TECHNICAL MEMORANDUM

May 1984

DRAFT

A TIME SERIES ANALYSIS
OF
SOUTH FLORIDA RAINFALL RECORDS

by

George Shih

DRE 182

RESOURCE PLANNING DEPARTMENT

SOUTH FLORIDA WATER MANAGEMENT DISTRICT

WEST PALM BEACH, FLORIDA

ACKNOWLEDGEMENTS

This report was prepared under Program 8020. Program Manager Mr. Robert L. Hamrick provided guidance and many valuable suggestions. Mr. Hamrick also reviewed the first draft and clarified many confusing points. Gratitude is extended to the members of the Resource Planning Staff for their assistance in the editorial process.

This publication was produced at an annual cost of $131.25 or $.26 per copy to inform the public. 500 191 Produced on recycled paper.

TABLE OF CONTENTS

A TIME SERIES ANALYSIS OF SOUTH FLORIDA RAINFALL RECORDS

SUMMARY

Other findings are:

This study indicates there is a high probability that, during the years of 1983-1984, south Florida will have above normal rainfall, and then during 1986-1987, it will have below normal rainfall. Other findings are:

(1) There may be 6 and 12 year cycles in south Florida rainfall. These basic cycles are fairly synchronized in south Florida.

(2) Attempts to use regression and ARIMA models for yearly rainfall forecasting were not successful.

(3) Extremes (peak or valley) in certain cycles show there are some predictabilities in both time and quantity. The forecast is based on the cyclic behavior of the extremes.

(4) For water supply planning, a 6-year drought-rainfall is recommended. The average minimum of the 6-year cycle is 49.68 inches a year (93.7% of over all average), about the same as the average minimum of a 12-year cycle.

INTRODUCTION

When a time series is not stationary in some way, patterns may be hidden in the time series. If a significant trend or cycle can be identified, it can help explain the variance involved in the time series. Reduction of variance is particularly important if the model is to be used as a planning or forecasting tool. Furthermore, the trend or cycle mode (and its associated parameters) often imply physical meanings that can help to identify the cause/effect relationships. One may even attempt to relate a shift in trend and/or cyclic parameters identified from historical data to certain physical changes in the system. Unless a satisfactory physical model is available to test the system response to these changes, time series analysis may be the next best thing to cast some light on the complicated climatological system of south Florida.

PURPOSE

In a previous study, rainfall data in south Florida was assumed to be independent and identically distributed. Tests of change in the mean showed that many rainfall stations did have a significant change during the period of record available. In this second phase of the study, the rainfall time series is not considered independent; rather , their trends or cyclic phenomena will be examined. It is hoped to :

1. Identify any significant patterns in the time series,
2. Quantify the reliability of the significant trends or cycles when identified,
3. Examine the geographic distribution of the trends or cycles among the stations, and
4. Apply these findings in a forecast mode.

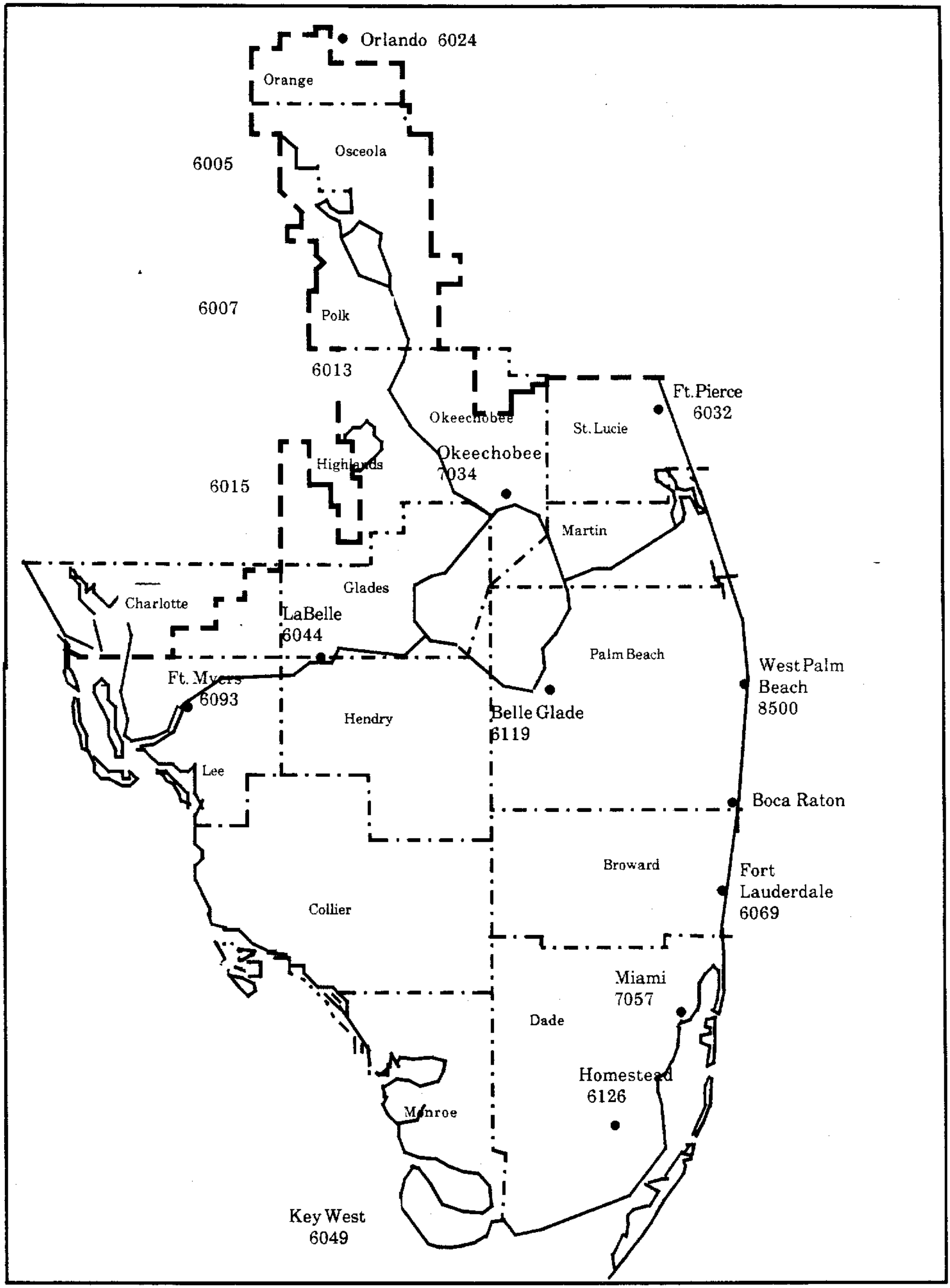
DATA SOURCE AND TREATMENT

There are over 600 monthly rainfall station records available in the Data Base of the District. Most of them have very short record length, started only recently, and are located in the population centers along the coast. Only 15 stations are selected for this analysis. The selection rationale is:

1. Early starting date,
2. Long record length, and
3. Even spatial distribution.

A map of the selected rainfall stations is shown as Figure 1.

Yearly data are used in this analysis. Missing data are filled in by correlation with nearby stations not in the selected 15 stations. The filling is done only to


Orlando 6024
Orange
Osceola
6005
6007
Polk
6013
Ft. Pierce
6032
Okeechobee
St. Lucie
Highlands
Okeechobee
7034
6015
Martin
Glades
Charlotte
LaBelle
6044
Palm Beach
West Palm
Beach
8500
Ft. Myers
6093
Hendry
Belle Glade
6119
Lee
Boca Raton
Broward
Fort
Lauderdale
6069
Collier
Miami
7057
Dade
Homestead
6126
Monroe
Key West
6049

extend the data up to 1982 for those stations that were discontinued, and for the random missing data within the period of record. In other words, the records are not extended to periods prior to the starting date of a station.

