National Symposium on Hydrology
Dec.16-18,1987, N.I.H., Roorkee

A COMPARATIVE STUDY OF RAINFALL SIMULATION USING TRANSITIONAL PROBABILITY MATRIX METHODS

S. Janakiraman* K. Venugopal,* and R. Sakthivadivel*

ABSTRACT

Daily rainfall values of North-East monsoon from September to December (122 days) are used for the formulation of Transitional Probability Matrices. Sixteen raingauge stations of Bhavani basin in India with lengths of record ranging from 23 years to 32 years are analysed. Synthetic generation is carried out using three different methods.

Daily rainfall values are converted into states by selecting suitable class intervals. Using the states thus arrived at transitional frequency and hence cumulative transitional probability is worked out for each station. In the first method, the class interval includes zero rainfall values as state 1. Generation is done with rectangular distributed random numbers in conjunction with the transitional probability matrices obtained continuously for one year.

In the second approach conditional probabilities are worked out for dry/wet days being followed by dry/wet days. The transitional probability matrix is arrived at taking into account only days with rainfall values greater than 0.1 mm. Synthetic generation is done in two parts. First part decides whether a particular day is dry or wet using conditional probability and rectangularly distributed random number. The second part decides the state of a wet day using cumulative transitional probability matrix.

In the third approach, wet and dry spells are identified in the historic sequence and fitted with geometric distributions. The transitional frequency matrix used in method 2 is used to fit two-parameter gamma distributions. Generation is done in two steps. First wet or dry spells are generated using the wet or dry spell distribution function. For wet spell duration the state for each day is determined from the cumulatives transitional probability matrix and rectangularly distributed random number.

The rainfall states generated are converted into rainfall values using mid value of the corresponding class interval. Synthetic generation is done for 100 years. The historic and synthetic rainfall values are compared by the Chi-square test, the F-test, Students' t-test and the median test and the results are discussed.

The first method gives best results except for the fact that the variance of the generated series is less. The fact that wet and dry spell sequences follow geometric distribution needs to be locked into as the same is not true for many of the stations.

INTRODUCTION

The tremendous increase in the scope and scale of planningand development of water resources systsems has led to a proliferation of mathematical models of hydrologic processes. The expanding rational hydrologic data bases, advances in hydrologic research and computer technology have provided an added impetus for the multiplication of hydrologic models.

*Centre for Water Resources, Anna University, Madras-600025

The synthesis of a rainfall series at a point is used in many applications. The technique of the Markov chain is one of the approaches often adopted for synthetic generation of daily rainfall. This study relates to such applications.

METHODOLOGY

For the synthesis of daily rainfall series, using Markov Chain Models, three approaches 1(a), 1(b) and 2 are studied.

A first order Markov Model is written as

$$pr(X_t=x_t/X_{t-1}=x_{t-1}\ldots\ldots X_1=x_1) = pr(X_t=x_t/X_{t-1}=x_{t-1}) \quad (1)$$

If $X_{t-1}=i$ and $X_t=j$, then the system has made a transition from state i to state j at the t^{th} step. The probabilities of the various transitions that may occur can be written as:

$$P_{ij} = pr\ (X_t=x_t/X_{t-1}=x_{t-1}) \quad (2)$$

The transitional probabilities are estimated from the transitional frequencies obtained from the historic data, using the following equation:

$$\hat{P}_{ij} = f_{ij}/F_i \quad (3)$$

where f_{ij} represents the frequency of transition from state, i, at time, t-1, to state, j, at time, t.

$$F_i = \sum_{j=1}^{K} f_{ij} \qquad \text{for } i=1,2,\ldots\ldots,K \quad (4)$$

and K is the number of states.

In approach 1(a), using the cumulative transitional probabilities estimated from the historic data and rectangularly distributed random numbers, daily rainfall states are generated. The generated states are then converted to rainfall values using the mean values of rainfall corresponding to each state. In this method, a day with a rainfall less than 0.1mm is treated as a dry day and is assigned the state 1.

In Approach 1(b), the conditional probabilities viz, the probability of a wet/dry day being followed by a dry/wet day are estimated from the historic data using the following relationship:

$$p_1 = pr\ (\text{wet day/wetday}) \quad (5)$$

$$p_0 = pr\ (\text{wet day/dry day}) \quad (6)$$

Using a random number and the estimated conditional probabilities, the state of day, t, to be wet/dry is determined. If the day, t, happens to be wet, then its state is identified with another random number using the modified cumulative transitional probability matrix. The cumulative transitional probability matrix estimated in Approach 1(a) is modified in such a way that the probability of occurrence of a dry day for day 't', for any state, i on day t-1, is zero.

In Approach 2(Jovanovic et. al., Smith et al.), the rainfall sequence is represented by a series of dry spell (m days) and wet spell (k days). The probability distribution of length of spells is assumed to be geometric i.e.,

$$pr\ (x=k) = (1-p_1)\ p_1^{(k-1)} \quad (7)$$

$$pr\ (x=m) = p_0\ (1-p_0)^{(m-1)} \quad (8)$$

Then these probabilities are summed up sequentially to obtain wet spell duration distribution function F(k) and dry spell duration distribution function G(m). The transitional frequency matrix estimated in Approach 1(a) is fitted by a series of two-parameter gamma distributions, one for each state i, whose probability density functions are given by

$$P_{ij}(x_1 \leqslant x \leqslant x_2) = \frac{\lambda_i^{\eta_i}}{\Gamma(\eta_i)} \int_{x_1}^{x_2} x^{\eta_i - 1}\ e^{-\lambda_i x}\ dx \qquad i = 1,2,\ldots,K \quad (9)$$

where λ_i and η_i are the two parameters for each state i and Γ is a gamma function. Using these two parameters the cumulative transitional probability matrix is reconstructed.

