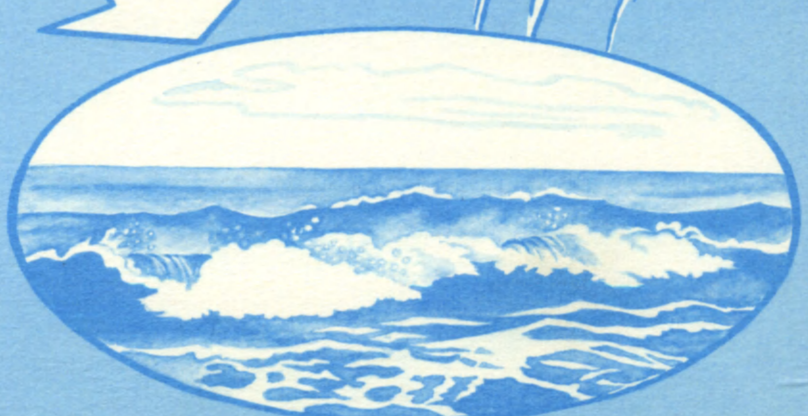
Technical Report No. 18

June, 1973

A Stochastic Approach to Space-time Modeling of Rainfall

by
Vijay Kumar Gupta



Technical Reports on
Natural Resource Systems

The University of Arizona
Tucson, Arizona 85721

A STOCHASTIC APPROACH TO SPACE-TIME MODELING OF RAINFALL

by

Vijay Kumar Gupta

Reports on Natural Resource Systems
Report No. 18

June 1973

Collaborative effort between the following Departments:
Hydrology and Water Resources
Systems and Industrial Engineering

University of Arizona
Tucson, Arizona 85721

PREFACE

This report constitutes the Ph.D. dissertation of the same title completed by the author in May, 1973, and accepted by the Department of Hydrology and Water Resources.

The candidate was partially supported by funds provided by the Office of Water Resources Research (Grant No. 14-31-0001-3708) and the National Science Foundation (Grant No. GK-35915). This support is gratefully acknowledged. However, the results are solely the responsibility of the candidate and his committee.

The report is another result of a continuing and informal effort between faculty and students on this campus in an area that we choose to call Natural Resource Systems. Competence in aspects of this subject may be found in many Colleges on campus. Vijay Gupta's dissertation seeks to show the extent to which the axiomatic approach of probability theory can be used to model the random space-time variability of hydrologic processes. Much work along these lines seems appropriate in the larger context of natural resource systems.

This report series constitutes an effort to communicate to practitioners and researchers the complete research results, including economic foundations and detailed theoretical development that cannot be reproduced in professional journals. These reports are not intended to serve as a substitute for the review and referee process exerted by the scientific and professional community in these journals.

Chester C. Kisiel and Lucien Duckstein

Departments of Hydrology & Water
Resources and Systems & Industrial
Engineering

Dedicated to

my Parents

ACKNOWLEDGMENTS

I would like to extend my deep gratitude to Dr. C. C. Kisiel, my major professor, along with Dr. L. Duckstein, for the constant encouragement, involvement and the general counsel they provided throughout the course of this study. I am grateful to the other members of my committee, namely, Drs. M. Fogel, J. Denny and J. L. Sanders in providing a constructive criticism on the written manuscript and for serving on the graduate committee. I am deeply indebted to Mr. David R. Dawdy for providing me with valuable comments throughout the course of this study, and serving on the graduate committee as a guest member.

I am grateful to Dr. P. Todorovic of Colorado State University for his very generous counsel in the early stages of this study.

To Dr. R. N. Bhattacharaya of the Department of Mathematics, University of Arizona, goes my deep indebtedness for his major help in proving Theorem 1 on the rate of convergence, and his willingness to be always available for some very valuable discussions that I had with him.

I acknowledge the Water Resources Research Center of The University of Arizona for generously allowing me to use their data for the numerical example.

I extend my appreciation to Mrs. Sarah Oordt for doing an excellent job of typing this manuscript.

This research was supported by an allotment grant on "Stochastic Space-time Models of the Rainfall-runoff Process" and title-1 matching grant on "Decision Analysis of Watershed Management Alternatives" from the office of W. R. R. of the U. S. Dept. of Interior.

TABLE OF CONTENTS

TABLE OF CONTENTS--Continued

LIST OF ILLUSTRATIONS

LIST OF TABLES

ABSTRACT

This study gives a phenomenologically based stochastic model of space-time rainfall. Specifically, two random variables on the spatial rainfall, e.g., the cumulative rainfall within a season and the maximum cumulative rainfall per rainfall event within a season are considered. An approach is given to determine the cumulative distribution function (c.d.f.) of the cumulative rainfall per event, based on a particular random structure of space-time rainfall. Then the first two moments of the cumulative seasonal rainfall are derived based on a stochastic dependence between the cumulative rainfall per event and the number of rainfall events within a season. This stochastic dependence is important in the context of the spatial rainfall process. A theorem is then proved on the rate of convergence of the exact c.d.f. of the seasonal cumulative rainfall up to the i^{th} year, $i \geq 1$, to its limiting c.d.f. Use of the limiting c.d.f. of the maximum cumulative rainfall per rainfall event up to the i^{th} year within a season is given in the context of determination of the 'design rainfall'. Such information is useful in the design of hydraulic structures.

Special mathematical applications of the general theory are developed from a combination of empirical and phenomenological based assumptions. A numerical application of this approach is demonstrated on the Atterbury watershed in the Southwestern United States.

CHAPTER 1

INTRODUCTION

The purpose of this study is to develop a stochastic model of the spatial rainfall process. The last decade and a half has witnessed an emergence of numerous efforts expended towards modeling of different aspects of stochastic hydrologic processes. In particular the processes of rainfall and streamflow have been modeled extensively because they form an integral part of the design and operation of water resources systems, such as detention storage reservoirs, dams, emergency spillways, flood control structures like spurs and barrages, natural recharge of aquifers, natural and artificial irrigation, to mention a few.

An important aim of any model of stochastic hydrologic process is to determine, if possible, the form of probability law that governs the outcome of random events of interest. Examples of such random events may include exceedance of maximum annual flow above the capacity of emergency spillway, flooding of a storage reservoir due to cumulative flows exceeding its capacity, and others. Such random events can be defined in terms of random variables, e.g., annual maximum or seasonal maximum peak flow, mean annual cumulative rainfall, etc., on the underlying hydrologic process.

In the present study, basically two random variables are considered for the process of spatial rainfall, e.g., the cumulative rainfall within a season and the maximum cumulative rainfall per rainfall event within a season. However, the same random variables can also be considered for the process of cumulative water yield, based on a simple transformation to calculate excess water yield from the cumulative rainfall per rainfall event. The nature of this transformation is also given in this study. The approach used here for the modeling stems from the mathematical theory of stochastic processes and to some extent is similar to the ones used in the past for the point rainfall process (Todorovic, 1968; Verschuren, 1968). Such an approach can be called phenomenological in the sense that most of the assumptions are justified in terms of the nature of rainfall phenomena. However, some assumptions are made to induce simplification in the approach.

To this end, a literature review of stochastic models in hydrology is given in Chapter 2. Only those models are considered that attempt to determine the form of the cumulative distribution function (c.d.f.) underlying random variables of interest. In order to give a general unifying perspective on the diverse stochastic models that exist in hydrology, models are classified into empirical, phenomenological and physical stochastic models. Specific models on the rainfall process and the process of excess water yield are then briefly reviewed under the guidelines of this classification.

The general theoretical considerations are then presented in Chapter 3. The theory is first developed within a fixed time interval (season). The salient features of the theory are as follows. The process of spatial rainfall is considered to be an intermittent process such that its evolution on the time axis is divided into an alternating sequence of random intervals called wet and dry intervals. During a wet interval rainfall occurs somewhere over a gaged region surrounding one or more basins, and during a dry interval no rainfall occurs over the region. The rainfall occurring within a wet interval forms a rainfall event which may correspond to a natural storm that may pass over the region. Based on this formulation, the number of rainfall events within a season is a positive integer valued random variable. The probability mass function (p.m.f.) of this can be obtained from the c.d.f. of random intervals denoting dry and wet intervals.

Next, within any given wet interval, a phenomenological approach is given to determine the c.d.f. of cumulative rainfall from a river basin. This approach is based on the stochastic process of number of storm cells that may occur over the basin, and the random variables denoting cumulative rainfall from these cells. Since the c.d.f. of the cumulative rainfall per event depends stochastically on the wet random interval, a stochastic dependence arises between the random variable denoting the cumulative rainfall per rainfall event and the random variable denoting number of such rainfall events within a season. This dependence structure makes the determination of the c.d.f.'s of cumulative rainfall and the maximum rainfall per event over a season analytically intractable. Therefore general expressions are derived for calculating the first two moments of the cumulative rainfall per season. Moreover the limiting behavior of these two c.d.f.'s are studied as follows.

A sequence of the same fixed season is considered from one year to another. Then the cumulative rainfall up to the i^{th} year within the same season is defined. A theorem is proved in regard to the rate of convergence of c.d.f. of the cumulative rainfall up to the i^{th} year; $i \geq 1$, within a season to its limiting c.d.f. Similarly the limiting c.d.f. of the maximum rainfall per event up to the i^{th} year $(i \uparrow \infty)$ within a fixed season is also given based on the results of Berman (1962). Since the exact c.d.f. of the maximum cumulative rainfall per event within a season is not obtainable in our case, therefore use of the limiting c.d.f. of the maximum cumulative rainfall per event in determination of the actual return period is indicated.

Specific theoretical applications of the general formulation of Chapter 3 are given in Chapter 4. Specifically the p.m.f. for the conditional probability of the number of cells given the wet duration of a rainfall event, is hypothesised to be a Poisson process in space. Then using a particular depth-area relationship for a cell and a two parameter gamma probability density function for the maximum cell depth, an analytical expression for the c.d.f. of the cumulative rainfall per rainfall event is given. In this derivation the wet interval is assumed to follow an exponential p.d.f. Then an approximate analytical expression is given for the expected cumulative spatial rainfall within a season. Extension to ungaged basins and multiple converging basins within the fixed gaged region is also given. Finally the limiting distribution of the maximum rainfall per rainfall event is derived using the derived expression for the c.d.f. of the cumulative rainfall per rainfall event.

A numerical demonstration of the specific results obtained in Chapter 4 is given in Chapter 5 on the Atterbury watershed near Tucson, Arizona. Summer rainfall is selected for the demonstrative purposes. Finally a summary of this study and recommendations for further research are given.

CHAPTER 2

LITERATURE REVIEW

The first attempts to treat hydrologic variables as statistical variables with an intention of estimating the frequency of occurrence of various magnitudes, were those dealing with floods (Horton, 1913; Fuller, 1914). The frequency analysis of precipitation, and in particular that of rainfall, was started around 1935 (Yarnell, 1935). Since then much effort has been expended in the realm of frequency analysis of rainfall and floods. All such frequency analyses have two major sources of error, (i) errors due to the assumed probability density function (p.d.f.) not conforming to the "true" p.d.f. underlying the population of floods and rainfall, and (ii) sampling errors due to non-representativeness of the record from which the numerical values of the parameters of the p.d.f. are estimated (Nash and Amorocho, 1966). Although both the above sources are of rather basic and great importance, one could argue that the uncertainty in the "true" form of frequency function or p.d.f. is of more fundamental importance than the sampling errors. To this effect, in regard to floods, Nash and Amorocho indicate that the magnitudes of floods corresponding to even very high return periods could be estimated with quite tolerable accuracy from relatively small samples, if one could be certain that the assumed form of the p.d.f. was correct.

This leads to a natural question as to how the true form of p.d.f. underlying any random variable can be determined? The answer lies in the approaches other than classical frequency analysis, that have been taken in the past. The remaining exposition discusses such approaches for the random variables defined on floods and rainfall. However, a classification of diverse approaches seems necessary to explain better the ideas behind them. Such is done prior to the review of specific models. Finally, note that since a p.d.f. may not always exist, therefore the following discussion generally considers cumulative distribution function (c.d.f.) of a random variable instead of a p.d.f.

2.1. A Classification of Modeling Approaches

All such models developed in the past, that attempt to determine the "true" form of the c.d.f. of some random variable of interest, are to some extent diverse in their respective approaches. Research has gone beyond the simple procedure of choosing some p.d.f. and fitting the data to it. For example, models of Todorovic (1968), Zelenhasic (1970), Eagleson (1972), Duckstein, Fogel and Kisiel (1972), etc., take a stochastic process orientation to arrive at the form of c.d.f.'s of variables like annual maximum flow, cumulative point rainfall within a season, etc. As a consequence of this it seems necessary to give a broad classification to these approaches, before attempting to review the models that are pertinent to this study. Our intent behind giving this classification is threefold,

(i) to justify the selection of models to be reviewed,

(ii) to give a unifying fabric to models that are purely statistical in comparison with the so-called "process oriented" models, and

(iii) to point out some of the issues involved in undertaking detailed modeling of a process to arrive at the "true" form of c.d.f., basically from a process oriented viewpoint.

Based on this, we classify models into three categories,

(i) Empirical or purely statistical models,

(ii) Phenomenological or process oriented stochastic models, and

(iii) Physically based stochastic models.

(i) In the first category, those models can be assigned, that take a purely statistical approach to estimate the "true" form of the c.d.f. of the random variable in question. Such a procedure has been summarized by Nash and Amorocho (1966) as, "In general such estimation involves choosing a particular form (double exponential, log-normal, etc.) for the frequency or probability distribution of the flood magnitude and using the available sample to estimate the numerical values of the parameters of this function." Such an approach is empirical because different forms of frequency functions are tried in this case, without any physical or phenomenological justification for doing so.

(ii) A phenomenological model to determine the "true" form of the c.d.f. of a random variable can be formulated in two ways; (1) to start from that particular process on which the random variable is defined, and then under a set of assumptions, determine the form of c.d.f. Such has been attempted in the past for the annual maximum peak flow, cumulative annual rainfall, etc., by Zelenhasic (1970) and Todorovic (1968) respectively. (2) To start from a finite number of processes that sequentially give rise to the process on which the random variables of interest are defined. For example in dealing with some random variable defined on streamflows, one could start with the process of global atmospheric circulation of moisture, heat, etc. which in turn gives rise to a process of local atmospheric disturbance, which in turn gives rise to the process of rainfall and/or snow, which in conjunction with the process of catchment dynamics leads to the process of streamflows. In this context, an ideal model would start from the process of global atmospheric disturbance to finally arrive at the c.d.f. of a random variable defined on streamflows. Alternatively, the starting phase can be one of the intermediate processes, e.g., starting from the process of rainfall and/or snow in conjunction with the transformation processes of rainfall to runoff (Kisiel, 1967) to finally derive the form of c.d.f. of a random variable on streamflows. Some attempts along these lines are the work of Woolhiser and Todorovic (1971) and Duckstein, et al. (1972).

(iii) The physically based stochastic models can be considered an extension of phenomenological models in the sense that the physical theories of fluid mechanics and thermodynamics, as the case may be, are used to take into account the behavior of physical variables. For example, instead of using a lumped parameter unit hydrograph with the stochastic rainfall process, if one uses the continuity and momentum equations for routing through a catchment, to arrive at the streamflow process, then such a model would be called physical. LeCam (1961) indicates of such a consideration leading to a so-called "true" model although he actually does not undertake such an approach. Eagleson's model (1972) can be placed in this category, since he uses a kinematic wave model in conjunction with stochastic rainfall inputs to finally arrive at the p.d.f. of maximum flow.

Having indicated a criterion for the classification of various approaches, a brief comparison is now given between these approaches. Firstly, to decide on the "true" form of a c.d.f. based on an empirical approach seems rather difficult, and moreover such a selection can sometimes be misleading. For example, if one could argue that a p.d.f. does not exist for some random variable under consideration, then any one of the selections of p.d.f.'s is as arbitrary as the other. In the case of maximum instantaneous flow within an arbitrary but fixed time interval, Zelenhasic (1970), based on a phenomenological approach indicates that the desired c.d.f. of maximum flow is not differentiable at the origin. The 'atom' at the origin is due to the fact that there is a non-zero probability (howsoever small) of having no exceedance above a pre-selected datum in every finite time interval. Another shortcoming in a purely empirical approach lies in the method of selection of one p.d.f. over the other. The most commonly used methods are the statistical goodness of fit tests, e.g., chi-square or Kolmogorov-Smirnoff, etc. Based on these tests, one can accept hypothesis regarding goodness of fit of a set of data by more than one type of p.d.f.'s. Therefore the selection of a p.d.f. based on a purely statistical testing can never guarantee it to be the "true" p.d.f.

A phenomenological approach uses the mathematical tools from the theory of stochastic processes to arrive at the form of c.d.f.'s of random variables. In other words, a particular hydrologic process is considered to be a stochastic process, and then the form of c.d.f.'s of random variables defined on this process are derived under a set of assumptions, which are to a large extent motivated by phenomenological considerations (for example, see Todorovic, 1968; Zelenhasic, 1970). However, since a stochastic process is defined in terms of a family or a sequence of random variables, one has to first determine the form of c.d.f.'s of the random variables which define a stochastic process, before attempting to derive the form of c.d.f.'s of random variables of interest defined on a stochastic process. Determination of the forms of c.d.f.'s of random variables that define a stochastic process leads to empiricism in a phenomenological approach. For example, consider the stochastic process defined in terms of a sequence of instantaneous flow exceedances within an arbitrary time interval (Zelenhasic, 1970). Using the phenomenological fact that the number of such exceedances is a random variable in any fixed finite time interval, Zelenhasic determines the c.d.f. of the maximum exceedance within this time interval. Empiricism in this approach lies in the fact of empirically fitting a c.d.f. to each exceedance, as done by Zelenhasic.

The above discussion on a phenomenological approach leads to the consideration of more than one stochastic process, as mentioned earlier. Such a consideration in the context of Zelenhasic's model would mean the determination of the form of c.d.f. of each exceedance. This can be determined by simultaneously considering the stochastic processes of rainfall over a river basin and that of the basin response. This in turn would lead to a phenomenological approach that would consist of three stochastic processes for determination of the form of c.d.f. of maximum instantaneous flow within an arbitrary but fixed time interval.

Besides determination of the form of c.d.f.'s of random variables which define a stochastic process, another important aspect of phenomenological modeling are the assumptions that go into it. Generally speaking, assumptions are motivated by two considerations, i.e., the extent to which such assumptions represent a phenomenon in the real world and an extent to which ease in mathematical manipulation is achieved from these assumptions. In this respect the results obtained from a phenomenological model are "true" only up to the extent to which the underlying assumptions are valid in the real world situation.

The contemporary literature in hydrology does not contain many physically based stochastic models. Presently the emphasis in research seems to be on building phenomenological models. However, a verification of the assumptions in the phenomenological models may only be attainable through physical principles. Most of the work on physical modeling falls in the category of deterministic models, which can be considered to be a very particular kind of stochastic models in the sense that they assign a probability one to every outcome. In this context, note that determinisim is frequently misunderstood. When considering a physical model, e.g., continuity equation, etc., one does have a deterministic relationship in mind. However if the inputs to a system are random, then even when continuity principle is applied to these inputs, the outputs will be stochastic. A classical example in hydrology is the area of rainfall-runoff models, which shall not be reviewed here.

A preliminary attempt along the lines of physically based stochastic modeling is that of Gupta (1972). This approach is not yet developed completely, but it attempts to numerically estimate the c.d.f. of a random variable defined on the streamflow process, by using stochastic rainfall process and a deterministic rainfall to runoff transformation. An analytical approach along physically based stochastic modeling is that of Eagleson (1972). Eagleson uses kinematic wave in conjunction with stochastic rainfall inputs to determine the c.d.f. of instantaneous flows. Since an analytical approach in this area can get mathematically very cumbersome, Eagleson makes expedient assumptions to arrive at some kind of analytical solutions as a first step.

Based on the above discussion, it seems that as detailed phenomenological or physical modeling of some process is attempted, the data requirements to validate such models also increase besides the increase in complexity of analysis. Thus far only a general idea behind stochastic modeling of hydrologic processes has been given. In the ensuing text, some specific models are reviewed.

2.2. Review of Specific Stochastic Models

In this section, the primary emphasis is on review of phenomenological models of the rainfall process. However, the process of excess water yield as derived from rainfall is also reviewed briefly. In particular the process of cumulative rainfall within an arbitrary but fixed time interval is considered. Moreover, a few of those models that consider other aspects of a rainfall process are also reviewed briefly, e.g., simulation of rainfall fields over a catchment.

To begin with, we consider the rainfall process as the process of our ultimate interest. After having reviewed the relevant models dealing with the rainfall, the process of excess water yield as derived from rainfall is considered next. The rainfall process has been analyzed in diverse ways in the literature, and different aspects of such have been considered in the past. Such approaches can be summarized as below.

(1) Modeling of rainfall process (after it reaches the ground) as a stochastic process. Two aspects on this have been considered, namely (a) the above as a point

process evolving in time and (b) as a spatial process evolving in time.

(2) Modeling of rainfall process from the local meteorological considerations of the atmosphere.

In regard to the rainfall simulation from the meteorological considerations Amorocho and Morgan (1971) give a model for the simulation of storm fields at the ground level. Their simulation is based on the convective storm model of Weinstein and Davis (1968). The inputs to this model are a number of meteorological parameters derived partly from atmospheric soundings and partly from local ground rainfall data. This model gives estimates of <u>total</u> storm rainfall, duration and area coverage, but does not permit a running computation of the temporal change in the precipitation field. However since this aspect of simulating rainfall from atmospheric conditions is not pertinent to our study, we shall not go into any further details of this model.

2.2.1. Stochastic Modeling of the Rainfall Process

<u>Point Rainfall Process</u>. In regard to the approaches on the stochastic modeling of the rainfall process, we first consider the work done on point rainfall process. Once again two aspects on the point rainfall process have been considered in the past, (i) to obtain the form of c.d.f.'s of random variables defined on point rainfall process and

(ii) simulation of synthetic traces of rainfall based on the emperical p.d.f.'s fitted to different random variables on this process.

Todorovic and Yevjevich (1969) give a comprehensive theoretical analysis of a point rainfall process. They consider $\xi(s) \geq 0$ as the rainfall intensity at some instance s, at a fixed point in space, within an arbitrary time interval $(t_0, t]$; $t_0 < t < \infty$. A storm event is defined as continuous precipitation between two non-rainy intervals, even though the total amount of precipitation and duration of some storms may be very small. Based on the phenomenologically motivated assumption regarding intermittency of the $\xi(s)$ process in time, i.e., $P\{\xi(s) = 0\} \neq 0$ (P is the probability measure), a sequence of storm events (as defined above) is obtained within $(t_0, t]$. Two important concepts should be noted here, (a) idea of an event (also explored by other researches, and is indicated later in this text,), and (b) a random number of such events within any arbitrary but fixed interval $(t_0, t]$. The cumulative process X(t) is given by the random integral,

$$X(t) = \int_{t_0}^{t} \xi(s)ds. \tag{2.1}$$

The following stochastic processes, defined as function of $\xi(s)$, are considered, (1) N(t), the number of complete storm events in time interval $(t_0, t]$, (2) $N(x)$, the maximum number of storm events after t_0, such that the total precipitation does not exceed the amount $x - x_0$; x_0 is the total amount up to time t_0; (3) $\{\tau_j; j = 1, 2, \ldots\}$, the terminal times of storm events which is a random sequence of points on the time scale, (4) X_j, the total precipitation for j storm events; (5) $Z_j = X_j - X_{j-1}$, the total precipitation during the j^{th} storm event, and (6) $\{X(t), t > t_0\}$ as given by Equation (2.1) as a stepwise non-decreasing cumulative function of $\xi(s)$.

Different cases are considered in regard to the derivation of the probability mass function (p.m.f.) of $N(t)$ based on various assumptions (note that $N(t)$ is a counting random variable). The derivation of the c.d.f. for $X(t)$ is given, using independence between $N(t)$ and $N(x)$, and assuming both as Poisson processes with intensity (rate) parameters λ_1 and λ_2 respectively. The p.d.f. of $x(t)$ for all $t > t_0$ is given by,

$$f_t(x) = \delta(x)e^{-\lambda_1 t} + \lambda_2\sqrt{\lambda_1 t/\lambda_2 x}\; e^{-(\lambda_1 t+\lambda_2 x)} I_1(2\sqrt{\lambda_1\lambda_2 tx}), \quad (2.2)$$

where $\delta(x) = 1$ for $x = 0$ and 0 otherwise. $I_1(2\sqrt{\lambda_1\lambda_2 tx})$ is the modified Bessel function of the first order.

Another aspect of the rainfall process considered in the past is related to the sequence of random variables

$\{\tau_j; j = 1,2,\ldots\}$ as indicated earlier. A new sequence $\{(\tau_{j-1},\tau_j); j = 1,2,\ldots\}$ denoting the time between the end of two consecutive storms can then be formed. Denote $T_j = (\tau_{j-1},\tau_j)$ for all $j = 1, 2, \ldots,$. T_j for all $j \geq 1$ can be considered as the sum of two random intervals, i.e., D_j denoting the dry period following the $(j-1)^{st}$ storm and C_j denoting the duration of the j^{th} storm. Hence $T_j = D_j + C_j$, for all $j = 1, 2, \ldots,$. Grace and Eagleson (1966) reviewed earlier models dealing with the aspect of so-called occurrence and non-occurrence of rainfall, in the light of the sequences $\{D_j; j = 1,2,\ldots\}$ and $\{C_j; j = 1,2,\ldots\}$ of random intervals. However these sequences are not considered for the storm as defined by Todorovic above, but for the daily rainfall occurrences. Therefore discretizing the time in this way makes the sequences of random intervals $\{D_j\}$ and $\{C_j\}$ positive integer valued. Crovelli (1971) mentions the work of several researchers who found that sequence in daily rainfall occurrence can be described by discrete Markov chains. Such include Gabriel and Neumann (1962), Caskey (1963), Weiss (1964), Green (1964), Feyerherm and Bark (1965), Wiser (1965) and Pattison (1965). Green (1964) looks at the sequences from the viewpoint of an alternating renewal process (as described by Cox, 1962) and derives more satisfactory results for the Tel-Aviv data, as compared to Gabriel and Neumann (1962). However he does not suggest the applicability of this model for every part of the world, since D_j and C_j may not always be independent and identically distributed (i.i.d.) as negative exponential for all $j \geq 1$, as assumed by Green in his model.

Another aspect regarding point rainfall process has been to look at the depth-duration relation. In our context this means the consideration of the joint p.d.f. of Z_j and C_j; $j = 1, 2, \ldots,$ denoting respectively the total rainfall and the storm duration for the j^{th} storm as given earlier. Grayman and Eagleson (1969) used empirical consideration to propose a two parameter gamma density to describe the conditional density of depth given duration. They use this empirical p.d.f. in conjunction with the p.d.f.'s of $\{D_j\}$ and $\{C_j\}$ to generate synthetic sequences of rainfall.

Crovelli (1971) proposed a common bivariate gamma p.d.f. to model Z_j and C_j for all $j \geq 1$. The density function $f(z,c)$ is given as

$$\left.\begin{aligned} f(z,c) &= \alpha\beta e^{-\beta c}(1 - e^{-\alpha z}) \quad \text{for} \quad 0 \leq \alpha z \leq \beta c \\ &= \alpha\beta e^{-\alpha z}(1 - e^{-\beta c}) \quad \text{for} \quad 0 \leq \beta c \leq \alpha z \end{aligned}\right\}. \tag{2.3}$$

Crovelli gives different statistical properties associated with the above p.d.f. including statistical estimators to estimate the parameters α and β that arise above, and finally derives the expected cumulative precipitation within a time interval $(0,t)$ using the theory of finite continuous time Markov chains (Parzen, 1967).

The work done in the past in regard to simulation of synthetic point rainfall sequences, based on the random intervals D_j, C_j, and the random variable Z_j, $j \geq 1$, is not reviewed here, since no simulation type modeling is being undertaken in this study. However, a few models dealing with such simulation are those of Pattison (1965), Grace and Eagleson (1966), Grayman and Eagleson (1969) and Sariahmed and Kisiel (1969).

