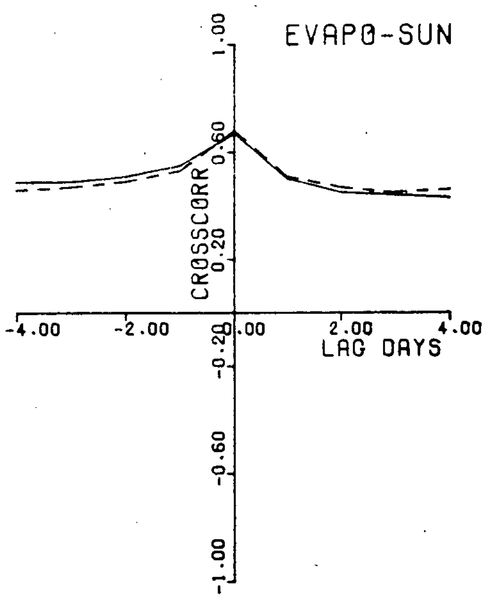
FIGURE 6.60: HISTORICAL (——) AND SIMULATED (---) CROSS-CORRELATION COEFFICIENTS GIVEN A DRY DAY

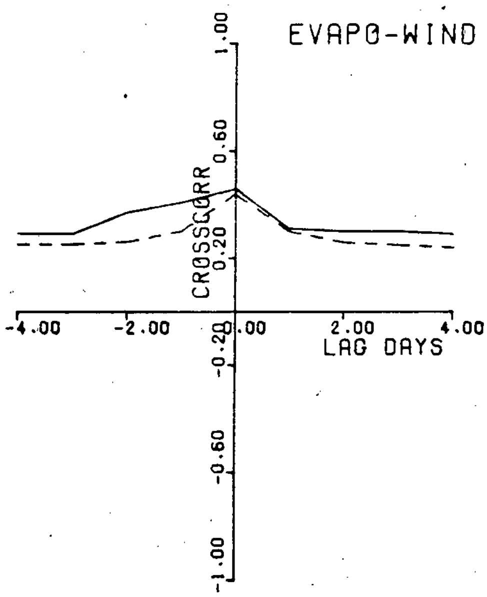


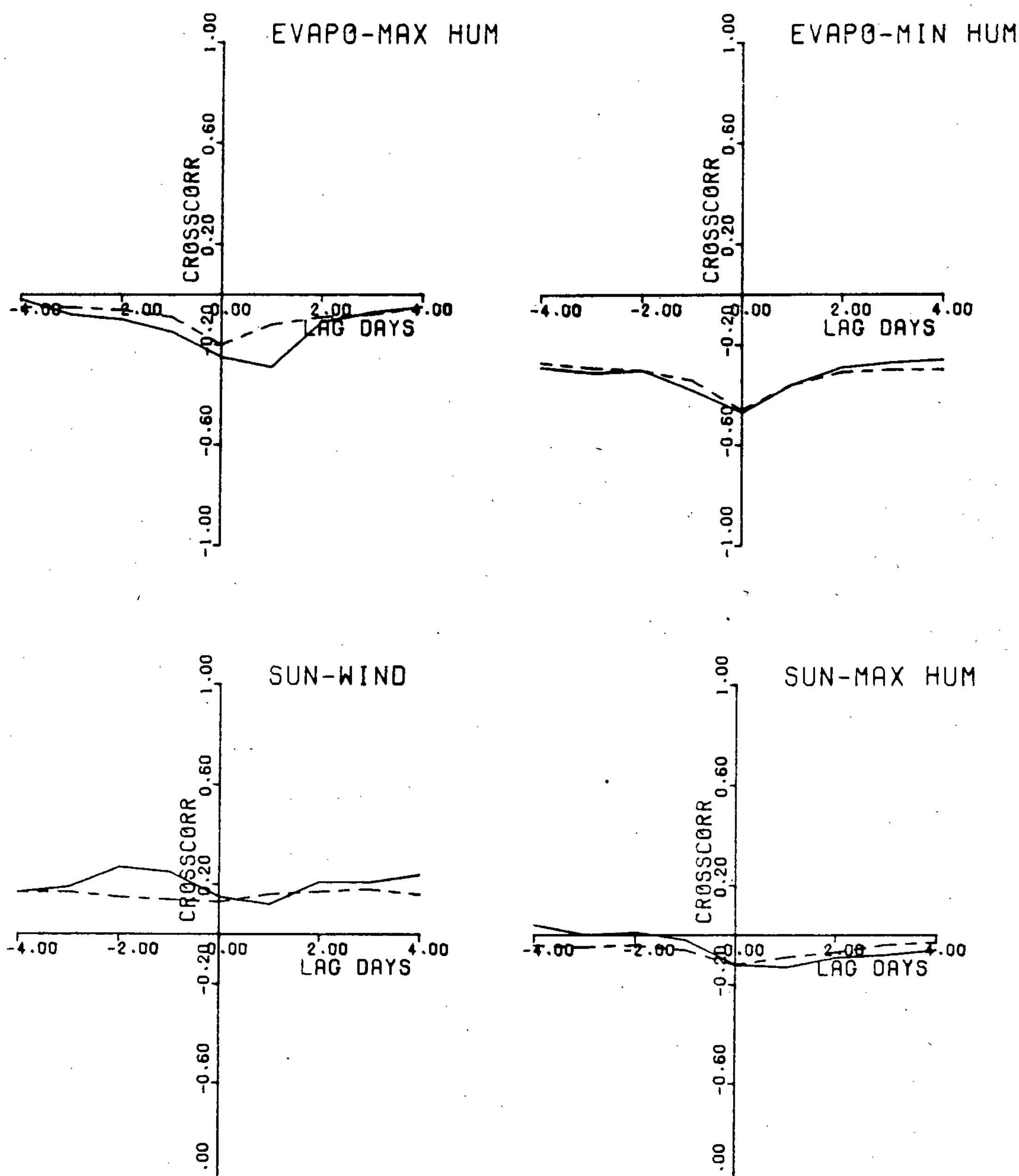

MAX TEMP-MIN HUM
MAX TEMP-WIND
MIN TEMP-EVAPO
MIN TEMP-SUN
CROSSCORR
LAG DAYS
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00

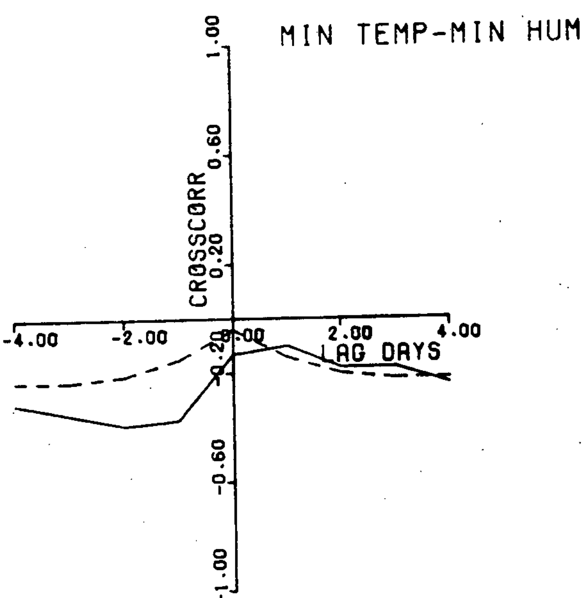

MIN TEMP-MAX HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

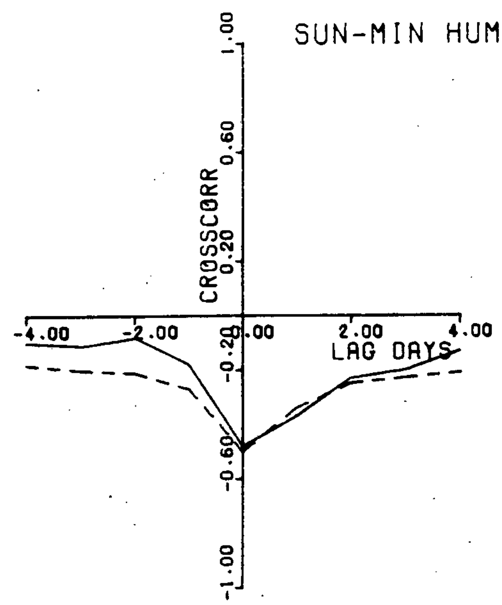

MIN TEMP-WIND
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

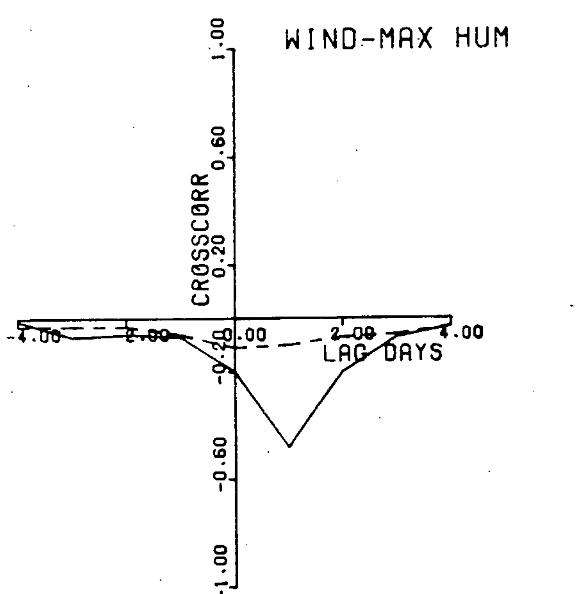

EVAPO-SUN
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

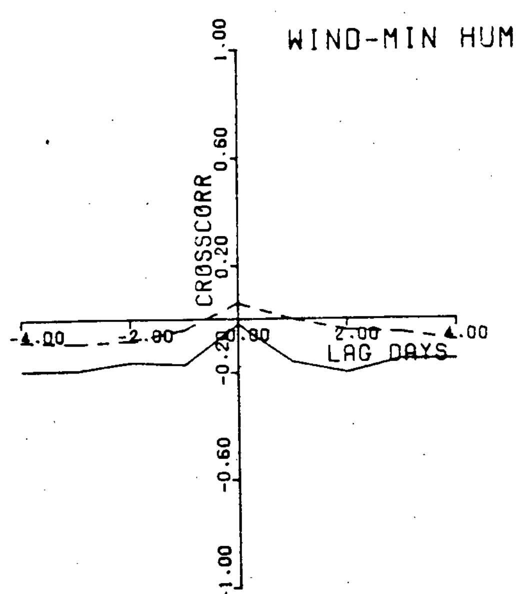

EVAPO-WIND
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

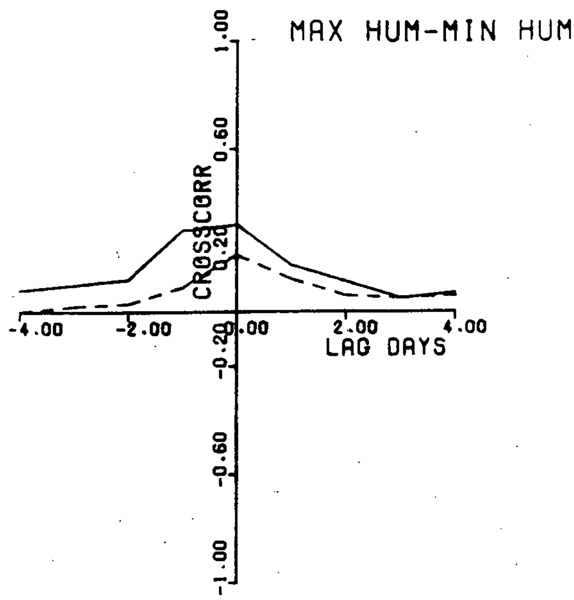

EVAPO-MAX HUM
EVAPO-MIN HUM
SUN-WIND
SUN-MAX HUM
CROSSCORR
LAG DAYS
-4.00
-2.00
0.00
2.00
4.00
-1.00
-0.60
-0.20
0.20
0.60
1.00


MIN TEMP-MIN HUM
CROSSCORR
LAG DAYS


SUN-MIN HUM
CROSSCORR
LAG DAYS


WIND-MAX HUM
CROSSCORR
LAG DAYS


WIND-MIN HUM
CROSSCORR
LAG DAYS


MAX HUM-MIN HUM
CROSSCORR
1.00
0.60
0.20
-0.20
-0.60
-1.00
-4.00
-2.00
0.00
2.00
4.00
LAG DAYS

lagged with evaporation and with sunshine duration. This weakness is also present on the conditioned sequences.

The problem that the generated values fall outside their respective admissible ranges can of course be easily overcome by simply setting the generated values to the appropriate boundary value whenever they fall outside the range. Such a procedure is easy to implement but it does change the parameter functions of the generated process (for example the mean). However, since the percentage of such points is quite small the resultant bias will be small.

A second possibility is to model some of the variables as being partly discrete and partly continuous (which is in fact what they are). This leads to severe theoretical complications, both in the specification of the joint distribution of the variables and in the derivation of estimates of the parameters. This approach will be investigated in a forthcoming research project.

A choice of model at this point is not straightforward as the performance of the models is neither perfect nor totally without merit but each model shows strengths and weaknesses. A criterion to base our preférence on any particular model can be based on factors such as:

(a) Implementation costs, i.e. derivation of theory for parameter estimation, complexity of model in terms of number of parameters needed and the computational simplicity aspect of the model.

(b) Preservation of climate variable properties by the model.

Model 3 is the more complex of the models, both in terms of the number of parameters and in computational difficulty. The model parameters are estimated iteratively and therefore more time consuming. From the model validation results obtained it is clear that this model performs best when the technique of ordinary least squares is used for parameter estimation.

Model 1 is the simplest in terms of the number of parameters, while Model 2 is computationally easier (the Fourier coefficients for the mean function are estimated as is Model 1 and the parameter of the autoregressive process is estimated by the lag one autocorrelation coefficient of each variable, previously computed during model identification and the covariance matrix is easily estimated).

Generally Model 1 appears to perform as well and sometimes better than the other models in describing some aspects of the climate variables, for example, the monthly standard deviations. However, the other models cannot be simply dismissed as more sophisticated model selection procedures may result in an improved performance by these models.

