DISAGGREGATION OF DAILY RAINFALL

by

Joanne S. Hershenhorn

A Thesis Submitted to the Faculty of the

DEPARTMENT OF HYDROLOGY AND WATER RESOURCES

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF SCIENCE
WITH A MAJOR IN HYDROLOGY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1984

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Joanne S. Hershenhorn

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

David A. Woolhiser
D. A. WOOLHISER
Adjunct Professor of
Hydrology and Water Resources

Sept 25, 1984
Date

ACKNOWLEDGMENTS

I would like to extend my deep gratitude to Dr. D. A. Woolhiser, my thesis advisor, for the constant encouragement, guidance and critical advice he provided throughout the course of this study.

I also wish to thank the entire staff associated with the Southwest Rangeland Watershed Research Center of the Agricultural Research Service, U.S. Department of Agriculture, for use of their data, publications and computer. Finally, a special word of appreciation to John J. Rocha for doing an excellent job of typing this manuscript.

Funding was provided through the USDA-ARS, Western Region, as part of the research program in Watershed Hydrology, in cooperation with the University of Arizona.

TABLE OF CONTENTS

TABLE OF CONTENTS--Continued

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS--Continued

LIST OF ILLUSTRATIONS--Continued

LIST OF SYMBOLS

English

D = the duration of an event

D_c = the duration of a complete event

D_c' = transformed event duration: $Dc' = \log_e(D_c)$ (or D' or d')

Dc_n = the duration of the n^{th} complete event

D_k = duration of the k^{th} event in a day

D_k = simulated duration of the k^{th} event in a day

D_n = duration of the n^{th} event

Dp = the duration of a partial event

i = year

j = a day in the year

K = total number of complete and partial events in a day

M = the rainfall occurrence process (wet or dry day)

$N(t)$ = counting process of the starting time of storm events

$N^*(t)$ = counting process of the ending time of storm events

N_{cj} = numbers of complete storms occurring on day j

Nc_{ij} = number of complete storms on day j, year i

Np_j = number of partial events occurring on day j

Np_{ij} = number of partial storms on day j, year i

R_1 = the ratio of the amount of the first event of the day to the daily total amount on days in which two events occur

R_2 = the ratio of the sum of the amounts of the second plus third events to the daily total on days in which three events occur

LIST OF SYMBOLS--Continued

R_3 = the ratio of the amount of the second event to the sum of the amounts of the second plus third events on days in which three events occur

R_4 = the ratio of the amounts of the sum of the third plus fourth events to the daily total amount on days in which four events occur.

R_5 = the ratio of the amount of the first event to the sum of the amounts of the first plus second events on days in which four events occur.

R_6 = the ratio of the amount of the third event to the sum of the amounts of the third plus fourth events on days in which four events occur

t = time, in cumulative hours, from some arbitrary time t = 0

Tc = rescaled starting time of complete events ($0 \leq Tc \leq 1$)

Tr = rescaled starting time of the r^{th} event in a day ($0 \leq Tr \leq 1$)

U(t) = process representing the time of day in military hours

Uc_ℓ = the starting time of the ℓ^{th} complete event in a day in military hours

U_k = starting time of the k^{th} event in a day

[x] = the integer part of the number x

X(t) = total amount of rainfall in the time interval (o,t)

X_k = amount of the k^{th} event in a day

$\hat{X}_k$ = simulated amount of the k^{th} storm in a day

X_n = total amount of rainfall in the n^{th} event

Y = the amount of a rainfall event

Yc = the amount of a complete rainfall event

Yc' (or y') = transformed event amount: $Yc' = Yc - .009$

Y" (or y") = transformed event amount: $Y'' = \log_e(Yc') = \log_e(Yc - .009)$

Yc_n = the amount of the n^{th} complete rainfall event

LIST OF SYMBOLS--Continued

Yp = the amount of rain in a partial event

Yp' = Transformed amount of a partial event: $Yp' = Yp - .009$

Zc = total daily rainfall amount due to complete events

Zc' = transformed daily rainfall amount for complete events: $Zc' = Zc - .009$

Zc_{ij} = total amount of rainfall contributed by complete events on day j of year i

Z_{ij} = total amount of rainfall on day j of year i

Zp_j = total rainfall amount contributed by partial events on day j

Zpr = the product of daily rainfall amounts (complete and partial events) on two consecutive rainy days

Zs = the sum of daily rainfall amounts (complete and partial events) on two consecutive rainy days

Greek

$\hat{\zeta}(t)$ = simulated rainfall process consisting of storm amounts and durations

$\eta(ij)$ = process representing the state of day j in year i; can be wet or dry

$\xi(t)$ = precipitation intensity at a point at time t

$\hat{\xi}(t)$ = simulated intensity process

τ_n = starting time, in cumulative hours, of the n^{th} rainfall event

τ_n^* = ending time, in cumulative hours, of the n^{th} event

ABSTRACT

This study represents an attempt to model the disaggregation of daily rainfall at a point into individual storms for the summer rainy season in southeast Arizona. The purpose of the model is to simulate the number of storms per day and the amount, duration and time of occurrence of each event while maintaining daily statistics.

Various storm ratios are introduced to describe the disaggregation of a daily amount into individual storm amounts. Marginal distributions are fit to the storm ratios and event starting times. On multiple-event days, starting times are obtained through the use of order statistics. Conditional distributions are fit to the number of storms per day, given a daily amount, and the duration of each event, given a storm amount.

The simulated data compare favorably with the observed data. Statistical tests indicate that the model adequately transforms daily rainfall into a sequence of individual storms.

CHAPTER 1

INTRODUCTION

Background

Current methods in the the development of water resources systems include the use of mathematical models to analyze possible future system behavior. Models can be used to simulate sequences reflecting many years with alternative design and management practices, enabling the evaluation of changes in these practices.

Although hydrology is only one component of agricultural production systems for urban drainage and runoff control facilities, and other hydraulic structures, peak rates and volumes of runoff are frequently controlling design factors. Infiltration, evaporation, surface runoff, soil erosion, and surface and subsurface transport of materials depend on rainfall as a driving input. In particular, greater emphasis is being placed on the use of physically-based infiltration models to estimate surface runoff, and these models are sensitive to the distribution of rainfall amounts in time periods as short as 5 minutes.

Breakpoint data, the observed time patterns of rainfall intensity, are obtained from weighing recording gages, and can be used to provide input to infiltration models. However, most National Weather Service (NWS) rain gages are the nonrecording type which measure only the total precipitation between observations, usually on a daily basis. When breakpoint data are available, it is time-consuming to convert them to a computer-compatible form. Rain charts must be coded, the information digitized, and programs written to transfer the data to usable files.

Lack of availability and the costliness associated with processing breakpoint data have prevented the widespread adoption of physically-based infiltration models by professionals in resource-management agencies. Although simpler empirical methods continue to be used, it is believed that an infiltration-based approach to hydrologic system modeling would prove superior.

One method of obtaining short-time period rainfall data is to disaggregate a daily rainfall amount into individual storms, and to further disaggregate the storms into intensity patterns. Much is known about the daily process. Daily rainfall data are readily available for many locations, and simulation procedures are well advanced. As spatial and seasonal variability can be accounted for in the daily process, if the disaggregation process developed at a point is relatively spatially invariant, it can be applied over areas experiencing similar weather, subject to minor modifications. While storm models based on aggregation techniques are available to simulate short-time period rainfall, it would be much more difficult to adjust the parameters of these models to maintain daily statistics at other locations.

Review of Literature

Most rainfall disaggregation modeling has focused on the disaggregation of annual to seasonal, seasonal to monthly, and monthly to daily rainfall amounts. Very little research has been carried out on the disaggregation of daily rainfall into representative storms.

Betson, Bales and Pratt (1980) described a scheme to disaggregate daily into hourly and hourly into 5-minute rainfall amounts. Transition probabilities were first determined from the data for PW/W and PD/D, the probabilities of occurrence of a wet hour given a wet hour, and of a dry hour given a dry hour. These transition probabilities can be adjusted for seasonality. A Weibull distribution was then used to independently generate hourly amounts, which were adjusted linearly to sum to the daily total. The same procedure, with different transition probabilities and Weibull parameters, was used to generate 5-minute rainfall amounts. The 5-minute amounts were added together to produce amounts for durations evenly divisible into an hour. Some shortcomings are evident if this model is to provide short-time rainfall sequences. Either a large number of transition probabilities must be estimated for hourly and 5-minute increment rainfall ocurrence, or the process is assumed to be stationary throughout the day. Also, the model does not explore the possibility of a dependency structure between duration and amount, or between the amount of one rainfall event to that of another, on a given day.

Most disaggregation models have been developed with streamflow processes in mind. Some of these techniques can be applied to the disaggregation of daily rainfall, although most cannot due to the high spatial and temporal variability and the intermittency associated with short-time period rainfall.

One model which has gained widespread use is Valencia and Schaake's linear disaggregation scheme (Valencia and Schaake, 1973). If a method were available to predict the number of storm events per day, and assuming the disaggregation could be represented by a linear process, Valencia and Schaake's model could be used to disaggregate a daily rainfall amount into representative storm amounts. However, their model does not address the intermittency associated with the daily rainfall process. It also does not account for the timing of rainfall bursts within the day, nor does it provide a mechanism for determining the duration associated with each amount.

As many of the subsequent efforts consist of modifications to Valencia and Schaake's basic model (Mejia and Rousselle, 1976; Tao and Delleur, 1976; Hoshi and Burges, 1978; Lane, 1979; Stedinger and Vogel, 1984), they will not be discussed.

Research Objectives

The objective of this research is to develop a parameter-efficient (parsimonious) model to simulate the amounts, durations, and times of occurrence of individual storm events within a day, given only the total amount of rainfall on that day and on the preceding and following days. The distributions of number of events (storms), storm amounts, storm durations and starting times obtained from the simulation model should closely resemble those from the data.

When used in conjunction with other models, such as Woolhiser and Osborn's (1984) intra-storm intensity model, the model developed herein will provide a description of short-time period rainfall.

This study is intended to serve as a pilot project to assess the feasibility of daily disaggregation models, and to develop basic techniques which can be applied for use in other locations.

CHAPTER 2

MATHEMATICAL MODEL

Introduction

In any attempt to model a stochastic process, one must choose among the various modeling approaches available: empirical, process-oriented or physically-based. The choice depends on the modeler's objectives and on the complexity of the processes to be described. The physical processes involved in the disaggregation of a daily rainfall amount are complex and not well understood at present. Also, because the model is being developed to provide better input to infiltration models, a physically-based description is not necessary.

An additional consideration in rainfall modeling is the high variability of rainfall processes, both in space and in time. One method of addressing this variability problem is to isolate the domain of variability, i.e., to describe a rainfall process as it varies in space at a constant time, or to describe a rainfall process as it varies in time for a constant space. The model developed herein is based on phenomenological and empirical considerations as applied to the point rainfall process.

Description of Rainfall Process

Let $\xi(t)$ denote the continuous process of precipitation intensity at a point for a given location. $\xi(t)$ will be equal to or greater than zero, and will take on positive values over random time intervals. A possible sample function of this process, in year i, is shown in Figure 2.1.

If the starting time of the n^{th} storm is denoted by τ_n, and the ending time of the n^{th} storm by τ_n^*, then we can define two counting processes:

$$N(t) = \sup\{n: \tau_n \leq t\}$$

$$N^*(t) = \sup\{n: \tau_n^* \leq t\}$$

Possible sample functions of these counting processes are shown in Figure 2.2.

Let $X(t)$ represent the total amount of rainfall in the time interval (o,t). By definition, then

$$X(t) = \int_o^t \xi(s)ds.$$

The total amount of rainfall in the n^{th} storm can be written as

$$X_n = \int_{\tau_n}^{\tau_n^*} \xi(s)ds$$

and the duration is $D_n = \tau_n^* - \tau_n$. A possible sample function of this process is shown in Figure 2.3.

If a complete event is defined as a rainfall event which begins and ends on the same day, then the number of complete storms occurring on day j is:

$$Nc_j = N^*\{24j\} - N\{24(j-1)\};\ j = 1, 2, \cdots$$

$$Nc_j = 0, 1, 2, \cdots$$

where the counting processes $N^*(t)$ and $N(t)$ are defined over a time scale of cumulative hours from some arbitrary time $t = 0$.

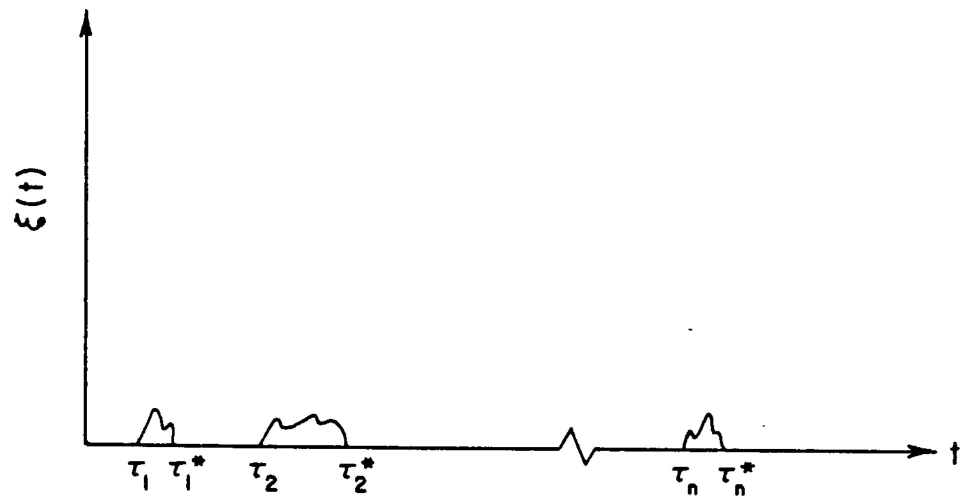


Figure 2.1. Possible sample function of the intensity process

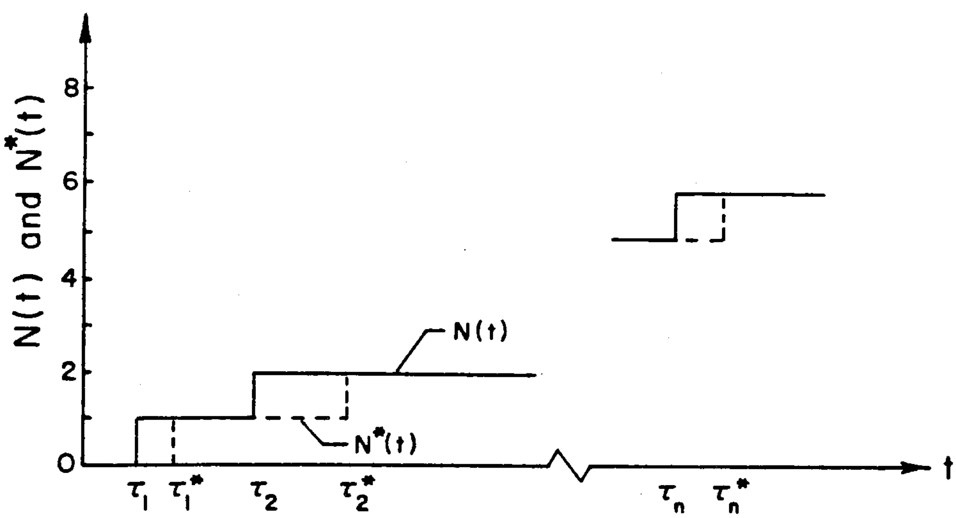


Figure 2.2. Possible sample functions of the counting processes

If a partial storm is defined as a rainfall event which begins on one day and ends on the next, then the number of partial storms occurring on day j is:

$$Np_j = N(24j) - N^*\{24(j-1)\} - Nc_j \; ; \; j = 1, 2, \cdots$$
$$Np_j = 0, 1, 2$$

If t is in cumulative hours from some arbitrary time $t = 0$, and i is the number of years from $t = 0$ $(i = 1, 2, \cdots)$, then j refers to the day in the year:

$$j = [\{t - (i-1)\, 365 \cdot 24\}/24] + 1$$
$$j = 1, 2, 3, \cdots$$

where [x] represents the integer part of the number x. Note that, by using this notation, the extra day in leap years is ignored.

Finally, to represent phenomena occurring within the day, define the variable U(t) such that:

$$U(t) = t - \{[t/24] \cdot 24\}$$

This function is shown in Figure 2.4. Note that for each time of beginning, τ_i, there is a corresponding $U(\tau_i)$, where $0 \leq U(\tau_i) \leq 24$.

The rainfall process is assumed to be stationary with possible cyclic variations on a seasonal basis. That is, the sample functions of the rainfall process for each year are assumed to be repeated samples of the same underlying process.

Let $\eta(ij) = 0$ if there is no precipitation on day j in year i, and $\eta(ij) = 1$ if there is precipitation. Note that if

$$\eta(ij) = 0, \text{ then } Nc_{ij} = Np_{ij} = 0$$
$$\eta(ij) = 1, \text{ then } Nc_{ij} + Np_{ij} \geq 1.$$

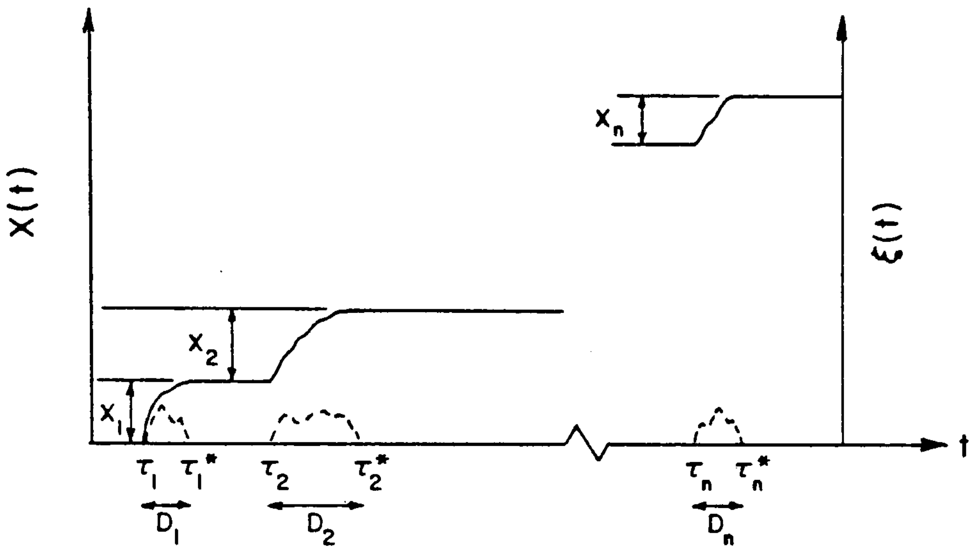


Figure 2.3. Possible sample function of the process X(t) (total rainfall amount)

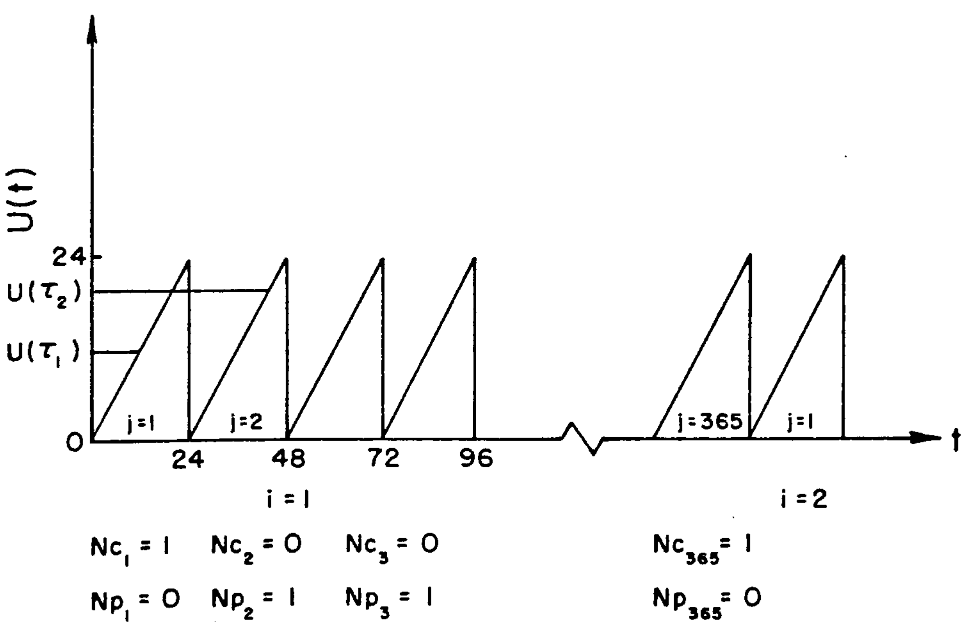


Figure 2.4. The function U(t)

The Disaggregation Process

Let Z_{ij} represent the total rainfall on day j of year i:

$$Z_{ij} = \int_{24(j-1)}^{24j} \xi(s)ds = X\{24j\} - X\{24(j-1)\}.$$

Thus, Z_{ij} for i = 1, 2, 3, • • •; j = 1, • • •, 365 represents a sequence of daily rainfall amounts.

Associated with each day that rain is observed, i.e., $\eta(ij) = 1$, are the random variables Z_{ij}, Nc_{ij}, Np_{ij}, U_k; k = 1, 2, • • •, $(Nc_{ij}, + Np_{ij})$, and the amount X_k and duration D_k of each complete or partial event during the day. A complete disaggregation model would require a description of the joint distribution of these random variables and an algorithm to generate them using Monte Carlo techniques, given only a simulated or observed sequence of daily rainfall amounts. The simulated rainfall process $\hat{\zeta}(t)$ will then consist of a sequence of pulses taking on positive values over random time intervals, as shown in Figure 2.5. The intensity of each pulse is equal to $\hat{X}_k/\hat{D}_k$, and the daily process, Z_{ij}, will preserve the stochastic properties of the real or simulated daily rainfall data, because $\sum \hat{X}_k = Z$.

The final step in the process is to disaggregate the storms into intensity patterns leading to the process $\hat{\xi}(t)$, as shown in Figure 2.6. An intra-storm disaggregation model that would transform the process $\hat{\zeta}(t)$ into $\hat{\xi}(t)$ for thunderstorm rainfall has been developed by Woolhiser and Osborn (1984). Bras and Rodriguez-Iturbe (1976) have also developed a procedure that appears to be more appropriate for longer duration storms. However, this transformation is not a part of the objectives of this research, and will not be considered further.

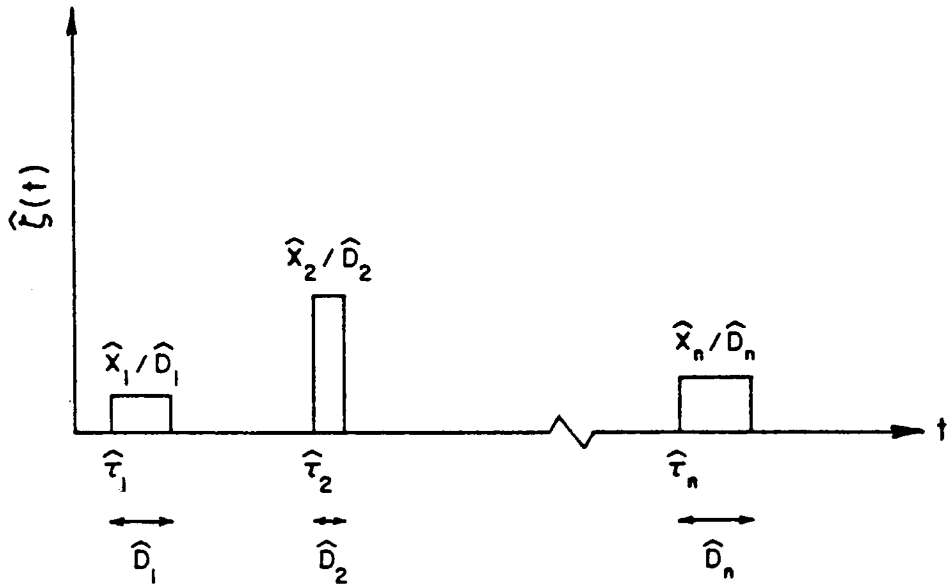


Figure 2.5. Simulated rainfall process

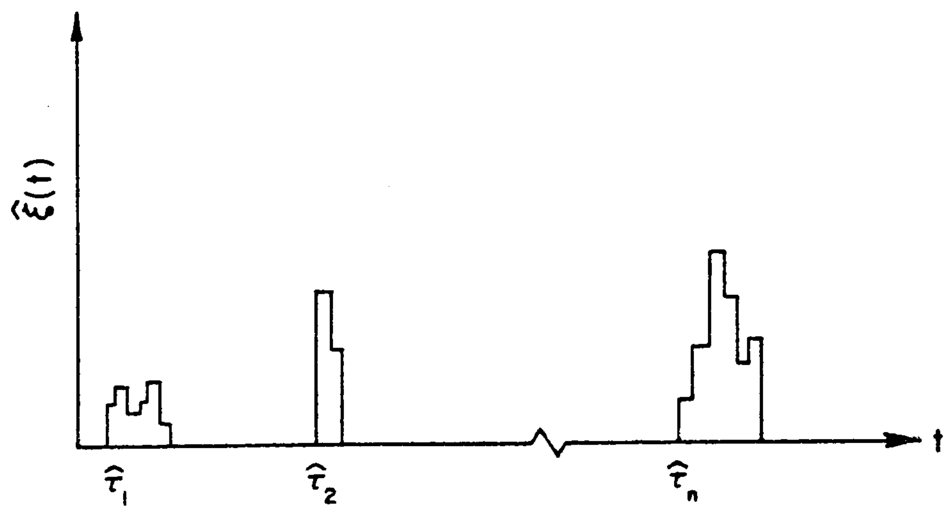


Figure 2.6. Simulated intra-storm intensity process

The overall objective of this research requires specifying the joint distribution function of a random number of random variables, some of which are continuous and some discrete. To make the problem tractable, certain simplifying assumptions will be required. Because this is the first attempt to disaggregate daily rainfall in this manner, there are no precedents, and an exploratory data analysis is required to identify possible simplifications. The simplifying assumptions made in this study, and the reasons for making them, are presented in Chapter 4.

CHAPTER 3

DESCRIPTION OF DATA

Location

Twenty-three years of summer breakpoint data, for the months of July and August, 1954-1976, were obtained from rain gage #5 on the Walnut Gulch Experimental Watershed, operated by the United States Department of Agriculture, Agricultural Research Service (USDA-ARS). This watershed, located in the Upper San Pedro River Valley of southeast Arizona, has an area of 58 mi^2 (150 km^2), and includes the city of Tombstone (see Figure 3.1). Walnut Gulch is an ephemeral stream which flows only after rainfall occurs on the watershed. It drains to the west, into the north-flowing San Pedro River. Rain gage #5 is located in the upper west portion of the watershed, approximately 1.7 miles (2.74 km) due east of the watershed outlet.

Climate

The climate of the Tombstone area, which is representative of southeast Arizona, is classified as semiarid, with an average annual precipitation of 11.5 inches (292.1 mm). Two rainy seasons are apparent, winter and summer. Approximately 65 percent of the annual precipitation, and virtually all of the runoff, occurs during the summer season (July through mid-September).

Summer thunderstorm rainfall is characterized by its short

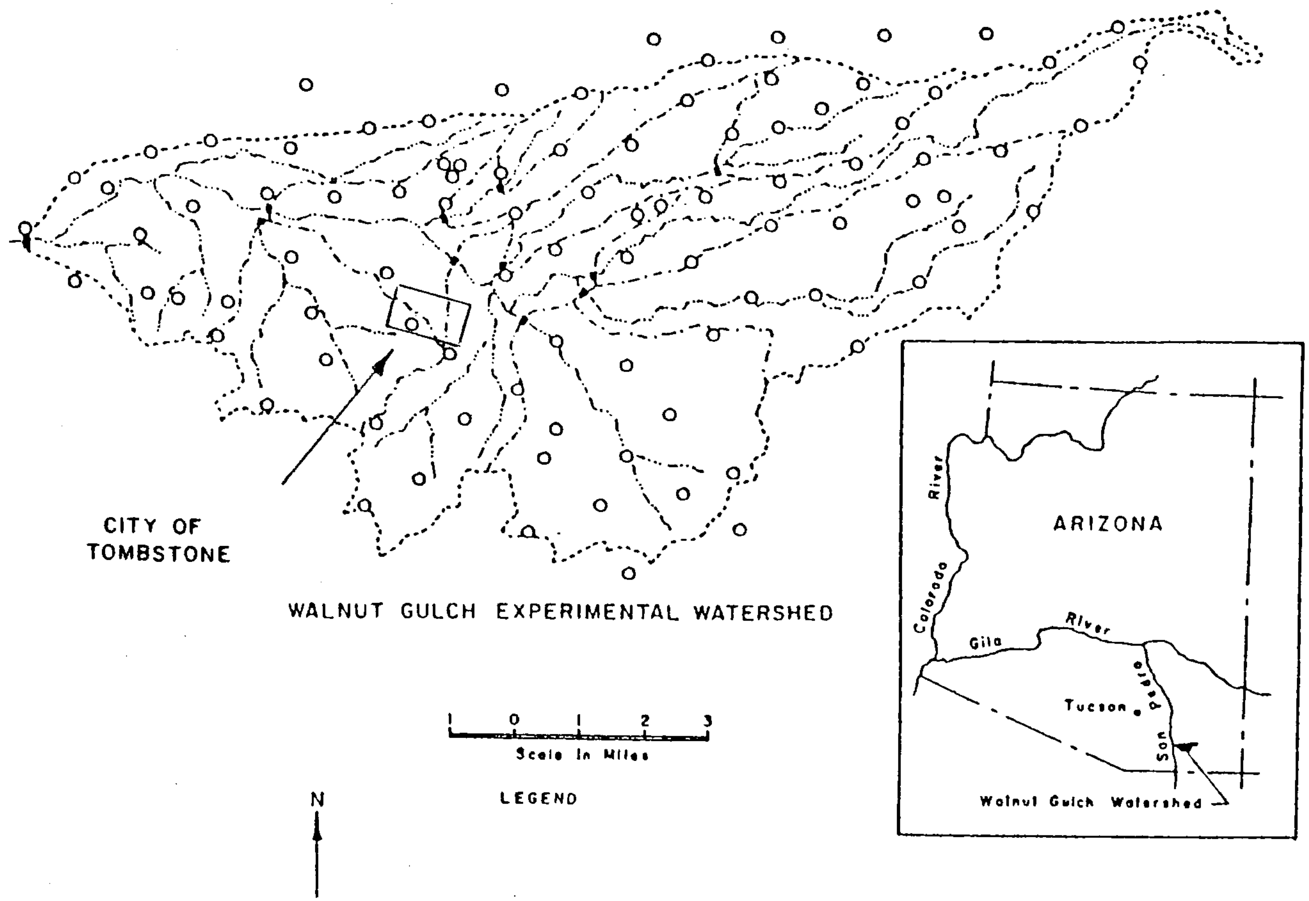


Figure 3.1. Location of the Walnut Gulch Watershed

duration and high intensity, a high spatial variability, and a strong diurnal pattern, with a maximum occurring in the late afternoon and early evening.

Two schools of thought have been formulated as to the source of moisture for summer rainfall. The first, widely accepted up through the early 1970's, focuses on the movement of moisture from the Gulf of Mexico into Arizona via the westward expansion of the Bermuda High (Bryson, 1957). The second school of thought, which gained popularity after weather satellite information became available, suggests the Pacific Ocean as a significant source of moisture (Hales, 1973). Both the Pacific Ocean and the Gulf of Mexico are now recognized as important sources of moisture for summer convective storms in southeast Arizona.

During the winter, large-scale cyclonic storms embedded in the jet stream bring rain to the region. The jet stream, a fast-moving current in the prevailing westerlies, typically (in winter) lies at 35° north latitude over Arizona. Daily departures from the mean bring the jet stream over the southeast portion of the state. Because most of the annual precipitation and runoff occur during the summer season, little research has been done on winter precipitation.

Instruments and Procedures

Rain gage #5 is a weighing, recording gage which produces a record of accumulated rainfall versus time. Every inch (25.4 mm) on the depth scale of the rain gage chart measures .333 inch (8.5 mm) of precipitation. The resolution is .006 inch (.15 mm), with an accuracy of .012 inch (.30 mm). Each inch on the time scale equals 125.22 minutes.

The resolution is 2.5 minutes, with an accuracy of 5.0 minutes (Chery and Kagan, 1975). The charts are changed once a week, and the time is set then, according to a technician's watch. Each gage is calibrated once a year, before the summer rainy season, for weighing accuracy (823 ± .01 grams equal 1.0 inch (25.4 mm) of water).

The charts are coded at the Southwest Rangeland Watershed Research Center, (SWRWRC) in Tucson, Arizona. The coder notes the date and beginning time of each event, then breaks the pen traces into line segments which reflect intensity patterns within the storm. Time corrections are distributed linearly throughout the recorded period. If a record is missing, estimates of beginning time, depth and duration are made from an isohyetal plot of the total rainfall measured by surrounding gages. After coding, the charts are digitized. The data are then checked for errors and put on a master file for further analysis.

Many factors affect the accuracy of raingage catch, the most important being wind. Other factors include evaporation, adhesion, and damage to the gage. Wind effects are due primarily to an increase in pressure on the windward side of the gage and a corresponding decrease in pressure and increase in acceleration over the top of the gage, with the formation of eddy currents over and within the orifice. Brooks (1938) reported that the catch of an unshielded 8-inch gage is about 5 percent less than the true precipitation. Two other consequences of the rain gage accuracy and processing methods include a resolution of .01 inch (.254 mm) and the nonmeasurement of many small events ($\leq$.05 inch, or 1.27 mm).