Process of Spatial Rainfall. In regard to the spatial rainfall process, two aspects similar to that of the point rainfall process have been considered in the past, namely, simulation of spatial rainfall (Sorman and Wallace, 1972) and analytical derivations of the form of c.d.f.'s of random variables defined on the process of spatial rainfall. The review in the ensuing text primarily considers the determination of the form of c.d.f.'s of random variables. However, some phenomenological aspects of the process of spatial rainfall are also reviewed, which are later used in Chapter 3 on theoretical considerations.

The possibility of a stochastic analysis of spatial rainfall is indicated by LeCam (1961). He identifies the shower cell as the basic element of an areal model. The cells occur in clusters that may conceivably correspond to a front. A bunch of cell clusters may correspond to what is called a storm. Note that this definition of a storm in an areal context differs sharply from that defined earlier for a point process. However the work of LeCam is too general to be put to specific modeling efforts. Further, since 1961, more has been learned in regard to spatial properties of storms, e.g., see Grayman and Eagleson (1971). The details of the salient features of spatial rainfall process as indicated by Grayman and Eagleson will be given later in this text.

Duckstein, et al. (1972) consider spatial rainfall in connection with prediction of cumulative rainfall for summer type storms on small watersheds in semi-arid regions. They consider the spatial rainfall process evolving in time as an intermittent process. However their definition of a rainfall event is operational rather than phenomenological, as given by LeCam. One of the two possible criteria for defining a rainfall event, is to consider n gages, with total rainfall $R_1, \ldots, R_n$, for a given day as specified by the U. S. National Weather Service. Now, an event is said to occur at any of the gages, if the mean precipitation $(\sum_{i=1}^{n} R_i)/n$ is greater than .5 inches and one gage records more than 1 inch. Other definitions of a rainfall event are of course possible. Having done this, the c.d.f. of the cumulative rainfall within a season is then derived based on the assumption of a Poisson p.m.f. for the random number of these events, and a negative binomial p.m.f. for the mean precipitation per event. The resulting process is a compound Poisson process, since it is assumed that the counting process and the random variable denoting average rainfall per event are mutually independent. Clearly the basic approach given above is the same as that of Todorovic and Yevjevich (1969), on the point rainfall process.

Another aspect of spatial rainfall is the so-called depth-area relationships. Such studies attempt to express the areal distribution of total rainfall from a "storm' under consideration, as a function of the point of maximum rainfall. Court (1961) gives a comprehensive review of the earlier work. Fogel and Duckstein (1969) also propose such a relationship for the convective thunderstorm cells for semi-arid summer rainfall. It seems that such relationship should not be done exclusive of the phenomenological basis behind a "storm type" under consideration (as given by meteorologists, namely convective, frontal, squall line, etc.). However these relationships as developed only for a thunderstorm cell in the semi-arid regions, can be called phenomenological. This conceptual basis is used later in Chapter 4 to derive the c.d.f. of the cumulative spatial rainfall from a cell.

Another aspect related to spatial rainfall considered in the past, are the spatial correlation structures of the rainfall depths in space. This aspect is not reviewed presently, because these studies are not phenomenologically oriented in the sense that such correlations are given for total observed rainfall amounts without considering the process of spatial rainfall as such. However, such works dealing with the spatial correlations center around trying different functional relations for correlations with an intent to justify one form over the other (Rodriguez-Iturbe, Vermarcke and Schaake, 1972). Such studies have been used as an aid in the design of data collection networks.

In regard to the simulation of spatial rainfall, Grayman and Eagleson (1971) attempt to simulate spatial rainfall using certain observed phenomenological features of different storm types (meteorologically speaking). Although simulation is not pertinent to the present effort, the phenemonenological features of storms as used by them are relevent. A comprehensive summary of this is given below.

Meteorological events can be classified by the scale or level of the event. Climatic scale is the largest. Next in size is the synoptic or marco-scale, for example a cyclone or a storm front. The next level is the mesoscale which generally has an area ranging from 25 to 5,000 square miles. The next and the smallest scale is cellular or micro-scale event, and is about 3 to 30 square miles in size. Grayman and Eagleson (1971) also review earlier works which noted the presence of the cellular activity within diverse storm types. They quote House (1969) for identifying two distinct levels of meso-scale activity, based on analysis of nine diverse storm types in the New England area. House calls these two levels as large meso-scale areas (LMSA) and small meso-scale areas (SMSA). A schematic representation of a typical storm type is given in Figure 2.1. In this figure four distinct levels of activities are indicated. They further summarize other statistical characteristics of levels of a storm type, e.g., histogram of cell durations, relationship between cell duration and cell intensity, etc. Finally, they indicate the random nature of a number of LMSA that develop within a synoptic level, and number of cells that develop within a LMSA, etc. However, it appears that a synthesis of the stochastic behavior of these storm properties is necessary.

Although in the above analysis, the convective summer type thunderstorm cells for semi-arid regions are not included, other works (Fogel and Duckstein, 1969) do indicate such cellular activity for summer type rainfall. However the different levels of activity for the air-mass type summer rainfall have not been identified to the best of our information.

Interestingly enough, part of the above features were speculated by LeCam (1961) in his attempt to give phenomenological reasons for modeling of spatial rainfall process. However, he only considered two levels of activity and described it as a very general two-stage clustering

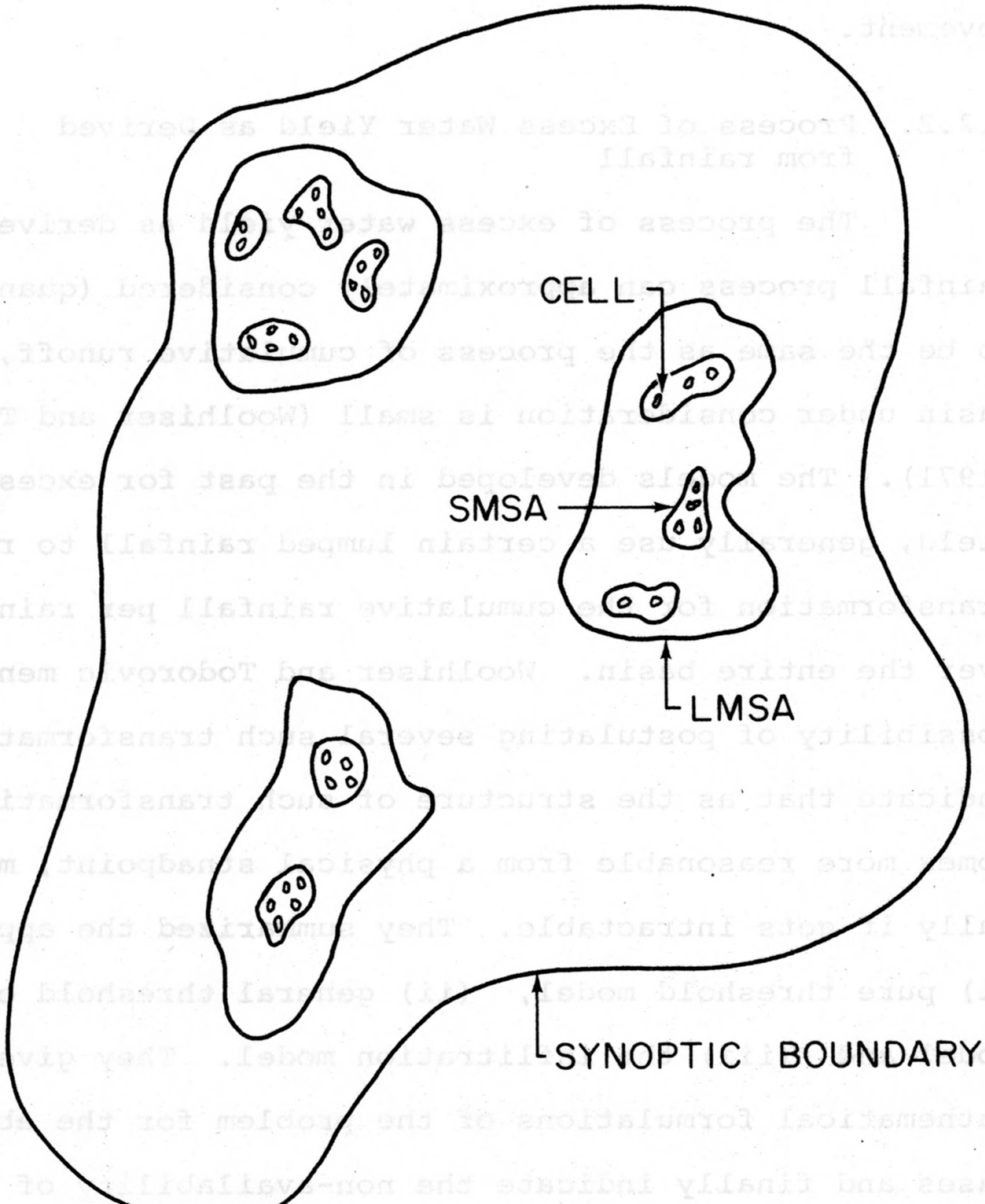


Figure 2.1. Schematic of Typical Storm.

process, having assumed that the cells do not undergo any movement.

2.2.2. Process of Excess Water Yield as Derived from rainfall

The process of excess water yield as derived from the rainfall process can approximately considered (quantitatively) to be the same as the process of cumulative runoff, if the basin under consideration is small (Woolhiser and Todorovic, (1971). The models developed in the past for excess water yield, generally use a certain lumped rainfall to runoff transformation for the cumulative rainfall per rainfall event over the entire basin. Woolhiser and Todorovic mention the possibility of postulating several such transformations, but indicate that as the structure of such transformations becomes more reasonable from a physical stnadpoint, mathematically it gets intractable. They summarized the approaches as, (i) pure threshold model, (ii) general threshold or storage model and (iii) the inflitration model. They give brief mathematical formulations of the problem for the above three cases and finally indicate the non-availability of a solution to second and third approaches given above. Finally, they hypothesis a chance mechanism to describe the runoff counting process from the point rainfall counting process.

Duckstein, et al. (1972) also use a pure threshold model to derive the c.d.f. of excess water yield. The pure threshold is simply given as follows. Let Z_j be a random variable denoting the excess water yield, corresponding to the cumulative rainfall per event Z_j, $j \geq 1$. Let ω^* be some constant, that depends on the physical characteristics of a watershed. Then based on the pure threshold model,

$$\left.\begin{aligned} Z_j' &= Z_i - \omega^* \quad \text{if} \quad Z_j > \omega^*, \quad j = 1, 2, \ldots \\ &= 0 \qquad\qquad \text{otherwise} \end{aligned}\right\} . \tag{2.4}$$

Another transformation given by Duckstein, et al. (1972) is the Soil Conservation Service formula. Clearly, once the c.d.f. of rainfall per event, denoted by Z_j is determined for all $j \geq 1$, the c.d.f. of the corresponding excess water yield can easily be determined based on some transformation of the form given by Equation (2.4). The c.d.f. of cumulative water yield within an arbitrary time interval is then determined, based on the sum of a random number of random variables, where each random variable denotes the excess water yield per event. However the derivation is given on the assumption of independence between the random variable denoting the number of such events and that denoting excess water yield per event.

In summary, the cumulative water yield given by the above approaches represents the free water accumulation rather than the surface runoff. As mentioned in the beginning, for small watersheds these two are nearly identical, but for large watersheds the attenuation of rainfall excess by watershed hydraulics must be considered.

2.3. Concluding Remarks

In the above review, an attempt is made to project the notion of phenomenological modeling of stochastic hydrologic processes, in particular that of the rainfall process. Table 2.1 gives a brief summary of the phenomenological models of the rainfall process that have been mentioned in the review.

Table 2.1. A Summary of Phenomenological Models of the Rainfall Process.

Type of Approach	Models of the Point Rainfall Process	Models of the Spatial Rainfall Process
Analytical	Todorovic (1968), Todorovic and Yevjevich (1969), Verschuren (1968), Gabriel and Neumann (1962), Green (1964) Crovelli (1971)	Fogel and Duckstein (1969), Duckstein, et al. (1972), Gupta (1972),
Simulation	Pattison (1965), Grace and Eagleson (1966), Grayman and Eagleson (1969), Sariahmed and Kisiel (1969)	Grayman and Eagleson (1971), Sorman and Wallace (1972), Amorocho and Morgan (1971)

The phenomenological approach to analytical modeling of the rainfall process is essentially 'event' based. However no universal definition of a rainfall event seems possible. An event is generally defined on the basis of the type of process available, the form of data and finally the mathematical approach taken to analyse the process.

An event based phenomenological approach to the modeling of the rainfall process has been along the lines of a random number of random variables. This approach has two aspects, (i) the nature of stochastic process to determine the p.m.f. of number of such events within a given time interval and (ii) the c.d.f. of random variables defined on an event, e.g., cumulative rainfall per event, cumulative excess water yield, etc.

In regard to the assumptions made in the past in such models, it is invariably assumed that the random variable denoting the number of events is independent of the random variables characterising magnitudes of the sequence of events. This assumption may be 'reasonable' under very restrictive conditions for the process of cumulative water yield and cumulative rainfall. For example in the case of the point rainfall process, where the rainfall bursts are of relatively very short duration in comparison with the duration between successive bursts, e.g., summer rainfall in arid lands, the assumption of independence seems reasonable. However for winter rainfall in some areas (Kao, Duckstein and Fogel, 1971), the above independence assumption is not reasonable. In such cases the durations of rainfall bursts give rise to a stochastic dependence between the random variables denoting the number of events and their magnitudes. Moreover if one is considering the rainfall process in space, then the assumption of independence does not seem at all reasonable, because of relatively long durations over which rainfall may occur over a region from a rainfall event.

Most of the modeling efforts with an aim to study the rainfall process analytically, have primarily dealt with the point rainfall process. The c.d.f.'s of rainfall magnitudes per rainfall event have also been assigned empirically based on data. In context of a spatial rainfall process, a need exists to consider the derivations of c.d.f.'s of rainfall per event using phenomenological characteristics of spatial rainfall. Such an attempt may ultimately lead to providing answers to the following, (i) Prediction of cumulative rainfall and water yield over a single or multiple basins within a gaged region, (ii) Extension to ungaged basins within this region, (iii) Impact of such considerations on the data requirements, both qualitative and quantitative.

In the next chapter on theoretical considerations, some of the above aspects are considered for the spatial rainfall process.

CHAPTER 3

THEORETICAL CONSIDERATIONS

3.1. Introduction

The rainfall process is the basic process for study in this exposition. The process of snow accumulation and snow melt are not considered. The development focuses on obtaining the analytical form of cumulative distribution functions (c.d.f.'s) of random variables under consideration.

The general approach gives a model to obtain the c.d.f.'s of cumulative rainfall and excess water yield from a rainfall event, and the c.d.f. of occurrence of a random number of such events within a fixed season. Based on the nature of a simple transformation that is used to convert the cumulative rainfall per event into excess water yield, the present development (in regard to cumulative surface runoff) can only apply over basins that are, say, roughly up to 200 mi^2 in area. This is because for basins of this size, the excess water yield is roughly the cumulative surface runoff, as mentioned in Chapter 2.

Four random variables are of interest in this study, namely, the cumulative rainfall and the excess water yield within a fixed season, and maximum cumulative rainfall and the excess water yield per event within a season. The c.d.f.'s of these random variables within a season (finite time interval), which represent our goal, cannot be obtained in analytical form in light of the stochastic dependence between the cumulative rainfall or excess water yield per event and the number of occurrence of such events within a fixed season. The specific nature of this dependence is indicated in the general approach, in which the following properties of these random variables are considered.

(1) The first two moments of the cumulative rainfall within a season;

(2) the rate of convergence of the c.d.f. of the cumulative rainfall within a fixed season up to j years; $j \geq 1$, to its limiting c.d.f.;

(3) the limiting c.d.f. of the maximum cumulative rainfall per event within a fixed season up to j years ($j \uparrow \infty$), which is given based on the results of Berman (1962).

The above derived results for the cumulative rainfall are equally applicable to the excess water yield.

Extension of each of the above results to multiple seasons is given under the assumption of independence of cumulative rainfall from one season to another.

3.2. General Approach

Let $(0,t_1]$, $(t_1,t_2]$, ..., etc. be a sequence of time intervals. For example, each time interval may typically represent a 'season'. Each time interval is assumed to be fixed and within each interval it is assumed that 'similar' kind of rainfall activity occurs. For example, the summer season in semi-arid lands may correspond to say $(0,t_1]$, within which the air-mass type of convective rainfall occurs. From one interval to another the rainfall activities can differ in regard to their statistical properties (discussed later in this chapter). The subsequent development is given for any one season say $(0,t_1]$. However, based on the assumption that the same general approach is applicable to other seasons as well, extension to multiple seasons would be indicated.

Let B denote the river basin under consideration and $R \supset B$ be some gaged region that contains the basin. For the present we don't discuss the nature of the region R, because it can vary with the type of rainfall activity under consideration and the extent of data availability. Moreover the concept of a region, which is important for extensions to ungaged basins, is elaborated on in Chapter 4. For the purposes of theoretical development, we treat R and in turn B as subsets of the two-dimensional Euclidean space.

For present development, we do not need to consider variations in altitude, but effect of such is indicated in Chapter 4.

The entire development is based on the assumption of a probability space (Ω, F, P), where Ω is sample space, F is the σ-algebra of the subsets of the sample space and P is a probability measure defined on it (for the details refer to Breiman, 1968). Further it is assumed that the stochastic processes that are analysed subsequently, are defined on the above probability space.

Let $\xi(u,v,s) \geq 0$, denote the rainfall intensity at some point $(u,v) \in R$ and at some instant $s \in (0,t_1]$. Since for every $(u,v) \in R$ (in particular $(u,v) \in B$) and every $s \in (0,t_1]$, $\xi(u,v,s)$ is a random variable, the family of random variables,

$$\{\xi(u,v,s) \geq 0; (u,v) \in R, s \in (0,t_1]\} \tag{3.1}$$

represents a random field. Assume that for every $(u,v) \in R$ and $s \in (0,t_1]$, $\xi(u,v,s)$ is integrable. The cumulative rainfall up to time t_1 from a river basin $B \subset R$, denoted by $X(t_1)$, $t_1 > 0$, is now given by the following integral,

$$X(t_1) = \int_0^{t_1} \int_B \int \xi(u,v,s)\,du\,dv\,ds. \tag{3.2}$$

$X(t_1)$, for all $t_1 > 0$, is a random variable and therefore the family of random variables $\{X(t_1);t_1 > 0\}$ represents a continuous parameter stochastic process. Since $\xi(u,v,s) \geq 0$, for all $(u,v) \varepsilon B$ and $s \varepsilon (0,t_1]$, $X(t_1)$, for $t_1 > 0$ has monotonically non-decreasing sample functions, i.e., for each $t' \varepsilon (0,t_1]$, and $\Delta t' > 0$, $X(t') \leq X(t'+\Delta t')$. The stochastic process $\{X(t_1);t_1 > 0\}$ forms the basic process of study in the ensuing text. Figure 3.1 represents a typical sample function of the process $\{X(t_1);t_1 > 0\}$.

It is assumed that $P\{\xi(u,v,s) = 0$, for all $(u,v) \varepsilon B$ and $s \varepsilon (0,t_1]\} \neq 0$. This assumption gives rise to the so-called intermittency in the rainfall process. Based on this intermittent nature of rainfall process, two sequences of 'dry' and 'wet' random intervals can be defined, such that no rainfall occurs anywhere over the region during 'dry' interval and rainfall occurs somehwere over the region during 'wet' interval. In order to avoid introducing new notation, we adopt the same notations as that for point rainfall process, introduced in the second chapter. Let $\{D_j;j\geq 1\}$ and $\{C_j;j\geq 1\}$ denote two sequences of random intervals, such that $D_j = \{s;\xi(u,v,s) = 0$, for all $(u,v) \varepsilon R\}$; $j \geq 1$, and $C_j = \{s;\xi(u,v,s) > 0$, for some $(u,v) \varepsilon R\}$; $j \geq 1$. Note that C_j and D_j for $j \geq 1$, represent random intervals within the fixed season $(0,t_1]$. Clearly, the above definition of the sequence $\{C_j\}$ does

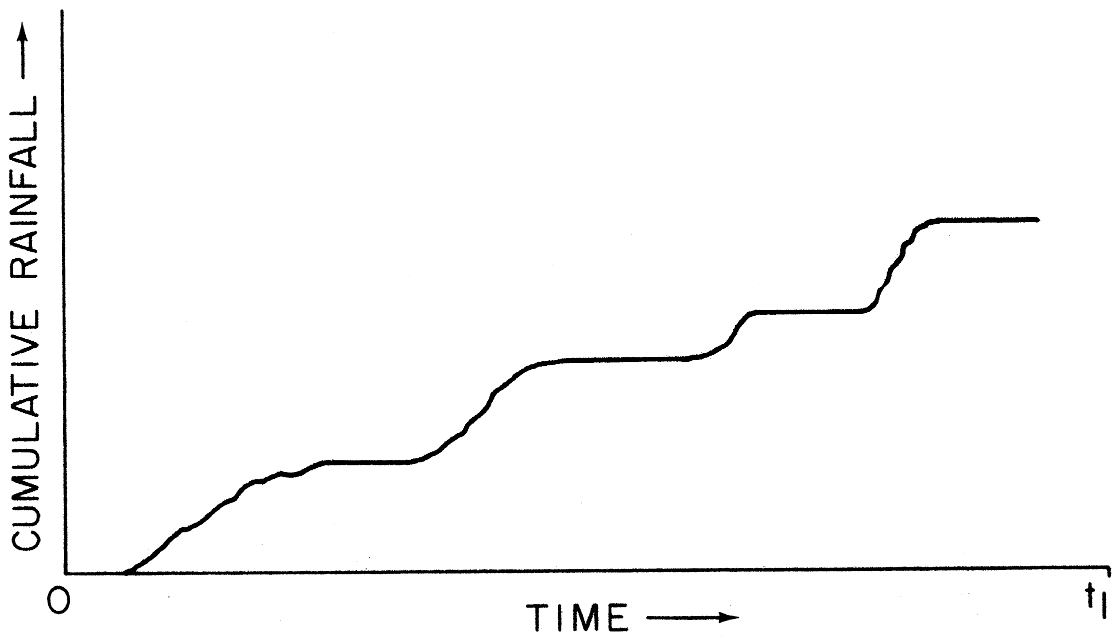


Figure 3.1. A Typical Sample Function of the Stochastic Process of the Cumulative Rainfall.

not require that $\xi(u,v,s) > 0$ for every $(u,v) \varepsilon R$, within the random intervals C_j, $j \geq 1$. Phenomenologically such a formulation is very realistic, because of the fact that spatial intermittency generally occurs in spatial rainfall.

The mathematical formulation of a random interval C_j; $j \geq 1$, can phenomenologically be interpreted as, the time interval from the commencement of rainfall anywhere over the region, up to the time of termination of rainfall everywhere over the region. The values assumed by the random wet intervals C_j's; $j \geq 1$, can be determined from a few recording rain gages covering the region. Moreover, if the radar observations are also available, then these may be used in conjunction with the rain gages to determine the values assumed by the random wet intervals. In our subsequent discussion, we refer to C_j's; $j \geq 1$, as <u>durations</u> of rainfall events. However, the delineation of C_j; $j \geq 1$, given above is more suitable for moving storm fronts. In the case of air-mass type of convective rainfall, subjective judgment may be needed for a criterion to determine the commencement and termination of rainfall over a region, that may constitute the duration of a rainfall event. This is because, the rainfall occurs locally in a random manner in space and time from thunderstorm cells (Fogel and Duckstein, 1969). In our context, some of these cells may be needed to be grouped as part of the same rainfall event, in case the cells occur adjacent to one another in time over the region. One such subjective criterion is demonstrated in the numerical case study given in Chapter 5.

The sequences $\{C_j\}$ and $\{D_j\}$ can be used to determine the probability mass function (p.m.f.) of number of events within a season $(0,t_1]$. Such is developed in the ensuing section.

3.2.1. Counting Process of Number of Rainfall Events

Firstly, the development is given for a finite time interval $(0,t_1]$. Then the limiting behavior is given by taking a sequence of years and considering the counting process within the fixed season $(0,t_1]$, from one year to the next. This limiting property of the counting process is used later in determining the limiting c.d.f.'s of the maximum cumulative rainfall and the excess water yield per rainfall event.

Assume that the random intervals in the sequence $\{C_j\}$ are mutually independent and identically distributed (i.i.d.) as C_1 and similarly in the sequence $\{D_j\}$ are i.i.d. as D_1. Moreover C_j and D_j; $j \geq 1$, are mutually independent. These assumptions can be justified heuristically as follows. Firstly, the sequences $\{D_j\}$ and $\{C_j\}$ are defined for the spatial rainfall within a fixed season $(0,t_1]$. It is assumed at the beginning, that the similar kind of rainfall activity takes place from the rainfall storms within the season, i.e., the air-mass type of convective rainfall in the summer season in the south-western United States. Therefore the random intervals C_j and D_j; $j \geq 1$, being identically distributed as C_1 and D_1 respectively within a fixed season seems reasonable. However, no phenomenological reason either for or against the assumption of independence seems possible at present. Therefore we assume independence to induce mathematical simplification in our treatment. However, note that such assumptions in regard to independence have also been made in the past for the point rainfall process (Grace and Eagleson, 1966; Todorovic, 1968).

Let $T_j = C_j + D_j$; $j \geq 1$. Then the random intervals T_j; $j \geq 1$, are mutually independent and identically distributed as T_1, since C_j are i.i.d as C_1 and D_j are i.i.d as D_1; $j \geq 1$. Assume that $E[T_1] < \infty$. The $\{T_j\}$ forms a renewal process, and let $N(t_1)$, $t_1 > 0$, denote the number of renewals (complete rainfall events) within $(0,t_1]$. $N(t_1)$ is defined as (Parzen, 1967, p. 133),

$$N(t_1) = \sup\{n; \sum_{j=1}^{n} T_j < t_1\}. \tag{3.3}$$

Now it follows from Equation (3.3), that the following random events are identical (Parzen, 1967, p. 133),

$$\{N(t_1) = n\} = \{\sum_{j=1}^{n} T_j < t_1 < \sum_{j=1}^{n+1} T_j\}. \qquad (3.4)$$

From Equation (3.4), the p.m.f. of the counting random variable $N(t_1)$ can be written as,

$$P\{N(t_1) = n\} = P\{\sum_{j=1}^{n} T_j < t_1\} - P\{\sum_{j=1}^{n+1} T_j < t_1\}. \qquad (3.5)$$

In other words, the probability distribution of the random variable $N(t_1)$ is completely determined by the c.d.f. of the random variables T_j; $j \geq 1$.

Let $m(t_1)$ denote the expectation of the random variable $N(t_1)$, then from Equation (3.5),

$$m(t_1) = E[N(t_1)] = \sum_{n=1}^{\infty} n(P\{\sum_{j=1}^{n} T_j < t_1\} - p\{\sum_{j=1}^{n+1} T_j < t_1\})$$

$$= \sum_{n=1}^{\infty} P\{\sum_{j=1}^{n} T_j < t_1\}. \qquad (3.6)$$

The expectation of a renewal counting process, also called its mean value function, completely determines its probability law (Parzen, 1967, p. 179). Therefore the second moment of $N(t_1)$, $m_2(t_1) = E[N^2(t_1)]$, is given by

$$m_2(t_1) = m(t_1) + \int_0^{t_1} m(t_1 - s)\, dm(s). \qquad (3.7)$$

Limiting behavior of the counting random variable $N(t_1)$ is given as follows. Let $\{N(t_1,j); j = 1,2,...\}$ be a sequence of counting random variables such that $N(t_1,j)$ denotes the number of rainfall events within the season $(0,t_1]$, during the j^{th} year, $j \geq 1$. Further assume that $\{N(t_1,j)\}$; $j \geq 1$, is an i.i.d. sequence. This assumption is reasonable, if one assumes that meteorology of the atmosphere from one year to another can be considered to be the same. Define partial sums $N_i(t_1)$ as $N_i(t_1) = N(t_1,1) + ... + N(t_1,i)$. Then $N_i(t_1)$ for all $i \geq 1$ denotes the total number of events within $(0,t_1]$ up to the i^{th} year.

