

## Exploratory Analysis of Precipitation Events with Implications for Stochastic Modeling

BARBARA G. BROWN<sup>1</sup>

*Department of Atmospheric Sciences, Oregon State University, Corvallis, OR 97331*

RICHARD W. KATZ

*Environmental and Societal Impacts Group, National Center for Atmospheric Research,<sup>2</sup> Boulder, CO 80307*

ALLAN H. MURPHY

*Department of Atmospheric Sciences, Oregon State University, Corvallis, OR 97331*

(Manuscript received 31 May 1984, in final form 17 September 1984)

### ABSTRACT

The use of a concept called a precipitation "event" to obtain information regarding certain statistical properties of precipitation time series at a particular location and for a specific application (e.g., for modeling erosion) is described. Exploratory data analysis is used to examine several characteristics of more than 31 years of primitive precipitation events based on hourly precipitation data at Salem, Oregon. A primitive precipitation event is defined as one or more consecutive hours with at least 0.01 inches (0.25 mm) of precipitation. The characteristics of the events that are considered include the duration, magnitude, average intensity and maximum intensity of the event and the number of hours separating consecutive events.

By means of exploratory analysis of the characteristics of the precipitation events, it is demonstrated that the marginal (i.e., unconditional) distributions of the characteristics are positively skewed. Examination of the conditional distributions of some pairs of characteristics indicates the existence of some relationships among the characteristics. For example, it is found that average intensity and maximum intensity are quite dependent on the event duration. The existence and forms of these relationships indicate that the assumption commonly made in stochastic models of hourly precipitation time series that the intensities (i.e., hourly amounts within an event) are independent and identically distributed must be violated. Again using exploratory data analysis, it is shown that the hourly intensities at Salem are, in fact, stochastically increasing and positively associated within a precipitation event.

### 1. Introduction

Precipitation occurrences and amounts and their characteristics are important factors in the analysis and modeling of many precipitation-sensitive phenomena (e.g., soil erosion, water resources management). In the case of soil erosion, several combinations of rainfall characteristics, soil temperature, and soil moisture status have been found to play important roles in controlling the timing and amount of soil loss on small agricultural watersheds in western Oregon (Istok and Kling, 1983). High-intensity rainfall and rainfall that occurs when the soil is saturated are two examples of factors that are important in causing soil erosion. Improved knowledge of the probabilities of occurrence of both the important factors and the combinations of factors would enhance the interpre-

tation of short-term erosion measurements and the ability to predict long-term erosion rates. However, in order to assign realistic probabilities to the occurrence of these conditions, it is first necessary to identify them in the available long-term climatic record.

One method of obtaining information regarding the characteristics of precipitation at a particular location and for a specific application (e.g., for modeling erosion) is the use of a concept called a precipitation "event." Such events are a convenient way of summarizing a time series of precipitation amounts into individual entities, defined so that they are meaningful in terms of a particular application. The concept of a precipitation event is not new. In fact, this approach has been applied in hydrology (e.g., Eagleson, 1970) and related fields for many years.

Basic characteristics of the precipitation events, such as duration and average intensity, can be calculated for each event, including those characteristics that are of interest for the specific application of concern (e.g., the factors that are important in causing

<sup>1</sup> Current address: Environmental and Societal Impacts Group, National Center for Atmospheric Research, Boulder, CO 80307.

<sup>2</sup> The National Center for Atmospheric Research is sponsored by the National Science Foundation.

erosion). An approach known as exploratory data analysis can provide useful information regarding the long-term probabilities of occurrence of precipitation events with particular characteristics. The exploratory data analysis approach can also be used to check whether certain properties, commonly assumed in stochastic models of precipitation time series, actually hold for real data.

In this paper we describe an exploratory analysis of the characteristics of precipitation events at a particular location in the Pacific Northwest and consider the implications of the results of the analysis for the stochastic modeling of precipitation. Section 2 consists of a brief description of the data used in the study and a discussion of the definition of a precipitation event and our analysis approach. The results of the exploratory analysis of some characteristics of precipitation events at Salem, Oregon, including marginal (i.e., unconditional) distributions of the precipitation event characteristics and conditional distributions of some pairs of characteristics, are described in Section 3. The implications of the results for the stochastic modeling of hourly precipitation time series are explored in Section 4. Finally, a summary and a discussion of current research are presented in Section 5.

## 2. Approach

The approach taken in this study consists of two steps. The first step is the condensation of a large quantity of hourly precipitation amounts into entities known as precipitation events. The second step is the exploratory statistical analysis of several characteristics of the events. These steps are discussed in this section following a brief description of the precipitation data base.

### a. Precipitation data base

The basic precipitation data used in this study consist of 31½ years (July 1948 to December 1979) of hourly precipitation amounts. These data were collected by the U.S. National Weather Service (NWS) at Salem, Oregon. The resolution of the hourly precipitation amounts is to the nearest 0.01 inch (0.25 mm) with a minimum measurable value of 0.01 inches (i.e., trace values are zeroed).

Salem is located in the Willamette Valley in western Oregon between the coastal and Cascade mountain ranges. Ninety percent of the average annual precipitation at Salem [41 inches (104 cm)] falls between 1 October and 31 May, with very little (none in some years) falling as snow. The analyses described in this paper were restricted to data from this so-called wet season (October to May). Of course, a more complete analysis would also consider the within-season variations of precipitation events.

### b. Precipitation event definition

For this study, a "precipitation event" is defined to be a period of one or more consecutive hours of precipitation followed by one or more hours of no precipitation. That is, two adjacent "wet" hours are considered to be part of the same precipitation event, whereas two wet hours that are separated by one or more dry hours are considered to belong to separate precipitation events. Specifically, let  $\{X_t: t = 1, 2, \dots\}$  denote a sequence of hourly precipitation amounts, where  $X_t$  is the precipitation amount in the  $t$ th hour ( $X_t \geq 0$ ). Then, for example, the sequence of precipitation amounts that satisfies

$$\begin{aligned} X_{t-l} > 0, \quad X_{t-l+1} = 0, \quad X_{t-l+2} = 0, \quad \dots, \quad X_t = 0, \\ X_{t+1} > 0, \quad X_{t+2} > 0, \quad \dots, \quad X_{t+k} > 0, \quad X_{t+k+1} = 0, \end{aligned} \quad (1)$$

contains a precipitation event that begins in the  $(t+1)$ th hour and ends in the  $(t+k)$ th hour and is preceded by  $l$  dry hours. A schematic example of the application of this precipitation event definition to a sequence of hourly precipitation amounts is presented in Fig. 1. For convenience, the hourly amounts of precipitation within an event (namely,  $X_{t+1}, X_{t+2}, \dots, X_{t+k}$ ) are referred to as "intensities."

