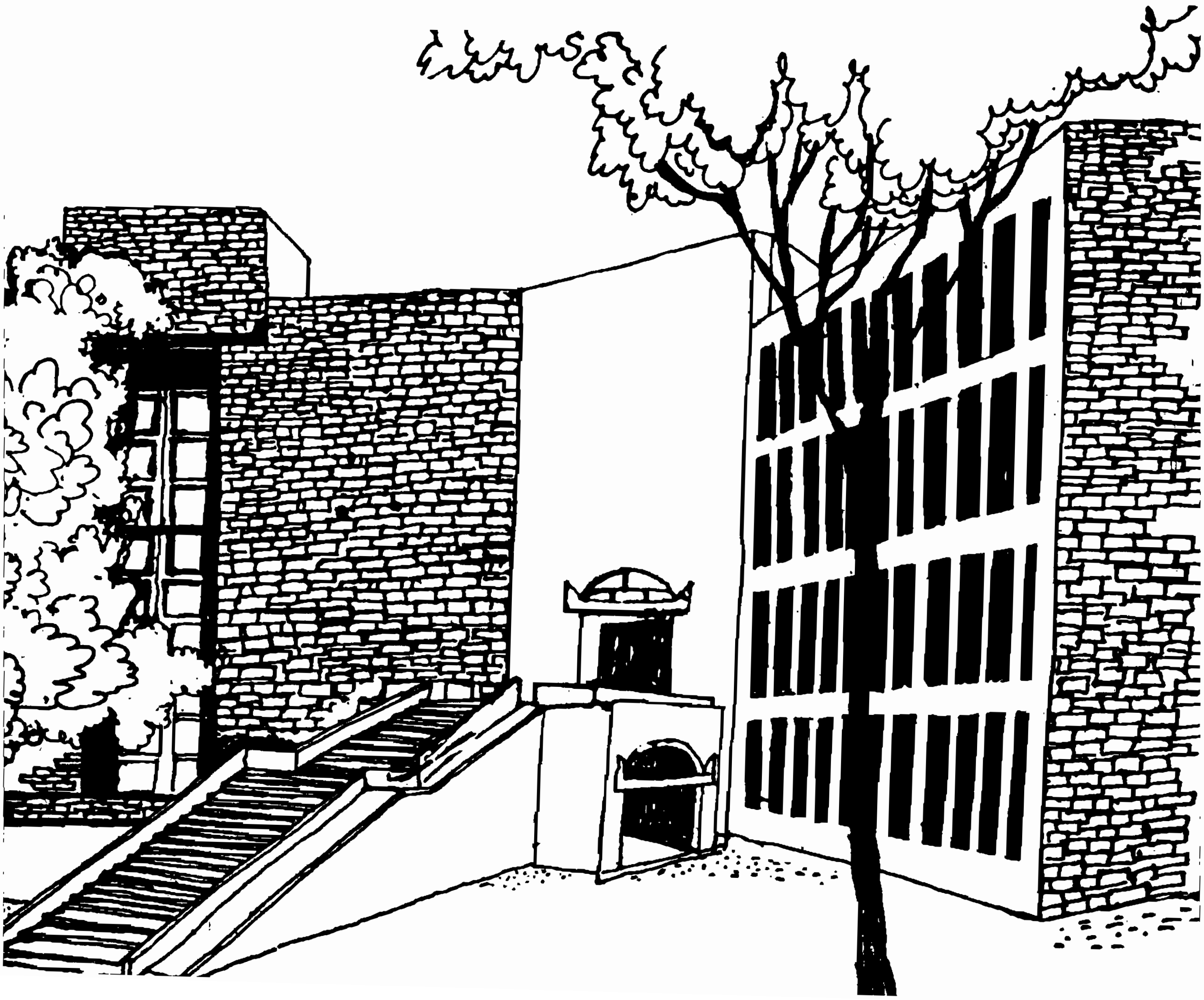
विद्याविनियोगाद्विकासः
IIM
AHMEDABAD

Working Paper



SYNTHETIC DAILY RAINFALL DATA GENERATION

By

Girja Sharan
M. Krishna Kumar

W P NO.1242
March 1995

INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD - 380 015
INDIA

Synthetic Daily Rainfall Data Generation

Girja Sharan **M. Krishna Kumar**

Centre for Management in Agriculture
Indian Institute of Management, Ahmedabad

Abstract

Long sequences of daily rainfall are often needed for simulation. These (when available) are cumbersome to input. Also in many cases historical data are too short to include all possible patterns. Hence the need for synthetic data. Discovering the stochastic structure underlying daily rains is the key to devising method for synthetic generation of data.

Some works have been reported that treat daily rains as multi-state Markov chain. This is useful in studies where one needs for instance the distribution of only the dry and wet spells etc. However, for use in simulation of run-off from a watershed, or for moisture budgeting and crop planning, or for scheduling of irrigation etc. one needs the magnitudes of rainfall and not just an interval. For these applications it appears necessary to look at daily rains as a Markov process as was done for instance by Carey and Haan for Kentucky. In this paper we report the results of using C&H method to generate synthetic data for Panchmahals district of Gujarat, a drought-prone area.

Properties of synthetic and historical data are compared.

Introduction

Long sequences of daily rainfall data are often needed for simulation. These (when available) are cumbersome to input. Hence the need for synthetic data. Discovering the stochastic structure underlying daily rains is the key to devising method for synthetic generation of data.

Review of Literature

There are several examples of daily rainfall being represented as a two state Markov chain (Gabriel 1962, Medhi 1985, Sharan 1988). In these works the aspect of interest was only the distribution of dry and wet spells. Accordingly two state description was adequate.

In many situations such as in moisture budgeting, estimation of daily run-off from a watershed, crop planning etc., one needs the magnitude of rainfall, not just the intervals. In such contexts, it is more appropriate to treat daily rains as Markov chain with continuous state space or a Markov process. One can partly fulfil this need by still retaining the chain structure, and dividing the state space more finely. But larger the number of states (divisions), longer the sequence of historical data required for parameter estimation. Allen and Haan (1975) treated daily rains as a seven state Markov chain. They stated that about 40 year data was required to get stable estimates.

What is desired therefore is a method that would yield magnitudes and yet not require unduly long historical sequence for parameter estimation.

Markov chain with continuous state space is defined as follows. If for all **n** and for all possible values of $\{x_n\}$ in $(-\infty,\infty)$

$$\begin{aligned} &Pr\{X_{n+1} \leq x \mid X_n = y, X_{n-1} = y_1, \ldots x_0 = y_n\} \\ &= Pr\{X_{n+1} \leq x \mid X_n = y\} \end{aligned} \tag{1}$$

then $\{X_n\}$ constitutes a Markov chain with continuous state space. In the case of daily rain the range of values of X_n will be $(0,\infty)$. The conditional distribution, on the right hand side of **equation 1**, defines the one-step-transition distribution function and is conventionally written as $P_1(y;x)$ or just $P(y;x)$. If **equation 1** is independent of value of **n**, the chain is said to be homogeneous.

Now, given the daily rainfall amount say **y**, on day **n**, one could get the distribution of daily rainfall on day **(n+1)** if one-step-transition distribution function, $P_1(y;x)$ is known using Chapman-Kolmogorov equation:

$$P_{n+1}(y;x) = \int_0^\infty P_n(y;z)\ P_1(z;x)\ dz \tag{2}$$

Equation 2 however is not very convenient, because the state space being continuous one would have to have an infinite number of such one-step-transition distribution functions.