Assuming that $m(t_1) = E[N(t_1,j)]$ for all $j \geq 1$ is finite, the following is true from the strong law of large numbers (Breiman, 1968, p.52), namely

$$\lim_{i\uparrow\infty} N_i(t_1)/i \overset{a.s.}{\rightarrow} m(t_1) \tag{3.8}$$

where 'a.s.' denotes 'almost sure' convergence. For every $\varepsilon > 0$, the criterion for the almost sure convergence is given as (Loève, 1955, p.115)

$$P\{\bigcap_{i}\bigcup_{n} |(N_{i+n}(t_1)/i) - m(t_1)| \geq \varepsilon\} = 0 \tag{3.9}$$

Moreover, it can be deduced from Equation (3.8) that the sequence $\{N_i(t_1)/i\}$ converges in probability, which is given as

$$\lim_{i\uparrow\infty} P\{|(N_i(t_1)/i) - m(t_1)| \geq \varepsilon\} \to 0. \tag{3.10}$$

Equation (3.10) is true because of the fact that a.s. convergence of a sequence of random variables implies convergence in probability (Loève, 1955, p. 116). This concludes the convergence behavior of the random variables $N_i(t_1)$; $i \geq 1$. We now return to Equation (3.2) in order to develop a model for determining the c.d.f. of cumulative rainfall from an event.

3.2.2. Cumulative Distribution Function of Cumulative Rainfall and Water Yield from an Event

Let $\{\tau_j; j\geq 1\}$ be a sequence of random variables denoting termination epochs of events, such that $P(\tau_j > \tau_{j-1}) = 1$ for all $j \geq 1$, and $\tau_0 \equiv 0$. Therefore according to the previously introduced notation $(\tau_{j-1}, \tau_j) = T_j$, for $j = 1, 2, \ldots,$. Now Equation (3.2) can be expressed as,

$$X(t_1) = \sum_{j=1}^{N(t_1)} \int_{\tau_{j-1}}^{\tau_j} \int_B \int \xi(u,v,s)\,dudvds$$

$$+ \int_{\tau(N(t_1))}^{t_1} \int_B \int \xi(u,v,s)\,dudvds. \tag{3.11}$$

As a first step, Equation (3.11) can be simplified by assuming that $\xi(u,v,s) \equiv 0$ for all $s \in [\tau(N(t_1),t_1)]$. This assumption is not serious in our context, since we are considering a fixed season $(0,t_1]$, and looking at all rainfall events that occur within this time. Therefore this assumption implies that we drop an incomplete storm at the tail end of a season, and only consider those rainfall events that terminate prior to the instant t_1. An added advantage of this assumption is that it also simplifies the analysis. Thus in view of this assumption, Equation (3.11) reduces to,

$$X(t_1) = \sum_{j=1}^{N(t_1)} \int_{\tau_{j-1}}^{\tau_j} \int\int_B \xi(u,v,s)\,dudvds. \tag{3.12}$$

A typical sample function of the process $\{X(t_1);t_1>0\}$, as given by Equation (3.12), is shown in Figure 3.2.

Define $Z(\tau_{j-1},\tau_j)$ for $j = 1, 2, \ldots,$ as

$$Z(\tau_{j-1},\tau_j) = \int_{\tau_{j-1}}^{\tau_j} \int\int_B \xi(u,v,s)\,dudvds; \quad j = 1, 2, \ldots \tag{3.13}$$

The sequence of random variables $Z(\tau_{j-1},\tau_j)$ for all $\tau_j < t_1$; $j \geq 1$, forms a discrete parameter stochastic process, and represents cumulative rainfall over the random interval (τ_{j-1},τ_j), $j \geq 1$, which is simply the cumulative rainfall per rainfall event. Moreover, assume that $\{Z(\tau_{j-1},\tau_j)\}$ is an i.i.d. sequence of random variables

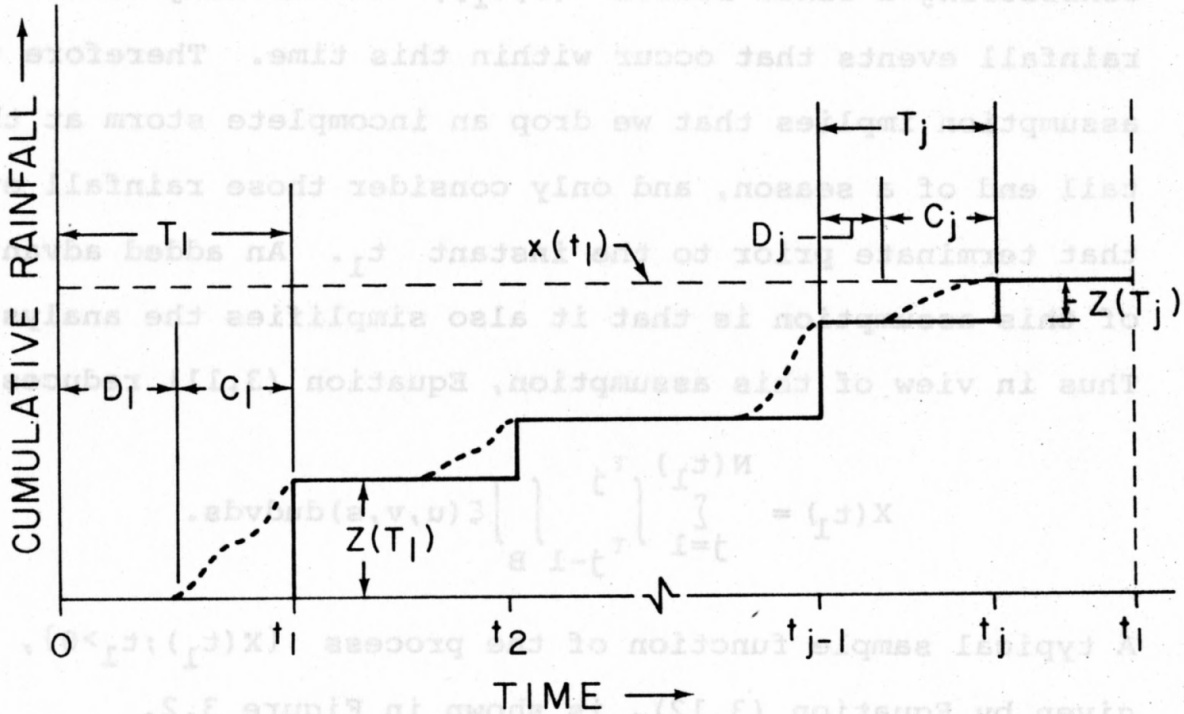


Figure 3.2. A Typical Sample Function of the Stochastic Process of Cumulative Rainfall as given by Equation (3.12).

within the fixed season $(0,t_1]$. Therefore rewriting $(\tau_{j-1},\tau_j) = T_j$; $j \geq 1$, $Z(T_j)$ for all $j \geq 1$, have the same distribution as $Z(T_1)$.

As indicated earlier, T_1 is a sum of two random intervals D_1 and C_1 and is given as $T_1 = D_1 + C_1$, where C_1 is the duration of a rainfall event and D_1 is the dry period preceding this event. Moreover $\xi(u,v,s) = 0$ for all $(u,v) \varepsilon R$, and $s \varepsilon D_1$. Therefore based on Equation (3.13), we denote $Z(T_1) = Z(C_1)$. However in case of excess water yield it is not so, because excess water yield depends on the dry duration preceding a rainfall event through the antecedent soil moisture conditions.

We first consider the random variable $Z(C_1)$ and outline a model to determine its c.d.f. Then using this model, the c.d.f. for the excess water yield is obtained.

The c.d.f. of the random variable $Z(C_1)$ for every $z > 0$, can be written as,

$$P\{Z(C_1) < z\} = \int_0^\infty P\{Z(c) < z | C_1 = c\} dF(c), \qquad (3.14)$$

where $F(c) = P\{C_1 < c\}$ is the c.d.f. of the random variable C_1, denoting random duration. In view of Equation (3.13), Equation (3.14) becomes

$P\{Z(C_1) < z\}$

$$= \int_0^\infty P\{\int_0^c \int_B \int \xi(u,v,s)\,du\,dv\,ds < z \mid C_1 = c\}dF(c). \qquad (3.15)$$

The conditional probability arising in Equation (3.15) is now determined.

Recall that the random duration C_1 of a rainfall event is defined with respect to a region R. Now, based on the phenomenological description of the spatial rainfall process given in Section 2.2.2 of Chapter 2, we assume during the random duration of a rainfall event, that the spatial rainfall occurs only from the storm cells and no rainfall occurs outside of these cells. This assumption is reasonable, because most of the high intensity rainfall takes place within these cells and the rainfall intensities outside these cells are very low (Grayman and Eagleson, 1971). Let $M(B_i, C_1)$ be a positive integer valued random variable, denoting the number of cell centers occurring in some basin $B_i \subset R$, $1 \leq i \leq r$, where r is the total number of disjoint (including contiguous) subbasins contained in the region R. A cell center is defined as the point of maximum rainfall within a cell. Although, $M(B_i, C_1)$ denotes the number of cell centers within $B_i \subset R$, in subsequent discussion we refer to it as denoting the number of cells. This is because we assume that each cell has only one point of maximum rainfall.

Now for a finite number of disjoint (including contiguous) subbasins contained in R, say $B_1, \ldots, B_r$; $r \geq 1$, we obtain finite sequence of random variables, given by

$$\{M(B_i, C_1); B_i \subset R, 1 \leq i \leq r\}, \tag{3.16}$$

denoting the number of storm cells that occur over B_i; $1 \leq i \leq r$, during the random duration C_1 of a rainfall event. In particular, $M(B, C_1)$ denotes the number of storm cells over the river basin B, where $B = \bigcup_{i=1}^{r_1} B_i$; $r_1 \leq r$, during the random duration C_1 of a rainfall event.

Now the conditional probability in Equation (3.15) may be expressed as

$$P\{\int_0^c \int_B \int \xi(u,v,s)\,du\,dv\,ds < z \mid C_1 = c\}$$

$$= P\{\sum_{n=1}^{M(B,c)} \int_{W_{2n-1}}^{W_{2n}} \int_{B_n'} \int \xi(u,v,s)\,du\,dv\,ds < z \mid C_1 = c\}, \tag{3.17}$$

where $\{(W_{2n-1}, W_{2n}); n = 1,2,\ldots\}$ is a sequence of random intervals denoting the random durations of cells. $B_n' \subset B$ for $n \geq 1$, denoting a subset of the basin is given as $B_n' = \{(u,v) \in B; \xi(u,v,s) > 0, s \in (W_{2n-1}, W_{2n})\}$, $n \geq 1$. Phenomenologically, Equation (3.17) can be interpreted as giving the conditional probability of the sum of rainfall from storm cells distributed over the river basin, given that the random duration C_1 of a rainfall event over the region is equal to c.

Let $\{Y_n; n \geq 1\}$ be a sequence of random variables denoting the total rainfall from cells. Then in Equation (3.17), we denote,

$$Y_n = \int_{W_{2n-1}}^{W_{2n}} \int_{B'_n}\int \xi(u,v,s)\,dudvds, \quad n = 1, 2, \ldots, \quad (3.18)$$

Now substituting Equation (3.18) into Equation (3.17), the conditional probability given by Equation (3.17) can be evaluated as

$$P\{\sum_{n=1}^{M(B,c)} Y_n < z | C_1 = c\}$$

$$= \sum_{k=1}^{\infty} P\{\sum_{n=1}^{k} Y_n < z, M(B,c) = k | C_1 = c\}$$

$$+ P\{M(B,c) = 0 | C_1 = c\}. \quad (3.19)$$

Equation (3.19) gives the most general framework to evaluate the conditional probability of the cumulative rainfall per event, given the duration of this event. It may be difficult to obtain any analytic forms of c.d.f.'s using Equation (3.19) in such generality. However further assumptions can be made in order to obtain analytical solutions. To a large extent, this forms the basis of mathematical applications presented in Chapter 4. Special cases including extension to ungaged basins, etc., are discussed in Chapter 4. We now give a derivation to obtain the c.d.f. of excess water yield.

Cumulative Distribution Function of the Excess Water Yield. The excess water yield can be obtained from the cumulative rainfall by use of the type of transformations discussed in Section 2.2.3 of Chapter 2. Moreover it was indicated in Chapter 2, that no theory for general threshold and infiltration model exists. We propose pure threshold model as our basis for transforming rainfall into excess water yield, but phenomenologically give a more meaningful random mechanism than that has been proposed in the past (Woolhiser and Todorovic, 1971). The choice of this transformation also restricts the applicability of this approach to moderate sized basins in regard to prediction of cumulative surface runoff. This is because of the assumption, that excess water yield can be considered approximately to be equal to the surface runoff only for small to mid-sized basins.

Let $\{Z'(T_j); j \geq 1\}$ be a sequence of i.i.d. random variables denoting excess water yield. $Z'(T_j)$ is distributed as $Z'(T_1)$ for all $j = 1, 2, \ldots,$. Define $Z'(T_1)$ as,

$$Z'(T_1) = Z(C_1) - g(D_1) \tag{3.20}$$

where the threshold value is given by $g(D_1)$, which denotes some functions for the preceding random dry period D_1. Strictly speaking, the threshold value may also depend upon the current storm conditions, e.g. duration, intensity, rainfall, etc., (Fogel, 1969), but presently we assume the preceding dry interval to be of importance during which soil moisture gets depleted.

Now the c.d.f. of $Z'(T_1)$, for any $z > 0$ can be obtained using Equations (3.14) and (3.20) as

$$P\{Z'(T_1) < z\} = P\{Z(C_1) < z + g(D_1)\}$$

$$= \int_0^\infty P\{Z(C_1) < z + g(d_1) | D_1 = d_1\} dF(d_1), \tag{3.21}$$

where $F(d_1) = P\{D_1 < d_1\}$. But $Z(C_1)$ is independent of the random interval D_1, hence Equation (3.21) reduces to,

$$P\{Z'(T_1) < z\}$$

$$= \int_0^\infty \int_0^\infty P\{Z(c) < z + g(d_1) | C_1 = c\} dF(c) dF(d_1), \tag{3.22}$$

where $F(c) = P\{C_1 < c\}$. The conditional probability in Equation (3.22) can be evaluated using Equation (3.19).

Finally, note that the conditional expectation and higher moments, etc., of $Z(C_1)$ can be obtained either using Equation (3.19) or directly as follows

$$E[Z(c)|C_1 = c] = \sum_{n=1}^{\infty} E[\sum_{j=1}^{n} Y_j I\{M(B,c) = n\}|C_1 = c]$$

$$= \sum_{n=1}^{\infty} \sum_{j=1}^{n} E[Y_j I\{M(B,c) = n\}|C_1 = c] \qquad (3.23)$$

where $I\{A\}$ is the indicator function of a set A. Equation (3.23) gives the conditional expectation of $Z(C_1)$. Similarly, higher conditional moments can be obtained.

We close the section on the general approach with the note that specific applications of the above will be given in Chapter 4.

3.3. Processes of the Cumulative Rainfall and Water Yield for a Season

3.3.1. Derivation of the First Two Moments

Ideally, we would like to obtain the c.d.f. of $X(t_1)$; $t_1 > 0$, as given by Equation (3.12). However, under the general dependence structure between $N(t_1)$ and $Z(T_j)$; $j \geq 1$, such is not obtainable. Therefore under this section we derive expressions for the first two moments of $X(t_1)$, namely $E[X(t_1)]$ and $E[X^2(t_1)]$. In its own right, such information can be of importance to an engineer.

Based on Equation (3.12),

$$E[X(t_1)] = E[\sum_{j=1}^{N(t_1)} Z(T_j) I\{\bigcup_{n=0}^{\infty} \{N(t_1) = n\}\}]. \qquad (3.24)$$

Now $\{N(t_1) = k\} \cap \{N(t_1) = j\} = \emptyset$ (Null Set) for $j \neq k$, and $I\{\bigcup_{n=0}^{\infty} \{N(t_1) = n\}\} = \sum_{n=0}^{\infty} I\{N(t_1) = n\} = 1$. Using this fact, Equation (3.24) becomes,

$$E[X(t_1)] = \sum_{n=1}^{\infty} E[\sum_{j=1}^{n} Z(T_j) I\{N(t_1) = n\}]. \qquad (3.25)$$

Since in Section 3.2.2., $Z(T_j)$ for $j \geq 1$, are assumed to be i.i.d. as $Z(T_1)$, therefore Equation (3.25) can be written as,

$$E[X(t_1)] = \sum_{n=1}^{\infty} nE[Z(T_1) I\{N(t_1) = n\}]. \qquad (3.26)$$

Substituting Equation (3.4) in Equation (3.26), we obtain

$$E[X(t_1)] = \sum_{n=1}^{\infty} nE[Z(T_1) I\{\sum_{j=1}^{n} T_j < t_1 < \sum_{j=1}^{n+1} T_j\}]$$

$$= \sum_{n=1}^{\infty} n\int_0^{t_1} E[Z(s) I\{\sum_{j=2}^{n} T_j < t_1 - s < \sum_{j=2}^{n+1} T_j\} | T_1 = s] dF(s) \qquad (3.27)$$

Note that the random variable $I\{\sum_{j=2}^{n} T_j < t_1 - s < \sum_{j=2}^{n+1} T_j\}$ is independent of the random variable T_1, and in turn is independent of $Z(T_1)$, since $Z(T_1)$ depends only on T_1. Hence Equation (3.27) reduces to,

$$E[X(t_1)]$$

$$= \sum_{n=1}^{\infty} n \int_0^{t_1} E[Z(s)|T_1 = s] P\{\sum_{j=2}^{n} T_j < t_1 - s < \sum_{j=2}^{n+1} T_j\} dF(s) \tag{3.28}$$

because $E[I\{A\}] = P\{A\}$ for any set A. From Equation (3.4), it follows that $P\{N(t_1) = n - 1\} = P\{\sum_{j=1}^{n-1} T_j < t_1 < \sum_{j=1}^{n} T_j\}$. Since the right hand side of this identity involves an event defined from the sum of n i.i.d. random variables T_j, $1 \le j \le n$, therefore the following is true,

$$P\{\sum_{j=2}^{n} T_j < t_1 < \sum_{j=2}^{n+1}\} = P\{\sum_{j=1}^{n-1} T_j < t_1 < \sum_{j=1}^{n} T_j\}$$

$$= P\{N(t_1) = n - 1\}. \tag{3.29}$$

Substituting Equation (3.29) into Equation (3.28),

$$E[X(t_1)] = \sum_{n=1}^{\infty} n \int_0^{t_1} E[Z(s)|T_1 = s] P\{N(t_1 - s) = n - 1\} dF(s)$$

$$= \int_0^{t_1} E[Z(s)|T_1 = s] \sum_{n=1}^{\infty} (n-1) P\{N(t_1 - s) = n - 1\} dF(s)$$

$$+ \int_0^{t_1} E[Z(s)|T_1 = s] \sum_{n=1}^{\infty} P\{N(t_1 - s) = n - 1\} dF(s)$$

or equivalently,

$$E[X(t_1)] = \int_0^{t_1} E[Z(s)|T_1 = s]m(t_1-s)dF(s)$$

$$+ \int_0^{t_1} E[Z(s)|T_1 = s]dF(s), \tag{3.30}$$

where $m(t_1-s) = E[N(t_1-s)]$ as defined in Section 3.2.1. The conditional expectation $E[Z(s)|T_1 = s]$ can be obtained from Equation (3.23).

The derivation of the $E[X^2(t_1)]$ can be done along similar lines as that for $E[X(t_1)]$. Therefore, we give the derivation in Appendix A. The final expression for $E[X^2(t_1)]$ is obtained as

$$E[X^2(t_1)] = \int_0^{t_1} E[Z^2(s_1)|T_1 = s_1]m(t_1-s_1)dF(s_1)$$

$$+ \int_0^{t_1} E[Z^2(s_1)|T_1 = s_1]dF(s_1)$$

$$+ \int_0^{t_1}\int_0^{t_1-s_2} \sum_{n=1}^{\infty} n(n-1)E[Z(s_1)|T_1 = s_1]E[Z(s_2)|T_2 = S_2]$$

$$P\{n(t_1-s_1-s_2) = n - 2\}dF(s_1)dF(s_2). \tag{3.31}$$

A special case for $E[X(t_1)]$ is now derived based on Equation (3.30).

In the case $Z(T_j)$'s; $j \geq 1$, are independent of T_j for all $j \geq 1$, and in turn are independent of the counting process $N(t_1)$, we get the case of the sum of a random number of i.i.d. random variables. Such a situation would arise when the $Z(T_j)$; $j \geq 1$, as given by Equation (3.13) depends only on the path of the sample function $\xi(u,v,s)$, and not on the random interval T_j. For this case denote $Z(T_j) = Z_j^*$; $j \geq 1$. Then $E[Z_1^*|T_1 = s] = E[Z_1^*]$. Substituting this in Equation (3.30),

$$E[X(t_1)] = E[Z_1^*]\{\int_0^{t_1} m(t_1 - s)dF(s) + \int_0^{t_1} dF(s)\}, \qquad (3.32)$$

but the expression within brackets in Equation (3.32) is the integral equation for the mean value function of a renewal counting process (Parzen, 1967, p. 171), which is given as

$$m(t_1) = F(t_1) + \int_0^{t_1} m(t_1 - s)dF(s). \qquad (3.33)$$

Substituting Equation (3.33) into Equation (3.32),

$$E[X(t_1)] = E[Z_1^*]m(t_1) = E[Z_1^*]E[N(t_1)], \qquad (3.34)$$

which is the expression that has been used earlier in hydrologic literature (Duckstein et al., 1972).

Also a general expression for $E[X(t_1)]$ has been given by Todorovic (1970), but he does not specify the "dependence" sturcture between the $Z(T_j)$; $j \geq 1$, and the $N(t_1)$. However his expression is given as,

$$E[X(t_1)] = \sum_{n=1}^{\infty} \int_{\{\tau_n < t_1\}} E[Z(T_n) | N(t_1)] dP. \tag{3.35}$$

Starting from Equation (3.35) we derive Equation (3.30) in Appendix B.

Similarly, as in the case of $E[X(t_1)]$, we can reduce Equation (3.31) to the expression for the variance of the sum of a random number of i.i.d. random variables. Denote $Z(T_j) = Z_j^*$; $j \geq 1$, as before, and note that $E[(Z_1^*)^2 | T_1 = s] = E[(Z_1^*)^2]$. Making this substitution in Equation (3.31) we obtain

$$E[X^2(t_1)] = E[(Z_1^*)^2] \int_0^{t_1} m(t_1 - s_1) dF(s_1) + E[(Z_1^*)^2] F(t_1)$$
$$+ E^2[Z_1^*] \sum_{n=1}^{\infty} n(n-1) \int_0^{t_1} \int_0^{t_1 - s_2} P\{N(t_1 - s_1 - s_2) = n - 2\} dF(s_1) dF(s_2). \tag{3.36}$$

Now the following identity follows from Equation (3.4),

$$\int_0^{t_1} \int_0^{t_1 - s_2} P\{N(t_1 - s_1 - s_2) = n - 2\} dF(s_1) dF(s_2) = P\{N(t_1) = n\}. \tag{3.37}$$

Substituting Equations (3.33) and (3.37) into Equation (3.36),

$$E[X^2(t_1)] = E[(Z_1^*)]m(t_1) + E^2[Z_1^*] \sum_{n=1}^{\infty} n(n-1)P\{N(t_1) = n\}$$

$$= E[(Z_1^*)^2]m(t_1) + E^2[Z_1^*](m_2(t_1) - m(t_1)), \qquad (3.38)$$

where $m_2(t_1)$ is $E[N^2(t_1)]$. Now let $\sigma^2[X(t_1)]$ denote the variance of $X(t_1)$, then,

$$\sigma^2[X(t_1)] = E[X^2(t_1)] - E^2[X(t_1)]. \qquad (3.39)$$

Substituting Equations (3.38) and (3.34) into Equation (3.39), the variance of $X(t_1)$ is given as,

$$\mathrm{Var}[X(t_1)] = m(t_1)\{E[(Z_1^*)^2] - E^2[Z_1^*]\}$$

$$+ E^2[Z_1^*]m_2(t_1) - E^2[Z_1^*]m^2(t_1) = m(t_1)\sigma^2[Z_1^*]$$

$$+ E^2[Z_1^*]\sigma^2[N(t_1)], \qquad (3.39)$$

where $\sigma^2[Z_1^*]$ and $\sigma^2[N(t_1)]$ denote the variance of Z_1^* and $N(t_1)$ respectively. $\sigma^2[N(t_1)]$ can be obtained from Equations (3.6) and (3.7). Equation (3.39) has also been used in hydrologic literature previously (Duckstein et al., 1972).

<u>Extension to More Than One Season</u>. The derivations presented above can be extended to more than one season.

For example, let $(0,t]$ be divided into two seasons, say $(0,t_1]$ and $(t_1,t]$. Let $X(t)$ denote the total cumulative rainfall up to time t. Then $X(t)$ can be written as,

$$X(t) = X(t_1) + X(t) - X(t_1), \tag{3.40}$$

where $X(t) - X(t_1)$, denotes the cumulative rainfall within the season $(t_1,t]$. Hence

$$E[X(t)] = E[X(t_1)] + E[X(t) - X(t_1)]. \tag{3.41}$$

Assuming that the random variables denoting the cumulative rainfall from one season to another are mutually independent, the variance of $X(t)$ denoted by $\sigma^2[X(t)]$ is given as,

$$\sigma^2[X(t)] = \sigma^2[X(t_1)] + \sigma^2[X(t) - X(t_1)]. \tag{3.42}$$

In Equations (3.41) and (3.42), the random variables $X(t_1)$ and $X(t) - X(t_1)$, can be non-identically distributed. Phenomenologically, such a situation can arise from differences in the type of rainfall storms that may occur over the region R from one season to another. For example, the occurrence of air-mass type of convective storms during summer season, and the frontal storms during winter season in the southwestern United States, can make the corresponding cumulative seasonal rainfall amounts non-identically distributed.

The derivations given in this section are also applicable to the excess water yield per event. This is due to the fact that in the previous derivations, $Z(T_1)$ denoting the cumulative rainfall per event is considered to be dependent on the random variable T_1, which is also true for $Z'(T_1)$ denoting the excess water yield per event.

This concludes our discussion on obtaining the first two moments of the cumulative rainfall and excess water yield within a finite time interval. Note that the derivations become 'involved' even for the second moment of $X(t_1)$. This in turn can give an indication in regard to the difficulties in obtaining an analytical form for the c.d.f. of $X(t_1)$ within a finite time interval.

3.4. Limiting Distribution and Convergence Rate for Cumulative Seasonal Rainfall

Let $\{X_i(t_1)\}$ be a sequence of random variables such that for every $i \geq 1$, $X_i(t_1)$ denotes the cumulative rainfall up to the i^{th} year within a fixed season $(0,t_1]$. Then for every $x > 0$, $\lim_{i\uparrow\infty} P\{X_i(t_1) < x\}$ gives the limiting c.d.f. of the random variables $X_i(t_1)$; $i \geq 1$. This is considered in the context of the rate of convergence of $P\{X_i(t_1) < x\}$; $i \geq 1$, to its limiting c.d.f., in the following theorem.

Some new notation is now introduced to achieve facility in giving the theorem. Let $Z_j = (Z(T_j) - E[Z(T_j)])/\sigma[Z(T_j)]$; $j \geq 1$, where $\sigma^2[Z(T_j)]$ is the variance of $Z(T_j)$. Clearly $E[Z_j] = 0$ and $\sigma^2[Z_j] = 1$; $j \geq 1$. Recall from Section 3.2.1 that $N(t_1,j)$ is defined as the number of events within $(0,t_1]$ during the j^{th} year, $j \geq 1$. Denote $N(t_1,j) = N(j)$ for all $j \geq 1$, and similarly $N_i = \sum_{j=1}^{i} N(j)$; $i \geq 1$. Let

$$X_i(t_1) = X(N_i) = \sum_{j=1}^{N_i} Z_j, \quad i = 1, 2, \ldots,. \tag{3.43}$$

Note that in the above notation we have <u>suppressed the variable</u> t_1, but it is understood that we are still referring to the same fixed interval $(0,t_1]$, for all the years.

