TR-148
1989

Stochastic Modeling of Rainfall-Runoff Process for Nonpoint Source Pollutant Load Estimation

M.A. Collins
R.O. Dickey

Texas Water Resources Institute

Texas A&M University

RESEARCH PROJECT COMPLETION REPORT

STOCHASTIC MODELING OF THE RAINFALL RUNOFF-PROCESS FOR NONPOINT SOURCE POLLUTANT LOAD ESTIMATION

Project Number - 02

(September 1, 1987 - August 31, 1989)

Grant Numbers

14-08-0001-G1451
14-08-0001-G1592

by

Michael A. Collins
Roger O. Dickey

The research on which this report is based was financed in part by the U.S. Department of the Interior, Geological Survey, through the Texas Water Resources Institute.

Technical Report No. 148
Texas Water Resources Institute
Texas A&M University
College Station, Texas 77843-2118

December, 1989

STOCHASTIC MODELING OF THE RAINFALL-RUNOFF PROCESS FOR NONPOINT SOURCE POLLUTANT LOAD ESTIMATION

ABSTRACT:

A stochastic simulation methodology was developed for the rainfall-runoff process to assist in the assessment of nonpoint source pollutant loads, particularly for ungaged watersheds where there is a scarcity or complete lack of historical data. The methodology was developed based on simulating individual rainfall-runoff events. A simulation model employed a rainfall simulator to stochastically generate rainfall event characteristics for input into basin hydrologic transformation functions which then predicted the corresponding runoff hydrography characteristics.

Also addressed was the impact of limited data availability on the ability to model the rainfall-runoff process. An evaluation was conducted to the degree to which committing valuable resources to expand the data base would provide measurable improvement in model results. Specifically, the probability of achieving certain levels of accuracy with the simulation model was statistically assessed as a function of the number of observed rainfall-runoff events used for model development. The probability of monitoring various numbers of rainfall-runoff events in specified time intervals was also established as an aid for planning field monitoring studies.

The simulation methodology was applied to a study watershed in the Lake Ray Hubbard reservoir drainage basin near Dallas, Texas. Regional rainfall characteristics were established using historical hourly data from the Federal Aviation Administration rain gage at Love Field Airport in Dallas, Texas. Hourly rainfall data were resolved into individual rainfall events and probability density functions were identified for event volume, time between events, and event duration. Linear hydrologic transformation functions were derived and incorporated into the simulation model by applying a unique stepwise least squares optimization procedure using observed data from the study watershed. Both total direct runoff and peak runoff rate were shown to be functions of rainfall event volume and a white noise component. Verification of the model was

achieved by statistically demonstrating that long-term simulation results and observed field data were drawn from the same underlying population.

ACKNOWLEDGEMENT

The work performed herein was sponsored in part by the City of Dallas Water Utilities Department, Research Division; the Texas Water Resources Institute of Texas A&M University; and Southern Methodist University. The opinions expressed herein are the opinions of the authors and should not be considered as the opinions or views of the sponsoring entities.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

CHAPTER 1
INTRODUCTION

Purpose

Effective and maximum protection and use of surface water sources is essential to long term resolution of the water supply needs of the nation, particularly for many urbanized areas which may rely almost exclusively upon surface water systems or which may be encountering limits upon their groundwater use. Nonpoint source (NPS) pollution can pose a significant threat to the long term quality and use of surface lakes and reservoirs for municipal and industrial water supply. Effective control is considerably hampered by the inability to adequately quantify NPS pollution because of its inherent complexity and the consequent time and financial resources that become necessary to adequately measure and monitor NPS loads in a particular region or watershed.

The rainfall-runoff process is the driving force behind generation of NPS pollution. Rainfall and subsequent runoff provide the energy and medium whereby soil is eroded and transported to surface waters as suspended sediment. Suspended sediment is the primary NPS pollutant (U.S. Environmental Protection Agency 1973) and it produces the high turbidity levels that characterize runoff events. Turbidity reduces light penetration in the water column and, thus, can decrease primary productivity. Furthermore, suspended sediment can be directly harmful to aquatic animals because the sediment particles can interfere with respiratory processes and cause abrasion damage

to the soft parts of sensitive tissues. Upon sedimentation, the particles can destroy nesting sites of fish, smother fish eggs, and alter habitats of benthic flora and fauna (Farnworth 1979).

Another, even more important, problem associated with suspended particles is their role in transporting nutrients and agricultural chemicals into surface waters. Phosphorus and nitrogen adsorbed to the suspended matter can induce excessive growth of algae and other aquatic plants. Metabolic by-products produced by these plants can impart objectionable tastes and odors to the water.

Agricultural organic chemicals, such as pesticides, also tend to be adsorbed to soil particles. The rainfall-runoff process and resulting soil erosion can result in the transport of these toxic chemicals into surface waters (U.S. Environmental Protection Agency 1973).