Carey and Haan (1978) have tried to overcome this difficulty, by dividing the daily rainfall on a day in three distinct states (as one would do if it were a three state chain). State one is zero rainfall. The non-zero values were put into second and third states. The division of non-zero range was so made that number of events in states two and three were approximately equal. Given the initial state, say **i**, (not magnitude of rainfall), on day **n**, the distribution function for rainfall **amount** (not state) next day **(n+1)**, is then given by:

$$P(X_{n+1} \leq X \mid X_n \epsilon i,\ \text{season } k) = P_{i1}^k + (1-P_{i1}^k)F(X \mid i,k) \tag{3}$$

where X_n = rainfall on day **n**, falling in state **i**,

P_{i1}^k = transition probability of no rain on day **n+1** given X_n is in state **i**, season **k**.

These are estimated by the method of maximum likelihood. Let f_{i1}^k be the historical frequency of transitions from state **i** to state one in season **k** and f_i^k be the total number of occurrences in state **i**, season **k**. Then $p_{i1}^k = f_{i1}^k / f_i^k$,

F(X/i,k) = transition distribution function when some rain does occur on day $n+1 \mid X_n \in i$, season k.

They found empirically that F(X/i,k) for their location, was gamma distribution function given below.

$$F(x/i,k) = \int_0^x \frac{(\lambda_{i,k})^{-\eta_{i,k}}}{\Gamma(\eta_{i,k})} U^{(\eta_{i,k}-1)} \exp(-U/\lambda_{i,k})\, dU \qquad (4)$$

While **equation 3** gives the expression for distribution function, it is not really necessary to use it for generation of synthetic data. That task is accomplished easier through a two step simulation procedure which involves numerical integration of **equation 4**.

Objectives

This work aims at using the Carey Haan method for generating a synthetic sequence of daily rainfall, of a drought prone region - Dahod Taluk of Panchmahals District in Gujarat. Rain occurs here generally over four months - June, July, August & September. There is practically no rainfall beyond this period. We have 31 years of daily rainfall records with the rainfall measured to the nearest millimetre.

Plan of Work

(1) First, we will determine whether transition density function for June, July, August & September are indeed **G** (η,λ),

(2) if so, we shall use these to generate synthetic data, and then

(3) compare the statistical properties of historical and synthetic data.

Analysis

Transition Probability Density Function

Daily rainfall magnitudes (x) were divided into three states:

State 1	No rain ($x = 0$)
State 2	$0 < x \leq 5$ mm
State 3	$x > 5$ mm

Using this division, the transition distribution pattern from a given initial state (1, 2 or 3) to wet states (2 or 3) was obtained from historical data for each of the four months separately.

For the purpose of illustration, the patterns for September (initial states 2 & 3) are shown in **figures 1 & 2.** Visual observation suggested the possibility of gamma being the underlying distribution as of course already indicated by Carey & Haan.

The estimates of parameters (η & λ) determined for all the four months are shown in **table 1.** The goodness of fit test results are also shown in the table.

As can be seen the fit is good in only six out of twelve cases. We shall ignore the lack of fit in thosé cases and use all the twelve gamma distributions for generation of synthetic data.

Table 1: Estimated Parameters

Season	Initial State	λ_i	η_i	Degree of freedom	Computed χ^2	Critical χ^2
Jun	1	16.98	0.71	2	6.032*	5.991
	2	31.53	0.46	2	4.497	5.991
	3	35.10	0.63	2	4.123	5.991
Jul	1	30.76	0.42	5	4.652	11.071
	2	45.06	0.28	5	20.191*	11.071
	3	47.45	0.43	5	6.993	11.071
Aug	1	34.99	0.34	5	16.352*	11.071
	2	29.97	0.31	5	17.170*	11.071
	3	53.25	0.35	10	20.307*	18.307
Spt	1	31.74	0.51	4	4.553	9.488
	2	89.11	0.15	2	26.309*	5.991
	3	33.79	0.71	6	4.991	12.592

* Significant at 0.05 level.

Note: August estimates are based on 31 years data, others on 25 years.

The synthetic traces of daily rainfall are generated in the following way.

1. Determine initial state = **i**.
2. Generate a uniform random number x (0 - 1).
3. If $x \leq P_{i1}^{k}$, then $X_{n+1} = 0$.