CHAPTER 7

7. SUMMARY AND CONCLUSIONS

This chapter gives a brief summary of the study performed followed by the main research findings and finally certain topics for future research are pointed out.

7.1 Summary.

Three stochastic models were considered to describe daily climate sequences. The technique employed was firstly to model rainfall using a first-order Markov chain with seasonal parameters to describe the occurrence of wet and dry days, while the Weibull distribution was used to describe the rainfall depth on wet days. The rainfall mean was allowed to vary seasonally. This model provides synthetic sequences of wet and dry days. Finally, the remaining climate variables were modelled according to the wet or dry status of each day. The three models considered describe the climate variables maximum and minimum temperature, evaporation, sunshine duration, windrun and maximum and minimum humidity.

The first model considered was proposed by Richardson (1981), where a stationary residual series is obtained by subtracting the seasonal mean and dividing by the standard deviation of each climate variable. A weakly stationary process suggested by Matalas (1967) is used to model the residual series. It is assumed that the residual series is normally distributed and that the serial correlation of each variable can be described by a first-order autoregressive process.

The second multivariate model considered regards each climate variable as the sum of two components, namely a deterministic seasonal mean and a stochastic error term, where the latter is described by an auto-regressive model of order one. The covariance matrix is used to describe the interdependence between the climate variables. Under the assumption of normality this covariance matrix is sufficient to specify the process.

The third model considered follows the same approach as the above model except that in this case each variable is not only conditioned on the wet or dry status of each day but also on the status of the preceding day. That is, one is assuming that the observed climate values on a rainy day, for example, differ significantly depending on whether the previous day was dry or wet.

For each of the above models, the seasonal parameter functions were approximated using a truncated Fourier representation.

7.2 Findings of this study.

The results of model validation for the rainfall model confirm the findings of Zucchini and Adamson (1984), namely that the assumptions regarding the characteristics of daily rainfall sequences, the rationale of model structure and the parameter estimation techniques are particularly successful in providing a model that can adequately reproduce the properties of daily rainfall sequences.

The requirement for an accurate simulation of the occurrence of wet and dry days is very important in the present study as these simulated sequences are used to determine the generation procedure to be adopted by the other climate variables. This component of the rainfall model was found to be very successful in preserving the characteristics of the occurrence of wet and dry days. Tests of the multivariate models for climate data showed that the models were capable of representing almost all the characteristics exhibited by the historical data. The pattern of serial correlation and cross-correlation in daily values of climate variables was closely described by the dependence structure in each of the models considered.

All models satisfactorily preserved the mean and standard deviation of minimum temperature, evaporation and sunshine duration. For the remaining climate variables, some of the models failed to preserve either the mean or the standard deviation structure of these variables. The maximum extreme values of the variables were mostly preserved by all models considered. The models failed however to preserve the minimum extreme values of the climate variables.

The results obtained have established that it is feasible to model climate variables, i.e. the properties exhibited by climate sequences can be reasonably accurately represented by a stochastic model. It is also established that there is certainly room for improvement. Details relating to some of the more important weaknesses of the models are given below.

7.3 Topics for future research.

A number of recommendations for future research are outlined.

7.3.1 Increase size of time series and number of stations.

The parameter estimates for the models considered were based on six years of historical data. This amount of data is not sufficient to be representative of the climatic occurrences. Dealing with a small data set leads to a further complication in semi-arid regions, namely that there are relatively few occurrences of rainfall thus, when conditioning the climate variables on the status of each day, there is a relatively large number of time periods with no observations. In the case of Model 3 (ML) and (LS) this problem is even more marked as one is conditioning the data set not only on the wet or dry status of each day, but also on the wet or dry status of the previous day.

To firmly establish the relative merits of the above models, the climate sequences of several weather stations, with similar and different climates, will have to be modelled and the resulting fits investigated.

7.3.2 Take into account the permissible boundaries of each climate variable.

One of the setbacks of the present models for climate sequences is that the minimum extreme values are not being preserved by the models. This problem arises because some climate variables lie within permissible boundaries, with some variables having a high frequency of values near or on an upper or lower limit. For example, sunshine duration has a lower limit of zero hours and many of the observed values of sunshine duration are zero or are close to zero. A model that takes into account the limits of each variable should be investigated.

7.3.3 More feasible models.

In the present study all variables were modelled in the same way. Furthermore, the Fourier approximations of the seasonal parameters were all of the same order. A more flexible model in which each variable is modelled differently depending on the properties it exhibits can be considered. For example, windrun might not show a significant difference between a sequence in which a dry day follows a wet day and a sequence in which a wet day follows a dry day. In such a case, the variable should not be conditioned on these sequences.

7.3.4 Develop procedures for dealing with missing values.

The serial correlation present in the climate variables does not allow one to simply discard missing values. Methods need to be developed to deal with this difficult problem. As the type of data considered in this study typically has a high proportion of missing observations, this problem is one which demands urgent attention.

7.3.5 Develop model selection techniques.

Presently no theoretical model selection techniques are available when dealing with data that exhibit properties such as the ones observed in the climate variables. Proper model selection techniques should be developed which ensure that the correct order of model components is selected.

7.4 Concluding remarks.

The climate variables investigated in this study have a very complex joint distribution. Each variable exhibits a number of distinctive features and in addition the variables are interdependent. Any model which is to usefully describe climate sequences must preserve these properties. This study has shown that it is feasible to model climate on a daily basis and that there are at least three models which can be used to do so. Furthermore a relatively short historical record suffices to estimate the model parameters. A number of weaknesses in the models have been identified and much research will need to be done to perfect certain components of the models. However, none of these difficulties appear to be insurmountable. There can be no doubt that even as they are, the three models considered here provide results which are adequate for practical purposes.

REFERENCES

ACTON, F.S. (1970). Numerical methods that work. Harper and Row, New York.

ALVEY, N.G. et al. (1983). GENSTAT: A general statistical program. Numerical Algorithms Group, Oxford.

BAKER, R.J. and NELDER, J.A. (1978). GLIM manual, (release 3). Numerical Algorithms Group, Oxford.

BARRY, R.G. and CHARLEY, R.J. (1968). Atmosphere, weather and climate. Methuen and Co. Ltd., London.

BOX, G. E. P. and JENKINS, G.M. (1970). Time series analysis: forecasting and control. Holden-Day, New York.

BRUHN, J.A., FRY, W.E. and FICK, G.W. (1980). Simulation of daily weather data using theoretical probability distributions. Journal of Applied Meteorology, 19, 1029-1036.

BUISHAND, T.A. (1977). Stochastic modelling of daily rainfall sequences. Meded. Landbouwhogesch, Wageningen, 77-83.

BUISHAND, T.A. (1978). Some remarks on the use of daily rainfall models. Journal of Hydrology, 36, 295-308.

CASKEY, J.E. (1963). A Markov chain model for the probability of precipitation occurrence in intervals of various length. Monthly Weather Review, 91, 298-301.

CHATFIELD, C. (1980). The analysis of time series: An introduction. (2nd ed.) Chapman and Hall, New York.

COCHRANE, D. AND ORCUTT, G.H. (1949). Application of least squares regression to relationships containing autocorrelated error terms. Journal of the American Statistical Association, 44, 32-61.

DIXON, W.J. et al. (1981). BMDP statistical software. University of California Press, Los Angeles.

DOBSON, A.J. (1983). An introduction to statistical modelling. Chapman and Hall, New York.

GABRIEL, K.R. and NEUMANN, J. (1962). A Markov chain model for daily rainfall occurrence at Tel Aviv. Quarterly Journal of the Royal Meteorological Society, 88, 90-95.

GOH, T.N. and TAN, K.J. (1977). Stochastic modelling and forecasting of solar radiation data. Solar Energy, 19, 755-757.

HAAN, C.T., ALLEN, D.M. and STREET, J.O. (1976). A Markov chain model of daily rainfall. Water Resources Research, 12, 443-449.

HOPKINS, J.W. and ROBILLARD, P. (1964). Some statistics of daily rainfall occurrences for the Canadian prairie provinces. Journal of Applied Meteorology, 3, 600-602.

ISON, N.T., FEYERHERM, A.M. and BARK, L.D. (1971). Wet period precipitation and the gamma distribution. Journal of Applied Meteorology, 10, 658-665.

LARSEN, G.A. and PENSE, R.B. (1982). Stochastic simulation of daily climatic data for agronomic models. Agronomy Journal, 74, 510-514.

LINHART, H. and ZUCCHINI, W. (1982). A method for selecting the covariates in analysis of covariates. South African Statistical Journal, 16, 97-112.

MATALAS, N.C. (1967). Mathematical assessment of synthetic hydrology. Water Resources Research, 3, 937-945.

MIELKE, P.W. (1973). Another family of distributions for describing and analyzing precipitation data. Journal of Applied Meteorology, 10, 275-280.

NELDER, J.A. and WEDDERBURN, R.W.M. (1972). Generalized linear models. Journal of the Royal Statistical Society, A, 135, 370-384.

NICKS, A.D. and HARP, J.F. (1980). Stochastic generation of temperature and solar radiation data. Journal of Hydrology, 48, 1-17.

RICHARDSON, C.W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resources Research, 17, 182-190.

ROLDAN, J. and WOOLHISER, D.A. (1982). Stochastic daily precipitation models: (1) A comparison of occurrence processes. Water Resources Research, 18, 1451-1459.

SMITH, R.E. and SCHREIBER, H.A. (1973). Point processes of seasonal thunderstorm rainfall, I, Distributions of rainfall events. Water Resources Research, 9, 871-884.

STERN, R.D. and COE, R. (1983). A model fitting analysis of daily rainfall data. Journal of the Royal Statistical Society, A, 147, 1-34.

TODOROVIC, P. and WOOLHISER, D.A. (1975). A stochastic model of n-day precipitation. Journal of Applied Meteorology, 14, 17-24.

WEISS, L.L. (1964). Sequences of wet or dry days described by a Markov chain probability model. Monthly Water Review, 92, 169-176.

WOOLHISER, D.A. and PEGRAM, G.G.S (1979). Maximum likelihood estimation of Fourier coefficients to describe seasonal variations of parameters in stochastic daily precipitation models. Journal of Applied Meteorology, 18, 34-42.

WOOLHISER, D.A. and ROLDAN, J. (1982). Stochastic daily precipitation models: (2) A comparison of distributions of amounts. Water Resources Research, 18, 1461-1468.

ZUCCHINI, W. amd ADAMSON, P.T. (1984a). The occurrence and severity of droughts in South Africa. Report to the Water Research Commission, No. 91/1/84.

ZUCCHINI, W. and ADAMSON, P.T. (1984b). The occurrence and severity of droughts in South Africa: Appendix 6. Report to the Water Research Commission, No. 91/1/84(a).

APPENDIX A

The choice of the Fourier approximation, L .

The order of approximation of the Fourier representation of a function, $\lambda(t)$, is always taken to be an odd integer. This restriction is made partly for programming convenience and partly for the following reason:

If we rewrite the Fourier representation of $\lambda(t,L)$ by its polar form, we get

$$\lambda(t,L) = \begin{cases} \alpha_0 + \sum_{i=1}^{P} \alpha_i \cos(2\pi i/NT\,((t-1) - \beta_i)) , & L \text{ odd} \\ \alpha_0 + \sum_{i=1}^{P} \alpha_i \cos(2\pi i/NT\,((t-1) - \beta_i)) + \alpha_p \cos\frac{2\pi p}{NT} \cdot (t-1) , & L \text{ even} \end{cases}$$

where

$$\alpha_0 = \gamma_1$$

$$\alpha_i = (\gamma_{2i}^2 + \gamma_{2i+1}^2)^{\frac{1}{2}} , \qquad i=1,2,\ldots,p$$

$$\beta_i = NT/2\pi i \arctan(\gamma_{2i+1}/\gamma_{2i}) \qquad i=1,2,\ldots,p$$

and P is the integer part of $(L-1)/2$. The α_i is called the amplitude and β_i is called the phase of the ith harmonic.

If L is even, then the highest harmonic does not have a phase parameter. Thus the quality of the fit of the model depends on the time origin selected. If L is odd we obtain the same degree of approximation for all time origins.

APPENDIX B

Properties of the Fourier series approximation.

We have used the Fourier representation of $\lambda(t)$ as the basis for obtaining approximations. Other representations are also feasible, e.g. polynomials or rational functions. There are several reasons for selecting the Fourier representation rather than other possibilities. Firstly, $\lambda(t)$ is known to be approximately sinusoidal in shape and consequently we can expect that even for small values of L, the approximation $\hat{\lambda}(t,L) \approx \lambda(t)$ will be reasonably accurate. Secondly, $\hat{\lambda}(t,L)$ is periodic, which is a property that $\lambda(t)$ is known to have. Thirdly, the individual components in the representation are orthogonal, which is a convenient mathematical property.

APPENDIX C

The Cholesky Decomposition.

For A an $(n \times n)$ symmetric, positive definite matrix, there exists a unique lower triangular matrix F with positive diagonal elements such that

$$A = FF^T .$$

This is known as the Cholesky decomposition. An algorithm to reduce a matrix A to its Cholesky decomposition is given below.

Notation.

f_{ij} is the ijth element of the matrix F.

a_{ij} is the ijth element of the matrix A.

Algorithm

Step 1: Set $f_{11} = \sqrt{a_{11}}$.

Step 2: For $j=2,3,\ldots,n$

Set $f_{j1} = a_{1j}/f_{11}$

Next j

Step 3: For $i=2,3,\ldots,n-1$

Set $f_{ii} = \sqrt{a_{ii} - \sum_{j=1}^{i-1} f_{ij}^2}$.

For $j = i+1, i+2,\ldots,n$

Set $f_{ji} = (a_{ij} - \sum_{k=1}^{i-1} f_{ik} f_{jk})/f_{ii}$.

Next j .

Next i .

Step 4: Set $f_{nn} = \sqrt{a_{nn} - \sum_{j=1}^{n-1} f_{nj}^2}$.

Step 5: End.

The elements of F above the main diagonal are defined to be zero. The above algorithm does not set them to zero, so if necessary the following step should be inserted immediately preceding "Next j" in Step 3:

Set $f_{ij} = 0$.

APPENDIX D

Obtaining maximum likelihood estimates in Model 3 (ML) .