CHAPTER FOUR

DATA ANALYSES

General Description of Historical Data

A storm is defined as any period in which the total rainfall is $\geq$.01 inch, or .254 mm, which does not contain an intervening period of zero intensity exceeding 10 minutes in duration. Any period of greater than 10 minutes, in which no increase in precipitation is measured, signifies an event ending (see Figure 4.1).

The data set consisted of 695 storms recorded at rain gage #5 in July and August, during the years 1954-1976, on the Walnut Gulch Watershed. Of these, 659 (94.82%) were complete events which begin and end on the same day, and 36 (5.18%) were partial events which begin on one day and end on the next (see Figure 4.2). A day is defined as that 24-hour period beginning at midnight (12 AM). About 1.5% of the data were estimated.

It is possible, for certain climatic regions, to define a day with a different beginning time, so that the probability of occurrence of a partial event is so low as to be considered negligible. Only complete events need be defined then. However, the data available for this study also require a method of analysis for partial events. Some general statistics are presented in Table 4.1.

The two groups were analyzed separately for two reasons:

(1) The statistics in Table 4.1 are different enough for

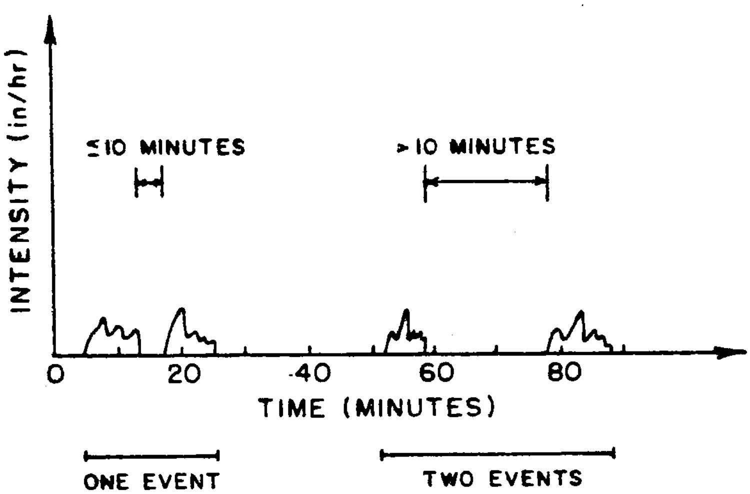


Figure 4.1. Definition of a storm

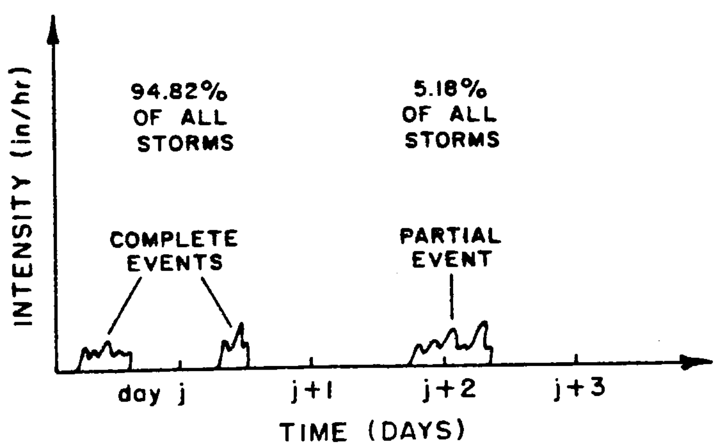


Figure 4.2. Definition of complete and partial events

Table 4.1. Statistics on duration and amount for summer thunderstorms on the Walnut Gulch Watershed (23 years of data)

	Type of event			
	Complete		Partial	
Number	659		36	
Amount	in	(mm)	in	(mm)
Mean	.199	(5.06)	.442	(11.24)
Standard deviation	.307	(7.80)	.413	(10.48)
Range	.01-2.99	(.25-75.94)	.02-1.36	(.50-34.54)
Duration (min)				
Mean	66.38		226.36	
Standard deviation	71.26		157.42	
Range	1.99-510.0		70.02-668.99	

complete and partial events to warrant a separate analysis.

(2) The modeling of partial events was anticipated to be more complex than the modeling of complete events, and it was hoped that some of the techniques developed for complete events could be applied to the modeling of partial events.

Model of Complete Events

In hydrologic modeling, abstraction or simplification is always necessary because of the complexities of hydrologic processes. A major assumption of this study is that the disaggregation processes described herein are stationary throughout the season of interest (July - August).

Analyses of 27 years of daily data indicated a periodic structure for daily rainfall occurrence and amount on the Walnut Gulch Watershed (Woolhiser and Econopouly, 1983). The parameters investigated were PW/W, the probability of a wet day following a wet day, PW/D, the probability of a wet day following a dry day, and mean daily amount. These parameters are shown in Table 4.2.

As the processes involved in disaggregating daily rainfall are derived from a daily amount, it is likely that these processes will also exhibit trends or periodicities. However, when examined over the entire year, the periodicity exhibited during the summer season becomes less significant, and the consequences of violating this assumption of stationarity are minimized (see Figure 4.3).

The 659 complete events were distributed over 450 rainy days. The majority occurred as a single storm in a day, and the largest number of occurrences observed in a day was six. A breakdown of the complete

Table 4.2. The parameters PW/W, PW/D, and mean daily rainfall amount for 14-day periods on the Walnut Gulch Watershed (27 years of data)

14-day period	PW/W	PW/D	$\bar{Z}$ (in)	$\bar{Z}$ (mm)
Jun 21 - Jul 4	.2381	.1042	.1867	4.742
Jul 5 - Jul 18	.5574	.2500	.2313	5.875
Jul 19 - Aug 1	.5751	.4270	.2828	7.183
Aug 2 - Aug 15	.3719	.2879	.3075	7.811
Aug 16 - Aug 29	.3750	.2519	.2524	6.411
Aug 30 - Sep 12	.4483	.1649	.2771	7.038

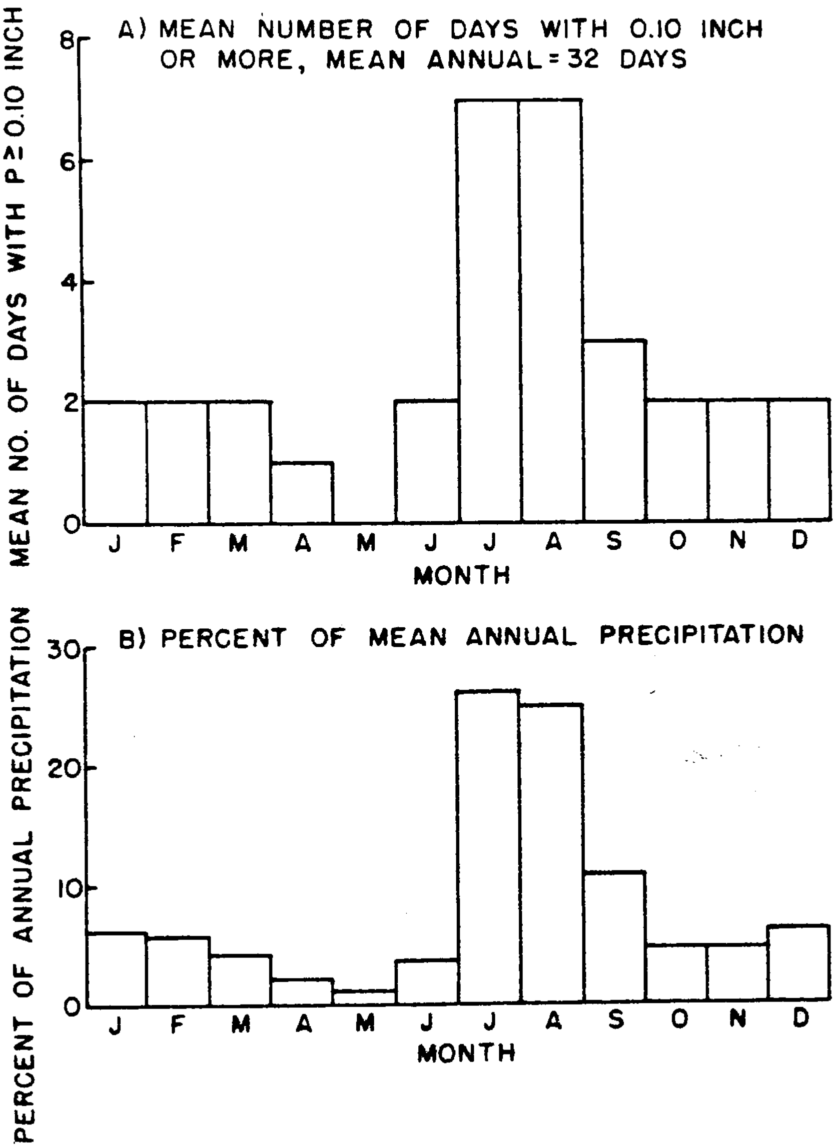


Figure 4.3. Monthly distribution of precipitation for Tombstone, Arizona (from Green and Sellers, Arizona Climate, Univ. of Arizona Press, 1964)

events, as well as the mean amount and duration of each class, is given in Table 4.3.

Joint Distribution of Number of Events per Day and Daily Rainfall Amount

The joint distribution function of the number of events per day, Nc_{ij}, and the daily rainfall amount, Zc_{ij}, can be written as a product of a conditional distribution and a marginal distribution:

$$H_{Nc_{ij},\, Zc_{ij}}(n,z) = G_{Nc_{ij}|Zc_{ij}}(n|z)F_{Zc_{ij}}(z)$$

From the assumption of stationarity from year to year, and within the months of July and August, the distribution function will not depend on the year, i, or the day, j. Therefore, these subscripts will be omitted throughout the remainder of the chapter. Also, as the lower threshold of observation is .01 inch (.254 mm), distributions were fit to the transformed variable Zc', where Zc' = Zc -.009. The expression is now:

$$H_{Nc,Zc'}(n,z) = G_{Nc|Zc'}(n|z)F_{Zc'}(z) \qquad (4.1)$$

or written out in longer form:

$$P(Nc \leq n,\ Zc' \leq z) = P(Nc \leq n|Zc' \leq z)P(Zc' \leq z)$$

with $0 \leq P \leq 1$.

If expressions can be found for the marginal distribution of daily amount, $F_{Zc}'(z)$, and the conditional distribution of number of events per day given daily amount, $G_{Nc|Zc}'(n|z)$, then the joint distribution is specified.

Several distributions have been fit to daily rainfall amounts. Das (1955) and Kotz and Neumann (1963) used forms of the gamma (Pearson Type III). Stidd (1953) and Beals (1954) found that raising the daily amounts to a power, 1/3 and 1/4, respectively, transformed them to a

Table 4.3. Mean amount and duration for each class of complete events

Number of events per day	Number of occurrences	Mean amount in	(mm)	Mean Duration min
1	308	.2107	(5.352)	69.874
2	96	.2293	(5.824)	66.219
3	30	.2024	(5.141)	64.097
4	12	.1052	(2.672)	37.420
5	3	.1200	(3.048)	120.273
6	1	.1283	(3.259)	12.510

normal distribution. Brakensiek (1958) fit a lognormal to daily amounts. Smith and Schreiber (1974) and Woolhiser and Roldân (1982) found a mixed exponential distribution appropriate.

Mielke and Johnson (1974) found that, for some hydrologic data, forms of the beta distribution rendered fits which were comparable to those obtained using the gamma distribution. The advantage of using these so-called beta-kappa distributions is that, unlike the gamma, the CDF's (cumulative distribution functions) can be written in closed form, making them easier to work with. Herein, the 3-parameter beta-kappa distribution will be referred to as the kappa (3).

Based on previous research and some exploratory analyses, the following five choices for the marginal distribution of daily rainfall amount were compared: exponential, mixed exponential, lognormal, kappa (3), and Weibull distributions. Table 4.4 lists some properties of these distributions.

Numerical optimization techniques, consisting of a mixed application of the Rosenbrock method, Palmer version, and the penalty function method (Palmer, 1969), were used to obtain maximum likelihood parameter estimates from the data when analytic expressions were not available. For distributions with the same number of parameters, the maximum log likelihood function value can be used to determine the best choice. A description of the log likelihood function is given in Appendix A. For distributions with a different number of parameters, the minimum Akaike Information Criterion (Akaike, 1974) can be used to choose between competing distributions. Calculation of the AIC is as follows:

Table 4.4. Properties of distributions tested for the marginal distribution of daily rainfall amount

Distribution	PDF	Number of parameters	Mean	Variance
Exponential	$f(z) = \frac{1}{\lambda} \exp[\frac{-z}{\lambda}]$	1	λ	λ^2
Lognormal ($Z'=\ln Z$)	$f(z') = \frac{1}{\sigma_{z'}\sqrt{2\pi}} \exp[-\frac{1}{2}(\frac{z'-m_{z'}}{\sigma_{z'}})^2]$	2	$m_{z'}$	$\sigma^2_{z'}$
Weibull	$f(z) = \frac{\gamma}{\alpha}(\frac{z}{\alpha})^{\gamma-1} \exp[(\frac{-z}{\alpha})^{\gamma}]$	2	$\alpha\Gamma(1+\gamma^{-1})$	$\alpha^2[\Gamma(1+2\alpha^{-1})-\{\Gamma(1+\gamma^{-1})\}^2]$
Mixed exponential	$f(z) = \alpha\frac{\exp[\frac{-z}{\beta}]}{\beta} + (1-\alpha)\frac{\exp[\frac{-z}{\gamma}]}{\gamma}$	3	$\alpha\beta + (1-\alpha)\gamma$	$\alpha\beta^2 + (1-\alpha)\gamma^2$
Kappa (3)	$f(z) = \frac{\alpha\theta}{\beta}(\frac{z}{\beta})^{\alpha\theta-1}[1+(\frac{z}{\beta})^{\theta}]^{-(\alpha+1)}$	3	$(\alpha\beta)$*$B(\alpha+\frac{1}{\theta},1-\frac{1}{\theta})$	$\alpha\beta^2$*$B(\alpha+\frac{2}{\theta},1-\frac{2}{\theta})$

*$B(\delta,\varepsilon)=[\Gamma(\delta)\Gamma(\varepsilon)/\Gamma(\delta+\varepsilon)]$ where Γ is the gamma function.

$$AIC = (-2)\log_e(\text{max likelihood function value}) + 2m$$

where m is the number of adjusted parameters. It should be noted that the data used consisted of daily rainfall amounts for days in which only complete events occurred.

The results shown in Table 4.5 indicate that the two-parameter Weibull distribution gives the best fit to daily rainfall amount for the given data. A chi-squared goodness-of-fit test indicates that the hypothesis of a Weibull distribution for the marginal distribution of daily rainfall amount cannot be rejected at the 95% level, as the statistic of 10.4805 obtained is less than the critical value $\chi^2_{.05,5} = 11.0758$. The distribution is written as: $F_{Zc'}(z)$

$$F_{Zc'}(z) = 1 - \exp[-(\tfrac{z}{\alpha})^{\gamma}] \tag{4.2}$$

$$\alpha = .2399 \; ; \; \gamma = .7364 \; ; \; z \geq .001$$

The theoretical distribution is compared with the empirical distribution function in Fig. 4.4. All empirical CDF's in this study are calculated as follows: The observed values are arranged by descending magnitude and assigned a rank, m, so that m = 1 for the item of largest magnitude. For N items, the exceedence probability is $P(Z > z = \frac{m}{N+1})$. The cumulative distribution function value, $P(Z \leq z) = 1 - P(Z > z) = 1 - \frac{m}{N+1}$. It now remains to determine an expression for $G_{Nc|Zc'}(n|z)$, the conditional distribution of number of events, given a daily amount.

The first distribution tested was the truncated Poisson (truncated at one, as we are not interested in the occurrence of zero events). Many have used a Poisson process to describe event arrival time. Todorovic and Yevjevich (1969) used a nonhomogeneous Poisson for the number

Table 4.5. Results of test of best fit for the marginal distribution of daily amount

Distribution	Number of parameters	Value of $\log_e$ likelihood function	AIC
Exponential	1	95.0009	-188.0018
Lognormal	2	117.6291	-231.2583
Weibull	2	134.1548**	-264.3096**
Mixed exponential	3	133.5140	-261.0281
Kappa(3)	3	129.5843	-253.2583

**Best values obtained.

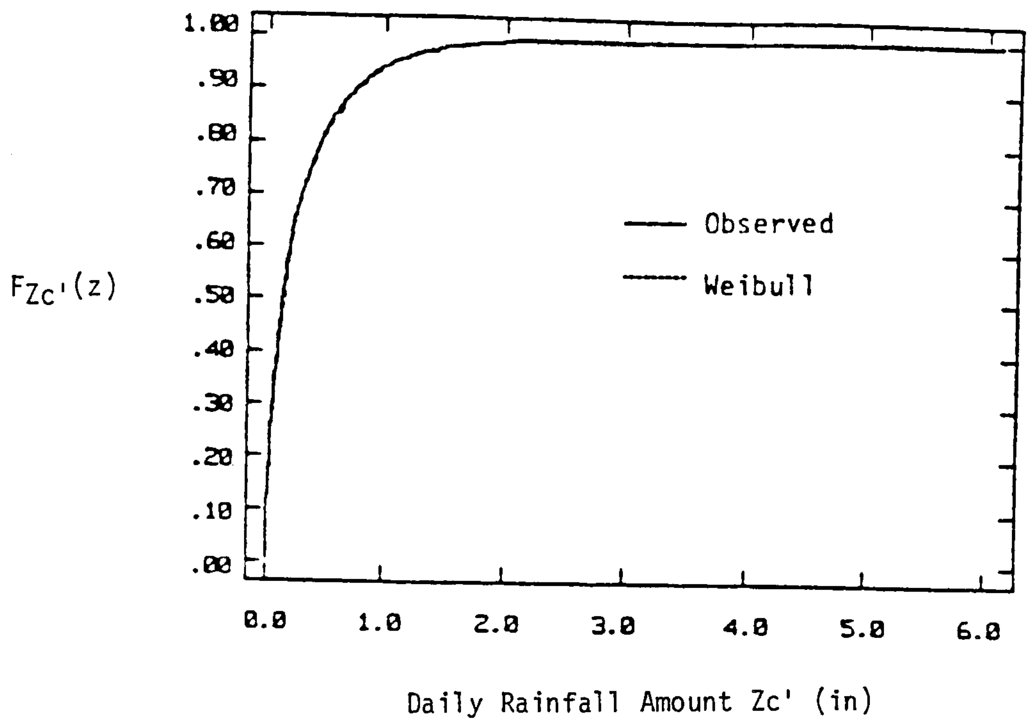


Figure 4.4. Observed and fitted distributions for daily rainfall amount

of storms per period. Fogel and Duckstein (1969) used a Poisson to describe the number of events per year over a 20 square-mile area. Mills (1982) also found the Poisson to fit event occurrence.

The expression for the truncated Poisson is:

$$P(Nc = n) = \frac{e^{-\mu}\mu^{n}}{n!(1-e^{-\mu})} \quad ; \quad n = 1, 2, 3, \cdots$$

Two approaches were taken to determine the parameter μ, the expected number of events per day: to let μ be a linear function of Z, $\mu = A + BZ$, and to let μ be a curvilinear function of Z, $\mu = A + BZ + CZ^2$. The conditional probability mass function of number of events per day, given daily amount, is then written as:

$$P(Nc = n|Zc' = z) = \frac{e^{-(A+Bz+Cz^2)}(A+Bz+Cz^2)^n}{n!(1-e^{-(A+Bz+Cz^2)})} ; \; n = 1, 2, 3, \cdots \qquad (4.3)$$

The joint distribution function is:

$$H_{Nc,Zc'}(n,z) = P\{Nc \leq n, \; Zc' \leq z\} = P(Nc \leq n|Zc' \leq z) \; P(Zc' \leq z)$$

$$= \sum_{i=1}^{n} \int_{o}^{z} \frac{\gamma}{\alpha}\left(\frac{z}{\alpha}\right)^{\gamma-1} \exp\left[\left(\frac{-z}{\alpha}\right)^{\gamma}\right] \frac{\exp\{-(A+Bz+Cz^2)\}(A+Bz+Cz^2)^{i}dz}{i! \; [1 - \exp\{-(A+Bz+Cz^2)\}]}$$

The equivalent of the joint density function can be written as:

$$f_{Nc,Zc'}(n,z) = \lim_{\Delta z \to o} \left[\frac{H_{Nc,Zc'}(n,z) - H_{Nc,Zc'}(n-1, \; z-\Delta z)}{\Delta z}\right]$$

$$= \left(\frac{\gamma}{\alpha}\right)\left(\frac{z}{\alpha}\right)^{\gamma-1} \exp\left[-\left(\frac{z}{\alpha}\right)^{\gamma}\right] \frac{\exp\{-(A+Bz+Cz^2)(A+Bz+Cz^2)^n}{n! \; [1 - \exp\{-(A+Bz+Cz^2\}]} \qquad (4.4)$$

Starting values of the parameters A, B and C were first estimated by the least squares method, then maximum likelihood estimates were obtained using the modified Rosenbrock numerical optimization techniques mentioned earlier. Herein, the use of numerical optimization procedures refers to this method.

As a first step in testing the Poisson, the data were divided into four classes, 1, 2, 3, and $\geq$ 4 events per day. An expression can then be obtained for the conditional distribution of daily amount, given that n events occurred, or $F_{Zc'|Nc}(z|N_c = n)$ for n = 1, 2, 3, $\geq$ 4, as follows:

$$P\{Zc' \leq z | Nc = n\} = P\{Zc' \leq z, Nc = n\}/P\{Nc = n\}$$

$$= \frac{H_{Nc,Zc'}(n,z) - H_{Nc}(n-1,z)}{H_{Nc,Zc'}(n,\infty) - H_{Nc,Zc'}(n-1,\infty)} = \frac{\int_0^z f_{Nc,Zc'}(n,z)\,dz}{\int_0^\infty f_{Nc,Zc'}(n,z)\,dz} \qquad (4.5)$$

The expression under the integral sign in the numerator of Equation (4.5) is the joint density of the number of events per day and daily amount, and can be expressed as in Equation (4.4) using the optimized parameter values obtained. The denominator is the marginal distribution of number of events per day, and can be evaluated using the same expression, but integrated (numerically) over all $z(0,\infty)$. If the hypothesis of a truncated Poisson is reasonable, then the conditional distribution $F_{Zc'|Nc}(z|n)$ given by Equation (4.5) and evaluated at n = 1, 2, 3, $\geq$ 4, should give results consistent with observed values.

An analysis showed that this model tends to underpredict the occurrence of low rainfall amounts for all classes of n = 1, 2, 3, and $\geq$ 4 events per day.

The next distribution tested was the geometric. The form is:

$$P(Nc = n) = p(1-p)^{n-1},\ n = 1, 2, 3, \cdots$$

and the expected value function $E[N_c]$ is equal to $\frac{1}{p}$. The initial hypothesis, that $\frac{1}{p} = A + BZ$, was later modified to $\frac{1}{p} = A \exp(BZ) + (1-A)$. The conditional probability mass function is written as:

$$P\{Nc = n|Zc' = z\} = \left[\frac{1}{A\ \exp(Bz) + (1-A)}\right]\left[1-\left\{\frac{1}{A\ \exp(Bz) + (1-A)}\right\}\right]^{n-1} \quad (4.6)$$

for n = 1, 2, 3, and $\geq$ 4 events per day. The same procedures that were used for the truncated Poisson were used to estimate parameters and to evaluate the conditional distribution of daily amount given that n events occurred.

Results of a simulation again indicated a tendency to overpredict the occurrence of one event per day and underpredict two per day. The data were then broken down into eight classes of daily amount. Using the midpoint of each class interval, and the parameter estimates already obtained (for the entire set), $P\{N_C = n|Z_C' = z \text{ midpoint}\}$ was determined for each class of n = 1, 2, 3, and $\geq$ 4 events per day. The geometric did very well in simulating the number of events per day for the classes of lower rainfall amounts; however, it consistently overpredicted the occurrence of one and underpredicted the occurrence of two events per day in the classes of higher rainfall amounts.

Apparently, the process is too complex to be adequately described by a 2-parameter distribution. However, as the geometric works well in the range of lower amounts, it is desirable to find a distribution that approaches the geometric at the lower limit. The shifted negative binomial distribution (SNBD) possesses this property (Buishand, 1977). The probability mass function is written as:

$$P(Nc = n) = \binom{n+r-2}{n-1} p^r(1-p)^{n-1} \ ; \ n = 1, 2, 3, \cdots \quad (4.7)$$

When r = 1, this reduces to the geometric distribution.

The method used to determine the parameters p and r is as follows: fix the upper and lower values, and allow p and r to vary with daily rainfall amount between these limits. The lower limits are set so that, given the minimum possible daily amount (.001 inch or .025 mm), the probability of getting one event per day is a certainty, i.e., $P\{N_c = 1 | Zc' = z_{min}) = 1$. This occurs when p = r = 1. To determine the upper limits, p_{min} and r_{max}, p and r are estimated over the two classes of highest daily rainfall amount using maximum likelihood techniques. Recall that the geometric distribution did poorly in the classes of higher daily rainfall amount. It was found that $p_{min} = .7228$ and $r_{max} = 2.3097$. A sketch is shown in Fig. 4.5.

An investigation showed that exponential curves proved to be adequate. The forms of p and r are now hypothesized as:

$$p = p_{min} + (1 - p_{min})e^{-B_1 Zc'}$$

$$r = r_{max} - (r_{max} - 1)e^{-B_2 Zc'}$$

Note that as Zc' approaches zero, both p and r approach 1.

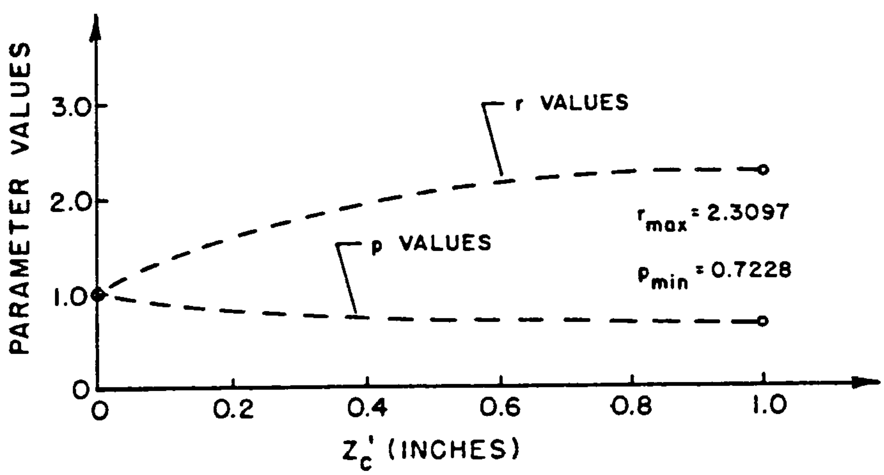


Figure 4.5. Variation in parameter values with daily amounts for shifted negative binomial distribution (SNBD)

The parameters B_1 and B_2 were estimated using numerical maximum likelihood techniques, and found to be 5.7939 and 9.5902, respectively. The final forms for p and r are:

$$\begin{aligned} p &= .7228 + .2772e^{-5.7939Zc'} \\ r &= 2.3097 - 1.3097e^{-9.5902Zc'} \end{aligned} \qquad (4.8)$$

The conditional probability is written as:

$$P\{Nc = n \mid Zc' = z\} = \binom{n+r-2}{n-1} p^r(1-p)^{n-1} \;; \quad n = 1, 2, 3, \cdots \qquad (4.9)$$

with the values for p and r as functions of Zc', as given in Equation (4.8).

The marginal probability mass function of N_C, obtained from a simulation, compares much better with observed values than the previous distributions tested (see Fig. 4.6). Although there is still a tendency to overpredict the occurrence of one event per day and underpredict the occurrence of two, the tendency is reduced. Given the complexities associated with simulating the number of events from a daily amount, this distribution was found to be adequate for the purposes of this model.

For a more detailed discussion on the method of hypothesis testing for this conditional probability distribution, see Appendix B.

Distributions Describing the Disaggregation of a Daily Rainfall Amount into N Amounts

Given a daily rainfall amount, Zc, and the number of storms, Nc = n, where n = 2, 3, 4, $\cdots$, the next step is to disaggregate Zc into n amounts of Yc_1, Yc_2, $\cdots$, Yc_n such that

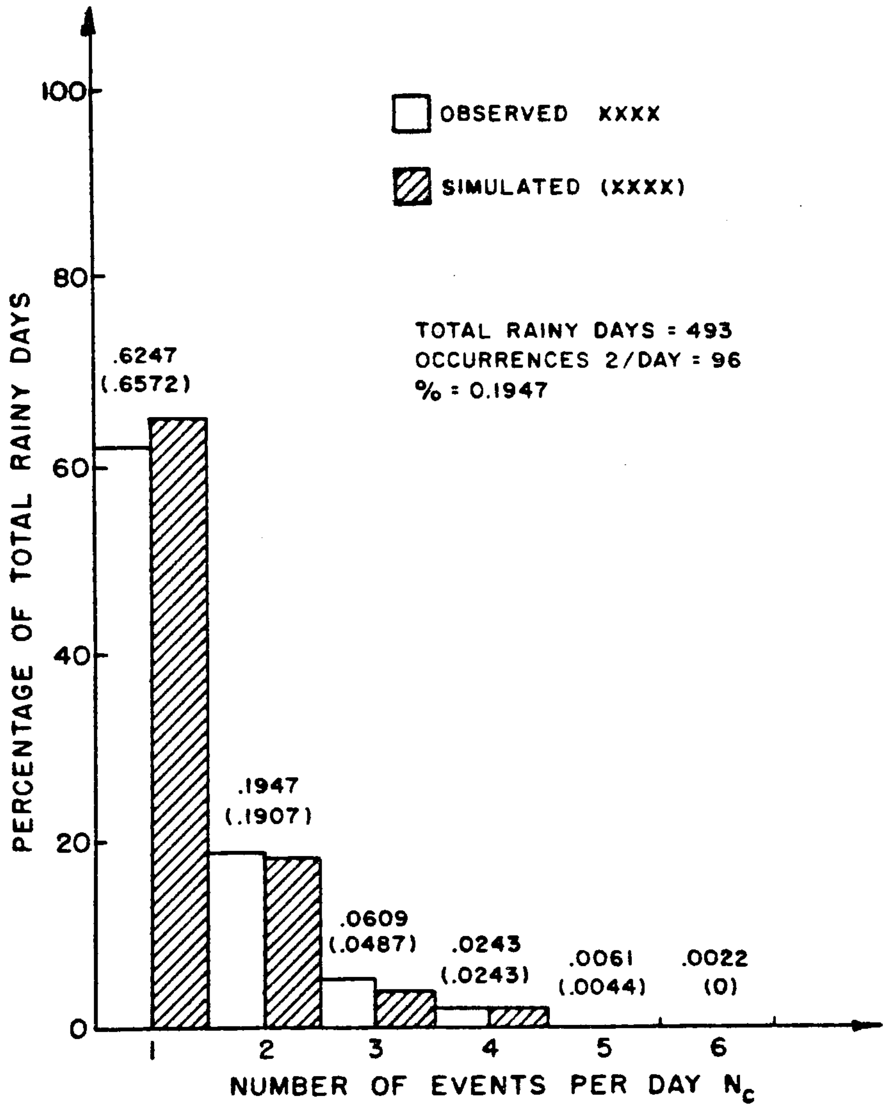


Figure 4.6. Results of simulation of conditional distribution of number of events, given daily amount, for complete events only

$$\sum_{i=1}^{n} Yc_i = Zc \qquad (4.10)$$

Note that, if only one rainfall event occurred, $Yc_1 = Zc$, and further disaggregation is not necessary. Also, as the greatest number of complete events observed in any day was six, there is no historical basis for disaggregating Zc into more than six events, and the method proposed will assume n = 6 as an upper limit.

Define the random variables R_1 through R_6 as follows:

a. Two events per day (96 occurrences)

let

$$R_1 = \frac{\text{amount of first event}}{\text{daily total}} = \frac{Yc_1}{Zc} \qquad (4.11)$$

b. Three events per day (30 occurrences)

let

$$R_2 = \frac{Yc_2 + Yc_3}{Zc} \; ; \; R_3 = \frac{Yc_2}{Yc_2 + Yc_3} \qquad (4.12)$$

c. Four events per day (12 occurrences)

let

$$R_4 = \frac{Yc_3 + Yc_4}{Zc} \quad R_5 = \frac{Yc_1}{Yc_1 + Yc_2} \; ; \; R_6 = \frac{Yc_3}{Yc_3 + Yc_4} \qquad (4.13)$$

The methods developed involve fitting distributions to the CDF's of the storm ratios defined above. To simplify matters, it is assumed that R_1 through R_6 are all independent random variables. While this assumption may not be strictly valid, it makes the problem tractable, and gives reasonable results.

Preliminary investigation indicates that plots of the CDF's of R_1 and R_2 are similar. As they are ratios which take on values from zero to one, and because the shape of the curves indicates a need for a flexible distribution, the beta was tested first. Parameters were estimated using numerical maximum likelihood techniques.