4. If $x > P_{i1}^{k}$, generate a random observation u from $F(./i,k)$ and set $X_{n+1} = u$ and determine the state of X_{n+1}.

5. Repeat steps 2-4 changing seasons when required.

For the month of August, 31 synthetic sequences were generated. Their statistical properties are compared with the 31 years of historical data. For June, July, September, only six sequences were generated. The properties are compared with the last six years of historical data, which was set aside for this purpose and not used for estimation of parameters.

Results

Synthetic **(S)** data is compared with the historical **(H)** data on following aspects.

(a) mean daily rainfall, maximum daily rainfall,

(b) distributions of wet and dry spells,

(c) frequency distribution of magnitudes of rainfall in each season (month).

Mean Daily Rainfall

The mean daily rainfall and standard deviation for each month are shown in **table 2.** Significant difference is seen in the standard deviation and maximum rainfall of July. The number of wet days and wet & dry spell lengths are almost similar.

Table 2 : Historical and Synthetic Data - Comparison

		Jun	Jul	Aug	Spt
Mean daily rainfall (mm)	H	3.3 (9.6)	10.4 (42.1)	7.9 (3.4)	2.3 (9.9)
	S	2.5 (8.5)	9.1 (21.7)	8.8 (3.3)	3.9 (13.2)
Max. daily rainfall (mm)	H	69.0	519.0	223.0	85.0
	S	81.9	141.3	144.4	74.5
Wet days (no.)	H	7.3	14.7	17.9	4.3
	S	7.7	14.2	18.2	5.0
Mean wet spell length (days)	H	3.0	3.6	5.4	2.5
	S	3.2	3.4	4.9	2.4
Mean dry spell length (days)	H	9.1	5.4	4.5	10.1
	S	8.2	3.9	3.7	7.8

Note: Values in parentheses are corresponding standard deviations.

Spells - Wet & Dry

Figures 3 & 4 show the frequency of wet and dry spells for August, obtained from historical and synthetic data. The mean spell lengths are almost equal for historical and synthetic data. The spells for other seasons are not plotted since we have only 6 years of data for comparison. The respective means are however shown in **table 2.**

Frequency Distribution

A frequency analysis is done to see whether the frequency of rainfall amounts in different intervals of historical and synthetic data are similar. For illustration, plot of August is shown in **figure 6**. Goodness of fit tests show satisfactory results for June (α = 0.05, degree of freedom ν = 2), September (α = 0.05, ν = 2) and August (α = 0.05, ν = 7). The fit was not satisfactory for July and therefore has also been shown here (**figure 5**).

Summary and Conclusions

Daily rainfall is viewed as a Markov chain with continuous state space.

A method developed earlier by Carey and Haan for Kentucky (USA), was tried out to see whether it will yield useful results in a drought prone region of Gujarat, India, where rainfall is limited to four months of a year and is highly erratic. The daily rain data of Dahod taluk in Panchmahals district was used as the case.

Each of the four months of monsoon is treated as a season. The daily rainfall amounts in each season were divided into three states. Transition density functions were developed for each state. Thus a total of 12 transition density functions were developed. Suitability of gamma distribution was tried. It was found that in six cases, the fit was good. It was poor in four cases and very poor in two others.

Using the empirically developed transition density functions (including those that were poor fits), traces of synthetic rainfall were generated.

Properties of synthetic and historical data were compared for patterns and means of wet and dry spells, mean and standard deviation of daily rainfall and frequency distribution of rainfall amounts.

Wet and dry spells distribution followed same pattern in synthetic and historical data.

However the daily maximum produced by the model were significantly less than those of the historical ones. Frequency of events produced by the model in the high rainfall class (greater than 5 mm) was higher than those of the historical ones.