Precipitation events formulated according to this definition can be termed "primitive" events. They are based on the simplest and most logical approach to defining precipitation events without the use of additional information (e.g., regarding the synoptic weather situation). Moreover, primitive events can easily be combined to create "composite" events that may be more appropriate for a particular application (e.g., for modeling the processes leading to soil erosion).

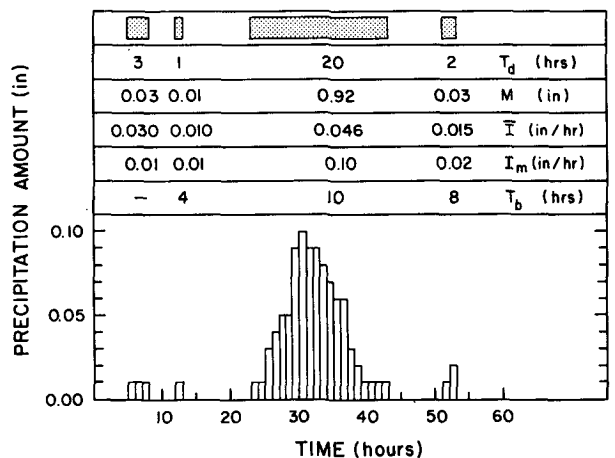


FIG. 1. Illustration of the calculation of five precipitation event characteristics for the four primitive precipitation events contained in a schematic time series of hourly precipitation amounts. The stippled areas at the top of the figure represent the time periods included in the four events.

Several characteristics of precipitation events are of interest, especially with regard to soil erosion. These characteristics include the duration  $T_d$ , magnitude  $M$ , average intensity  $\bar{I}$ , and maximum intensity  $I_m$  of the event, and the time between events  $T_b$ . The *duration* of an event is the total number of consecutive hours of precipitation, the *magnitude* is the total amount of precipitation that occurs during the event (i.e., the sum of the hourly intensities), the *average intensity* is the average rate of precipitation per hour during the event, the *maximum intensity* is the maximum hourly intensity within the event, and the *time between events* is the number of dry hours separating the event from the previous event. For the precipitation sequence just presented [see (1) and Fig. 1], the precipitation event characteristics are computed as follows:

$$T_d = k, \tag{2}$$

$$M = \sum_{j=1}^k X_{t+j}, \tag{3}$$

$$\bar{I} = \frac{1}{k} \sum_{j=1}^k X_{t+j} = M/T_d, \tag{4}$$

$$I_m = \max_{1 \leq j \leq k} X_{t+j}, \tag{5}$$

$$T_b = l. \tag{6}$$

Similar characteristics have been considered in other studies, but in some cases have been designated by other names. For example, our variable "magnitude" is denoted "accumulated rainfall depth" by Nguyen and Rousselle (1981) and others.

Additional characteristics that would be useful for a particular application could also be calculated for each precipitation event. For example, for the purpose of estimating soil erosion it might be of interest to know the total amount of precipitation that occurred in the 24 hours prior to the beginning of each event, as an index of the saturation level of the soil. Calculation of this additional characteristic would require more information about the precipitation record than is contained in the event characteristics actually considered, but it could easily be accomplished when the hourly amounts are first combined into events. Other types of characteristics could similarly be computed.

*c. Analysis methods*

The hourly precipitation amounts for the wet season at Salem were summarized by combining them into primitive precipitation events. A total of 8523 precipitation events were thus formulated from the 31½-year record of hourly precipitation amounts. Each of the five precipitation event characteristics just described (duration, magnitude, average intensity, maximum intensity, and time between events) was calculated for each event.

Some techniques of exploratory data analysis were used to examine the characteristics of the primitive precipitation events at Salem. This approach to data analysis is particularly appropriate due to the statistical properties of precipitation data. In particular, precipitation data generally have quite skewed distributions, so that the application of classical statistical methods based on the Gaussian distribution usually is inappropriate. Exploratory data analysis, moreover, relies on techniques that generally are less sensitive to the occurrence of extreme values (i.e., outliers). Finally, exploratory data analysis provides a wide array of useful methods for displaying, examining, and comparing data; these methods are particularly appropriate in summarizing large data sets for highly variable quantities. Kleiner and Graedel (1980) review the application of exploratory data analysis to geophysical data.

The exploratory analysis of the characteristics of the Salem primitive precipitation events consisted of two basic components. First, the statistical properties of the individual characteristics, including marginal distributions and order statistics, were examined. Second, the relationships between some pairs of the characteristics were evaluated through the examination of conditional distributions of the characteristics. The results of these analyses are presented in Section 3. Exploratory data analysis was also employed to investigate the statistical characteristics of the hourly precipitation intensities and, in particular, to determine whether the hourly intensities are independent and identically distributed (Section 4).

**3. Results of exploratory analysis**

*a. Individual characteristics*

The individual precipitation event characteristics are considered in this section in two ways, in terms of frequency distributions and in terms of order and return statistics, each of which provides a different type of information. The empirical marginal frequency distributions can be used to obtain general descriptions of the distributional properties of the characteristics. The order and return statistics allow estimation of the probabilities of equaling or exceeding specified values of the characteristics in a 31-year period and in a single year, respectively. The marginal distributions and order and return statistics of the Salem primitive event characteristics are described in the following subsections.

**3. Results of exploratory analysis**

*a. Individual characteristics*

The individual precipitation event characteristics are considered in this section in two ways, in terms of frequency distributions and in terms of order and return statistics, each of which provides a different type of information. The empirical marginal frequency distributions can be used to obtain general descriptions of the distributional properties of the characteristics. The order and return statistics allow estimation of the probabilities of equaling or exceeding specified values of the characteristics in a 31-year period and in a single year, respectively. The marginal distributions and order and return statistics of the Salem primitive event characteristics are described in the following subsections.