The rainfall station at Key West, MRF 6049, is of particular interest. It has the longest length of record (150 years). It is far away from the land mass and is presumed to have minimal impact by human activities. It also seems to be in a different climatological regime from the rest of the stations. This station has a mean of 38.42 inches per year, and a 9.24 inch standard deviation of the rainfall series.

A synthetic yearly time series intended to represent the average District rainfall is derived from the data of the selected stations. Basically, it is an arithmetic means of all the "over 50 years" stations. This synthetic time series is used as "average" District rainfall time series. The mean of this average series is 53.01 inches, and standard deviation is 7.20 inches. This series is denoted by ALL*.

ANALYSIS STRATEGY

1. Select a generic model, such as Regression, Stochastic, or others as appropriate.
2. Because seasonal (yearly) cycles are well established, use yearly totals to detect multi-year patterns.
3. Examine the variance reduction by the specific model to compare with others to define the relative model efficiency.
4. When a promising model is found, apply the model in a forecast mode.

MODEL TESTED

The basic comparison model is the mean, e.g., one can always use the monthly mean or yearly mean for filling the missing data or for future planning.

Any further sophistication of the model will have to be tested against the sample mean, variance, and degree of freedom involved to see the efficiency of the selected model.

1. Linear Regression Model. Linear regression equations of different types, with yearly rainfall as dependent variable and time, or various transformations thereof as independent variables, are tested in a stepwise regression analysis. As expected, the regression model is not an efficient model for rainfall. No regression equation is found to be superior to the mean. The standard error of estimation by regression equation is no better than the standard deviation of the sample. For example, the "best" regression equation tested for averaged yearly rainfall series for the District is in the form of

$$\text{Rainfall (inch)} = a_1 + a_2 t^2 + a_3 t^{-2}$$

where t denotes time in year.

The squared correlation (R2) is 0.044 The standard error of estimation is 7.13 inches, while the standard deviation of the original series is 7.20 inches. The small improvement cannot justify the use of the regression model over the sample mean. It is concluded that the regression model alone is not useful.

2. Stochastic Models. Hidden cycles in a time series can be detected in the frequency domain by spectrum analysis and Fourier series analysis, etc. It is a foregone conclusion that yearly rainfall cycles exist in south Florida due to the relative positions between the sun and the earth. Hence, multi-year cycles are the target of the investigation. When yearly rainfall data were used in these investigations, two prominent cycles, around 6 and 11 years, showed up. There is a gap with no rainfall series showing the probability of an 8-year cycle. Some station records are not long enough to demonstrate a periodic phenomenon of more than 11 years. Still the analysis shows that there may be a cyclic period around 29 years.

Being unable to sort out the cause-effect interrelationship from local peculiarities, it is commonly thought that the multi-year cyclic phenomena are results of global and/or solar system factors. If these factors are controlling the cycles, then within south Florida, these cycles should show some kind of commonality in terms of length and synchronization. Variation of the cycle lengths as discussed above indicates that either the phenomena are not results of a common cause or else the local effects have modified and blurred the common cause. Furthermore, not knowing the physical forces behind these cyclic phenomena, one should also be cautious about getting spurious cycles; analogous to spurious correlation in regression analysis. Because, given any (time) series not purposely randomly generated, one can always find some kind of periodicities in the series. Hence the "cycles" identified visually should go through some kind of statistical significance test. For a particular parameter to be statistically significant for explanation of the variance of the data points, this parameter should at least explain $Fc \times 1/N \times 100$ percent of the variance, where Fc is the critical F test value at a given confidence and N is the number of data points. At 90% confidence level with data points over 120, Fc = 2.75, Table 1 shows the significant cycles for all the stations tested. Note that for each cycle, there are two parameters involved in explaining the variance. Also it can be shown that the variance explained by a given cycle is one half of the squared amplitude associated with the cycle. Viewing the small percentage of variance explainable by each cycle in Table 1 indicates that the amplitudes associated with these signficant cycles are still very small. Some representative plots of variance spectrum are included in Appendix A.

One can also investigate the time series in the time domain, in addition to the frequency domain discussed above. Autocorrelation of lag 1 through N years were examined. It basically displayed the peak correlation around six and ten years. Because of negative correlation, the plots of the correlogram can be confusing.

TABLE 1. SIGNIFICANT CYCLES						
Stations	Cycles in Months		Percent Variance Explained			
ALL*	6/2.11	12/55.19	70/0.59			
MRF 6049	6/1.56	12/25.12	36/0.33	42/0.35	118/0.38	156/0.40
MRF 6126	12/42.47					
MRF 7057	6/1.75	12/25.64	42/0.82	72/0.74	112/0.71	352/1.20
MRF 6069	6/1.71	12/25.87	38/0.86			
MRF 6093	6/5.40	12/46.14				
MRF 6044	6/4.18	12/46.82				
MRF 6119	6/1.82	12/42.85				
MRF 8500	12/33.10	74/1.44				
MRF 6015	6/5.33	12/41.69	66/0.92			
MRF 7034	6/1.88	12/36.64	184/1.19			
MRF 6032	6/2.34	12/25.54	42/0.60	68/0.92		
MRF 6013	6/5.25	12/44.00	64/0.63			
MRF 6007	6/6.57	12/38.30	72/0.54			
MRF 6005	6/10.45	12/37.68				
MRF 6024	6/5.91	12/35.72	68/0.77			

However, these plots are very useful in ARIMA model identifications. Some correlegrams are attached in Appendix B.

Plots of moving average can also help in identifying hidden cycles. Moving average has the effect of smoothing out the short-term irregularities when the right numbers of periods are used in computing the moving average. Peaks and valleys can be amplified in the time scale. When the moving averages are plotted in relation to calendar time, as shown in Appendix C, one can also see whether the particular cycles are synchronized among stations. In general, it is observed that wet and dry runs are fairly well synchronized in the area with some locational variation to shift the peaks (or valleys) among the stations. Examination of these graphs shows that 6 and 12 year cycles are relatively well synchronized and have larger amplitude.

From the above mentioned spectrum, correlogram, moving average, and synchronization analysis, it is fairly consistent that there is a 6 year cycle; but the next fundamental cycle varies from 10 to 12 years and reduces to marginal significance. This is also observed when significant cycles of each station are plotted in a map. The Kissimmee River Valley shows a 6 year cycle, the lower east coastal area shows a mix of 3 and 10 year cycles, while the area southwest of the District, lacking good data, does not show any particular cycle.

Judging from the results of spectrum analysis and correlogram plots, it can be concluded that ARIMA (Dixon, 1981) type models may not be very useful. Nevertheless, a Box-Jenkins time series analysis using annual data was attempted. The results of the "best" type of model and the model forecast estimation error are shown in Appendix D. It can be seen, for the models tested,that the forecast values are not a better estimate than the means. This indicates that quantitative forecast models of this type may not be worth the effort. Furthermore, there is no geographical pattern shown in the results of this model analysis.