Rectangularly distributed random number is used to read the duration of wet and dry spells successively from the corresponding distribution functions. The reconstructed transitional probability matrix based on two-parameter gamma distribution is modified as specified in Approach 1(b) and then is used for the generation of rainfall values with random numbers generated, for the duration of wet spell alone.

DATA BASE

For the simulation of daily rainfall the data of sixteen raingauging stations in Coonoor sub-basin of Bhavani Basin (Tamil Nadu, India) is used. Figure 1 shows the location of all the stations in Coonoor sub-basin and the Table I gives the name of each station along with the period of record. For the purpose of analysis here, North-East monsoon seasonal daily rainfall spread over 122 days from September to December has been used. The data for this work has been obtained from the Tamil Nadu Electricity Board.

COMPUTATIONAL PROCEDURE

To illustrate the methodology only one station out of 16 stations viz., Carolina is selected. The generation process is carried out for a period of 100 years.

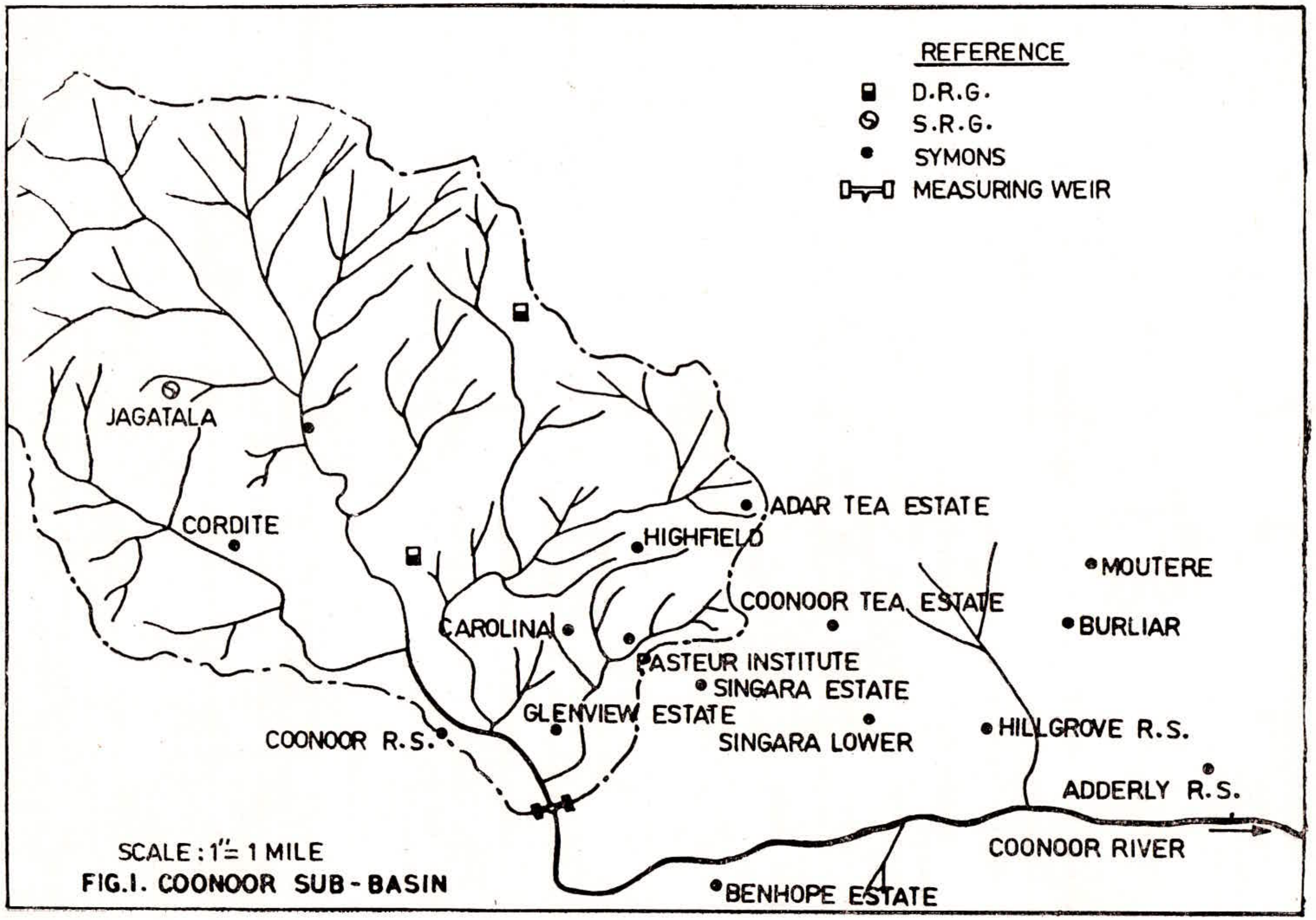
Approach 1(a)

Daily rainfall values of Carolina for the monsoon season are grouped into 18 intervals known as states as shown in Table II. The eighteen states now constitute the state of the First Order Markov Chain Model. Using equation (3), the transitional probabilities are calculated and hence the cumulative transitional probability matrix.