To this end, we give a lemma in regard to the sequence $\{N(j)\}$, which is used in the theorem.

LEMMA. Let $N(j)$ be a sequence of i.i.d. random variables with mean m and variance σ^2. Assume that $E|N(1)|^{2p} < \infty$, for some $p \geq 1$. Denote $N_i = N(1) + \ldots + N(i)$. Then (a) $(N_i - m_i)|\sigma\sqrt{i}$ converges in distribution to the standard normal, and (b)

$$\lim_{i\uparrow\infty} E\left(\frac{N_i - m_i}{\sigma\sqrt{i}}\right)^{2n} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^{2n} e^{-x^2/2} dx, \quad 1 \leq n \leq p \tag{3.44}$$

and the right hand side is bounded by some function of p, say $b(p)$, because all moments of the standard normal distribution are bounded.

Proof. (a) follows simply from the classical central limit theorem. For the proof of (b), see Bhattacharaya (1973).

We now give the theorem on the rate of convergence of the sequence $\{X(N_i)\}$.

THEOREM 1. Let $\{Z_i\}$ be an i.i.d. sequence of random variables with mean zero and variance one. Put $X(n) = Z_1 + \ldots + Z_n$. Let $\{N(j)\}$ be a sequence of positive integer valued i.i.d. random variables, $N(j)$; $j \geq 1$, is not independent of Z_i; $i \geq 1$, and such that $E[N(j)]^{2p} < \infty$ for some $p \geq 1$. Let $N_i = N(1) + \ldots + N(i)$. If $\Phi(x)$ denotes the c.d.f. of a standard normal distribution, then

$$\sup_x \left| P\left\{\frac{X(N_i)}{\sqrt{N_i}} < x\right\} - \Phi(x)\right| = O(1/i)^{p/(6p+1)}. \tag{3.45}$$

Proof. Let $\varepsilon > 0$ be arbitrary. By Chebyshev's inequality,

$$P\{|N_i - mi| > \varepsilon mi\} = \delta(\varepsilon, i) \leq E[N_i - mi]^{2p}/(\varepsilon mi)^{2p}. \tag{3.46}$$

Using the Lemma in the right hand side, gives

$$\delta(\varepsilon,i) \leq E[N_i - mi]^{2p}/(\varepsilon mi)^{2p}$$

$$\leq b(p)\sigma^{2p} i^p/(\varepsilon mi)^{2p} = a_1/\varepsilon^{2p} i^p \tag{3.47}$$

where $a_1 = b(p)\sigma^{2p}/m^{2p}$ is some constant. Now, for any $-\infty < x < \infty$,

$$P\{\frac{X(N_i)}{\sqrt{N_i}} < x\} = \sum_{n=0}^{\infty} P\{\frac{X(n)}{\sqrt{n}} < x, N_i = n\}$$

$$= \sum_{|n-mi|<mi\varepsilon} P\{\frac{X(n)}{\sqrt{n}} < x, N_i = n\}$$

$$+ \sum_{|n-mi|\geq mi\varepsilon} P\{\frac{X(n)}{\sqrt{n}} < x, N_i = n\}. \tag{3.48}$$

Based on Equation (3.46), Equation (3.48) is written as,

$$|P\{\frac{X(N_i)}{\sqrt{N_i}} < x\} - \sum_{|n-mi|<mi\varepsilon} P\{\frac{X(n)}{\sqrt{n}} < x, N_i = n\}|$$

$$\leq P\{|N_i - mi| > m\varepsilon i\} = \delta(\varepsilon,i). \tag{3.49}$$

Let $n_1 = [m(1-\varepsilon)i]$ and $n_2 = [m(1+\varepsilon)i]$, where [---] denotes the integral part of the number within square brackets Now for $|n - mi| < \varepsilon mi$,

$$P\{\frac{X(n)}{\sqrt{n}} < x, N_i = n\} \leq P\{X(n_1) < x\sqrt{n_2} + \rho, N_i = n\} \tag{3.50}$$

where $\rho = \max_{n_1<n\leq n_2} |\sum_{n_1<k\leq n} Z_k|$. Similarly, we get,

$$P\{\frac{X(n)}{\sqrt{n}} < x, N_i = n\} \geq P\{X(n_1) < x\sqrt{n_1} - \rho, N_i = n\}. \tag{3.51}$$

From Kolmogorov's inequality (Breiman, 1968, p. 65),

$$P\{\rho \geq (\varepsilon)^{1/3}\sqrt{n_1}\} \leq (n_2 - n_1)/n_1(\varepsilon)^{2/3}. \tag{3.52}$$

Now let G be the event $\{\rho < (\varepsilon)^{1/3}\sqrt{n_1}\}$ and H the event $\{|N_i - m_i| < \varepsilon mi\}$, then from Equations (3.49), (3.50), (3.51) and (3.52), it follows that

$$P\{\frac{X(N_i)}{\sqrt{N_i}} < x\} \leq P\{\frac{X(n_1)}{\sqrt{n_1}} < x\sqrt{\frac{n_2}{n_1}} + (\varepsilon)^{1/3}, G \cap H\}$$

$$+ \frac{(n_2-n_1)}{n_1(\varepsilon)^{2/3}} + \delta(\varepsilon,i) \tag{3.53}$$

and,

$$P\{\frac{X(N_i)}{\sqrt{N_i}} < x\} \geq P\{\frac{X(n_1)}{\sqrt{n_1}} < x - (\varepsilon)^{1/3}, G \cap H\} - \delta(\varepsilon,i)$$

$$\geq P\{\frac{X(n_1)}{\sqrt{n_1}} < x - (\varepsilon)^{1/3}\} - \frac{(n_2-n_1)}{n_1(\varepsilon)^{2/3}} - 2\delta(\varepsilon,i). \tag{3.54}$$

Hence it follows from Equations (3.53) and (3.54), that

$$P\{\frac{X(n_1)}{\sqrt{n_1}} < x - (\varepsilon)^{1/3}\} - \frac{2\varepsilon^{1/3}}{(1-\varepsilon)} - 2\delta(\varepsilon,i) \leq P\{\frac{X(N_i)}{\sqrt{N_i}} < x\}$$

$$\leq P\{\frac{X(n_1)}{\sqrt{n_1}} < x\sqrt{\frac{1+\varepsilon}{1-\varepsilon}} + (\varepsilon)^{1/3}\}$$

$$+ \frac{2\varepsilon^{1/3}}{1-\varepsilon} + \delta(\varepsilon,i). \tag{3.55}$$

Based on the Berry-Eseen theorems (see Bhattacharaya, 1973) for the sum of i. i.d. random variables with zero mean and unit variance, the left and right sides of the inequality given by Equation (3.55) can be written as,

$$P\{\frac{X(n_1)}{\sqrt{n_1}} < x - (\varepsilon)^{1/3}\} \geq \Phi(x) - (\varepsilon)^{1/3} - \frac{1.6\beta_3}{\sqrt{n_1}(\beta_2)^{3/2}} \tag{3.56}$$

assuming that $\beta_3 = EZ_1^3 < \infty$ and $\beta_2 = EZ_1^2 < \infty$. Now from Lipschitz condition for the function $\Phi(\cdot)$, we get,

$$\Phi(x - (\varepsilon)^{1/3}) \geq \Phi(x) - (\varepsilon)^{1/3}/\sqrt{2\pi}. \tag{3.57}$$

Similarly, we obtain,

$$P\{\frac{X(n_1)}{\sqrt{n_1}} < x\sqrt{\frac{1+\varepsilon}{1-\varepsilon}} + (\varepsilon)^{1/3}\} \leq \Phi(x\sqrt{\frac{1+\varepsilon}{1-\varepsilon}}) + \frac{(\varepsilon)^{1/3}}{\sqrt{2\pi}}$$

$$+ \frac{1.6\beta_3}{\sqrt{n_1}(\beta_2)^{3/2}}. \tag{3.58}$$

Since $\sqrt{\frac{1+\varepsilon}{1-\varepsilon}}$ is of the order of $(1 + \varepsilon)$, hence,

$$\Phi(x\sqrt{\frac{1+\varepsilon}{1-\varepsilon}}) = \Phi(x(1+\varepsilon)) = \Phi(x) + \frac{1}{\sqrt{2\pi}}\int_x^{x(1+\varepsilon)} e^{-u^2/2}du. \tag{3.59}$$

Let $y = u/x$, then

$$\frac{1}{\sqrt{2\pi}}\int_x^{x(1+\varepsilon)} e^{-u^2/2}du = \frac{1}{\sqrt{2\pi}}\int_1^{1+\varepsilon} e^{-y^2x^2/2}xdy$$

$$\leq \frac{1}{\sqrt{2\pi}} e^{-x^2/2}x\varepsilon \leq \frac{1}{\sqrt{2\pi}}e^{-1/2}\varepsilon. \tag{3.60}$$

Combining Equations (3.58), (3.59) and (3.60),

$$P(\frac{X(n_1)}{\sqrt{n_1}} < x\sqrt{\frac{1+\varepsilon}{1-\varepsilon}} + (\varepsilon)^{1/3}) \leq \Phi(x) + \frac{1}{\sqrt{2\pi}}e^{-1/2}\varepsilon$$

$$+ \frac{(\varepsilon)^{1/3}}{\sqrt{2\pi}} + \frac{1.6\beta_3}{\beta_2\sqrt{n_1}}. \tag{3.61}$$

Substituting Equations (3.56), (3.57) and (3.61) into Equation (3.55),

$$\Phi(x) - \frac{(\varepsilon)^{1/3}}{\sqrt{2\pi}} - \frac{2(\varepsilon)^{1/3}}{(1-\varepsilon)} - \frac{1.6\beta_3}{\sqrt{n_1}(\beta_2)^{3/2}} - 2\delta(\varepsilon,i)$$

$$\leq P\{\frac{X(N_i)}{\sqrt{N_i}} < x\} \leq \Phi(x) + \frac{(\varepsilon)^{1/3}}{\sqrt{2\pi}} + \frac{{}^{1/3}e^{-1/2}}{\sqrt{2\pi}}$$

$$+ \frac{1.6\beta_3}{\beta_2\sqrt{n_1}} + \frac{2(\varepsilon)^{1/3}}{1-\varepsilon} + \delta(\varepsilon,i)$$

or equivalently,

$$|P\{\frac{X(N_i)}{\sqrt{N_i}} < x\} - \Phi(x)| \leq (\varepsilon)^{1/3}(\frac{e^{-1/2}}{\sqrt{2\pi}} + \frac{1}{\sqrt{2\pi}} + \frac{2}{1-\varepsilon})$$

$$+ 2\delta(\varepsilon,i) + \frac{1.6\beta_3}{\beta_2\sqrt{mi(1-\varepsilon)}}. \tag{3.62}$$

Note that the right hand side in Equation (3.62) is independent of x, $(1-\varepsilon)$ is of the order of one. Therefore,

$$\sup_x |P\{\frac{X(N_i)}{\sqrt{N_i}} < x\} - \Phi(x)| \leq \varepsilon^{1/3}a_2 + \frac{a_3}{\sqrt{i}} + 2\delta(\varepsilon,i), \tag{3.63}$$

where $a_2 = 2 + (e^{-1/2} + 1)/\sqrt{2\pi}$, $a_3 = 1.6\beta_3/\beta_2\sqrt{m}$. Substituting Equation (3.47) into Equation (3.63),

$$\sup_x |P\{\frac{X(N_i)}{\sqrt{N_i}} < x\} - \Phi(x)| \leq a_2\varepsilon^{1/3} + a_3(i)^{-1/2}$$

$$+ a_1(i)^{-p}(\varepsilon)^{-2p}. \tag{3.64}$$

Minimum of the bound on the right hand side is now obtained by differentiating it with respect to ε. Thus

$$\frac{\varepsilon}{d\varepsilon}(a_2\varepsilon^{1/3} + a_3(i)^{-1/2} + a_1(i)^{-p}(\varepsilon)^{-2p}) = 0$$

or

$$\varepsilon = (2pa_1/a_2)^{3/(6p+1)}(i)^{-3p/(6p+1)} \sim (i)^{-3p/(6p+1)} \quad (3.65)$$

Substituting Equation (3.65) into the r.h.s. of Equation (3.64), the order of the error bound on the r.h.s. is given as

$$\sup_x \left|P\left\{\frac{X(N_i)}{\sqrt{N_i}} < x\right\} - \Phi(x)\right| = O(1/i)^{p/(6p+1)},$$

which was to be proved.

COROLLARY. Theorem of Anscombe (1952) follows from the above as a special case. However, the added restriction $E[N(j)]^{2p} < \infty$ is not required. Recall from Section 3.2.1, that $\{N_i\}$ denote the partial sums of i.i.d. random variables and it is shown in Equation (3.10), that N_i/i converges in probability. This implies that $\delta(\varepsilon,i) \to 0$ as $i \uparrow \infty$. Since $\varepsilon > 0$ is arbitrary, therefore from Equation (3.55), it can be concluded that,

$$P\{\frac{X(N_i)}{\sqrt{N_i}} < x\} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-u^2/2}du. \tag{3.66}$$

Operationally speaking, such convergence rates can be used to give a quantitative estimate of the deviation of the c.d.f. of the cumulative seasonal rainfall up to the i^{th} year $i \geq 1$, from its limiting c.d.f. Such considerations are necessary, if the c.d.f. of the cumulative seasonal rainfall is not obtainable in an analytical form. However, the rate of convergence as given by our theorem, provides a loose bound. For example, in case $p = 1$, the order of the bound is $(1/i)^{1/7}$. For large values of p, the order of this bound converges to $(1/i)^{1/6}$, for all $i \geq 1$. Therefore the results given here can only be considered as a first step towards such efforts. Moreover, in light of the obtained order of the bound, it can only be regarded as being more of a mathematical interest than operational interest.

Although the above discussion is given in terms of cumulative rainfall, the results also hold good for the excess water yield.

Cumulative Rainfall from Multiple Seasons. In this section, we give a derivation to obtain the limiting c.d.f. of $X(t)$, $t > 0$, such that the interval $(0,t]$ is divided into more than one season, say $(0,t_1]$, $(t_1,t_2]$, etc., and within each season the random variables denoting cumulative rainfall $X(t_1)$, $X(t_2) - X(t_1)$, ..., are mutually independent but <u>non-identically</u> distributed.

For demonstrative purposes, $(0,t]$ is divided into two seasons, $(0,t_1]$, $(t_1,t]$. Extension to more than two seasons follows identically to the development given below.

Let $\{T_j'\}$ be a sequence of i.i.d. random variables within $(t_1,t]$, representing the random time intervals between termination of successive rainfall events. However $\{T_j'\}$ within $(t_1,t]$ can have a different c.d.f. than that of the c.d.f. of $\{T_j\}$ defined within the interval $(0,t_1]$. Let $N_i(t_1,t)$ denote the counting process of number of rainfall events within $(t_1,t]$ up to the i^{th} year. If $m(t_1,t) = E[N(t_1,t)] < \infty$, then for every $\varepsilon > 0$, the following is true based on Equation (3.10),

$$P\{|(N_i(t_1,t)/i) - m(t_1,t)| > \varepsilon\} \to 0. \tag{3.67}$$

Moreover assume that the general approach to derive the c.d.f. of cumulative rainfall per event as given in Section 3.2.2 also holds well for the season $(t_1,t]$. Let $\{Z(T_j')\}$ denote the cumulative rainfall per event within $(t_1,t]$, such that $Z(T_j')$ within $(t_1,t]$ and $Z(T_j)$ within $(0,t_1]$ are non-identically distributed.

Based on the above assumptions, $X(t_1)$ and $X(t) - X(t_1)$ are also non-identically distributed. Let

$N_i(t) = N_i(t_1) + N_i(t_1,t)$. Denoting $X(t_1,t) = X(t) - X(t_1)$, the c.d.f. of $X_i(t)$ is given as,

$$P\{\frac{X_i(t)}{\sqrt{N_i(t)}} < x\} = P\{\frac{X_i(t_1)}{\sqrt{N_i(t)}} + \frac{X_i(t_1,t)}{\sqrt{N_i(t)}} < x\}$$

$$= P\{\frac{X_i(t_1)}{\sqrt{N_i(t_1)}}\sqrt{\frac{N_i(t_1)}{N_i(t)}} + \frac{X_i(t_1,t)}{\sqrt{N_i(t_1,t)}}\sqrt{\frac{N_i(t_1,t)}{N_i(t)}} < x\}. \tag{3.68}$$

But

$$\left.\begin{aligned} &\lim_{i\uparrow\infty} \frac{N_i(t_1)}{N_i(t)} \overset{a.s.}{\rightarrow} \frac{m(t)}{m(t_1)+m(t_1,t)} \\ &\text{and} \\ &\lim_{i\uparrow\infty} \frac{N_i(t_1,t)}{N_i(t)} \overset{a.s.}{\rightarrow} \frac{m(t_1,t)}{m(t_1)+m(t_1,t)} \end{aligned}\right\}. \tag{3.69}$$

Now both random variables on the left hand side in Equation (3.68) converge to a normal c.d.f., with means zero and variances $m(t)/(m(t_1) + m(t_1,t))$ and $m(t_1,t)/(m(t_1) + m(t_1,t))$. Therefore their sum will converge to a standard normal distribution with mean zero and variance one. This follows from the well known fact that sum of independent normal random variables is again normal, and its mean and variance is given by the sums of means and variances of the original random variables respectively. Hence,

$$\lim_{i\uparrow\infty} P\{\frac{X_i(t)}{\sqrt{N_i}(t)} < x\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du. \qquad (3.70)$$

3.5. Limiting Distribution for Maximum Cumulative Rainfall per Event

Extreme valued distributions of hydrologic variables have been used in the past in connection with the determination of the average return periods of extreme random events defined in terms of these hydrologic variables. However, it has been recognized that the concept of average return period is very misleading, since the actual return periods can be much less than the average (Kendall, 1959). Nonetheless, the derivations of the actual or average return periods in the past have been considered only for the cases where the c.d.f. of the random variable (maximum seasonal or annual flow or rainfall, etc.) is know for a finite time (generally a year or a season). In our case, since we cannot obtain an expression for the c.d.f. of the maximum cumulative rainfall per event within a season $(0,t_1]$, because of dependence between $Z(C_1)$ and $N(t_1)$, we would show the use of a limiting distribution to obtain the necessary information on the actual return periods. In this spirit, we also recommend that the concept of the average return period should be completely dispensed with in hydrology, since actual return periods can be directly obtained for the cases where the c.d.f. of a random variable over a year or a season is known

(Todorovic and Zelenhasic, 1970), without going through the route of the average return period.

To this end, we first give the treatment to obtain the limiting c.d.f. of the maximum rainfall per event, and then present its application in a separate sub-section in the spirit of obtaining actual return periods.

The treatment is first presented for a single season $(0,t_1]$, and is then extended to multiple seasons. Now define $V(N_i)$ as,

$$V(N_i) = \sup_{i \leq j \leq N_i} Z(T_j). \tag{3.71}$$

Recall from Section 3.2.1 that $N_i = N_i(t_1)$ is the random variable denoting the number of rainfall events within $(0,t_1)$ up to the i^{th} year, $i \geq 1$.

Now consider the sequence of i.i.d. random variables $Z(T_j)$, $j \geq 1$. Genedenko (1942) was the first one to give the necessary and sufficient conditions for the existence of limiting c.d.f. of maximum of a <u>fixed</u> number of i.i.d. random variables. Let

$$V(n) = \max_{i \leq j \leq n} Z(T_j). \tag{3.72}$$

Then,

$$P\{V(n) < x\} = [P\{Z(T_j) < x\}]^n = [F(x)]^n. \tag{3.73}$$

If there exists a c.d.f. $G(x)$, and sequences $\{u_n\}$ and $\{v_n\}$ such that $u_n > 0$ for $n \geq 1$, and

$$\lim_{n\uparrow\infty} P\{(V(n) - v_n)u_n^{-1} < x\} = G(x) \tag{3.74}$$

at each continuity point x of $G(x)$, then $G(x)$ belongs to one of the following types, i.e.,

$$\left.\begin{array}{lll} G_1(x) = \exp[-x^{-\alpha}] & \text{for} & x > 0, \quad \alpha > 0 \\ G_1(x) = 0 & \text{for} & x \leq 0 \end{array}\right\} \tag{3.75}$$

$$\left.\begin{array}{lll} G_2(x) = \exp[-(-x)^{\alpha}] & \text{for} & x > 0; \quad \alpha > 0 \\ G_2(x) = 0 & \text{for} & x \leq 0 \end{array}\right\} \tag{3.76}$$

$$G_3(x) = \exp[-e^{-x}] \qquad -\infty < x < \infty. \tag{3.77}$$

Since type three, i.e., $G_3(x)$, is relevant to our case, we give the necessary and sufficient conditions for the convergence of the c.d.f. of $V(n)$ to the third type. For this case the constants $\{u_n\}$ and $\{v_n\}$ satisfy the relations,

$$1 - F(v_n) \sim n^{-1} \quad \text{and} \quad 1 - F(u_n+v_n) \sim (ne)^{-1} \qquad (3.78)$$

for $v_n \leq v_{n+1}$ and $n \geq 1$. The above results of Genedenko are reproduced from Berman (1962).

The limiting c.d.f. of the maximum of a random number of random variables under the condition that $Z(T_j)$ are dependent on the counting process $N(t_1)$, is given by Berman (1962). We state Berman's theorem below.

THEOREM 2. If $\{N_i\}$ and $Z(T_j)$ are not necessarily independent of each other, and if there exists a positive number m, such that $N_i|i \to m$ in probability, then for every x,

$$\lim_{i \uparrow \infty} P\{u_i^{-1}(V(N_i) - v_i) < x\} = [G(x)]^m \qquad (3.79)$$

where $G(x)$ is one of the three types as given earlier by Equations (3.75), (3.76) and (3.77).

Recall that we established the convergence of $N_i|i$ to $m(t_1)$ in probability in Equation (3.10). Therefore in our case if $\{Z(T_j)\}$ satisfy the convergence criterion given by Equation (3.74), then Berman's theorem can be used to derive the limiting c.d.f.

We now extend the above results to two seasons. For the case of two seasons $(0,t_1]$ and $(t_1,t]$, define

$$V(N_i(t)) = \max\left(\sup_{1\leq j\leq N_i(t_1)} Z(T_j), \sup_{1\leq j\leq N_i(t_1,t)} Z(T_j')\right). \tag{3.80}$$

Then from the independence of the rainfall events between two seasons, it follows from Theorem 2, that

$$\lim_{i\uparrow\infty} P\{u_i^{-1}(V(N_i(t)) - v_i) < x\}$$

$$= [G(x)]^{m(t_1)} [G'(x)]^{m(t_1,t)} \tag{3.81}$$

where $G'(x)$ is that for the sequence $\{Z(T_j')\}$, as $G(x)$ for $\{Z(T_j)\}$ given by Equation (3.74), and $m(t_1,t) = E[N(t_1,t)]$, is the expected number of rainfall events within the season $(t_1,t]$.

3.5.1. Return Period: What Does it Really Mean?

We first give the mathematical formulation behind determination of the actual and the average return periods of an extreme event, and then indicate the role of a limiting c.d.f. within this framework.

Recall from Section 3.2.1, that $N(i) = N(t,i)$ denotes the number of rainfall events within the season $(0,t_1]$, during the i^{th} year, $i \geq 1$. Let $V(N(i))$ denote the maximum cumulative rainfall per event during the i^{th} year, i.e.,

$$V(N(i)) = \sup_{1 \leq j \leq N(i)} Z(T_j); \quad i = 1, 2, \ldots,. \tag{3.82}$$

Now define a positive integer valued random variable K, as

$$K = \min\{i; V(N(i)) > x\}. \tag{3.83}$$

Thus K denotes the number of years it takes to have the first exceedance of the maximum seasonal rainfall per event. From Equation (3.83) it follows that

$$P\{K = k\}$$

$$= P\{V(N(1)) < x, \ldots, V(N(k-1)) < x, V(N(k)) > x\}. \tag{3.84}$$

Classically, the average return period is defined to be $E[K]$. Based on the assumption that $V(N(i))$; $i \geq 1$, are i.i.d. random variables, $E[K]$ is obtained as,

$$E[K] = \sum_{k=1}^{\infty} kP\{K = k\} = 1/P\{V(N(k)) > x\}. \tag{3.85}$$

A serious implication of specifying the average return period as the design criterion is pointed out by Kendall (1959). Based on the work of Kendall, the actual return period is defined as the period during which no exceedance takes place. Let n' denote the actual return period, and α' the probability that no exceedance occurs during n'. $(1-\alpha')$ is also called the risk. Then from Equation (3.84),

$$\alpha' = P\{K > n'\} = \sum_{n=n'+1}^{\infty} P\{K = n\}$$

$$= \sum_{n=n'+1}^{\infty} P\{V(N(1)) < x, \ldots, V(N(n-1)) < x, V(N(n)) > x\}. \tag{3.86}$$

Using the assumption in Equation (3.86), that $V(N(i))$, denoting the maximum cumulative rainfall per event within $(0,t_1)$ during i^{th} year; $i \geq 1$, are i.i.d. as $V(N(1))$,

$$\alpha' = \sum_{n=n'+1}^{\infty} [P\{V(N(1)) < x\}]^{n-1} \cdot P\{V(N(1)) > x\}$$

$$= \sum_{n=n'+1}^{\infty} P\{V(N(1)) < x\}^{n-1} - \sum_{n=n'+1}^{\infty} P\{V(N(1)) < x\}^{n}$$

$$= P\{V(N(1)) < x\}^{n'}, \tag{3.87}$$

Based on Equation (3.87), the actual return period n' can be determined corresponding to a specified level of risk $(1 - \alpha')$, provide the c.d.f. of $V(N(1))$ during a season is known.

Now, in order to point out the danger in using the average return period as the design period, let n' in

Equation (3.87) denote the average return period. Substituting Equation (3.85) into Equation (3.87), the corresponding risk is obtained as,

$$(1 - \alpha') = 1 - (1 - \frac{1}{n'})^{n'}. \tag{3.88}$$

For $n' = 25$, $(1 - \alpha') = .65$. Taking limits on both sides of Equation (3.88) as $n' \to \infty$, $(1 - \alpha') = 1 - e^{-1}$ (Kendall, 1959). In other words, there is a 65% chance of having at least one exceedance of an extreme event, if the design period of twenty-five years of a structure is taken to be the average return period.

It is now clear from Equation (3.87) that the concept of an average return period should be dispensed with in the hydrologic design practices, since it is a misleading and a redundant concept.

In case $P\{V(N(1) < x\}$ is not obtainable, as is the case with us, then limiting c.d.f. can be used to obtain information on the design period of a structure. Such is discussed below.

Consider a sequence of real numbers $\{u_n\}$ and $\{v_n\}$ as given in Equation (3.77), and rewrite Equation (3.86) as,

$$1 - \alpha_n' = P\{V(N(1)) < u_n x + v_n, \ldots, V(N(n)) < u_n x + v_n\}$$

$$= P\{\max_{1 \le i \le n} V(N(i)) < u_n x + v_n\}. \tag{3.89}$$

Substituting Equation (3.71) into Equation (3.89),

$$1 - \alpha_n' = P\{\sup_{1 \le j \le N_n} Z(T_j) < u_n x + v_n\}. \tag{3.90}$$

Taking limits as $n \uparrow \infty$, on both sides of Equation (3.90)

$$\lim_{n \uparrow \infty} (1 - \alpha_n') = \lim_{n \uparrow \infty} P\{\sup_{1 \le j \le N_n} Z(T_j) < u_n x + v_n\}. \tag{3.91}$$

But the right hand side of Equation (3.91) is the limiting distribution as given by Equation (3.79). Therefore substituting Equation (3.79) and (3.10) into Equation (3.91),

$$1 - \alpha'' = \lim_{n \uparrow \infty} (1 - \alpha_n') = [G(x)]^{m(t_1)}. \tag{3.92}$$

The left hand side denotes the risk corresponding to the worst possible event that can ever occur. Thus using Equation (3.92), one can compute the value of the worst possible rainfall per event corresponding to a specified value of the risk. However in order to compute a design rainfall corresponding to an actual finite return period, Equation (3.91) can be used to give an approximate result, provided a large value of the actual return period is desired and some value of the risk is specified.