3. Cycles of Wet and Dry Years. Another way to look at a cyclic phenomenon is simply to find the mean values of a given cycle. For example, to find the mean values of a 6 year cycle, values at years 1, 7, 13, 19, . . . are added up to find the mean for the first position; and years 2, 8, 14, 20, . . . for the second position, similar to finding the monthly mean in a monthly time series. Figure 2 shows some of these cyclic means from the average series. It can be seen that 6, 10, 11, and 12 year cycles have larger peak-to- valley amplitudes. In order to derive more information from these cyclic models, some statistics were plotted in Figures 3a through 3c for the average series. Figure 3a is a plot of variation coefficient of the cycle mean and shows that some cyclic models may be able to reduce the prediction variance compared to the mean, but the reduction is not significant. In other words, cyclic models, just like ARIMA models, are not a more efficient model for year after year prediction than the sample mean. Next, it is attempted to see if these models will be able to shed any light on the extreme (wet or dry) years. Figure 3b is a plot of the variation coefficient of each annual mean for 6 and 12 year cycles of the same series. Observing the low variation coefficient at the extreme values, by referring to Figure 2, it is clear that any variance explained by the model is realized mostly by reducing the variance at the extremes. To take data size and variance into consideration, the standard deviation scores (Z scores) from the series mean for the extremes of a given cycle are plotted, as shown in Figure 3c. It is found that the extreme Z scores are significant only at certain cycle years. It is reasoned that when a cycle exists, the peak (or valley) will tend to fall into the same cyclic position when the cyclic model is correct. Hence, it is concluded that this type of cyclic model may have utility in identifying the cyclic length, quantifying cyclic extremes, and predicting return of extremes.

Because of the short record length, each cyclic position in a cycle model may have only a few data points (typically less than 10). Student "t" criteria at 90%