1) MARGINAL DISTRIBUTIONS

The empirical frequency distributions of the values of the individual precipitation event characteristics (i.e., the marginal distributions) can be conveniently

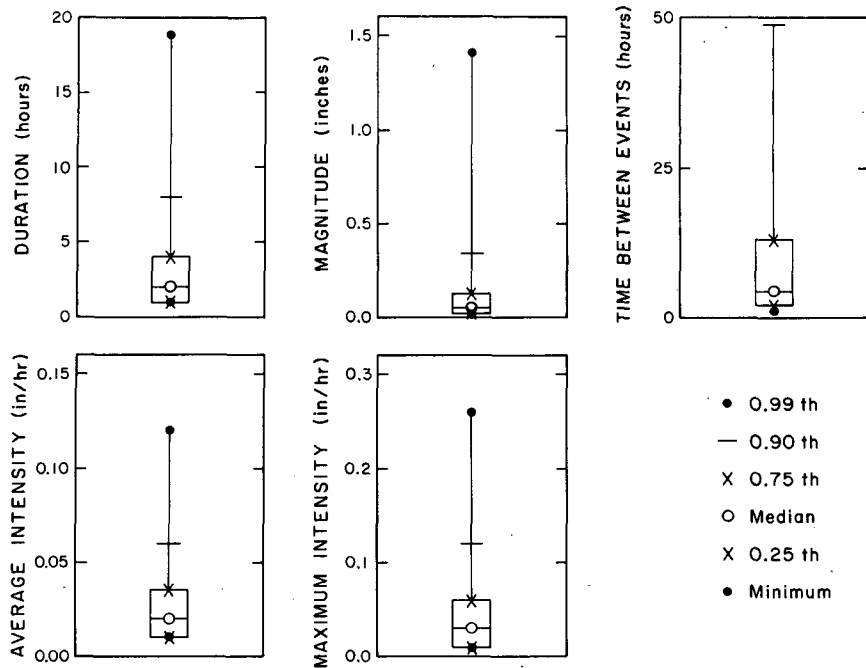


FIG. 2. Box plots of the distributions of precipitation event characteristics for wet season at Salem. The key for identifying the quantiles shown is given in the lower right corner.

displayed using box plots (Tukey, 1977). These displays provide a concise method of presenting several properties of the distributions simultaneously. Box plots of the distributions of the five precipitation event characteristics  $T_d$ ,  $M$ ,  $T_b$ ,  $\bar{I}$ , and  $I_m$  [defined by (2)–(6)] at Salem are shown in Fig. 2.

A major feature that is illustrated by the box plots in Fig. 2 is the skewed nature of the distributions of the five characteristics. That is, most of the precipitation events had relatively small values of the characteristics, whereas a few events had relatively large values. For example, the upper quartile (i.e., the 0.75th quantile) of magnitude is 0.13 inches and the 0.99th quantile is 1.41 inches; in other words, 75% of the events produced less than 0.13 inches of precipitation, whereas 1% produced more than 1.41 inches. This result is not unexpected since the skewed nature of the distributions of precipitation variables has been noted in previous studies (e.g., Neyman and Scott, 1967; Biondini, 1976).

The box plots in Fig. 2 can be used to obtain other general information about the distributions of the precipitation event characteristics. For example, note that the quantiles of the distribution of maximum intensity are approximately twice as large as the corresponding quantiles of the distribution of average intensity. This result suggests that, on average, the largest hourly rainfall intensity during a precipitation event is approximately twice as large as the average hourly intensity. Moreover, specific probability estimates may be taken directly from the box plots (e.g.,

the estimated probability of occurrence of a precipitation event with a maximum intensity exceeding 0.12 inches per hour is 0.10).

## 2) ORDER AND RETURN STATISTICS

The extreme values of the characteristics that will be exceeded with a specified expected frequency in a 31-year period (the period of record, rounded down from 31½ years) and in a single year can be estimated using order and return statistics, respectively. The order statistics are simply the values of a particular characteristic ranked from largest to smallest, whereas the return statistics consist of every 31st order statistic. Thus, the  $n$ th order statistic for a particular characteristic is an estimate of the value that the characteristic would be expected to equal or exceed  $n$  times in a 31-year period, whereas the  $n$ th return statistic is an estimate of the value that the characteristic would be expected to equal or exceed  $n$  times in a single year. For example, a precipitation event that lasts as long or longer than the tenth duration order statistic would be expected to occur approximately ten times in 31 years. A precipitation event that lasts as long or longer than the tenth duration return statistic would be expected to occur about ten times in a single year.

The first 100 maximum intensity order and return statistics are shown in Fig. 3. As would be expected based on the maximum intensity box plot in Fig. 2, the first few order statistics of maximum intensity are

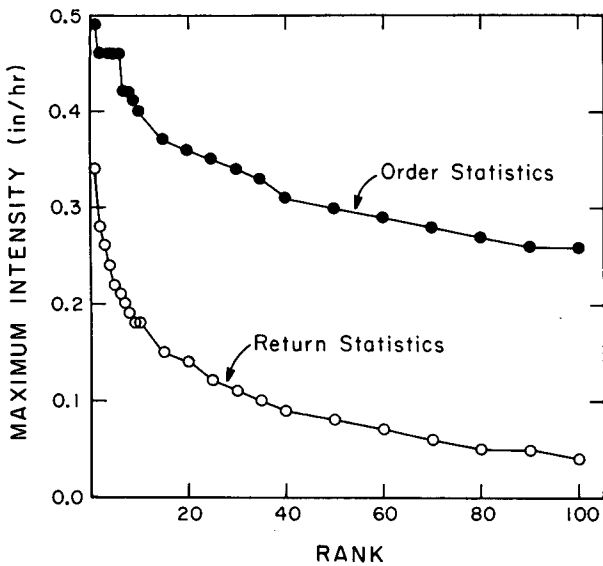


FIG. 3. Maximum intensity order and return statistics for wet season at Salem. All of the first 10 points are plotted, every fifth point is plotted for ranks between 10 and 40, and every tenth point is plotted for ranks larger than 40.

very large and subsequent values are much smaller. Also, the slope of the order statistic curve is quite steep for the first few values and much flatter for later values. This result is due to the relatively long right-hand tail of the distribution of maximum intensity illustrated in Fig. 2. The return statistic curve has a similar shape. However, since a larger portion of the range of maximum intensity is covered by the return statistics, the slope is even steeper for the early return statistics than for the early order statistics. The curves in Fig. 3 provide a large amount of information about the distribution of extreme values of maximum intensity. For example, based on these statistics, an event with a maximum intensity of at least 0.40 inches per hour ( $\text{in h}^{-1}$ ) can be expected to occur about ten times in 31 years; an event with a maximum intensity of at least 0.20  $\text{in h}^{-1}$  can be expected seven times in a single year. The curves of the order and return statistics for the other precipitation event characteristics are quite similar in shape to the maximum intensity curves shown in Fig. 3 and are not presented here because of space limitations.