Approach 1(b)

For the station viz., Carolina,

No. of wet days given the previous day wet	=	947
No. of dry days given the previous day wet	=	568
i.e., No. of wet/dry days given the previous day wet	=	1515
No. of wet days given the previous day dry	=	562



SCALE : 1" = 1 MILE

FIG.I. COONOOR SUB-BASIN

TABLE I. RAINFALL STATIONS IN COONOOR SUB-BASIN WITH THE PERIODS OF RECORD

Sl. No.	Name of station	Years of record
1	Carolina	32
2	Coonoor Tea Estate	32
3	Cordite	32
4	Singara Estate	32
5	Burliar	31
6	Moutere	31
7	Highfield	28
8	Adar Tea Estate	26
9	Adderly R.S.	26
10	Hillgrove R.S.	26
11	Pasteur Institute	26
12	Benhope	25
13	Glenview Estate	25
14	Jagatala	25
15	Singara Lower	25
16	Coonoor R.S.	23

TABLE II. DAILY RAINFALL VALUES WITH CORRESPONDING STATES AND PREASSIGNED (MID-VALUES) RAINFALL VALUES FOR EACH STATE-CAROLINA

Daily rainfall interval (mm)	State	Midvalues (mm)
0.00	1	0.00
0.10 - 2.00	2	1.05
2.10 - 4.00	3	3.05
4.10 - 6.00	4	5.05
6.10 - 8.00	5	7.05
8.10 - 10.00	6	9.05
10.10 - 20.00	7	15.05
20.10 - 30.00	8	25.05
30.10 - 40.00	9	35.05
40.10 - 50.00	10	45.05
50.10 - 60.00	11	55.05
60.10 - 80.00	12	70.05
80.10 - 100.00	13	90.05
100.10 - 120.00	14	110.05
120.10 - 140.00	15	130.05
140.10 - 160.00	16	150.05
160.10 - 180.00	17	170.05
180.10 - 260.00	18	220.05

No. of dry days given the previous day dry = 1795
i.e., No. of wet/dry days given the previous day dry = 2357

The estimated conditional probabilities are:

pr(wet day/wet day) = 0.6251
pr(wet day/dry day) = 0.2384
pr(dry day/wet day) = 0.3749
pr(dry day/dry day) = 0.7616

The modified cumulative transitional probability matrix is prepared as explained earlier.

Approach 2

Simulation of wet and dry spell sequences

The transitional frequencies as well as the conditional probabilities of wet/dry days being folowed by wet/dry days at carolina are presented in Table III.

From the historical records the number of continuous dry spells and the number of continuous wet spells and their frequencies are found out. The maximum length of dry spell (m) = 39 days and the maximum length of wet spell (k) = 20 days.

The theoretical probability distribution of wet spell with length, k= 1,2,....20 days and dry spell with lengths, m=1,2,....39 days are found using equations (7) and (8) with $p_1=0.6251$ and $p_0=0.2384$. The probabilities, cumulative probabilities and the frequencies of theoretical dry and wet spells are computed. The cumulative probabilities of theoretical dry and wet spells respectively form the distribution functions for the dry spell, G(m) and wet spell, F(k). These distribution functions are represented in Figures 2(a) and 2(b). Simulation of dry/wet spells is initiated by assuming that the first spell is wet.

Synthesis of rainfall values

For each rainfall state, i, the observed transitional frequency is fitted by gamma distribution. The parameters, λ_i and η_i are calculated using the following relations:

$$\bar{x} = \frac{\sum_{j=1}^{K} f_{ij}.M_j}{\sum_{j=1}^{K} f_{ij}} \tag{10}$$

$$S_i^2 = \frac{\sum_{j=1}^{K} f_{ij}\ (M_j - x_i)^2}{\sum_{j=1}^{K} f_{ij}} \tag{11}$$

$$\lambda_i = \frac{\bar{x}_i}{S_i^2} \tag{12}$$

$$\eta_i = \bar{x}_i . \lambda_i \tag{13}$$

TABLE III. SUMMARY OF THE OBSERVED TRANSITIONAL FREQUENCIES AND CORRESPONDING CONDITIONAL PROBABILITIES OF RAINFALL EVENTS AT CAROLINA

Month	Previous day	Wet	Dry	Total	p_1	$1-p_1$	p_0	$1-p_0$
September	Wet	196	158	354	0.5537	0.4463		
	Dry	168	406	574			0.2927	0.7073
	Total	364	564	928				
October	Wet	336	164	500	0.6720	0.3280		
	Dry	165	295	460			0.3587	0.6413
	Total	501	459	960				
November	Wet	255	127	382	0.6675	0.3325		
	Dry	115	431	546			0.2106	0.7894
	Total	370	558	928				
December	Wet	129	106	235	0.5689	0.4511		
	Dry	96	629	725			0.1324	0.8676
	Total	225	735	960				
Season	Wet	947	568	1515	0.6251	0.3749		
	Dry	562	1795	2357			0.2384	0.7616
	Total	1509	2363	3872				