With this, we conclude the discussion on general theoretical considerations. The next chapter considers some specific theoretical applications of the general theory given here.

CHAPTER 4

THEORETICAL APPLICATIONS

4.1. Introduction

This chapter is devoted to a few specific theoretical applications of the general theoretical considerations given in Chapter 3. The treatment given here is in the spirit of obtaining analytical solutions.

In particular, the c.d.f.'s of the random variables $Z(T_1)$ and $Z'(T_1)$, denoting the cumulative rainfall and the excess water yield per event respectively, are derived under a set of hypotheses and assumptions. Specifically, the derivation is based on the following hypotheses;

(1) the stochastic field of the distribution of number of storm cells over any sub-basin within the region, given the duration of a storm (rainfall event), is a Poisson field (the word 'field' is in reference to identifying a process in two dimensional parametric space);

(2) the maximum rainfall depth within a cell (cell center) as a random variable follows a two parameter gamma p.d.f.;

(3) the storm duration denoted by C_1, follows an exponential p.d.f. The derivation is firstly given with

reference to a single sub-basin within the region, and then generalized to multiple basins within the region. A schematic extension of these results to an ungaged basin within the region is also indicated.

Based on the above derivations for cumulative rainfall per event, an approximate expression is obtained for the average cumulative rainfall within a season. Finally, an expression is obtained for the limiting distribution of the maximum cumulative rainfall per event, using the results given in Section 3.5 of Chapter 3.

Our intent in this chapter is mainly demonstrative. However, many hypotheses are given a phenomenological interpretation. Finally, all possible applications of the theoretical considerations given in Chapter 3 are not considered here, but only the important ones.

4.2. Stochastic Process of Spatial Rainfall

Recall from Section 3.2.2 of Chapter 3, that the c.d.f. of random variable $Z(C_1)$, denoting cumulative rainfall per event with duration C_1, may be obtained by combining Equations (3.14) and (3.19) as

$$P\{Z(C_1) < z\} = \int_0^\infty P\{Z(c) < z | C_1 = c\} dF(c)$$

$$= \int_0^\infty \sum_{k=1}^\infty P\{\sum_{n=1}^k Y_n < z, M(B,c) = k | C_1 = c\} dF(c)$$

$$+ \int_0^\infty P\{M(B,c) = 0|C_1 = c\}dF(c), \qquad (4.1)$$

where $M(B,C_1)$; $B \subset R$, denotes the number of 'cell centers' within some basin B contained in the region R, and Y_n denotes the spatial rainfall from the n^{th} cell, $n \geq 1$. Moreover, in Section 3.2.2 of Chapter 3, we referred to $M(B,C_1)$ as denoting the number of cells instead of cell centers, based on the assumption that each cell has only one cell center (point of maximum rainfall). In the subsequent discussion also, we will refer to $M(B,C_1)$ as denoting the number of cells. Within our formulation, Equation (4.1) represents the most general expression for obtaining the c.d.f. of $Z(C_1)$.

To this end, assume that $M(B,C_1)$ and Y_n, for all $n \geq 1$, are conditionally independent, where the conditioning random variable C_1 denotes the duration of a rainfall event. Phenomenologically speaking, this assumption means that given the duration of a rainfall event, the number of storm cells that occur over the basins and the corresponding rainfall amounts from each cell, are mutually independent. Based on this assumption, Equation (4.1) can be rewritten as,

$$P\{Z(C_1) < z\}$$

$$= \int_0^\infty \sum_{k=1}^\infty P\{\sum_{n=1}^k Y_n < z|C_1 = c\}P\{M(B,c) = k|C_1 = c\}dF(c)$$

$$+ \int_0^\infty P\{M(B,c) = 0 | C_1 = c\} dF(c). \tag{4.2}$$

In order to determine the c.d.f. $Z(C_1)$, as given by Equation (4.2), the following probabilities must be determined, namely,

(1) $P\{M(B,c) = k | C_1 = c\}$, $k = 0, 1, \ldots,$;

(2) $P\{\sum_{n=1}^{k} Y_n < z | C_1 = c\}$ and

(3) $F(c) = P\{C_1 < c\}$.

Such is considered in the ensuing sub-sections.

4.2.1. Random Field of the Number of Storm Cells

The conditional probability $P\{M(B,c) = k | C_1 = c\}$, $k = 0, 1, 2, \ldots,$ gives the p.m.f. of number of storm cells over the basin B, given that the duration of a storm over the region is c. Recall that the random duration C_1 is defined <u>with respect to the entire region</u> $R \supset B_i$, $1 \leq i \leq r$, and conditionally in the random field $\{M(B_i,C_1); B_i \subset R, 1 \leq i \leq r\}$, the duration behaves as a fixed parameter.

More precisely, let $\mu(B_i,c)$ be a finite measure for all $B_i \subset R$, $1 \leq i \leq r$, i.e., $\mu(B_i,c) < \infty$, and c is some positive real number. Recall from Section 3.2.2 that $B_1, \ldots, B_r$ represent the toral number of mutually disjoint (including contiguous) sub-basins within the region R and $B = \bigcup_{i=1}^{r_1} B_i$, $r_1 \leq r$. Hydrologically, these sub-basins can be conceived of as divisions of a single basin, or multiple basins with a common drainage outlet or multiple basins with multiple outlets. Such is shown schematically in Figure 4.1. Now, following Takács (1956), assume the following, which will be given a phenomenological interpretation afterwards,

(i) $P\{M(B_i,c) = 0|C_1 = c\} \neq 1$, if $\mu(B_i,c) > 0$ for any $B_i \subset R$, and,

$$\sum_{k=0}^{\infty} P\{M(B_i,c) = k|C_1 = c\} = 1, \quad k = 0, 1, 2, \ldots,.$$

(ii) The conditional probability $P\{M(B_i,c) = k|C_1 = c\}$ depends only on the measure $\mu(B_i,c)$ for all $B_i \subset R$, $k = 0, 1, \ldots,.$

(iii) If B_1 and $B_2 \subset R$ and $B_1 \cap B_2 = \emptyset$ (Null Set), then

$$P\{M(B_1 \cup B_2,c) = k|C_1 = c\}$$

$$= P\{M(B_1,c) + M(B_2,c) = k|C_1 = c\},$$

$$= \sum_{j=0}^{k} P\{M(B_1,c) = j|C_1 = c\}P\{M(B_2,c) = k - j|C_1 = c\},$$

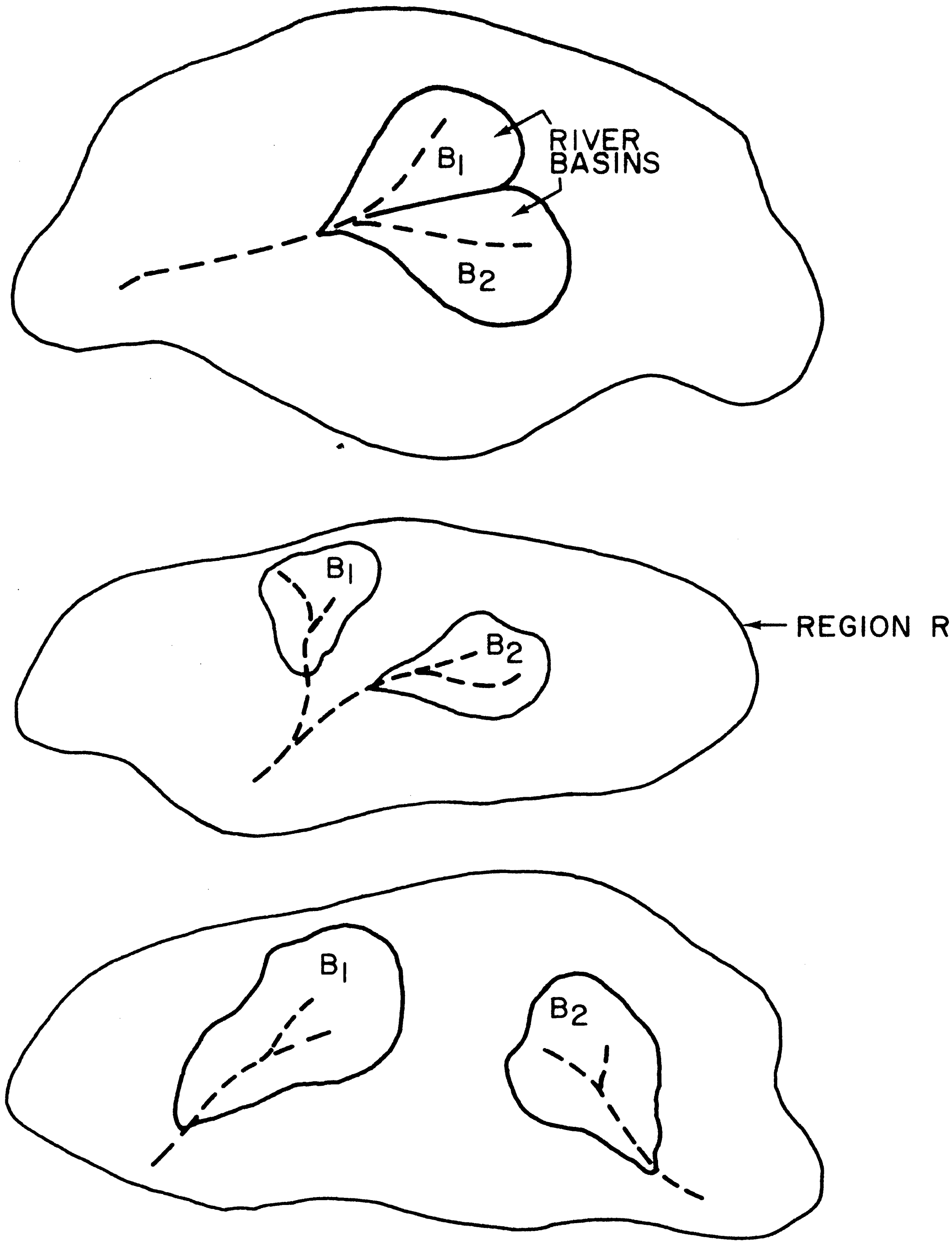


Figure 4.1. A Schematic View of three Different Configurations of River Basins Within a Region.

i.e., the random variables $M(B_1,C_1)$ and $M(B_2,C_1)$ are conditionally independent.

(iv)
$$\lim_{\mu(B_i,c)\to 0} \frac{P\{M(B_i,c)\geq 1|C_1=c\}}{P\{M(B_i,c)=1|C_1=c\}} = 1; \quad i \leq i \leq r.$$

Under the above assumptions (i) to (iv), the conditional distribution of the random variable $M(B_i,C_1)$ for every $B_i \subset R$, $1 \leq i \leq r$, follows a Poisson p.m.f. (Takács, 1956). Mathematically this can be expressed as,

$$P\{M(B_i,c) = k|C_1 = c\} = \exp(-\mu(B_i,c))[\mu(B_i,c)]^k/k!;$$

$$k = 0, 1, 2, \ldots, \quad B_i \subset R; \quad 1 \leq i \leq r. \tag{4.3}$$

Phenomenologically, the above assumptions (i) to (iv) giving rise to a Poisson p.m.f., can be explained as follows.

The finite measure $\mu(B_i,c)$, denotes the conditional expectation of the number of storm cells over the sub-basin B_i, $1 \leq i \leq r$. Now, according to the first assumption, given that the conditional expectation $\mu(B_i,c) > 0$ for any $B_i \subset R$, there is a positive probability less than 1, that one or more cells can occur over the sub-basin. The second assumption implies that the probability of occurrence of cells over any sub-basin B_i; $1 \leq i \leq r$, depends only on the measure $\mu(B_i,c)$ defined over that sub-basin. For example, it may depend on the area of the sub-basin. The third assumption states that the random variables denoting the number of cells, over disjoint sub-basins are conditionally independent. This assumption may be more justifiable for sub-basins that are 'far apart'. This can be argued according to the nature of spatial evolution of these cells. For example, the cells occurring over two disjoint sub-basins from within two different small meso-scale areas (SMSA) as mentioned in Section 2.2.1 of Chapter 2, may be assumed to be independent. However, if the cells occur over two disjoint sub-basins from within the same SMSA, then there may be a dependence in the number of cells, that occur over each of these two sub-basins. Nonetheless, as a first approximation, we accept the validity of the third assumption. Finally, according to the fourth assumption, as the expected number of cells over a basin becomes small, the probability of occurrences of more than one cell also gets small. In other words, over very small sub-basins, two or more cells cannot occur simultaneously. This assumption may be violated by fusion of two or more cells. However, such considerations can complicate our analysis considerably; yet two or more cells fused together can be viewed as one cell only.

The second assumption given above is very important phenomenologically, in the sense that non-homogenity within the region R can be incompassed in the theory. Such non-homogenity can arise due to elevation differences between the sub-basins within the region, which in turn can influence the p.m.f. of the occurrence of number of cells over these sub-basins. Influences of this kind have been observed empirically (Duckstein, Fogel and Thames, 1973; Grayman and Eagleson, 1971). Presently, we assume a simple form for the measure $\mu(B_i, c)$, but still introduce non-homogenity between the sub-basins $B_1, \ldots, B_r$, as given below, namely,

$$\mu(B_i, c) = \alpha_i c \nu(B_i), \quad 1 \leq i \leq r \tag{4.4}$$

where $\alpha_i \neq \alpha_j$; $i \neq j$, $1 \leq i$, $j \leq r$, are positive real numbers and $\nu(B_i)$ is the area of the sub-basin B_i, $1 \leq i \leq r$. α_i can be interpreted as the rate of occurrence of cells per unit area per unit time, within the sub-basin B_i, $1 \leq i \leq r$. Different basins can have different rates of the occurrence of cells, due to non-homogenity within the region R. However, the simple form of measure as given by Equation (4.4) is assumed only for the demonstrative purposes. More general type of measures can be assigned as well, if empirical evidence warrants it.

This concludes the discussion on the random field of the number of storm cells. Next, the spatial distribution of rainfall from individual cells is considered.

4.2.2. Cumulative Distribution Function of the Spatial Rainfall from Storm Cells

The conditional probability of the total spatial rainfall from k cells; $k \geq 1$, with respect to the storm duration is given as $P\{\sum_{n=1}^{k} Y_n < z | C_1 = c\}$, $z > 0$. Let us assume that (i) the random variables Y_n; $n \geq 1$, are i.i.d. and (ii) Y_n, $n \geq 1$, are independent of storm duration. Phenomenologically, the first assumption may not be very sound, especially for the frontal or squall line type of storms, where multiple cells occur within a SMSA. For such situations the successive cells that develop in time, are influenced by the cells preceding them in regard to the moisture content (Byers and Braham, 1949). On the other hand, for the air-mass type of convective rainfall in arid regions, cells occur more or less in isolation and therefore the first assumption may be more reasonable for the air-mass type of thunderstorm cells. The second assumption seems reasonable as a first approximation, but strictly speaking the dependence between Y_n; $n \geq 1$, and the storm duration C_1 does arise through the random durations of individual cells. In this light the independence can be attributed to the fact that Y_n, depends only on the path of the sample function $\xi(u,v,s) \geq 0$, and not on the random duration of the n^{th} cell denoted by (W_{2n-1}, W_{2n}); $n \geq 1$, (see Section 3.2.2, Chapter 3). Phenomenologically, such would be true, when most of the rainfall from a cell occurs due to high rainfall intensities within a time interval much smaller than the total cell duration, as is the case with summer rainfall in the Southwestern United States.

Recall from Section 3.2.2 of Chapter 3, that Y_n; $n \geq 1$ is given by Equation (3.18) as,

$$Y_n = \int_{W_{2n-1}}^{W_{2n}} \int_{B_n'}\int \xi(u,v,s)\,dudvds, \tag{4.5}$$

where $B_n' = \{(u,v);\xi(u,v,s) > 0,$ for all $s \in (W_{2n-1},W_{2n})\}$ Thus B_n' for $n = 1, 2, \ldots,$ represents the 'areal extent' of a cell. Define a new family of random variables given as,

$$\{\xi_n'(u,v) > 0; (u,v) \in B_n', B_n' \subset R\}, \tag{4.6}$$

where,

$$\xi_n'(u,v) = \int_{W_{2n-1}}^{W_{2n}} \xi(u,v,s)\,ds; \quad n \geq 1. \tag{4.7}$$

Clearly, the random variable $\xi_n'(u,v) > 0$, represents the cumulative rainfall (with respect to time) within a cell for any point in space $(u,v) \in B_n'$, and the family of random variables given by Equation (4.6) represents a random field for all $n \geq 1$. Substituting Equation (4.7) into Equation (4.6),

$$Y_n = \int_{B_n'}\int \xi_n'(u,v)\,du\,dv. \tag{4.8}$$

Now, the determination of the c.d.f. of Y_n; $n \geq 1$, involves hypothesizing some model for the random field given by Equation (4.6).

The work done in the past in regard to modeling of the random field given by Equation (4.6), falls in the realm of depth-area relationships. However, most of the work along these lines does not have a phenomenological orientation in the sense, that such relationships are not given for the individual cells. A few exceptions to this are the models proposed by Fogel and Duckstein (1969), Woolhiser and Schwalen (1959) and Osborn (1970), but these models only deal with the air-mass type of convective rainfall cells. For our purpose, any one of such models can be used, as far as they characterize the spatial rainfall distribution within a cell.

The basic idea behind such models is that the random variable $\xi_n'(u,v)$ at every $(u,v) \in B_n'$ is expressed as some deterministic function of the maximum rainfall depth denoted by the random variable $\xi_n(u_0,v_0)$, $(u_0,v_0) \in B_n'$. These models seem to represent an over simplified view of the spatial rainfall. This is mainly due to the fact that these relationships are purely empirical and are heavily conditioned by the type of rainfall data that is generally available to form these empirical relationships (Davis, Kisiel and Duckstein, 1972).

In the absence of any more empirical evidence, on the phenomenological behavior of the cumulative spatial rainfall from a cell, other than the one given above, we assume that <u>any one</u> of the depth-area relationships is a reasonable model for describing the random field given by Equation (4.6). Let $g(u,v)$ for all $(u,v) \varepsilon B_n'$; $n \geq 1$, be a real valued deterministic function, and $\xi_n(u_0,v_0)$; $n \geq 1$, be the center rainfall depth. Then the random field given by Equation (4.6) can be expressed as,

$$\{\xi_n(u,v) \overset{n}{\equiv} \xi_n(u_0,v_0)g(u,v);(u,v) \ \varepsilon \ B_n'\}; \quad n \geq 1. \tag{4.9}$$

Substituting Equation (4.9) into Equation (4.8), the cumulative rainfall per cell denoted by Y_n is given as,

$$Y_n = \int_{B_n'}\!\!\int \xi_n(u_0,v_0)g(u,v)dudv$$

$$= \xi_n(u_0,v_0)\int_{B_n'}\!\!\int g(u,v)dudv; \quad n \geq 1. \tag{4.10}$$

Now, using the assumption made earlier in this section, that Y_n; $n \geq 1$, are i.i.d. random variables, we conclude from Equation (4.10) that $\xi_n(u_0,v_0)$; $n \geq 1$, are i.i.d. as, say, ξ_0, and $B'_n \overset{n}{\equiv} B'$. Recall that B'_n; $n \geq 1$, denotes the areal extent of a cell, and the equivalence relation implies that the areal extent of all cells is the same. This assumption is not unreasonable, because empirical evidence to this effect exists in the literature for at least air-mass type of connective rainfall (Osborn, 1970).

In view of the above, Equation (4.10) can be written as,

$$Y_n = \xi_0 g'(\nu(B')); \quad n \geq 1$$

where, (4.11)

$$g'(\nu(B')) = \int_{B'}\int g(u,v)\,du\,dv \; .$$

Note that $\nu(B')$ denotes the measure of the set B', i.e., the area covered by a cell.

Based on the form of empirical functions $g(u,v)$ proposed earlier, for example see Fogel and Duckstein (1969), the cumulative spatial rainfall from a cell would appear something similar to given in figure 4.2. However from Equation (4.11) it is clear that under our formulation, a complex form of function $g(u,v)$ would not add to the generality, but can only make its integration over B' impossible in a closed form. Therefore for demonstrative purpose, we assume

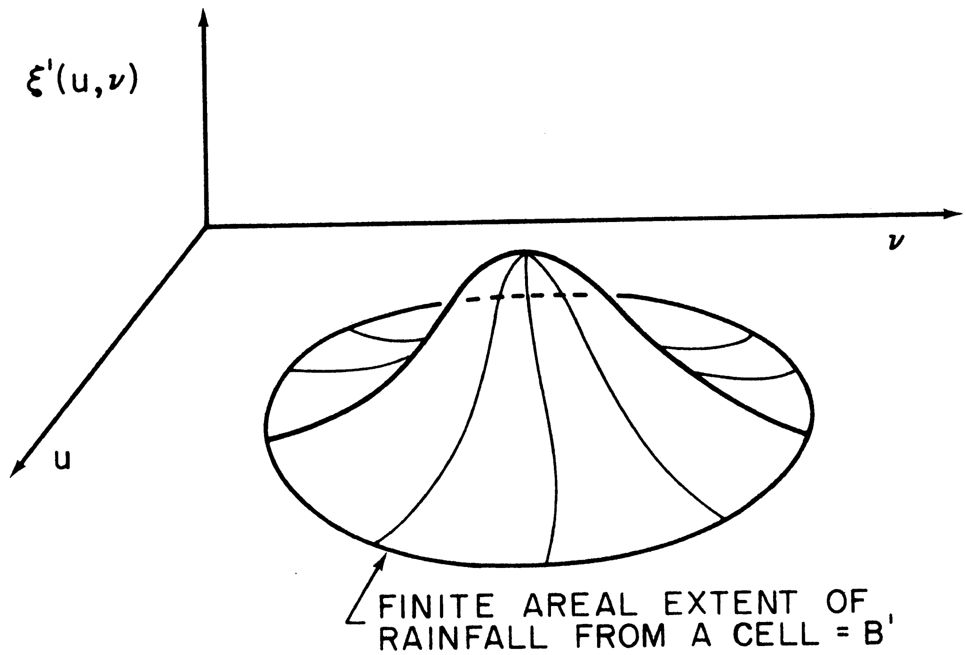


Figure 4.2. A Conceptualization of the Distribution of the Cumulative Spatial Rainfall in a Cell.

a linear functional relation as schematically shown in Figure 4.3, which makes the spatial rainfall from a cell appear as a symmetric cone. However, empirical relation of this kind have been proposed in the past by Osborn (1970). Assume that the base of the cone is circular with radius r_c. The assumption of circularity is not necessary, but is made only to simplify computations. Now the cumulative spatial rainfall per cell Y_n; $n \geq 1$, is simply the volume of a cone with radius r_c and height ξ_0, which is given by Equation (4.11) as,

$$Y_n = \xi_0 \pi r_c^2/3; \quad n \geq 1. \tag{4.12}$$

Equation (4.12) gives an expression for the total cumulative spatial rainfall from a cell. However in actuality, we are only interested in that <u>fraction</u> of the total cumulative rainfall per cell, that falls over a sub-basin under consideration. The reason for introducing a 'fraction' is clear from the fact that depending on the <u>location of</u> the cell center in relation to the sub-basin, the total cumulative rainfall from a cell does not always occur over the sub-basin. Thus the proportion or fraction of the total cumulative rainfall per cell that falls over the basin can be assumed to be a bounded random variable, denoted by Q. The upper and lower bounds of 1 and 0 respectively, on the

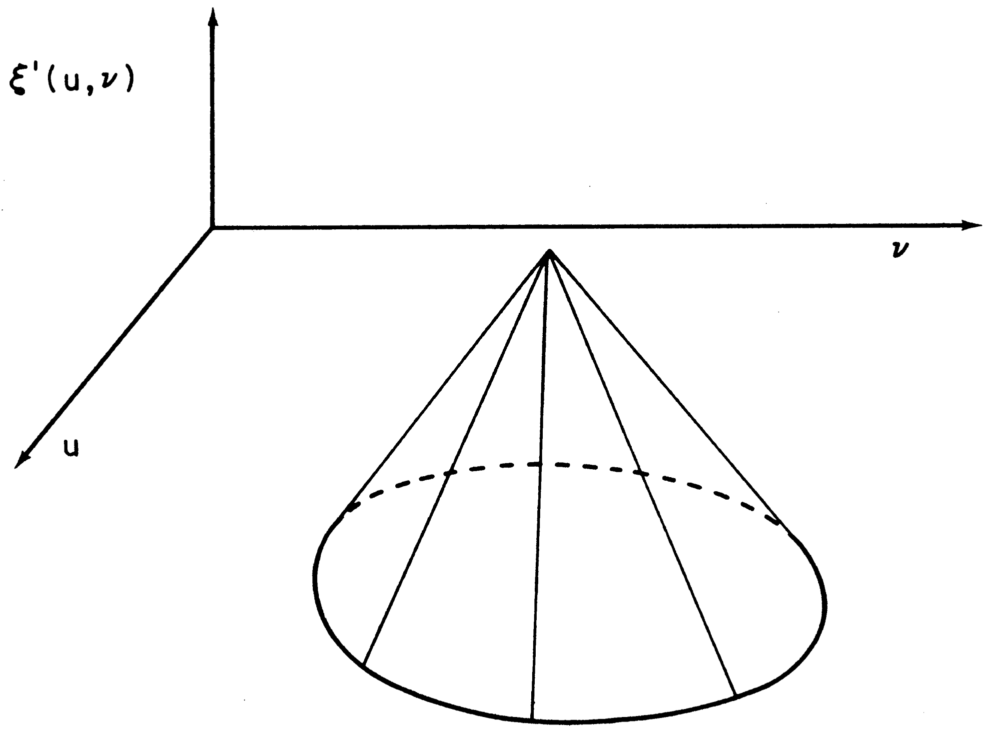


Figure 4.3. A Conceptualization of the Distribution of the Cumulative Spatial Rainfall in a Cell, as Assumed in This Study.

random variable Q are obvious from the fact, that either the entire cumulative rainfall occurs over the basin or none occurs at all. The randomization is schematically shown in Figure 4.4. Now, if the random variables Y_n; $n \geq 1$, denote the cumulative rainfall per cell that can occur over a sub-basin B_i; $1 \leq i \leq r$, then in view of Equation (4.12) and the above assumption, Y_n is given by,

$$Y_n = \xi_0 \pi r_c^2 Q/3; \quad n \geq 1. \tag{4.13}$$

Note that Equations (4.12) and (4.13) represent two different sequences of random variables, but we don't indicate them by separate notations, since no confusion seems possible. In the subsequent discussion, any reference to Y_n would only imply Equation (4.13), unless stated otherwise.

To this end, assume that ξ_0 follows a two parameter gamma p.d.f., denoted by $f(y)$ and for every $y > 0$, given as,

$$f(y) = \lambda^{n_0} y^{n_0-1} e^{-\lambda y}/(n_0-1)!; \quad \lambda > 0, \quad n_0 = 1,2, \ldots, \tag{4.14}$$

where λ and n_0 are the parameters of the gamma p.d.f. Our basis for assuming a gamma p.d.f. is purely empirical (see Chapter 5), without any phenomenological justification.