Given the likelihood function

$$\ell(\psi) = -T \log(\sqrt{2\pi}) - C(DD)/2 \log \sigma_e^2(DD) -$$

$$C(WW)/2 \log \sigma_e^2(WW) - C(DW)/2 \log \sigma_e^2(DW) -$$

$$C(WD)/2 \log \sigma_e^2(WD) - 1/2[1/\sigma_e^2(DD) \sum_{t\in N(DD)} \beta^2(DD) +$$

$$1/\sigma_e^2(WW) \sum_{t\in N(WW)} \beta^2(WW) + 1/\sigma_e^2(DW) \sum_{t\in N(DW)} \beta^2(DW) +$$

$$1/\sigma_e^2(WD) \sum_{t\in N(WD)} \beta^2(WD)] \quad ,$$

the maximum likelihood estimates of the parameters Θ^{DD}, Θ^{WW}, Θ^{DW}, Θ^{WD}, $\sigma_e^2(DD)$, $\sigma_e^2(WW)$, $\sigma_e^2(DW)$ $\sigma_e^2(WD)$, α_j^D, α_j^W, are obtained by differentiating $\ell(\psi)$ with respect to the various parameters and setting the derivatives equal to zero.

The parameter estimates for Θ^{DD} , $\sigma_e^2(DD)$ and α_j^D are derived here.

Setting

$\partial\ell(\psi)/\partial\Theta^{DD} = 0$ gives that

$$- 1/2 \left\{ 2/\sigma_e^2(DD) \sum_{t \in N(DD)} \beta(DD) \cdot \left[- S_{i,t-1} + \mu_{t-1}^D \right] \right\} = 0 .$$

Now $\beta(DD) = S_{i,t} + \Theta^{DD}(- S_{i,t-1} + \mu_{t-1}^D) - \mu_t^D$

and so

$$- \sum_{t \in N(DD)} (S_{i,t} - \mu_t^D)(\mu_{t-1}^D - S_{i,t-1}) = \Theta^{DD} .$$

$$\sum_{t \in N(DD)} (-S_{i,t-1} + \mu_{t-1}^D)^2$$

giving that

$$\hat{\Theta}^{DD} = \frac{\sum_{t \in N(DD)} (\hat{\mu}_t^D - S_{i,t})(\hat{\mu}_{t-1}^D - S_{i,t-1})}{\sum_{t \in N(DD)} (\hat{\mu}_{t-1}^D - S_{i,t-1})^2} .$$

Now

$$\partial \ell(\psi)/\partial \sigma_e^2(DD) = \frac{-C(DD)}{2\sigma_e^2(DD)} + 1/2\sigma_e^4(DD) \sum_{t \in N(DD)} \beta^2(DD)$$

and setting this equal to zero gives that

$$\hat{\sigma}_e^2(DD) = 1/C(DD) \cdot \sum_{t \in N(DD)} \beta^2(DD) .$$

The parameter estimate for α_j^D is given by

$$\partial \ell(\psi)/\partial \alpha_j^D = 0$$

that is,

$$1/\sigma_e^2(DD) \sum_{t\in N(DD)} \beta(DD)[\varphi_j(t) + \Theta^{DD} \varphi_j(t-1)] +$$

$$1/\sigma_2^2(DW) \sum_{t\in N(DW)} \beta(DW)[\Theta^{DW} \varphi_j(t-1)] +$$

$$1/\sigma_e^2(WD) \sum_{t\in N(WD)} \beta(WD)[- \varphi_j(t)] = 0$$

that is

$$- 1/\sigma_e^2(DD) \sum_{t\in N(DD)} [(S_{i,t} - \Theta^{DD} S_{i,t-1}) - \mu_t^D$$

$$+ \Theta^{DD} \mu_{t-1}^D] \varphi_j(t) + \Theta^{DD}/\sigma_e^2(DD) \sum_{t\in N(DD)} [(S_{i,t} -$$

$$\Theta^{DD} S_{i,t-1}) - \mu_t^D + \Theta^{DD} \mu_{t-1}^D] \varphi_j(t-1) +$$

$$\Theta^{DW}/\sigma_e^2(DW) \sum_{t\in N(DW)} [(S_{i,t} - \Theta^{DW} S_{i,t-1}) - \mu_t^W +$$

$$\Theta^{DW} \mu_{t-1}^D] \varphi_j(t-1) - 1/\sigma_e^2(WD) \sum_{t\in N(WD)}$$

$$[(S_{i,t} - \Theta^{WD} S_{i,t-1}) - \mu_t^D + \Theta^{WD} \mu_{t-1}^W] \varphi_j(t) = 0 .$$

Let $A = -1/\sigma_e^2(DD) \sum_{t \in N(DD)} [(S_{i,t} - \Theta^{DD} S_{i,t-1}) \varphi_j(t) -$

$\Theta^{DD} (S_{i,t} - \Theta^{DD} S_{i,t-1}) \varphi_j(t-1)] + \Theta^{DW}/\sigma_e^2(DW) \; .$

$\sum_{t \in N(DW)} [(S_{i,t} - \Theta^{DW} S_{i,t-1}) - \mu_t^W] \varphi_j(t-1) -$

$1/\sigma_e^2(WD) \sum_{t \in N(WD)} [(S_{i,t} - \Theta^{WD} S_{i,t-1}) + \Theta^{WD} \mu_{t-1}^W] \varphi_j(t) \; .$

Then

$$A = 1/\sigma_e^2(DD) \sum_{t \in N(DD)} [- \mu_t^D + \Theta^{DD} \mu_{t-1}^D] [\varphi_j(t) -$$

$$\Theta^{DD} \varphi_j(t-1)] - \Theta^{DW}/\sigma_e^2(DW) \sum_{t \in N(DW)} [\Theta^{DW} \mu_{t-1}^D] \varphi_j(t-1) +$$

$$1/\sigma_e^2(WD) \sum_{t \in N(WD)} [- \mu_t^D] \varphi_j(t) \; .$$

$$= 1/\sigma_e^2(DD) \sum_{t \in N(DD)} \Bigg[- \alpha_j^D \varphi_j(t) - \sum_{\substack{k=1 \\ k \neq j}}^{L} \alpha_k^D \varphi_k(t) +$$

$$\Theta^{DD} (\alpha_j^D \varphi_j(t-1) + \sum_{\substack{k=1 \\ k \neq j}}^{L} \alpha_k^D \varphi_k(t-1)) \Bigg] [\varphi_j(t) -$$

$$\Theta^{DD} \varphi_j(t-1)] - \Theta^{DW}/\sigma_e^2(DW) \sum_{t \in N(DW)} \Bigg[\Theta^{DW} (\alpha_j^D \varphi_j(t-1) +$$

$$\left.\sum_{\substack{k=1\\k\neq j}}^{L} \alpha_k^D \varphi_k(t-1)\right] \varphi_j(t-1) + 1/\sigma_e^2(WD) \sum_{t\in N(WD)} \left[- \alpha_j^D \varphi_j(t) - \right.$$

$$\left.\sum_{\substack{k=1\\k\neq j}}^{L} \alpha_k^D \varphi_k(t)\right] \varphi_j(t) .$$

Let

$$M = + 1/\sigma_e^2(DD) \sum_{t\in N(DD)} \left[- \sum_{\substack{k=1\\k\neq j}}^{L} \alpha_k^D \varphi_k(t) + \right.$$

$$\left.\Theta^{DD} \left(\sum_{\substack{k=1\\k\neq j}}^{L} \alpha_k^D \varphi_k(t-1)\right)\right] [\varphi_j(t) - \Theta^{DD} \varphi_j(t-1)] -$$

$$\Theta^{DW}/\sigma_e^2(DW) \sum_{t\in N(DW)} \left[\Theta^{DW} . \sum_{\substack{k=1\\k\neq j}}^{L} \alpha_k^D \varphi_k(t-1)\right] \varphi_j(t-1) +$$

$$1/\sigma_e^2(WD) \sum_{t\in N(WD)} \left[- \sum_{\substack{k=1\\k\neq j}}^{L} \alpha_k^D \varphi_k(t)\right] \varphi_j(t) .$$

Then

$$M - A = \alpha_j^D [1/\sigma_e^2(DD) \sum_{t\in N(DD)} [\varphi_j(t) - \Theta^{DD} \varphi_j(t-1)]^2$$

$$+ \Theta^{DW}/\sigma_e^2(DW) \sum_{t\in N(DW)} \Theta^{DW} \varphi_j^2(t-1)$$

$$+ \Theta^{WD}/\sigma_e^2(WD) \sum_{t\in N(WD)} \varphi_j^2(t)]$$

and so

$$\hat{\alpha}_j^D = (M - A) \, . \, [1/\sigma_e^2(DD) \sum_{t \in N(DD)} [\varphi_j(t) - \Theta^{DD} \, \varphi_j(t-1)]^2 +$$

$$\Theta^{DW}/\sigma_e^2(DW) \sum_{t \in N(DW)} \Theta^{DW} \, \varphi_j^2(t-1) + \Theta^{WD}/\sigma_e^2(WD) \sum_{t \in N(WD)} \varphi_j^2(t)]^{-1} \quad .$$

Similarly, estimates of the other parameters can be obtained.

APPENDIX E

Obtaining ordinary least squares estimates in Model 3 (LS).

Ordinary least squares for the parameter Θ^{DD}, Θ^{WW}, Θ^{WD}, Θ^{DW}, α_j^D and α_j^W are obtained by minimizing the following function:

$$P = \sum_{t\in N(DD)} (S_{i,t} - \mu_t^D - \Theta^{DD} (S_{i,t-1} - \mu_{t-1}^D))^2 +$$

$$\sum_{t\in N(WW)} (S_{i,t} - \mu_t^W - \Theta^{WW} (S_{i,t-1} - \mu_{t-1}^W))^2 +$$

$$\sum_{t\in N(DW)} (S_{i,t} - \mu_t^W - \Theta^{DW} (S_{i,t-1} - \mu_{t-1}^D))^2 +$$

$$\sum_{t\in N(WD)} (S_{i,t} - \mu_t^D - \Theta^{WD} (S_{i,t-1} - \mu_{t-1}^W))^2$$

with respect to the parameters.

The results will be shown here for Θ^{DD} and α_j^D .

Now

$$\partial P/\partial\Theta^{DD} = 0 \quad \text{gives}$$

$$\sum_{t\in N(DD)} (S_{i,t} - \mu_t^D - \Theta^{DD} (S_{i,t-1} - \mu_{t-1}^D))(S_{i,t-1} - \mu_{t-1}^D) = 0$$

and so

$$\hat{\Theta}^{DD} = \frac{\sum_{t\in N(DD)} (S_{i,t} - \mu_t^D)(S_{i,t-1} - \mu_{t-1}^D)}{\sum_{t\in N(DD)} (S_{i,t-1} - \mu_{t-1}^D)^2} .$$

And

$\partial P/\partial \alpha_j^D = 0$ gives

$$\sum_{t\in N(DD)} [S_{i,t} - \mu_t^D - \Theta^{DD} (S_{i,t-1} - \mu_{t-1}^D)][-\varphi_j(t)$$

$$+ \Theta^{DD} \varphi_j(t-1)] + \sum_{t\in N(DW)} [S_{i,t} - \mu_t^W -$$

$$\Theta^{DW} (S_{i,t-1} - \mu_{t-1}^D)] [\Theta^{DW} \varphi_j(t-1)] + \sum_{t\in N(WD)}$$

$$[S_{i,t} - \mu_t^D - \Theta^{WD} (S_{i,t-1} - \mu_{t-1}^W)] [-\varphi_j(t)] = 0$$

that is,

$$\sum_{t\in N(DD)} (S_{i,t} - \Theta^{DD} . S_{i,t-1})(- \varphi_j(t) + \Theta^{DD} \varphi_j(t-1)) +$$

$$\sum_{t\in N(DW)} (S_{i,t} - \mu_t^W - \Theta^{DW} S_{i,t-1})(\Theta^{DW} \varphi_j(t-1)) -$$

$$\sum_{t\in N(WD)} (S_{i,t} - \Theta^{WD}(S_{i,t-1} - \mu_{t-1}^W))(\varphi_j(t)) =$$

$$\sum_{t\in N(DD)} (\mu_t^D - \Theta^{DD} \mu_{t-1}^D)(-\varphi_j(t) + \Theta^{DD} \varphi_j(t-1)) -$$

$$\sum_{t\in N(DW)} (\Theta^{DW})^2 \mu_{t-1}^D \varphi_j(t-1) - \sum_{t\in N(WD)} \mu_t^D \varphi_j(t) .$$

Denote LHS of above equation by A , then

$$A = \sum_{t \in N(DD)} \Big(\alpha_j^D \varphi_j(t) + \sum_{\substack{k=1 \\ k \neq j}}^{L} \alpha_k^D \varphi_k(t) - \Theta^{DD} \alpha_j^D \varphi_j(t-1) -$$

$$\Theta^{DD} \sum_{\substack{k=1 \\ k \neq j}}^{L} \alpha_k^D \varphi_k(t-1)\Big)\big(-\varphi_j(t) + \Theta^{DD} \varphi_j(t-1)\big) -$$

$$\sum_{t \in N(DW)} (\Theta^{DW})^2 \Big(\alpha_j^D \varphi_j(t-1) + \sum_{\substack{k=1 \\ k \neq j}}^{L} \alpha_k^D \varphi_k(t-1)\Big) \varphi_j(t-1) -$$

$$\sum_{t \in N(WD)} \Big(\alpha_j^D \varphi_j(t) + \sum_{\substack{k=1 \\ k \neq j}}^{L} \alpha_k^D \varphi_k(t)\Big) \varphi_j(t) .$$

Let

$$M = \sum_{t \in N(DD)} \Big(\sum_{\substack{k=1 \\ k \neq j}}^{L} \alpha_k^D \varphi_k(t) - \Theta^{DD} \sum_{\substack{k=1 \\ k \neq j}}^{L} \alpha_k^D \varphi_k(t-1)\Big)\big(-\varphi_j(t) -$$

$$\Theta^{DD} \varphi_j(t-1)\big) - \sum_{t \in N(DW)} (\Theta^{DW})^2 \sum_{\substack{k=1 \\ k \neq j}}^{L} \alpha_k^D \varphi_k(t-1)\, \varphi_j(t-1) -$$

$$\sum_{t \in N(WD)} \sum_{\substack{k=1 \\ k \neq j}}^{L} \alpha_k^D \varphi_k(t)\, \varphi_j(t) .$$

Then

$$A - M = \sum_{t \in N(DD)} (\alpha_j^D \varphi_j(t) - \Theta^{DD} \alpha_j^D \varphi_j(t-1)) \,.$$

$$(- \varphi_j(t) + \Theta^{DD} \varphi_j(t-1)) - \sum_{t \in N(DW)} (\Theta^{DW})^2 \alpha_j^D (\varphi_j(t-1))^2 -$$

$$\sum_{t \in N(WD)} \alpha_j^D (\varphi_j(t))^2 \,.$$

And so

$$\hat{\alpha}_j^D = (A - M) \left\{ \sum_{t \in N(DD)} (\varphi_j(t) - \Theta^{DD} \varphi_j(t-1))^2 - \right.$$

$$\left. (\Theta^{DW})^2 \sum_{t \in N(DW)} (\varphi_j(t-1))^2 - \sum_{t \in N(WD)} (\varphi_j(t))^2 \right\}^{-1} \,.$$

Similarly, estimates for other parameters can be obtained.