*b. Pairs of characteristics*

As noted in Section 1, it is frequently of interest to have information about the probability of occurrence of various combinations of factors (e.g., the occurrence of high-intensity precipitation over a long period of time). If two factors (i.e., characteristics) are statistically independent, information about their joint behavior can simply be obtained using the

marginal distributions of the two characteristics. However, if the characteristics are not independent, it is necessary to examine the actual joint distributions of the characteristics in order to obtain estimates of these probabilities.

Hence, it is of interest to consider the joint distributions of pairs of precipitation event characteristics in order to investigate the types of relationships (i.e., dependencies) that may exist. In particular, the relationships between duration and magnitude, duration and average intensity, and duration and maximum intensity are examined in this section. Another type of dependence that could be investigated is the relationship between these four characteristics,  $T_d$ ,  $M$ ,  $\bar{I}$ , and  $I_m$ , and the time between events  $T_b$ . However, previous evaluations of these data have provided no indication that this type of dependence exists. In this section consideration is limited to the observed relationships among the event characteristics. As an alternative approach, theoretical relationships among characteristics can be derived on the basis of stochastic models for hourly precipitation time series. These theoretical relationships will be discussed in Section 4.

The conditional distributions of the characteristics given a wide range of values of duration can provide some insight into the types of relationships that exist between the precipitation event characteristics and the duration of the event. These distributions can most easily be displayed by means of conditional quantiles as in Fig. 4. Similar displays based on "moving" statistics have been employed by Cleveland and Kleiner (1975) and Katz (1978).

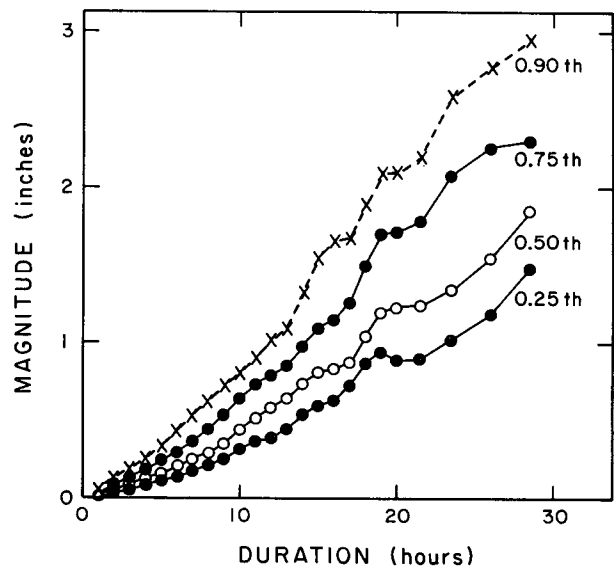


FIG. 4. Quantiles of the conditional distributions of magnitude given duration for wet season at Salem.

### 1) MAGNITUDE GIVEN DURATION

Figure 4 contains quantiles of the conditional distributions of magnitude  $M$  given values of duration  $T_d$  ranging between 1 and  $\sim 30$  hours. For example, the median magnitude for events lasting 10 hours is about 0.45 inches. The conditional quantiles are connected by lines so that four curves are presented: the 0.25th, 0.50th, 0.75th, and 0.90th quantiles of the conditional distributions. The curves were smoothed using hanning (Tukey, 1977) in order to make the trends with increasing duration more visible. Note that due to the smaller numbers of events with large durations, the concomitant quantiles are for conditional distributions given an interval of duration values rather than a single value.

The curves in Fig. 4 indicate that there is a strong relationship between precipitation event duration and magnitude; namely, the longer the event, the greater the amount of precipitation accumulated. This relationship, evidenced by the increasing conditional quantiles of magnitude with increasing duration, should have been expected (e.g., see Eagleson, 1970). In fact, if intensity were independent of duration, a roughly linear increase in the conditional median of magnitude with increasing duration would be present. The variability of magnitude also appears to depend on the duration, as indicated by the increase in the distance between the upper and lower quartiles (i.e., the interquartile range) as the duration becomes longer.

### 2) AVERAGE INTENSITY GIVEN DURATION

The quantiles of the conditional distributions of average intensity  $\bar{I}$  given duration  $T_d$  are displayed in Fig. 5. The trend in these quantiles is not as strong

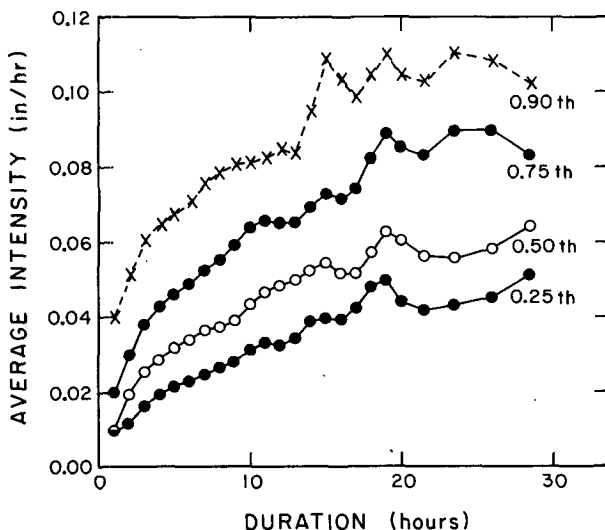


FIG. 5. As in Fig. 4 but for average intensity.