Figure 1

Distribution Pattern of Rainfall Amounts in september
(Transition from wet state (2) to wet states (2&3))

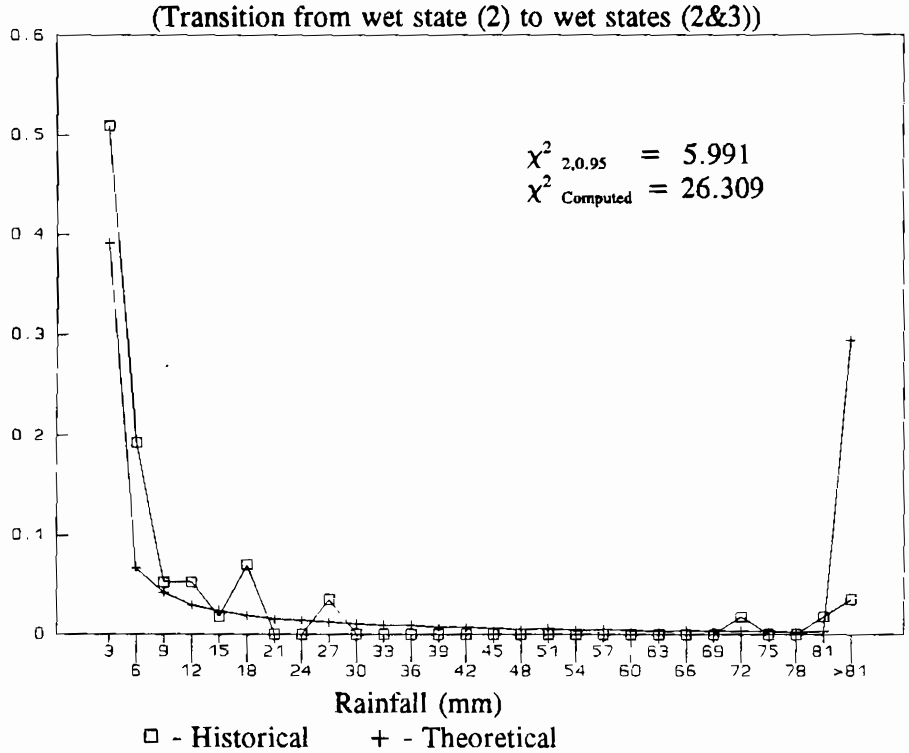


Figure 2

Distribution Pattern of Rainfall Amounts in September
(Transition from wet state (3) to wet states (2&3))