APPENDIX F

Programs.

```
C       PROGRAM 1
C       ---------
C
C
C       THIS PROGRAM CALCULATES THE VECTORS REQUIRED FOR
C       PARAMETER ESTIMATION.  THESE VECTORS ARE :
C
       INTEGER          NT
       PARAMETER        NT = 365
       INTEGER          NY
       PARAMETER        NY = 7
       INTEGER          NRT
       PARAMETER        NRT = 30
       INTEGER          IND
       INTEGER          N (NT)
       INTEGER          NR (NT)
       INTEGER          NW (NT)
       INTEGER          NWW (NT)
       INTEGER          ND (NT)
       INTEGER          NDW (NT)
       INTEGER          R (NRT, NT)
       INTEGER          RAIN
       INTEGER          T

  5    FORMAT ( )
 15    FORMAT (7 (I4))
 25    FORMAT (14 (I5))

C      THE REQUIRED VECTORS ARE CALCULATED.
C
       DO 10, I = 1, NT
          N (I) = 0
          NR (I) = 0
          NW (I) = 0
          NWW (I) = 0
          ND (I) = 0
          NDW (I) = 0
          DO 70, J = 1, NRT
             R (J, I) = 0
 70       CONTINUE
 10    CONTINUE
       IND = -1
       DO 20, J = 1, NY
       DO 30, I=1, NT
          READ 5, RAIN
          IF (RAIN .EQ. 0) THEN
             N(I) = N(I) + 1
             IF (IND .EQ. 0) THEN
                ND(I) = ND(I) + 1
             ELSEIF (IND .EQ. 1) THEN
                NW(I) = NW(I) + 1
                IND = 0
```

```
                 ELSEIF (IND .EQ. -1) THEN
                    IND = 0
                 ENDIF
              ELSEIF (RAIN .GT. 0) THEN
                 NR(I) = NR(I) + 1
                 R (NR(I), I) = RAIN
              IF (IND .EQ. 0) THEN
                 NDW(I) = NDW(I) + 1
                 IND = 1
              ELSEIF (IND .EQ. 1) THEN
                 NWW(I) = NWW(I) + 1
              ELSEIF (IND .EQ. -1) THEN
                 IND = 1
              ENDIF
           ELSEIF (RAIN .LT. 0) THEN
              IND = -1
           ENDIF
 30        CONTINUE
 20     CONTINUE
        DO 40, I = 1, NT
           N(I) = N(I) + NR(I)
           ND(I) = ND(I) + NDW(I)
           NW(I) = NW(I) + NWW(I)
 40     CONTINUE
C
C       THE VECTORS CALCULATED ARE WRITTEN OUT.
C
        DO 50, I = 1, NT
           WRITE (10,15) I, NW(I), NWW(I), ND(I), NDW(I),
     &                   N(I), NR(I)
 50     CONTINUE
        DO 60, T = 1, NT
           WRITE (20, 25) (R (I, T), I = 1, NR (T))
 60     CONTINUE
        STOP
        END


$C     PROGRAM 2
$C     ---------
$C
$C
$C     THIS IS A GLIM PROGRAM TO ESTIMATE THE FOURIER
$C     COEFFICIENTS FOR A FUNCTION WHICH HAS BEEN
$C     APPROXIMATED BY FOURIER SERIES.  THE MODEL BEING
$C     FITTED IS GIVEN BY :
$C                      GAMMA(I) * PHI(I, T)
$C              WHERE :
$C                      PHI(1, T) = 1
$C                      PHI(2, T) = COS(2*PI*T / 365)
$C                      PHI(3, T) = SIN(2*PI*T / 365)
$C                      PHI(4, T) = COS(4*PI*T / 365)
$C                      ETC.
$C                      FOR T = 1, 2, ........, 365
```

```
$CAL X1 = %SIN (2 * %PI / 365 * (T - 1) + %PI / 2) $
$CAL X2 = %SIN (2 * %PI / 365 * (T - 1)) $
$CAL X3 = %SIN (2 * %PI / 365 * 2 * (T - 1) + %PI / 2) $
$CAL X4 = %SIN (2 * %PI / 365 * 2 * (T - 1)) $
$CAL X5 = %SIN (2 * %PI / 365 * 3 * (T - 1) + %PI / 2) $
$CAL X6 = %SIN (2 * %PI / 365 * 3 * (T - 1)) $
$CAL X7 = %SIN (2 * %PI / 365 * 4 * (T - 1) + %PI / 2) $
$CAL X8 = %SIN (2 * %PI / 365 * 4 * (T - 1)) $
$CAL X9 = %SIN (2 * %PI / 365 * 5 * (T - 1) + %PI / 2) $
$CAL X10 = %SIN (2 * %PI / 365 * 5 * (T - 1)) $
$CAL X11 = %SIN (2 * %PI / 365 * 6 * (T - 1) + %PI / 2) $
$CAL X12 = %SIN (2 * %PI / 365 * 6 * (T - 1)) $
$YVAR NT
$C
$C BINOMIAL ERROR
$C
$ERROR B N
$C
$C LOGIT LINK
$C
$LINK G
$C
$C FITTING HIERARCHY MODELS
$FIT : $
$DISPLAY M A D$
$FIT : + X1, X2 $
$DISPLAY M A R D$
$FIT X1, X2 : + X3, X4 $
$DISPLAY M A D$
$FIT X1, X2, X3, X4 : + X5, X6 $
$DISPLAY M A D$
$FIT X1, X2, X3, X4, X5, X6 : + X7, X8 $
$DISPLAY M A D$
$FIT X1, X2, X3, X4, X5, X6, X7, X8 : + X9, X10 $
$DISPLAY M A D$
$FIT X1, X2, X3, X4, X5, X6, X7, X8, X9, X10 : + X11, X12 $
$DISPLAY M A D$
$C
$C
$C THE RUNSTREAM TO SUBMIT ABOVE PROGRAM
$C
$C  $ECHO
$C  $UNITS 365
$C  $DATA T P P N NT P P
$C  $READ
$C  ADD THE DATA SET
```

```
C PROGRAM 3
C ---------
C
C
C THIS PROGRAM FINDS THE PARAMETER ESTIMATES FOR
C THE DISTRIBUTION OF RAINFALL DEPTHS.
C THE FOLOWING SUBROUTINES ARE USED WITHIN THE
C PROGRAM :
C COSSIN ---  COMPUTE THE COSINE AND SINE TERMS OF THE
C             FOURIER SERIES
 C SOLN   ---  COMPUTES THE SOLUTION TO A SYSTEM OF LINEAR
 C        ---  EQUATIONS
 C COEVAR ---  COMPUTES THE COEFFICIENT OF VARIATION
        EXTERNAL            COSSIN
        EXTERNAL            SOLN
        EXTERNAL            COEVAR
 C NUMBER OF PARAMETERS IN THE MEAN FUNCTION MODEL
        INTEGER             NPARM
        PARAMETER           NPARM = 3
        INTEGER             NT
        PARAMETER           NT = 365
        INTEGER             NR (NT)
 C MATRIX WITH RAINFALL DEPTHS OBSERVED
        INTEGER             R (50, NT)
        INTEGER             T
 C CURRENT ITERATION
        INTEGER             ITER
        REAL                COEFF
 C VECTOR OF 1ST PARTIAL DERIVATIVES
        REAL                DERIV (NPARM)
 C VECTOR OF 2ND PARTIAL DERIVATIVES
        REAL                DERIV2 (NPARM, NPARM)
        REAL                SOLUTN (NPARM)
 C MATRIX WITH FOURIER SERIES COSINE & SINE TERMS
        REAL                PHI (NPARM, NT)
 C AVERAGE OBSERVED RAINFALL IN EACH PERIOD
        REAL                Q (NT)
        REAL                TERM0
        REAL                TERM1
 C VECTOR OF PARAMETER ESTIMATES
        REAL                PAR (NPARM)
        REAL                PI
        PARAMETER           PI = 3.141593
 C CORRESPONDING AMPLITUDES
        REAL                AM (0 : NPARM)
        REAL                TA
        REAL                TB
 C CORRESPONDING PHASES
        REAL                PH (NPARM)
 C CURRENT ESTIMATE OF THE MEAN
        REAL                F (NT)
 C OBSERVED DAILY STANDARD DEVIATIONS
        REAL                SO (NT)
 C FITTED DAILY STANDARD DEVIATIONS
        REAL                SF (NT)
 C CONVERGENCE CRITERION
        REAL                DELTA
```

```
      DO 10, T = 1, NT
         READ 5, NR(T)
 10   CONTINUE
      DO 20, T = 1, NT
         READ 25, (R (I, T), I = 1, NR(T))
 20   CONTINUE
      CALL COSSIN (PHI, NPARM, NT)
      DO 30, T = 1, NT
         IF (NR(T) .GT. 0) THEN
            TERM0 = 0
            DO 40, I = 1, NR (T)
               TERM0 = TERM0 + R (I, T)
 40         CONTINUE
            Q (T) = TERM0 / NR (T)
         ENDIF
 30   CONTINUE
      DO 50, I = 1, NPARM
         TERM0 = 0
         TERM1 = 0
         DO 60, T = 1, NT
            IF (NR (T) .NE. 0) THEN
               TERM0 = TERM0 + Q (T) * PHI (I, T)
               TERM1 = TERM1 + PHI (I, T) ** 2
            ENDIF
 60      CONTINUE
         PAR (I) = TERM0 / TERM1
 50   CONTINUE
      ITER = 0
      DO 100, N = 1, 50
         DO 70, I = 1, NPARM
            DERIV (I) = 0
            DO 80, J = 1, I
               DERIV2 (I, J) = 0
 80         CONTINUE
 70      CONTINUE
         DO 90, T = 1, NT
            TERM0 = PAR (1)
            DO 180, I = 2, NPARM
               TERM0 = TERM0 + PAR (I) * PHI (I, T)
180         CONTINUE
            F (T) = TERM0
 90      CONTINUE
         DO 110, T = 1, NT
            IF (NR (T) .GT. 0) THEN
               DO 120, I = 1, NPARM
                  DERIV (I) = DERIV (I) - NR (T) *
     &                        (Q (T) - F (T)) * PHI (I, T)
                   DO 130, J = 1, I
                      DERIV2 (I, J) = DERIV2 (I, J) +
     &                                NR (T) * PHI (I, T)
     &                                * PHI (J, T)
 130               CONTINUE
 120           CONTINUE
             ENDIF
 110      CONTINUE
          DO 140, I = 1, NPARM
             DO 150, J = I + 1, NPARM
                DERIV2 (I, J) = DERIV2 (J, I)
 150         CONTINUE
 140      CONTINUE
          CALL SOLN (DERIV, DERIV2, SOLUTN)
```

```
            DO 160, I = 1, NPARM
               PAR (I) = PAR (I) - SOLUTN (I)