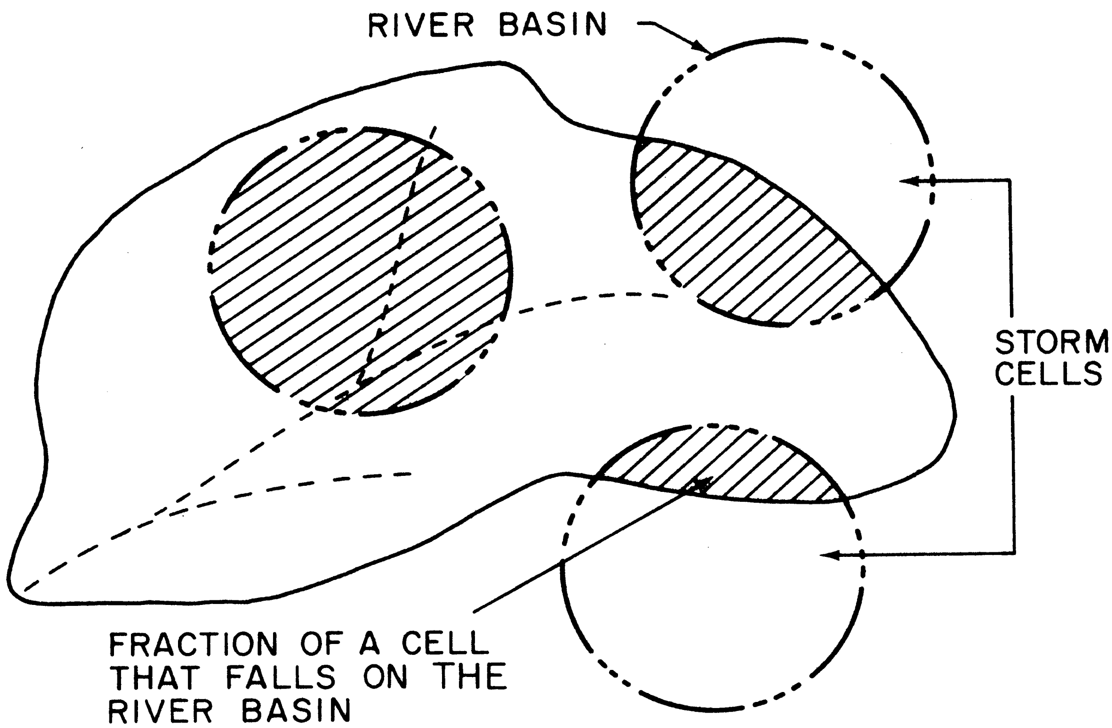


Figure 4.4. A Schematic Representation of Randomness in the Cumulative Rainfall per Cell over a Basin Due to Cell Location.

However, a gamma p.d.f. does have a good amount of flexibility in the sense, that for different values of λ and n_0, it can give a wide variety of shapes.

In regard to the random variable Q, assume that Q is uniformly distributed betwen 0 and 1, which is a special case of a beta density function. A phenomenological reasoning can be given to this effect. In case the basin under consideration is much larger in comparison with the areal extent of a cell, then a beta density skewed towards the upper bound would be more suitable, since a large percentage of cells would tend to fall completely within the basin. On the other hand, if a basin is comparable to the areal extent of a cell, then a uniform p.d.f. may seem more reasonable. However our assumption of a uniform p.d.f. is more for achieving analytical and computational simplification. The p.d.f. of Q, denoted by $f(q)$, is given as,

$$\left.\begin{aligned} f(q) &= 1 \quad \text{for} \quad 0 \leq q \leq 1 \\ &= 0 \quad \text{otherwise} \end{aligned}\right\}. \tag{4.15}$$

Now assuming that ξ_0 and Q are mutually independent, and substituting Equations (4.14) and (4.15) into Equation (4.13), the c.d.f. of Y_n; $n \geq 1$, can be obtained as

$$P\{Y_n < y\} = \int_0^1 P\{\tfrac{1}{3}\xi_0 q\pi r_c^2 < y | Q = q\} f(q)dq \tag{4.16}$$

$$= \int_0^1 P\{\xi_0 < (3y/\pi q r_c^2)\} dq$$

$$= \int_0^1 [1 - e^{-\lambda y'/q} (\sum_{j=0}^{n_0-1} (\lambda y'/q)^j / j!)] dq; \quad n \geq 1 \qquad (4.17)$$

where $y' = 3y/\pi r_c^2 \geq 0$.

The c.d.f. of $\sum_{n=1}^{k} Y_n$, which is our goal as indicated in the beginning of this section, can be obtained using the following fact. Consider k i.i.d. random variables, each gamma distributed with parameters λ and n_0. Then their sum is also a gamma with parameters λ and n_0k (Parzen, 1967). Using this fact and Equation (4.17), for any $y > 0$, the c.d.f. of $\sum_{n=1}^{k} Y_n$; $K = 1, 2, \ldots,$ is given by,

$$P\{\sum_{n=1}^{k} Y_n < y\} = \int_0^1 [1 - e^{-\lambda y'/q} (\sum_{j=0}^{n_0k-1} (\lambda y'/q)^j / j!)] dq. \qquad (4.18)$$

Equation (4.18) can be rewritten as,

$$P\{\sum_{n=1}^{k} Y_n < y\} = \int_0^1 [\int_0^{\lambda y'/q} e^{-w} w^{n_0k-1} dw | (n_0k-1)!] dq \qquad (4.19)$$

where $y' = 3y/\pi r_c^2$ as given before.

This result concludes this sub-section on obtaining the c.d.f. of the cumulative spatial rainfall. In the next section, the formulations given in this section and in

Section 4.2.1, are used to obtain the c.d.f. of cumulative rainfall per event.

4.3. Cumulative Distribution Functions of Cumulative Rainfall and Water Yield from an Event

We can use Equation (4.2) to obtain the c.d.f. of the cumulative rainfall per event. Then the c.d.f. of the excess water yield can be obtained by substituting Equation (4.2) into Equation (3.22), as mentioned in Section 3.2.2 of Chapter 3.

Recall from Section 3.2.2 that the basin $B = \bigcup_{i=1}^{r_1} B_i$; $1 \leq r_1 \leq r$. Assume that $r_1 = 1$. Then $B = B_1$, i.e., we first consider only one sub-basin B_1 contained in the region R. Then using Equation (3.2), the c.d.f. of the cumulative rainfall can be derived for one sub-basin. The result is then generalized to multiple basins, i.e., when $B = \bigcup_{i=1}^{r_1} B_i$, $2 < r_1 \leq r$.

To this end, we start by explicitly identifying the three c.d.f.'s that consititue Equation (4.2). Assume that the random duration of a rainfall event, denoted by C_1, has an exponential p.d.f. Phenomenologically speaking, such assumption seems reasonable for the air-mass type of convective storms over a small region, say up to 200 mi^2 or so, because the number of multiple cells that can occur within the random duration is small and each cell is of a short duration say up to 15-20 minutes or so. An empirical verification of this assumption is given in Chapter 5, for the Atterbury watershed in the Southwestern United States. On the other hand this assumption does not seem reasonable for the durations of storms associated with moving fronts, since such storms 'persist' for a relatively long time. In any case, the results given below can always be modified if C_1 does not follow an exponential p.d.f. Now in view of this assumption, the exponential p.d.f. of C_1 with parameter $\beta_1 > 0$, can be expressed as,

$$f(c) = \beta_1 e^{-\beta_1 c}; \quad c \geq 0. \tag{4.20}$$

Next, we consider the Poisson p.m.f., which gives the conditional (on C_1) distribution of the number of storm cells over the basins B_i; $1 \leq i \leq r$. Let $\alpha_1 \nu(B_1) = \alpha_1'$ in Equation (4.4). Then replacing $\alpha_1 \nu(B_1)$ by α_1' in Equation (4.3), the conditional distribution of the number of cells that occur over B_1, is given by,

$$P\{M(B_1,c) = k | C_1 = c\} = e^{-\alpha_1' c}(\alpha_1' c)^k / k!; \quad k \geq 0. \tag{4.21}$$

Finally, the c.d.f. of cumulative rainfall from k cells, $k = 0, 1, \ldots,$ can be obtained from Equation (4.19).

Based on the assumption made in Section 4.2, that $\{Y_n\}$, denoting spatial rainfall per cell, are independent of the storm duration C_1, and replacing B by B_1 in Equation (4.2), it modifies to

$$P\{Z(C_1) < z\} = \int_0^{\infty} P\{M(B_1,c) = 0|C_1 = c\}dF(c)$$

$$+ \int_0^{\infty} \sum_{k=1}^{\infty} P\{\sum_{n=1}^{k} Y_n < z\}P\{M(B_1,c) = K|C_1 = c\}dF(c). \tag{4.22}$$

Substituting Equations (4.19), (4.20) and (4.21) into Equation (4.22), and denoting $z' = 3z|\pi r_c^2$, Equation (4.22) is written as,

$$P\{Z(C_1) < z\} = \int_0^{\infty} \beta_1 e^{-(\alpha_1'+\beta_1)c} dc$$

$$+ \int_0^{\infty} \sum_{k=1}^{\infty} \int_0^1 \int_0^{z'\lambda/q} (e^{-w} w^{n_0 k-1} dwdq/(n_0 k-1)!)$$

$$\cdot (e^{-\alpha_1' c}(\alpha_1' c)^k/k!)\beta_1 e^{-\beta_1 c} dc. \tag{4.23}$$

Equation (4.23) may be transformed into,

$$P\{Z(C_1) < z\} = \frac{\beta_1}{\alpha_1'+\beta_1} + \int_0^1 \int_0^{z'\lambda/q} [\sum_{k=1}^{\infty} (e^{-w} w^{n_0 k-1} dw/(n_0 k-1)!)$$

$$\cdot (\alpha_1')^k \int_0^{\infty} c^k e^{-(\alpha_1'+\beta_1)c} dc/k!]dq$$

$$= \frac{\beta_1}{\alpha_1'+\beta_1}[1 + \int_0^1\int_0^{z'\lambda/q} \sum_{k=1}^{\infty} (e^{-w} w^{n_0 k-1} dw/(n_0k-1)!)(\frac{\alpha_1'}{\alpha_1'+\beta_1})^k dq] \tag{4.24}$$

Equation (4.24) can be rewritten as,

$$= \frac{\beta_1}{\alpha_1'+\beta_1}[1 + \int_0^1\int_0^{z'\lambda/q} e^{-w}\{\sum_{k=1}^{\infty} [w^{n_0 k-1}/(n_0k-1)!][\frac{\alpha_1'}{\alpha_1'+\beta_1}]^k\}]dwdq \tag{4.25}$$

The infinite series in Equation (4.25) can be summed up as below.

$$\sum_{k=1}^{\infty} [w^{n_0 k-1}/(n_0k-1)!][\frac{\alpha_1'}{\alpha_1'+\beta_1}]^k$$

$$= \theta \sum_{k=1}^{\infty} (w\theta)^{n_0 k-1}/(n_0k-1)! \tag{4.26}$$

where,

$$\theta^{n_0} = \frac{\alpha_1'}{\alpha_1'+\beta_1}, \quad \text{or} \quad \theta = (\frac{\alpha_1'}{\beta_1'+\alpha_1})^{1/n_0}. \tag{4.27}$$

The sum of the infinite series occurring on the right hand side of Equation (4.26) can be expressed as follows (Parzen, 1967, p. 175),

$$\sum_{k=1}^{\infty} (w\theta)^{n_0 k-1}/(n_0k-1)! = \frac{1}{n_0}\sum_{j=0}^{n_0-1} \delta^j \exp(w\theta\delta^j) \tag{4.28}$$

where,

$$\delta^j = \exp(2\pi ij/n_0); \quad i = \sqrt{-1}, \quad 0 \leq j \leq n_0. \tag{4.29}$$

Substituting Equation (4.28) into Equation (4.25),

$P\{Z(C_1) < z\}$

$$= \frac{\beta_1}{\alpha_1'+\beta_1}[1 + \int_0^1\int_0^{z'\lambda/q}(e^{-w}/n_0)\sum_{j=0}^{n_0-1}\delta^j\exp(w\theta\delta^j)dwdq]$$

$$= \frac{\beta_1}{\alpha_1'+\beta_1}\{1 + (\theta/n_0)\int_0^1\sum_{j=0}^{n_0-1}\delta^j\int_0^{z'\lambda/q}\exp[-w(1-\theta\delta^j)dwdq\}$$

$$= \frac{\beta_1}{\alpha_1'+\beta_1}\{1 + (\theta/n_0)\int_0^1\sum_{j=0}^{n_0-1}\frac{\delta^j}{1-\theta\delta^j}[1 - e^{-\lambda z'(1-\theta\delta^j)/q}]dq$$

$$= \frac{\beta_1}{\alpha_1'+\beta_1}\{1 + (1/n_0)\sum_{j=0}^{n_0-1}\frac{\theta\delta^j}{1-\theta\delta^j}[1 - \int_0^1 e^{-\ z'(1-\theta\delta^j)/q}dq]\}. \tag{4.30}$$

The integral in Equation (4.30) cannot be obtained in an analytical form. In regard to the behavior of this integral and its possible simplification in the form of an infinite series, see Bromwich (1926, p. 336). However, Equation (4.30) can be expressed more conveniently in view of the following simplification, namely,

$$1 + (1/n_0)\sum_{j=0}^{n_0-1} \theta\delta^j/(1-\theta\delta^j) = (1/n_0)\sum_{j=0}^{n_0-1} 1/(1-\theta\delta^j). \qquad (4.31)$$

Where the expression within the summation sign on the right hand side of Equation (4.31) is the sum of an infinite series for complex numbers, since $|\theta\delta^j| < 1$, (see Churchill, 1960, p. 132). Hence the right hand side in Equation (4.31) can be written as,

$$(1/n_0)\sum_{j=0}^{n_0-1} 1/(1-\theta\delta^j) = (1/n_0)\sum_{j=0}^{n_0-1}\sum_{k=0}^{\infty}(\theta\delta^j)^k$$

$$= \sum_{k=0}^{\infty}(\theta^k/n_0)\sum_{j=0}^{n_0-1}(\delta^j)^k. \qquad (4.32)$$

Now consider the following identity (Parzen, 1967, p. 176), namely,

$$\left.\begin{aligned} &(1/n_0)\sum_{j=0}^{n_0-1}(\delta^j)^k = 1 \quad \text{if } k \text{ is a multiple of } n_0 \\ &(1/n_0)\sum_{j=0}^{n_0-1}(\delta^j)^k = 0 \quad \text{otherwise} \end{aligned}\right\}. \qquad (4.33)$$

Substituting Equation (4.33) into Equation (4.32),

$$(1/n_0)\sum_{j=0}^{n_0-1} 1/(1-\theta\delta^j) = \sum_{k=0}^{\infty}\theta^{kn_0} = 1/(1-\theta^{n_0})$$

$$= (\alpha_1'+\beta_1)/\beta_1. \qquad (4.34)$$

Since from Equation (4.27), $\theta^{n_0} = \alpha_1'(\alpha_1'+\beta_1) < 1$. Substituting Equation (4.34) into Equation (4.30), the c.d.f. of the cumulative rainfall per event becomes,

$$P\{Z(C_1 < z\}$$

$$= 1 - \frac{\beta_1}{(\alpha_1'+\beta_1)n_0} \sum_{j=0}^{n_0-1} \frac{\theta\delta^j}{1-\theta\delta^j} \int_0^1 e^{-\lambda z'(1-\theta\delta^j)/q} dq. \qquad (4.35)$$

Equation (4.35) represents a 'proper' c.d.f. in the sense that,

$$\lim_{z'\to\infty} P\{Z(C_1) < z\} = 1. \qquad (4.36)$$

Moreover taking the limit as $z' \to 0$, in Equation (4.35) it follows from Equation (4.34), that

$$\lim_{z'\to 0} P\{Z(C_1) < z\} = \beta_1/(\alpha_1'+\beta_1). \qquad (4.37)$$

From Equation (4.37), it is clear that the c.d.f. of $Z(C_1)$ is not differentiable at the origin, hence the c.d.f. of $Z(C_1)$ does not have a p.d.f. Phenomenological significance to the existence of this 'atom' at the origin of the c.d.f. of $Z(C_1)$, in the context of extension to ungaged basins, will be given later.

The c.d.f. of the cumulative rainfall for one sub-basin as given by Equation (4.35), can now be generalized to <u>multiple basins</u>. In order to accomplish this generalization, we first indicate a very important property of the Poisson field giving the p.m.f. of number of cells within our region R. The generalization then follows very simply from this property.

Recall that r is the total number of sub-basins within our region R. Let $1 < r_1 \leq r$ be an integer, such that the basin B is comprised of the union of $B_1, \ldots, B_{r_1}$, i.e., $B = \bigcup_{i=1}^{r_1} B_i$. Assume that $B_1, \ldots, B_{r_1}$ are mutually disjoint (including contiguous), i.e., $B_j \cap B_k = \emptyset$, $j \neq k$. Therefore from the third assumption on page 86, it follows that the random variables $M(B_i, C_1)$ and $M(B_j, C_1)$, $i \neq j$, are conditionally independent for $1 \leq i$, $j \leq r_1$. Hence based on Equation (4.3), it can be shown that the conditional p.m.f. of the random variable $M(\bigcup_{j=1}^{r_1} B_j, C_1)$ has a Poisson distribution with the conditional expectation $\mu(\bigcup_{j=1}^{r_1} B_j, c)$ given by,

$$\mu(\bigcup_{j=1}^{r_1} B_j, c) = E[M(\bigcup_{j=1}^{r_1} B_j, c) | C_1 = c] = E[\sum_{j=1}^{r_1} M(B_j, c) | C_1 = c]$$

$$= \sum_{j=1}^{r_1} E[M(B_j, c) | C_1 = c] = \sum_{j=1}^{r_1} \alpha_j c \nu(B_j) = \sum_{j=1}^{r_1} \alpha_j' . \qquad (4.38)$$

If we let $\alpha' = \sum_{j=1}^{r_1} \alpha_j \nu(B_j)$, then for the basin B, the conditional p.m.f. of the number of cells can be written as,

$$P\{M(B,c) = k | C_1 = c\} = e^{-\alpha' c}(\alpha' c)^k/k!; \quad k \geq 0. \quad (4.39)$$

Note that Equation (4.39) is identical to Equation (4.21), except that the parameter α_1' of Equation (4.21) is now replaced by the parameter α', given by Equation (4.38). Therefore the c.d.f. of the cumulative rainfall over the basin $B = \bigcup_{i=1}^{r_1} B_i$, is given by Equation (4.35), with parameter α_1' replaced by parameter $\alpha' = \sum_{j=1}^{r_1} \alpha_j'$.

A particular case of the c.d.f. of $Z(C_1)$ as given by Equation (3.35), can be obtained by assuming the p.d.f. of the random variable Q, given by Equation (4.15) to be degenerate at 1. This assumption on Q would correspond to the case of having relatively very small storm cells as compared to the size of the river basin. Based on this assumption, Equation (3.35) reduces to the following analytical form, namely,

$$P\{Z(C_1) < z\} = 1 - \frac{\beta_1}{(\alpha_1'+\beta_1)n_0} \sum_{j=1}^{n_0-1} \frac{\theta\delta^j}{1-\theta\delta^j} e^{-\lambda z'(1-\theta\delta^j)}. \quad (4.40)$$

Finally note, that all the parameters in Equations (4.35) and (4.40) are phenomenologically meaningful. This provides a contrast to an empirical approach, where the parameters in the fitted c.d.f. generally have only a statistical significance.

Cumulative Distribution Function of the Excess Water Yield. Having obtained the c.d.f. of cumulative rainfall per event, we now obtain for special watershed conditions the c.d.f. of the excess water yield, denoted by $Z'(T_1)$. Recall, that Equation (3.20) in Chapter 3, gives the excess water yield as,

$$Z'(T_1) = Z(C_1) - g(D_1) \tag{4.41}$$

where $Z(C_1)$ and $g(D_1)$ are mutually independent random variables as is indicated in Chapter 3. Now, based on the phenomenological considerations, we first assume a functional form for $g(D_1)$ as given below.

Assume that $g(D_1)$ is primarily governed by the antecedent conditions, i.e., the longer the dry duration, that produces the soil moisture depletion, the higher the threshold $g(D_1)$, but not exceeding some constant, say equal to $\ell_1 > 0$. On the other hand, the shorter the dry duration the lower is the threshold $g(D_1)$, say equal to $\ell_1 - \ell_2$, $\ell_1 > \ell_2 > 0$. A mathematical function satisfying these requirements is assumed to be given by,

$$g(D_1) = \ell_1 - \ell_2 e^{-\omega_0 D_1} \tag{4.42}$$

where ω_0 is some decay constant.

Assume for the demonstrative purpose that D_1 follows an exponential p.d.f. with parameter $\beta_2 > 0$. Then the p.d.f. of D_1 can be expressed as,

$$f_{D_1}(d_1) = f(d_1) = \beta_2 e^{-\beta_2 d_1}, \quad d_1 > 0. \tag{4.43}$$

Now the c.d.f. of the excess water yield, $Z'(T_1)$, corresponding to the c.d.f. of the cumulative rainfall per event, $Z(C_1)$, can be obtained by substituting Equations (4.43), (4.42) into Equation (3.21), that is,

$$P\{Z'(T_1) < z\}$$

$$= \int_0^\infty \beta_2 e^{-\beta_2 d_1} P\{Z(C_1) < z + \ell_2 - \ell_1 - e^{-\omega_0 d_1}\} dd_1. \tag{4.44}$$

In particular, the c.d.f. of $Z'(T_1)$, corresponding to the c.d.f. of $Z(C_1)$, given by Equation (3.35) can be obtained by substituting Equation (3.35) into Equation (4.44). However, after making this substitution, any further simplification of Equation (4.44) is not possible. This conclusion mirrors the analytical difficulties in obtaining the similar c.d.f.'s of water yields, by using phenomenologically more realistic watershed models.

4.3.1. Schematic Extension to an Ungaged Basin

Extension of preceding results to an ungaged basin can be achieved in the context of the existence of a gaged region R surrounding the basin. Such has been assumed throughout in our treatment. The concepts behind such an extension are given below.

We assume that the ungaged basin say B_u is contained within our region R. Then it follows from this assumption that the assumed conditional p.m.f. of spatial distribution of the storm cells, as given in Section 4.2, is also applicable to the ungaged basin. The parameter α_u for B_u, given by Equation (4.4), can be estimated from the estimate of parameters for the gaged basins and topography of the region, using if necessary, some subjective judgment. Recall that the random intervals C_1 and D_1 denoting the random duration of a storm and the dry period preceding a storm respectively, have been defined with respect to the region R; hence their c.d.f.'s are also valid for the ungaged basin. The remaining treatment can be carried out in the same way as that for the gaged basins.

In the above context, we emphasise that the formulation of a framework in the context of a region surrounding one or more sub-basins under consideration, is essential.

In such a situation the storm characteristics are determined with respect to the region. Of course, such determinations can be made only on the basis of whatever data is available. Then extensions to ungaged basins can be carried out as indicated above. In this context, the occurrence of an atom in the c.d.f. of the cumulative rainfall, as given by Equation (4.35) is very reasonable, because it gives the probability that no storm cell occurs over a sub-basin, when a storm occurs over the region.

With this we close this section on the determination of the c.d.f.'s of the cumulative rainfall and the water yields from a rainfall event. The next section deals with the derivation of the expected cumulative rainfall within some season, $(0,t_1]$.

4.4. Expected Cumulative Rainfall Within a Fixed Season

Recall from Section 3.3 in Chapter 3, that the general expressions for the $E[X(t_1)]$ and $E[X^2(t_1)]$ denoting the first and second moments of cumulative rainfall within a fixed season $(0,t_1]$, are given by Equations (3.30) and (3.31). In this section, we demonstrate the use of Equation (3.30) in regard to obtaining $E[X(t_1)]$,

The expression for expected value of $X(t_1)$ as given by Equation (3.30), is

$$E[X(t_1)] = \int_0^{t_1} E[Z(s)\,|\,T_1 = s]m(t_1 - s)dF(s)$$

$$+ \int_0^{t_1} E[Z(s)\,|\,T_1 = s]dF(s), \qquad (4.45)$$

where, $F(s) = P\{T_1 < s\}$.

We first restrict ourselves to the summer rainfall in semi-arid lands, such as the Southwestern United States.

Recall from Section 3.2 in Chapter 3 that $T_1 = C_1 + D_1$. Using Equations (4.43) and (4.20) for the p.d.f. of D_1 and C_1 respectively, the p.d.f. of T_1 for any $s > 0$, can be obtained as (Feller, 1971, p. 7),

$$f_{T_1}(s) = f(s) = (e^{-\beta_2 s} - e^{-\beta_1 s})\beta_1\beta_2(\beta_1-\beta_2), \quad \beta_1 \neq \beta_2. \quad (4.46)$$

Equation (4.49) represents the convolution of the p.d.f.'s of two non-identically distributed exponential random variables. Using Equation (4.45) as such, an analytical expression for $E[x(t_1)]$ is not possible. However, based on a phenomenological justification, the following simplification can be achieved to obtain an approximate expression for $E[x(t_1)]$.

Purely from phenomenological considerations, $E(D_1) >> E(C_1)$, because the air-mass type of convective rainfall events have very short durations. This in turn implies that, $\beta_1 >> \beta_2$. Such has also been observed for the point rainfall in Boston, Massachusetts (Grayman and Eagleson, 1969). Hence Equation (4.46) can be approximated as,

$$f(s) \simeq \beta_2 e^{-\beta_2 s}; \quad s > 0. \tag{4.47}$$

Equation (4.47) states that the random interval T_1 follows approximately an exponential p.d.f. Substituting Equation (4.47) into Equation (3.5) in Chapter 3, it can be shown that $N(t_1)$ follows a Poisson p.m.f. with parameter β_2. Note that such has been empirically fitted by Duckstein et al. (1972), for the point rainfall process, in Tucson, Chicago and New Orleans.

Now $E[Z(s)|T_1 = s]$ can be calculated by using Equation (3.23) given in Chapter 3, and the assumptions in regard to $\{Y_n\}$ and $M(B_1,C_1)$ given in Section 4.2, as follows:

$$E[Z(s)|T_1 = s] = \int_0^s E[Z(c)|C_1 = c]f(c)dc$$

$$= \int_0^s E[Y]E[M(B_1,c)|C_1 = c]f(c)dc. \tag{4.48}$$

Based on Equations (4.13) and (4.14), $E[Y] = n_0\pi r_c^2/6\lambda$. Since B_1 is sub-basin under consideration, therefore from

Equation (4.4), $E[M(B_1,c)|C_1 = c] = \alpha_1 c \nu(B_1) = \alpha_1' c$. Substituting these two equations into Equation (4.48),

$$E[Z(s)|T_1 = s] = \int_0^s (n_0 \pi r_c^2/6\lambda)\alpha_1' c \beta_1 e^{-\beta_1 c} dc. \tag{4.49}$$

Now substituting Equation (4.49), along with the approximation given by Equation (4.47), into Equation (4.45), the average cumulative seasonal rainfall is obtained as,

$$E[X(t_1)] \simeq \int_0^{t_1}\int_0^s (n_0 \pi r_c^2/6\lambda)\alpha_1' c \beta_1 e^{-\beta_1 c} \beta_2 (t_1 - s)\beta_2 e^{-\beta_2 s} ds$$

$$+ \int_0^{t_1}\int_0^s (n_0 \pi r_c^2/6\lambda)\alpha_1' c \beta_1 e^{-\beta_1 c} \beta_2 e^{-\beta_2 s} ds. \tag{4.50}$$

Equation (4.50) upon simplification gives,

$$E[X(t_1) \simeq (n_0 \pi r_c^2 \alpha_1'/6\lambda)\{[(1+t_1)(1-e^{-\beta_2 t_1})/\beta_2]$$

$$+[(1 - e^{-(\beta_2+\beta_1)t_1} - e^{-(\beta_2+\beta_1)t_1}(\beta_1+\beta_2)t_1)/(\beta_1+\beta_2)^2]$$

$$- [(1+t_1)(1 - e^{-(\beta_1+\beta_2)t_1})/(\beta_1+\beta_2)]$$

$$- [(1 - e^{-\beta_2 t_1} - \beta_2 t_1 e^{-\beta_2 t_1})/\beta_2^2]\}. \tag{4.51}$$

Now let us consider, for example, the case of winter rainfall in the Southwestern United States. The winter rainfall is less intense than the summer rainfall but tends to persist for longer duration than the summer rainfall (Kao et al., 1971). Therefore let us assume that C_1' for winter rainfall is gamma with parameter β_0 and say n_1, and D_1' is gamma with parameters β_0 and say $k_0 n_1$, i.e., $E(D_1') = K_0 E(C_1')$; $K_0 > 1$. Further we assume that the remaining assumptions outlined in Section 4.2 also hold for winter rainfall. Then $T_1' = D_1' + C_1'$ has a gamma p.d.f. with parameters β_0 and $(K_0+1)n_1$ is equal to say n_2. The p.d.f. of T_1' can be expressed as

$$f_{T_1'}(s) = f(s) = \beta_0^{n_2} s^{n_2-1} e^{-\beta_0 s}/n_2!. \tag{4.52}$$

The expectation of $N(t_1,t]$ in the winter season $(t_1,t]$ corresponding to the p.d.f. of T_1 given by Equation (4.52) can be obtained as (Parzen, 1967, p. 177),

$$E[N(t_1,t)] = \beta_1(t-t_1)/n_2 + (1/n_2)\sum_{j=1}^{n_2-1} [\varepsilon_1^j/(1 - \varepsilon_1^j)]$$

$$\cdot \exp[-\beta_1(t-t_1)(1 - \varepsilon_1^j)], \tag{4.53}$$

where $\varepsilon_1^j = \exp(2\pi ij/n_2)$, $i = \sqrt{-1}$. Now the conditional expectation $E[Z(s)|T_1' = s]$ can be calculated from Equation

(4.48), and then $E[x(t,t_1)]$ can be obtained using Equations (4.52) and (4.53). We do not actually perform these computations here since an analytical form for $E[x(t,t_1)]$ is not obtainable.