The fit obtained is less than satisfactory (see Fig. 4.7a). However, if the lower part of the curve could be raised and the higher part lowered, a better fit is expected. This can be accomplished by the addition of a Fourier series term to the beta distribution:

$$f_{R_1}(r) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} r^{\alpha-1}(1-r)^{\beta-1} + \theta \sin 2\Pi r \; ; \qquad (4.14)$$

$$\alpha, \beta > 0 \; ; \; 0 \leq r \leq 1$$

θ must be defined so that at $r = .75$

$$|\theta \sin 2\pi r| < \frac{\Gamma(\alpha+B)}{\Gamma(\alpha)\Gamma(\beta)} r^{\alpha-1} (1-r)^{\beta-1}$$

This distribution has three parameters, α, β, and θ, and will be referred to as the beta-Fourier. It satisfies the two requirements for a density function that

(1) $f_{R_1}(r) \geq 0$

(2) $\int_0^\infty f_{R_1}(r)dr = 1$

For further discussion on this, see Appendix C.

Maximum likelihood parameter estimates were obtained for both $f_{R_1}(r)$ and $f_{R_2}(r)$, and the $\log_e$ likelihood function values were calculated to be $L_{R_1} = 9.8888$ and $L_{R_2} = 3.0749$, respectively. A total of six parameters are involved, three for each distribution. We want to test the hypothesis H_0 of a common parameter set, thus reducing the number of parameters needed from six to three. If $L_{R_1+R_2} = 12.5910$ is the $\log_e$ likelihood function value obtained when the parameters are optimized over all values of R_1 and R_2, and

$$\text{Log } \lambda = [L_{R_1} + {}_{R_2}] - (L_{R_1} + L_{R_2}) = 12.5910 - 12.9637$$

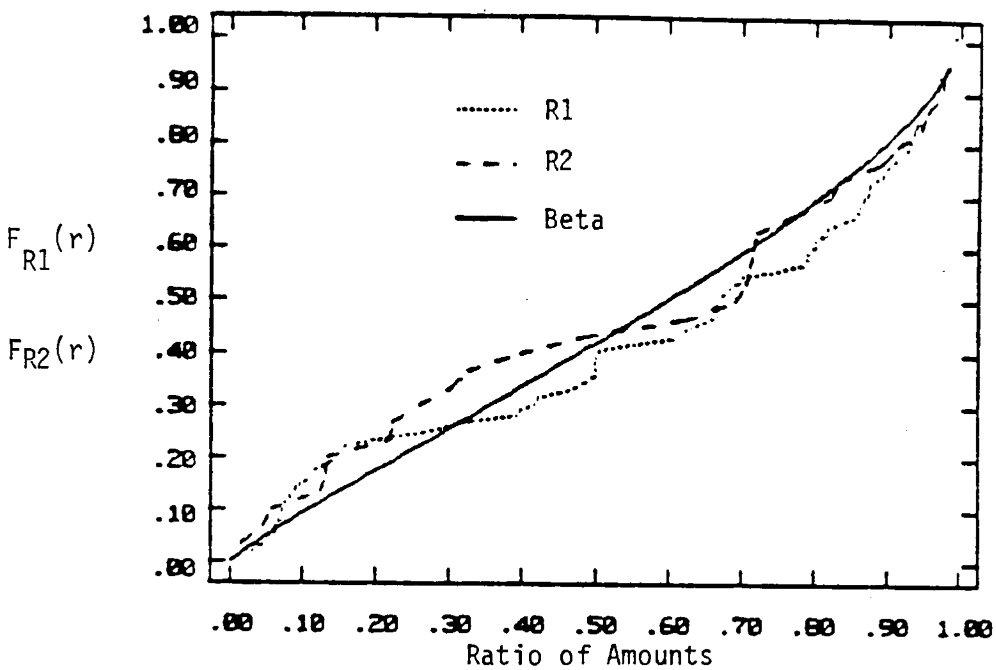
then the statistic $-2[\text{Log } \lambda]$ is approximately χ^2 distributed as n becomes infinite, with degrees of freedom equal to the number of parameters that is determined by the hypothesis H_0 (Hoel, 1971). The hypothesis is rejected for large values of the statistic. In this case,

$$-2(12.5910 - 12.9637) = .7454.$$

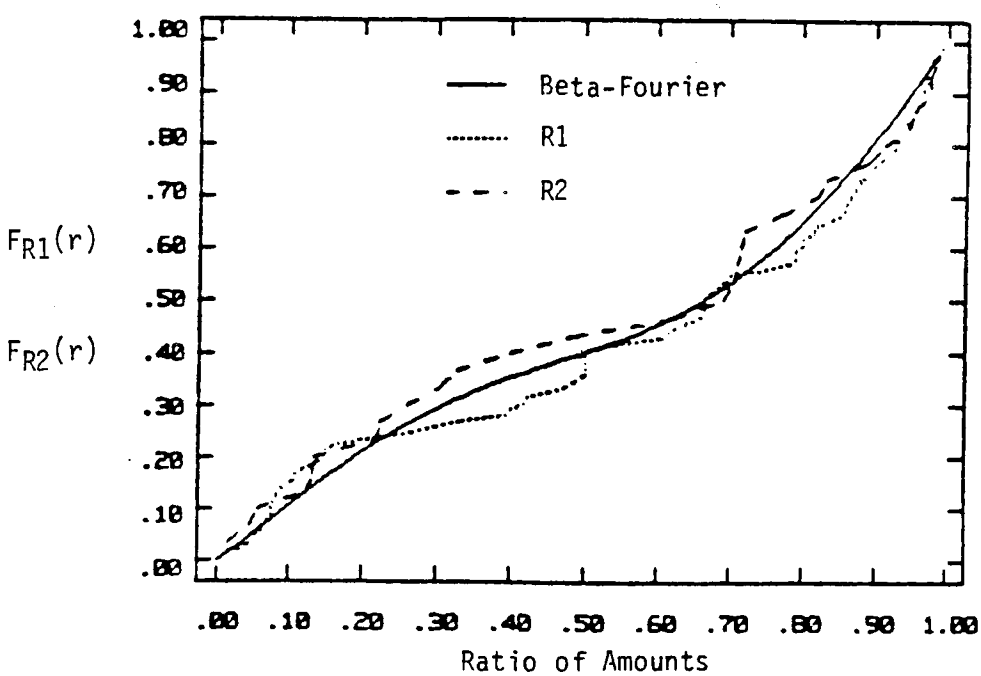
The χ^2 statistic for $\nu = 3$ (degrees of freedom) and $\alpha = .05$ is $\chi^2_{.05,3} = 7.8153$. Since the value obtained is less than this critical value, H_0 cannot be rejected. Thus, the distributions of R_1 and R_2 can be written as in Equation (4.14), with parameters as follows: $\alpha = 1.2514$, $\beta = .9045$, $\theta = .0819$. The CDF is obtained by numerically integrating Equation (4.14) (see Fig. 4.7b).

The plots of the observed CDF's for the random variables R_3, R_4, R_5, and R_6 are shown by the dotted lines in Figures 4.8 through 4.11. As the sample size for each is small (30, 12, 12, and 12, respectively) and, due to the shapes of the observed distributions, it is not unreasonable to hypothesize a uniform distribution to describe each. K-S goodness-of-fit tests were run to test these hypotheses. The results are shown in Table 4.6. As the largest differences between the observed and hypothesized CDF's are all smaller than the critical values, the hypothesis of a uniform distribution for each cannot be rejected at the 95% level.

Due to a lack of data for the classes of five and six events per day (3 and 1 occurrences, respectively), some ad hoc assumptions are necessary to describe the disaggregation process. The first is that, on



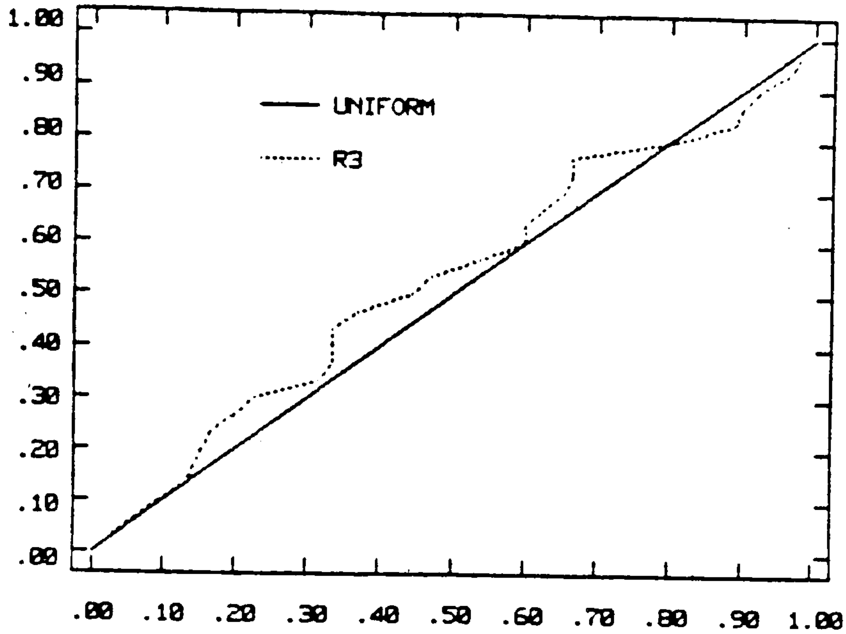
A) Beta distribution



B) Beta-Fourier distribution

Figure 4.7. Observed and fitted distributions of storm ratios R_1 and R_2

$F_{R3}(r)$



Ratio of Amounts R3

Figure 4.8. Observed and fitted distribution of storm ratio R3

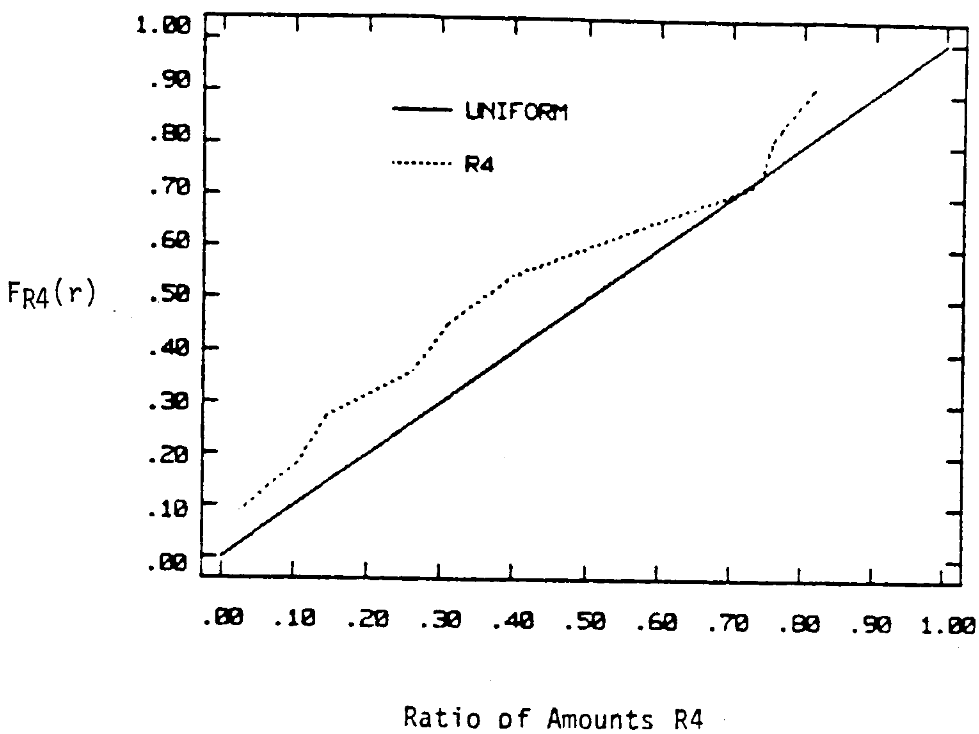


Figure 4.9. Observed and fitted distribution of storm ratio R4

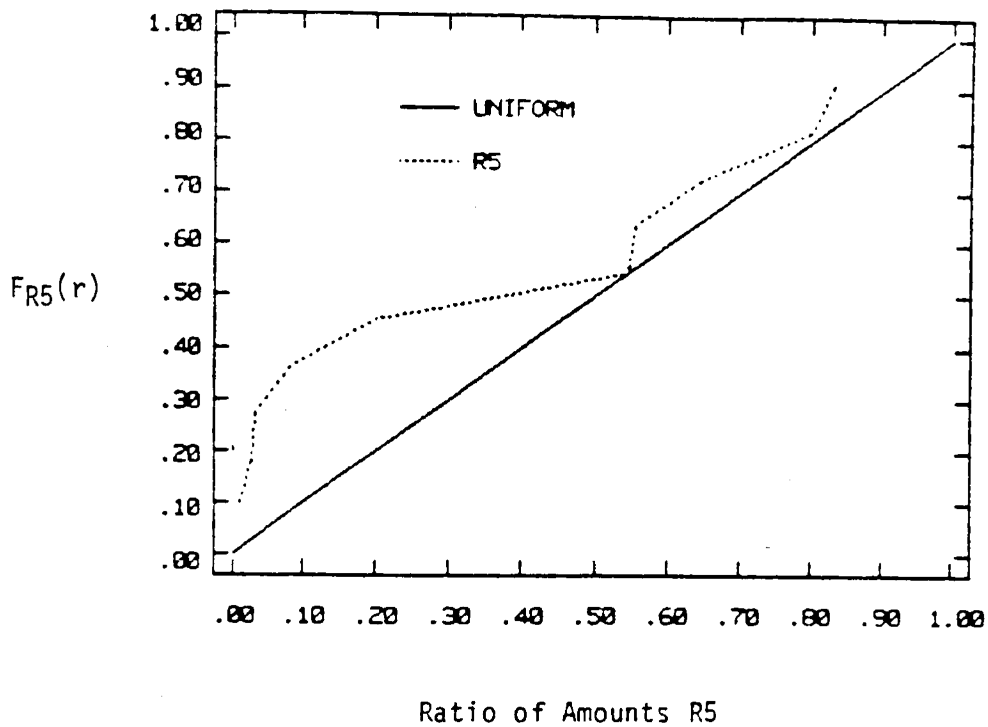


Figure 4.10. Observed and fitted distribution of storm ratio R5

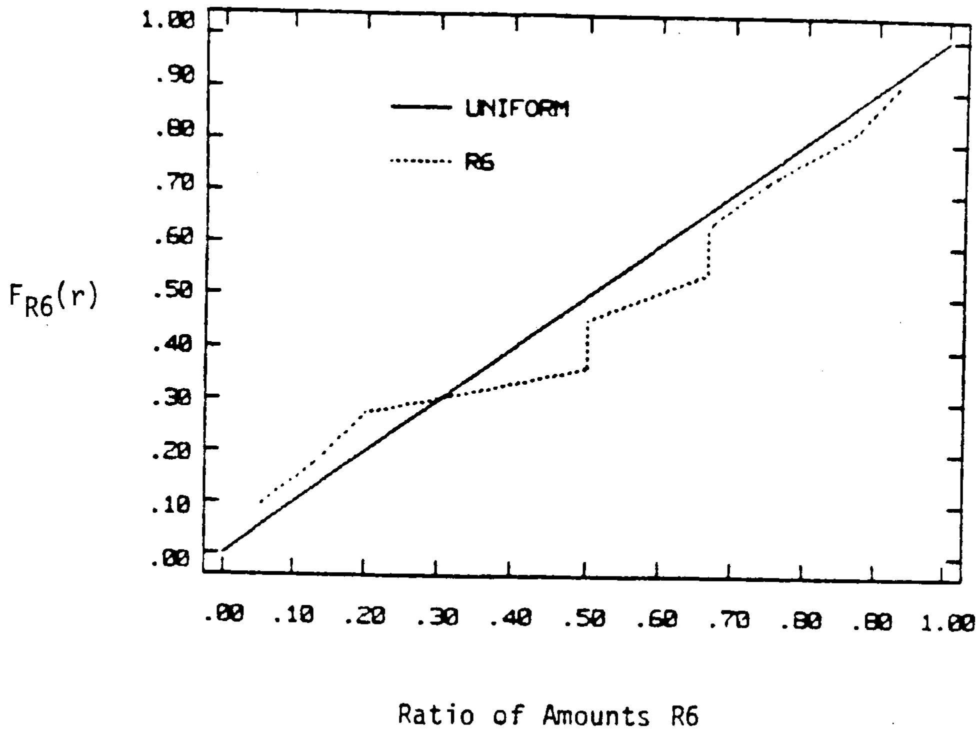


Figure 4.11. Observed and fitted distribution of storm ratio R6

Table 4.6. Kolmogorov-Smirnov goodness-of-fit test for the uniform distribution fit to the CDF's of R_3 R_4, R_5 and R_6

Random variable	Number of observations	Critical K-S; $\alpha = 0.05$	Maximum $\|D_2\|$ observed
$R_3 = \frac{Yc_2}{Yc_2 + Yc_3}$	30	.240	.118
$R_4 = \frac{Yc_3 + Yc_4}{Zc}$	12	.382	.121
$R_5 = \frac{Yc_1}{Yc_1 + Yc_2}$	12	.382	.198
$R_6 = \frac{Yc_3}{Yc_3 + Yc_4}$	12	.382	.115

days in which five events occur, the distribution of the ratio of the first plus second amounts ($1^{st} + 2^{nd}$) to the daily total is uniform. It is further assumed that the 1^{st} and 2^{nd} amounts can be disaggregated in the same manner as the 1^{st} and 2^{nd} amounts for days in which two events occur, and that the 3^{rd}, 4^{th} and 5^{th} amounts follow the scheme already developed for the 1^{st}, 2^{nd} and 3^{rd} amounts on days in which three events occur.

For days in which six events occur, it is assumed that the distribution of the ratio of the sum of the first three amounts to the daily total amount is uniform. Each set of three events is then disaggregated according to the scheme already developed for three events per day.

Joint Density Function of Event Amount and Duration

Given the daily rainfall amount, Zc, and that n storms have occurred, with the amounts Yc_1, Yc_2, • • •, Yc_n, it is necessary to simulate the duration associated with each event, Dc_1, Dc_2, • • • Dc_n. The joint density function of event amount and duration is written as:

$$h_{Yc',Dc}(y,d) = g_{Dc|Yc'}(d|y)f_{Yc'}(y) \qquad (4.15)$$

where $g_{Dc|Yc'}(d/y)$ is the conditional density function of event duration given an amount, and $f_{Yc}'(y)$ is the marginal density of the transformed event amount Yc', where Yc' = Yc -.009.

Four choices for $f_{Yc}'(y)$ were compared: the kappa(3), mixed exponential, lognormal, and Weibull density functions. Parameters were estimated using maximum likelihood techniques. The Weibull distribution has both the maximum $\log_e$ likelihood function value and minimum AIC (see

Table 4.7), and is the best choice. A χ^2 goodness-of-fit test indicated that the hypothesis of a Weibull, to describe the marginal distribution of event amount $f_{Y_C}'(y)$, cannot be rejected at the 95% level, as the calculated statistic of 10.4805 was less than the critical statistic $\chi^2_{05,5} = 11.0758$. The graphs of the observed and Weibull CDF's are shown in Figure 4.12. The distribution is written as:

$$F_{Y_C'}(y) = 1 - \exp\left[-\left(\frac{y}{\alpha}\right)^{\gamma}\right]$$

$$f_{Y_C'}(y) = \frac{\gamma}{\alpha}\left(\frac{y}{\alpha}\right)^{\gamma-1} \exp\left[-\left(\frac{y}{\alpha}\right)^{\gamma}\right] \tag{4.16}$$

$$y \geq .001''$$

$$\alpha = .6428 \quad ; \quad \gamma = .1424$$

It now remains to determine a form for $g_{D_C|Y_C}'(d|y)$.

Many have found a significant correlation between storm depth (amount) and duration. Grace and Eagleson (1966) fit a Weibull distribution to duration and used a linear regression relationship to predict depth from duration. They fit a beta distribution to the residual percent spread. Sariahmed (1969) also fit a Weibull to duration and used a regression to predict depth from duration using a Weibull distribution for the residual percent spread. The residual percent spread is defined as:

$$\%H = \frac{\text{depth residual}}{\text{depth from regression equation}}$$

The approach taken here is to predict event duration from an amount. Many functional relations between duration and amount were tested for linear dependency; these included:

$$Dc \text{ vs } Yc'$$

$$\log_e Dc \text{ vs } Yc'$$

Table 4.7. Comparison of distributions for the marginal distribution of storm rainfall amount

Distribution	Number of parameters	$\log_e$ likelihood function value	AIC*
Kappa(3)	3	495.0176	-984.0352
Mixed exponential	3	490.9962	-975.9924
Lognormal	2	480.9519	-957.9038
Weibull	2	502.2669**	-1000.5338**

*AIC = -2 $\log_e$ (max. likelihood function value + 2m) where m is the number of parameters

**Best values.

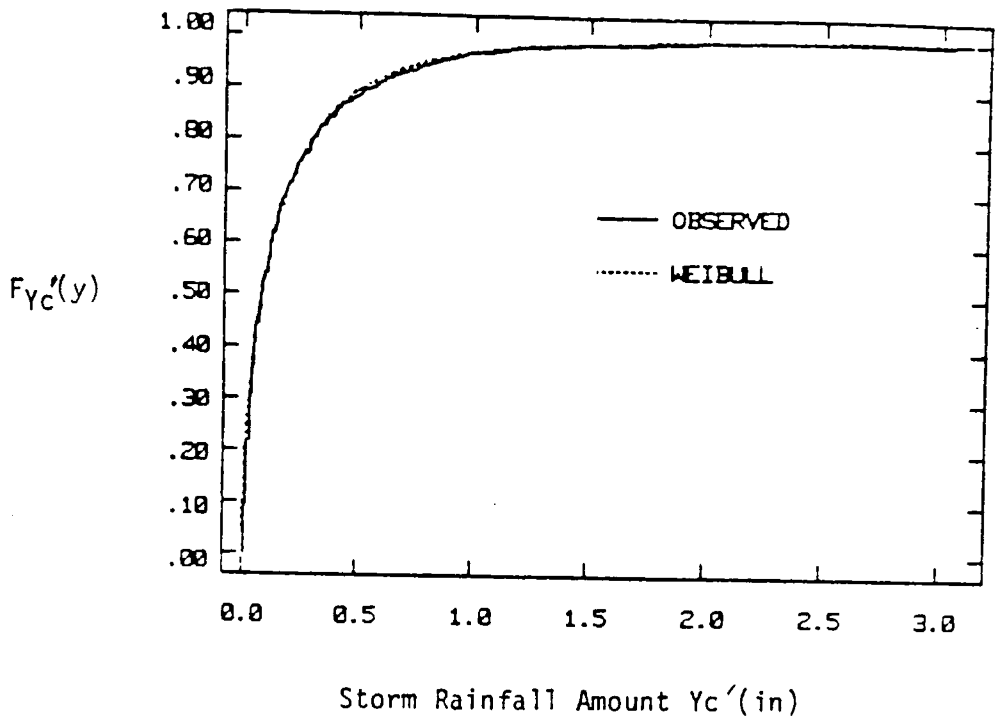


Figure 4.12. Observed and fitted distribution for storm rainfall amount, complete events

$$\log_e Dc \text{ vs } Yc'^{1/2}$$

$$\log_e Dc \text{ vs } Yc'^{1/3}$$

$$\log_e Dc \text{ vs } \log_e Yc'$$

The best linear relationship was found between $\log_e$ Dc vs $\log_e$ Yc'.

To simplify the notation, let Y represent Yc. Define a new random variable Y" where Y" = $\log_e$ Y'. Recall that Y' = Y -.009. Then

$$Y' = e^{Y''}, \text{ and } \frac{dY'}{dY''} = e^{Y''}$$

Thus, Y" is a function of a random variable and has a density function

$$g_{Y''}(y'') = f_{Y'}(y')\left|\frac{dy'}{dy''}\right|.$$

Thus

$$g_{Y''}(y'') = \frac{\gamma}{\alpha}\left(\frac{y'}{\alpha}\right)^{\gamma-1} \exp\left[-\left(\frac{y'}{\alpha}\right)^{\gamma}\right]e^{y''}$$

or

$$g_{Y''}(y') = \frac{\gamma}{\alpha}\left(\frac{y'}{\alpha}\right)^{\gamma-1} \exp\left[-\left(\frac{y'}{\alpha}\right)^{\gamma} + y''\right] \qquad (4.17)$$

Also, define a new random variable D' = $\log_e$ D where D represents Dc. Then assume the conditional density of D', given Y", is normal, with an expected value function which is a linear function of Y":

$$E[d'|y''] = A + By''$$

A linear regression was run on d' versus y", and the following were obtained (see Figure 4.13):

slope = .3785 = B

intercept = 4.6460 = A

standard error of estimate = .8885

r^2 = .2478

To test for the linearity of the regression, a correlation ratio test was used (Kendall and Stuart, 1979). For large values of n, the statistic

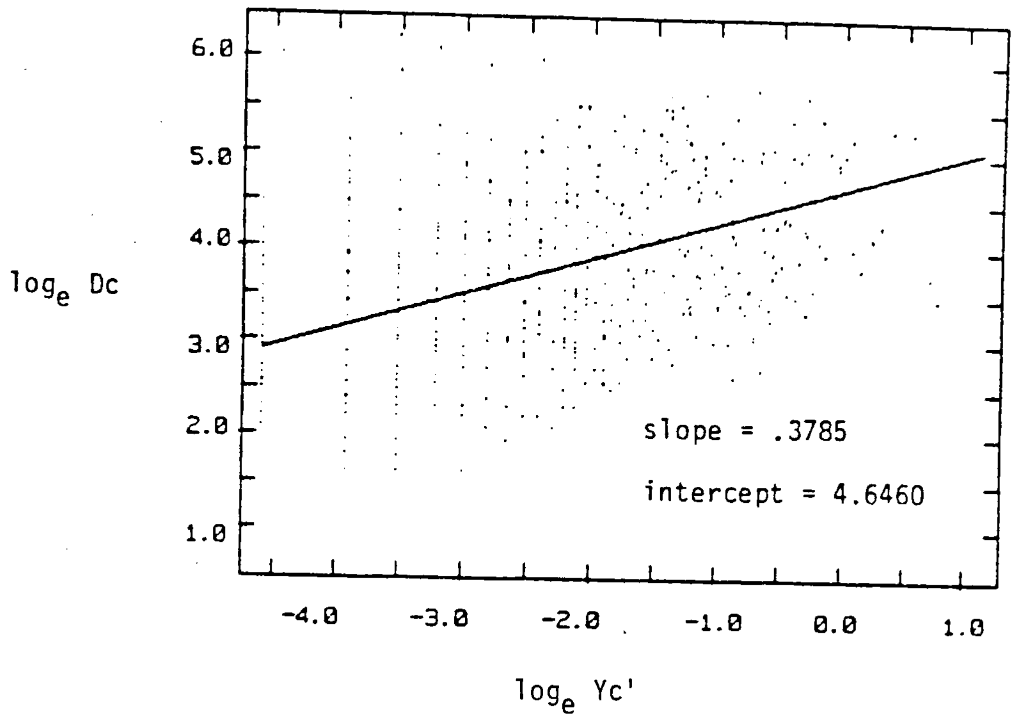


Figure 4.13. Scattergram of $\log_e$ duration vs. $\log_e$ of amount, with regression line for complete events

$$\frac{(e^2-r^2)/(k-2)}{(1-e^2)/(n-k)} \tag{4.18}$$

is approximately F-distributed with (k-2) and (n-k) degrees of freedom; where k is the number of classes, n is the number of observations in each class, r^2 is the coefficient of determination and

$$e^2 = \frac{\sum_{i=1}^{k} n_i(\overline{\ln Dc_i} - \overline{\ln Dc})^2}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (\ln Dc_{ij} - \overline{\ln Dc})^2} = \frac{\sum_i n_i(\overline{\ln Dc_i})^2 - n(\overline{\ln Dc})^2}{\sum_i \sum_j (\ln Dc_{ij})^2 - n(\overline{\ln Dc})^2} \tag{4.19}$$

In the equation above, $\overline{\ln Dc_i}$ is the mean of the i^{th} class, $\overline{\ln Dc}$ is the mean of all observations, j refers to the j^{th} member in class i, and n_i is the number of observations in class i. The hypothesis is rejected for large values of the statistic. At α = 0.05, the critical statistic is 1.8300. The value obtained was 1.7798, so the hypothesis cannot be rejected at the 95% level. Calculations are shown in Table 4.8.

To test the hypothesis that the conditional density of d' given y" is normal, the values of y" were divided into four classes, and a χ^2 test was run on the residuals from the regression, d'-E[d'], for each class. The classes are shown in Table 4.9.

For each class, the hypothesis is that the residuals are distributed normally, N(0,.8885), about the regression line. At $\chi^2_{.05,6}$, the critical statistic is 12.5982. The four values obtained were 9.4543, 9.6526, 3.0975, and 5.4383 for classes 1 through 4, respectively. This hypothesis cannot be refected for any class at the 95% level.

Thus, the conditional density is written as:

$$g_{Dc'|Yc''}(d'|y'') = .3785\, y'' + 4.6460 + \varepsilon(0,.8885) \tag{4.20}$$

Table 4.8. Calculations for the correlation ratio test, for the hypothesis of linearity of the regression, complete events

Interval $\log_e (Yc_i - .009)$			$\overline{\text{Log}_e\ Dc_i}$	$(\log_e Dc_i)^2$	n_i	$n_i(\overline{\log_e Dc_i})^2$
-4.9999	to	-4.5000	2.9150	8.4961	57	484.2754
-4.4999	to	-4.0000	0	0	0	0
-3.9999	to	-3.5000	3.1343	9.8240	130	1277.1232
-3.4999	to	-3.000	3.7194	13.8340	33	456.5224
-2.9999	to	-2.5000	3.6467	13.2983	87	1156.9563
-2.4999	to	-2.0000	3.8258	14.6371	74	1083.1418
-1.9999	to	-1.5000	3.8148	14.5525	68	989.5680
-1.4999	to	-1.0000	4.4278	19.6052	62	1215.5246
- .9999	to	- .5000	4.0934	16.7559	48	804.2843
- .4999	to	0	4.6950	22.0431	36	793.5523
0	to	.4999	4.7825	22.8724	12	274.4688
.5000	to	.9999	4.5341	20.5577	5	102.7885
1.0000	to	1.49999	4.4529	19.8285	2	39.6570
TOTAL					614	8677.8625

$$e^2 = \frac{8677.8625 - 614(3.7217)^2}{613(1.0236)^2} = .2694$$

mean $\log_e$ Dc = 3.7217

variance $\log_e$ Dc = $(1.0236)^2$

$r^2 = .2478$; $n = 614$; $k = 12$

$$\frac{(e^2 - r^2/(k - 2)}{(1 - e^2)/(n - k)} = \frac{(.2694 - .2478)/10}{(1 - .2694)/602} = \frac{.00216}{.00121} = 1.7798$$

$F_{10,602} = 1.83$ at 95% level

Table 4.9. Classes for the test of normality of residuals about the regression line

Class	Interval $(\log_e (Y_c - .009))$			Size
1	$-\infty$	to	-3.5000	187
2	-3.4999	to	-2.000	194
3	-1.9999	to	-1.000	130
4	- .9999	to	∞	103

A transformation is performed to obtain the duration Dc as follows:

$$Dc = \exp(Dc'). \tag{4.21}$$

Distribution of Event Starting Times

The time of day of event occurrence is important to certain rainfall-driven processes. For instance, on days in which multiple events occur, antecedent soil moisture conditions affect infiltration rates.

If Uc_ℓ, $\ell = 1, 2, \cdots, n$ represents the starting time of the ℓ^{th} complete event on a given day, in military hours, then the ratio

$$Tc = \frac{Uc_\ell}{24} \tag{4.22}$$

takes on the values zero through one inclusive, representing this starting time. The observed CDF's of the starting times for the classes of one, two, three and four events per day are shown by the dotted lines in Figures 4.14 through 4.17, respectively. As the classes of five and six events per day are small (3 and 1 occurrences), the ad hoc assumption is that the distribution which describes the starting times of classes of one through four events per day also describes the starting times of classes of five and six events per day.

The plots of the CDF's of event starting times indicate that a flexible distribution is needed. As the value of the ratio Tc_ℓ is between zero and one, a beta was tried with maximum likelihood parameter estimates. The fit is highly unsatisfactory (see Figure 4.18a). It appears that more than two parameters will be needed.