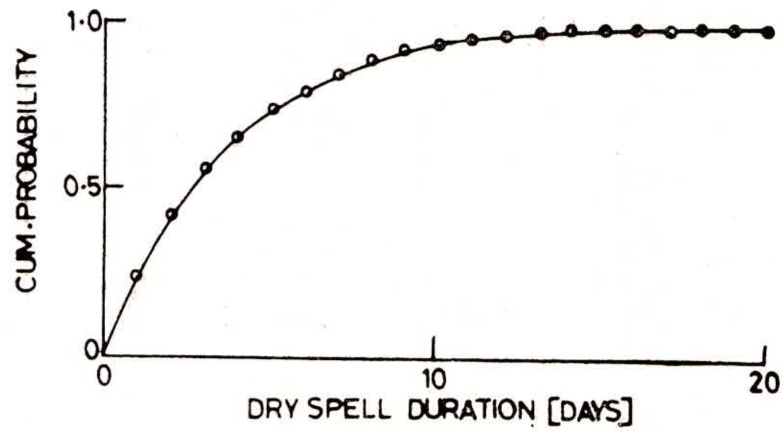


FIG. 2 (a) DRY SPELL DURATION DISTRIBUTION FUNCTION - CAROLINA

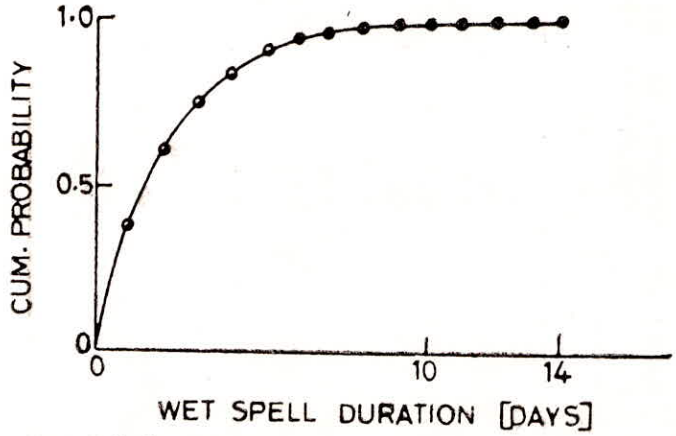


FIG. 2 (b). WET SPELL DURATION DISTRIBUTION FUNCTION - CAROLINA for $i = 1,2..,K$ and M_j is the middle value of rainfall interval.

Typical plots of observed and theoretical transitional frequency as obtained from gamma distribution for the first nine states (i=1 to 9) are shown in Figure 3. The values of λ_i and η_i are plotted against rainfall intervals as shown in Figure 4. Since the variation of λ_i is small, a mean value, namely, λ is taken instead of λ_i. The plot of η_i is fitted by a straight line and the values of η_i are read for each interval. The computations of λ and η_i are made on the above lines for the remaining stations and all the values are plotted in Figure5.

Using the mean value of λ and the interpolated value of η_i, the cumulative transitional probabilities are estimated using equation (9) and the cumulative transitional probabilities are modified as in Approach 1(b).

RESULTS AND DISCUSSIONS

The acceptability of stochastic models can be judged by studying whether the generated data fits very closely with the historic data or not. The historic and generated data are compared by conducting statistical tests such as Chi-square test, Student's t-test and the median test. The summary of results of statistical tests are furnished in Table IV

TABLE IV. SUMMARY OF THE STATISTICAL TEST RESULTS FOR HISTORIC AND SYNTHETIC SEQUENCES

Sl. No.	Parameter	Test conducted	Approaches 1(a)	1(b)	2
1	Frequency of states	Chi-square Test	16	16	Nil
2	Frequency of wet spell	Chi-square Test	*	3	4
3	Frequency of dry spell	Chi-square Test	*	1	Nil
4	Seasonal Rainfall	F and t-test	4	1	Nil
5	Average wet days in a season	F and t-test	15	13	*
6	Seasonal rainfall	Median Test	*	16	8

Note: Figures indicate the number of stations for which the test has been successful