This research seeks to develop a statistically based simulation methodology for the rainfall-runoff process to assist in the assessment of NPS pollutant load generation in small watersheds that are typified by sporadic and ill-defined runoff and by limited and difficult to obtain field data. By describing the rainfall-runoff behavior in probabilistic terms, variability in pollutant loads can be assessed with regard to the need for additional field data in view of the intended use of pollutant load estimates and the likely increase in the certainty of pollutant load estimates that may or may not be expected through the development of additional field data. The Monte Carlo based stochastic simulation strategy of this research provides a means to describe the expected runoff volumes and flow rates, and then ultimately the pollutant loads, in a probabilistic fashion as a function of primary driving forces and constraints while incorporating varying levels of information about those forces and constraints.

Background

Nonpoint Source Pollutant Load Assessment

Even though stochastic modeling of the rainfall-runoff process is the specific objective of this research, it is to be tailored in such a way as to facilitate NPS pollutant load estimation. Therefore, it seems appropriate to review briefly current methods for NPS load assessment. There are three current strategies for the estimation of NPS pollution generation in a given watershed:

(1) Average regional or national scale data, such as that developed in the National Urban Runoff Program (Niedzialkowski and Athayde 1985, U.S. Environmental Protection Agency 1983) are used as surrogates for site specific data;

(2) Site specific data are collected to deduce basically empirical estimates of the NPS loads; and

(3) A deterministic model of the rainfall-runoff process coupled with empirically oriented pollutant load generation functions is used to attempt to deterministically simulate the NPS pollution generation process.

Use of average regional or national scale data, even when enhanced by identification of statistical correlations such as those currently under investigation by the USGS (Driver and Lyston 1986), does not identify the mechanisms at work in the NPS pollution process: such identification is seen by some (Ellis 1986; Sonnen 1986; Terstriep, Noel, and Bender 1986) as crucial to reliable prediction of NPS pollution. Average or statistical correlations do little to explain why NPS loads do in fact assume certain magnitudes. Furthermore, empirical NPS pollution data and relations not having their origin in actual data from a watershed under study will often be suspect because of the recognized

and considerable variability that such data demonstrate from one site to the next and even from one event to the next at the same site (Roesner 1982).

On the other hand, comprehensive and reliable site specific data bases for NPS pollution are difficult to develop. Site specific field studies are extremely costly and, since NPS pollution is strongly rainfall event dependent, sampling programs are logistically difficult and/or equipment intensive (Jennings 1982). Consequently, there is always difficulty in funding and conducting a field study of sufficient scope to develop data of the required detail (Huber et al 1979, Huber 1986).

Existing simulation models, such as SWMM (Huber et al. 1977), CREAMS (U.S. Department of Agriculture 1984), SWRRB (Arnold et al. 1986, Arnold and Williams 1985), and others (U.S. Army Corps of Engineers 1974, Marsalek 1986, U.S. Environmental Protection Agency 1971) can be effective in realistically estimating individual event NPS pollution, but they are very data intensive in their requirements for calibration and verification. A sizeable site-specific data base involving meteorologic, hydrologic, or hydraulic conditions is essential (Brown 1975, Huber 1986, Urbonas and Roesner 1986). Development of such data bases may require more time and financial resources than available. Furthermore, and importantly, such simulation models attempt to describe the generation of NPS pollution in a deterministic fashion: no specific attempt is made to treat key factors as being inherently random and thus incapable, at least from a practical standpoint, of a deterministic description.

Recognition of the inherent uncertainty in NPS pollution generation due to the underlying stochastic nature of the rainfall-runoff process may provide a more direct solution to development of the level and type of information that is needed in certain applications. Despite the obvious limitations of a

deterministic view of the NPS pollution problem, only limited attention has been given to nondeterministic modeling of NPS pollution (Hemain 1986, Marsalek 1986, North Central Texas Council of Governments 1984, U.S. Department of Agriculture 1985) and that has focused upon statistical modeling rather than stochastic modeling, that is, use of statistical correlations or frequency curves as the basis for deducing relations among factors influencing NPS pollution rather than constructing a model incorporating random processes (Driver and Lystrom 1986, Mancini and Plummer 1986, Medina 1979, Roles and Jonker 1985). Such statistical models do not allow interacting factors to adequately describe NPS load variability since the relations used are ones which represent, effectively, a statistical smoothing.

A need exists for new strategies that address the fundamental stochastic nature of NPS pollutant generation due to the underlying stochastic nature of the rainfall-runoff process.