The mixed beta distribution has five parameters, and is written as:

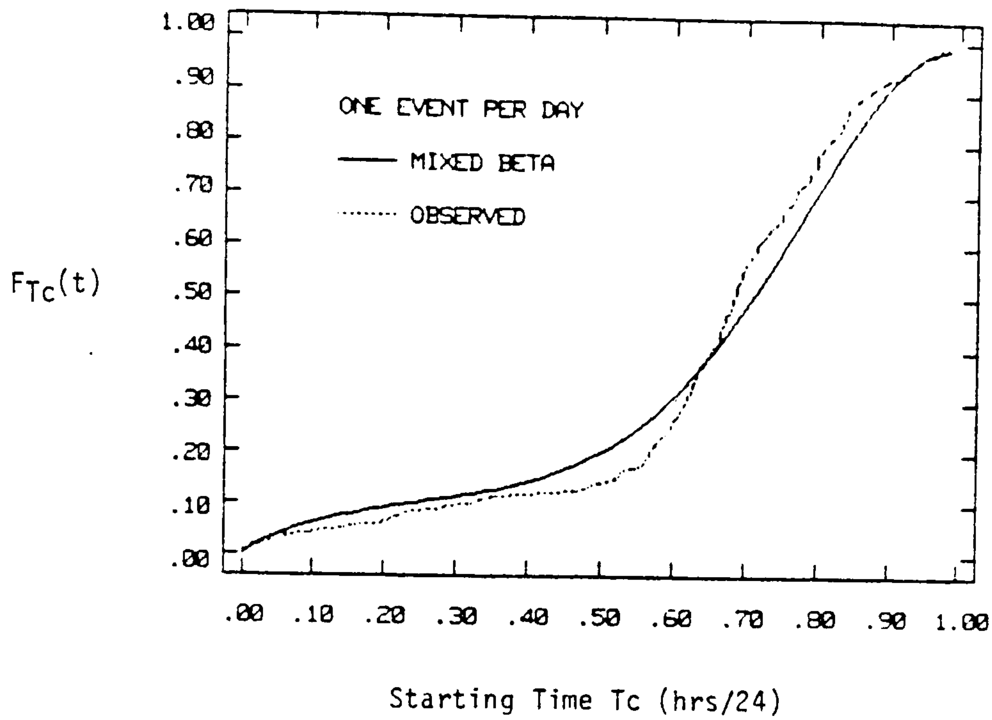
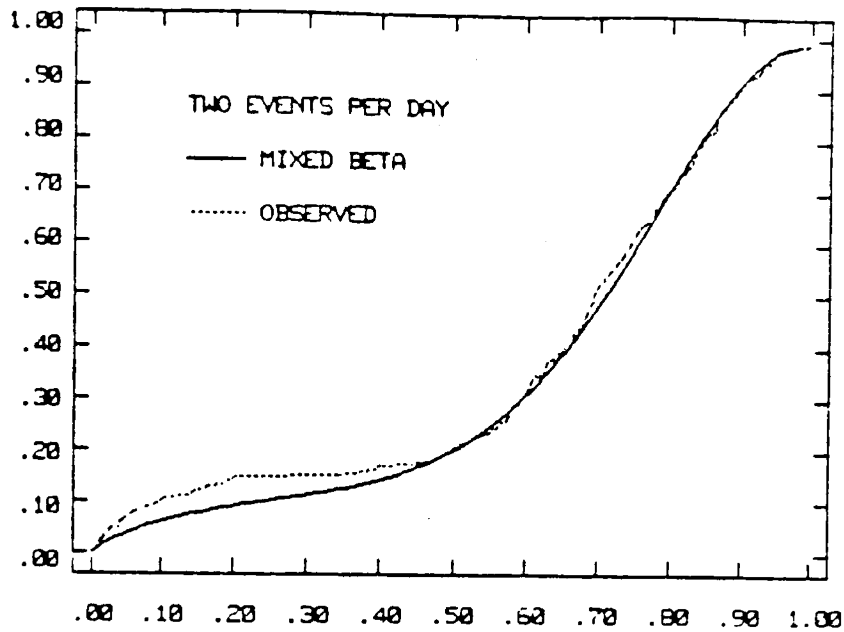


Figure 4.14. Observed and fitted distribution for starting time, one event per day

$F_{Tc}(t)$



Starting Time Tc (hrs/24)

Figure 4.15. Observed and fitted distribution for starting time, two events per day

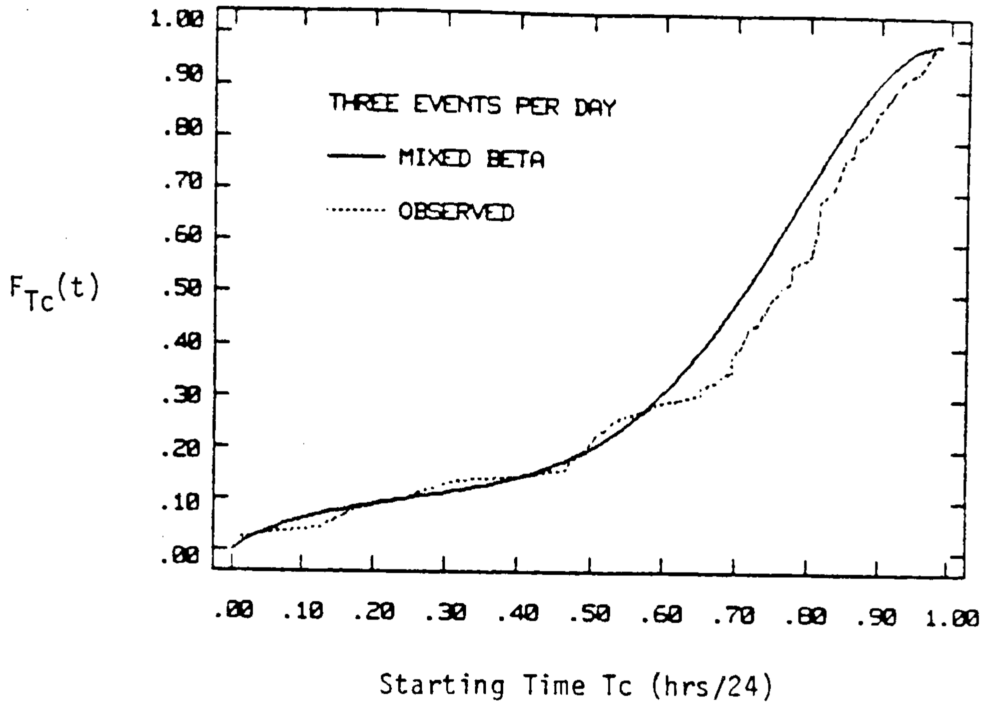


Figure 4.16. Observed and fitted distribution for starting time, three events per day

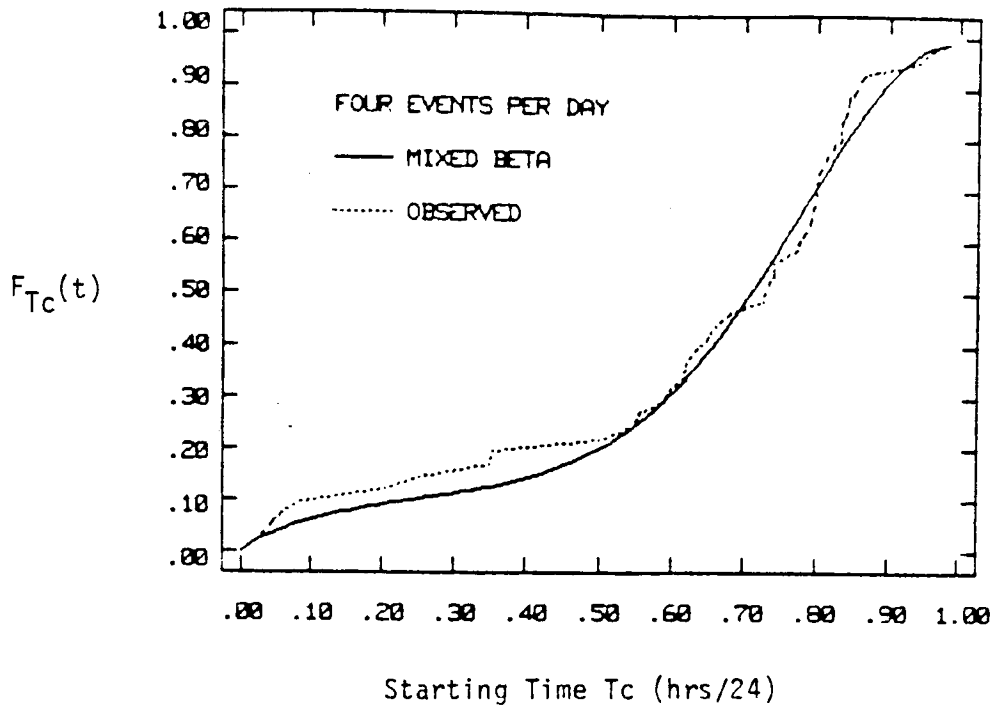


Figure 4.17. Observed and fitted distribution for starting time, four events per day

$$F_{Tc}(t) = \omega\left[\frac{\Gamma(\alpha_1+\beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)}t^{\alpha_1-1}(1-t)^{\beta_1-1}\right] + (1-\omega)\left[\frac{\Gamma(\alpha_2+\beta_2)}{\Gamma(\alpha_2)\Gamma(\beta_2)}t^{\alpha_2-1}(1-t)^{\beta_2-1}\right] \tag{4.23}$$

$$\alpha_1,\ \alpha_2,\ \beta_1,\ \beta_2, > 0 \quad ; \quad 0 < \omega < 1 \quad ; \quad 0 \leq t \leq 1$$

The fit obtained was much better than with the simple beta distribution (see Fig. 4.18b).

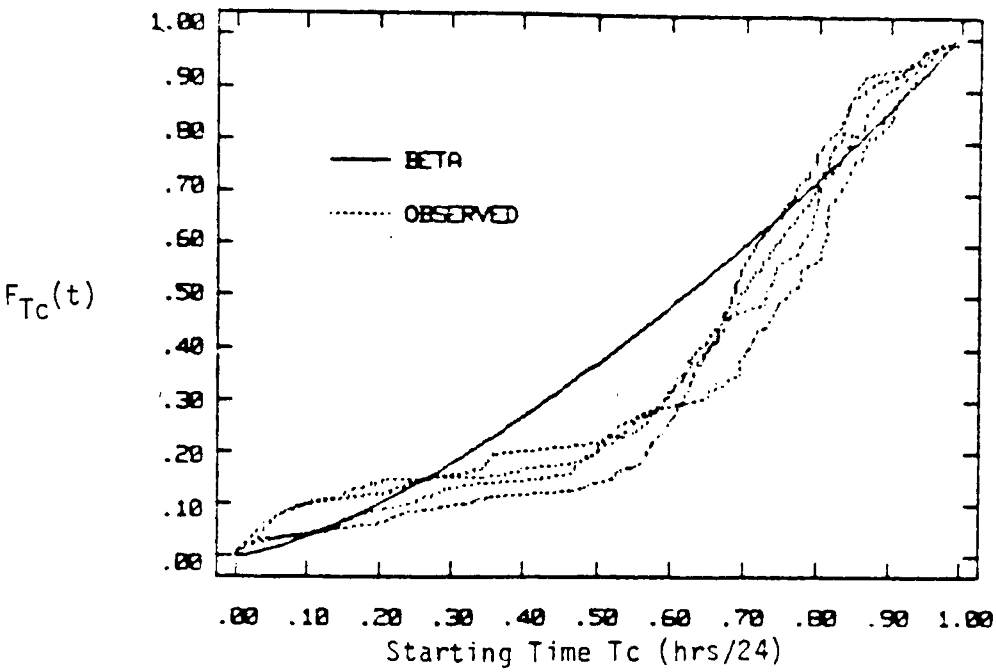
Although the number of parameters is larger than desirable, there are many advantages to using this distribution. A plot of the probability density function (pdf) of starting times $f_{Tc}(t)$ indicates high frequencies of occurrence in the late afternoon, and just before midnight for summer thunderstorm rainfall in southeast Arizona (see Figure 4.19). The mixed beta distribution allows for the preservation of this bimodal tendency.

Also, the use of order statistics enables a relatively easy method of predicting starting times on multiple event days. Let $F_{Tr}(t)$ represent the CDF of the starting time of the r^{th} event on a day in which n events occur. The general form is:

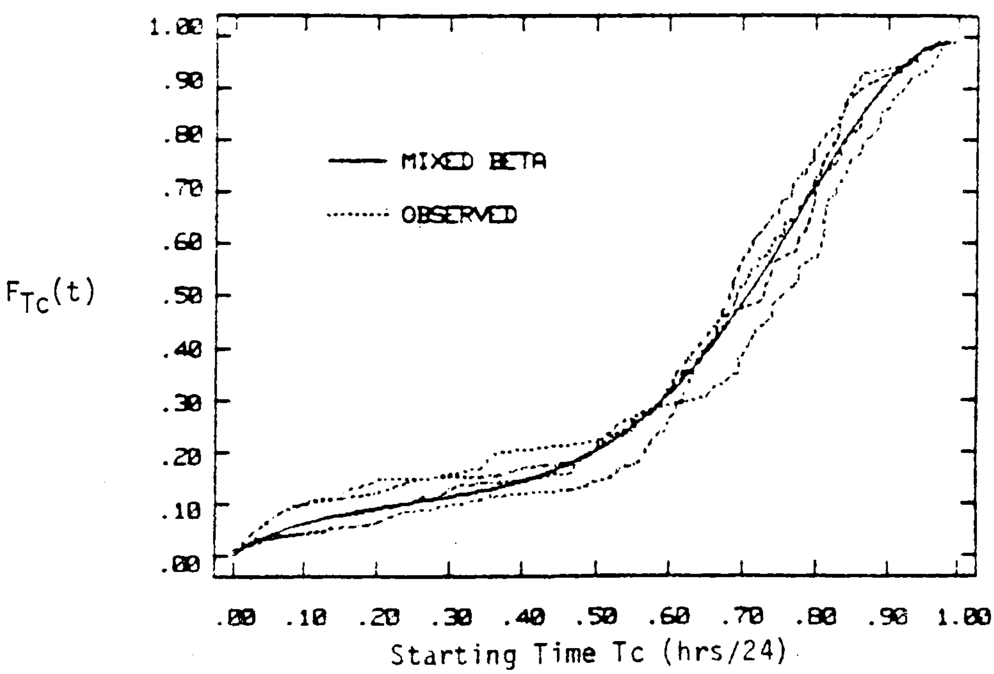
$$F_{Tr}(t) = \frac{n!}{(r-1)!(n-r)!}\left[\sum_{i=0}^{n-r}\frac{\binom{n-r}{i}(-1)^i}{i+r}(F_{Tc})^{i+r}\right] \tag{4.24}$$

where F_{Tc} is the CDF of all event starting times, as given in Equation (4.23).

Given that n events occur in a day, it is only necessary to generate r = 1, 2, • • •, n values according to Equation (4.24), order these from lowest to highest, and multiply by 24 to get the appropriate starting times. See Table 4.10 for the mathematical expressions for the



A) Beta distribution



B) Mixed beta distribution

Figure 4.18. Observed and fitted distributions of starting time for all events

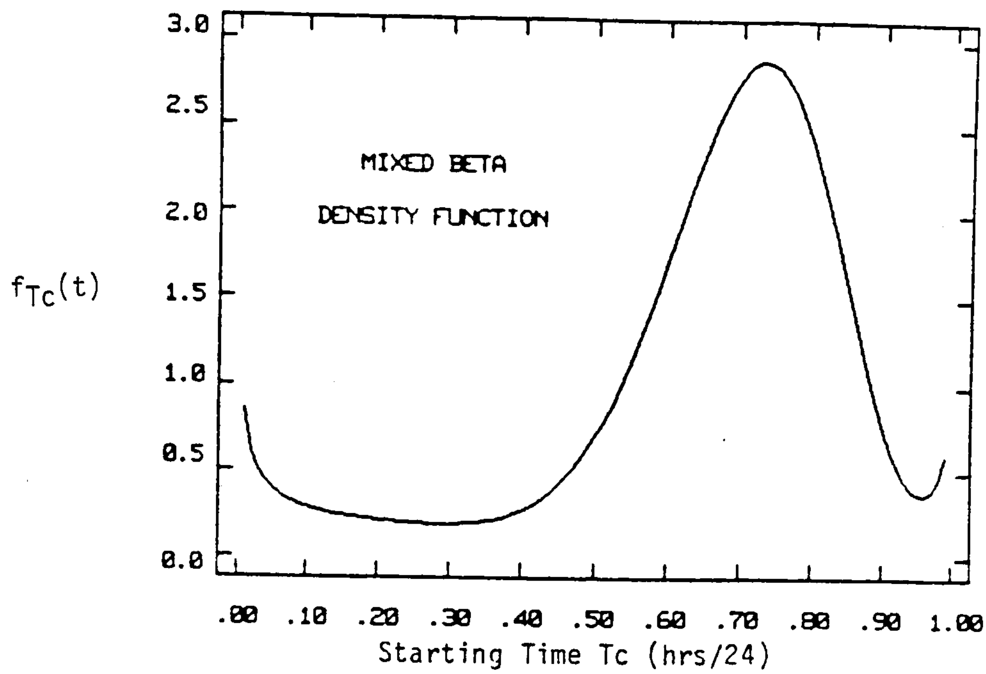


Figure 4.19. Mixed beta density function fit to starting time for all complete events

distributions of starting times obtained by using order statistics for the occurrence of one through four events in a day.

Thus, it is desirable to test two hypotheses concerning event starting times.

The first, H_0, is that the mixed beta distribution, shown in Equation (4.23) with parameters optimized over the entire set of complete storms, describes the observed CDF's of starting times for the classes of one, two, three, and four events per day (refer back to Figures 4.14 through 4.17). A K-S goodness-of-fit test was run for each of the four classes, and the results are shown in Table 4.11. The hypothesis cannot be rejected at $\alpha = 0.05$ for the classes of two, three, and four events per day. Although the hypothesis is rejected at $\alpha = 0.05$ for the class of one event per day, it cannot be rejected at $\alpha = 0.01$. It was decided not to reject H_0 under these conditions. The mixed beta distribution, as given in Equation (4.23), describes the starting times of all complete events. Maximum likelihood parameter estimates are:

$$\alpha_1 = .6389$$

$$\beta_1 = 3.2895$$

$$\alpha_2 = 6.2318$$

$$\beta_2 = 2.3816$$

$$\omega = .1483$$

The second hypothesis we wish to test, H_0, is that on a day in which n events occur, the distribution of the starting times of each of these events can be obtained using order statistics. To test this hypothesis, K-S goodness-of-fit tests were run on the observed and

Table 4.10. Distributions of event starting times obtained by using order statistics

Number of events per day	Order of occurrence	Cumulative Distribution Function
2	1st	$F_{T1} = 2F - F^2$
	2nd	$F_{T2} = F^2$
3	1st	$F_{T1} = 3F - 3F^2 + F^3$
	2nd	$F_{F2} = 3F^2 - 2F^3$
	3rd	$F_{T3} = F^3$
4	1st	$F_{T1} = 4F - 6F^2 + 4F^3 - F^4$
	2nd	$F_{T2} = 6F - 8F^3 + 3F^4$
	3rd	$F_{T3} = 4F^3 - 3F^4$
	4th	$F_{T4} = F^4$

*NOTE: F is F_{TC}, as given in Equation 4.23.

Table 4.11. Kolmogorov-Smirnov goodness-of-fit test for mixed beta distribution fit to event starting times

Class of events	Number of observations	Critical K-S value; α = 0.05	Max. $\lvert D_2 \rvert$ observed
one/day	308	0.078	0.090*
two/day	192	0.098	0.055
three/day	90	0.143	0.141
four/day	48	0.196	0.081

*At α = 0.01, the critical K-S statistic is .0929; the hypothesis cannot be rejected at α = 0.01.

predicted CDF's (see Figures 4.20 through 4.22). Results are shown in Table 4.12.

For the class of one event per day, it has already been shown that the hypothesis cannot be rejected at $\alpha = 0.01$. With the exception of the first event on days in which three occur, this hypothesis cannot be rejected at $\alpha = 0.05$. However, H_0 cannot be rejected for the first event of three per day at $\alpha = 0.01$. Therefore, it was also decided not to reject H_0 under these conditions.

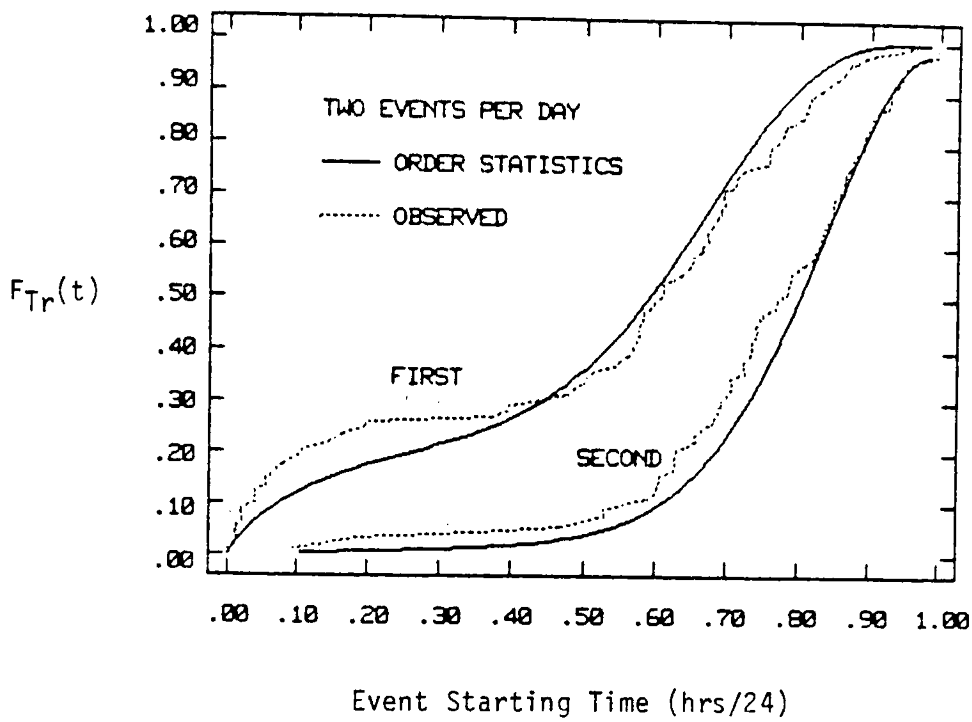


Figure 4.20. Observed and fitted distributions for starting times of two events per day using order statistics

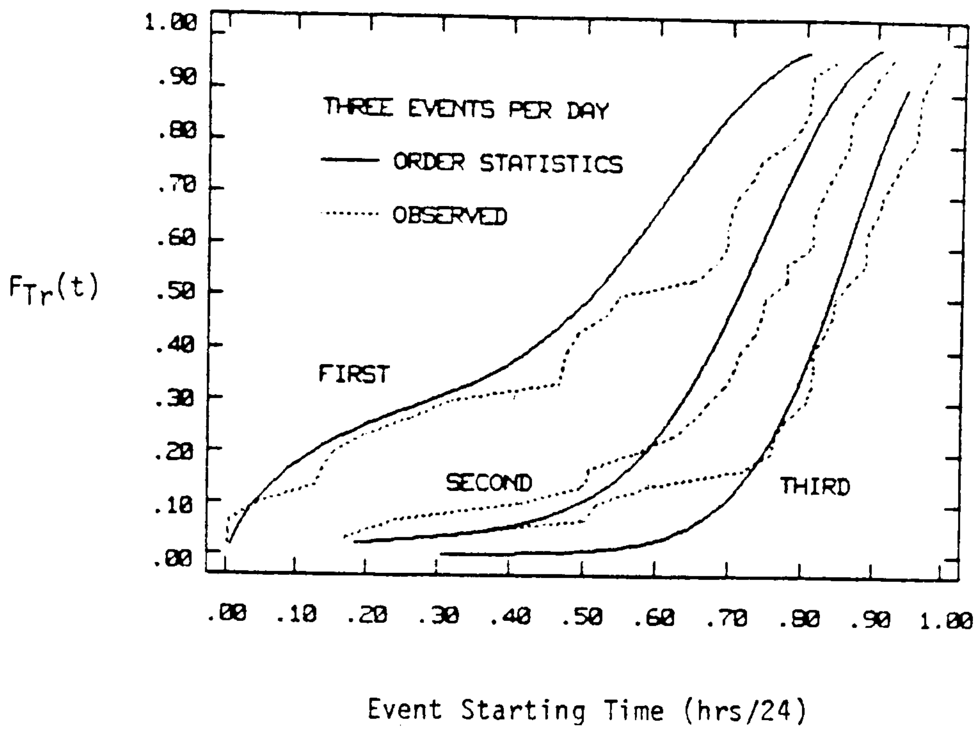


Figure 4.21. Observed and fitted distributions for starting times of three events per day using order statistics

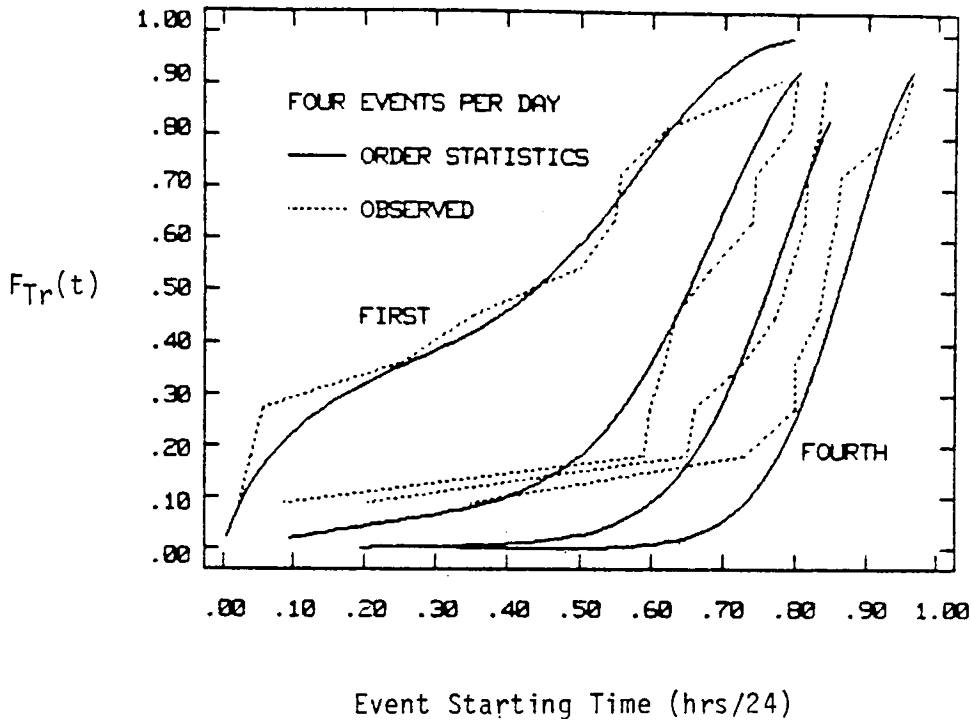


Figure 4.22. Observed and fitted distributions for starting times of four events per day using order statistics

Table 4.12. Kolmogorov-Smirnov goodness-of-fit test on the use of order statistics to describe event starting times

Class of events	Order in day	Number of observations	Critical K-S $\alpha = 0.05$	Max. $\lvert D_2 \rvert$ observed
two/day	1st	96	.139	.083
	2nd	96	.139	.121
three/day	1st	30	.240	.261*
	2nd	30	.240	.229
	3rd	30	.240	.145
four/day	1st	12	.382	.110
	2nd	12	.382	.158
	3rd	12	.382	.094
	4th	12	.382	.216

*At $\alpha = 0.01$, the critical K-S statistic is .290; the hypothesis cannot be rejected at $\alpha = 0.01$.

Model of Partial Events

Occurrence

If Zc_j is the daily rainfall amount contributed by complete events on day j, and Zp_j is the daily amount contributed by partial events, then $Z_j = Zp_j + Zc_j$ is the total daily rainfall amount on day j. Given two consecutive nonzero daily rainfall amounts, Z_j and Z_{j+1}, it is desirable to determine whether the occurrence of a partial event is dependent on some functional relationship of these amounts. The two relationships examined are the sum and product of daily amounts, $Zs = Z_j + Z_{j+1}$ and $Zpr = Z_j \cdot Z_{j+1}$, respectively.

The first hypothesis to be tested, H_0, is that the probability of occurrence of a partial event is independent of the sum of amounts, Zs.

Let $P\{M = m\}$ be the probability mass function of the random variable m where:

$$m = \begin{cases} 0 & \text{if no partial event occurs} \\ 1 & \text{if a partial event occurs} \end{cases}$$

If $g_{Zs|M}(z|m)$ is the conditional density of the sum of daily amounts, given the state of occurrence, then the equivalent of the joint density can be written as:

$$h_{Zs,M}(z,m) = g_{Zs|M}(z|m)\ P\{M = m\}\ ;\ m = 0,\ 1$$

However, if the hypothesis of independence is true, the joint density can be written as the product of the marginals:

$$h_{Zs,M}(z,m) = f_{Zs}(Z)\ P\{M=m\} \tag{4.25}$$

Thus, if the conditional $g_{Zs|M}(z|o)$ is the same as $g_{Zs|M}(z|1)$, then the joint distribution can be expressed in terms of the marginals. Therefore, H_0, the hypothesis of independence, cannot be rejected.

The test of H_0 involves looking for a significant difference between Zs on days when a partial event occurred and on days when no partial event occurred. Of the 217 occurrences of W-W days (two consecutive wet days) in the data set, 36 contained partial events, and 181 did not.

The rank-sum test is a nonparametric test which can be used to test if two samples come from the same distribution (Hoel, 1971). It is based on the randomization principle; i.e., in determining the critical region for the test statistic, the distribution is considered under all possible permutations of the observed values rather than on a new set of values, as in the traditional testing for a specified distribution. For example, let $x_1', x_2', \cdots, x_n'$ represent the random sample of Zs for days on which a partial event occurred, and $y_1', y_2', \cdots. y_n'$ represent the random sample of Z_s on a day in which no partial event occurred. If each sample is arranged in order of increasing magnitude, the ordered sets $x_1, x_2, \cdots, x_n$ and $y_1, y_2, \cdots, y_n$ are obtained. If the two ordered sets are combined into a single ordered set, a typical arrangement might be $y_1, y_2, x_1, y_3, x_2, y_4, \cdots$. The ranks of the x values are written down, and Q is the sum of those numbers. In the given example, $Q = 3+5+ \cdots$. For large sample sizes, the statistic

$$\tau = \frac{Q - E(Q)}{\sigma_Q} \tag{4.26}$$

is approximately standard normal, where

$$E(Q) = \frac{n_1(n_1+n_2+1)}{2} \; ; \; \sigma_Q^2 = \frac{n_1 n_2(n_1+n_2+1)}{12} \tag{4.27}$$

In Equation (4.27), n_1 is the number of observations of W-W days with a partial event, and n_2 is the number of observations of W-W days without a partial event. The value obtained was:

$$\tau = \frac{4275 - 3924}{344.055} = 1.0202$$

A two-sided test is used and, at $\alpha = 0.05$, the critical value x is $|x| < 1.96$. Since the value obtained does not lie in the critical region, the hypothesis of independence of occurrence of a partial event and sum of daily rainfall amounts on W-W days, H_0, cannot be rejected at the 95% confidence level.

The second hypothesis to be tested, H_0, is that the probability of occurrence of a partial event is independent of the product of amounts, Zpr.

The form of the test is the same as above; however, it involves looking for a significant difference between Zpr on days when a partial event occurs and on days when no partial event occurs. The value obtained for the rank-sum test, as given by Equation (4.26), was

$$\tau = \frac{4108 - 3924}{344.055} = .5348$$

As this value is less than the critical value, $|x| < 1.96$, H_0 cannot be rejected at the 95% level.

Thus, it has been shown that the occurrence of a partial event is independent of the sum and product of daily rainfall amounts on two consecutive wet days. Occurrence can be determined in a probabilistic manner from the data. As 36 of 217 W-W day transitions contained partial events, $P\{m = 1\} = (36/217) = 0.1659$, given two consecutive wet days.

The number of partial events in a day is limited to zero, one, or two. If $P\{m = 1\} = .1659$ is the probability of occurrence of a partial event which begins on day j, then, by definition, this event ends on day j + 1, and therefore, a partial event also occurs on day j + 1. The scheme used to predict occurrence is also used to predict the number of occurrences per day.

Marginal Distribution of Partial Storm Amounts

The mean amount of each partial event (.4425 in. or 11.24 mm) is more than twice as great as the mean amount of a complete event (.1992 in. or 5.06 mm; refer back to Table 4.1). However, each partial event can be separated into two events based on day of occurrence, producing a new set of 72 events with a smaller mean amount. Herein, the term "partial event" will refer to this new class of separated storms, unless stated otherwise.

It is desirable to test the hypothesis H_0 that the marginal distribution of amounts for the partial events, $F_{Y_p'}(y)$, is no different from the marginal distribution of amounts for complete events, $F_{Y_c'}(y)$, as given in Equation (4.16). The variable Yp' is defined to be Yp' = Yp -.009. The likelihood ratio test is used.

Let L_{Y_c} be the $\log_e$ likelihood function value obtained when H_0 is true ($\alpha = .6428$, $\gamma = .1424$); and let L_{Y_p} be the $\log_e$ likelihood function value when the parameters are optimized only over the set of partial events ($\alpha = .6117$; $\gamma = .1487$). Then

$$\log \lambda = L_{Y_c} - L_{Y_p} \quad = 56.66619 - 56.666214$$

and the statistic -2(56.66619 - 56.666214) is approximately equal to zero. The two distributions are essentially indistinguishable. Thus, H_0 cannot be rejected at $\alpha = 0.05$, and the marginal distribution of amounts of partial events, $F_{Y_p'}(y)$, is described by Equation (4.16).

Joint Density Function of Amount and Duration

The joint density function of amount and duration is written as:

$$h_{Y_p', D_p}(y,d) = g_{D_p|Y_p'}(d|y) f_{Y_p'}(y)$$

and an expression for $f_{Y_p}'(y)$ has been found. The approach taken to determine the conditional density of duration given the amount, $g_{D_p|Y_p}(d|y)$, is the same as for complete events.

A linear regression was fit to the $\log_e$ of duration on the $\log_e$ amount, and the following were obtained (see Figure 4.23):

slope = .3296

intercept = 5.1624

standard error of estimate = .7755

$r^2 = .3066$

A correlation ratio test was run to test the hypothesis of linearity of the regression. Results are shown in Table 4.13. At $\alpha = 0.05$, $F_{4,66} = 2.52$. As the calculated statistic of .0429 is less than this critical value, the hypothesis cannot be rejected at the 95% level.