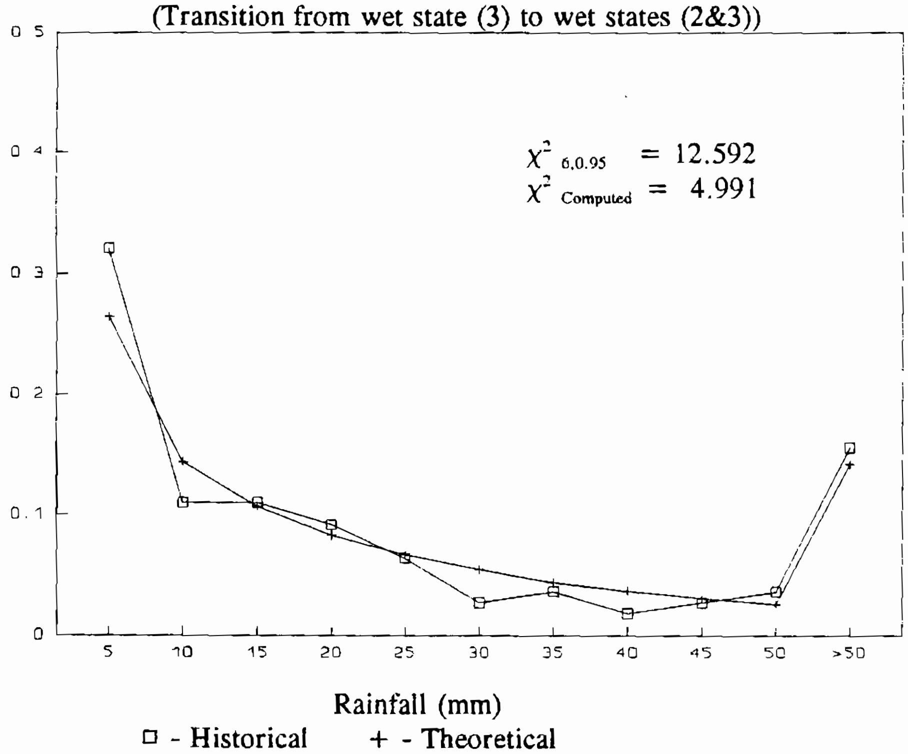


Figure 3

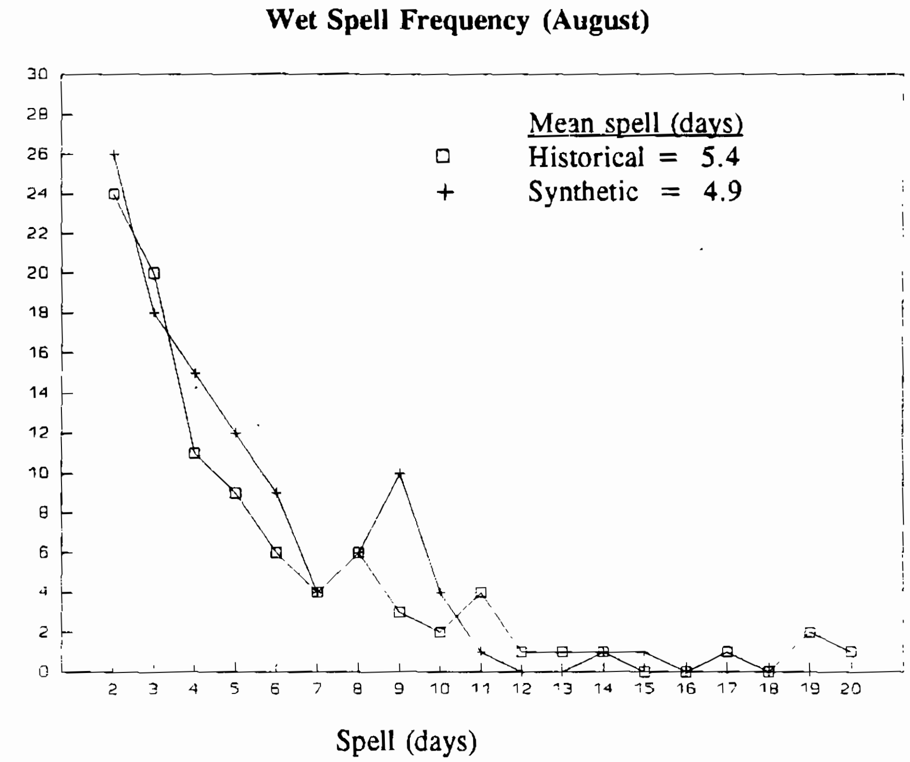