With this we close this section on obtaining an expression for the expectation of seasonal rainfall within a fixed season. The next section is devoted to the limiting distribution of the maximum cumulative rainfall per event.

4.5. Limiting Distribution of the Maximum Cumulative Rainfall per Rainfall Event

In order to derive the limiting c.d.f. of the maximum of a random number of $Z(C_j)$'s; $j \geq 1$, denoting the cumulative rainfall per rainfall event, we select the expression for the c.d.f. of $Z(C_1)$, as given by Equation (4.40), which is a particular case of Equation (4.35). Recall from Section 4.3 that Equation (4.40) is applicable to those basins, which are large in comparison with the size of the storm cells. The reason for selecting Equation (4.40) is, that Equation (4.35) induces mathematical difficulties in obtaining the limiting c.d.f. of maximum of $Z(C_j)$'s; $j \geq 1$. Further, in order to keep the computations simple, take $n_0 = 1$ in Equation (4.40). Then the c.d.f. of $Z(C_1)$ as given by Equation (4.40) simplifies to,

$$P\{Z(C_1) < z\} = 1 - \frac{\alpha_1'}{\alpha_1'+\beta_1} e^{-\lambda z'\beta_1/(\alpha_1'+\beta_1)}$$

$$= 1 - \frac{\alpha_1'}{\alpha_1'+\beta_1} e^{-\lambda z \beta_1 3/\pi r_c^2(\alpha_1'+\beta_1)} \tag{4.54}$$

since $z' = 3z/\pi r_c^2$. Let a constant θ_1 be given as

$$\theta_1 = 3\lambda\beta_1/\pi r_c^2(\alpha_1'+\beta_1). \tag{4.55}$$

In view of Equation (4.55), Equation (4.54) can be simply written as,

$$P(Z(C_1) < z) = 1 - \frac{\alpha_1'}{\alpha_1'+\beta_1} e^{-\theta_1 z}. \tag{4.56}$$

Recall from Equation (3.72) in Chapter 3, that

$$V(n) = \max_{1 \le j \le n} Z(T_j) = \max_{1 \le j \le n} Z(C_j)$$

and the limiting c.d.f. of $V(n)$ is given by Equation (3.74). In view of Equation (3.74), let the sequences $\{u_n\}$ and $\{v_n\}$ be defined as,

$$\left.\begin{aligned} u_n &= \frac{1}{\theta_1} \\ v_n &= \frac{1}{\theta_1}\log n \end{aligned} \quad ; \quad n \ge 1 \right\}. \tag{4.58}$$

Then the limiting c.d.f. of $V(n)$ can be obtained as,

$$\lim_{n\uparrow\infty} P\{V(n) < u_n z + v_n\} = \lim_{n\uparrow\infty} [1 - \frac{\alpha_1'}{\alpha_1'+\beta_1} e^{-(z+\log n)}]^n$$

$$= \lim_{n\uparrow\infty} [1 - \frac{\alpha_1' e^{-z}}{(\alpha_1'+\beta_1)n}]^n = \exp[- \frac{\alpha_1'}{\alpha_1'+\beta_1} e^{-z}] \qquad (4.60)$$

Now the limiting c.d.f. of $V(N_n)$, denoting the maximum cumulative rainfall per event of a random number of rainfall events within $(0,t_1]$ up to the n^{th} year; $n \geq 1$, can be obtained from Equation (3.79) in Chapter 3. Therefore substituting Equation (4.60) into Equation (3.79), the limiting c.d.f. is given by,

$$\lim_{n\uparrow\infty} P[V(N_n) < u_n z + v_n]$$

$$= \exp[- \frac{m(t_1)\alpha_1'}{\alpha_1'+\beta_1} e^{-z}] \qquad (4.61)$$

where $m(t_1) = E[N(t_1)]$.

Derivation of the limiting c.d.f. for the more general case, i.e., when $n_0 > 1$ in Equation (4.40), is not considered here. Such considerations would involve more detailed analysis in regard to the selection of the constants $\{u_n\}$ and $\{v_n\}$ then that given for the special case of $n_0 = 1$.

This concludes the chapter on theoretical applications. In the next chapter an example is considered to give a numerical demonstration of a few of the applications developed in this chapter.

CHAPTER 5

NUMERICAL EXAMPLE, SUMMARY AND RECOMMENDATIONS

5.1. Numerical Example

This section is devoted to a numerical demonstration of a few important aspects of the mathematical applications given in Chapter 4. Specifically, the intent behind this numerical case study is two fold; (i) empirical justification of the c.d.f.'s of random variables that were used in Chapter 4 to develop mathematical models, and (ii) computational demonstration of a few specific analytical results obtained in the earlier sections.

The Atterbury watershed is selected for this numerical demonstration. This watershed is located near Tucson, Arizona and is about 18 mi.^{2} in area. It has a fairly dense network of both recording and non-recording raingages. A plan view of the Atterbury watershed is given in Figure 5.1.

The fixed season is taken to be the summer season, during which the air mass type of convective rainfall occurs in and near Tucson. Typically, the summer season is comprised of the interval from the second week of July to about the middle of September. Therefore the fixed interval

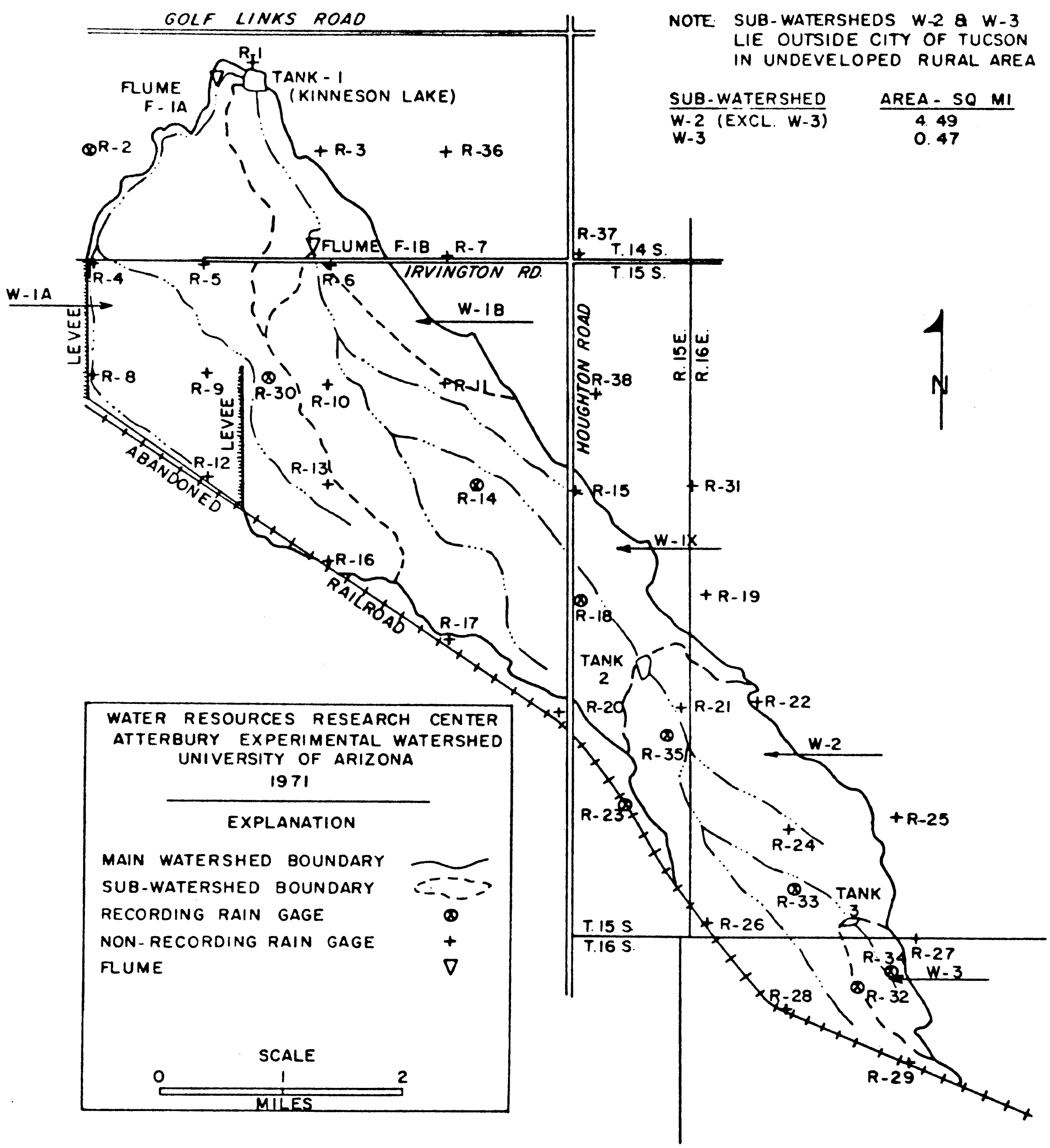


Figure 5.1. A Plan View of the Atterbury Watershed.

$(0,t_1]$, as given in the theoretical considerations, is now assumed to be the summer season.

The summer rainfall due to air mass type of convective storms has very high spatial variability and generally the storm cells occur in isolation both in time and space. Moreover, the durations of these cells, on the average, is in the order of thirty minutes or so. Therefore the following subjective criterion is used to first select the random durations of rainfall events. Twelve hour periods are chosen from 0001 to 1200 hours (noon) and from 1201 to 2400 hours (midnight). The random duration of the rainfall events are then selected from within these twelve hour periods as follows. Continuous rainfall data are observed simultaneously at the recording rain gages R-2 and R-32 shown in Figure 5.1. Then the duration of a rainfall event is the time interval from the beginning of rainfall at any one of these two rain gages to the time when the rainfall terminates at both of these two rain gages. This information is selected for twenty-two rainfall events from within eight years of data as given in Table 5.1. The corresponding values of the random durations are also indicated in Table 5.1. Now, these durations are the values assumed by the random variable C_1. An exponential p.d.f. is hypothesised for C_1 in Chapter 4. Figure 5.2 shows the fitted

Table 5.1. Summer Rainfall Data for Atterbury Watershed

Date	Gage R-32		Gage R-2		Duration C_1	Cell Center Depth ξ_0	Gage
	Begins	Ends	Begins	Ends	Hours	Inches	
1969							
9/4	1900	2120	2100	2300	4.00	1.12	R-21
8/11	-	-	-	-	-	0.75	R-20
8/7	1740	1830	-	-	.83	1.03	R-21
7/6	-	-	1430	1500	.50	0.62	R-5
1968							
8/19	2100	2130	1800	1900	5.66	-	-
	2245	2315	2215	2340		-	-
8/7	1600	1620	1630	1720	1.33	0.95	R-6
7/30	-	-	1830	1900	.50	0.85	R-16
7/21	-	-	1715	1735	.33	1.12	R-6
1967							
8/23	1430	1445	-	-	.15	0.90	R-24
7/29	-	-	2000	2130	1.50	1.18	R-14
7/17	0620	0930				2.10	R-21
	-	-	0630	0930	3.17	1.37	R-6
	-	-	-	-		1.91	R-16
7/12	0115	0345	0130	0340	2.50	1.51	R-21
7/10	2100	2130	2130	2200	1.00	0.92	R-23
1966							
8/19	-	-	1900	1915	.25	1.11	R-11
8/16	1715	1800	-	-	.75	1.28	R-20
1965							
9/12	1400	1515	-	-	1.25	1.23	R-17
8/29	0430	0500	-	-	.50	1.27	R-21
8/17	-	-	0030	0200	1.50	1.40	R-1
7/16	-	-	1800	1900	1.00	2.08	R-9
1964							
7/22	-	-	-	-	-	1.91	R-29
7/4	2030	2100	-	-		0.71	R-32
	-	-	1940	2100	1.33	1.50	R-1
	-	-	-	-		1.30	R-20
7/12	1700	1900	-	-	2.00	1.23	R-29
1963							
7/28	1830	1940	-	-	1.17	1.26	R-24
	-	-	1900	1940		1.45	R-13
1961							
8/22	1600	2300	-	-		1.93	R-18
	-	-	2020	2240	6.66	2.26	R-9
	-	-	-	-		1.56	R-26

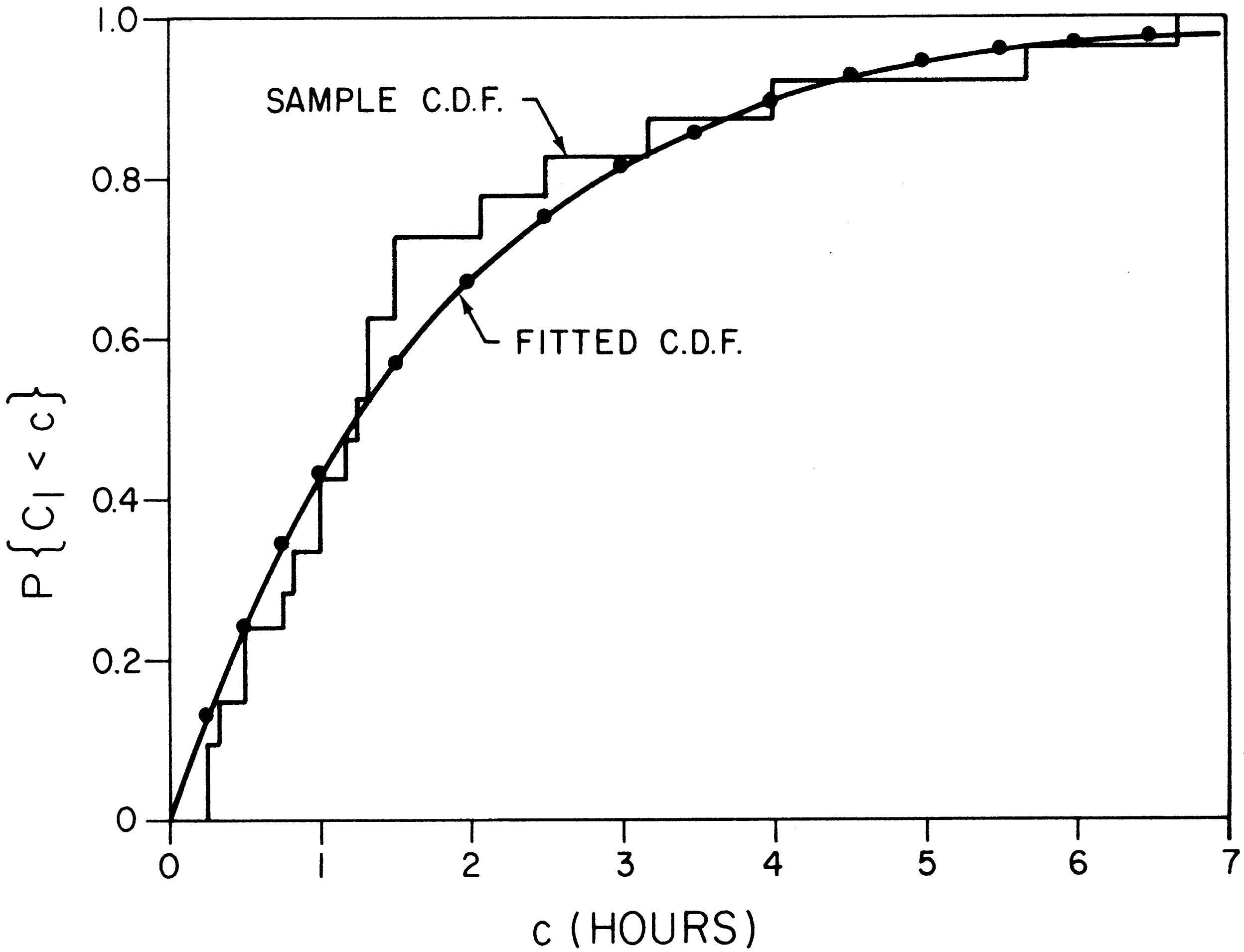


Figure 5.2. Empirical and the Fitted c.d.f.'s for the Duration of a Rainfall Event.

exponential c.d.f. and the sample c.d.f. of C_1, with the estimated parameter $\hat{\beta}_1 = .558$ (hours).

Next, in regard to the random variable ξ_0, denoting the maximum cell depth (cell center depth), the following is done to obtain the observed values. Firstly, isohyets are drawn for the total rainfall per cell, using data from non-recording rain gages. Then the maximum observed rainfall amount is obtained for the center of a cell isohyet by taking the recorded rainfall amount from the rain gage that is closest to the isohyetal center. Table 5.1 gives the value of the maximum cell rainfall depths in inches, and the corresponding rain gages at which these values are observed. A two parameter gamma p.d.f. is hypothesised for ξ_0 in Chapter 4. The parameters λ and n_0 are estimated from the observed data based on the method of moments. The sample c.d.f. and the fitted two parameters gamma c.d.f. are shown in Figure 5.3, with the estimated parameters $\hat{\lambda} = 7.34$ and $\hat{n}_0 = 10$.

Finally, in regard to the conditional p.m.f. of the occurrence of number of cells over a sub-basin B_1, recall that a Poisson distribution is hypothesised in Section 4.2.1 of Chapter 4. However, the parameter α_1', that denotes the conditional expectation (on C_1) of the number of cells over B_1, cannot be estimated properly in view of having

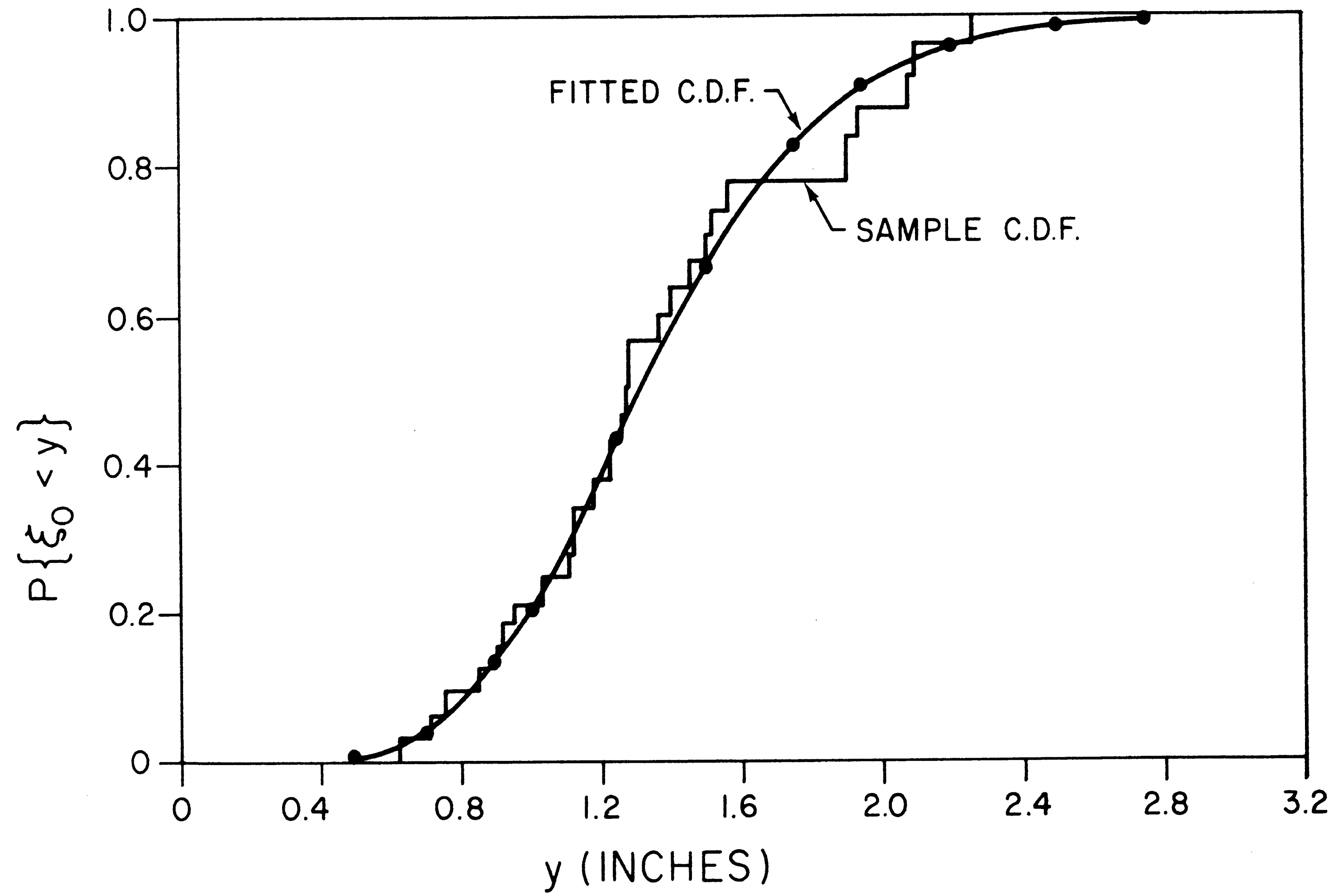


Figure 5.3. Empirical and Fitted c.d.f.'s for the Maximum Cell Depth.

only twenty-two rainfall events. Therefore the following procedure is recommended, which seems important both from the viewpoint of estimating α_1' and verifying the hypothesis of a Poisson distribution.

Let $M(B_1)$ denote the random variable giving the number of cells per rainfall event. Then the p.m.f. of $M(B_1)$ can be obtained by integrating the conditional p.m.f. over the random duration C_1 as follows

$$P\{M(B_1) = k\} = \int_0^\infty P\{M(B_1,c) = k | C_1 = c\} f(c) dc;$$

$$k \geq 0. \tag{5.1}$$

Substituting Equations (4.21) and (4.20) from Chapter 4, into Equation (5.1),

$$P\{M(B_1 = k\} = \int_0^\infty [e^{-\alpha_1' c} (\alpha_1' c)^k / k!] \beta_1 e^{-\beta_1 c} dc$$

$$= [\alpha_1' / (\alpha_1' + \beta_1)]^k \beta_1 / (\alpha_1' + \beta_1); \quad k = 0, 1, 2, \ldots, \tag{5.2}$$

Equation (4.2) represents a geometric distribution for the unconditional p.m.f. of the number of cells that can occur from a rainfall event over the basin B_1. Strictly from the data viewpoint, it is easier to obtain the p.m.f. of $M(B_1)$, since only the non-recording rain gages can be used to interpolate the isohyets of the cells, which in turn are used to determine the number of cells per rainfall event. Use of Equation (5.2) in the estimation of α_1' is given below.

In the first place, note that no gaged region surrounding the Atterbury watershed is considered. However, the actual basin from which the cells can also contribute to the cumulative rainfall over the Atterbury sub-basin, is bigger than the Atterbury sub-basin. This is because of the fact, that those storm cell centers that occur outside the boundary of Atterbury within a certain distance from it, can also contribute to the cumulative rainfall over the Atterbury sub-basin. To make this distinction clear, we denote Atterbury sub-basin by B_a and the basin surrounding Atterbury, from which a cell can contribute to the cumulative rainfall over Atterbury, by B_1. Clearly $B_1 \supset B_a$. Assume that a cell has a fixed radius of three miles, i.e., $r_c = 3$ (Osborn, 1970). Then the boundary of the basin B_1 can be identified by drawing a boundary parallel to the Atterbury sub-basin boundary, at a distance of three miles from it. The area enclosed by the basin B_1 is then approximately equal to 128 mi.^{2}.

With the above formulation in view, we now estimate α_1' of the Poisson distribution corresponding to the basin B_1. Fogel and Duckstein (1969) give the information in regard to sixty-four rainfall events observed on Atterbury sub-basin. Their definition of an event implies the occurence of at least one storm cell center over the Atterbury. We assume that the p.m.f. of number of cells based on the definition of a rainfall event as given by Fogel and Duckstein, corresponds to the unconditional p.m.f. of the number of cells,,as given by Equation (5.2). Now out of the sixty-four rainfall events as given by Fogel and Duckstein (1969), fifty-four of them had only one cell center, nine of them had two and three of them had three cell centers located over Atterbury sub-basin. Based on the data, there are, on the average, 1.23 cell centers that can occur over Atterbury from a rainfall event. This average value, denoted by $E[M(B_a)]$, is obtained from Equation (5.2) to be,

$$\left.\begin{aligned} & E[M(B_a)] = \beta_1/\alpha'_a = 1.23 \\ \text{or,} \quad & \\ & \hat{\alpha}'_a = \hat{\beta}_1/1.23 = .471 \end{aligned}\right\} . \qquad (5.3)$$

Thus within our formulation, .471 denotes an estimate of the average number of cell centers that can occur over Atterbury, given the duration of a rainfall event; provided the rainfall durations have an exponential p.d.f. Recall that α'_a represents the parameter of the Poisson p.m.f. for the Atterbury sub-basin, B_a. Now, using

Equation (5.3) in conjunction with the assumption, that the parameters α'_a and α'_1 for the sub-basin B_a and the basin B_1 respectively, are proportional to the areas of the basins, $\hat{\alpha}'_1$ can be obtained as $\hat{\alpha}'_1 = .471 \times 128/18 = 3.35$.

Finally, an empirical verification of the unconditional p.m.f. of the number of cells over B_a, is not possible presently, because no information is available on the relative frequency of zero cell centers over the Atterbury sub-basin. However the relative frequencies for 1, 2 and 3 cells do suggest a geometric type p.m.f., as obtained in Equation (5.2).

In summary, the following estimates of the parameters are obtained to be used in Equations (4.35) or (4.40) giving the c.d.f. of the cumulative rainfall per rainfall event; $\hat{\alpha}'_1 = 3.35$, $\hat{\beta}_1 = .558$, $\hat{\lambda} = 7.34$, $\hat{n}_0 = 10$, $r_c = 3$ and area of $B_a = 18$ mi.^{2}. However, a small but important modification is required in Equations (4.35) and (4.40), which arises due to the size and shape of the Atterbury sub-basin in relation to the size and shape of the thunderstorm cells. Recall from Section 4.2.2 in Chapter 4, that the upper bound on the random variable Q denoting the fraction of the cumulative rainfall from a cell that can occur over a basin, is set at 1. But presently, Atterbury has an area of only 18 mi.^{2}, whereas a cell has an area of 28.3 mi.^{2} (corresponding to $r_c = 3$ miles), therefore a maximum of roughly

70% of the total rainfall from a cell can ever occur over Atterbury. Thus the upper bound on Q becomes .7 in case of Atterbury sub-basin. In view of this, Equation (4.35) modifies to,

$P\{Z(C_1) < z\}$

$$= 1 - \frac{\beta_1}{(\alpha_1' + \beta_1)n_0} \sum_{j=0}^{n_0-1} \frac{\theta\delta^j}{1-\theta\delta^j} \int_0^{.7} e^{-\lambda z'(1-\theta\delta^j)/q} dq. \quad (5.4)$$

A plot of the c.d.f. of $Z(C_1)$, as given by Equation (5.4) is shown in Figure 5.4, corresponding to different values of z in mi.^{2} inches.