* Test not conducted

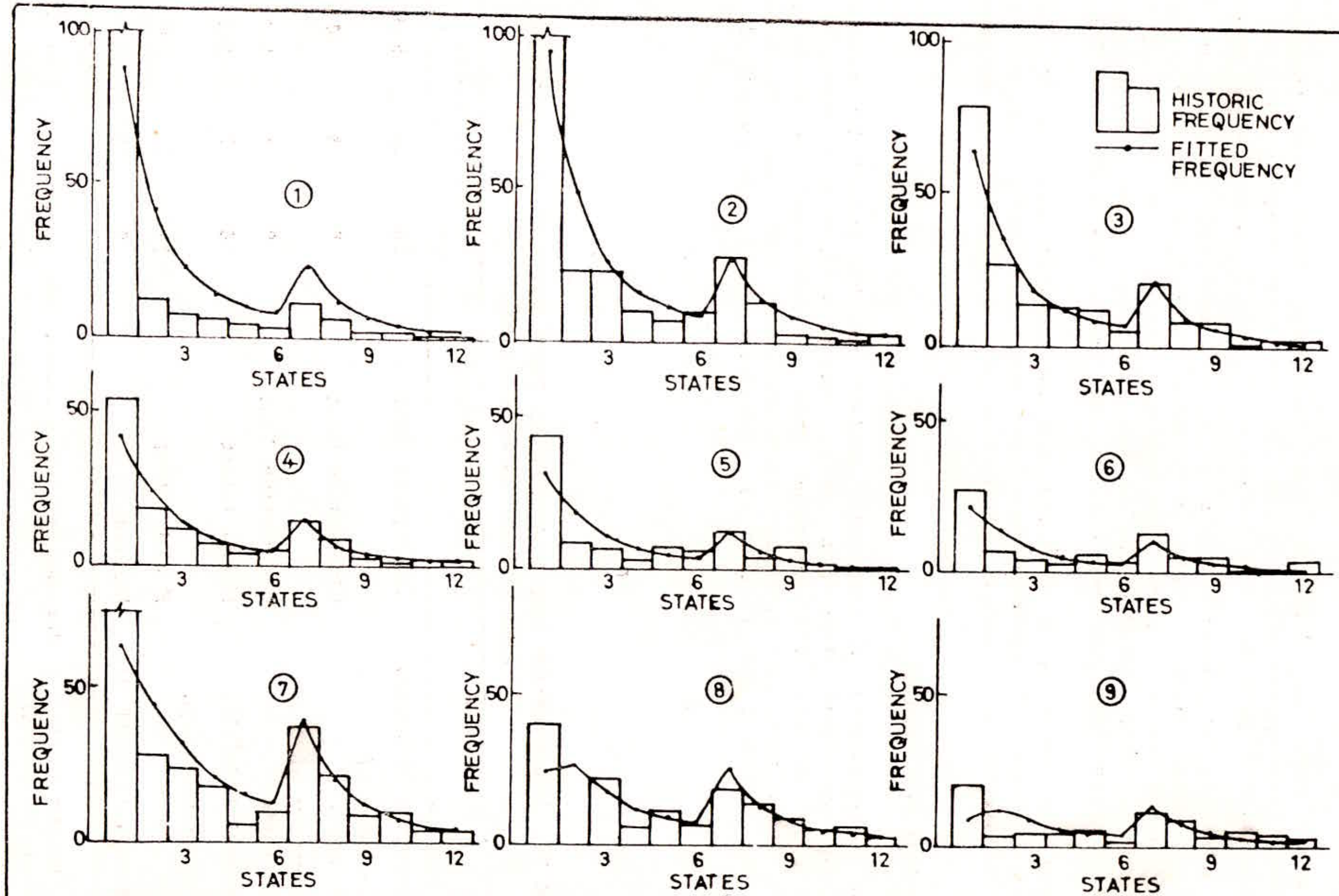


FIG. 3. THEORETICAL GAMMA DISTRIBUTION FOR STATES 1 TO 9 CAROLINA

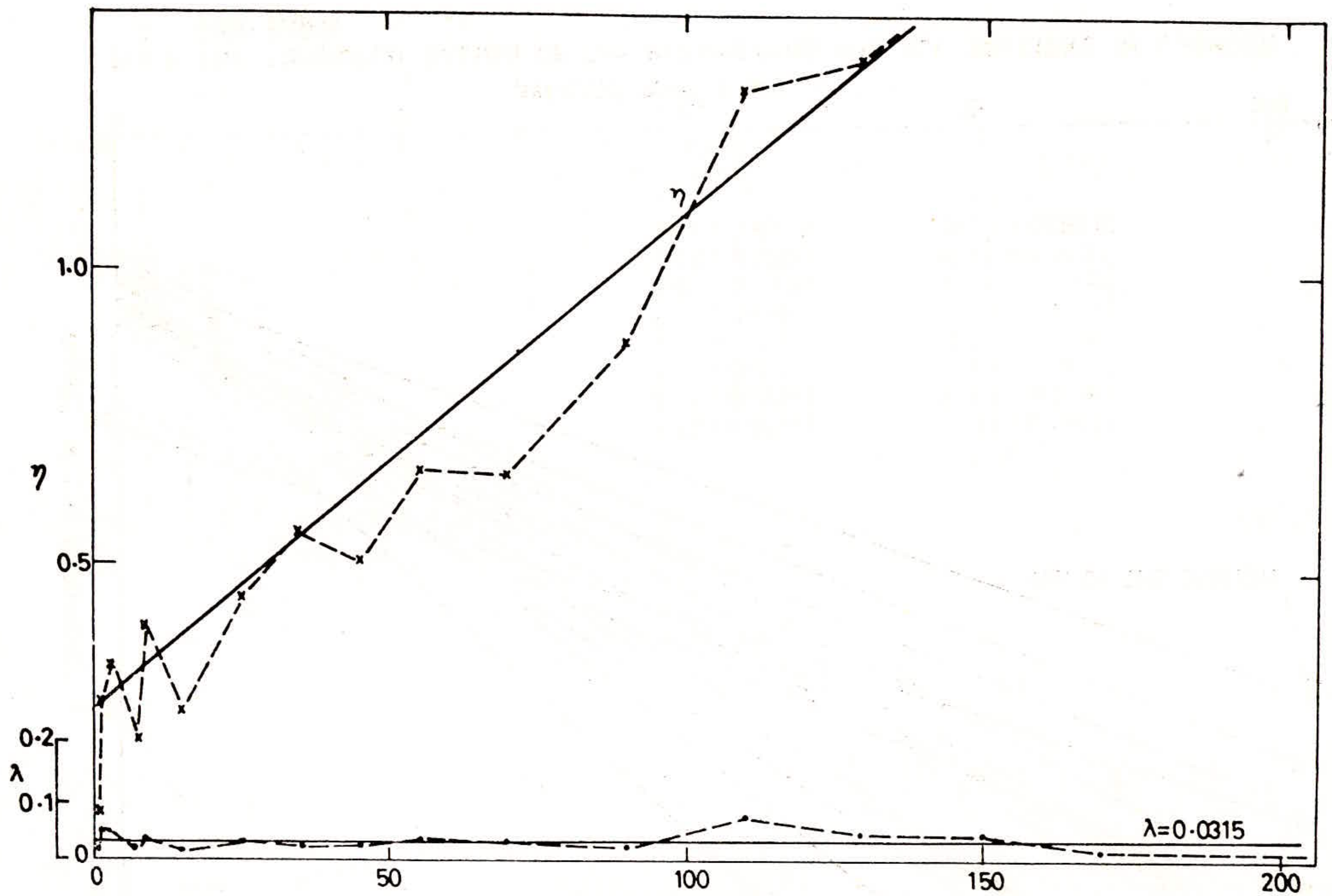


FIG. 4. PARAMETERS OF GAMMA DISTRIBUTION - CAROLINA

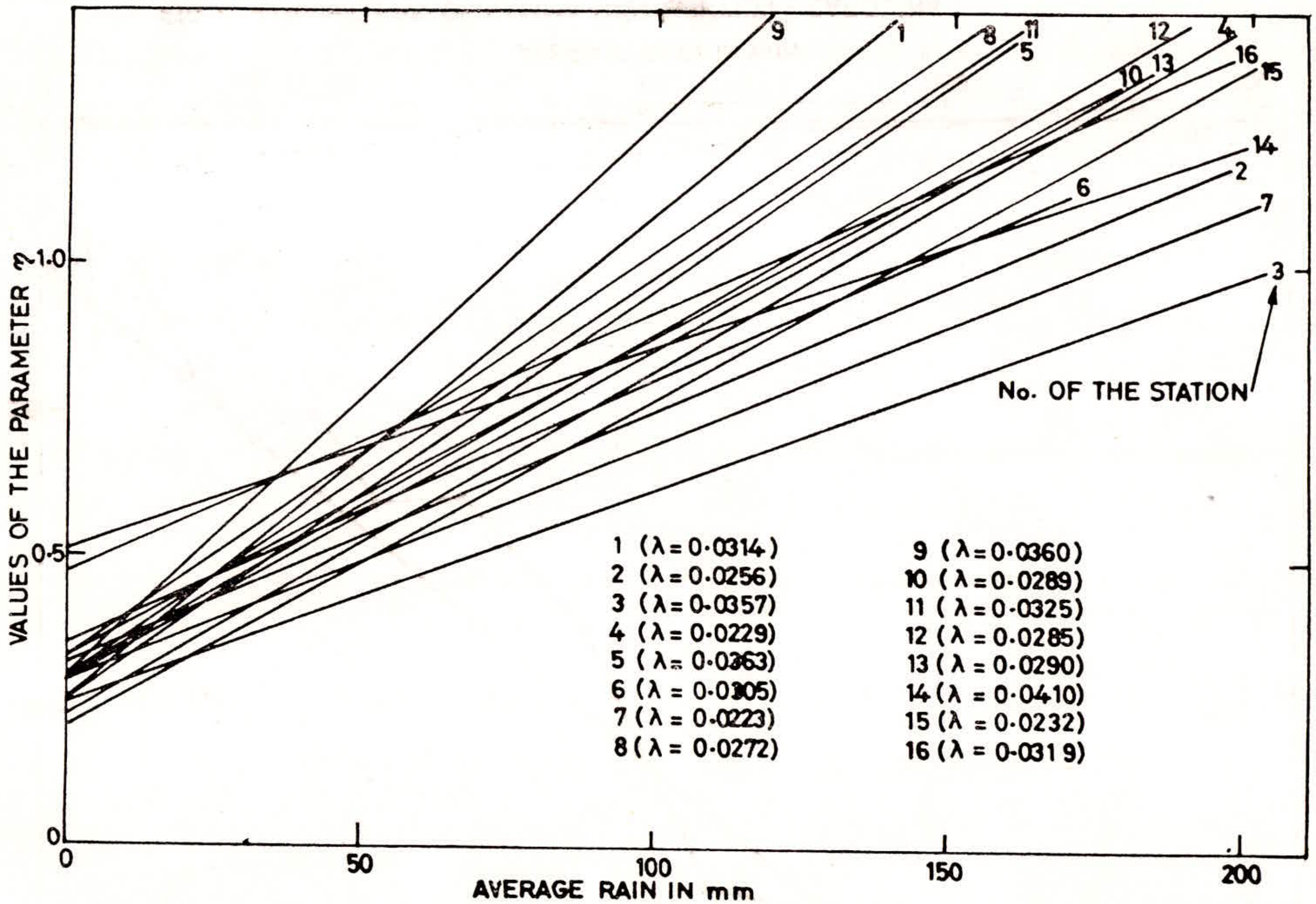


FIG.5. THE COMPUTED VALUES OF THE PARAMETERS FOR ALL STATIONS IN COONOOR SUB-BASIN

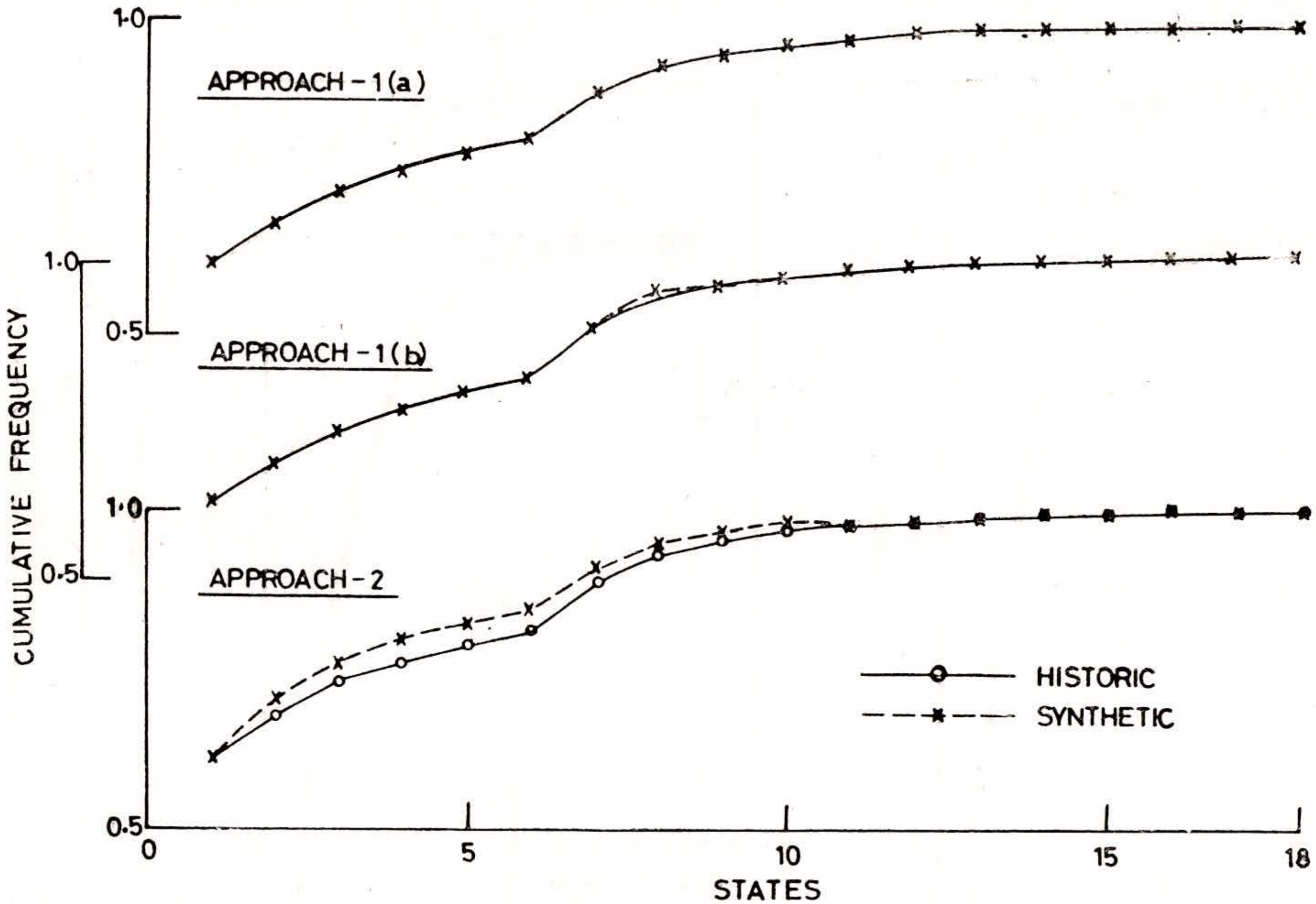


FIG. 6. CUMULATIVE FREQUENCY CURVES FOR DAILY RAINFALL FOR THE SEASON-CAROLINA

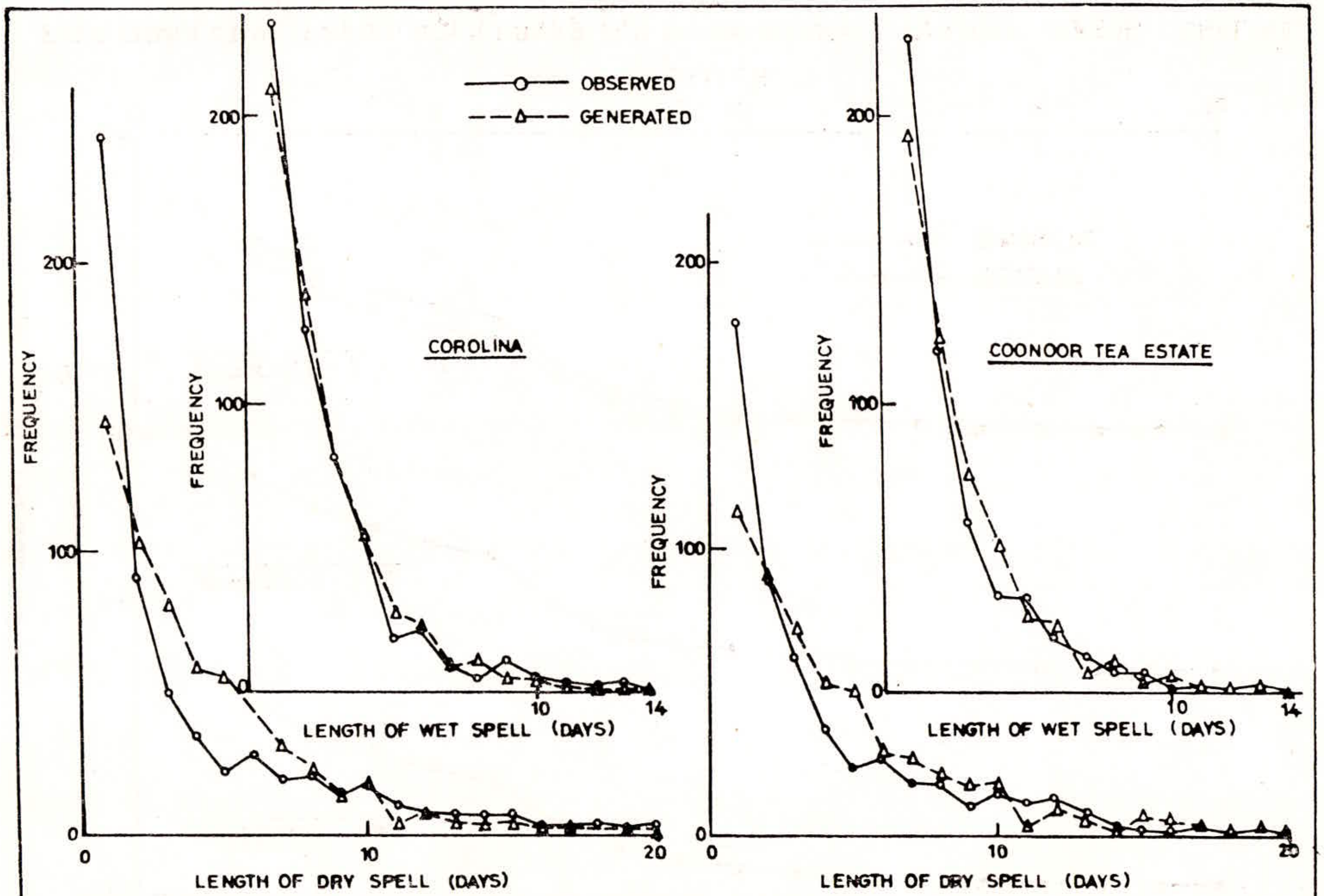


FIG. 7. OBSERVED AND SIMULATED DISTRIBUTIONS OF LENGTHS OF SPELLS - APPROACH 1 (b)