Figure 4

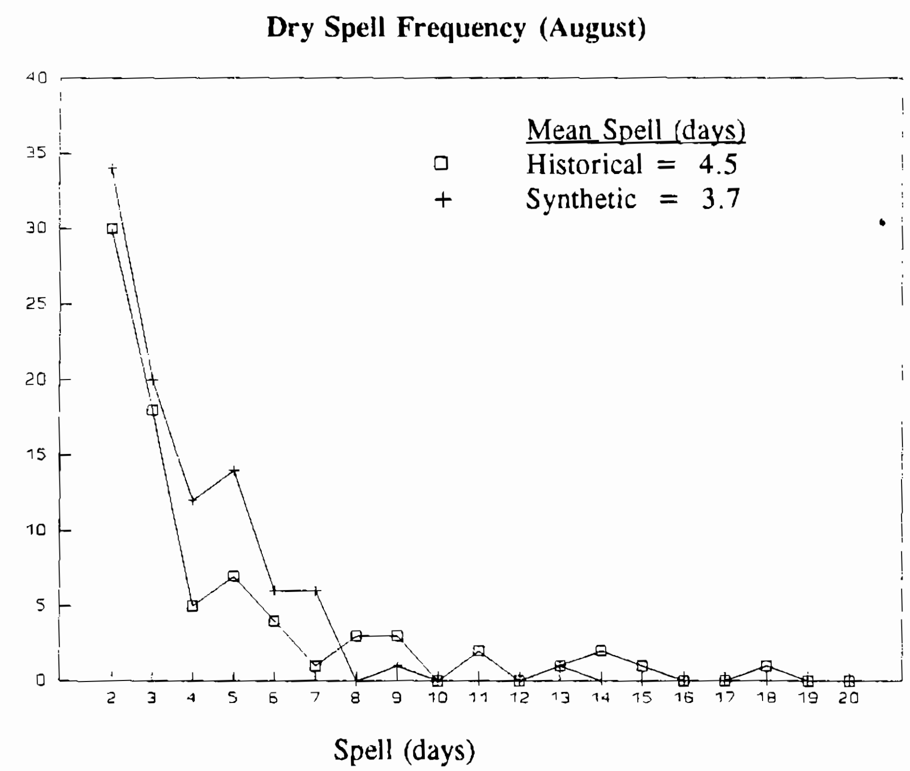


Figure 5