160         CONTINUE
            ITER = ITER + 1
            DELTA = 0
C TESTING FOR CONVERGENCE
            DO 170, I = 1, NPARM
               DELTA = DELTA + ABS (DERIV (I))
170         CONTINUE
            IF (DELTA .LE. 0.0001) THEN
               GOTO 2
            ENDIF
100      CONTINUE
2        DO 200, T = 1, NT
            TERMO = PAR (1)
            DO 220, I = 2, NPARM
               TERMO = TERMO + PAR (I) * PHI (I, T)
220         CONTINUE
C COMPUTED FITTED VALUES
            F (T) = TERMO
200      CONTINUE
C PRINTING OUT OBSERVED AND FITTED DAILY MEANS
         DO 300, T = 1, NT
            PRINT 25, T, Q (T), F (T)
300      CONTINUE
         PRINT 25, 'PAR', (PAR (I), I = 1, NPARM)
         CALL COEVAR (R, F, NR, COEFF, NT, NPARM)
         D = ATAN2 (-1 * PAR(3), PAR(2))
         DD = -1 * D * NT / (2 * PI * 1)
C COMPUTE THE AMPLITUDE AND PHASE REPRESENTATION
         AM (0) = PAR (1)
         K = (NPARM - 1) / 2
         DO 190, I = 1, K
            TA = PAR (2 * I)
            TB = PAR (2 * I + 1)
            C = NT / (2 * PI * I)
            A = ATAN (TB / TA)
C           PRINT 25, 'A', A
            AM (I) = SQRT (TA ** 2 + TB ** 2)
            IF (TA .LT. 0) THEN
               PH (I) = A + PI
            ELSEIF (TA .EQ. 0) THEN
               IF (TB .GE. 0) THEN
                  PH (I) = PI / 2
               ELSE
                  PH (I) = 1.5 * PI
               ENDIF
            ELSEIF (TA .GT. 0) THEN
               IF (TB .GE. 0) THEN
                  PH (I) = A
               ELSE
                  PH (I) = A * 2 * PI
               ENDIF
            ENDIF
            PH (I) = PH (I) * C
190      CONTINUE
         FF = 2 * PI / NT
         PRINT 25, 'AM', (AM (I), I = 0, K)
         PRINT 25, 'PH', (PH (I), I = 1, K)
```

```
C COMPUTING THE FITTED AND OBSERVED STANDARD DEVIATIONS
        DO 310, T = 1, NT
           SF (T) = COEFF * F (T)
           DO 320, I = 1, NR (T)
              SO (T) = (R (I, T) - Q (T)) ** 2 + SO (T)
320        CONTINUE
           SO (T) = SQRT (SO (T) / (NR (T) - 1))
           PRINT 25, T, SO (T), SF (T)
310     CONTINUE
        STOP
        END




C PROGRAM 4
C ---------
C
C
C PROGRAM TO SIMULATE RAINFALL SEQUENCES
C
        INTEGER         NT
        PARAMETER       NT = 365
C NUMBER OF PARAMETERS IN THE MEAN FUNCTION
        INTEGER         NPARM
        PARAMETER       NPARM = 3
C TOTAL NUMBER OF OBSERVATIONS TO BE SIMULATED
         INTEGER         TIME
         PARAMETER       TIME = 7300
 C STATE OF THE PRESENT  DAY
         INTEGER         PSTATE
 C STATE OF THE PREVIOUS DAY
         INTEGER         STATE
         INTEGER         K
         INTEGER         T
         INTEGER         A
 C THE STARTING SEED FOR A RANDOM NUMBER GENERATOR
         INTEGER         SEED (9)
         REAL            RAIN
 C VECTOR WITH PARAMETER ESTIMATES FOR THE PROBABILITY
 C OF A WET DAY OCCURING GIVEN A DRY DAY HAS OCCURRED
 C AND FOR THE PROBABILITY THAT A WET DAY OCCURS GIVEN
 C A WET HAS OCCURRED
         REAL            GAMMA (2, NPARM)
 C MATRIX WITH FOURIER SERIES COSINE AND SINE TERMS
         REAL            PHI (NPARM, NT)
         REAL            PI
 C UNIFORM RANDOM NUMBER BETWEEN 0 AND 1
         REAL            UNIFOR
 C THE AMPLITUDES OF THE FOURIER REPRESENTATION OF THE
 C MEAN FUNCTION
         REAL            AMM (0 : NPARM)
 C THE PHASES OF THE FOURIER REPRESENTATION OF THE
 C MEAN FUNCTION
         REAL            PHM (NPARM)
 C THE COEFFICIENT OF VARIATION
         REAL            C
```

```
C       MAIN PROGRAM
C       ------------

  5     FORMAT ( )
 15     FORMAT (F7.3, 4 (F9.3), 2 (F10.3), F9.3)

C INPUTTING THE DATA SETS
        DO 50, I = 1, 2
           READ 5, (GAMMA (I, J), J = 1,NPARM)
 50     CONTINUE
        READ 5, (AMM (I), I = 0,1)
        READ 5, (PHM (I), I = 1, 1)
        READ 5, C
C COMPUTING THE FOURIER SERIES TERMS
        CALL COSSIN (PHI, NPARM, NT)
        SEED (8) = 6831145
        SEED (9) = 33501959
        A = 8
C INITIAL STATE OF THE DAY IS DRY
        STATE = 1
        DO 20, T = 1, TIME
           K = T
           DO 30, I = 1, 100
              IF ( K .GT. 365) THEN
                 K = K - 365
              ELSE
                 GOTO 2
              ENDIF
 30        CONTINUE
C COMPUTE THE PROBABILITY OF GETTING A WET DAY GIVEN
C THE STATE OF THE PREVIOUS DAY
 2           CALL PIEST (NPARM, GAMMA, STATE, K, PHI, PI)
C GENERATE A RANDOM UNIFORM NUMBER BETWEEN 0 AND 1
           UNIFOR = UNIF (SEED, A)
            IF (UNIFOR .LT. PI) THEN
 C STATE OF PRESENT DAY IS WET
               PSTATE = 2
            ELSE
 C STATE OF PRESENT DAY IS DRY
               PSTATE = 1
            ENDIF
            IF (PSTATE .EQ. 1) THEN
               RAIN = 0
            ELSE
               RAIN = 1
            ENDIF
 C IF RAIN = 1 THEN THE DEPTH OF RAINFALL IS SIMULATED
            CALL DEPTH (SEED, NPARM, RAIN, K, AMM, PHM, C)
            WRITE (10, 15) RAIN
 C STATE OF THE PREVIOUS DAY IS UPDATED
            IF (PSTATE .NE. STATE) THEN
               STATE = PSTATE
            ENDIF
  20     CONTINUE
         STOP
         END
```

```
C PROGRAM 5
C ---------
C
C
C PROGRAM TO CONDITION THE DATA SETS INTO THE DRY OR
C WET STATUS OF THE DAY
C
C TOTAL NUMBER OF OBSERVATIONS
      INTEGER          TIME
      PARAMETER        TIME = 2190
C NUMBER OF VARIABLES
      INTEGER          NV
      PARAMETER        NV = 7
      REAL             RAIN (2190)
      REAL             CLIMA (NV)
C VECTOR CONTAINING MISSING VALUE NOTATION
      REAL             MISS (NV)

   5    FORMAT ( )
  15    FORMAT (7 (F 10.3))

        DO 10, J = 1, NV
           MISS (J) = -999.0000
  10    CONTINUE
        DO 100, I = 1, TIME
           READ 5, RAIN (I)
 100    CONTINUE
        DO 20, I = 1, TIME
           READ 5, (CLIMA (K), K = 1, NV)
           IF (RAIN (I) .EQ. 0) THEN
              WRITE (10, 15) (CLIMA (K), K = 1, NV)
              WRITE (20, 15) (MISS (K), K = 1, NV)
           ELSEIF (RAIN (I) .GT. 0) THEN
              WRITE (10, 15) (MISS (K), K = 1, NV)
              WRITE (20, 15) (CLIMA (K), K = 1, NV)
           ELSEIF (RAIN (I) .EQ. -999) THEN
              WRITE (10, 15) (MISS (K), K = 1, NV)
              WRITE (20, 15) (MISS (K), K = 1, NV)
           ENDIF
  20    CONTINUE
        STOP
        END
```

```
C PROGRAM 6
C ---------
C
C
C PROGRAM TO COMPUTE THE DAILY MEANS OF THE OBSERVED SERIES.
C THIS IS DONE FOR THE TWO CONDITIONED DATA SERIES IN ONE RUN.

C NUMBER OF VARIABLES
      INTEGER       NV
      PARAMETER     NV = 7
C NUMBER OF YEARS OBSERVED
      INTEGER       NY
      PARAMETER     NY = 6
      INTEGER       NT
      PARAMETER     NT = 365
       INTEGER       DENOM (NV, NT)
       REAL          CLIMA (NV)
       REAL          MEAN (NV, NT)

  5    FORMAT ( )
C 5    FORMAT (3 (10X), F 10.3)
 15    FORMAT (7 (F 10.4))

       DO 80, M = 1, 2
          DO 10, I = 1, NV
             DO 20, J = 1, NT
                MEAN (I, J) = 0
                DENOM (I, J) = NY
 20          CONTINUE
 10       CONTINUE
          DO 30, J = 1, NY
             DO 40, I = 1, NT
                READ 5, (CLIMA (K), K = 1, NV)
                DO 50, K = 1, NV
                   IF (CLIMA (K) .NE. -999) THEN
                      MEAN (K, I) = MEAN (K, I) + CLIMA (K)
                   ELSE
                      DENOM (K, I) = DENOM (K, I) - 1
                   ENDIF
 50             CONTINUE
 40          CONTINUE
 30       CONTINUE
          DO 60, I = 1, NT
             DO 70, K = 1, NV
                IF (MEAN (K, I) .NE. 0) THEN
                   MEAN (K, I) = MEAN (K, I) / DENOM (K, I)
                ELSE
                   MEAN (K, I) = -999.0000
                ENDIF
 70          CONTINUE
 60       CONTINUE
          IF (M .EQ. 1) THEN
             DO 90, I = 1, NT
                WRITE (10, 15) (MEAN (K, I), K = 1, NV)
 90          CONTINUE
          ELSE
             DO 100, I = 1, NT
                WRITE (20, 15) (MEAN (K, I), K = 1, NV)
100          CONTINUE
          ENDIF
 80    CONTINUE
```