To test the hypothesis that the conditional density of the $\log_e$ Dp given $\log_e$ Yp' is normal, the $\log_e$ (amounts) were divided into three classes. For each class, the hypothesis is that the residuals are normally distributed N(0, .7755) about this line. At $\chi^2_{.05,1}$, the critical statistic is 3.9. The three values obtained were 1.7075, 1.8326, and 1.7329, respectively. The hypothesis cannot be rejected for any class

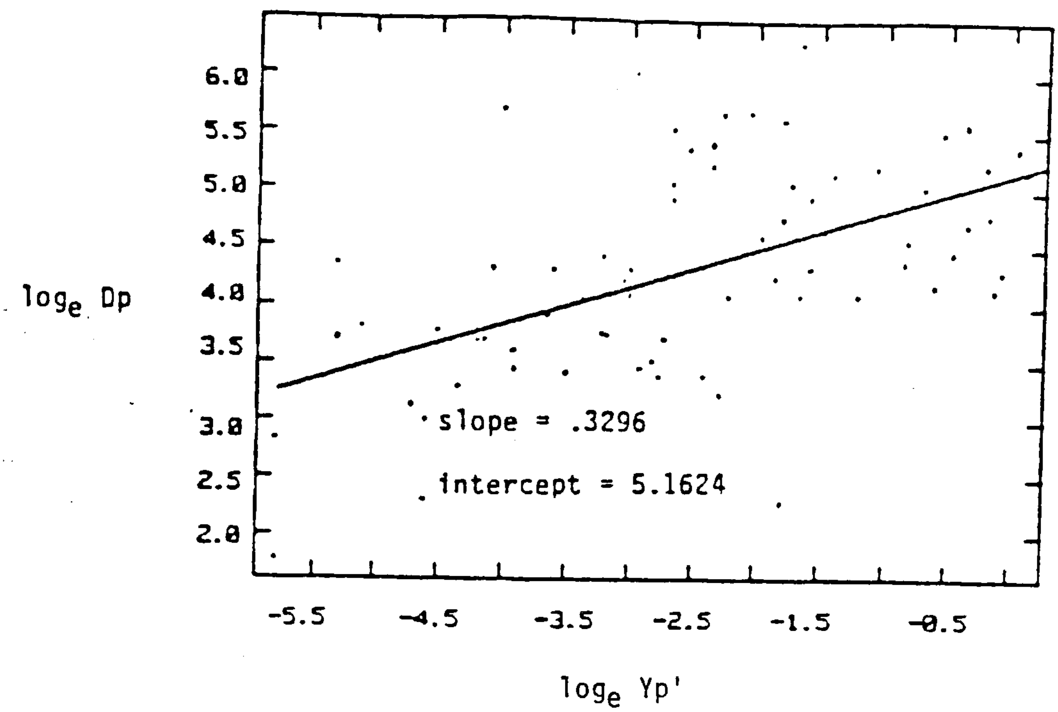


Figure 4.23. Scattergram of $\log_e$ of duration vs. $\log_e$ of amount, with regression line for partial events

Table 4.13. Calculations for correlation ratio test for hypothesis of linearity of the regression, partial events

Interval $\log_e$ (Yp_i -.009)	$\overline{\log_e Dp_i}$	$(\overline{\log_e Dp_i})^2$	n_i	$n_i(\overline{\log_e Dp_i})^2$
-5.7999 to -4.8000	3.2956	10.8611	5	54.3056
-4.7999 to -3.8000	3.6371	13.2283	11	145.5111
-3.7999 to -2.8000	4.1039	16.8420	14	235.7879
-2.7999 to -1.8000	4.6488	21.6109	15	324.1631
-1.7999 to -0.8000	4.6815	21.9160	14	306.8236
-0.7999 to +0.1200	4.8942	23.9534	13	311.3941
TOTAL			72	1377.9854

$$e^2 = \frac{1377.9854 - 72(4.3450)^2}{71(.9247)^2} = .3084$$

$$\text{mean } \log_e Dp = 4.3450$$

$$\text{variance } \log_e Dp = (.9247)^2$$

$$r^2 = .3066 \ ; \ n = 72 \ ; \ k = 6$$

$$\frac{(e^2 - r^2)/(k - 2)}{(1 - e^2)/(n - k)} = \frac{(.3084 - .3066)/4}{(1 - .3084)/66} = \frac{.00045}{.0105} = .0429$$

$$F_{4,66} = 2.52 \text{ at the 95\% level}$$

at the 95% level.

The conditional density is written as follows:

$$g_{\log_e D_p | \log_e Y_{p'}}(\log_e d, \log_e y) = .3296 \log_e Y_{p'} + 5.1624 + \varepsilon(0, .7755) \tag{4.28}$$

A transformation is performed to obtain the duration Dp as follows:

$$Dp = \exp(\log_e Dp).$$

As a partial event has been defined as either ending or beginning at midnight, once the duration has been predicted, the time of occurrence within the day is specified.

CHAPTER 5

SIMULATION PROCEDURES

Introduction

Tests in Chapter 4 showed that certain aspects of rainfall processes can be described by appropriate mathematical models. Some parts of the model have already been tested (Chapter 4). A simulation was run to test the previously untested model hypotheses. If the simulated data compare favorably with the observed data, the model is considered reasonable.

Hypothesis testing procedures consist of goodness-of-fit tests applied to various distributions, when appropriate. Histograms of the actual and simulated data will be compared to test the joint distribution of number of events per day and daily amount, because an analytic expression could not be found for the marginal distribution of number of events per day. Marginal distributions will be compared for the amounts, durations, and starting times of the complete and partial rainfall events.

The computer program XDISAG, shown in Appendix D, reads in real or simulated daily rainfall amounts and disaggregates each into the number of events per day (NUM). It then simulates the amount (AMT), duration (DURATN) and starting time (STIME) of each event. Subroutines called include XAMTS, DUR, GENST, OVRLP, BETAMX, and BETAF. Some of the subroutines call other subroutines. Those called include XINTP, GAM, and TIME. A flowchart of the main program is shown in Appendix D.

The program was run on a Digital Equipment Corporation PDP-11 computer at the Southwest Rangeland Watershed Research Center in Tucson. It took 130 seconds to simulate 23 years of disaggregated daily data. Input consisted of the daily data used to construct the model: i.e., data from the Walnut Gulch Watershed for the months of July and August, years 1954 - 1976. A description of the simulation procedures follows.

Simulation of Rainfall Processes

Conditional Distribution of Number of Events per Day Given a Daily Amount

A year of daily amounts is read in and stored in the array Z(i,j), where i is the year, and j is the day. The disaggregation process is applied to each rainy day, j, in the season of interest.

Given a daily rainfall amount, Z_j, it is first necessary to simulate the type (complete or partial) and number of events which occurred on day j. The number of partials, NPDC, is generated first, followed by the total number of events, NUM. The number of complete events is NUM -NPDC.

For simulation purposes, a partial event is considered as the sum of two complete events. That this assumption is not unreasonable was shown in Chapter 4, pages 73 - 74. Thus, if a partial event is simulated at the end of day j, a partial event must also occur at the beginning of day j+1. The term "partial event" refers to the separate portions which occur on day j and on day j+1, unless specified otherwise.

Because a partial event can occur at the end of day j only if there is rain on day (j+1), Z_j and Z_{j+1} are looked at concurrently. Given two consecutive rainy days, Z_j, $Z_{j+1} \neq 0$, the probability of a partial event occurring at the end of day j is .1659. A uniform random number (CRAN) is generated. If CRAN $\leq$ .1659, a partial event occurred at the end of day j. If CRAN > .1659, no partial event occurred. The variable NPDCE, the number of partials at the end of the current day j, takes on the value zero or one, accordingly.

A partial event occurs at the beginning of day j only if a partial occurred at the end of day (j-1). The variable NPDBE, the number of partial events at the end of the previous day, (j-1), is zero if no partial event occurred, and one if a partial did occur. The total number of partials on day j is NPDC = NPDBE + NPDCE.

It now remains to simulate the number of complete events for day j. The conditional distribution of number of complete events given a daily amount, $G_{Nc|Zc'}(n|z)$, can be described by the shifted negative binomial distribution (SNBD):

$$P(N=n|Z'=z) = \binom{n+r-2}{n-1} p^r (1-p)^{n-1} \; ; \; n = 1, 2, 3, \cdots \tag{5.1}$$

where

$$\begin{aligned} p &= .7228 + (.2772)\exp(-5.7939Zc') \\ r &= 2.3097 - (1.3097)\exp(-9.5902Zc') \end{aligned} \tag{5.2}$$

Because only 36 partial events occurred in 23 years, and as a method has already been developed to simulate the number of partial events on day j, it was assumed that the scheme developed to simulate the number of complete events per day can also be used to simulate the total number of events per day, NUM.

To summarize, Equation (5.1), with the parameters as given in Equation (5.2), is used to simulate the total number of events on day j, NUM. The number of partial events, NPDC, is simulated using the method described in the preceding paragraphs. The number of complete events per day is NUM-NPDC, provided NUM $\geq$ NPDC. Otherwise, the number of complete events is zero.

Subroutine NCEVNT is used to generate the total number of daily events, with the upper limit set at six per day, the largest number observed in the record. The joint distribution of number of events per day, given a daily amount, is written as the product of the conditional and marginal distributions (see Equation (4.1)). As the marginal distribution was calculated for Z -.009, the conditional was also calculated using this transformation. Thus, .009 is subtracted from the total daily amount before the number of events is calculated, and is added back in afterward.

The transformed daily amount, Z -.009, is passed to the subroutine NCEVNT, and the parameters p and r are calculated according to Equation (5.2). For each daily amount, the probabilities that NUM = 1, 2, 3, 4 and 5 are calculated as needed. A uniform random number (RAN1) is generated, and this is tested against each probability as follows. Let P1 = P(NUM=1), P2 = P(NUM=2), and so on. If RAN1 $\leq$ P1, then NUM = 1, and one event per day occurs. If P1 < RAN1 $\leq$ P2, NUM = 2, and so on. If P5 < RAN1, NUM = 6. The value of NUM, the total number of events on day j, is passed back to the main program.

A comparison between the observed and simulated data is shown in Table 5.1 and Figure 5.1. A visual observation indicates a good fit.

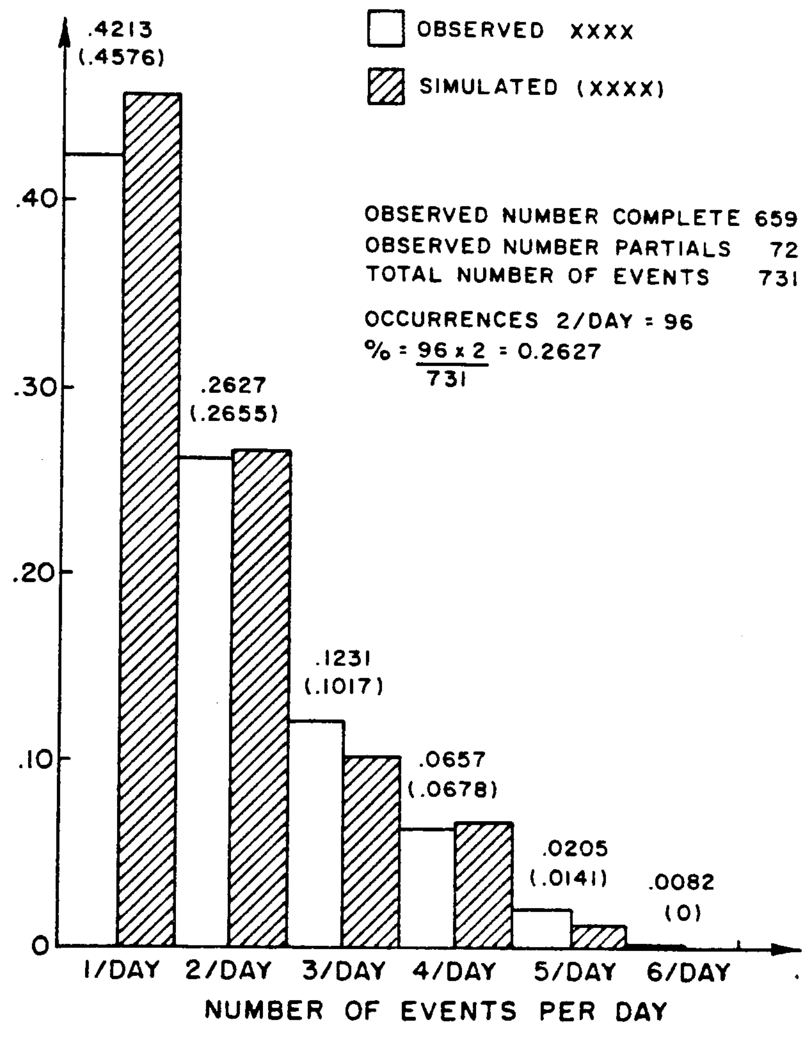
The Disaggregation of a Daily Rainfall Amount into N Amounts

Subroutine XAMTS is used to disaggregate the daily rainfall

Table 5.1. Occurrences of type and number of events per day for observed and simulated data (simulated data in parentheses)

Number of occurrences	Complete 659 (642)	*Partial 36 (33)
One/day	308 (324)	35 (29)
Two/day	96 (94)	1 (4)
Three/day	30 (24)	
Four/day	12 (12)	
Five/day	3 (2)	
Six/day	1 (0)	

*A partial event here refers to both portions on day j and (j+1).



NOTE: A partial event refers only to that portion which occurred on day j

Figure 5.1. Histogram of number of events per day

amount Z into N amounts. The storm ratios R_1 through R_6, as given in Chapter 4, are used to disaggregate Z, and the distributions used include the uniform and beta-Fourier.

As the random number generator in the computer generates variables which are uniformly distributed, it is necessary to transform a uniform random variable to a beta-Fourier distributed random variable. The inverse transformation method is used as follows.

The CDF $F_R(r)$ is defined over the range zero to one. A uniform random number, say x, is generated, and $F_R(r)$ is set equal to this number : $x = F_R(r)$. The variable r (or, in this case, storm ratio) is uniquely determined by the number generated, and for every x_0 which is generated, it is possible to find the corresponding r_0 by the inverse function of F, if it is known: $r_0 = F^{-1}(x_0)$. $F^{-1}(x)$ is the mapping of x into the domain of r (Naylor, Balintfy, Burdick and Chu, 1966) (see Figure 5.2).

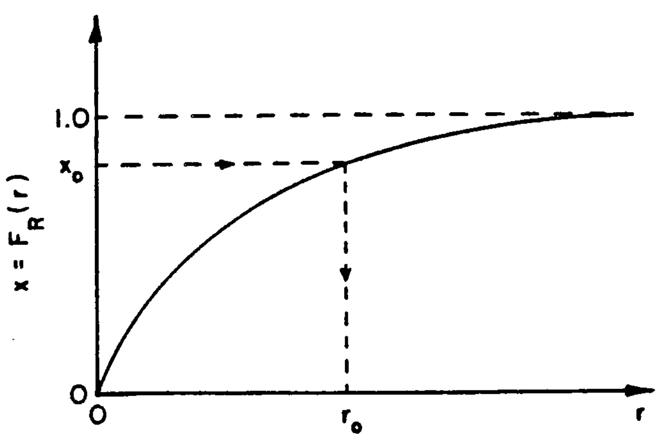


Figure 5.2. Graphical representation of the inverse transformation method

The CDF values for the beta-Fourier distribution cannot be found using an analytic expression. However, they can be obtained by numerically integrating the density function. The procedure is as follows.

The density function in Equation (4.14) is numerically integrated from zero to one in increments of .0001. For every value of the CDF which is a multiple of .02, that value, $F_R(r)$, and its corresponding value $r = F_R^{-1}(r)$, are stored in the arrays YVAR and XVAR, respectively. The subroutine BETAF, called at the beginning of XDISAG, calculates all these values and stores them in the stated arrays. The arrays are passed to the subroutine XAMTS, as well as the number of events per day and the daily rainfall amount. The r values corresponding to the inverse transformation $r = F_R^{-1}(r)$ are obtained from interpolating between the appropriate r values corresponding to the values which bracket $x_0 = F_R(r)$.

An example of the method used to generate a beta-Fourier distributed random variable is given below. Some stored values of the arrays XVAR and YVAR, where XVAR is the r value corresponding to $F_R(r)$, and YVAR is the value of $F_R(r)$, are:

XVAR(1) = 0	YVAR(1) = 0
XVAR(2) = .0222	YVAR(2) = .02
XVAR(3) = .0427	YVAR(3) = .04
XVAR(4) = .0623	YVAR(4) = .06
XVAR(5) = .0815	YVAR(5) = .08
XVAR(6) = .1006	YVAR(6) = .10

Assume the uniform random number x_0 = .0768 is generated. Let Y = .0768/.02 = 3.84 (the denominator of Y is the interval at which the CDF values are tabled). Let Q be the integer value of Y; in this case, Q = 3. If one looks at the (Q+1) and (Q+2) values in YVAR (.06 and .08, respectively), it is only necessary to interpolate between the corresponding values in XVAR to get the desired r_0 corresponding to $x_0 = F_R(r) = .0768$. In this case, $r_0 \doteq .0784$. This method is used to generate beta-Fourier distributed random variables.

The disaggregation scheme is as follows. A beta-Fourier distribution is used to describe R_1 and R_2 where:
$R_1 = \frac{Y_1}{Z}$ on days in which two events occur, and $R_2 = \frac{Y_2 + Y_3}{Z}$ on days in which three events occur. A uniform distribution is used for all other disaggregation processes.

On days in which two events occur, a uniform random variable is generated and transformed to a beta-Fourier distributed random variable, and this is the value of R_1. The amount of the first event is then $Y_1 = R_1 Z$, and the amount of the second event is $Y_2 = Z - Y_1$.

On days in which three events occur, the amount of the second plus third events is $Y_2 + Y_3 = R_2 Z$, where R_2 has been generated and transformed as described above. The amount of the first event is $Y_1 = Z - R_2 Z$. The amount of the second event is $Y_2 = R_3(R_2 Z)$ where R_3 is uniformly distributed and $R_3 = \frac{Y_2}{Y_2 + Y_3}$. The amount of the third event is then $Y_3 = Z - Y_1 - Y_2$.

For four events per day, the storm ratio R_4 is uniformly distributed, and $R_4 = \frac{Y_3 + Y_4}{Z}$. The sum of the amounts of the third plus fourth events is SUM34 = $R_4 Z$. The sum of the amounts of the first plus second events is SUM12 = Z - SUM34. The ratios R_5 and R_6, where $R_5 = \frac{Y_1}{Y_1 + Y_2}$ and $R_6 = \frac{Y_3}{Y_3 + Y_4}$, are also uniformly distributed. Thus, the amounts of the events are as follows: $Y_1 = R_5 SUM12$; $Y_2 = SUM12 - Y1$; $Y_3 = R_6 SUM34$; and $Y_4 = SUM34 - Y_3$.

On days in which five events occur, a uniform random number, say RAN1, is generated and set equal to the ratio of the sum of the amount of the first two events to the daily total amount:

$$RAN1 = \frac{Y_1 + Y_2}{Z} .$$

The sum of the amounts of the first two events is: SUM12 = RAN1•Z. These amounts are then disaggregated according to the scheme already developed for two events per day. The sum of the amounts of the third, fourth, and fifth events is SUM345 = Z - SUM12, and these are disaggregated according to the scheme developed for three events per day.

In a similar manner, on days in which six events occur, a uniform random number, say RAN2, is used to predict the sum of the first three event amounts: SUM123 = RAN2•Z. Then SUM456 = Z-SUM123. SUM123 and SUM456 are each disaggregated using the method developed for three events per day. The disaggregated amounts are passed back to the main program in the array AMT.

The CDF's of the observed and simulated event amounts are shown in Figures 5.3 and 5.4, respectively, for the complete and partial events. Let H_0 be the hypothesis that the observed and simulated distribution of the amounts of the complete events come from the same parent distribution. A Kolmogorov-Smirnov goodness-of-fit test is used, and the statistic $|D_2|$ is the maximum value of the difference between the two CDF's. As the test is not whether a certain sample comes from a stated theoretical distribution, but rather if two samples are from the same (unspecified) population, the test statistic must be adjusted

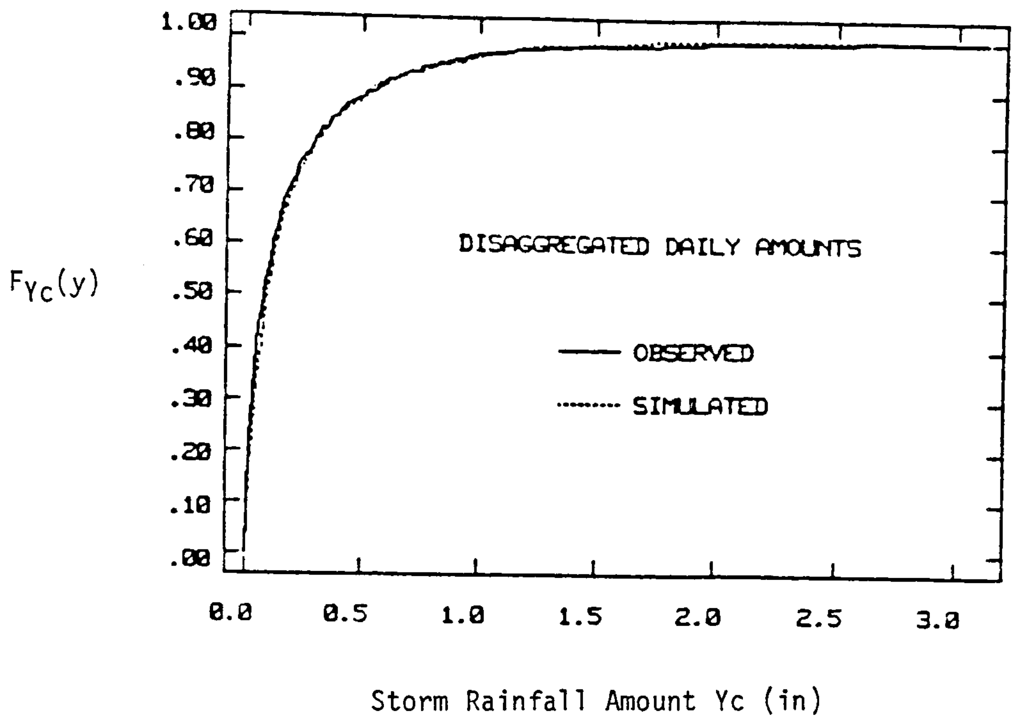


Figure 5.3. Observed and simulated distributions of storm rainfall amount for complete events

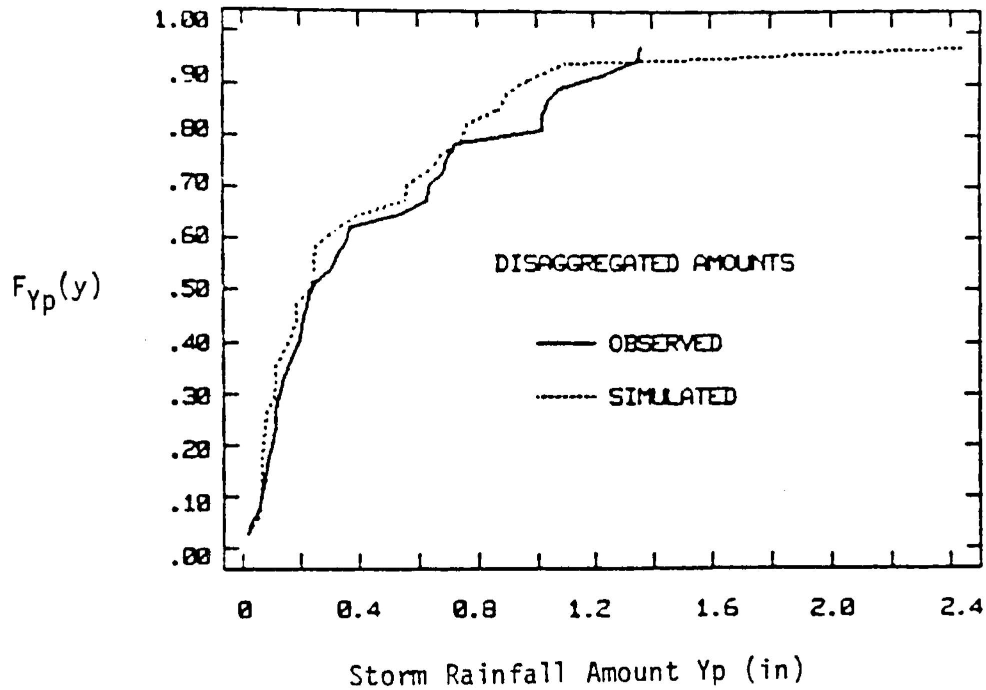


Figure 5.4. Observed and simulated distributions of storm rainfall amount for partial events

accordingly. At α = 0.05, the critical statistic is $1.36/\sqrt{n}$; and the and the value of n used is $n = \frac{n_1 n_2}{n_1 + n_2}$, where n_1 and n_2 are the sizes of the two samples (Benjamin and Cornell, 1970, p. 510). In this case, $n = \frac{659(642)}{659 + 542} = 325.1945$; and the critical statistic at α = 0.05 is .0754. As the maximum $|D_2|$ observed is .0748, H_0 cannot be rejected at the 95% confidence level.

Each partial event is assumed to consist of the sum of both portions occurring on days j and j + 1. Let H_0 be the hypothesis that the observed and simulated distributions of partial event amounts come from the same population. The value of n is $n = \frac{36 \times 33}{36 + 33} = 17.2174$. The critical statistic, at α = 0.05, is .3278. As the largest observed value of $|D_2|$ is .1105, H_0 cannot be rejected at the 95% confidence level. The observed and simulated marginal distributions of rainfall event amount compare favorably.

Joint Distribution of Event Amount and Duration

Once the rainfall amounts have been simulated for each of the N events in a day, it is necessary to assign a duration to each event. Because the joint density function is written as the product of the conditional and marginal density functions (see Equation (4.15)), and the marginal density function was obtained for the transformed variable Y -.009, where Y represents the event rainfall amount, the conditional density was also obtained for the transformed variable. Thus, .009 is subtracted from each event amount before the subroutine DUR is called, and is added in afterward.

Subroutine DUR generates the duration associated with each event amount. As it was shown in Chapter 4 that partial events (separate portions on days j and j+1) have longer durations than complete events, but can be described in an analogous manner, the same scheme is used for each, but the parameters differ accordingly.

The array containing the values of the disaggregated event amounts on day j is passed to subroutine DUR, as well as the number of partial events occurring. The expected value of the $\log_e$ of each duration is obtained as follows: $E[\log_e(\text{duration})] = A + B \log_e (\text{amount})$, with the parameters as specified in Chapter 4. Because the conditional distribution of the $\log_e$ (duration) is normally distributed about the expected value of the $\log_e$ (amount), a standard normal random variate is generated according to the following equation: $x = (-2 \log_e r_1)^{1/2} \cos(2\pi r_2)$, where r_1 and r_2 are two uniformly distributed independent random variables (Naylor, Balintfy, Burdick and Chu, 1966, p. 95). This value of x is multiplied by the standard error of estimate of the regression, and the product is added to the expected value of the $\log_e$ (duration) already obtained. This product is actually the residual value about the regression line as given by the expected value function. The value obtained when the residual is added to the expected value is the $\log_e$ (duration), which is transformed to get the predicted event duration: $\exp(\log_e \text{duration}) = \text{duration}$.

The same procedure is used for each event amount on day j, and the corresponding event durations are passed back to the main program in the array DURATN.

The CDF's of the observed and simulated event durations are shown in Figures 5.5 and 5.6, respectively, for the complete and

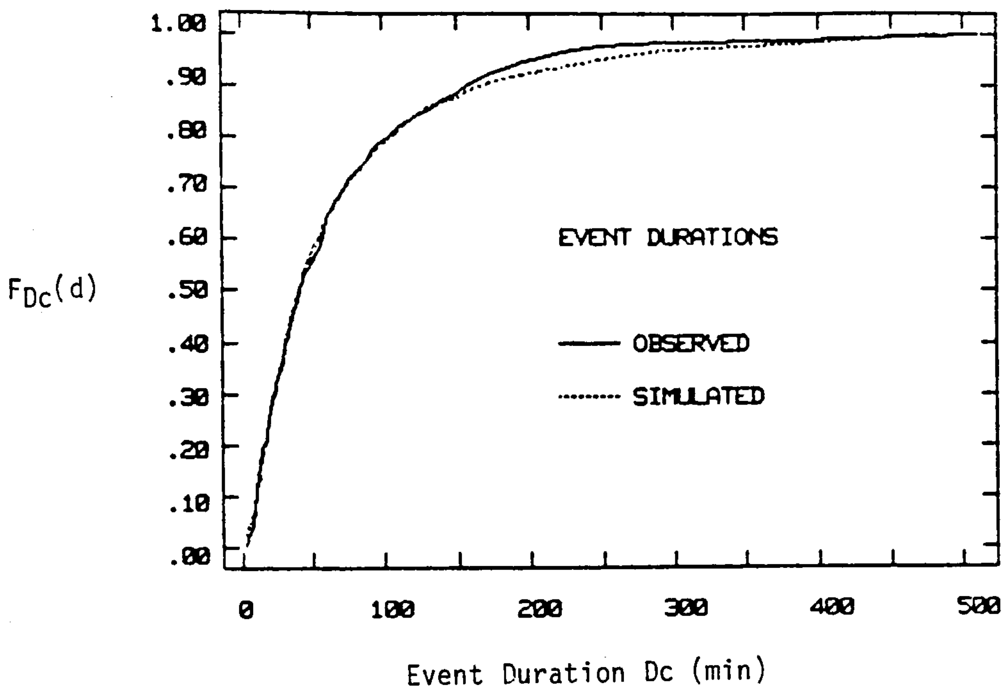


Figure 5.5. Observed and simulated distributions of event duration for complete events

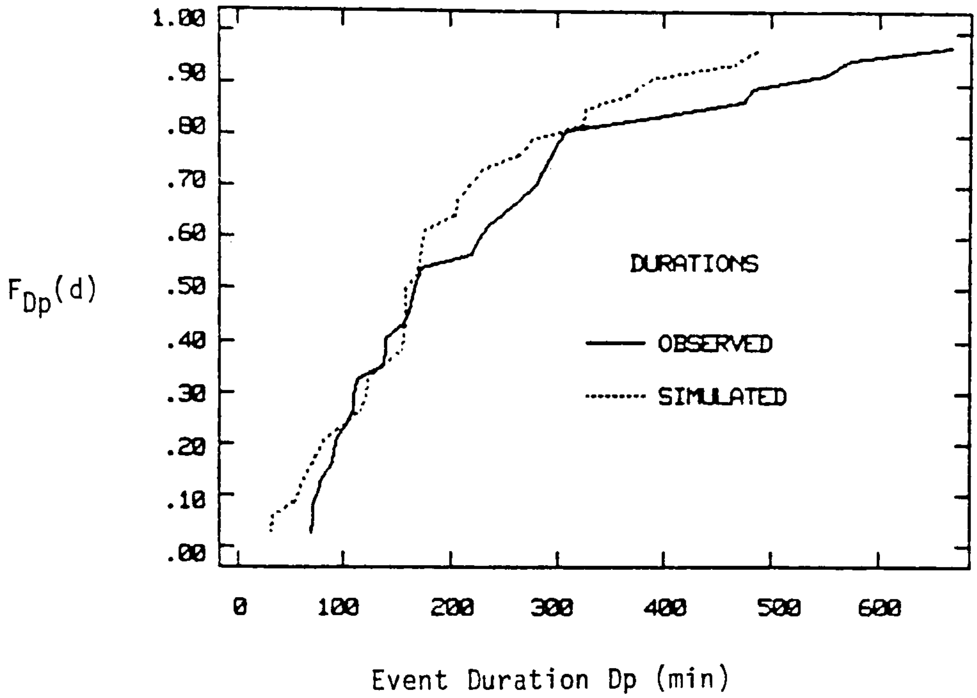


Figure 5.6. Observed and simulated distributions of event duration for partial events

partial events. Let H_0 be the hypothesis that the observed and simulated marginal distributions of durations of complete events come from the same parent distribution. At $\alpha = 0.05$, the critical statistic, using $n = \frac{659(642)}{659 + 642}$, is .0754. The maximum $|D_2|$ observed is .027. As this value is less than the critical statistic, H_0 cannot be rejected at the 95% confidence level.

Let H_0 be the hypothesis that the 36 observed and 33 simulated partial event durations come from the same distribution. At $\alpha = 0.05$, and for $n = 17.2174$, the critical statistic is .3278. As the maximum $|D_2|$ observed is .1431, H_0 cannot be rejected at the 95% confidence level. Therefore, the observed and simulated durations compare favorably.

Event Starting Times

It has been shown in Chapter 4 that event starting times can be described by a mixed beta distribution, and that on multiple event days, the starting times can be obtained using order statistics. Subroutine GENST is used to generate event starting times. The number of events per day, as well as the number of partials and when they occur (at the beginning or end of the day), are passed from the main program. Subroutine BETAMX, called at the beginning of XDISAG, is used to calculate values of the CDF of the mixed beta distribution in a similar method as the values for the beta-Fourier distribution were obtained. The CDF values for the mixed beta, stored in the arrays XVAL and YCDF, are also passed to the subroutine GENST.