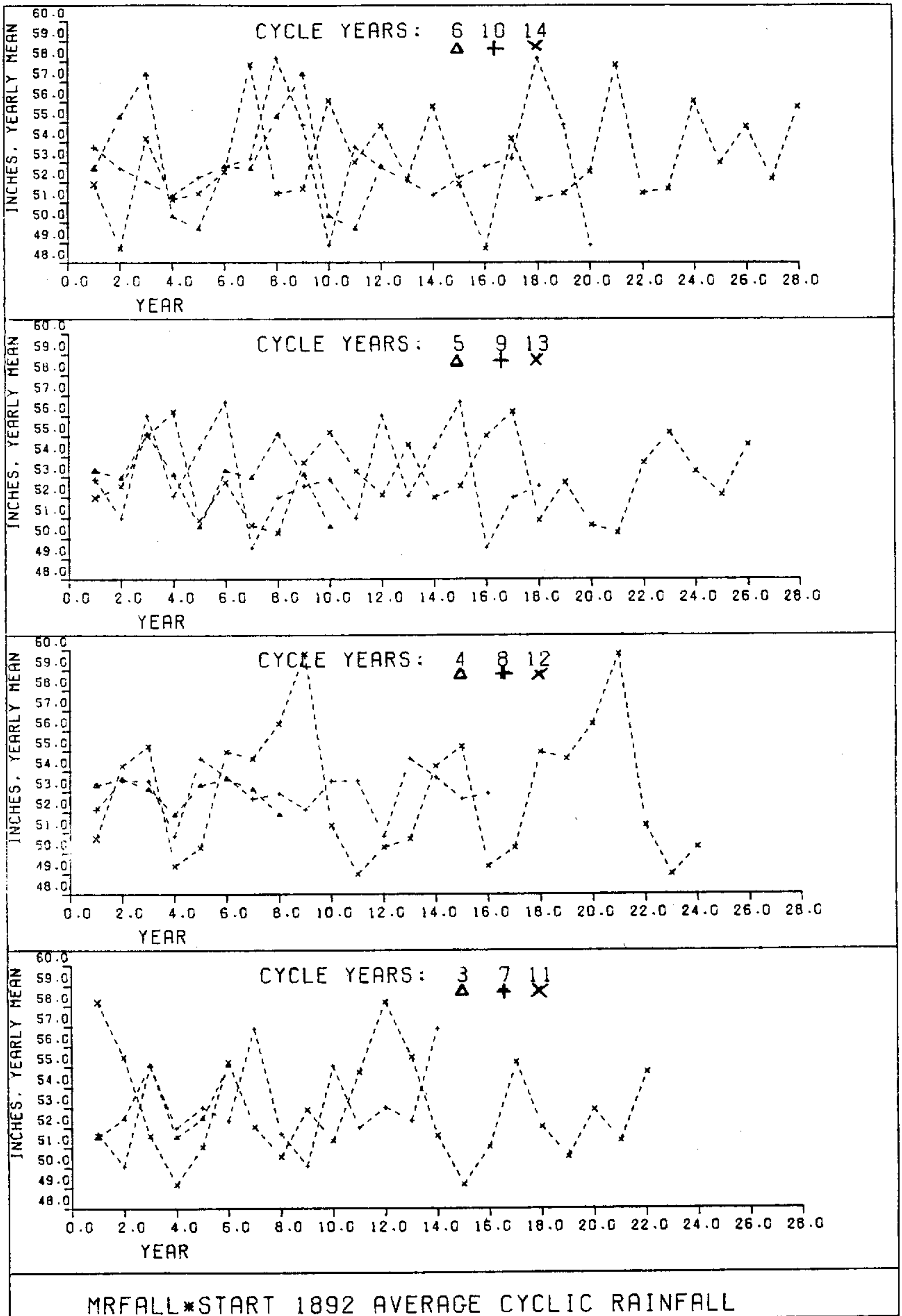


Figure ~~1~~ 2 Cyclic Rainfall For Series All*

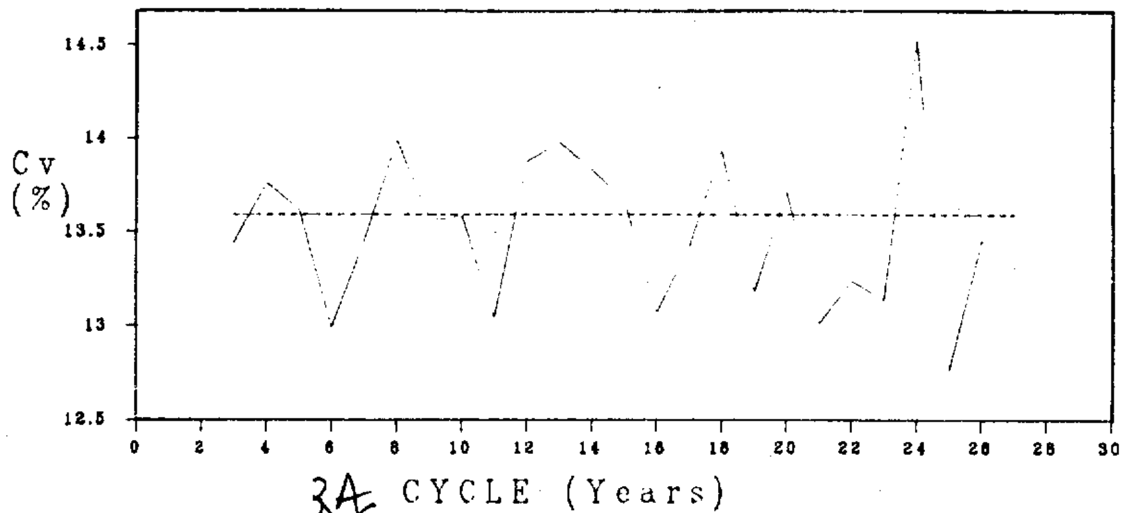


Fig. ~~2a~~ 3A Cv OF CYCLE MEAN

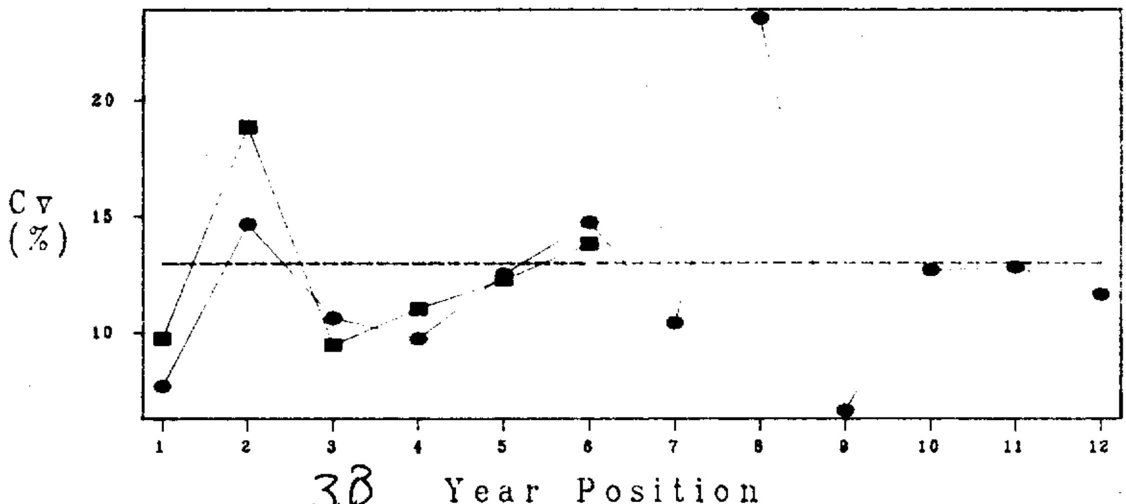


Fig. ~~2b~~ 3B Cv OF 6 & 12 YEAR CYCLES

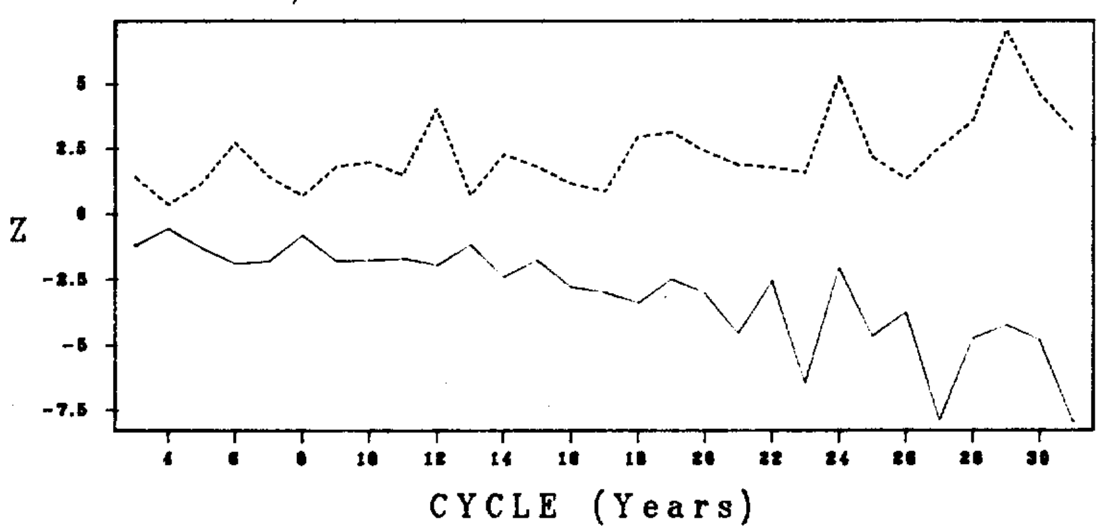


Fig. ~~2c~~ Z SCORES OF EXTREMES
3C

significance level is used for discrimination. To use these models in identifying significant cycles, some obvious facts are observed:

1) Extreme values are monotonically increasing (for wet side) or decreasing (for dry side) as cycle lengths increase. If the values are not monotonic, either the cycle length is not right or data variability is so high that the cyclic characteristic is unidentifiable. Since "t" can take data variance and degrees of freedom into consideration, it is used as the discriminator. The first significant cycle is tested against the mean at 90% confidence level. After a cycle is found, the same procedure can then be used to identify the next significant cycle with the monotonic requirement.

2) The model does not assume that wet cycles and dry cycles should necessarily be the same length. Hence, wet and dry cycles may be identified separately. From Figure 3c it can be seen that 6, 12, and 24 year cycles are significant on the wet side and 6 and 18 year cycles are significant on the dry side.

It is interesting to plot the mean extreme value of N year cycle at 1/N percentile on a probability paper as shown in Appendix E. The validity of this procedure as a method of distribution-free frequency analysis may deserve further investigation.

The next step is to attempt to use the model in a real time forecast of the extreme wet or dry years in a cycle. This is done by looking into the historical data and estimating the probabilities that extreme conditions occurred in the same cycle position. For example, in the average series for a 12 year cycle in the wet side, the wettest position is at the 9th position. When the historical average rainfall is divided into consecutive groups of 12 years, there are 7 groups. The wettest years occurred at the 8th, 9th, 11th, 3rd, 8th, 8th, and 6th positions of 1 through 7 cycles, respectively. It can be seen that 9th and 8th positions have more than the random chance of being the wettest year in the cycle. The 9th position had 2/7 = 28.57% to be a wettest year in comparison to the random chance of 1/12 = 8.33%. In other words, 9th position had about three times greater than random opportunity to be the wettest year in a cycle. Assuming an exponential distribution, the probability of the forecasted year to be the extreme of the cycle is computed and shown in Table 2 for the average rainfall series. Formulae used in this table are attached in Appendix F. The results of individual station analyses are shown in Appendix G.

An examination of Table 2 shows that 1984 appears to be the peak for 3, 6, 12, and 24 year cycles. Since different cycles show different expected rainfall, a legitimate question is: "What is the expected rainfall for 1984?" The model is not able to answer this question; however, using the expected rainfall from the longest cycle (24 years in this case) usually provides the most significant estimate (62.81 inches). Of course, depending on application, other criteria such as the expected rainfall corresponding to the least quotient of standard deviation, divided by percent probability, can be used.

One aspect of the forecast method, using multiple and separate wet and dry cycles, is the possible coincidence of both wet and dry extreme conditions in the same year. It indicates the destructive interference of these cycles. The rainfall of that particular year will have no better prediction than the mean.

Six and 12 year cycles deserve special attention because of the large amplitudes and time synchronization of these cycles. The 6 year cycle is relatively symmetrical with 3 years wet and 3 years dry, which is reflected in the historical "run." The 12 year cycle is not simply two repetitions of the6 year cycle, it has two valleys of the same magnitude and one dominating peak. The amplitude of both the valleys in the 12 year cycle is about the same as that of the 6 year cycle, while the amplitude of the 12-year peak increased 2.4 inches over that of the 6 year cycle.