PROGRAM 7

GENSTAT PROGRAM TO ESTIMATE THE PARAMETERS OF THE FOURIER APPROXIMATIONS FOR THE MEAN AND STANDARD DEVIATION FUNCTIONS.

```
'REFERENCE / NUNN = 10'    GLIMMODELS
'UNITS'        $365
 'VARIATE'      TIME = 1...365
  :             X(1...12) $ 365
 'READ'         OBSN (1...7)
 'SCALAR'       PI = 3.141592654.
 'COMMENT'      FOURIER TERMS ARE COMPUTED
 'CALCULATE'    X(1, 3, 5, 7, 9, 11) = COS (2 * PI / 365 *
                (1, 2, 3, 4, 5, 6) * (TIME - 1))
  :             X(2, 4, 6, 8, 10, 12) = SIN (2 * PI / 365 *
                (1, 2, 3, 4, 5, 6) * (TIME - 1))
  :             TM = -TIME
 'TERMS'        OBSN (1), X(1...12)
 'COMMENT'      MODEL SPECIFICATIONS :
 'Y / ERROR = NORMAL, LINK = IDENTITY' OBSN (1)
 'COMMENT'      HIERARCHICAL MODELS ARE FITTED FOR THE
  :             THE PURPOSE OF MODEL SELECTION
 'FIT / PRINT = CA'
 'ADD / PRINT = CAU, ANDEV = I' X(1), X(2); FVAL = FITTED
 'ADD / PRINT = CA, ANDEV = I'  X(3), X(4)
  :                             X(5), X(6).
  :                             X(7), X(8)
  :                             X(9), X(10)
 'ADD / PRINT = CA, ANDEV = T' X(11), X(12)
 'COMMENT' SECOND VARIABLE BEING FITTED
 'TERMS'        OBSN (2), X(1...12)
 'Y / ERROR = NORMAL, LINK = IDENTITY' OBSN (2)
 'FIT / PRINT = CA'
 'ADD / PRINT = CAU, ANDEV = I' X(1), X(2); FVAL = FITTED
 'ADD / PRINT = CA, ANDEV = I'  X(3), X(4)
  :                             X(5), X(6)
  :                             X(7), X(8)
  :                             X(9), X(10)
 'ADD / PRINT = CA, ANDEV = T' X(11), X(12)
 'TERMS'        OBSN (3), X(1...12)
 'Y / ERROR = NORMAL, LINK = IDENTITY' OBSN (3)
 'FIT / PRINT = CA'
 'ADD / PRINT = CAU, ANDEV = I' X(1), X(2); FVAL = FITTED
 'ADD / PRINT = CA, ANDEV = I'  X(3), X(4)
  :                             X(5), X(6)
  :                             X(7), X(8)
  :                             X(9), X(10)
 'ADD / PRINT = CA, ANDEV = T' X(11), X(12)
 'TERMS'        OBSN (4), X(1...12)
 'Y / ERROR = NORMAL, LINK = IDENTITY' OBSN (4)
 'FIT / PRINT = CA'
 'ADD / PRINT = CAU, ANDEV = I' X(1), X(2); FVAL = FITTED
 'ADD / PRINT = CA, ANDEV = I'   X(3), X(4)
   :                             X(5), X(6)
   :                             X(7), X(8)
   :                             X(9), X(10)
  'ADD / PRINT = CA, ANDEV = T' X(11), X(12)
  'TERMS'       OBSN (5), X(1...12)
```

```
'Y / ERROR = NORMAL, LINK = IDENTITY' OBSN (5)
'FIT / PRINT = CA'
'ADD / PRINT = CAU, ANDEV = I' X(1), X(2); FVAL = FITTED
'ADD / PRINT = CA, ANDEV = I'  X(3), X(4)
 :                             X(5), X(6)
 :                             X(7), X(8)
 :                             X(9), X(10)
'ADD / PRINT = CA, ANDEV = T' X(11), X(12)
'TERMS'       OBSN (6), X(1...12)
'Y / ERROR = NORMAL, LINK = IDENTITY' OBSN (6)
'FIT / PRINT = CA'
'ADD / PRINT = CAU, ANDEV = I' X(1), X(2); FVAL = FITTED
'ADD / PRINT = CA, ANDEV = I'  X(3), X(4)
 :                             X(5), X(6)
 :                             X(7), X(8)
 :                             X(9), X(10)
'ADD / PRINT = CA, ANDEV = T' X(11), X(12)
'TERMS'       OBSN (7), X(1...12)
'Y / ERROR = NORMAL, LINK = IDENTITY' OBSN (7)
'FIT / PRINT = CA'
'ADD / PRINT = CAU, ANDEV = I' X(1), X(2); FVAL = FITTED
'ADD / PRINT = CA, ANDEV = I'  X(3), X(4)
 :                             X(5), X(6)
 :                             X(7), X(8)
 :                             X(9), X(10)
'ADD / PRINT = CA, ANDEV = T' X(11), X(12)
```

```
C PROGRAM 8
C ---------
C
C
C PROGRAM TO COMPUTE THE STANDARS DEVIATIONS FROM THE
C RESIDUAL TIME SERIES.
C
C NUMBER OF VARIABLES
      INTEGER              NV
      PARAMETER            NV = 7
      INTEGER              NY
      PARAMETER            NY = 6
C NUMBER OF YEARS OBSERVED
      INTEGER              NT
      PARAMETER            NT = 365
       INTEGER              DENOM (NV, NT)
       REAL                 RESID (NV)
       REAL                 MEAN (NV, NT)

   5   FORMAT (7 (F 11.4))
  15   FORMAT (7 (F 11.2))

       DO 80, M = 1, 2
          DO 10, I = 1, NV
             DO 20, J = 1, NT
                MEAN (I, J) = 0
                DENOM (I, J) = NY
  20         CONTINUE
  10      CONTINUE
          DO 30, J = 1, NY
             DO 40, I = 1, NT
                READ 5, (RESID (K), K = 1, NV)
                DO 50, K = 1, NV
                   IF (RESID (K) .NE. -999) THEN
                      MEAN (K, I) = MEAN (K, I) +
     &                              RESID (K) ** 2
                   ELSE
                      DENOM (K, I) = DENOM (K, I) - 1
                   ENDIF
  50            CONTINUE
  40         CONTINUE
  30      CONTINUE
          DO 60, I = 1, NT
             DO 70, K = 1, NV
                IF (MEAN (K, I) .NE. 0) THEN
                   MEAN (K, I) = MEAN (K, I) / DENOM (K, I)
                ELSE
                   MEAN (K, I) = -999.0000
                ENDIF
  70         CONTINUE
  60      CONTINUE
          IF (M .EQ. 1) THEN
             DO 90, I = 1, NT
                WRITE (10, 15) (MEAN (K, I), K = 1, NV)
  90         CONTINUE
          ELSE
             DO 100, I = 1, NT
                WRITE (20, 15) (MEAN (K, I), K = 1, NV)
 100         CONTINUE
          ENDIF
  80   CONTINUE
C      STOP
```

```
'COMMENT'   PROGRAM 9
 :          ---------
 :
 :          GENSTAT PROGRAM TO COMPUTE THE AUTOCORRELATION
 :          WITHIN EACH VARIABLE AND TO PLOT THE CORRELOGRAM.
 :          THE PARTIAL AUTOCORRELATION AND THE PARTIAL
 :          AUTOCORRELATION PLOT IS ALSO GIVEN
 :
'REFERENCE' AUTOCORRELATION
'UNITS'     $ 2190
'VARIATE'   LAG = 0...15
 :          BVL = -1.0, +1.0, *, *
'COMMENT'   5% BOUNDARY LINES
 :          ULINE, ULINE2, ULINE3, ULINE4, ULINE5 $ 16
 :          ULINE6, ULINE7, MLINE, LLINE, LLINE2 $ 16
 :          LLINE3, LLINE4, LLINE5, LLINE6, LLINE7 $ 16
'READ'      RESID (1...7)
'CALC'      ULINE, LLINE = 2, -2 / SQRT (NVAL (RESID(1)))
 :          ULINE2, LLINE2 = 2, -2 / SQRT (NVAL (RESID(2)))
 :          ULINE3, LLINE3 = 2, -2 / SQRT (NVAL (RESID(3)))
 :          ULINE4, LLINE4 = 2, -2 / SQRT (NVAL (RESID(4)))
 :          ULINE5, LLINE5 = 2, -2 / SQRT (NVAL (RESID(5)))
 :          ULINE6, LLINE6 = 2, -2 / SQRT (NVAL (RESID(6)))
 :          ULINE7, LLINE7 = 2, -2 / SQRT (NVAL (RESID(7)))
 :          MLINE = 0
'DERIVE / ORDER = 15'  ACFRES (1...7) = ACF (RESID (1...7); S)
 :                     PACFRES (1...7) = PACF (ACFRES
                                  (1...7); PEV; PC)
'PRINT / P'    ACFRES (1...7) $ 12.5
'HEADING' HY1 = ''ACF OF RESIDUALS''
 :        HY2 = ''PARTIAL ACF OF RESIDUALS''
 :        HX = ''LAG''
 :        HPPPP = ''PPPP''
 :        HLAB1 = ''---*''
'GRAPH /INTX=Y,ATX=HX,ATY=HY1, BV = BVL'
        ULINE, MLINE, LLINE, ACFRES (1); LAG $ HPPPP; HLAB1
'GRAPH /INTX=Y,ATX=HX,ATY=HY2, BV = BVL'
        ULINE, MLINE, LLINE, ACFRES (1); LAG $ HPPPP; HLAB1
'GRAPH /INTX=Y,ATX=HX,ATY=HY1, BV = BVL'
        ULINE2, MLINE, LLINE2, ACFRES (2); LAG $ HPPPP; HLAB1
'GRAPH /INTX=Y,ATX=HX,ATY=HY2, BV = BVL'
        ULINE2, MLINE, LLINE2, ACFRES (2); LAG $ HPPPP; HLAB1
'GRAPH /INTX=Y,ATX=HX,ATY=HY1, BV = BVL'
        ULINE3, MLINE, LLINE3, ACFRES (3); LAG $ HPPPP; HLAB1
'GRAPH /INTX=Y,ATX=HX,ATY=HY2, BV = BVL'
        ULINE3, MLINE, LLINE3, ACFRES (3); LAG $ HPPPP; HLAB1
'GRAPH /INTX=Y,ATX=HX,ATY=HY1, BV = BVL'
        ULINE4, MLINE, LLINE4, ACFRES (4); LAG $ HPPPP; HLAB1
'GRAPH /INTX=Y,ATX=HX,ATY=HY2, BV = BVL'
        ULINE4, MLINE, LLINE4, ACFRES (4); LAG $ HPPPP; HLAB1
'GRAPH /INTX=Y,ATX=HX,ATY=HY1, BV = BVL'
        ULINE5, MLINE, LLINE5, ACFRES (5); LAG $ HPPPP; HLAB1
'GRAPH /INTX=Y,ATX=HX,ATY=HY2, BV = BVL'
        ULINE5, MLINE, LLINE5, ACFRES (5); LAG $ HPPPP; HLAB1
'GRAPH /INTX=Y,ATX=HX,ATY=HY1, BV = BVL'
        ULINE6, MLINE, LLINE6, ACFRES (6); LAG $ HPPPP; HLAB1
'GRAPH /INTX=Y,ATX=HX,ATY=HY2, BV = BVL'
        ULINE6, MLINE, LLINE6, ACFRES (6); LAG $ HPPPP; HLAB1
'GRAPH /INTX=Y,ATX=HX,ATY=HY1, BV = BVL'
        ULINE7, MLINE, LLINE7, ACFRES (7); LAG $ HPPPP; HLAB1
'GRAPH /INTX=Y,ATX=HX,ATY=HY2, BV = BVL'
        ULINE7, MLINE, LLINE7, ACFRES (7); LAG $ HPPPP; HLAB1
```

```
'COMMENT'   PROGRAM 10
 :          ----------
 :
 :
 :          GENSTAT PROGRAM TO COMPUTE THE CROSS
 :          CORRELATIONS BETWEEN THE VARIABLES.
 :
'REFERENCE' CROSSCORRELATION
'UNITS'     $ 2190
'COMMENT'   LAGS OF UP TO 15 TO BE COMPUTED
'VARIATE'   LAG = 0 ... 15
 :          BVL = -1.0, +1.0, *, *
 :          CCFRES (1 ...49)
 :          RES (1...7)
'READ'      RES (1...7)
'COMMENT'   COMPUTING THE CROSS CORRELATIONS
'DERIVE / ORDER = 15' CCFRES (1...7) = CCF (RES (1);
                      (RES (1); RES (1...7); S)
  :       CCFRES (8...14) = CCF (RES (2); RES (1...7); S)
  :       CCFRES (15...21) = CCF (RES (3); RES (1...7); S)
  :       CCFRES (22...28) = CCF (RES (4); RES (1...7); S)
  :       CCFRES (29...35) = CCF (RES (5); RES (1...7); S)
  :       CCFRES (36...42) = CCF (RES (6); RES (1...7); S)
  :       CCFRES (43...49) = CCF (RES (7); RES (1...7); S)
```

```
C PROGRAM 11
C ----------
C
C
C PROGRAM TO COMPUTE MATRICES A AND B WHEN MODEL 1 IS
C USED.
C
C SUBROUTINE TO INVERT A MATRIX
        EXTERNAL            INVERT
C SUBROUTINE TO TRANSPOSE A MATRIX
        EXTERNAL            TRANSP
C SUBROUTINE TO MULTIPLY MATRICES
        EXTERNAL            MULT
C SUBROUTINE TO SUBTRACT MATRICES
        EXTERNAL            SUBTR
C SUBROUTINE TO FIND THE CHOLESKY DECOMPOSITION
        EXTERNAL            CHOLES
C NUMBER OF VARIABLES
        INTEGER             NV
        PARAMETER           NV = 7
        REAL                LAG0 (NV, NV)
        REAL                LAG1 (NV, NV)
        REAL                INV (NV, NV)
        REAL                A (NV, NV)
        REAL                B (NV, NV)
        REAL                TRSP (NV, NV)
        REAL                TERM (NV, NV)
```

```
5       FORMAT ( )
15       FORMAT (7 (F 10.4))

        DO 20, I = 1, NV
           READ 5, (LAG0 (I, J), J = 1, NV)
20      CONTINUE
        DO 30, I = 1, NV
           READ 5, (LAG1 (I, J), J = 1, NV)
30      CONTINUE
        CALL INVERT (LAG0, INV)
        CALL TRANSP (LAG1, TRSP, NV)
        CALL MULT (LAG1, INV, A, NV, NV, NV, NV)
        CALL MULT (A, TRSP, TERM, NV, NV, NV, NV)
        CALL SUBTR (LAG0, TERM, NV)
        CALL CHOLES (TERM, B, NV)
        DO 40, I = 1, NV
           PRINT 15, (A (I, J), J = 1, NV)
40      CONTINUE
        DO 50, I = 1, NV
           PRINT 15, (B (I, J), J = 1, NV)
50      CONTINUE
        STOP
        END


C PROGRAM 12
C ----------
C
C
C PROGRAM TO GENERATE SYNTHETIC CLIMATE SEQUENCES USING
C MODEL 1.
C
        INTEGER         NT
        PARAMETER       NT = 365
C NUMBER OF VARIABLES
        INTEGER         NV
        PARAMETER       NV = 7
C NUMBER OF PARAMETERS IN THE MEAN FUNCTION
        INTEGER         NPARM
        PARAMETER       NPARM = 3
C TOTAL NUMBER OF VALUESTO BE SIMULATED
        INTEGER         TIME
        PARAMETER       TIME = 7300
C STATE OF THE DAY (I.E. DRY OR WET)
        INTEGER         STATE
        INTEGER         T
        INTEGER         AA
        INTEGER         SEED (9)
        REAL            RAIN
C VECTOR WITH PARAMETER ESTIMATES FOR THE OCCURRENCE
C OF RAINFALL
        REAL            GAMMA (2, NPARM)
C FOURIER TERMS
        REAL            PHI (NPARM, NT)
```

```
      REAL                PI
      REAL                STDDEV
      PARAMETER           STDDEV = 1
      REAL                MEAN
      PARAMETER           MEAN = 0
      REAL                A (NV, NV)
      REAL                B (NV, NV)
      REAL                RES (NV, 1)
      REAL                NRAND (NV, 1)
      REAL                UNIFOR
C ESTIMATES OF THE MEAN FUNCTION
      REAL                ALPHA (2, NPARM, NV)
C ESTIMATES OF THE STANDARD DEVIATION FUNCTION
      REAL                PSI (2, NPARM, NV)
      REAL                OBSN (NV)
C AMPLITUDE
      REAL                AMM (0 : NPARM)
C PHASE
      REAL                PHM (NPARM)
      REAL                C

C       MAIN PROGRAM
C       ------------

  5     FORMAT ( )
 15     FORMAT (F7.3, 4 (F9.3), 2 (F10.3), F9.3)

C READ IN MODEL PARAMETER ESTIMATES
        CALL READ (A, B, GAMMA, PSI, ALPHA, NPARM, NV,
     &            AMM, PHM, C)
C COMPUTE FOURIER TERMS
        CALL COSSIN (PHI, NPARM, NT)
        DO 10, I = 1, NV
           RES (I, 1) = 0
 10     CONTINUE
        SEED (1) = 115487
        SEED (2) = 257599
        SEED (3) = 6799501
        SEED (4) = 6237235
        SEED (5) = 7822589
        SEED (6) = 26875327
        SEED (7) = 69972223
        SEED (8) = 6831145
        SEED (9) = 33501959
        AA = 8
C INITIAL STATE OF THE DAY
        STATE = 1
        DO 20, T = 1, TIME
           KK = T
           DO 30, I = 1, 100
              IF ( KK .GT. 365) THEN
                 KK = KK - 365
              ELSE
                 GOTO 2
              ENDIF
 30        CONTINUE
```

```
C GENERATE RAINFALL VALUE
 2          CALL PIEST (NPARM, GAMMA, STATE, KK, PHI, PI)
            UNIFOR = UNIF (SEED, AA)
            IF (UNIFOR .LT. PI) THEN
               STATE = 2
            ELSE
               STATE = 1
            ENDIF
C GENERATE NORMAL RANDOM NUMBER WITH 0 MEAN AND STANDARD
C DEVIATION OF UNITY
            DO 50, K = 1, NV
               NRAND (K, 1) = GRAND (SEED, STDDEV, MEAN, K)
 50         CONTINUE
C           PRINT 5, (NRAND (K, 1), K = 1, NV)
C GENERATE CLIMATE SEQUENCE
            CALL MODEL (NRAND, STATE, NPARM, ALPHA, PSI, A, B,
     &                  RES, PHI, KK, OBSN)
            IF (STATE .EQ. 1) THEN
               RAIN = 0
            ELSE
               RAIN = 1
             ENDIF
 C IF IT RAINED THEN THE RAINFALL DEPTH IS GENERATED
             CALL DEPTH (SEED, NPARM, RAIN, KK, AMM, PHM, C)
             WRITE (10, 15) RAIN, (OBSN (I), I = 1, NV)
  20      CONTINUE
          STOP
          END
```