Similarly Equation (4.40), which represents a particular case of Equation (3.35), in view of Equation (5.4) modifies to,

$P\{Z(C_1 < z\}$

$$= 1 - \frac{\beta_1}{(\alpha_1' + \beta_1)n_0} \sum_{j=0}^{n_0-1} \frac{\theta\delta^j}{1-\theta\delta^j} e^{-\lambda z'(1-\theta\delta^j)/.7}. \quad (5.5)$$

A plot of Equation (5.5), corresponding to different values of z in mi.^{2} inches is given in Figure 5.5.

Now, the modified form of Equation (4.35) as given by Equation (5.4), corresponds to the case where the random effect of the location of a cell on the cumulative rainfall

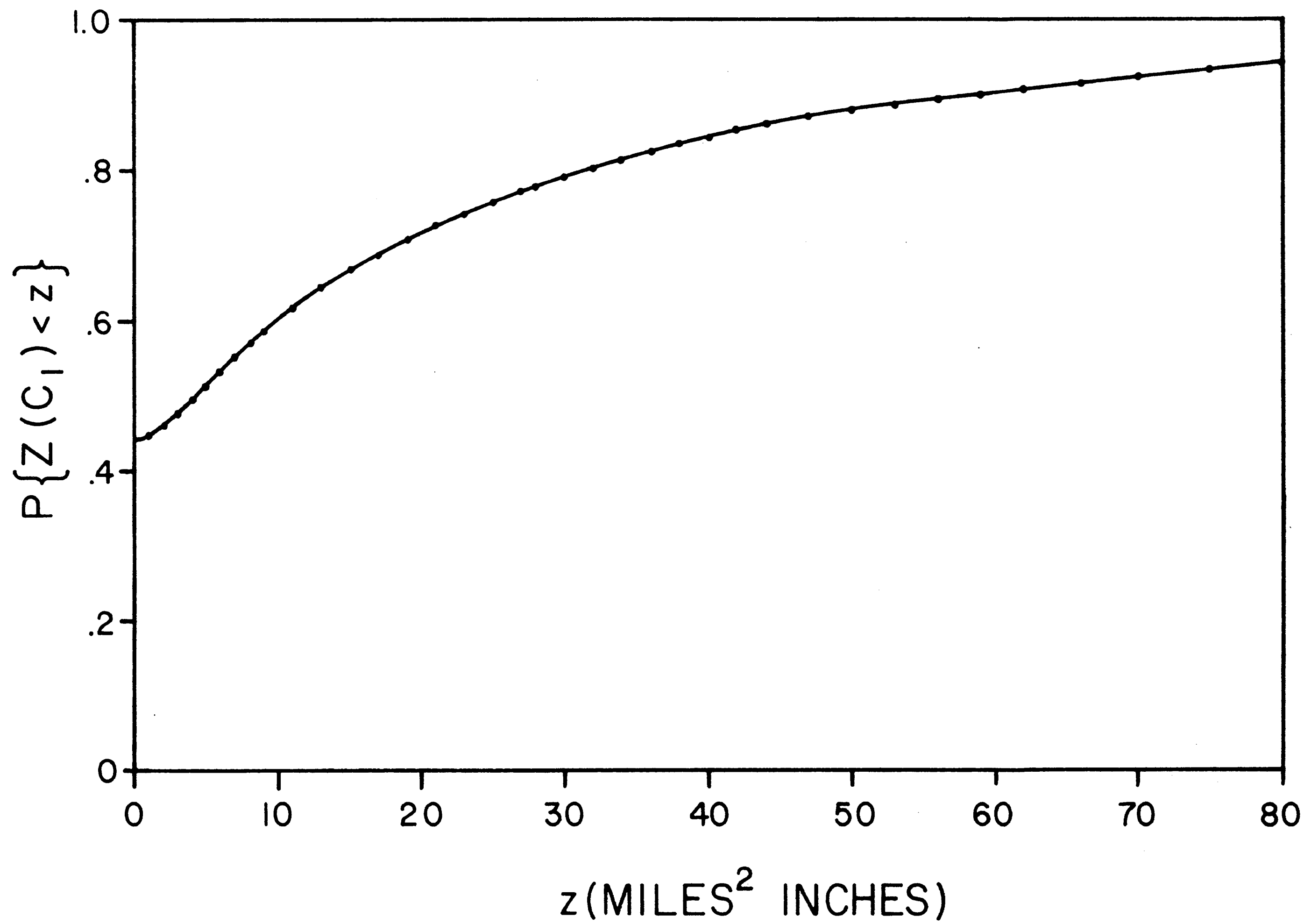


Figure 5.4. c.d.f. of the Cumulative Rainfall per Rainfall Event: Considering the Random Effect of Cell Location.

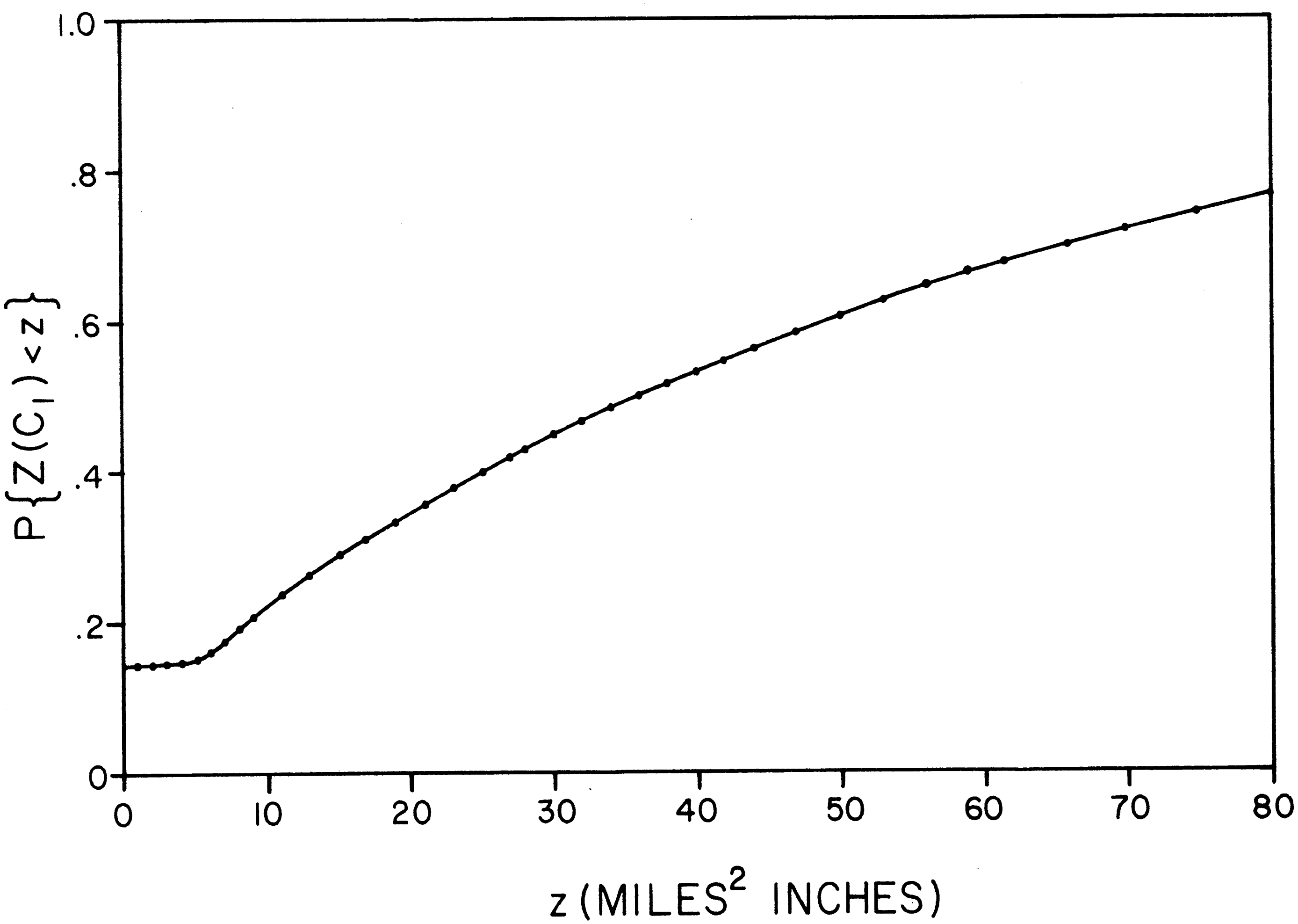


Figure 5.5. c.d.f. of the Cumulative Rainfall per Rainfall Event: Not Considering the Random Effect of Cell Location.

that can occur on the basin has been considered. On the other hand, Equation (5.5) is obtained on the assumption that the location of a cell does not affect the resultant cumulative rainfall over the basin, which is reasonable for those basins that are large in relation to the cell size. The influence of location of a cell on the c.d.f. of the cumulative rainfall per event from a small basin can be ascertained from Figures 5.4 and 5.5.

The 'atom' at the origin of the c.d.f. of $Z(C_1)$ corresponds to the probability that no cell occurs over the basin B_1 during the duration of a rainfall event over a region R that contains B_1. Since no such region has been identified presently, therefore the existence of this 'atom' is kind of hypothetical. Moreover, the random duration C_1 is also defined with respect to a region, but in the above example it has been computed with respect to the Atterbury watershed. In this context, the c.d.f.'s of $Z(C_1)$ given in Figures 5.4 and 5.5 should only be considered demonstrative.

Having obtained the c.d.f. of $Z(C_1)$, the limiting c.d.f. of the maximum cumulative rainfall per rainfall event can be computed. However, for Atterbury, $n_0 = 10$, and the special case derived in Section 4.5 of Chapter 4, is for $n_0 = 1$. Therefore we do not give numerical computations for the limiting c.d.f. corresponding to Equation (5.5), since for $n_0 > 1$, the sequences $\{u_n\}$ and $\{v_n\}$ will have to be redetermined.

This concludes the numerical demonstration of the important results obtained in Chapter 4. Ensuing section contains a comprehensive summary of this study.

5.2. Summary and Recommendations

In summary, this study introduces a phenomenologically oriented approach for building stochastic models of the cumulative rainfall in a space-time construct. The general theoretical developments given in Chapter 3 have the following important features in comparison with the previous work on the point rainfall process, and in their own right, namely,

(1) It represents a generalization of the previous work on the point rainfall process in two ways, i.e.,

(i) considerations leading to a particular stochastic dependence between the cumulative rainfall amounts per rainfall event and the number of such rainfall events within a season, through the random durations of rainfall events. The *dependence* seems particularly important, when considering *spatial* rainfall over some region. However for the point rainfall process this dependence may not be very significant;

(ii) a phenomenological approach leading to the derivation of the c.d.f. of the cumulative rainfall per rainfall event, as opposed to empirically fitting a c.d.f. to the cumulative rainfall per rainfall event.

(2) An analytical result on the error bound for the convergence of the sum of a random number of random variables, not necessarily independent, is given.

(3) Use of the limiting c.d.f. of the maximum of a random number of random variables (cumulative rainfall per event), when the two are not necessarily independent, is given in the context of determining the 'design rainfall.

(4) Extension of the above results to multiple seasons within an year, such that the cumulative rainfall amounts per event between any two seasons are non-identically distributed.

Possible extensions of this study are possible along the following lines, such as,

(1) Derivation of the c.d.f. of the cumulative rainfall per event at a <u>fixed</u> point in space, using a space-time framework as presented here. This would in turn relate the previous work on the point rainfall process to the present work on the space-time rainfall.

(2) Use of this approach to simulate space-time rainfall. The use of a theoretical framework as given here,

in the rainfall simulation would be an improvement over some previous simulation procedures, such as adopted by Grayman and Eagleson (1971), who simultaneously use both the empirical c.d.f.'s as well as the expected values of random variables to simulate rainfall in space and time. The simulated rainfall in turn may be used to estimate the c.d.f.'s of random variables defined on streamflows using an approach such as proposed by Gupta (1972).

(3) Possibility of improving the error bound on the convergence of the c.d.f. of the seasonal cumulative rainfall, to its limiting c.d.f. A 'tight' error bound can be used to determine the c.d.f. of cumulative rainfall over a few years period.

(4) Considerations on determining the limiting c.d.f. of the maximum cumulative rainfall per event, using the general expressions of the c.d.f. of the cumulative rainfall per event as given by Equations (4.35) and (4.40).

(5) Considerations on other random variables defined on the process of space-time rainfall, than the ones considered here. Such may be the time of occurrence of the maximum rainfall per event within a season, limiting behavior of its c.d.f.; crossing properties of the partial sums defined in terms of the normalized

cumulative rainfall within a fixed season over a sequence of years, which in turn may be useful to analyse quantitative aspects of droughts defined in terms of the rainfall amounts.

(6) Application of this approach to other basins in the United States and other areas of the world with an intent to verify some of the hypothesis and assumptions given here.

(7) Possibility of relaxing some assumptions made in Chapter 4, to arrive at more 'realistic' forms of the c.d.f.'s for the cumulative rainfall per event.

APPENDIX A

DERIVATION OF THE SECOND MOMENT OF $X(t_1)$

$$E[X^2(t_1)] = E[[\sum_{j=1}^{N(t_1)} Z(T_j)]^2 I\{\bigcup_{n=0}^{\infty}\{N(t_1) = n\}\}].$$

Since $Z(T_j)$; $j \geq 1$ are i.i.d., therefore

$$E[X^2(t_1)] = \sum_{n=1}^{\infty} E[\{nZ^2(T_1) + n(n-1)Z(T_1)Z(T_2)\}I\{N(t_1) = n\}]$$

$$= \sum_{n=1}^{\infty}\int_0^{t_1}\int_0^{t_1-s_2} E[nZ^2(s_1) + n(n-1)Z(s_1)Z(s_2)|T_1 = s_1, T_2 = s_2]$$

$$\cdot\ P\{N(t_1 - s_1 - s_2) = n - 2\}dF(s_1)dF(s_2)$$

$$= \int_0^{t_1}\int_{s_1}^{t_1}\sum_{n=1}^{\infty} nF[Z^2(s_1)|T_1 = s_1]P\{N(t_1-s_1-s_2) = n-2\}dF(s_2)dF(s_1)$$

$$+ \int_0^{t_1}\int_0^{t_1-s_2}\sum_{n=1}^{\infty} n(n-1)E[Z(s_1)|T_1 = s_1]E[Z(s_2)|T_2 = s_2]$$

$$\cdot\ P\{N(t_1) = t_1 - s_1 - s_2\}dF(s_1)dF(s_2).$$

This upon simplification gives

$$E[X^2(t_1)] = \int_0^{t_1} E[Z^2(s_1)|T_1 = s_1]\{E[N(t_1)] + 1\}dF(s_1)$$

$$+ \int_0^{t_1}\int_0^{t_1-s_2} \sum_{n=1}^{\infty} n(n-1)E[Z(s_1)|T_1 = s_1]E[Z(s_2)|T_2 = s_2]$$

$$\cdot\ P\{N(t_1 - s_1 - s_2) = n - 2\}dF(s_1)dF(s_2).$$

APPENDIX B

A SECOND DERIVATION OF THE EXPECTATION OF $X(t_1)$

Consider the identity given by Todorovic (1970) as,

$$E[X(t_1)] = \sum_{n=1}^{\infty} \int_{\{\tau_n < t_1\}} E[Z(T_n) | N(t_1)] dP.$$

Since $Z(T_n)$ depends on $N(t_1)$ only through T_n; $n > 1$, therefore,

$$E[X(t_1)] = \sum_{n=1}^{\infty} \int_0^{t_1} \int_{\{\tau_{n-1} < t_1 - s\}} E(Z(T_n) | T_n = s) dP dF(s).$$

Since T_n; $n \geq 1$ are i.i.d. as T_1, therefore

$$E[X(t_1)] = \int_0^{t_1} \sum_{n=1}^{\infty} E[Z(T_1) | T_1 = s] [\int_{\{\tau_{n-1} < t_1 - s\}} dP] dF(s)$$

$$= \int_0^{t_1} E[Z(T_1) | T_1 = s] dF(s) \sum_{n=1}^{\infty} P\{\tau_{n-1} < t_1 - s\}.$$

In view of Equation (3.6), the above equation reduces to,

$$E[X(t_1)] = \int_0^{t_1} E[Z(T_1)\,|\,T_1 = s]dF(s)\{1 + E[N(t_1-s)]\}$$

$$= \int_0^{t_1} E[Z(T_1)\,|\,T_1 = s]m(t_1-s)dF(s)$$

$$+ \int_0^{t_1} E[Z(T_1)\,|\,T_1 = s]dF(s).$$

LIST OF SELECTED SYMBOLS

Symbol	Meaning
$x \in A$	x is an element of the set A
$\subset (\supset)$	is contained in (contains)
$\bigcup_{i=1}^{n} A_i$	the union of the sets $A_1, \ldots, A_n$
$\bigcap_{i=1}^{n} A_i$	the intersection of the sets $A_1, \ldots, A_n$
$\emptyset$	the empty set
$\|x\|$	absolute value of x
$\uparrow$	is increasing to
O	of the order of
$I\{A\}$	the indicator function of a set A
P	probability measure
R	a geographic region
$B_i;\ 1 \leq i \leq r$	r mutually disjoint sub-basins within a region
C_j	random duration of the j^{th} rainfall event
D_j	random dry duration preceding the j^{th} rainfall event

τ_j	termination epoch of the j^{th} rainfall event
$T_j = (\tau_{j-1}, \tau_j)$	random interval between the terminations of the $(j-1)^{st}$ and the j^{th} rainfall events
$Z(T_j) = Z(C_j)$	the cumulative rainfall from the j^{th} rainfall event
$Z'(T_j)$	the excess water yield from the j^{th} rainfall event
$N(t_1)$	the number of rainfall events within a fixed season $(0,t_1]$
$N(t_1,j) = N(j)$	the number of rainfall events within $(0,t_1]$ <u>during</u> the j^{th} year
$N_i(t_1) = N_i = \sum_{j=1}^{i} N(j)$	the number of rainfall events within a fixed $(0,t_1]$ <u>up to</u> the i^{th} year
$X(t_1)$	the cumulative rainfall from a basin B under consideration $(\subset R)$ within $(0,t_1]$
$M(B,C_1)$	the number of storm cells that occur over a basin $B(\subset R)$, during the random duration C_1 of a rainfall event
Y_n	the cumulative rainfall from the n^{th} storm cell

$V(N(i)) = \sup_{1\leq j\leq N(i)} Z(T_j)$	the maximum rainfall among $Z_1, \ldots, Z(T_{N(i)})$ within $(0,t_1]$ <u>during</u> the i^{th} year
$V(N_i) = \sup_{1\leq j\leq N_i} Z(T_j)$	the maximum rainfall among $Z_1, \ldots, Z(T_{N_i})$ within $(0,t_1]$ <u>up</u> <u>to</u> the i^{th} year
$F(\cdot)$	denotes the c.d.f. of a random variable (same notation is used for the c.d.f.'s of all the random variables)
$f(\cdot)$	denotes the p.d.f. of a random variable
$E[\cdot]$	denotes the mathematical expectation of a random variable
$E[N(t_1)] = m(t_1) = m$	mathematical expectation of the number of rainfall events within $(0,t_1]$
$\nu(B_i)$	denotes the area of the i^{th} sub-basin $(\subset R)$

LIST OF REFERENCES

Amorocho, J., and Morgan, D., "Convective Storm Field Simulation for Distributed Catchment Model", Porceedings of the International Symposium on Mathematical Modeling in Hydrology, International Institute for Scientific Hydrology, Warsaw, Poland (1971).

Anscombe, F. J., "Large-Sample Theory of Sequential Estimation", Proc. Camb. Phil. Soc., V. 48 (1952), p. 600.

Berman, S. M., "Limiting Distribution of the Maximum Term in Sequences of Dependent Random Variables", Ann. Math. Stat., V. 33 (1962), pp. 894-908.

Bhattacharaya, R. N., Dept. of Math., Univ. of Arizona, Typed Manuscripts (1973).

Breiman, L., "Probability", Addison-Wesley Publishing Co. (1963).

Bromwich, T., "An Introduction to the Theory of Infinite Series", Macmillan and Co., Ltd. (1926).

Byers, H. R., and Braham, R. R., "The Thunderstorm", Report of the Thunderstorm Project, Washington, D. C. (1949).

Caskey, J. E. Jr., "A Markov Chain Model for the Probability of the Precipitation Occurrences in Intervals of Various Length", Monthly Weather Review, V. 91 (1963), pp. 298-301.

Churchill, R. V., "Complex Variables and Applications", McGraw-Hill Book Co. (1960).

Court, A., "Area-Depth Rainfall Formulas", J. Geophysical Research, No. 66 (1961), pp. 1823-1831.

Cox, D. R., "Renewal Theory", Methuen and Co., Ltd. (1962).

Crovelli, R. A., "Stochastic Models for Precipitation", Ph. D. Dissertation, Colorado State University (1971).

Davis, D. R., Kisiel, C. C., and Duckstein, L., "Optimum Design of Mountainous Raingage Network Using Decision Theory", Proceedings, Symp. on Dist. of Precip. in Mountainous Areas, Geilo, Norway (1972), pp. 416-420.

Duckstein, L., Fogel, M., and Kisiel, C. C., "A Stochastic Model of Runoff Producing Rainfall for Summer Type Storms", Water Resources Research, V. 8, No. 2 (1972), pp. 410-421.

Duckstein, L., Fogel, M., and Thames, J., "Elevation Effects on Rainfall: A Stochastic Model", J. of Hydrology, V. 18, No. 1 (1973), pp. 21-36.

Eagleson, P. S., "Dynamics of Flood Frequency", Water Resources Research, V. 8, No. 4 (1972), pp. 878-898.

Feller, W., "An Introduction to Probability Theory and its Applications, Vol. II", Second Edition, John Wiley and Sons, Inc. (1971).

Feyerherm, A. M., and Bark, L. D., "Statistical Methods for Persistent Precipitation Patterns", J. Appl. Met., V. 4 (1965), pp. 320-328.

Fogel, M., "Effect of Storm Rainfall Variability on Runoff from Small Semi-arid Watersheds", Transactions, A.S.C.E., V. 12, No. 6 (1969).

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

Fuller, W. E., "Flood Flows", Transactions, A.S.C.E., V. 77, Paper 1293 (1914).

Gabriel, K. R., and Neumann, J., "A Markov Chain Model for Daily Rainfall Occurrences at Tel-Aviv", Quart. J. Royal Met. Soc., V. 88 (1962), pp. 90-95.

Genedenko, B. V., "Sur la distribution Limite du derme maximum d'une série aléatoire", Ann. Math. V. 44 (1942), pp. 423-453.

Grace, R. A., and Eagleson, P. S., "The Synthesis of Short-Time Inherent Rainfall Sequences", R. M. P. Lab. Report No. 91, M.I.T. (1966).

Grayman, W. M., and Eagleson, P. S., "Streamflow Record Length for Modeling Catchment Dynamics", R. M. P. Lab. Report No. 114, M.I.T. (1969).

Grayman, W. M., and Eagleson, P. S., "Evaluation of Radar and Raingage Systems for Flood Forcasting", R. M. P. Lab. Report No. 138, M.I.T. (1971).

Green, J. R., "A Model for Rainfall Occurrence", J. Roy. Stat. Soc., Series B, V. 26 (1964), pp. 345-353.

Gupta, V. K., "Transposition of Storms for Estimating Flood Probability Distributions", Hydrology Paper No. 59, Colorado State University (1972).

Horton, R. E., "Frequency of Recurrence of Hudson River Floods", U. S. Weather Bureau Bull. Z (1913).

House, R. A., "Characteristics of Mesoscale Precipitation Areas", M. S. Thesis, M.I.T. (1969).

Kao, S., Duckstein, L., and Fogel, M., "A Probabilistic Model of Winter Rainfall", Paper Presented at A.G.U. Fall National Meeting, San Francisco, California (1971).

Kendall, G. R., "Statistical Analysis of Extreme Values", Proceedings of Symposium No. 1, Spillway Design Floods, National Research Council of Canada, Ottawa (1959).

Kisiel, C. C., "Transformation of Deterministic and Stochastic Processes in Hydrology", Proceedings, International Symposium, Fort Collins, Colorado (1967), pp. 600-607.

LeCam, L., "A Stochastic Description of Precipitation", IV Berkley Symposium on Mathematical Statistics and Probability, V. 3 (1961), pp. 165-186.

Loève, M., "Probability Theory", D. Van Nostrand Co., Inc. (1955).

Morgan, D. L., "A Cumulus Convective Model Applied to Thunderstorm Rainfall in Arid Regions", U. S. Corps. of Engineers, Research Note No. 1 (Dec. 1970).

Nash, J. E., and Amorocho, J., "The Accuracy of the Prediction of Floods of High Return", Water Resources Research, V. 2, No. 2 (1966), p. 191.

Osborn, H. B., "Schematic Representation of Thunderstorm Rainfall", U.S.D.A.-A.R.S. and S.C.S. Watershed Modeling Workshop, Tucson, Arizona (March 1970).

Parzen, E., "Stochastic Processes", Holden-Day Publishing Co., Inc. (1967).

Pattison, A., "Synthesis of Hourly Rainfall Data", Water Resources Research, V. 1, No. 4 (1965), pp. 489-498.

Rodriguez-Iturbe, I., Venmarcke, E. H., and Schaake, J. C., "Problems of Analytical Methods in Hydrologic Data Collection", Intl. Symp. on Uncertainties in Hyd. and Water Res. Systems, V. 2 (1972), pp. 451-472.

Sariahmed, A., and Kisiel, C. C., "Synthesis of Sequences of Summer Thunderstorms Volumes for the Atterbury Watershed in Tucson Area", Proc. IASH Symp. Use of Analog Digital Computers in Hydrology, V. 2 (1969), pp. 439-447.

Sorman, U. A., and Wallace, J. R., "Digital Simulation of Thunderstorm Rainfall", School of Civil Engineering, Georgia Institute of Technology (1972).

Takács, L., "On Secondary Stochastic Processes Generated by a Multidimensional Poisson Process", Address delivered at Colloquium on Stochastic Processes, Balaton, Sept. 13-15 (1956).

Todorovic, P., "A Mathematical Study of Precipitation Phenomenon", Eng. Res. Center, Colorado State University (1968).

Todorovic, P., "On Some Problems Involving Random Number of Random Variables", Ann. Math. Stat., V. 41, No. 3 (1970), pp. 1059-1063.

Todorovic, P., and Yevjevich, V., "Stochastic Process of Precipitation", Hydrology Paper No. 35, Colorado State University (1969).

Todorovic, P., and Zelenhasic, E., "A Stochastic Model for Flood Analysis", Water Resources Research, V. 6, No. 6 (1970), pp. 1641-1648.

Verschuren, J. P., "A Stochastic Analysis of Precipitation", Ph. D. Dissertation, Colorado State University (1968).

Weinstein, A. I., and Davis, L. G., "A Parameterised Numerical Model of Cumulus Convection", Report No. 11 to the National Science Foundation NSF-OA 777, Dept. of Meteorology, Pennsylvania State University (1968).

Weiss, L. L., "Sequences of Wet and Dry Days Described by a Markov Chain Probability Model", Monthly Weather Review V. 92 (1964), pp. 169-176.

Wiser, E. H., "Modified Markov Probability Models of a Sequences of Precipitation Events", Monthly Weather Review, V. 93 (1965), pp. 511-516.

Woolhiser, D. A., and Schwalen, H. C., "Area-Depth Frequency Relations for Thunderstorm Rainfall in Southern Arizona", Univ. Ariz. Agr. Exp. Sta. Tech. Rpt. No. 567 (1959).

Woolhiser, D. A., and Todorovic, P., "A Stochastic Model of Sediment Yield for Ephemeral Streams", Paper Presented at Symp. on Statistical Hydrology, Intl. Assoc. for Statistics in Physical Sciences, Tucson, Arizona (1971).

Yarnell, D. P., "Rainfall Intensity-Frequency Data", Department of the Interior, Miscellaneous Publication No. 204, Washington (1935).

Zelenhasic, E., "Theoretical Probability Distributions for Flood Peak", Hydrology Paper No. 42, Colorado State University (1970).