The inverse transformation method, as described in the section on disaggregating event amounts, is used to generate a mixed beta distributed random variable. As the variable generated is defined over the range zero to one, the values obtained are multiplied by 24 to get the event starting times in military hours.

Because the starting times are generated in an independent manner (independent of each other and of event durations), subroutine OVRLP (overlap) was introduced to ensure that no two events occur simultaneously. It also ensures that no complete event runs over midnight, partial events at the end of the day end at midnight, and partial events at the start of the day begin at midnight.

The following situations can arise, and must be corrected in subroutine OVRLP:

(1) On a day in which there is no partial event at the end, a complete event may run over midnight.

(2) There may be less than 10, but greater than zero, minutes of no rainfall between events.

(3) An event may end after the beginning time of the next event.

If a complete event runs over midnight, the starting time is moved to an earlier time in the day, such that the event ends 5 minutes before midnight.

If the time between events is less than 10 minutes, or if the events overlap, the starting time of the earlier event is moved back in the day so that a 10-minute interval exists between the end of the earlier event and the beginning of the following one.

During subroutine OVRLP, a shift to an earlier starting time may cause an event to begin on the previous day. If it is the first event of the day, the starting time is set at zero hours (military time), and the following event is made to begin 10 minutes after the first event ends.

If there is an overlap introduced due to the procedures in the paragraph above, or if the event which has a starting time on the previous day is not the first event of the day, the following occurs: the starting time of the first event is set at zero hours. All subsequent events in the day are made to begin 20 minutes after the end of the most recent event, unless there is a partial event at the end of the day. In this case, the same procedure is used, but the partial event is made to begin so that it ends at midnight.

The procedures in OVRLP, while shifting some starting times, maintain the number of events per day and event durations.

It has been shown in Chapter 4 that the distribution of event starting times can be described by a mixed beta. However, it is necessary to determine if subroutine OVRLP significantly affects starting times on multiple-event days.

Kolmogorov-Smirnov goodness-of-fit tests were run to test the hypotheses that, on multiple event days and for complete events, the observed and simulated starting times come from the same distributions. Table 5.2 lists the value of n used, the critical statistic at $\alpha = 0.05$, and the largest $|D_2|$ value observed for each event on days in which 2, 3, and 4 events per day occur. In each case, as the largest observed

Table 5.2. Kolmogorov-Smirnov goodness-of-fit tests for observed and simulated starting times

Number of events per day	Event number in day	$n = \frac{n_1 n_2}{n_1 + n_2}$	Critical K-S static at $\alpha = 0.05$	Maximum $\|D_2\|$ observed
2	1	47.4947	.1973	.1435
	2	47.4947	.1973	.0884
3	1	13.33	.3725	.3111
	2	13.33	.3725	.3491
	3	13.33	.3725	.2995
4	1	6.0	.5552	.1077
	2	6.0	.5552	.2685
	3	6.0	.5552	.3349
	4	6.0	.5552	.2131
	Partial event at day's end	17.2174	.3278	.1775

value of $|D_2|$ is less than the critical statistic, the hypotheses cannot be rejected at the 95% confidence level. The observed and simulated CDF's of event starting times are shown in Figures 5.7 through 5.15

A K-S test was also run to test the hypothesis that the observed and simulated distributions of partial event starting times come from the same parent distribution. In this case, a partial event comprises both portions on day j and day j + 1, and only the starting time on day j is examined. Again, at $\alpha = 0.05$, the hypothesis cannot be rejected. The statistics are given at the bottom of Table 5.2, and the observed and simulated CDF's are shown in Figure 5.16.

It is therefore concluded that subroutine OVRLP does not significantly alter the distribution of event starting times.

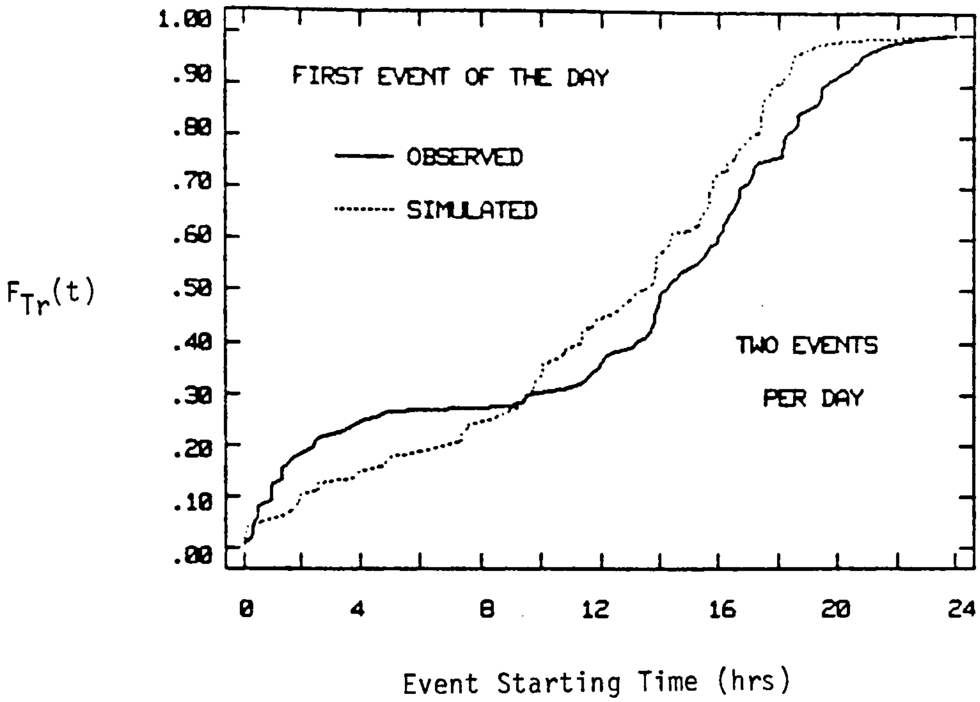


Figure 5.7. Observed and simulated distributions of event starting time, for the first event, on days in which two complete events occur

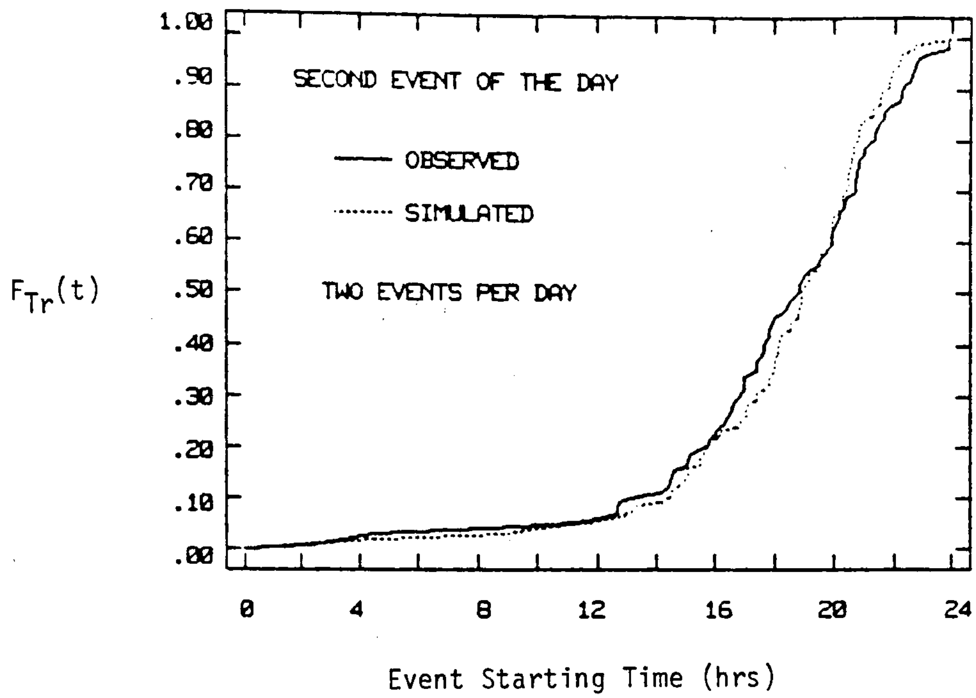


Figure 5.8. Observed and simulated distributions of event starting time, for the second event, on days in which two complete events occur

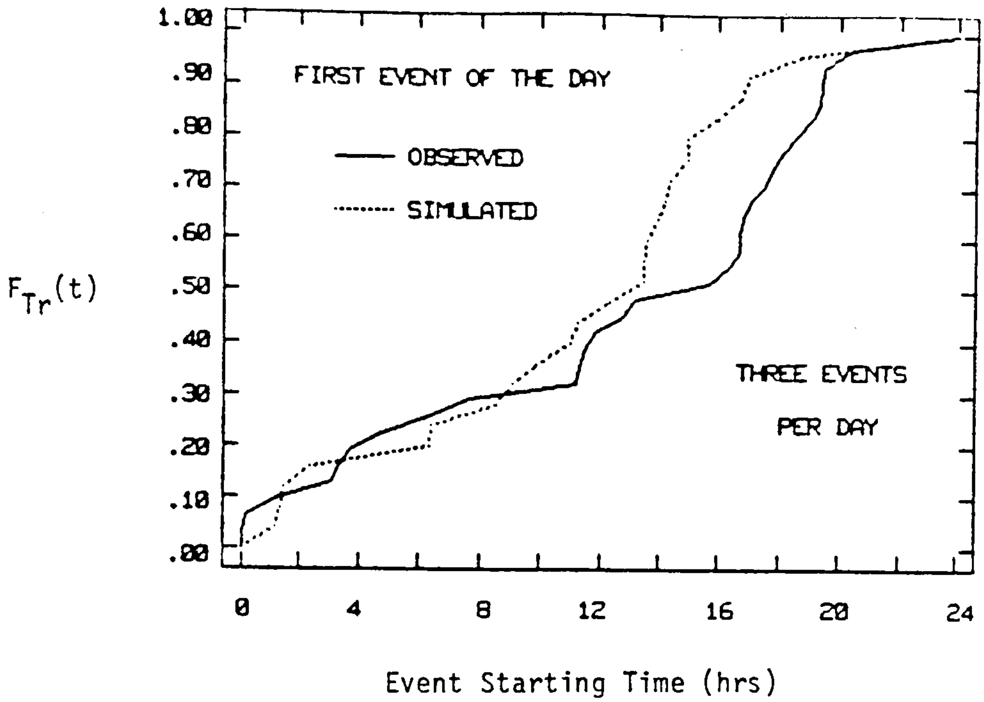


Figure 5.9. Observed and simulated distributions of event starting time, for the first event, on days in which three complete events occur

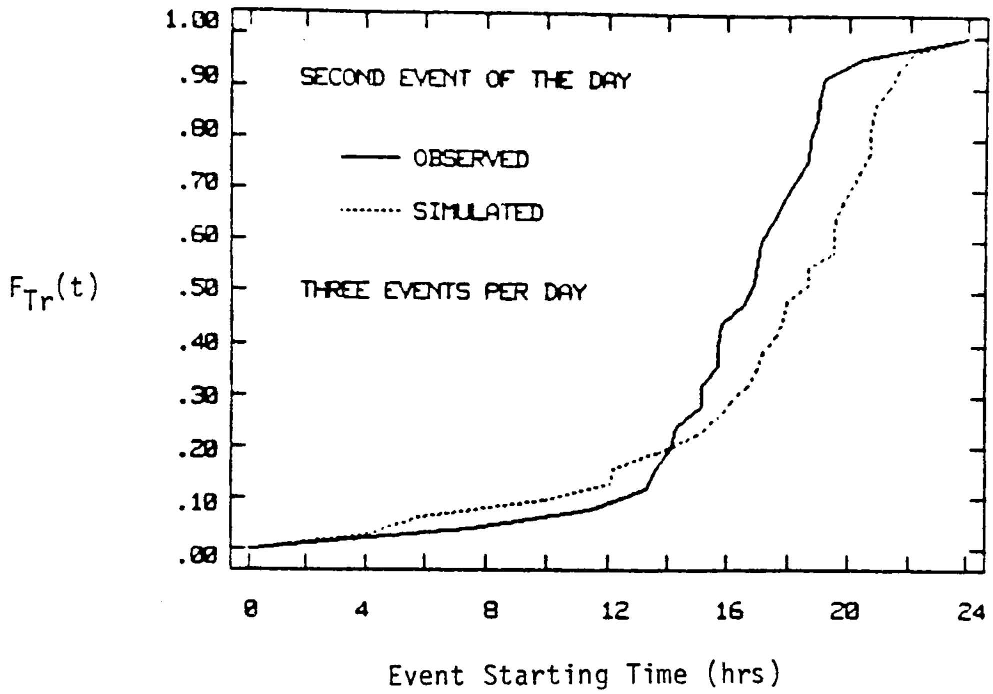


Figure 5.10. Observed and simulated distributions of event starting time, for the second event, on days in which three complete events occur

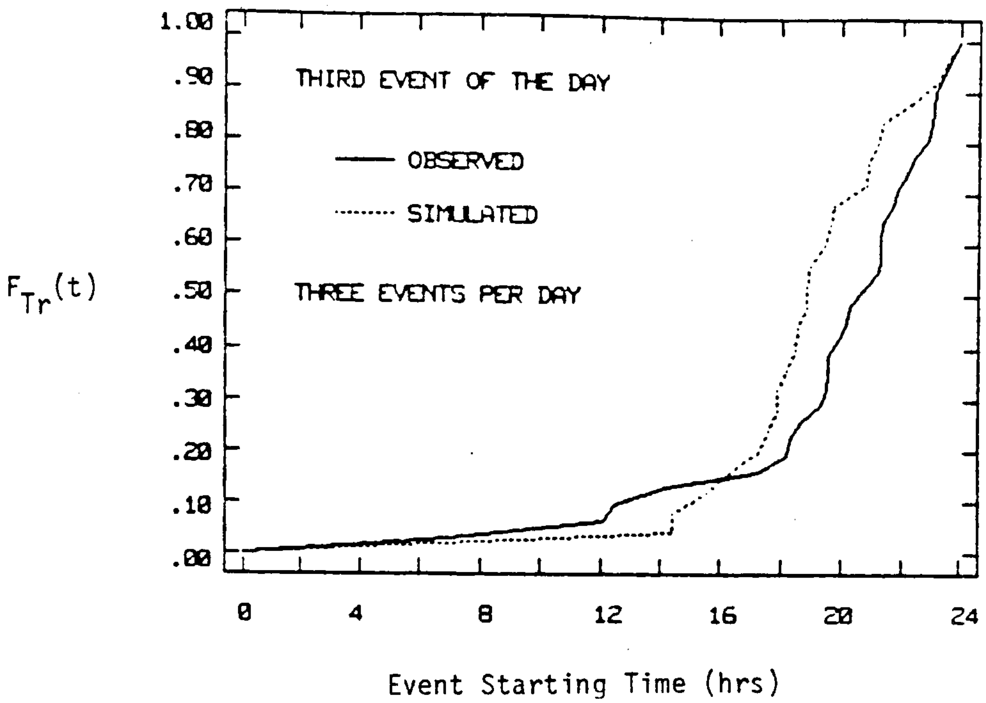


Figure 5.11. Observed and simulated distributions of event starting time, for the third event, on days in which three complete events occur

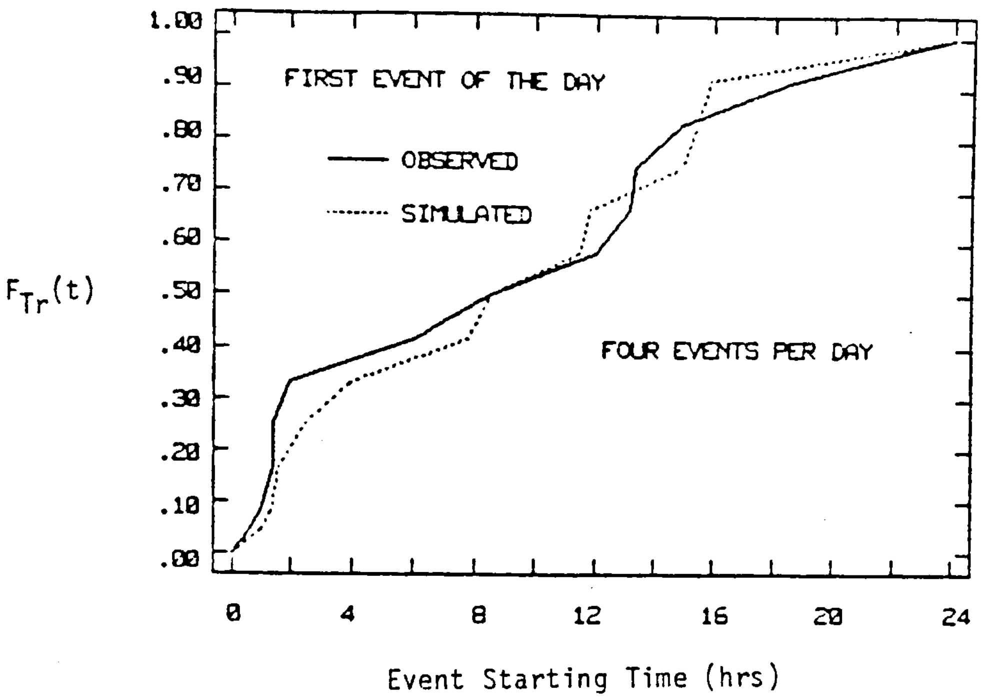


Figure 5.12. Observed and simulated distributions of event starting time, for the first event, on days in which four complete events occur

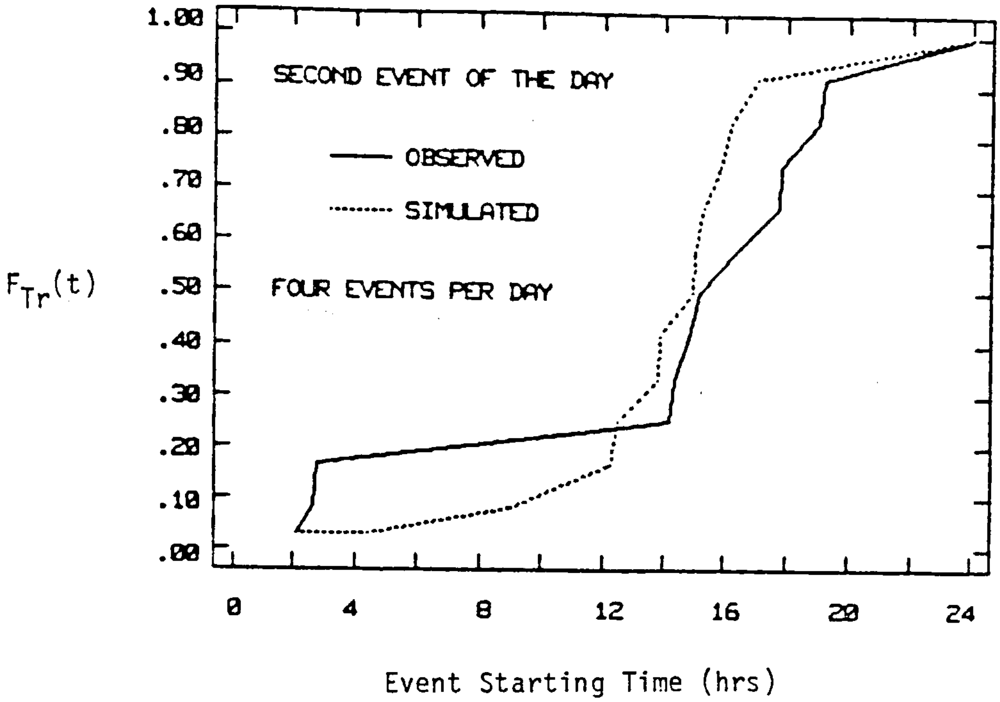


Figure 5.13. Observed and simulated distributions of event starting time, for the second event, on days in which four complete events occur

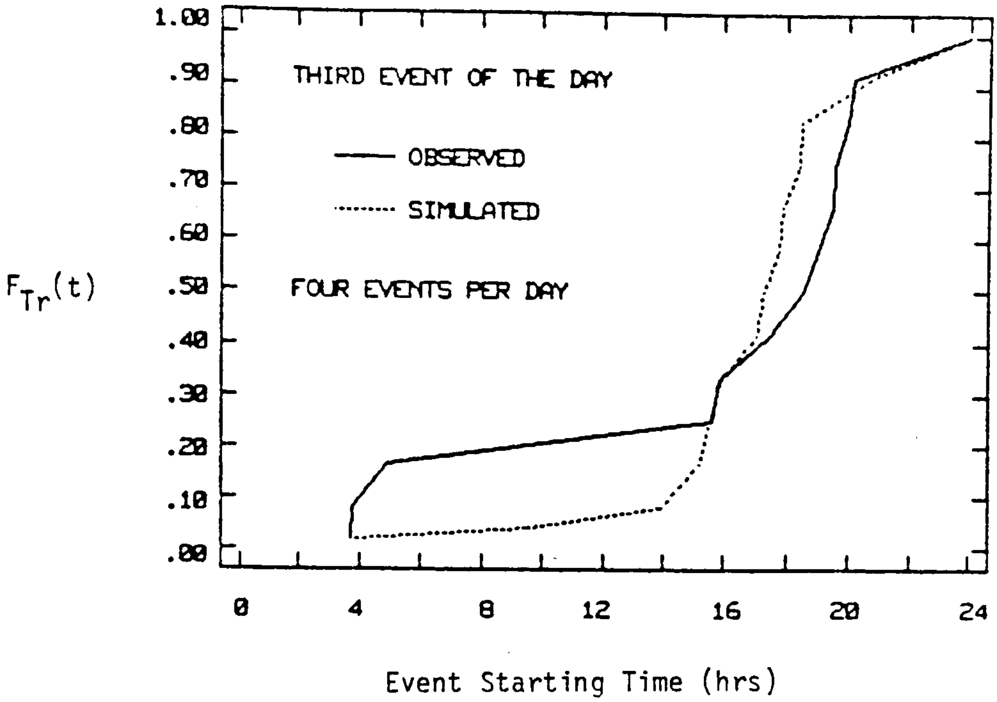


Figure 5.14. Observed and simulated distributions of event starting time, for the third event, on days in which four complete events occur

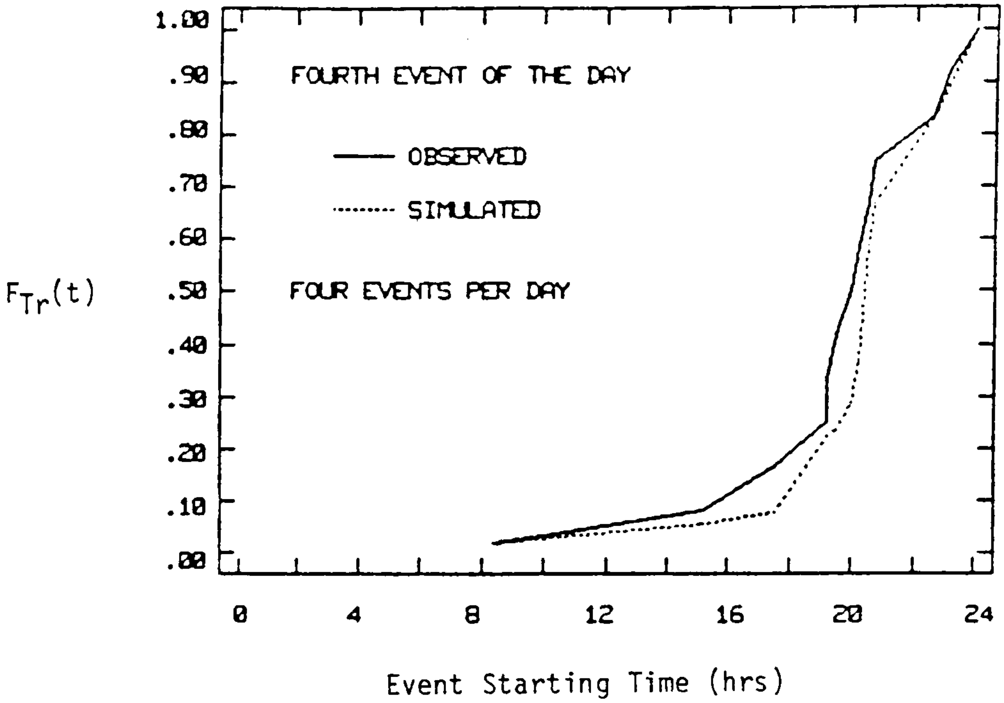


Figure 5.15. Observed and simulated distributions of event starting time, for the fourth event, on days in which four complete events occur

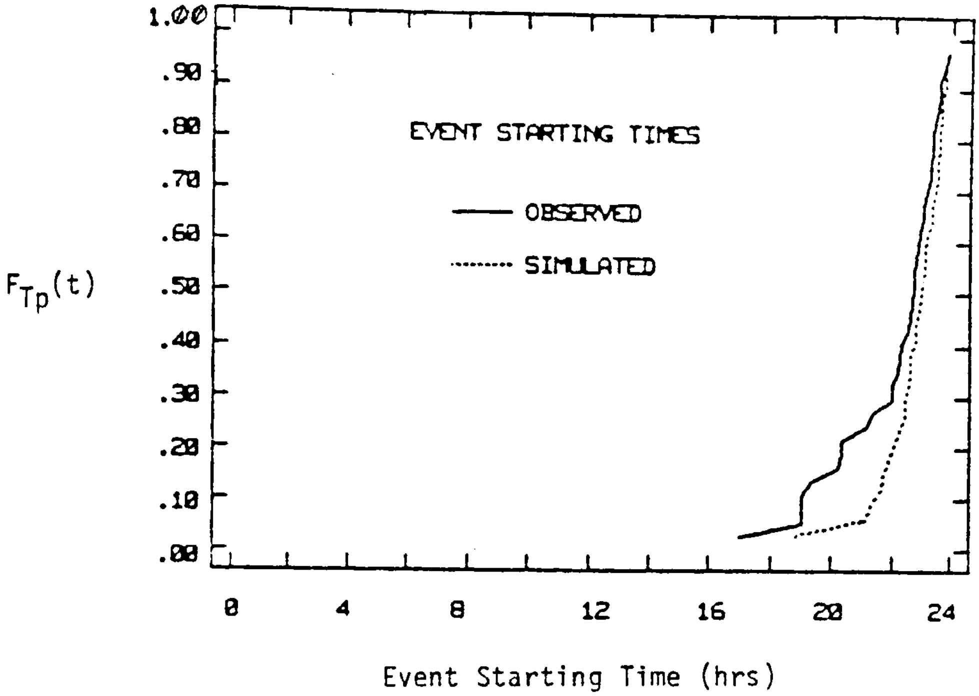


Figure 5.16. Observed and simulated distributions of event starting time for partial events which begin at the end of the day

CHAPTER 6

DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS

Discussion

The objective of this research was to develop a model to simulate the number of storm rainfall events per day and the amount, duration, and starting time of each event, given the daily rainfall amount and the amounts on the preceding and following days. Simulated distributions for each give a good fit to the observed data, as was shown in Chapter 5.

A useful simulation model should be parameter-efficient, and should not require much computer-processing time. Although the number of parameters (19) is larger than desirable, it is not unreasonable, given the complexity of the processes studied. Also, the computer-simulation time (130 seconds for 23 years of summer data) is reasonable.

Model input consists of real or simulated daily rainfall amounts, and output consists of the number of events per day and the amount, duration, and starting time of each event. This output can be used as input for an intra-storm intensity model to generate sequences of breakpoint data (Woolhiser and Osborn, 1984). The breakpoint data can then be used as input to physically-based infiltration models, which provide better estimates of surface runoff and related processes than the empirical techniques often used.

It was hoped that the disaggregation processes developed would be spatially invariant for areas with a similar climate. To see if this is so, data from an appropriate location should be obtained, model parameters adjusted accordingly, and the results of the simulation compared with the observed data.

Some model inconsistencies are apparent. The conditional distribution of event duration, given an amount, assumes the marginal distribution of event amount to be constant for all events. However, the use of the storm ratios R_1 through R_6, for the disaggregation of the daily amount into individual event amounts, indicates that the distribution of storm amount varies on multiple-event days.

Also, the distribution of the sums of event amounts should be consistent with the distribution of the daily amount. For certain distributions, it is relatively easy to show this is true. However, for the distributions in question (uniform, beta-Fourier, and Weibull), such a proof quickly leads to intractable expressions.

Finally, it is assumed that the storm ratios R_1 through R_6 are mutually independent. However, for the case of three events per day, the ratios used are:

$$R_2 = \frac{Y_2 + Y_3}{Z} \; ; \; R_3 = \frac{Y_2}{Y_2 + Y_3}$$

As the same random variable $(Y_2 + Y_3)$ is used, both in the numerator of R_2 and in the denominator of R_3, this assumption is questionable. The same issue can be raised concerning the use of ratios in the case of four, five, and six events per day. Again, although methods are available to show independence, the problem becomes intractable for the distributions considered.

Because the methods discussed in Chapter 4 provide reasonable results and are not too complex, and with the understanding that those inconsistencies are present, it was decided to maintain these structures in the model.

Some other shortcomings are noted as follows. The hypotheses of a linear relationship between the logs of duration and amount for complete and partial events are weak. Also, the small sample size of the multiple-event days for four, five, and six events per day (12, 3, and 1 occurrences, respectively) weakens the validity of the assumptions made concerning these events.

Conclusions

This study indicates that it is feasible to model the disaggregation of daily rainfall amounts into individual storms, and to simulate the amount, duration, and starting time of each event. When combined with an intra-storm intensity model, a reasonable method of simulating breakpoint data is provided.

Although many shortcomings are present, it is felt that they are necessary, both to prevent the model from becoming too complex and to restrict the number of parameters. No simple mathematical tools are presently available to model many of the daily rainfall processes. As a first attempt to model the disaggregation of daily rainfall, the procedures described give reasonable results.

Recommendations

The model could be improved if a simpler method could be found to describe the conditional distribution of number of events per day, given a daily amount. Parameter-estimating techniques which were used are inexact, time-consuming, and difficult to evaluate, statistically.

An improvement could also be made if a stronger relationship could be found between event duration and amount. Furthermore, the data for storm amounts could be broken into classes based on magnitude, and the durations within each class examined for any differences which might be present.

To test the hypothesis of spatial invariance of the disaggregation processes, data should be obtained from several other climatologically-similar locations, new parameters estimated, and the simulation results compared with the observed data. Different distributions and parameter-estimating techniques may be required.

Finally, another test of the model's usefulness would be to combine it with an intra-storm intensity model, thus resulting in simulated breakpoint data which could be used as input to runoff models. The simulated results could then be compared with the observed runoff data to test the rainfall-runoff model sensitivity to short-time period rainfall inputs.

The disaggregation of daily rainfall data has many applications. Some potential uses include: the modeling of storm runoff in an urban setting and the associated design of storm drainage systems; the prediction of expected rates of infiltration, evaporation and evapotranspiration and the availability of water to seeds and seedlings; and such runoff-related processes as the movement of pesticides and nutrients from agricultural areas.

APPENDIX A

THE LOG LIKELIHOOD FUNCTION

The (log) likelihood function gives the relative likelihood of having observed a particular sample as a function of a specified parameter set. If θ represents a set of parameters of the distribution of X, and $x_1, x_2, \cdots, x_n$ represent a sample of X, then the likelihood function is written as:

$$L(\theta | x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f_X(x_i | \theta)$$

and the log likelihood function is:

$$\log[L(\theta | x_1, x_2, \cdots, x_n)] = \log[\prod_{i=1}^{n} f_X(x_i | \theta)] = \sum_{i=1}^{n} \log[f_X(x_i | \theta)]$$

The (log) likelihood function is used to determine parameter estimates, the best estimate of θ being the value $\hat{\theta}$ which causes the (log) likelihood function to be a maximum. Similarly, when choosing between competing distributions, the best choice is the distribution which causes the (log) likelihood function to be a maximum, assuming that the distributions have the same number of parameters.

One technique frequently used to find a maximum for a parameter value is to take the partial derivative of the log likelihood function with respect to that parameter, set the expression equal to zero, and solve for the parameter. This method often leads to intractable expressions, however, and numerical optimization techniques are frequently used.

APPENDIX B

DISCUSSION ON HYPOTHESIS TESTING FOR THE CONDITIONAL PROBABILITY OF NUMBER OF EVENTS PER DAY GIVEN DAILY AMOUNT

Let H_0 be the hypothesis we wish to test, that the observed sample of number of events per day, given daily amount, is a random sample from a population described by a shifted negative binomial distribution (SNBD), with the parameters p and r as follows:

$$P(N_C = n|Z_C' = z) = \binom{n+r-2}{n-1}p^r(1-p)^{n-1};\ n = 1, 2, 3, \cdots$$

where

$$p = .7228 + .2772 \exp[-5.7939Zc']$$

$$r = 2.3097 - 1.3097 \exp[-9.5902Zc']$$

Let H_1 be the alternate hypothesis that the observed values do not come from such a population.

One method commonly used to test the validity of H_0 is to examine the goodness-of-fit between the observed and hypothesized distributions, and the two tests most often used for this purpose are the chi-squared (χ^2) and Kolmogorov-Smirnov (K-S) goodness-of-fit tests. The section below discusses why these tests are not applicable to the testing of H_0.

The χ^2 statistic D_1 is related to the histogram's deviations from the predicted values. It has (k-r-1) degrees of freedom in which k is the number of classes considered, and r is the number of parameters estimated from the data. A guideline often used in determining k is to choose intervals for which the expected number of occurrences is equal to, or greater than, 5. In this case, because the number of observed classes is 6, and the last two classes of 5 and 6 events per day should be lumped together due to their infrequent occurrence, k takes on the value of 5. As there are 4 parameters estimated from the data, 2 each for p and r, the number of degrees of freedom is equal to 5-4-1 = 0, which is meaningless. The χ^2 test, then, is not applicable in this situation for two reasons: because the number of allowable classes is small, and the number of occurrences in the higher classes is small.