PROGRAM 13

PROGRAM TO COMPUTE THE VARIANCE OF THE VARIABLES. PROCEDURE 2D IS USED.

```
/PROBLEM
/INPUT          VARIABLE = 7.
                FORMAT = '(7(F10.4))'.
/VARIABLE       NAMES = MAXTEMP, MINTEMP, EVAPO, SUNSHINE,
                        WIND, MAXHUM, MINHUM.
                MISSING =  -999.0000, -999.0000, -999.0000,
                           -999.0000, -999.0000, -999.0000,
                           -999.0000.
/PRINT          NO COUNT.
 /END
```

```
  PROGRAM 14
C ----------

C PROGRAM TO GENERATE CLIMATE SEQUENCES WHEN MODEL 2 IS
C USED

        INTEGER        NT
        PARAMETER      NT = 365
C NUMBER OF VARIABLES
        INTEGER        NV
        PARAMETER      NV = 7
C NUMBER OF PARAMETERS IN THE MEAN FUNCTION
        INTEGER        NPARM
        PARAMETER      NPARM = 3
C TOTAL NUMBER OF VALUES TO BE GENERATED
        INTEGER        TIME
        PARAMETER      TIME = 7300
C NUMBER OF STATES OF RAIN
        INTEGER        NP
        PARAMETER      NP = 2
C PRESENT STATE
        INTEGER        PSTATE
C PREVIOUS STATE
        INTEGER        STATE
        INTEGER        K
        INTEGER        T
        INTEGER        A
        INTEGER        SEED (9)
NUMBER OF CONSECUTIVE WET SEQUENCE
        INTEGER        WSEQ
```

```
NUMBER OF CONSECUTIVE DRY SEQUENCE
        INTEGER         DSEQ
        INTEGER         COUNT
        REAL            RAIN
C PARAMETERS OF OCCURRENCE OF RAINFALL
        REAL            GAMMA (2, NPARM)
C FOURIER TERMS
        REAL            PHI (NPARM, NT)
        REAL            PI
C PARAMETERS FOR THE AUTOREGRESSIVE PROCESS
        REAL            RAU (NP, NV)
        REAL            DECOMP (NP, NV, NV)
        REAL            DECOM (NP, NV, NV)
        REAL            RES (2, NV)
        REAL            UNIFOR
        REAL            RESULT (NP, 1, NV)
C PARAMETERS FOR THE MEAN FUNCTION
        REAL            ALPHA (2, NPARM, NV)
        REAL            OBSN (NV)
C AMPLITUDES
        REAL            AMM (0 : NPARM)
C PHASES
        REAL            PHM (NPARM)
C COEFFICIENT OF VARIATION
        REAL            C

C       MAIN PROGRAM
C       ------------

  5     FORMAT ( )
 15     FORMAT (F7.3, 4 (F9.3), 2 (F 10.3), F9.3)

        CALL DATA (GAMMA, RAU, ALPHA, NPARM, NV, NP,
     &  AMM, PHM, C)
        CALL CHOLES (DECOMP)
C COMPUTE THE COVARIANCE MATRIX
        CALL COVAR (DECOMP, DECOM)
C COMPUTE THE FOURIER TERMS
        CALL COSSIN (PHI, NPARM, NT)
        DO 10, I = 1, NV
           DO 40, J = 1, 2
              RES (J, I) = 0
 40        CONTINUE
 10     CONTINUE
        SEED (1) = 115487
        SEED (2) = 257599
        SEED (3) = 6799501
        SEED (4) = 6237235
        SEED (5) = 7822589
        SEED (6) = 26875327
        SEED (7) = 69972223
        SEED (8) = 6831145
        SEED (9) = 33501959
        A = 8
C INITIAL STATE OF THE DAY
        STATE = 1
C NUMBER OF CONSECUTIVE WET VALUES
        WSEQ = 1
C NUMBER OF CONSECUTIVE DRY VALUES
        DSEQ = 2
        DO 20, T = 1, TIME
           K = T
```

```
          DO 30, I = 1, 20
             IF ( K .GT. 365) THEN
                K = K - 365
             ELSE
                GOTO 2
             ENDIF
 30       CONTINUE
C GENERATE RAINFALL VALUE
 2       CALL PIEST (NPARM, GAMMA, STATE, K, PHI, PI)
          UNIFOR = UNIF (SEED, A)
          IF (UNIFOR .LT. PI) THEN
             PSTATE = 2
             WSEQ = WSEQ + 1
             IF (PSTATE .NE. STATE) THEN
                COUNT = DSEQ
                DSEQ = 1
              ELSE
                 COUNT = 1
              ENDIF
           ELSE
              PSTATE = 1
              DSEQ = DSEQ + 1
              IF (PSTATE .NE. STATE) THEN
                 COUNT = WSEQ
                 WSEQ = 1
              ELSE
                 COUNT = 1
              ENDIF
           ENDIF
 C GENERATE NORMAL RANDOM NUMBER
           CALL RANDOM (DECOM, RESULT, PSTATE, SEED)
 C GENERATE CLIMATE VALUE
           CALL OMODEL (RESULT, STATE, NPARM, ALPHA, RAU,
     &                 RES, PHI, COUNT, K, OBSN, PSTATE)
           IF (PSTATE .EQ. 1) THEN
              RAIN = 0
           ELSE
              RAIN = 1
           ENDIF
 C IF IT RAINS THEN RAINFALL DEPTH IS GENERATED
           CALL DEPTH (SEED, NPARM, RAIN, K, AMM, PHM, C)
           WRITE (10, 15) RAIN, (OBSN (I), I = 1, NV)
           IF (PSTATE .NE. STATE) THEN
              STATE = PSTATE
           ENDIF
  20     CONTINUE
         STOP
         END
```

```
C PROGRAM 15
C ---------

C PROGRAM TO COMPUTE THE SETS OF POSSIBLE RAIN - NORAIN
C SEQUENCES

C TOTAL NUMBER OF OBSERVATIONS
        INTEGER            TIME
        PARAMETER          TIME = 2190
        INTEGER            SEQ (4, TIME)
        INTEGER            COUNT (4)
        INTEGER            PREV
C       INTEGER            RAIN
        REAL               RAIN
        REAL               CLIMA

  5     FORMAT ( )
 15     FORMAT (14 (I5))

        PREV = 0
        DO 20, J = 1, 4
           COUNT (J) = 0
 20     CONTINUE
        DO 10, I = 1, TIME
           READ 5, CLIMA
           IF (CLIMA .EQ. 0) THEN
              RAIN = 0
           ELSEIF (CLIMA .GT. 0) THEN
              RAIN = 1
           ELSEIF (CLIMA .EQ. -999) THEN
              RAIN = 2
           ENDIF
           IF ((RAIN .NE. 2) .AND. (PREV .NE. 2)) THEN
              IF (RAIN .EQ. PREV) THEN
                 IF (RAIN .EQ. 0) THEN
                    COUNT (1) = COUNT (1) + 1
                    SEQ (1, COUNT (1)) = I
                 ELSE
                    COUNT (2) = COUNT (2) + 1
                    SEQ (2, COUNT (2)) = I
                 ENDIF
              ELSE
                 IF (RAIN .EQ. 0) THEN
                    COUNT (4) = COUNT (4) + 1
                    SEQ (4, COUNT (4)) = I
                 ELSE
                    COUNT (3) = COUNT (3) + 1
                    SEQ (3, COUNT (3)) = I
                 ENDIF
              ENDIF
           ENDIF
           PREV = RAIN
 10     CONTINUE
        DO 30, J = 1, 4
           WRITE (10, 5) COUNT (J)
           WRITE (10, 15) (SEQ (J, I), I = 1, COUNT (J))
 30     CONTINUE
        STOP
        END
```

```
C PROGRAM  16
C ----------

C PROGRAM TO COMPUTE PARAMETER ESTIMATES FOR MODEL 3

C SUBROUTINE TO COMPUTE FOURIER TERMS
        EXTERNAL          COSSIN
C SUBROUTINE TO COMPUTE MEAN FUNCTION
        EXTERNAL          MUEST
C SUBROUTINE TO COMPUTE ESTIMATE FOR AUTOREGRESSIVE PROCESS
        EXTERNAL          RAUEST
C SUBROUTINE TO COMPUTE ESTIMATE FOR STANDARD DEVIATION
        EXTERNAL          SIGEST
C SUBROUTINE TO COMPUTE ESTIMATE FOR MEAN FUNCTION
        EXTERNAL          ALPEST
        INTEGER           NT
        PARAMETER         NT = 365
C NUMBER OF VARIABLES
        INTEGER           NV
        PARAMETER         NV = 7
C TOTAL NUMBER OF OBSERVATIONS
        INTEGER           TIME
        PARAMETER         TIME = 2190
C NUMBER OF PARAMETERS IN MEAN FUNCTION
        INTEGER           NPARM
        PARAMETER         NPARM = 3
        INTEGER           COUNT (4)
        INTEGER           SEQ (4, TIME)
        INTEGER           CONV
        REAL              PHI (NPARM, 0 : NT)
C ESTIMATES OF MEAN FUNCTION
        REAL              ALPHA (2, NV, NPARM)
        REAL              MU (2, NV, 0 : NT)
        REAL              CLIMA (2, NV, 0 : TIME)
C ESTIMATES OF AUTOREGRESSIVE PROCESS
        REAL              RAU (4, NV)
C ESTIMATES OF STANDARD DEVIATION
        REAL              SIGMA (4, NV)
```

```
C PRESENT ESTIMATES FOR MEAN FUNCTION
        REAL                  PALPHA (2, NV, NPARM)
C PRESENT ESTIMATES FOR AUTOREGRESSIVE PROCESS
        REAL                  PRAU (4, NV)
C PRESENT ESTIMATES FOR STANDARD DEVIATION
        REAL                  PSIGMA (4, NV)
C CONVERGENCE CRITERION
        REAL                  DELTA

 5      FORMAT ( )
15      FORMAT (14 (I5))
25      FORMAT (' CONVERGENCE ACHIEVED BY VARIABLE : ', I10)
35      FORMAT (' CONVERGENCE NOT MET IN 55 ITERATNS
    6           BY VAR : ', I10)
45      FORMAT (7 (F10.4))

        DO 10, M = 1, 2
           DO 20, L = 1, NPARM
              READ 5, (ALPHA (M, K, L), K = 1, NV)
20         CONTINUE
10      CONTINUE
        DO 30, J = 1, 4
           READ 5, COUNT (J)
           READ 15, (SEQ (J, I), I = 1, COUNT (J))
30      CONTINUE
        DO 40, M = 1, 2
           DO 50, T = 1, TIME
              READ 5, (CLIMA (M, K, T), K = 1, NV)
50         CONTINUE
40      CONTINUE
        DO 70, K = 1, NV
           DO 80, M = 1, 2
              DO 90, N = 1, 50
                 IF (CLIMA (M, K, N) .NE. -999) THEN
                    CLIMA (M, K, 0) = CLIMA (M, K, N)
                    GOTO 80
                 ENDIF
90            CONTINUE
80         CONTINUE
70      CONTINUE
C COMPUTE FOURIER TERMS
        CALL COSSIN (PHI, NPARM, NT)
        DO 220, L = 1, NPARM
           PHI (L, 0) = PHI (L, NT)
220     CONTINUE
        DO 100, K = 1, NV
           CONV = 0
           DO 110, N = 1, 55
C COMPUTE MEAN FUNCTION GIVEN ESTIMATES OF FOURIER SERIES
              CALL MUEST (MU, NPARM, NV, NT, ALPHA, PHI, K)
C ESTIMATE PARAMETER FOR AUTOREGRESSIVE PROCESS
              CALL RAUEST (RAU, CLIMA, MU, COUNT, SEQ,
    6                      NT, TIME, K)
C ESTIMATE STANDARD DEVIATION
              CALL SIGEST (SIGMA, RAU, CLIMA, MU, COUNT,
    &                      SEQ, NT, NV, TIME, K)
C ESTIMATE MEAN FUNCTION PARAMETERS
              CALL ALPEST (ALPHA, SIGMA, RAU, MU, PHI,
    &                      COUNT, SEQ, NT CLIMA, TIME, K)
C TEST FOR CONVERGENCE
```

```
            IF (N .NE. 1) THEN
               CONV = 1
               DO 120, J = 1, 4
                  DELTA = ABS ((PRAU (J, K) - RAU (J, K))
     &                        / RAU (J, K))
                  IF (DELTA .GT. 0.0001) THEN
                     CONV = 0
                  ENDIF
                  DELTA = ABS ((PSIGMA (J, K) - SIGMA (J, K))
     &                         / SIGMA (J, K))
                  IF (DELTA .GT. 0.0001) THEN
                     CONV = 0
                  ENDIF
120            CONTINUE
               DO 130, M = 1, 2
                  DO 140, L = 1, NPARM
                     DELTA  =ABS ((PALPHA (M, K, L) -
     &                             ALPHA (M, K, L))
     &                            / ALPHA (M, K, L))
                     IF (DELTA .GT. 0.001) THEN
                        CONV = 0
                     ENDIF
140               CONTINUE
130            CONTINUE
            ENDIF
            IF (CONV .EQ. 1) THEN
               PRINT 25, K
               GOTO 100
            ELSE
               DO 150, J = 1, 4
                  PRAU (J, K) = RAU (J, K)
                  PSIGMA (J, K) = SIGMA (J, K)
150            CONTINUE
               DO 160, M = 1, 2
                  DO 170, L = 1, NPARM
                     PALPHA (M, K, L) = ALPHA (M, K, L)
170               CONTINUE
160            CONTINUE
            ENDIF
110      CONTINUE
         PRINT 35, K
100   CONTINUE
      DO 230, K = 1, NV
      CALL MUEST (MU, NPARM, NV, NT, ALPHA, PHI, K)
230   CONTINUE
      CALL RESID (RAU, MU, COUNT, SEQ, CLIMA, NT)
      DO 180, M = 1, 2
         DO 190, L = 1, NPARM
            PRINT 5, (ALPHA (M, K, L), K = 1, NV)
190      CONTINUE
180   CONTINUE
      DO 200, J = 1, 4
         PRINT 5, (RAU (J, K), K = 1, NV)
200   CONTINUE
      DO 210, J = 1, 4
         PRINT 5, (SIGMA (J, K), K = 1, NV)
210   CONTINUE
```

```
C PROGRAM 17
C ---------

C PROGRAM TO ESTIMATE THE CORRELATION MATRIX

C TOTAL NUMBER OF OBSERVATIONS
        INTEGER                  TIME
        PARAMETER                TIME = 2190
C NUMBER OF VARIABLES
        INTEGER                  NV
        PARAMETER                NV = 7
        INTEGER                  COUNT (4)
        INTEGER                  SEQ (4, TIME)
        REAL                     TERM (5)
        REAL                     CORR (NV, NV)
        REAL                     RESI (NV, TIME)
        REAL                     VARI (NV)