TABLE 2. WET AND DRY YEAR FORECAST

CYCLE (YEAR)	WET (PEAK OF CYCLE) YEAR Position	TIMING YEAR	TIMING Prob. %	AMOUNT (Inches) Expected RF	AMOUNT (Inches) St. Dev.	DRY (VALLEY OF CYCLE) YEAR Position	TIMING YEAR	TIMING Prob. %	AMOUNT (Inches) Expected RF	AMOUNT (Inches) St. Dev.
STATION: MRF ALL START: 1892 MEAN: 53.01 ST. DEV: 7.20										
3	3	1984	45	55.08	6.74					
		1983/84	72							
6	3	1984	32	57.37	5.43	5	1986	30	49 68	6.11
		1983/84	61				1985/86	55		
							1985/87	79		
12	9	1984	28	59.80	3.96	11	1986	27	49.39	4.82
		1983/84	54				1985/86	54		
							1985/87	66		
18						11	1992	21	45.51	4.67
24	21	1984	33	62.81	2.94					

Hence, for water supply, a 6 year drought rainfall may be good for a 10 year project lifetime. For water storage system planning, a 12 year rainfall cycle may be an appropriate base for storage capacity and residence time computations.

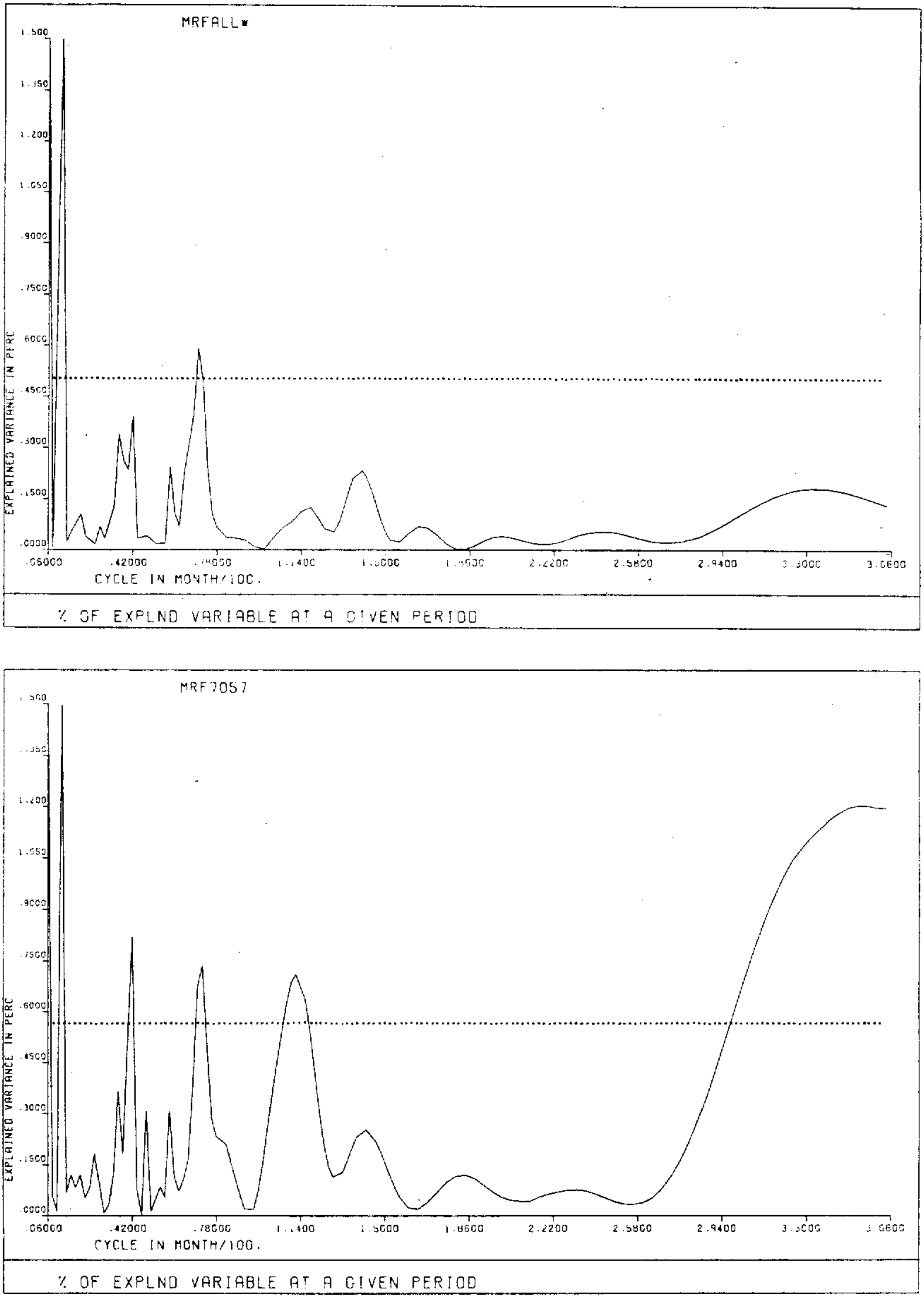
In a previous study (Shih, 1983) detecting the rainfall change in the same area, using data ending in 1981, it was found that yearly rainfall from 1970 to 1981 was about 5 inches less than the yearly rainfall prior to 1970 in the available records. Decreased tropical cyclone activity in the 1970-1981 period accounted for 2.68 inches of rainfall reduction. Positioning the 12 year cycle in real time can account for another one inch of rainfall reduction which is the maximum amount explainable from the cyclic study. This leaves 1.32 inches, or 26.4% of the rainfall reduction, to be accounted for by other factors.

The picture is more confusing when looking at stations individually. There are 10 stations out of 15 that show 6 year cycles, and only 4 stations that show 12 year cycles, even though the majority of the stations have 10 to 13 year cycles. There is no particular pattern of geographic distribtuion of these cycles. However, there is a strong indication from the individual stations that in recent years, peak of cycles will occur in 1983-84 and valley of cycles will occur in 1985-1986. While the study failed to suggest a model for year-to-year rainfall forecasting, the cycle-extreme does provide a better guess at upcoming wet and dry years.