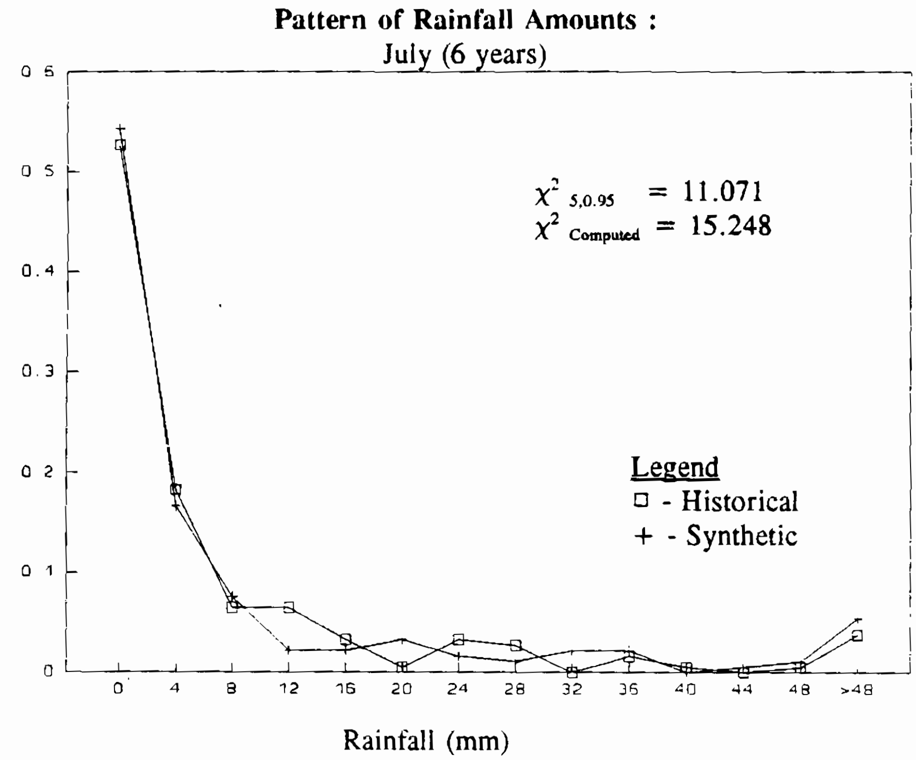
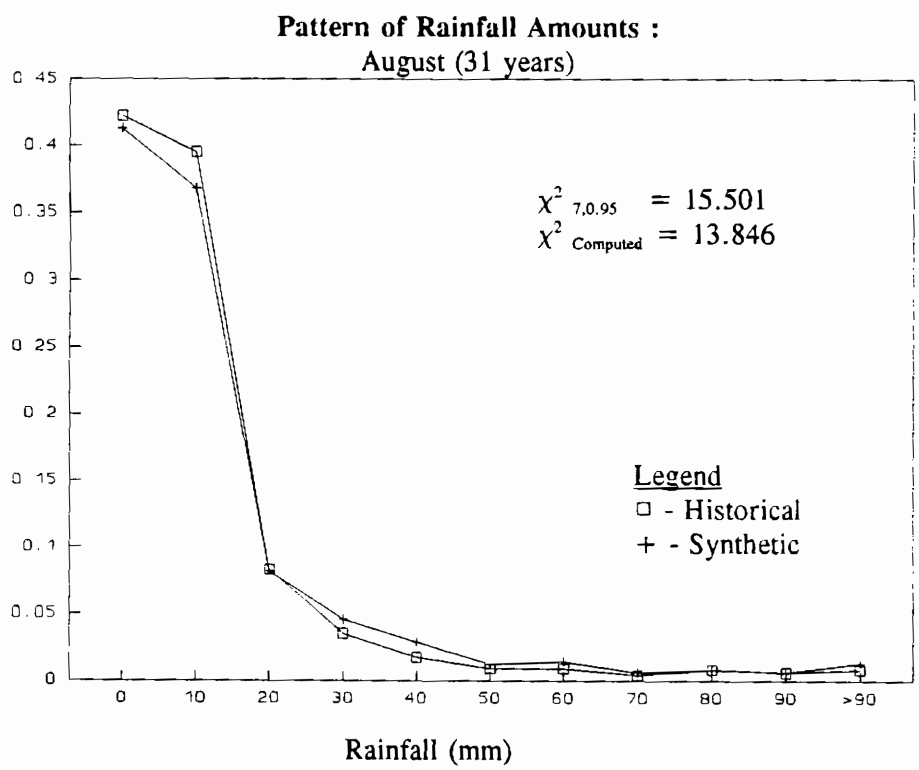