  5     FORMAT ( )
 15     FORMAT (14 (I5))
 25     FORMAT (7 (F10.6))
 35     FORMAT (4 (F 11.4), /, 3 (F 15.4))

        DO 20, T = 1, TIME
           READ 5, (RESI (K, T), K = 1, NV)
 20     CONTINUE
        DO 30, J = 1, 4
           READ 5, COUNT (J)
           READ 15, (SEQ (J, I), I = 1, COUNT (J))
 30     CONTINUE
        DO 40, I = 1, NV
           CORR (I, I) = 1
 40     CONTINUE
        DO 50, N = 1, 4
           DO 60, K = 1, NV
              DO 70, J = K+1, NV
                 DO 120, II = 1, 5
                    TERM (II) = 0
120              CONTINUE
                 DO 80, T = 1, COUNT (N)
                    I = SEQ (N, T)
                    TERM (1) = RESI (K, I) * RESI (J, I)
     &                         + TERM (1)
                    TERM (2) = TERM (2) + RESI (K, I)
                    TERM (3) = TERM (3) + RESI (J, I)
                    TERM (4) = TERM (4) + RESI (K, I) ** 2
                    TERM (5) = TERM (5) + RESI (J, I) ** 2
 80              CONTINUE
                 TERM (1) = TERM (1) / COUNT (N)
                 TERM (4) = SQRT ((TERM (4) / COUNT (N)) -
     &                      (TERM (2) / COUNT (N)) ** 2)
                 VARI (K) = TERM (4) ** 2
                 TERM (5) = SQRT ((TERM (5) / COUNT (N)) -
     &                      TERM (3) / COUNT (N)) ** 2)
                 TERM (2) = TERM (2) * TERM (3) /
     &                      COUNT (N) ** 2
```

```
                    TERM (1) = TERM (1) - TERM (2)
                    TERM (4) = TERM (4) * TERM (5)
                    CORR (K, J) = TERM (1) / TERM (4)
 70              CONTINUE
 60           CONTINUE
              TERM (2) = 0
              TERM (4) = 0
              DO 10, T = 1, COUNT (N)
                 I = SEQ (N, T)
                 TERM (2) = TERM (2) + RESI (7, I)
                 TERM (4) = TERM (4) + RESI (7, I) ** 2
 10           CONTINUE
              VARI (7) = (TERM (4) / COUNT (N)) - (TERM (2) /
     &                          COUNT (N)) ** 2
              PRINT 35, (VARI (K), K = 1, NV)
              DO 110, K = 1, NV
                 PRINT 25, (CORR (K, J), J = K, NV)
110           CONTINUE
 50        CONTINUE
           STOP
           END
```

```
C PROGRAM 18
C ----------

C PROGRAM TO GENERATE CLIMATE SEQUENCES WHEN USING MODEL 3.

        INTEGER          NT
        PARAMETER        NT = 365
C NUMBER OF VARIABLES
        INTEGER          NVAR
        PARAMETER        NVAR = 7
C NUMBER OF PARAMETERS IN THE MODEL FOR THE MEAN FUNCTION
        INTEGER          NPARM
        PARAMETER        NPARM = 3
C TOTAL NUMBER OF VALUES TO BE GENERATED
        INTEGER          TIME
        PARAMETER        TIME = 7300
C NUMBER OF RAIN - NO-RAIN SEQUENCES
        INTEGER          NP
        PARAMETER        NP = 4
C PRESENT STATE OF THE DAY
        INTEGER          PSTATE
C PREVIUOUS STATE OF THE DAY
        INTEGER          STATE
        INTEGER          K
        INTEGER          T
        INTEGER          A
        INTEGER          SEED (9)
        INTEGER          SEQ
        REAL             RAIN
C PARAMETER ESTIMATES FOR THE OCCURRENCE OF RAINFALL
        REAL             GAMMA (2, NPARM)
```

```
C FOURIER TERMS
        REAL             PHI (NPARM, NT)
        REAL             PI
C PARAMETER  ESTIMATES FOR THE AUTOREGRESSIVE PROCESS
        REAL             RAU (NP, NVAR)
        REAL             DECOMP (NP, NVAR, NVAR)
        REAL             DECOM (NP, NVAR, NVAR)
        REAL             RES (2, NVAR)
        REAL             UNIFOR
        REAL             RESULT (NP, 1, NVAR)
C PARAMETER ESTIMATES FOR THE MEAN FUNCTION
        REAL             ALPHA (2, NPARM, NVAR)
        REAL             OBSN (NVAR)
C AMPLITUDES FOR THE MODEL OF RAINFALL DEPTH
        REAL             AMM (0 : NPARM)
C PHASES FOR THE MODEL OF RAINFALL DEPTH
        REAL             PHM (NPARM)
C THE COEFFICIENT OF VARIATION
        REAL             C

C       MAIN PROGRAM
C       ------------

  5     FORMAT ( )
 15     FORMAT (F7.3, 4 (F9.3), 2 (F10.3), F9.3)

        CALL DATA (GAMMA, RAU, ALPHA, NPARM, NVAR, NP,
     &  AMM, PHM, C)
C COMPUTE THE CHOLESKY DECOMPOSITION OF THE CORRELATION
C MATRIX
        CALL CHOLES (DECOMP)
C COMPUTE THE COVARIANCE MATRIX
        CALL COVAR (DECOMP, DECOM)
C COMPUTE THE FOURIER SERIES TERMS
        CALL COSSIN (PHI, NPARM, NT)
        DO 10, I = 1, NVAR
           DO 40, J = 1, 2
              RES (J, I) = 0
 40        CONTINUE
 10     CONTINUE
        SEED (1) = 115487
        SEED (2) = 257599
        SEED (3) = 6799501
        SEED (4) = 6237235
        SEED (5) = 7822589
        SEED (6) = 26875327
        SEED (7) = 69972223
        SEED (8) = 6831145
        SEED (9) = 33501959
        A = 8
C INITIAL STATE OF THE DAY = DRY
        STATE = 1
        DO 20, T = 1, TIME
           K = T
           DO 30, I = 1, 100
              IF ( K .GT. 365) THEN
                 K = K - 365
```

```
                 ELSE
                    GOTO 2
                 ENDIF
 30           CONTINUE
C GENERATE RAINFALL VALUE
C
C COMPUTE THE PROBABILITY FOR A WET DAY FOLLOWING A WET
C DAY OR THE PROBABILITY FOR A WET DAY FOLLOWING A DRY
C DAY
 2           CALL PIEST (NPARM, GAMMA, STATE, K, PHI, PI)
C GENERATE A UNIFORM RANDOM NUMBER BETWEEN 0 AND 1
              UNIFOR = UNIF (SEED, A)
              IF (UNIFOR .LT. PI) THEN
                 PSTATE = 2
                 IF (PSTATE .NE. STATE) THEN
C THAT IS : HAVE THE SEQUENCE DRY-WET
                    SEQ = 3
                 ELSE
C THAT IS HAVE THE SEQUENCE WET-WET
                    SEQ = 2
                 ENDIF
              ELSE
                 PSTATE = 1
                 IF (PSTATE .NE. STATE) THEN
C THAT IS HAVE THE SEQUENCE WET-DRY
                     SEQ = 4
                  ELSE
 C THAT IS HAVE THE SEQUENCE DRY-DRY
                     SEQ = 1
                  ENDIF
               ENDIF
 C GENERATE A NORMAL RANDOM NUMBER
               CALL RANDOM (DECOM, RESULT, SEQ, SEED)
 C GENERATE CLIMATE VALUES
               CALL MODEL (RESULT, STATE, NPARM, ALPHA, RAU,
     &                            RES, PHI, SEQ, K, OBSN, PSTATE)
               IF (PSTATE .EQ. 1) THEN
                  RAIN = 0
               ELSE
                  RAIN = 1
               ENDIF
 C IF IT RAINS THEN RAINFALL DEPTH VALUE IS GENERATED
               CALL DEPTH (SEED, NPARM, RAIN, K, AMM, PHM, C)
               WRITE (10, 15) RAIN, (OBSN (I), I = 1, NVAR)
               IF (PSTATE .NE. STATE) THEN
                  STATE = PSTATE
               ENDIF
  20       CONTINUE
           STOP
           END
```

```
C PROGRAM 19
C ----------

C PROGRAM TO ESTIMATE PARAMETERS FOR MODEL 3 USING
C ORDINARY LEAST SQUARES ESTIMATES

C SUBROUTINE TO COMPUTE FOURIER SERIES TERMS
        EXTERNAL            COSSIN
        EXTERNAL            MUEST
C SUBROUTINE TO COMPUTE MEAN FUNCTION GIVEN THE FOURIER
C COEFFICIENT ESTIMATES
        EXTERNAL            RAUEST
C SUBROUTINE TO ESTIMATE THE MEAN FUNCTION PARAMETERS
        EXTERNAL            ALPH
        INTEGER             NT
        PARAMETER           NT = 365
C NUMBER OF VARIABLES
        INTEGER             NV
        PARAMETER           NV = 7
C NUMBER OF OBSERVATIONS
        INTEGER             TIME
        PARAMETER           TIME = 2190
C NUMBER OF PARAMETERS IN THE MEAN FUNCTION
        INTEGER             NPARM
        PARAMETER           NPARM = 3
        INTEGER             COUNT (4)
        INTEGER             SEQ (4, TIME)
C CONVERGENCE CRITERION
        INTEGER             CONV
C FOURIER TERMS
        REAL                PHI (NPARM, 0 : NT)
C PARAMETER ESTIMATES FOR THE MEAN FUNCTION
        REAL                ALPHA (2, NV, NPARM)
        REAL                MU (2, NV, 0 : NT)
        REAL                CLIMA (2, NV, 0 : TIME)
C PARAMETER ESTIMATES FOR THE AUTOREGRESSIVE PROCESS
        REAL                RAU (4, NV)
C VECTOR OF PRESENT ESTIMATES FOR THE MEAN FUNCTION
        REAL                PALPHA (2, NV, NPARM)
C VECTOR OF PRESENT ESTIMATES FOR THE AUTOREGRESSIVE PROCESS
        REAL                PRAU (4, NV)
        REAL                DELTA

  5     FORMAT ( )
 15     FORMAT (14 (I5))
 25     FORMAT (' CONVERGENCE ACHIEVED BY VARIABLE : ', I10)
 35     FORMAT (' CONVERGENCE NOT MET IN 55 ITERATIONS
    &             BY VAR : ', I10)
 45     FORMAT (7 (F10.4))

        DO 10, M = 1, 2
           DO 20, L = 1, NPARM
              READ 5, (ALPHA (M, K, L), K = 1, NV)
 20        CONTINUE
 10     CONTINUE
        DO 30, J = 1, 4
           READ 5, COUNT (J)
           READ 5, (SEQ (J, I), I = 1, COUNT (J))
 30     CONTINUE
        DO 40, M = 1, 2
           DO 50, T = 1, TIME
              READ 5, (CLIMA (M, K, T), K = 1, NV)
```

```
 50         CONTINUE
 40      CONTINUE
C DETERMINING THE VALUE OF THE CLIMATE VARIABLES AT TIME
C ZERO
         DO 70, K = 1, NV
            DO 90, M = 1, 2
               DO 80, N = 1, 50
                  IF (CLIMA (M, K, N) .NE. -999) THEN
                     CLIMA (M, K, 0) = CLIMA (M, K, N)
                     GOTO 90
                  ENDIF
 80            CONTINUE
 90         CONTINUE
 70      CONTINUE
C COMPUTE FOURIER TERMS
         CALL COSSIN (PHI, NPARM, NT)
         DO 220, L = 1, NPARM
C SET THE FOURIER TERM AT TIME ZERO TO BE THAT AT TIME
C 365
            PHI (L, 0) = PHI (L, NT)
220      CONTINUE
         DO 100, K = 1, NV
            CONV = 0
            DO 110, N = 1, 55
C COMPUTE THE MEAN FUNCTION
               CALL MUEST (MU, NPARM, NV, NT, ALPHA, PHI, K)
C ESTIMATE THE PARAMETER OF THE AUTOREGRESSIVE PROCESS
               CALL RAUEST (RAU, CLIMA, MU, COUNT, SEQ,
     &                          NT, TIME, K)
C ESTIMATE THE PARAMETERS OF THE MEAN FUNCTION
               CALL ALPH (ALPHA, RAU, MU, PHI, COUNT, SEQ, NT,
     &                         CLIMA, TIME, K)
               IF (N .NE. 1) THEN
                  CONV = 1
                  DO 120, J = 1, 4
C TEST FOR CONVERGENCE
                     DELTA = ABS ((PRAU (J, K)-RAU (J, K))
     &                            / RAU (J, K))
C                    PRINT *, 'DELTA', DELTA
                     IF (DELTA .GT. 0.0001) THEN
                        CONV = 0
                     ENDIF
120               CONTINUE
                  DO 130, M = 1, 2
                     DO 140, L = 1, NPARM
                        DELTA  =ABS ((PALPHA (M, K, L)
     &                              - ALPHA (M, K, L))
     &                             / ALPHA (M, K, L))
C                         PRINT *, 'DELTA', DELTA
                          IF (DELTA .GT. 0.0001) THEN
                             CONV = 0
                          ENDIF
 140                    CONTINUE
 130                 CONTINUE
                  ENDIF
```

```
                IF (CONV .EQ. 1) THEN
                   PRINT 25, K
                   GOTO 100
                ELSE
C UPDATING THE PRESENT PARAMETER ESTIMATES
                   DO 150, J = 1, 4
                      PRAU (J, K) = RAU (J, K)
150                CONTINUE
                   DO 160, M = 1, 2
                      DO 170, L = 1, NPARM
                         PALPHA (M, K, L) = ALPHA (M, K, L)
170                   CONTINUE
160                CONTINUE
                ENDIF
110          CONTINUE
             PRINT 35, K
             CALL MUEST (MU, NPARM, NV, NT, ALPHA, PHI, K)
100       CONTINUE
C COMPUTING THE RESIDUAL TIME SERIES
          CALL RESID (RAU, MU, COUNT, SEQ, CLIMA, NT)
          DO 180, M = 1, 2
             DO 190, L = 1, NPARM
                PRINT 5, (ALPHA (M, K, L), K = 1, NV)
190          CONTINUE
180       CONTINUE
          DO 200, J = 1, 4
             PRINT 5, (RAU (J, K), K = 1, NV)
200       CONTINUE
          DO 210, M = 1, 2
             DO 230, I = 1, NT
                PRINT 45, (MU (M, K, I), K = 1, NV)
230          CONTINUE
210       CONTINUE
          STOP
          END
```

18 MAY 1987