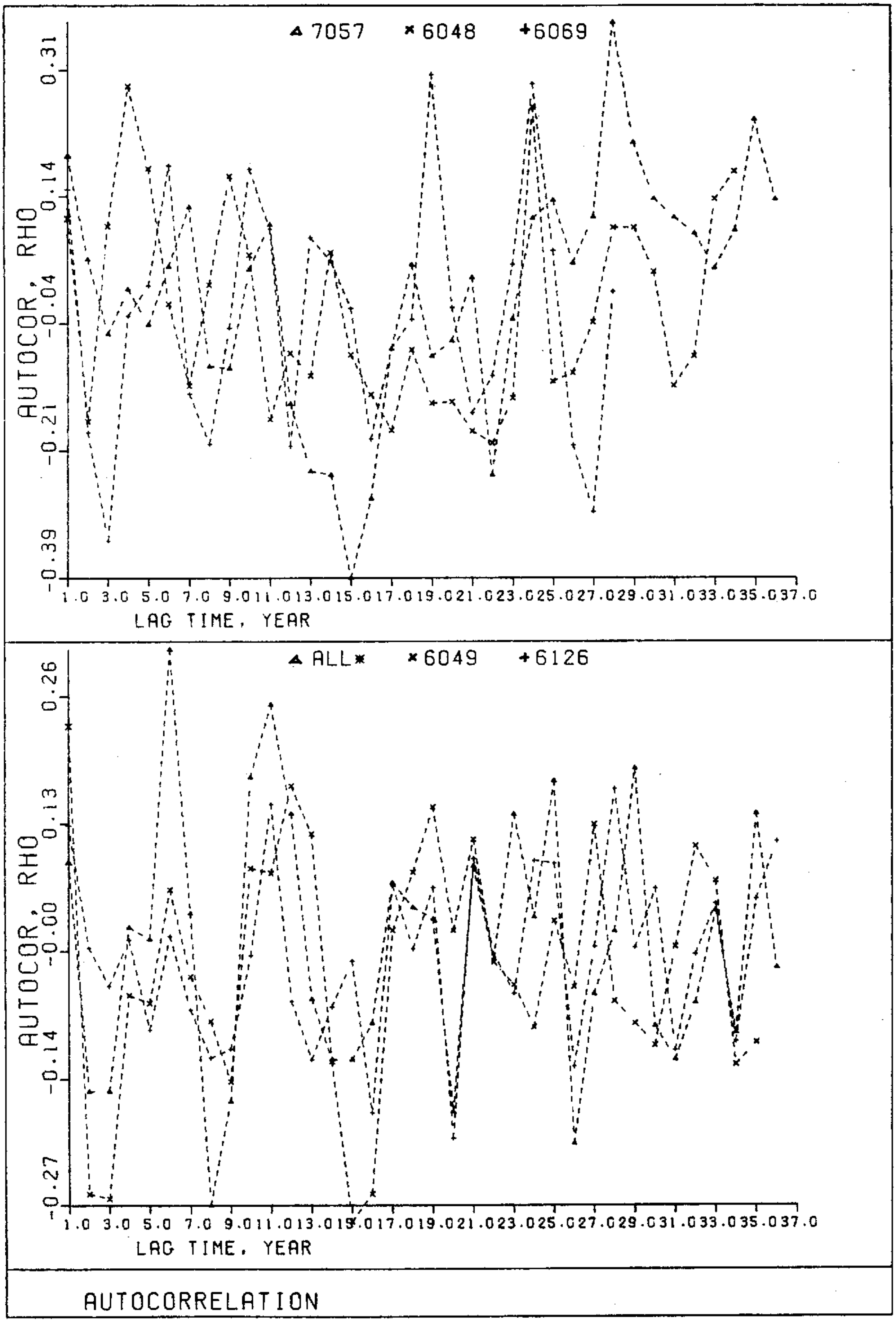
LITERATURE CITED

1. Dixon, W.J., 1981. BMDP Statistical Software 1981. University of California Press. Berkely, Ca.

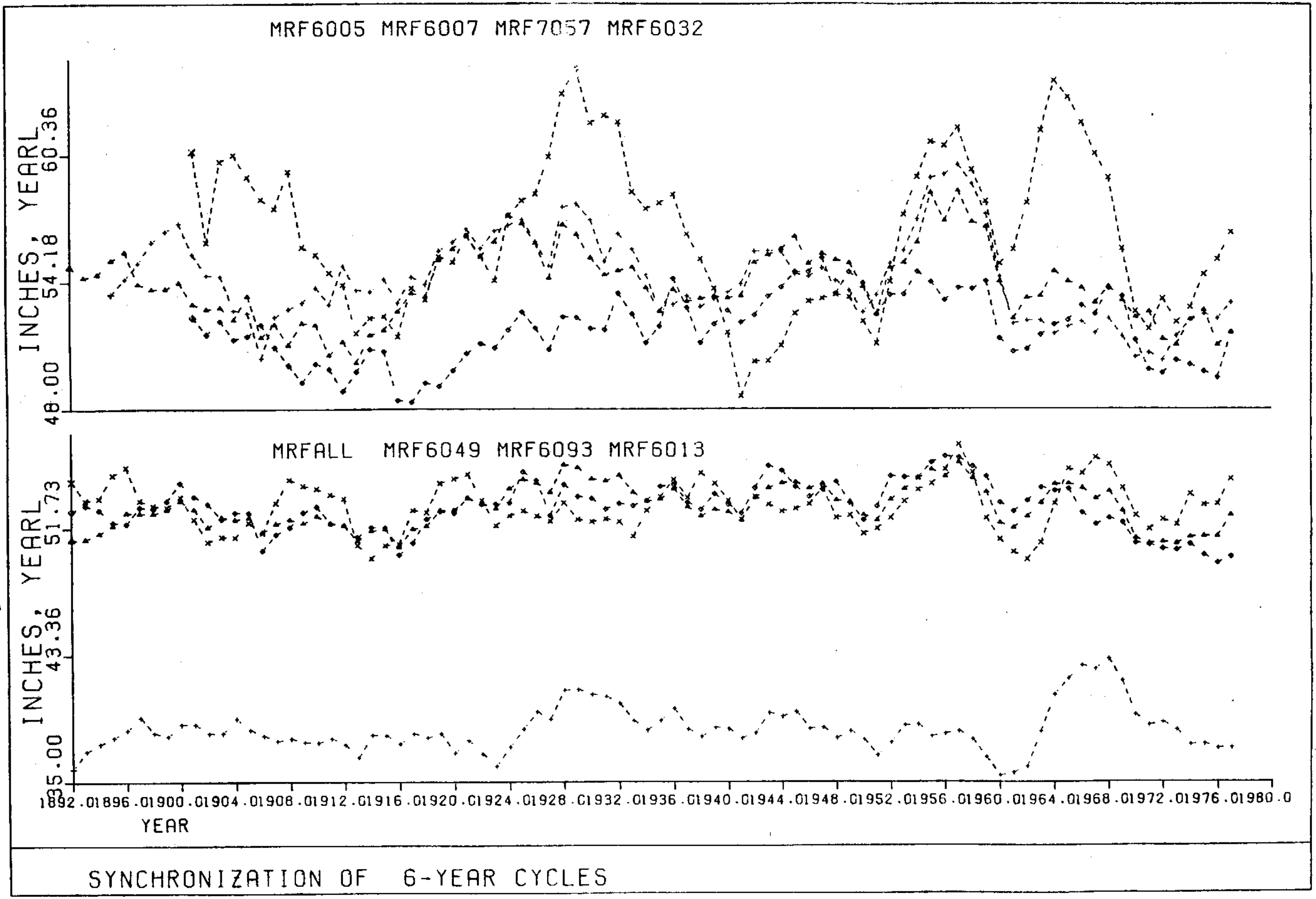
2. Shih, G., 1983. Data Analysis to Detect Rainfall Changes in South Florida. South Florida Water Management District, Resource Planning Department Technical Memorandum.