The K-S statistic D_2 concentrates on the deviations between the observed and hypothesized cumulative distribution functions. The test is strictly valid only for continuous distributions and when the model is hypothesized wholly independent of data (Benjamin and Cornell, 1970). There is no quantitative method to adjust the distribution in a manner analogous to subtracting the degrees of freedom in the χ^2 test; however, the same method is often adopted. For the same reasons as mentioned in the χ^2 test, the K-S test is also not applicable to the stated problem.

Due to the limitations imposed by the data, and because the parameters were estimated in an unusual manner, an appropriate statistical test could not be found. Based on a visual observation of the observed and hypothesized histograms, it was decided not to reject H_0.

APPENDIX C

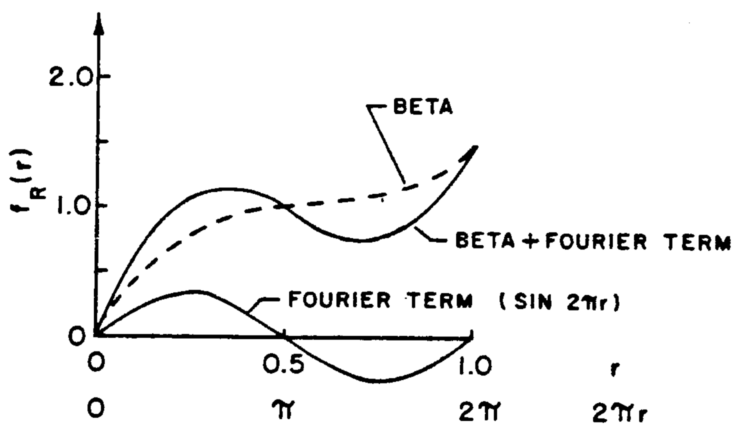
THE BETA-FOURIER DISTRIBUTION

The beta-Fourier density function is written as:

$$f_R(r) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} r^{\alpha-1}(1-r)^{\beta-1} + \theta \sin 2\pi r \;;$$

$$\alpha,\ \beta > 0 \;;\; 0 \leq r \leq 1$$

and may be thought of as adding two curves together.



The diagram above is a schematic, and is not drawn to scale.

A proper density function must satisfy two requirements:

(1) $f_R(r) \geq 0$

(2) $\int_0^{\infty} f_R(r)dr = 1$

To show (1), note that $f_R(r)$ will always be greater than zero if:

$$|\theta \sin 2\pi r| < \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} r^{\alpha-1}(1-r)^{\beta-1}$$

To show (2), we examine the beta-Fourier density function in two parts:

$$\int_0^{\infty} f_R(r)dr = \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} r^{\alpha-1}(1-r)^{\beta-1}dr + \int_0^1 \theta \sin 2\pi r \, dr$$

The first expression on the right side of the equal sign is the beta distribution, and this is equal to one. The second expression is evaluated as follows:

$$\int_0^{\infty} \theta \sin 2\pi r \, dr = \theta \int_0^{r=1} \sin 2\pi r \, dr = \frac{\theta}{2\pi} \int_{r=0}^{r=1} \sin 2\pi r \, 2\pi dr$$

$$= \frac{\theta}{2\pi} \left[-\cos 2\pi r \Big|_{r=0}^{r=1}\right] = \frac{\theta}{2\pi} \left[-\cos 2\pi + \cos 0\right]$$

$$= \frac{\theta}{2\pi} \left[-1 + 1\right] = \frac{\theta}{2\pi} (0) = 0$$

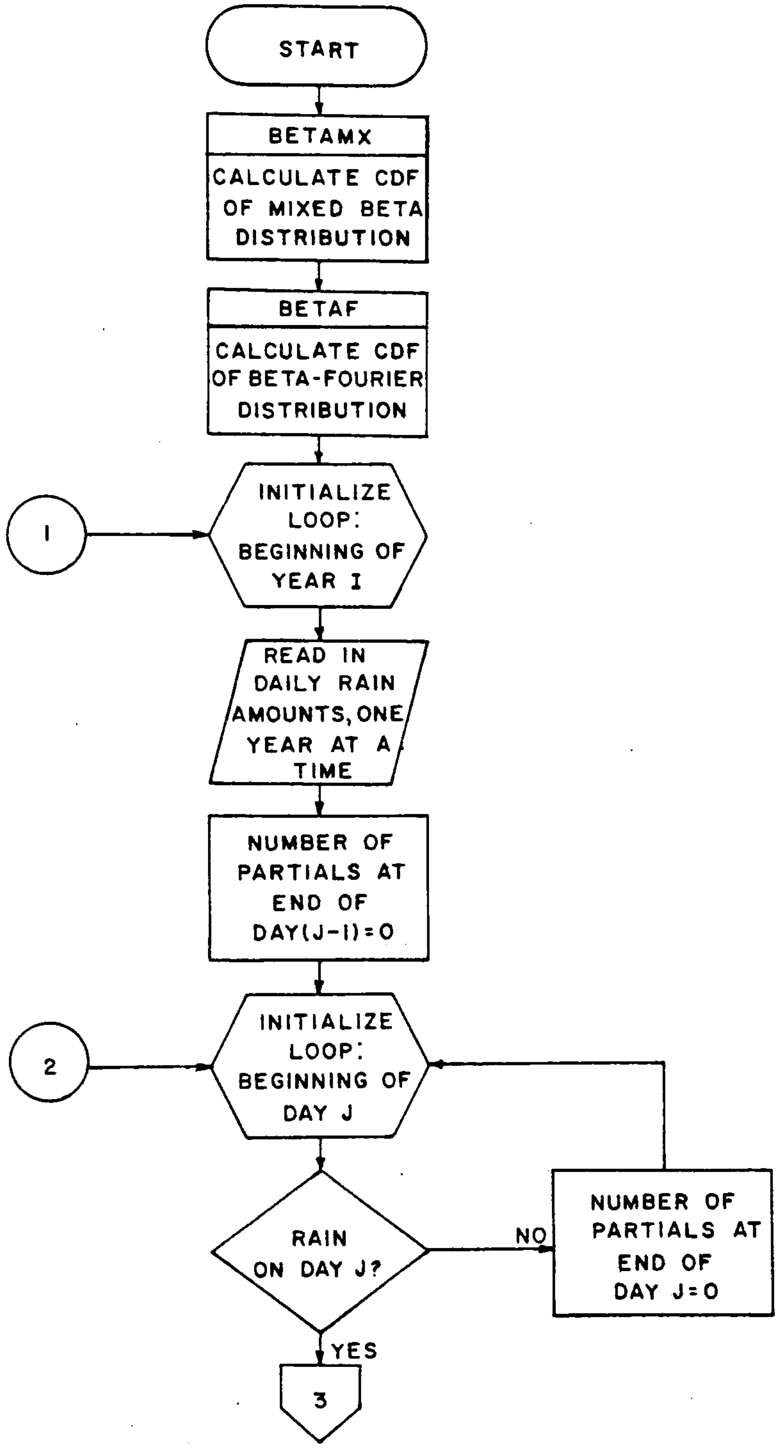
So: $\int_0^{\infty} f_R(r)dr = 1 + 0 = 1.$

Thus, if θ is properly restricted, the beta-Fourier is a proper distribution.