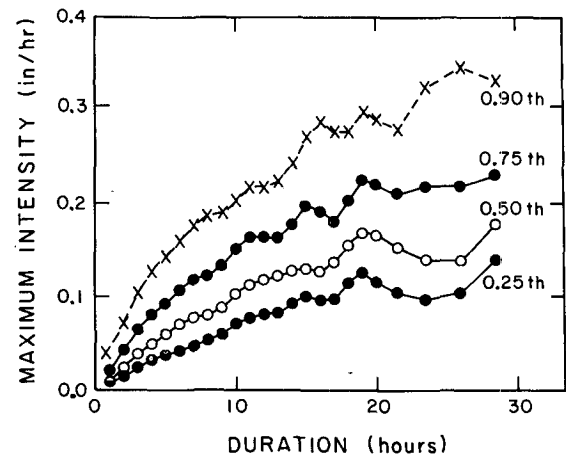


FIG. 6. As in Fig. 4 but for maximum intensity.

as the trend for the quantiles of magnitude given duration. [Note that since  $\bar{I}$  is related to  $M$  by (4), this difference in trends is to be expected.] Nevertheless, there clearly is a relationship between these two characteristics, as indicated by the gradual increase in the conditional quantiles of average intensity with increasing values of duration. For example, the median value of average intensity for events lasting 5 hours is only slightly more than  $0.03 \text{ in h}^{-1}$ , whereas the median value of average intensity for events lasting 20 hours is about  $0.06 \text{ in h}^{-1}$ .

### 3) MAXIMUM INTENSITY GIVEN DURATION

The relationship between maximum intensity  $I_m$  and duration  $T_d$  also is quite strong, as indicated by the curves in Fig. 6. The existence of this relationship also should have been expected, as will be discussed in Section 4. As shown in Fig. 6, the conditional quantiles and the interquartile range of maximum intensity are increasing functions of duration, at least for events lasting 20 hours or less. Apparently, the probability of a large maximum intensity value is greater for long-duration precipitation events than for short-duration events. For example, less than 25% of the events lasting 5 hours had a maximum intensity of at least 0.10 inches, whereas this percentage is greater than 50 for events lasting 10 hours. The implications of this result and others described in this section for the stochastic modeling of hourly precipitation time series are considered in Section 4.

### 4. Implications for stochastic modeling of precipitation

By means of exploratory data analysis, certain relationships among the various precipitation event characteristics have been detected. In this section, the sequence of hourly precipitation amounts is viewed as a realization of a stochastic process, and we attempt to identify through exploratory data analysis

the properties of this stochastic process that are responsible for these relationships among the precipitation event characteristics. Note that the stochastic modeling approach provides an alternative method for estimating the probabilities of occurrence of precipitation events with specified characteristics. If an adequate model were available, then it could be used to generate the desired probabilities, either analytically or by simulation. Waymire and Gupta (1981) and the Committee on Precipitation, AGU Hydrology Section (1984) provide recent reviews of stochastic precipitation models.

The simplest stochastic models for hourly precipitation amounts assume that, given a precipitation event occurs from the  $(t + 1)$ th hour through the  $(t + k)$ th hour, the hourly precipitation amounts,  $X_{t+1}, X_{t+2}, \dots, X_{t+k}$ , within the event (i.e., the intensities) are conditionally independent and identically distributed (iid), say with common distribution function (df)  $F$  (e.g., Nguyen and Rousselle, 1981). First, the nature of the theoretical relationships among certain precipitation event characteristics is considered under this iid requirement for the intensities. For simplicity and to avoid any confounded effects due to seasonal nonstationarity, only hourly precipitation data for the month of January at Salem are analyzed (i.e., 31 Januaries from 1949 through 1979).

*a. Theoretical event relationships*

1) AVERAGE INTENSITY GIVEN DURATION

The conditional distribution of average intensity  $\bar{I}$  given the duration  $T_d$ , under the iid assumption for the intensities, is discussed first. Given a precipitation event of, say, duration  $T_d = k$  hours, it follows from well-known properties of the sample mean that the average intensity  $\bar{I}$  of this event has conditional expected value

$$E(\bar{I}|T_d = k) = \mu, \tag{7}$$

and conditional variance

$$\text{var}(\bar{I}|T_d = k) = \frac{\sigma^2}{k}, \tag{8}$$

where  $\mu$  and  $\sigma^2$  are the expected value and variance for the df  $F$  of hourly intensities. In other words, if the intensities are iid, then the average intensity of a precipitation event has conditional expected value (7) that is constant (i.e., it does not depend on the given duration of the event) and has conditional variance (8) that decreases as the given duration increases.

However, because the distributions are skewed, the diagrams employed in Section 3 to portray the conditional distributions rely on quantiles, rather than means and variances. Under the assumption that the intensities are iid with a positively skewed df  $F$  and given a precipitation event of duration  $T_d = k$  hours,

the average intensity  $\bar{I}$  has conditional median that gradually increases to the same limiting value  $\mu$  as the conditional mean (7) and has conditional interquartile range that decreases at a rate approximately proportional to  $1/k$  for large duration  $k$ . To actually compute these theoretical conditional quantiles, it is further assumed that the df  $F$  of the hourly intensities is a member of the gamma family with density function

$$F'(x) = \lambda^\alpha x^{\alpha-1} e^{-\lambda x} / \Gamma(\alpha), \quad x > 0, \tag{9}$$

where  $\alpha > 0$  is the "shape" parameter,  $\lambda > 0$  is the "scale" parameter, and  $\Gamma$  denotes the gamma function (e.g., Johnson and Kotz, 1970, Chapter 17). Under this assumption that  $F$  is a gamma df and given a precipitation event of duration  $T_d = k$  hours, the average intensity  $\bar{I}$  has conditional distribution that is gamma with shape parameter  $\alpha' = k\alpha$  and the same scale parameter  $\lambda' = \lambda$ . Using this result, the conditional quantiles can be calculated by inverting the gamma df (9) with the appropriate shape and scale parameters.

Figure 7a shows the theoretical quantiles for the conditional distribution of average intensity given

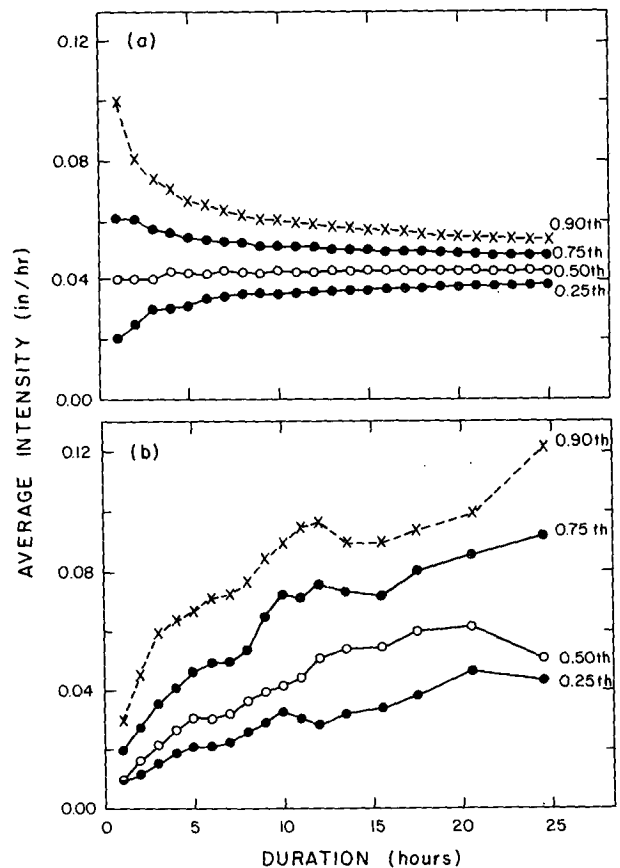


FIG. 7. Quantiles of the conditional distributions of average intensity given duration for January at Salem: (a) theoretical, (b) actual.