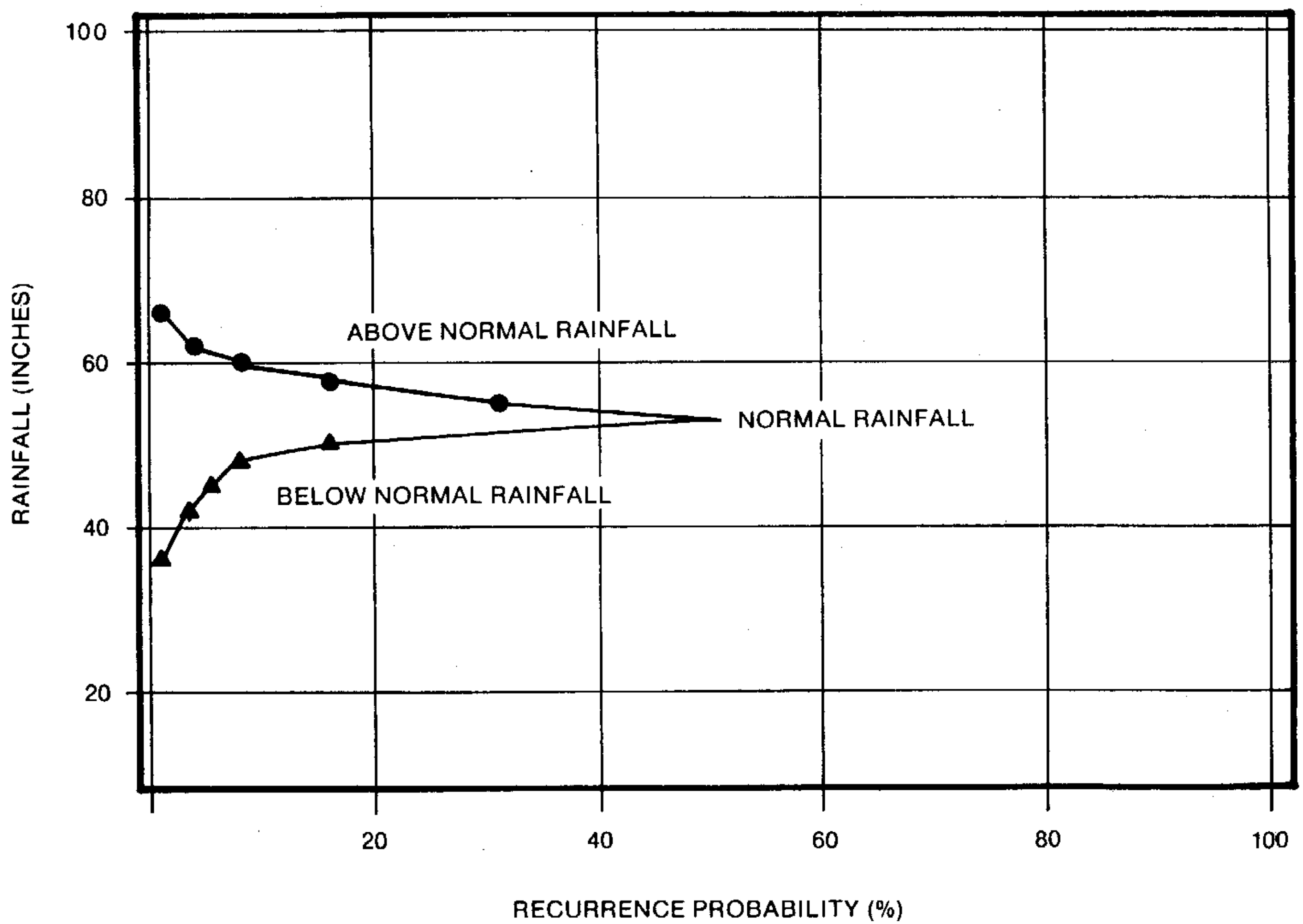
Appendix A   Sample Variance Spectrum



Appendix B Correlogram of Some Stations



Appendix C Synchronization of 6-Year Cycle



Appendix E Annual Rainfall Probability For Series All*

APPENDIX D
"Best" ARIMA Model

Station	Standard Deviation	Estimation Errors (In.)	ARIMA (P.D.Q.)		
ALL*	7.20	7.08		D = 6;	Q = 6, 11
MRF 6049	9.24	10.81	P = 8;	D = 8;	Q = 8
MRF 6093	9.65	10.01		D = 10;	Q = 10
MRF 6013	9.45	10.72	P = 6;	Q = 6	
MRF 6005	9.67	11.00	P = 6;	Q = 6	
MRF 6007	9.28	12.77	P = 11		
MRF 7057	13.91	14.60	P = 28		
MRF 6032	9.64	9.93	P = 6;	D = 11	
MRF 6015	10.17	12.49	P = 17;	Q = 2	
MRF 6126	11.36	13.49	P = 11		
MRF 6069	13.31	19.34	P = 6;	D = 12	
MRF 6119	10.24	12.66	P = 5		
MRF 8500	12.42	10.60	P = 24		
MRF 6044	8.16	10.14	P = 1;	Q = 1	
MRF 7034	9.83	11.21	P = 6, 17		
MRF 6024	10.08	13.31	P = 10, 29		

APPENDIX F

MODEL FOR WET AND DRY YEARS

1. Let X_i, $i=1,\text{---},N$ be the yearly rainfall of a station. There are N years of record. The mean and standard deviation of the series are x and Sx, respectively. The variation coefficient is

$$Cvx = Sx/x.$$

2. To study the L cycle year, the X_i series is arranged as follows:

$$\begin{array}{cccccc}
x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\
x_7 & x_8 & x_9 & x_{10} & x_{11} & x_{12} \\
x_{13} & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & .
\end{array}$$

or let it be denoted by

$$\begin{array}{cccccc}
Y_{11} & Y_{12} & Y_{13} & Y_{14} & Y_{15} & Y_{16} \\
Y_{21} & Y_{22} & Y_{23} & Y_{24} & Y_{25} & Y_{26} \\
Y_{31} & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & .
\end{array}$$

3. Let n_j, $j = 1,\text{---},L$ be the number of entries in each column. It is obvious $n_j = N/L$ or $n_j = N/L + 1$. L from 3 to N/3 are studied.
4. Some statistics of each column are computed. Let

$$Y_j = \sum_{k=1}^{n_j} \frac{Y_{kj}}{n_j}$$

$$Sy_j = \left[\sum_{k=1}^{n_j} \frac{(Y_{kj} - Y_j)^2}{n_j - 1} \right]^{\frac{1}{2}}$$

$$Cv_{yj} = \frac{Sy_j}{Y_j}$$

It is by observing that Cv_{yj} of the peaks and valleys of known important cycles reduced to about one half ($\frac{1}{2}$) of the average Cvx that promoted the further investigation of this model.

5. Let

$$t_j = \frac{(Y_j - X)}{\left[(N-1)S_x^2 + (n_j - 1)Sy_j^2 \right]^{\frac{1}{2}}} \left[\frac{Nn_j(N + n_j - 2)}{N + n_j} \right]^{\frac{1}{2}}$$

6. The first (shortest) significant cycle requires that

$$(i)\ Sy_j < Sx$$

$$(ii)\ Cvy_j < Cvx$$

$$(iii)\ |t_j| > t_c$$

t_c is at one tail 90% significance with ($N + n_j$) degrees of freedom. The absolute value is to be applicable for both wet and dry years. For a given cycle, there may be more than one that satisfies these requirements, but only those at the extremes (peak or valley) are of interest. The maximum t_j (for the wet year) and minimum t_j (for the dry year) is used for the selection of peak and valley. Those extremum for each cycle are denoted by capital subscripts.

7. The succeding significant cycle requires that

$$(i)\ Sy_M < Sx$$

$$(ii)\ Cvy_M < Cvx$$

$$(iii)\ |t_M| \geq |t_S|$$

$$(iv)\ |Y_M| \geq |Y_S|$$

where subscripts M and S denote cycle length. Requirements (*iii*) and (*iv*) are because of the monotonic nature for magnitude of recurrence events.

8. Year position of a cycle is simply counting where the peak or valley falls in the sequence of a cycle.

9. For the timing forecast of a year, refer to No. 2. and extend to the future for the wet or dry year forecast. For example, if a station started in the year 1892, for a 12 year cycle the wettest year falls in position 9, the latest cycle started in 1976 (1892 + 12x7). Since the first position of the cycle started in 1976, the 9th position is 1984.