Figure 6



References

1. Sharan Girja and Kayastha Sandeep. **Lift irrigation in Panchmahal.** Centre for Management in Agriculture Monograph series no.144. IIMA. 1990.

2. Adamowski K. and Smith A.F. **Stochastic generation of rainfall**. Journal of Hydraulics Division. ASCE, vol.98. 1982.

3. Carey Daniel I. and Haan C.T. **Markov process for simulating daily point rainfall.** Journal of the Irrigation and Drainage Division. ASCE, vol.104. 1978.

4. Krishna Kumar M. **Development of rigid operating rules for a single-multipurpose reservoir by simulation**. M.E. Dissertation, L.D. College of Engineering, Ahmedabad. 1991.

5. Clarke R.T. **Mathematical models in hydrology**. FAO Irrigation and Drainage Paper 19. Rome. 1984.

6. Kottegoda N.T. **Stochastic water resources technology**. The McMillan Press Ltd. London. 1980.

7. Kottegoda N.T. **Stochastic five daily streamflow model**. Journal of the Hydraulics Division. ASCE, vol.98. 1972.

8. Lettenmaier Dennis P. **Synthetic streamflow forecast generation**. Journal of Hydraulic Engineering. ASCE, vol.110. 1984.

9. Mutreja K.N. **Applied hydrology**. Tata McGraw Hill Publishing Co. Ltd. New Delhi.

10 Payne Kip, Neuman W.R., and Kerri K.D. **Daily streamflow simulation**. Journal of Hydraulic Division. ASCE, vol.95. 1969.

11. Srikanthan Ratnasingham, and McMohan Thomas A. **Stochastic generation of monthly streamflows**. Journal of Hydraulic Division. ASCE, vol.108. 1982.

12. Fishman George S. **Concepts & methods in discrete event digital simulation**. John Wiley & Sons. New York. 1973.

13. U.S.Department of Commerce. **Handbook of mathematical functions with formulas, graphs & mathematical tables**. National Bureau of Standards. 1968.

14. Medhi J. **Stochastic Processes**. Wiley Eastern Limited. New Delhi. 1982.