duration, under the assumption that the intensities are iid with a gamma df. These quantiles were computed using parameters  $\alpha$  and  $\lambda$  of the gamma df that were estimated from all of the January intensities at Salem. Figure 7b shows the corresponding actual conditional quantiles of average intensity for January at Salem, which are quite similar to the actual quantiles based on the entire wet season (Fig. 5). Note that the theoretical and actual conditional quantiles of average intensity behave quite differently as functions of the given event duration.

## 2) MAXIMUM INTENSITY GIVEN DURATION

We now consider the conditional distribution of maximum intensity  $I_m$  given the duration  $T_d$ , under the iid assumption for the intensities. Given a precipitation event of, say, duration  $T_d = k$  hours, it follows directly from the definition of the sample maximum that the maximum intensity  $I_m$  of this event has conditional distribution

$$\Pr(I_m \leq x | T_d = k) = [F(x)]^k. \quad (10)$$

Equation (10) implies that the  $p$ th quantile,  $x_p(k)$  say, of this conditional distribution of  $I_m$  is given by

$$x_p(k) = F^{-1}(p^{1/k}). \quad (11)$$

In other words,  $x_p(k)$  is the  $(p^{1/k})$ th quantile of the df  $F$ . Thus the theoretical quantiles can be computed directly from  $F$ , without any assumption regarding its parametric form. The exact behavior of the theoretical quantiles of the conditional distribution of maximum intensity as the duration increases depends on the specific functional form of the df  $F$ . By analogy to the known asymptotic behavior of the distribution of the maximum in the case of  $F$  being an exponential df (David, 1981, p. 263), for large duration  $k$  the maximum intensity should have a conditional median that increases approximately like  $\ln k$  and a conditional interquartile range that is approximately constant.

Figure 8a shows the theoretical quantiles for the conditional distribution of maximum intensity given duration, under the assumption that the intensities are iid with common df  $F$ . These quantiles were computed by (11), using the empirical df based on all of the January intensities at Salem to estimate  $F$ . Figure 8b shows the corresponding actual conditional quantiles of maximum intensity for January at Salem, which are quite similar to the actual quantiles based on the entire wet season (Fig. 6). Note that the theoretical and actual conditional quantiles of maximum intensity behave somewhat differently as functions of the given event duration.

### b. Check on iid assumption

These discrepancies between the theoretical and observed conditional quantiles of average intensity  $\bar{I}$

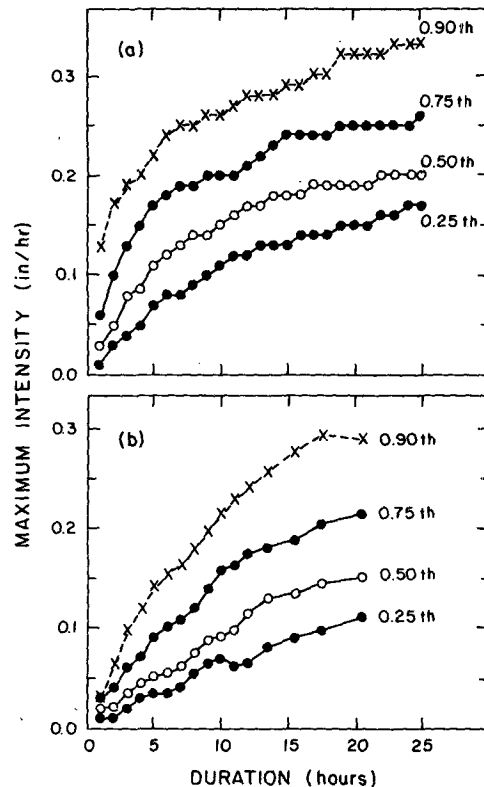


FIG. 8. As in Fig. 7 but for maximum intensity.

and maximum intensity  $I_m$ , given the event duration  $T_d$ , suggest that the assumption that the intensities,  $X_{t+1}, X_{t+2}, \dots, X_{t+k}$  say (using the example in Section 2b, part 1), are iid must be violated. By studying the marginal and joint distributions of these intensities, we check on the assumptions of independence and of identical distributions. First, the requirement that  $X_{t+1}, X_{t+2}, \dots, X_{t+k}$  each have identical distributions is examined.

## 1) IDENTICAL DISTRIBUTIONS

Figure 9 shows how the distribution of intensity on the  $j$ th hour of the event varies as  $j$  increases ( $j = 1, 2, \dots$ ). Increasing trends in the quantiles as the event progresses are evident, with the trends being most marked for the upper quartile and the 90th percentile. In other words, the intensities have distributions whose central tendency as measured by the median, and variability as measured by the interquartile range, both increase as the event progresses, with the increase in variability being more rapid than the increase in central tendency. It is apparent that the intensity on the  $(j+1)$ th hour within the event is stochastically larger than the intensity on the  $j$ th hour within the event; that is,

$$\Pr(X_{t+j} \geq x) \leq \Pr(X_{t+j+1} \geq x) \quad (12)$$



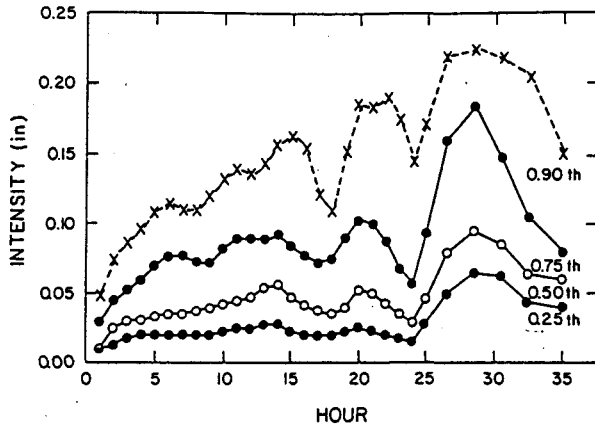


FIG. 9. Quantiles of the distributions of intensity as a function of the hour within a precipitation event for January at Salem.