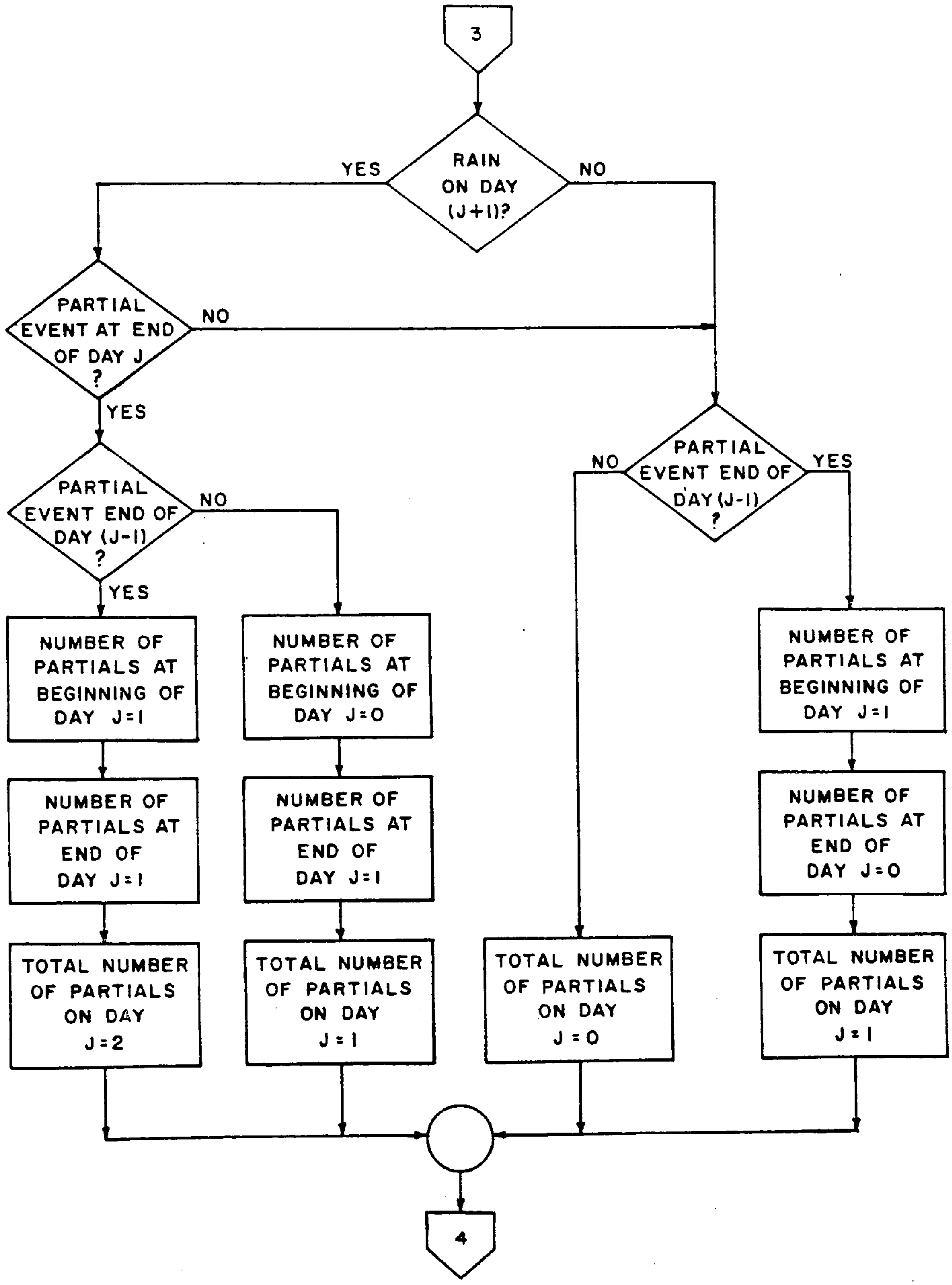
APPENDIX D

COMPUTER PROGRAM

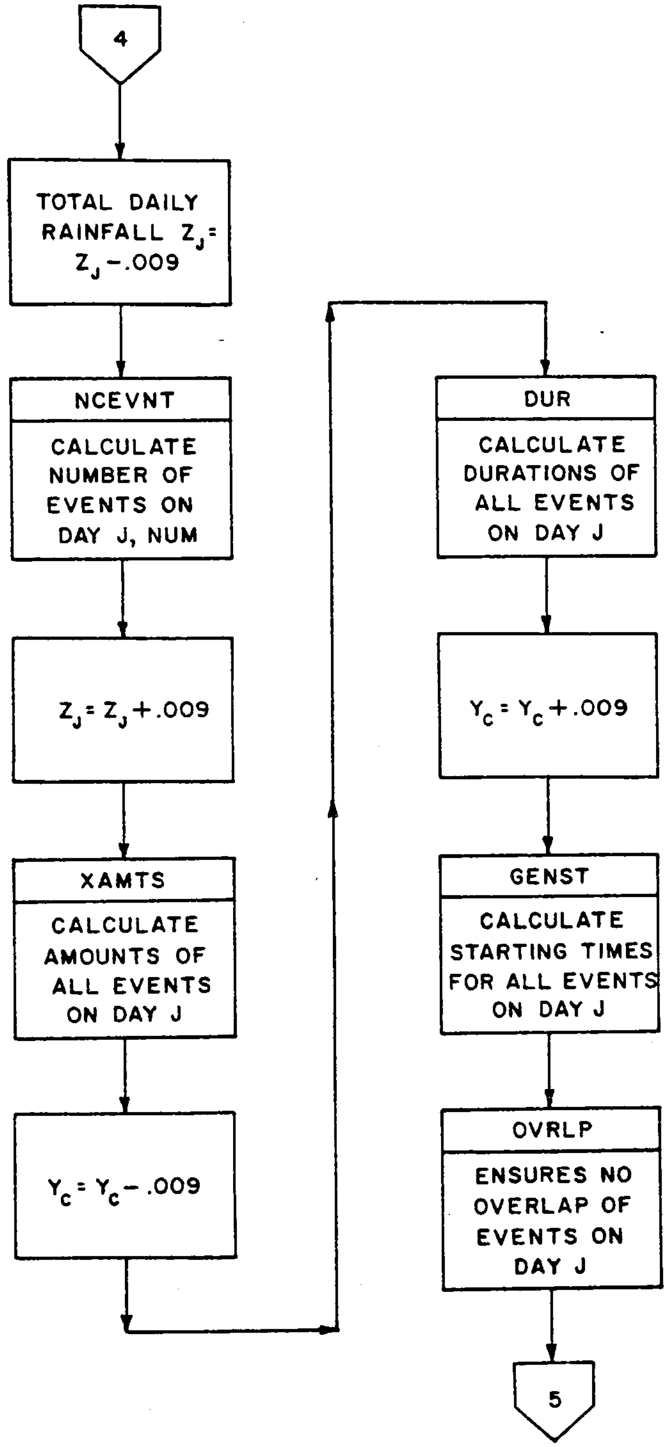
APPENDIX D
FLOWCHART FOR XDISAG-MAIN PROGRAM ONLY



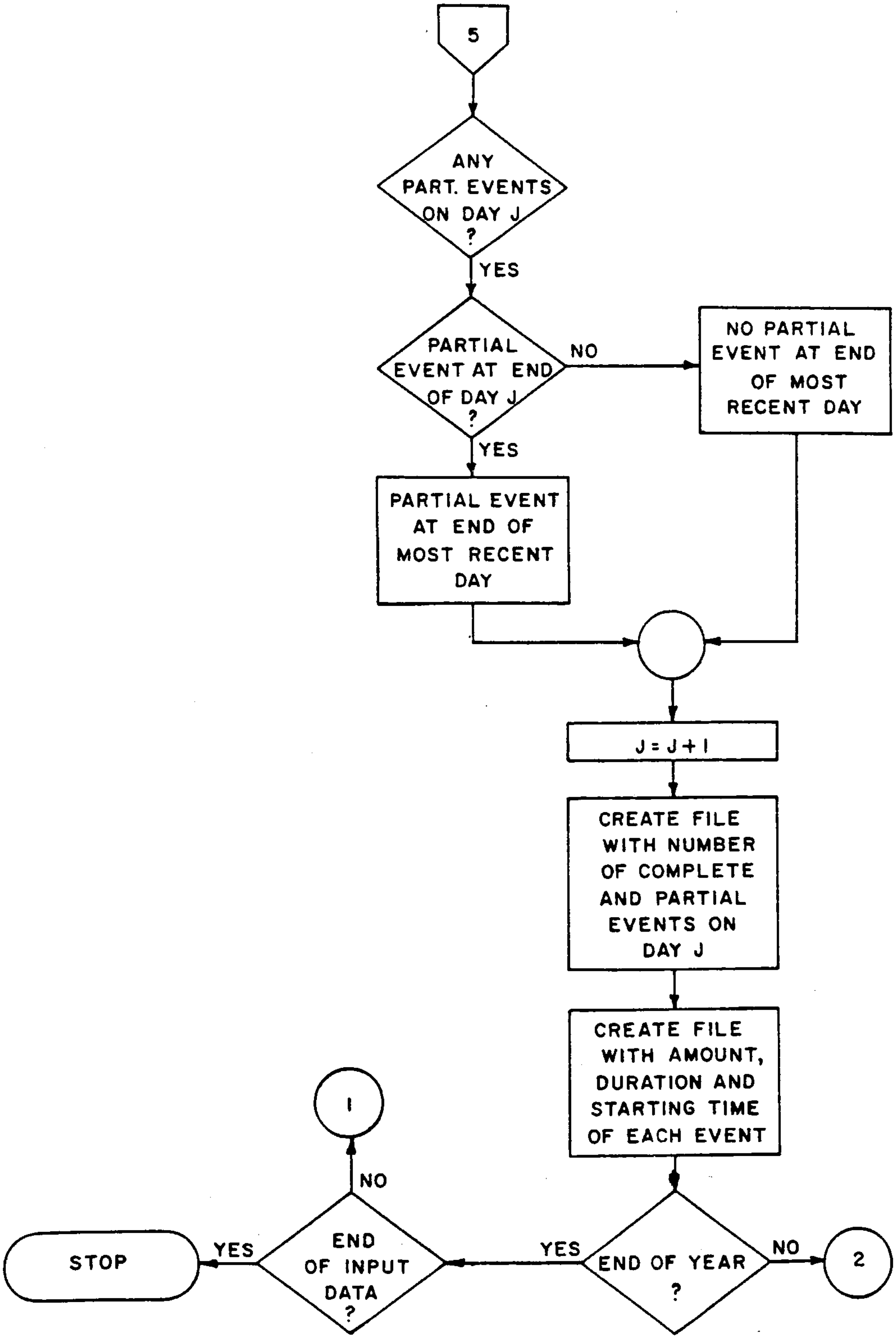
FLOWCHART FOR XDISAG-CONTINUED



FLOWCHART FOR XDISAG-CONTINUED



FLOWCHART FOR XDISAG-CONTINUED



```
                              PROGRAM XDISAG
C
C THIS PROGRAM DISAGGREGATES DAILY RAINFALL AMOUNTS INTO NUMBER OF
C EVENTS,AND SIMULATES THE AMOUNT,DURATION AND STARTING TIME OF
C EACH EVENT
C
C INPUT CONSISTS OF DAILY RAINFALL AMOUNTS FROM GAGE 5 ON THE WALNUT
C GULCH WATERSHED
C
      COMMON KRAN,LRAN
      BYTE NAM1(11),NAM2(11),NAM3(11)
      INTEGER HRS
      DIMENSION Z(23,64),XVAL(51),YCDF(51),XVAR(51),YVAR(51),AMT(6),
     1DURATN(6),STIME(6),HRS(6),MIN(6),ISTAT(6),XNO(800),A(800),
     2D(800),S(800),XNPA(800)
      DATA CECON,NY,ND/.1659,23,64/
C
C Z IS AN ARRAY CONTAINING DAILY RAINFALL AMOUNTS.THE FIRST ARGUMENT RE-
C FERS TO THE YEAR,STARTING WITH 1953;SO 1953 IS YEAR 1,AND SO ON,UP
C THROUGH YEAR 23(1976).THE SECOND ARGUMENT REFERS TO THE DAY;DAY 1
C CORRESPONDS TO JUNE 30,AND DAY 64 TO SEPT 1.
C
C XVAL IS AN ARRAY CONTAINING STARTING TIME VALUES,RANGING FROM
C ZERO TO ONE.ONE CORRESPONDS TO 24 HOURS MILITARY TIME.
C
C YCDF IS AN ARRAY CONTAINING THE CDF VALUES CORRESPONDING TO THE
C STARTING TIMES IN XVAL
C
C XVAR IS AN ARRAY CONTAINING RATIOS OF AMOUNTS,FOR DISAGGREGATED
C DAILY RAINFALL
C
C YVAR IS AN ARRAY CONTAINING THE CDF VALUES CORRESPONDING TO THE
C RATIOS OF DISAGGREGATED AMOUNTS IN XVAR
C
C AMT IS AN ARRAY CONTAINING THE DISAGGREGATED AMOUNTS FOR EACH OF
C THE EVENTS IN A DAY
C
C DURATN IS AN ARRAY CONTAINING THE DURATIONS OF THE DISAGGREGATED
C RAINFALL EVENTS
C
C STIME IS AN ARRAY CONTAINING THE STARTING TIMES OF THE DISAGGREGATED
C EVENTS
C
C HRS IS AN ARRAY CONTAINING TIME OF DAY,IN CLOCK HOURS
C
C MIN IS AN ARRAY CONTAINING MINUTES
C
C ISTAT TELLS WHETHER THE TIME IS AM OR PM
```

```
C
C CECON IS THE PROBABILITY OF A CONTINUING EVENT.THE VALUE IS CONSTANT
C AND EQUAL TO 0.1659
C
C NY IS THE NUMBER OF YEARS OF DATA READ IN.
C
C ND IS THE NUMBER OF DAYS OF RAINFALL DATA.TWO EXTRA DAYS ARE ADDED ON,
C TO ACCOUNT FOR CONTINUING EVENTS ON THE FIRST OR LAST DAY OF THE PER-
C IOD OF INTEREST
C
C I IS THE YEAR NUMBER:I=1 FOR THE FIRST YEAR READ IN
C
C NPDBE IS THE NUMBER OF PARTIAL EVENTS AT THE END OF THE PREVIOUS DAY.
C IT CAN BE EQUAL TO ZERO OR ONE
C
C NPDC IS THE NUMBER OF PARTIAL EVENTS ON THE CURRENT DAY.IT CAN EQUAL
C ZERO,ONE OR TWO
C
C NPDCB IS THE NUMBER OF PARTIAL EVENTS AT THE BEGINNING OF THE CURRENT
C DAY.IT CAN EQUAL ZERO OR ONE.
C
C NPDCE IS THE NUMBER OF PARTIAL EVENTS AT THE END OF THE CURRENT DAY.
C IT CAN EQUAL ZERO OR ONE.
C
C ZAMT IS THE TOTAL DAILY AMOUNT OF RAINFALL
C
C NUM IS THE TOTAL NUMBER OF EVENTS GENERATED IN A DAY
C
C GENERATE SEVERAL RANDOM NUMBERS TO MINIMIZE THE STARTING EFFECTS
C
      KRAN=0
      LRAN=0
      DO 47 K=1,354
         R=RAN(KRAN,LRAN)
 47   CONTINUE
      CALL BETAMX(XVAL,YCDF)
      CALL BETAF(XVAR,YVAR)
C
C CALL INPUT FILE
C
      NV=64
      CALL EDROLD(21,NAM1,NV,64,64,10,'INPUT FILE',IEND)
      IF(IEND.NE.0)STOP
      I=1
      N=1
      N2=1
C
C READ IN DAILY AMOUNTS OF RAINFALL,ONE YEAR AT A TIME
C
```

```
 2      READ(21,END=900)(Z(I,J),J=1,ND)
        WRITE(6,5)I
 5      FORMAT(5X,'YEAR:',5X,I2)
        NPDBE=0
        DO 100 J=1,ND-1
           IF(Z(I,J).EQ.0)GO TO 36
           IF(Z(I,J+1).EQ.0.)GO TO 38
C
C DETERMINE IF A PARTIAL EVENT OCCURS AT THE END OF THE DAY
C
           CRAN=RAN(KRAN,LRAN)
           IF(CRAN.GT.CECON)GO TO 38
           IF(NPDBE.EQ.1)GO TO 34
           NPDCB=0
           NPDCE=1
           NPDC=1
           GO TO 52
 34        NPDCB=1
           NPDCE=1
           NPDC=2
           GO TO 52
 36        NPDBE=0
           GO TO 100
 38        IF(NPDBE.EQ.0)GO TO 50
           NPDC=1
           NPDCB=1
           NPDCE=0
           GO TO 52
 50        NPDC=0
C
C GENERATE THE NUMBER OF EVENTS OCCURRING ON DAY J
C
 52        ZAMT=Z(I,J)-.009
           CALL NCEVNT(ZAMT,NPDC,NUM)
           ZAMT=ZAMT+.009
C
C GENERATE THE AMOUNT,DURATION,AND STARTING TIME OF EACH OF THE EVENTS
C
           CALL XAMTS(NUM,ZAMT,XVAR,YVAR,AMT)
           DO 70 K=1,NUM
             AMT(K)=AMT(K)-.009
 70        CONTINUE
           CALL DUR(NUM,NPDC,NPDBE,AMT,DURATN)
           DO 75,K=1,NUM
             AMT(K)=AMT(K)+.009
 75        CONTINUE
           CALL GENST(NUM,NPDC,NPDBE,XVAL,YCDF,DURATN,STIME)
           CALL OVRLP(NUM,NPDC,NPDBE,STIME,DURATN)
C
```

```
C STORE THE TOTAL NUMBER OF EVENTS PER DAY AND THE NUMBER OF PARTIALS
C WHICH OCCUR
C
          XNO(N)=NUM
          XNPA(N)=NPDC
          N=N+1
C
C STORE THE AMOUNT,DURATION,AND STARTING TIME GENERATED FOR EACH EVENT
C
          DO 80 N4=1,NUM
             A(N2)=AMT(N4)
             D(N2)=DURATN(N4)
             S(N2)=STIME(N4)
             N2=N2+1
 80       CONTINUE
          CALL TIME(STIME,NUM,HRS,MIN,ISTAT)
C
C PRINT OUT SIMULATED DATA FOR DAY J
C
          WRITE(6,85)J,ZAMT,NUM,NPDC
 85       FORMAT(1X,//,2X,'DAY',5X,I2,15X,'RAINFALL AMOUNT:',5X,F8.6/
     1    5X,'NUMBER OF EVENTS:',5X,I2,5X,'NUMBER OF PARTIALS:',
     2    5X,I2)
          DO 88 L=1,NUM
            IF(NUM.EQ.1)AMT(L)=ZAMT
            WRITE(6,87)L,HRS(L),MIN(L),ISTAT(L),AMT(L),DURATN(L)
 87         FORMAT(2X,'EVENT',1X,I2,3X,'START:',3X,I2,':',I2,
     1      2X,A2,3X,'AMT:',2X,F9.6,3X,'DUR:',2X,F6.2)
 88       CONTINUE
          IF(NPDC.EQ.0)GO TO 99
          IF(NPDC.EQ.2)GO TO 94
          IF(NPDBE.EQ.1)GO TO 92
          WRITE(6,90)
 90       FORMAT(5X,'CONTINUING EVENT AT END OF DAY')
          GO TO 98
 92       WRITE(6,93)
 93       FORMAT(5X,'CONTINUING EVENT AT BEGINNING OF DAY')
          GO TO 98
 94       WRITE(6,95)
 95       FORMAT(5X,'TWO CONTINUING EVENTS')
 98       IF(NPDCE.EQ.0)GO TO 99
          NPDBE=1
          GO TO 100
 99       NPDBE=0
 100     CONTINUE
        I=I+1
        IF(I.LE.NY)GO TO 2
 900    WRITE(6,905)I-1
 905    FORMAT(5X,I2,5X,'YEARS OF DATA READ IN')
```

```
      N=N-1
      N2=N2-1
C
C CREATE TWO NEW FILES.THE FIRST CONTAINS THE TOTAL NUMBER OF EVENTS
C IN A DAY,AND THE NUMBER OF PARTIALS WHICH OCCUR.THE SECOND CONTAINS
C THE AMOUNT,DURATION,AND STARTING TIME OF EACH EVENT
C
      NV=2
      CALL EDRNEW(54,NAM2,NV,2,2,8,'NEW FILE',IEND)
      NV=3
      CALL EDRNEW(55,NAM3,NV,3,3,8,'NEW FILE',IEND)
      DO 907 K2=1,N
         WRITE(54)XNO(K2),XNPA(K2)
 907  CONTINUE
      DO 908 K3=1,N2
         WRITE(55)A(K3),D(K3),S(K3)
 908  CONTINUE
 909  END
              SUBROUTINE NCEVNT(AMOUNT,NP,NC)
C
C THIS SUBROUTINE GENERATES THE NUMBER OF EVENTS PER DAY,NC,
C GIVEN THE DAILY AMOUNT OF RAIN,FROM A CONDITIONAL SHIFTED
C NEGATIVE BINOMIAL DISTRIBUTION (SNBD)
C
C AMOUNT IS THE TOTAL DAILY RAINFALL AMOUNT
C
C NP IS THE NUMBER OF PARTIAL EVENTS IN THE DAY
C
C NC IS THE TOTAL NUMBER OF EVENTS,COMPLETE AND PARTIAL
C
C X1,X2,Y1,AND Y2 ARE THE PARAMETERS ESTIMATED FROM THE DATA
C
      COMMON KRAN,LRAN
      DATA X1,X2,Y1,Y2/.7228,5.7939,2.3097,9.5902/
      RAN1=RAN(KRAN,LRAN)
      Q=1.-X1
      T=Y1-1.
      P=X1+(Q*EXP((-X2)*AMOUNT))
      R=Y1-(T*EXP((-Y2)*AMOUNT))
      RP2=R+2.
      RP3=R+3.
      RP4=R+4.
      RP5=R+5.
C
C CALCULATE P(N=1)
C
      P1=P**R
      IF(RAN1.LE.P1)GO TO 50
C
```

```
C CALCULATE P(N=2)
C
        P2=R*P1*(1.-P)
        PX2=P1+P2
        IF(RAN1.LE.PX2)GO TO 60
C
C CALCULATE P(N=3)
C
        CALL GAM(R,GR)
        CALL GAM(RP2,GRP2)
        P3=(GRP2/(2.*GR))*P1*(1.-P)**2.
        PX3=PX2+P3
        IF(RAN1.LE.PX3)GO TO 70
C
C CALCULATE P(N=4)
C
        CALL GAM(RP3,GRP3)
        P4=(GRP3/(6.*GR))*P1*(1.-P)**3.
        PX4=PX3+P4
        IF(RAN1.LE.PX4)GO TO 80
C
C CALCULATE P(N=5)
C
        CALL GAM(RP4,GRP4)
        P5=(GRP4/(24.*GR))*P1*(1.-P)**4.
        PX5=PX4+P5
        IF(RAN1.LE.PX5)GO TO 90
C
C THE MAXIMUM NUMBER OF EVENTS IN A DAY IS SIX
C
        NC=6
        GO TO 100
 50     IF(NP.EQ.2)GO TO 60
        NC=1
        GO TO 100
 60     NC=2
        GO TO 100
 70     NC=3
        GO TO 100
 80     NC=4
        GO TO 100
 90     NC=5
        GO TO 100
 100    RETURN
        END
                SUBROUTINE XINTP(X1,X2,Y1,Y2,CNUM,XVAL)
C
C THIS SUBROUTINE PERFORMS A LINEAR INTERPOLATION
C
```

```
      Y=Y2-Y1
      T=CNUM-Y1
      FRAC=T/Y
      X=X2-X1
      XVAL=X1+FRAC*X
      RETURN
      END
              SUBROUTINE GAM(A,F)
C
C THIS SUBROUTINE CALCULATES THE GAMMA FUNCTION OF A AND RETURNS F
C
      GAMF(X)=1.-.5748646*X+.9512363*X*X-.6998588*X*X*X
      1+.4245549*X*X*X*X-.1010678*X*X*X*X*X
      AD=A
      NC=IFIX(A)
      AT=1.0
      TA=A
      IF(NC.EQ.0)GO TO 14
      DO 13 L=1,NC
         AT=AT*AD
 13   AD=AD-1.
 14   F=GAMF(AD)*AT/TA
      RETURN
      END
              SUBROUTINE TIME(THRS,NC,HRS,MIN,ISTAT)
C
C THIS SUBROUTINE CONVERTS MILITARY TIME TO CLOCK TIME
C
      INTEGER HRS
      DIMENSION THRS(6),HRS(6),MIN(6),ISTAT(6)
      DO 30 K=1,NC
         MINTOT=(THRS(K)*60.)+.5
         HRS(K)=MINTOT/60
         IF(HRS(K).EQ.24)GO TO 10
         MIN(K)=MINTOT-(HRS(K)*60.)
         IF(HRS(K).LT.12)GO TO 10
         HRS(K)=HRS(K)-12
         ISTAT(K)='PM'
         IF(HRS(K).NE.0)GO TO 5
         HRS(K)=12
  5      GO TO 20
 10      ISTAT(K)='AM'
         IF(HRS(K).EQ.24)GO TO 18
         IF(HRS(K).NE.0)GO TO 20
 18      HRS(K)=12
 20      IF(MIN(K).NE.60)GO TO 30
         HRS(K)=HRS(K)+1
         MIN(K)=0
 30   CONTINUE
```

```
      RETURN
      END
              SUBROUTINE OVRLP(NUM,NP,NPDBE,ST,DUR)
C
C THIS SUBROUTINE CHECKS FOR OVERLAP OF DAILY RAINFALL EVENTS.BY
C DEFINITION,THERE MUST BE AT LEAST TEN MINUTES BETWEEN STORM EVENTS.
C IF AN OVERLAP IS FOUND,THE EARLIER EVENT IS MADE TO BEGIN SO THAT
C IT ENDS TEN MINUTES BEFORE THE NEXT EVENT BEGINS.
C
C IF AN EVENT RUNS OVER MIDNIGHT,THE BEGINNING TIME IS MOVED UP SO
C THAT IT ENDS 5 MINUTES BEFORE MIDNIGHT,UNLESS IT'S A PARTIAL EVENT.
C
C A PARTIAL EVENT AT THE END OF THE DAY ENDS AT MIDNIGHT.A PARTIAL
C EVENT AT THE BEGINNING OF THE DAY BEGINS AT MIDNIGHT.
C
C IF AN EVENT IS MOVED UP SO THAT IT BEGINS ON THE PREVIOUS DAY,IT IS
C MADE TO BEGIN 10 MINUTES AFTER THE PREVIOUS EVENT ENDS.
C
C IF AN OVERLAP IS INTRODUCED BECAUSE OF THE SCHEME ABOVE,OR IF
C MORE THAN ONE EVENT OCCURS ON THE PREVIOUS DAY,THE FOLLOWING OCCURS:
C THE FIRST EVENT BEGINS AT MIDNIGHT,AND ALL SUBSEQUENT EVENTS START
C 20 MINUTES AFTER THE PREVIOUS EVENT ENDS,UNLESS THERE'S A PARTIAL
C AT THE END OF THE DAY
C
      DIMENSION ST(6),DUR(6),END(6)
C
C CONVERT STARTING AND ENDING TIMES FROM HOURS TO MINUTES
C
      NCX=NUM
      DO 10 J=1,NUM
         ST(J)=60.*ST(J)
         END(J)=ST(J)+DUR(J)
 10   CONTINUE
      IF(NP.EQ.0)GO TO 15
      IF(NP.EQ.2.OR.(NP.EQ.1.AND.NPDBE.EQ.0))GO TO 13
 12   ST(1)=0.00
      END(1)=ST(1)+DUR(1)
      GO TO 15
 13   ST(NUM)=1440.-DUR(NUM)
      END(NUM)=1440.
      IF(NP.EQ.2)GO TO 12
C
C MAKE SURE NO COMPLETE EVENT RUNS OVER MIDNIGHT
C
 15   IF(END(NUM).LE.1440.)GO TO 20
      XMIN=END(NUM)-1440.
      ST(NUM)=ST(NUM)-(XMIN+5.)
 20   IF(NUM.EQ.1)GO TO 41
 22   DO 40 M=NUM,2,-1
```

```
      IF(ST(M).GT.(END(M-1)+10.))GO TO 40
      IF(ST(M).GT.END(M-1))GO TO 30
      XOVR=END(M-1)-(ST(M))
      ST(M-1)=ST(M-1)-XOVR-10.
23    IF(ST(M-1).GT.0.)GO TO 25
      ST(M-1)=0.
      END(M-1)=ST(M-1)+DUR(M-1)
      ST(M)=END(M-1)+10.
      END(M)=ST(M)+DUR(M)
      IF(M.NE.2)GO TO 42
      IF(M.EQ.NCX)GO TO 45
      IF(ST(M+1).LT.END(M)+10.)GO TO 42
      GO TO 40
25    END(M-1)=ST(M-1)+DUR(M-1)
      GO TO 40
30    XOVR=ST(M)-END(M-1)
      Y=10.-XOVR
      ST(M-1)=ST(M-1)-Y
      GO TO 23
40    CONTINUE
41    IF(ST(1).GE.0.)GO TO 45
42    ST(1)=0.
      END(1)=ST(1)+DUR(1)
      IF(NP.EQ.2.OR.(NP.EQ.1.AND.NPDBE.EQ.0))NUM=NUM-1
      DO 43 K=2,NUM
         ST(K)=END(K-1)+20.
         END(K)=ST(K)+DUR(K)
43    CONTINUE
45    DO 50 J=1,NUM
         ST(J)=ST(J)/60.
50    CONTINUE
60    NUM=NCX
      RETURN
      END

            SUBROUTINE XAMTS(NUM,DAMT,XVAR,YVAR,VAR)
C
C THIS SUBROUTINE DISAGGREGATES A DAILY AMOUNT OF RAINFALL
C INTO REPRESENTATIVE AMOUNTS, FOR EACH EVENT OF THE DAY.
C
      COMMON KRAN,LRAN
      DATA XINC/.02/
      DIMENSION XVAR(51),YVAR(51),VAR(6)
C
C XINC IS THE INCREMENT BY WHICH THE CDF TABLE IS SET UP IN THE
C SUBROUTINE BETAF
C
C DAMT IS THE TOTAL DAILY AMOUNT OF RAINFALL
```

```
C
C NUM IS THE NUMBER OF EVENTS IN A DAY
C
C XVAR AND YVAR ARE THE TABLED VALUES OF THE CDF OF THE BETA WITH
C ADDED FOURIER TERM
C
C VAR IS AN ARRAY CONTAINING DISAGGREGATED AMOUNTS FOR EACH OF THE
C EVENTS IN A DAY.THE SCHEME IS SET UP DISAGGREGATE UP TO SIX
C EVENTS PER DAY.
C
      DAILY=DAMT
      IF(NUM.EQ.1)GO TO 998
      TRY=0.
      COUNT=0.
C
C ON DAYS IN WHICH TWO EVENTS OCCUR,THE RATIO OF THE
C AMOUNT OF THE FIRST EVENT TO THE DAILY TOTAL AMOUNT IS DES-
C CRIBED BY A BETA DISTRIBUTION WITH AN ADDED FOURIER TERM.THE
C SAME DISTRIBUTION IS USED TO DESCRIBE THE RATIO OF THE AMOUNTS
C OF THE (SECOND + THIRD) EVENTS TO THE TOTAL DAILY AMOUNT,FOR
C DAYS IN WHICH THREE EVENTS OCCUR
C
C CALCULATIONS FOR SIX EVENTS PER DAY
C
C THE DAILY AMOUNT IS FIRST DISAGGREGATED INTO TWO AMOUNTS,SPLIT1
C AND SPLIT2,ACCORDING TO A UNIFORM DISTRIBUTION
C
      IF(NUM.NE.6)GO TO 10
      SIXRAN=RAN(KRAN,LRAN)
      SPLIT1=SIXRAN*DAMT
      SPLIT2=DAMT-SPLIT1
      TRY=1.
      DAMT=SPLIT1
      GO TO 20
C
C CALCULATIONS FOR FIVE EVENTS PER DAY
C
C THE DAILY AMOUNT IS FIRST DISAGGREGATED INTO TWO AMOUNTS, BRK1
C AND BRK2,ACCORDING TO A UNIFORM DISTRIBUTION
C
 10   IF(NUM.NE.5)GO TO 15
      FIVRAN=RAN(KRAN,LRAN)
      BRK1=FIVRAN*DAMT
      BRK2=DAMT-BRK1
       DAMT=BRK1
      COUNT=1.
      GO TO 20
 15   IF(NUM.EQ.4)GO TO 80
      GO TO 20
```

```
 18     COUNT=2.
 20     RAN1=RAN(KRAN,LRAN)
        Y=RAN1/XINC
        N1=INT(Y)+1
        N2=N1+1
        XA1=XVAR(N1)
        XA2=XVAR(N2)
        CDF1=YVAR(N1)
        CDF2=YVAR(N2)
        CALL XINTP(XA1,XA2,CDF1,CDF2,RAN1,RATIO)
        IF((NUM.EQ.3).OR.(NUM.EQ.6))GO TO 40
        IF(NUM.EQ.5.AND.COUNT.EQ.2)GO TO 35
C
C CALCULATIONS FOR TWO EVENTS PER DAY
C
C RATIO IS THE RATIO OF THE FIRST EVENT TO THE DAILY TOTAL AMOUNT
C
        VAR(1)=DAMT*RATIO
        VAR(2)=DAMT-VAR(1)
        IF(NUM.EQ.2)GO TO 99
        IF(NUM.EQ.5.AND.COUNT.EQ.1.)GO TO 18
 35     VAR(3)=BRK2-(BRK2)*RATIO
        REMAIN=BRK2-VAR(3)
        RAN2=RAN(KRAN,LRAN)
        VAR(4)=RAN2*REMAIN
        VAR(5)=REMAIN-VAR(4)
        GO TO 99
 40     IF(NUM.EQ.6.AND.TRY.EQ.2.)GO TO 45
C
C CALCULATIONS FOR THREE EVENTS PER DAY
C
C RATIO IS THE RATIO OF THE (SECOND+THIRD) EVENTS TO THE DAILY TOTAL
C
C REMAIN IS THE SUM OF THE AMOUNTS OF THE (SECOND+THIRD) EVENTS
C
        VAR(1)=DAMT-(DAMT*RATIO)
        REMAIN=DAMT-VAR(1)
C
C THE RATIO OF THE AMOUNT OF THE SECOND EVENT TO THE SUM OF THE
C AMOUNTS OF THE (SECOND + THIRD) EVENTS,ON DAYS IN WHICH THREE
C EVENTS OCCUR,IS UNIFORMLY DISTRIBUTED
C
        RAN2=RAN(KRAN,LRAN)
        VAR(2)=RAN2*REMAIN
        VAR(3)=REMAIN-VAR(2)
        IF(NUM.EQ.3)GO TO 99
        TRY=2.
        GO TO 20
 45     VAR(4)=SPLIT2-(SPLIT2*RATIO)
```

```
      REMAIN=SPLIT2-VAR(4)
      RAN2=RAN(KRAN,LRAN)
      VAR(5)=RAN2*REMAIN
      VAR(6)=REMAIN-VAR(5)
      GO TO 99
C
C CALCULATIONS FOR FOUR EVENTS PER DAY.
C THE RATIO OF THE SUM OF THE AMOUNTS OF THE (THIRD + FOURTH)
C EVENTS TO THE TOTAL DAILY AMOUNT IS UNIFORMLY DISTRIBUTED.THE
C RATIOS OF:1)THE FIRST TO THE (FIRST + SECOND) AND 2)THE THIRD
C TO THE (THIRD + FOURTH) ARE ALSO UNIFORMLY DISTRIBUTED.
C
 80   RAN34=RAN(KRAN,LRAN)
      AMT34=DAMT*RAN34
      AMT12=DAMT-AMT34
      RAN3=RAN(KRAN,LRAN)
      VAR(3)=RAN3*AMT34
      VAR(4)=AMT34-VAR(3)
      RAN2=RAN(KRAN,LRAN)
      VAR(2)=RAN2*AMT12
      VAR(1)=AMT12-VAR(2)
C
C IF THE (TOTAL DAILY AMOUNT)/NUM IS LESS THAN .01 INCH,DECREASE THE
C NUMBER OF EVENTS BY ONE,AND ADJUST THE AMOUNTS
C
  99  IF((DAILY/NUM).GT..01)GO TO 399
      NUM=DAILY/.01
      SUM=0.
      DO 390 K1=1,NUM-1
         VAR(K1)=.01
         SUM=SUM+VAR(K1)
 390  CONTINUE
      VAR(NUM)=DAILY-SUM
      GO TO 999
C
C THE MINIMUM AMOUNT OF RAIN WHICH CAN BE MEASURED PER EVENT IS .01
C INCH.IF LESS THAN .01 INCH IS GENERATED,TAKE SOME FROM THE LARGEST
C EVENT OF THE DAY AND ADD IT TO THE SMALL EVENT TO GET .01 INCH
C
 399  DO 400 I=1,NUM
         IF(VAR(I).LT..01)GO TO 501
 400  CONTINUE
      GO TO 999
C
C FIND THE LARGEST EVENT AMOUNT OF THE DAY
C
 501  JBIG=1
      DO 510 J1=2,NUM
         IF(VAR(J1).LT.VAR(JBIG))GO TO 510
```

```
            JBIG=J1
 510     CONTINUE
         DO 540 K=1,NUM
            IF(VAR(K).GE..01)GO TO 540
            RAMT=.01-VAR(K)
            VAR(K)=.01
            VAR(JBIG)=VAR(JBIG)-RAMT
 540     CONTINUE
         GO TO 999
 998     VAR(1)=DAILY
 999     DAMT=DAILY
         RETURN
         END
                  SUBROUTINE DUR(NC,NP,NPDBE,AMOUNT,DURAT)
C
C THIS SUBROUTINE GENERATES A DURATION FOR AN EVENT,GIVEN AN
C AMOUNT
C
C THE SAME METHOD IS USED FOR COMPLETE AND PARTIAL EVENTS BUT THE
C PARAMETERS ARE DIFFERENT
C
         DIMENSION AMOUNT(6),DURAT(6),A(6)
         DATA PI,SLOP,DINTCP,SERES,SLOP1,DINT1,SERES1/
         13.14159,.3785,4.6460,.8885,.3296,5.1624,.7755/
         COMMON KRAN,LRAN
C
C SLOP,DINTCP AND SERES ARE THE SLOPE,INTERCEPT AND STANDARD
C ERROR OF ESTIMATE OF THE REGERESSION OF THE LN OF DURATION
C ON THE LN OF AMOUNT,FOR COMPLETE EVENTS
C
C SLOP1,DINTCP1,AND SERES1 ARE FOR PARTIAL EVENTS
C
         NCX=NC
         IF(NP.EQ.0)GO TO 24
         IF(NP.EQ.2)GO TO 11
         IF(NPDBE.EQ.1)GO TO 10
         M=NC
         NC=NC-1
         K=1
         GO TO 12
 10      M=1
         K=2
         GO TO 12
 11      M=1
         NC=NC-1
         K=2
 12      A(M)=ALOG(AMOUNT(M))
         R1=RAN(KRAN,LRAN)
```

```
      R2=RAN(KRAN,LRAN)
      X1=(SQRT((-2.)*ALOG(R1)))*COS(2.*PI*R2)
      Q1=DINT1+(SLOP1*A(M))+(X1*SERES1)
      DURAT(M)=EXP(Q1)
      IF(DURAT(M).GT.480.)DURAT(M)=480.
      IF(M.EQ.NC+1.AND.NP.EQ.2)GO TO 200
      IF(NC.EQ.1.AND.NP.EQ.2)GO TO 15
      IF(NP.EQ.1.AND.NCX.EQ.1)GO TO 200
      GO TO 25
 15   M=2
      GO TO 12
 24   M=0
      K=1
 25   IF(NC.EQ.0)NC=1
      IF(NC.LT.K)NC=K
      DO 100 I=K,NC
         A(I)=ALOG(AMOUNT(I))
         R1=RAN(KRAN,LRAN)
         R2=RAN(KRAN,LRAN)
C
C GENERATE A STANDARD NORMAL RANDOM VARIATE ,X1
C
         X1=(SQRT((-2.)*ALOG(R1)))*COS(2.*PI*R2)
C
C THE NATURAL LOG OF THE DURATION IS NORMALLY DISTRIBUTED ABOUT
C THE LINE Y=.3785*(LN(AMT)+4.6460 WITH STANDARD ERROR
C OF ESTIMATE EQUAL TO .8885
C
C
C LN(DUR)=.3785*(LN(AMT))+4.6460+(EPSILON*.8885)
C
         Q1=DINTCP+(SLOP*A(I))+(X1*SERES)
         DURAT(I)=EXP(Q1)
         IF(DURAT(I).LE.480.0)GO TO 100
         DURAT(I)=480.00
 100  CONTINUE
      IF(M.EQ.0)GO TO 200
      IF(M.EQ.NC+1.AND.NP.EQ.1)GO TO 200
      IF(M.EQ.1.AND.NP.EQ.1)GO TO 200
      M=NC+1
      GO TO 12
 200  NC=NCX
      RETURN
      END
              SUBROUTINE GENST(NC,NP,NPDBE,XST,CDFST,DURAT,ST)
C
C THIS SUBROUTINE GENERATES STARTING TIMES FOR RAINFALL EVENTS.A UNIFORM
C RANDOM VARIABLE IS TRANSFORMED TO A MIXED BETA DISTRIBUTED RANDOM VAR-
C IABLE.
```

```
C
      COMMON KRAN,LRAN
      DATA XINC/.02/
      DIMENSION XST(51),CDFST(51),DURAT(6),ST(6)
C
C XINC IS THE INCREMENT BY WHICH THE CDF TABLE IS SET UP IN THE SUB-
C
C ROUTINE BETAMX
C
C NC IS THE NUMBER OF COMPLETE EVENTS
C
C NP IS THE NUMBER OF PARTIAL EVENTS
C
C NPDBE IS THE NUMBER OF PARTIAL EVENTS AT THE END OF THE PREVIOUS DAY
C
C XST AND CDFST ARE THE TABLED VALUES OF STARTING TIMES AND CORRES-
C PONDING CDF VALUES,AT CDF INCREMENTS OF .02
C
C DURAT  IS AN ARRAY CONTAINING DURATIONS
C
C ST IS AN ARRAY CONTAINING STARTING TIMES FOR THE NC EVENTS
C
      NCX=NC
      IF(NP.EQ.0)GO TO 20
      IF(NP.EQ.2)GO TO 11
      IF(NPDBE.EQ.1)GO TO 10
      K=1
      M=NC
      IF(NC.EQ.1)GO TO 48
      NC=NC-1
      GO TO 30
 10   ST(1)=0.00
      IF(NC.EQ.1.AND.NPDBE.EQ.1)GO TO 60
      K=2
      GO TO 30
 11   ST(1)=0.00
      K=2
      M=NC
      IF(NP.EQ.2.AND.NC.EQ.2)GO TO 48
      NC=NC-1
      GO TO 30
 20   K=1
 30   IF(NC.LT.K)NC=K
      DO 35 M1=K,NC
         XRAN=RAN(KRAN,LRAN)
         X=XRAN/XINC
C
C XRAN/INC TELLS WHICH TABULATED VALUES TO INTERPOLATE BETWEEN TO
C GET STARTING TIME
```

```
C
            N1=INT(X)+1
            N2=INT(X)+2
            ST1=XST(N1)
            ST2=XST(N2)
            CDF1=CDFST(N1)
            CDF2=CDFST(N2)
            CALL XINTP(ST1,ST2,CDF1,CDF2,XRAN,STIM)
            ST(M1)=STIM
 35      CONTINUE
C
C SORT THE GENERATED STARTING TIMES FROM EARLY TO LATE
C
         CALL SORT1(ST,NCX,-1)
         DO 45 I=1,NCX
            ST(I)=24*ST(I)
 45      CONTINUE
         IF(NP.EQ.0)GO TO 60
         IF(NP.EQ.1.AND.NPDBE.EQ.1)GO TO 60
 48      ST(NCX)=(1440.-DURAT(NCX))/60.
 60      NC=NCX
         RETURN
         END

                SUBROUTINE BETAMX(XNUM,YNUM)
C
C THIS SUBROUTINE COMPUTES THE CUMULATIVE DISTRIBUTION FUNCTION OF
C EVENT STARTING TIMES,ACCORDING TO THE MIXED BETA DISTRIBUTION.
C THE TRAPEZOIDAL APPROXIMATION TECHNIQUE IS USED TO NUMERICALLY
C INTEGRATE THE MIXED BETA DENSITY FUNCTION.
C
C STARTING TIMES AND CORRESPONDING CDF VALUES ARE STORED IN THE ARRAYS
C XNUM AND YNUM,RESPECTIVELY,AND PASSED BACK TO THE CALLING SUBROUTINE.
C
C VALUES STORED ARE THOSE CORRESPONDING TO CDF INCREMENTS OF .02,
C STARTING AT 0.00 AND INCREASING UP TO 1.00.
C
         BYTE NAM(11)
         DIMENSION XNUM(51),YNUM(51)
         DATA A1,B1,A2,B2,OM/.6709,3.2267,6.3076,2.4190,.1519/
C
C A1,B1,A2,B2, AND OM ARE MAX. LIKELIHOOD PARAMETER ESTIMATES FOR THE
C MIXED BETA DENSIITY FUNCTION.OM IS THE WEIGHTING FACTOR.
C
         C1=A1+B1
         C2=A2+B2
         CALL GAM(C1,G1)
         CALL GAM(A1,G2)
         CALL GAM(B1,G3)
```

```
      CALL GAM(C2,G4)
      CALL GAM(A2,G5)
      CALL GAM(B2,G6)
      BET1=G1/(G2*G3)
      BET2=G4/(G5*G6)
      P1=A1-1.
      Q1=B1-1.
      P2=A2-1.
      Q2=B2-1.
      EPS=.001
      YVAL=.02
      X=.00005
      W=1.-OM
      N=2
      CUM1=0.00
      XNUM(1)=0.00
      YNUM(1)=0.00
      R1=X
C
C DESIRED CDF VALUES ARE FOR INCREMENTS OF .02.THE INTERVAL OF NUMERICAL
C INTEGRATION IS CHOSEN SO THAT THE MIDPOINT WILL BE SOME MULTIPLE OF .02,
C FOR EXAMPLE,.02,.04,.06,ETC.
C
C EPS IS EPSILON
C
C N IS THE ARRAY INDEX FOR XNUM AND YNUM
C
C BDEN1 IS THE VALUE OF THE DENSITY FUNCTION AT THE LEFT END OF THE
C INTERVAL
C
C BDEN2 IS THE VALUE AT THE RIGHT END
C
C R1 AND R2 ARE THE LEFT AND RIGHT ENDPOINTS OF THE INTERVAL
C
C CUM1 IS THE VALUE OF THE CDF UP TO BUT NOT INCLUDING THE CURRENT
C INTERVAL OF INTEGRATION.
C
C CUM2 IS THE CURRENT CDF VALUE
C
C B IS THE DESIRED CDF VALUE (AT INCREMENTS OF .02)
C
      BDEN1=((OM*BET1*(R1**P1)*((1.-R1)**Q1))+(W*BET2*(R1**P2)*
      1((1.-R1)**Q2)))
      B=YVAL
      DO 100 J=1,10000
         IF(N.GT.51)GO TO 999
         R2=R1+.0001
         BDEN2=((OM*BET1*(R2**P1)*((1.-R2)**Q1))+(W*BET2*(R2**P2)
     1   *((1.-R2)**Q2)))
         CUM2=CUM1+((BDEN1+BDEN2)/2.)*(R2-R1)
```

```
C
C IF CUM2 IS WITHIN EPSILON OF B,ACCEPT THE VALUE AND STORE IN ARRAY
C  YNUM
C
          IF(ABS(CUM2-B).LE.EPS)GO TO 70
C
C IF CUM2 IS LARGER THAN B,REDUCE INTERVAL SIZE BY A HALF AND REDO CALCU-
C LATIONS
C
          IF(CUM2-B.GT.0.)GO TO 80
          R1=R2
          BDEN1=BDEN2
          CUM1=CUM2
          GO TO 100
 70       XNUM(N)=(R1+R2)/2.
          YNUM(N)=B
          N=N+1
          B=B+YVAL
          R1=R2
          BDEN1=BDEN2
          CUM1=CUM2
          GO TO 100
 80       XHALF=(R2-R1)/2.
          R2=R1+XHALF
          BDEN2=((OM*BET1*(R2**P1)*((1.-R2)**Q1))+(W*BET2*(R2**P2)
     1    *((1.-R2)**Q2)))
          CUM2=CUM1+((BDEN1+BDEN2)/2.)*(R2-R1)
          IF(ABS(CUM2-B).LE.EPS)GO TO 70
          IF((CUM2-B).GT.0.)GO TO 80
          R1=R2
          BDEN1=BDEN2
          CUM1=CUM2
 100  CONTINUE
 999  XNUM(51)=1.00
      YNUM(51)=1.00
      RETURN
      END

             SUBROUTINE BETAF(XVAR,YVAR)
C
C THIS SUBROUTINE CALCULATES THE CUMULATIVE DISTRIBUTION FUNCTION OF
C RATIOS OF AMOUNTS OF INDIVIDUAL RAIN EVENTS TO THE DAILY TOTAL AMOUNT.
C
C IT CAN BE USED TO PREDICT TWO SEPARATE CASES:1)ON DAYS IN WHICH TWO
C EVENTS OCCUR,IT CAN PREDICT THE RATIO OF THE FIRST EVENT TO THE DAILY
C TOTAL AMOUNT.2)ON DAYS IN WHICH THREE EVENTS OCCUR,IT CAN PREDICT THE
C RATIO OF THE (SECOND+THIRD) EVENTS TO THE DAILY TOTAL.
C
C THE DISTRIBUTION USED IS THE BETA WITH A FOURIER TERM ADDED ON.THE PA-
C RAMETERS ARE MAXIMUM LIKELIHOOD ESTIMATES.
```

```
C
C THE TRAPEZOIDAL APPROXIMATION IS USED TO NUMERICALLY INTEGRATE THE
C DENSITY FUNCTION.RATIOS AND CORRESPONDING CDF VALUES ARE STORED IN THE
C ARRAYS XVAR AND YVAR,RESPECTIVELY,AND ARE PASSED BACK TO THE MAIN
C PROGRAM.
C
C VALUES STORED ARE THOSE CORRESPONDING TO CDF INCREMENTS OF .02,START-
C ING AT 0.00 AND INCREASING UP TO 1.
C
      DIMENSION XVAR(51),YVAR(51)
      DATA A,B,C,PI/1.25141,.90452,.08194,3.14159/
C
C A AND B ARE PARAMETER ESTIMATES FOR THE BETA;C IS THE AMPLITUDE ESTI-
C MATE IN THE FOURIER TERM.
C
      D=A+B
      CALL GAM(D,G1)
      CALL GAM(A,G2)
      CALL GAM(B,G3)
      BET=G1/(G2*G3)
      P1=A-1.
      P2=B-1.
      EPS=.001
      YVAL=.02
      X=.00005
      CUM1=0.
      XVAR(1)=0.0
      YVAR(1)=0.0
      N=2
      R1=X
C
C THIS SUBROUTINE IS VERY SIMILAR TO BETAMX,EXCEPT FOR THE EXPRESSION
C FOR THE DENSITY FUNCTION.
C
      BDEN1=BET*(R1**P1)*((1.-R1)**P2)+C*2.*PI*COS(2.*PI*R1)
      B=YVAL
      DO 100 J=1,10000
         IF(N.GT.51)GO TO 999
         R2=R1+.0001
         BDEN2=BET*(R2**P1)*((1.-R2)**P2)+C*2.*PI*COS(2.*PI*R2)
         CUM2=CUM1+((BDEN1+BDEN2)/2.)*(R2-R1)
C
C IF CUM2 IS WITHIN EPSILON OF B,ACCEPT THE VALUE AND STORE IN ARRAY
C YVAR
C
         IF(ABS(CUM2-B).LE.EPS)GO TO 70
C
C IF CUM2 IS LARGER THAN B,REDUCE INTERVAL SIZE BY A HALF,AND REDO
C CALCULATIONS
```

```
C
      IF(CUM2-B.GT.O.)GO TO 80
      R1=R2
      BDEN1=BDEN2
      CUM1=CUM2
      GO TO 100
 70   XVAR(N)=(R1+R2)/2.
      YVAR(N)=B
      N=N+1
      B=B+YVAL
      R1=R2
      BDEN1=BDEN2
      CUM1=CUM2
      GO TO 100
 80   XHALF=(R2-R1)/2.
      R2=R1+XHALF
      BDEN2=BET*(R2**P1)*((1.-R2)**P2)+C*2.*PI*COS(2.*PI*R2)
      CUM2=CUM1+((BDEN1+BDEN2)/2.)*(R2-R1)
      IF(ABS(CUM2-B).LE.EPS)GO TO 70
      IF((CUM2-B).GT.0)GO TO 80
      R1=R2
      BDEN1=BDEN2
      CUM1=CUM2
 100  CONTINUE
 999  CONTINUE
      NV=2
      RETURN
      END
```

LIST OF REFERENCES

Akaike, H., "A New Look at the Statistical Model Identification," IEEE Transactions on Automatic Control, Vol. 19, No. 6, pp. 716-723, 1974.

Beals, G. A., "Specification of Daily Precipitation through Synoptic Climatology," MS Thesis, MIT, 1954.

Benjamin, J. R., and Cornell, C. A., "Probability, Statistics and Decision for Civil Engineers," McGraw-Hill Book Company, New York, 1970.

Betson, R. P., Bales, J., and Pratt, H. E., "User's Guide to TVA-HYSIM," US EPA, EPA-600/7-80-048, pp. 40-44, 1980.

Brakensiek, D. L., "Fitting a Generalized Log-Normal Distribution to Hydrologic Data," Transactions AGU, Vol. 39, pp. 467-473, 1958.

Bras, R. L., and Rodriguez-Iturbe, I., "Rainfall Generation: A Nonstationary Time-Varying Multidimensional Model," Water Resources Research, Vol. 12, No. 3, pp. 450-456, 1976.

Brooks C. F., "Need for Universal Standards for Measuring Precipitation, Snowfall and Snow Cover," IASH Bulletin, No. 23, pp. 7-58, 1938.

Bryson, R. A., "The Annual March of Precipitation in Arizona, New Mexico, and Northwestern Mexico," Technical Report No. 6, Institute of Atmospheric Physics, University of Arizona, Tucson, 1957.

Buishand, T. A., "Stochastic Modeling of Daily Rainfall Sequences," Mededelingen Landbouwhogeschool, Wageningen, 77-3, 1977.

Chery, D. L., and Kagan, R. S., "An Overview of the Precipitation Processing System at the Southwest Watershed Research Center," National Symposium on Precipitation Analysis for Hydrologic Models, AGU, Davis, CA, pp. 48-59, 1975.

Das, S. C., "The Fitting of Truncated Type III Curves to Daily Rainfall Data," Australian Journal of Physics, Vol. 8, pp. 298-304, 1955.

Fogel, M. M., and Duckstein, L., "Point Rainfall Frequencies in Convective Storms," Water Resources Research, Vol. 5, No. 6, pp. 1229-1237, 1969.

Grace, R. A., and Eagleson, P. S., "Synthesis of Short-Time Increment Rainfall Sequences," Hydrodynamic Lab. Report No. 91, MIT, 1966.

Green, C. R., and Sellers, W. D., "Arizona Climate," The Univ. of Arizona Press, Tucson, 1964.

Hales, J. E., "Southwestern United States Summer Monsoon Source - Gulf of Mexico or Pacific," NOAA Technical Memorandum, NW-SWR-84, 1973.

Hoel, P. G., "Introduction to Mathematical Statistics," John Wiley & Sons, Inc., New York, 1971.

Hoshi, K., and Burges, S. J., "The Impact of Seasonal Flow Characteristics and Demand Patterns on Required Reservoir Storage," Journal of Hydrology, Vol. 37, pp. 241-260, 1978.

Kendall, M., and Stuart, A., "The Advanced Theory of Statistics," Vol. 2, Charles Griffin & Company, Ltd., London, 1977.

Kotz, S., and Neumann, J., "On the Distribution of Precipitation Amounts for Periods of Increasing Length," Journal of Geophysical Research, Vol. 8, pp. 3635-3640, 1963.

Lane, W. L., "Applied Stochastic Techniques (LAST Computer Package) User Manual," Bureau of Reclamation, US Department of the Interior, Denver, CO, 1979.

Mejia, J. M., and Rousselle, J., "Dissaggregation Models in Hydrology, Revisited," Water Resources Research, Vol. 12, No. 2, pp. 185-186, 1976.

Mielke, P. W., Jr., and Johnson, E. S., "Some Generalized Beta Distributions of the Second Kind Having Desirable Application Features in Hydrology and Meteorology," Water Resources Research, Vol. 10, No. 2, pp. 223-226, 1974.

Mills, W. C., "Stochastic Modeling of Rainfall for Deriving Distributions of Watershed Input," In: Statistical Analysis of Rainfall and Runoff, ed. V. P. Singh, Water Resources Publications, Littleton, CO., pp. 103-118, 1982.

Naylor, T. H., Balintfy, J. L., Burdick, D. S., and Chu, K., "Computer Simulation Techniques," John Wiley & Sons, Inc., New York, 1966.

Palmer, J. R., "An Improved Procedure for Orthogonalizing the Search Vectors in Rosenbrock's and Swann's Direct Search Optimization Methods," Computer Journal, Vol. 12, pp. 69-71, 1969.

Sariahmed, A., "Synthesis of Sequences of Summer Thunderstorm Volumes for the Atterbury Watershed in the Tucson Area," MS Thesis, University of Arizona, 1969.

Smith, R. E., and Schreiber, H. A., "Point Processes of Seasonal Thunderstorm Rainfall. 2. Rainfall Depth Probabilities," Water Resources Research, Vol. 10, No. 3, pp. 418-428, 1974.

Stedinger, J. R., and Vogel, R. M., "Disaggregation Procedures for Generating Serially Correlated Flow Vectors," Water Resources Research, Vol. 20, No. 1, pp. 47-56, 1984.

Stidd, C. K., "Cube-Root-Normal Precipitation Distributions," Transactions AGU, Vol. 34, No. 1, pp. 31-35, 1953.

Tao, P. C., and Delleur, J. W., "Multistation, Multiyear Synthesis of Hydrologic Time Series by Disaggregation," Water Resources Research, Vol. 12, No. 6, pp. 1303-1312, 1976.

Todorovic, P., and Yevjevich, V., "Stochastic Process of Precipitation," CSU Hydrology Papers, No. 35, Colorado State University, 1969.

Valencia, D. R., and Schaake, J. C., Jr., "Disaggregation Processes in Stochastic Hydrology," Water Resources Research, Vol. 9, No. 3, pp. 580-585, 1973.

Woolhiser, D. A., and Econopouly, T. W., unpublished research, 1983.

Woolhiser, D. A., and Osborn, H. B., "A Stochastic Model of Thunderstorm Rainfall Intensities," unpublished, 1984.

Woolhiser, D. A., and Roldán, J., "Stochastic Daily Precipitation Models. 2. A Comparison of Distribution of Amounts," Water Resources Research, Vol. 18, No. 5, pp. 1461-1468, 1982.