   Note that the forecast of a wet year from one cycle may coincide with the forecast of a dry year from another cycle when:

$$STY + CW \cdot N + PW = STY + CD \cdot M + PD$$

where:

STY	=	Starting year of the station
CW,CD	=	Cycle length of wet and dry cycle, respectively
N,M	=	Integer multipliers
PW,PD	=	Year position of peak (wet) and valley (dry) in the cycles CW and CD, respectively

   It is obvious that the equation can be true only when CW and CD are not multipliers of each other.

10. The probability of the wettest (or driest) year of the cycle falling in the forecasted year is computed by assuming an exponential distribution. Let L be the cycle length of interest, say L=6 years. Refering to 2 above, Y_{kj} is replaced by rank R_{ki} of Y_{kj} in each row, For example, R_{ki} = rank of Y_{kj} in Y_{kj}, j =1,...,L. R_{ki} takes one of the integral values between 1 and L. Y_{kj} is transformed into R_{ki} as

$$\begin{array}{cccccc} R_{11} & R_{12} & R_{13} & R_{14} & R_{15} & R_{16} \\ R_{21} & R_{22} & R_{23} & R_{24} & R_{25} & R_{26} \\ R_{31} & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \end{array}$$

Suppose position year p, say p=4, is indentified to be the wettest (or driest) year of the cycle. Then the probability of the wettest yer expected to fall in thes year is

$$P(\%) = \frac{\left[\dfrac{\sum_{k=1}^{n_p} \left(\frac{1}{L}\right)^{(L - R_{kp})}}{n_p} \right]}{\sum_{l=1}^{L} \left(\frac{1}{L}\right)^{l-1}}$$

and the probability of the driest year expected to fall in this year is

$$P(\%) = \frac{\left[\dfrac{\sum_{k=1}^{n_p} \left(\frac{1}{L}\right)^{(R_{kp} - 1)}}{n_p} \right]}{\sum_{l=1}^{L} \left(\frac{1}{L}\right)^{l-1}}$$

The exponential distribution assumption is checked by actual percentage of years that the extrema occured in the predicted year from the short cycles where most data points are available.

Appendix G. Wet and Dry Year Forecasts of Individual Stations

CYCLE	WET (PEAK OF CYCLE)					DRY (VALLEY OF CYCLE)				
		TIMING		AMOUNT (Inches)			TIMING		AMOUNT (Inches)	
(YEAR)	YEAR Position	YEAR	Prob. %	Expected RF	St. Dev.	YEAR Position	YEAR	Prob. %	Expected RF	St. Dev.
STATION: MRF 6049 START: 1833 MEAN: 38.42 ST. DEV: 9.24										
6	2	1984	31	32.30	8.93					
13						2	1990/91	25	33.45	6.80
15	2	1984	30	47.64	10.30					
17						10	1995	24	30.61	4.11
25						6	1988	17	27.94	4.32
STATION: MRF 6093 START: 1892 MEAN: 53.27 ST. DEV: 9.65										
10						4	1985	21	46.49	8.99
12	9	1982/84	68.94	61.57	9.25					
STATION: MRF 6013 START: 1892 MEAN: 53.03 ST. DEV: 9.45										
6	3	1984	21	57.14	9.24	5	1986	36	47.76	7.97
10	8	1989	14	59.14	8.91	10	1991	24	45.52	6.60
18	3	1984	20	64.34	8.30					
21	14	1989	26	65.33	5.48					
24						17	2004/05	32	43.08	3.40
STATION: MRF 6005 START: 1892 MEAN: 54.05 ST. DEV: 9.67										
6						5	1986	30	48.79	7.14
18						11	1992	20	44.37	6.71
25						11	2002	33	42.86	5.19
29	10	1988	32	64.54	9.33					
STATION: MRF 6007 START: 1895 MEAN: 54.30 ST. DEV: 9.28										
5						2	1986	41	49.07	7.11
10	9	1983/84	25	60.58	7.36					
17						11	1994/96	40	45.44	6.74
24						22	1988	32	44.05	2.65
27	9	1993	64	72.53	2.28					

CYCLE (YEAR)	WET (PEAK OF CYCLE) YEAR Position	WET TIMING YEAR	WET TIMING Prob. %	WET AMOUNT (Inches) Expected RF	WET AMOUNT (Inches) St. Dev.	DRY (VALLEY OF CYCLE) YEAR Position	DRY TIMING YEAR	DRY TIMING Prob. %	DRY AMOUNT (Inches) Expected RF	DRY AMOUNT (Inches) St. Dev.
STATION: MRF 7057 START: 1901 MEAN: 56.57 ST. DEV: 13.91										
6	5	1983/84	59	64.93	11.05					
12	12	1983/84	18	67.91	7.15	2	1985/86	18	46.60	6.35
24	12	1983/84	33	73.21	4.51					
STATION: MRF 6032 START: 1901 MEAN: 51.97 ST. DEV: 9.64										
6	6	1984	37	57.80	6.91	4	1988	34	49.10	7.46
15	3	1992	21	59.88	2.64					
21	15	1999	32	62.40	2.08					
23						16	1985/86	33	45.28	1.22
STATION: MRF 6015 START: 1907 MEAN: 52.21 ST. DEV: 10.17										
5						5	1986	44	45.34	8.23
15	1	1982/84	40	61.93	8.49	10	1991/92	22	40.64	5.14
STATION: MRF 6126 START: 1910 MEAN: 62.34 ST. DEV: 11.36										
4	3	1984	35	67.15	10.49					
6						4	1985	31	56.94	6.57
STATION: MRF 6069 START: 1912 MEAN: 61.71 ST. DEV: 13.31										
6	1	1984	36	68.10	10.65	3	1986	32	56.00	11.36
11	3	1992	46	75.42	12.94					
16						12	1986/87	27	43.80	4.62
STATION: MRF 6119 START: 1924 MEAN: 56.79 ST. DEV: 10.24										
6	1	1984	38	60.27	5.95	2	1985	49	48.62	8.16
9	1	1987	31	64.99	6.59					
12	11	1994/95	47.93	67.18	4.48					
18						15	1992	63	43.95	6.41

CYCLE (YEAR)	WET (PEAK OF CYCLE)					DRY (VALLEY OF CYCLE)				
	YEAR Position	TIMING		AMOUNT (Inches)		YEAR Position	TIMING		AMOUNT (Inches)	
		YEAR	Prob. %	Expected RF	St. Dev.		YEAR	Prob. %	Expected RF	St. Dev.
	STATION: MRF 8500 START: 1928 MEAN: 62.93 ST. DEV: 12.42									
6	3	1984	42	72.38	9.07	5	1986	40	53.24	6.49
12	9	1984	25	74.99	7.42					
	STATION: MRF 6044 START: 1929 MEAN: 51.74 ST. DEV: 8.16									
7						7	1984/85	33.60	46.11	4.92
13	12	1992	25	58.71	6.22					
17	17	1996	31	61.08	2.95					
	STATION: MRF 7034 START: 1937 MEAN: 42.81 ST. DEV: 9.83									
8	5	1989	20	47.14	7.01					
10	3	1990/91	25	51.51	3.85	9	1985	23	35.09	4.83
15						10	1991		31.08	6.51
	STATION: MRF 6024 START: 1916 MEAN: 50.26 ST. DEV: 10.08									
6						4	1985	26	47.90	5.83
9						7	1986	16	45.85	2.56
12	9	1984	18	57.46	6.84					
13						1	1994	20	43.95	4.38
16						7	1986	48	40.99	5.26