for every  $x, j = 1, 2, \dots, k$ . In particular, these results make it clear that the distributions of the intensities are not identical, implying that stochastic models of hourly precipitation time series should allow the  $df F$  to vary as a function of the hour  $j$  within the event. Of course, a more detailed analysis might reveal other ways in which the distribution of hourly intensities varies. For instance, it might vary diurnally (i.e., depending on the time of day) or it might depend on the duration of the precipitation event (e.g., being stochastically smaller on the last hour of an event).

2) INDEPENDENCE

We now consider the assumption that the intensities,  $X_{t+1}, X_{t+2}, \dots, X_{t+k}$ , are independent. Figure 10 portrays the relationship between the intensities on the first and second hours of an event (i.e., between  $X_{t+1}$  and  $X_{t+2}$ ). The quantiles of the conditional distribution of  $X_{t+2}$ , given  $X_{t+1}$ , increase as the given value of  $X_{t+1}$  increases, with the increase being more marked the higher the quantile. A similar degree of positive association characterizes the relationships between the intensities on the second and third hours within an event (i.e., between  $X_{t+2}$  and  $X_{t+3}$ ) and between intensities on later hours (figures not included in this paper). In particular, these results make it clear that the intensities are not independent, implying that stochastic models of hourly precipitation time series should allow for a certain degree of positive dependence. We note that Nguyen (1983) has proposed one particular form of stochastic model that allows the hourly intensities to be correlated.

c. Comparison of theoretical and empirical event relationships

The effects of these violations of the iid requirement for intensities are now discussed, in an attempt to

explain the discrepancies between the theoretical and empirical quantiles of the conditional distributions of average intensity  $\bar{I}$  and maximum intensity  $I_m$  given the event duration  $T_d$ . The differences illustrated in Fig. 7 between the theoretical and empirical quantiles of the conditional distributions of average intensity  $\bar{I}$ , given the event duration  $T_d$ , are considered first. Second, the possible causes of the discrepancies between the theoretical and empirical quantiles of maximum intensity  $I_m$ , given the event duration  $T_d$  (Fig. 8), are discussed.

1) AVERAGE INTENSITY GIVEN DURATION

We recall from Fig. 7 that the theoretical conditional median of average intensity slightly increases at first and then is roughly constant for larger durations, whereas the actual conditional median increases quite sharply. In addition, the theoretical conditional interquartile range of average intensity decreases rapidly at first and then more slowly, whereas the actual conditional interquartile range increases quite markedly. The positive association among the intensities would only have the effect of increasing the conditional interquartile range, leaving the conditional median unchanged (this result is analogous to the inflation of the variance by positive dependence discussed by Jones, 1975). Thus, the increasing trend in the actual conditional median of average intensity as the duration increases must be attributable to the fact that the intensities are not identically distributed, but rather are stochastically increasing (12). On the other hand, the increasing trend in the actual conditional interquartile range of average intensity is attributable to both the fact that the intensities are stochastically increasing and that they are positively associated.

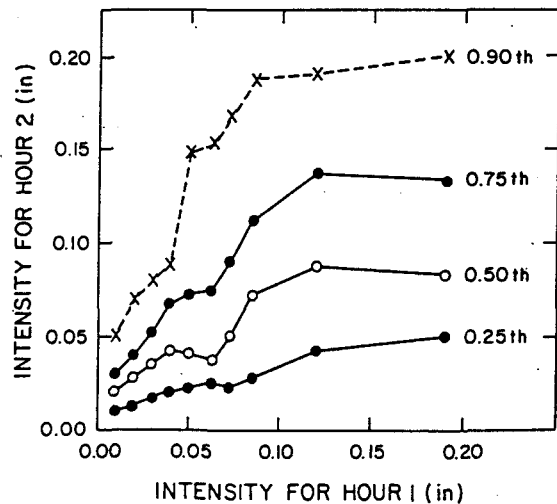


FIG. 10. Quantiles of the conditional distributions of intensity on the second hour of a precipitation event given the intensity of the first hour for January at Salem.

## 2) MAXIMUM INTENSITY GIVEN DURATION

As shown in Fig. 8, the actual conditional median of maximum intensity increases more slowly at first than the theoretical conditional median and remains smaller for all durations. In addition, the actual conditional interquartile range increases more slowly at first than the theoretical conditional interquartile range but then increases more rapidly to nearly coincide with the theoretical range for larger durations. The positive association among the intensities would reduce the conditional median, but could either increase or decrease the conditional interquartile range (decrease it slightly if the  $F$  were exponential). Thus, the actual conditional median of maximum intensity is reduced both because the intensities are stochastically increasing and because they are positively associated. On the other hand, the actual conditional interquartile range is reduced for smaller event durations, probably primarily because the intensities are stochastically increasing.

## 5. Summary and concluding remarks

The results presented in the preceding sections provide a general statistical description of precipitation events based on hourly precipitation data at Salem, Oregon. Some conclusions that can be drawn regarding the characteristics of primitive precipitation events at Salem include the following:

- 1) The marginal distributions of the characteristics are strongly positively skewed. That is, precipitation events with relatively small values of the characteristics occur most frequently and events with large values of the characteristics occur relatively infrequently. This result is consistent with previous statistical descriptions of precipitation characteristics.
- 2) The extreme values of the precipitation event characteristics rapidly decrease from very large initial values to much smaller subsequent values, as evidenced by curves of order and return statistics.
- 3) The values of some pairs of precipitation event characteristics are not independent. For example, the values of precipitation event magnitude, average intensity, and maximum intensity are all related to the duration of the precipitation event.

Because of the relationships between average intensity and duration and between maximum intensity and duration at Salem, the assumption commonly made in stochastic models of hourly precipitation time series that the intensities are independent and identically distributed must be violated. By means of exploratory analysis, we have shown that the hourly intensities are, in fact, stochastically increasing and are positively associated within a precipitation event. These statistical properties of the hourly intensities have been employed to help explain the discrepancies between the actual relationships among event char-

acteristics and the theoretical relationships based on independent and identically distributed intensities.