The historic and synthetic cumulative frequencies of states for Carolina for all the three approaches are shown in Figure 6. The observed and generated frequencies of wet and dry spells for two stations viz. Carolina and Coonoor Tea Estate are shown in Figure 7.

The historic and generated mean and standard deviation for all the stations using three Approaches viz. 1(a), 1(b) and 2 are presented in Table V.

CONCLUSIONS

It is clear from Tables IV and V that Approach 1(a) gives the best results except for the fact that the variance of the generated series is less.

Approach 1(b) which requires computation of conditional probabilities and modification of cumulative transitional probability matrix does not improve the quality of the generated series.

In Approach 2 assumption is made that the wet/dry frequencies follow geometric distribution (Jovanovic et. al., Smith et. al.,) but it is found that this is not true for most of the stations analysed.

REFERENCES

1. Haan, C.T., Statistical Methods in Hydrology, The Iowa State University Press, Ames, 1977.
2. Hughes,A. and Grawoig,D., Statistics, A foundation for Analysis, Addison-Wesley Publishing Co., U.S.A., 1971.
3. Janakiraman, S., Event type Markov Model for rainfall simulation in Bhavani Basin, M.E. Thesis, Anna University, Madras, India, July, 1982.
4. Jovanovic, S., Dakkak, A.R., Cabrik, M. and Brajkovic, M., Simulation of Daily Rainfall Series using Markov Chain Models, Proceedings of the Warsaw Symposium, Mathematical Models in Hydrology, Symposium Volume, 1, I.A.H.S.-A.I.S.H., Publication No.101, July, 1971, pp.110-120.
5. Khanal, N.N. and Hamrick, R.L., A stochastic model for Daily Rainfall Data Synthesis, Proceedings Symposium on Statistical Hydrology held at Tucson, Arizona, Aug.-Sept., 1971, Miscellaneous Publication No.1275, pp.197-210.
6. Sakthivadivel,R. and Dharaneeswaran, V., Stochastic Simulation of Daily Rainfall in Madras Basin, Tamil Nadu, Proceedings of the Symposium in Water resources development, Vol. I, New Delhi, April, 1980.
7. Smith, R.E. and Schreiber, H.A., Point Processes of Seasonal Thunderstorm Rainfall 1. Distribution of Rainfall events, Water Resources Research, Vol. 9, No.4, August, 1973, pp. 871-884.