In this paper we have considered the characteristics of "primitive" precipitation events at one location in the Pacific Northwest. These events were created using only the information contained in the precipitation record and without consideration of any particular application. Of course, it usually would be desirable to apply a precipitation event definition that is specifically selected for a certain application. For example, we are currently investigating the application of a precipitation event definition that is appropriate for estimating the effects of precipitation in causing soil erosion in the Pacific Northwest. The erosion-specific events would be derived from primitive precipitation events by taking the union of any events separated by less than six hours, based on studies of the hydrologic response of soil to precipitation. Preliminary results of exploratory analyses of the characteristics of the erosion-specific events at five locations in the Pacific Northwest (including Salem) are described in two reports (Brown *et al.*, 1983a,b). It is of interest to note that the conclusions listed earlier in this section regarding primitive events at Salem apparently also characterize, in general, the erosion-specific events at the five locations. Precipitation events that are appropriate for other applications (e.g., forest fire management) could similarly be constructed.

The concept of an "event" also can be used to summarize other types of meteorological time series, such as wind speed or temperature observations, in a way that is both convenient and meaningful. For example, the occurrence and intensity of periods of freezing temperatures are of importance in characterizing erosion potential. These periods can easily be investigated by condensing the temperature record into "freezing" events, where each freezing event consists of a period of temperatures that are below freezing followed by a period of temperatures that are above freezing. Characteristics of both cold and warm portions of a long series of freezing events can provide considerable information about periods of freezing and thawing at a particular location. An exploratory analysis of the characteristics of freezing events at several locations in the Pacific Northwest is currently being undertaken and preliminary results have been described in two reports (Brown *et al.*, 1983c; Brown *et al.*, 1984).

The tools of exploratory data analysis used in this study have produced a substantial amount of information about the statistical properties of precipitation events. The results of the analyses have provided general descriptions of the precipitation event characteristics and insights into the relationships among the characteristics. Clearly, the use of exploratory data analysis is advantageous and rewarding in studies in which the variables possess both high variability

and asymmetric distributions, and in which the objective is to obtain a general description of the statistical properties of a phenomenon. Moreover, the analyses described in this paper could easily be used to study other types of climatological events, including other types of precipitation events that may be appropriate for different applications.

*Acknowledgments.* Bruce A. Peterson provided much of the computer and data base support for this project. His contribution is gratefully acknowledged. The authors would also like to thank Jonathan D. Istok, Larry L. Boersma, and Moyle E. Harward for their roles in initiating, supporting, and aiding this research. This work was conducted at Oregon State University and was supported by funds provided by the STEEP (Solutions to Environmental and Economic Problems) program and by the Oregon Agricultural Experiment Station. STEEP is administered by the Science and Education Administration, Cooperative Research, U.S. Department of Agriculture.

#### REFERENCES

- Biondini, R., 1976: Cloud motion and rainfall statistics. *J. Appl. Meteorol.*, **15**, 205-224.
- Brown, B. G., J. D. Istok, R. W. Katz and A. H. Murphy, 1983a: Statistical analysis of climatological data to characterize erosion potential: 1. Precipitation events in western Oregon. Oregon Agricultural Experiment Station Spec. Rep. No. 686, Oregon State University, 287 pp.
- , —, — and —, 1983b: Statistical analysis of climatological data to characterize erosion potential: 2. Precipitation events in eastern Oregon/Washington. Oregon Agricultural Experiment Station Spec. Rep. No. 687, Oregon State University, 178 pp.
- , —, — and L. Boersma, 1983c: Statistical analysis of climatological data to characterize erosion potential: 3. Freezing events in western Oregon. Oregon Agricultural Experiment Station Spec. Rep. No. 688, Oregon State University, 146 pp.
- , —, —, — and —, 1984: Statistical analysis of climatological data to characterize erosion potential: 4. Freezing events in eastern Oregon/Washington. Oregon Agricultural Experiment Station Spec. Rep. No. 689, Oregon State University, 117 pp.
- Cleveland, W. S., and B. Kleiner, 1975: A graphical technique for enhancing scatterplots with moving statistics. *Technometrics*, **17**, 447-454.
- Committee on Precipitation, AGU Hydrology Section, 1984: A new interdisciplinary focus on precipitation research. *Eos*, **65**, 377-380.
- David, H. A., 1981: *Order Statistics*, 2nd ed. Wiley & Sons, 360 pp.
- Eagleson, P. S., 1970: *Dynamic Hydrology*. McGraw-Hill, 462 pp.
- Istok, J. D., and G. F. Kling, 1983: Effect of subsurface drainage on runoff and sediment yield from an agricultural watershed in western Oregon, U.S.A. *J. Hydrol.*, **65**, 279-291.
- Johnson, N. L., and S. Kotz, 1970: *Distributions in Statistics: Continuous Univariate Distributions—1*. Wiley & Sons, 300 pp.
- Jones, R. H., 1975: Estimating the variance of time averages. *J. Appl. Meteorol.*, **14**, 159-163.
- Katz, R. W., 1978: Persistence of subtropical African droughts. *Mon. Wea. Rev.*, **106**, 1017-1021.
- Kleiner, B., and T. E. Graedel, 1980: Exploratory data analysis in the geophysical sciences. *Rev. Geophys. Space Phys.*, **18**, 699-717.
- Neyman, J., and E. L. Scott, 1967: Some outstanding problems relating to rain modification. *Proc. Fifth Berkeley Symp. on Mathematical Statistics and Probability*, **5**, 293-326.
- Nguyen, V.-T.-V., 1983: Stochastic modeling of rainfall temporal patterns. *Preprints Second Int. Meeting on Statistical Climatology*, National Institute of Meteorology and Geophysics, Lisbon, 17.1.1.-17.1.4.
- , and J. Rousselle, 1981: A stochastic model for the time distribution of hourly rainfall depth. *Water Resour. Res.*, **17**, 399-409.
- Tukey, J. W., 1977: *Exploratory Data Analysis*. Addison-Wesley, 688 pp.
- Waymire, E. C., and V. J. Gupta, 1981: The mathematical structure of rainfall representations. I. A review of the stochastic rainfall models. *Water Resour. Res.*, **17**, 1261-1272.