Stochastic Rainfall-Runoff Modeling

Information about stream flow rates is essential for the analysis and design of water resource systems and NPS pollution management strategies. Stream flows are subject to the fundamental uncertainties of natural hydrologic phenomena and, as such, are the result of a mechanism with underlying random or stochastic components. Therefore, in order to address the reliability of water resource systems, including NPS pollution management systems, and to assess the risks of failure to achieve system objectives, it is necessary to describe the probabilistic nature of stream flow rates.

The basic data used to generate the statistics to describe the probabilistic nature of stream flows comes from historical stream gaging records. These records are likely to span a relatively short period of time, commonly less than

25 to 50 years, and can be thought of as only one of the infinite number of possible realizations of the underlying stochastic process. Thus, future sequences of stream flows are likely to differ significantly from the observed historical sequence. A common technique is to extend the historical record to a longer sequence, or multiple sequences of a given duration, by synthesizing or generating stream flow data using time series stochastic modeling techniques.

Some of the first applications of time series modeling techniques for stream flow synthesis were conducted by Fiering (1962), Matalas (1967), and Fiering and Jackson (1971). Since this early work, a vast body of literature has been generated on the subject. There are several excellent recent references that summarize the theory and application of time series analysis in hydrology (Bras and Rodriguez-Iturbe 1985; El-Shaarawi and Esterby 1982; Hipel 1985; Kottegoda 1980; Loucks, Stedinger, and Haith 1981; and McCuen and Snyder 1986). Additional theoretical development of time series analysis can be found in the classic work by Box and Jenkins (1970) and in Bhat (1984). Time Series models of the autoregressive (AR), moving-average (MA), autoregressive moving-average (ARMA), and autoregressive integrated moving-average (ARIMA) types have been used to generate synthetic stream flow sequences for application to the optimization of designs and development of operating strategies for water resource systems. These models require a record of historical streamflows spanning at least 20 to 30 years.

Unfortunately, many practical problems of great interest deal with watersheds where historical stream gaging data are limited or non-existent. In addition, the statistics of stream flow can be altered greatly by anthropogenic changes in land cover, land use, stream flow regulation, and groundwater use. These factors can render the historic sequence of flows almost useless for predicting future events or, at least, make questionable the stationarity

assumption typically invoked when performing time series analysis on hydrologic data.

Where stream flow data are available but limited (a minimum of 5 years of data is usually required) and long-term historical climatological data are available, time series stochastic modeling of the rainfall-runoff process has sometimes been used. The usual procedure is to apply a linear transfer function model with a noise component for the residuals. White noise and autocorrelated noise components have both been used. The theory of linear transfer function-noise models has been discussed in detail by Box and Jenkins (1970). The transfer function typically addresses the cross-correlation between stream flow at some desired point as the response variable with stream flow at upstream points and/or rainfall within the watershed as input time series (Chang, Kavvas, and Delleur 1982; Mimikou and Rao 1982; and Sharma 1985). The transfer function sometimes involves cross-correlation between stream flow as the response variable with rainfall and antecedent rainfall effects as inputs (Caroni, Mannocchi, and Ubertini 1982), or with rainfall at multiple gage sites and temperature at multiple sites as inputs (Hipel, McLeod, and Noakes 1982; Thompstone, Hipel, and McLeod 1985; Tong, Thanoon, and Gadmundsson 1985).

For ungaged watersheds where there is a complete lack of historical streamflow data, or where recent non-stationarity has been induced by the activities of man, a common alternate procedure is to estimate the statistical characteristics of stream flow rates using a suitable deterministic rainfall-runoff model to simulate the watershed response to a stochastic rainfall input. The stochastic properties of the rainfall input are estimated from long-term historical precipitation data collected at a rain gage in the same geographic and climatic

region as the study watershed. This is the approach to be used in this work where a small, ungaged watershed is studied.

Much work on the stochastic-deterministic modeling of the rainfall-runoff process has focused on the theoretical development of probability density functions (pdf's) for runoff volume given some simple assumed pdf's for pertinent rainfall characteristics. Simplified basin transfer functions are used in this type of analysis to facilitate the theoretical computations.

A common technique has been to model the basin response as a single linear reservoir (Frind 1969, Klemes 1974, and Singh and Birsoy 1977). Other researchers have extended this work by using a cascade of linear reservoirs (Tarboton, Bras, and Puente 1987). The development of linear reservoir theory for simulation of basin storage effects has been discussed in detail by Chow (1964) and Singh (1988).

Linear reservoirs are used to simulate only the time lag associated with basin storage effects. This approach does not deal with the complex task of estimating runoff volume when given the input rainfall volume, which involves consideration of the many basin hydrologic factors that influence the transformation of rainfall to runoff. These factors include interception, evapotranspiration, infiltration, and accretion of soil water to the deep groundwater, all of which abstract or remove some rainfall before it can run off. Some researchers have avoided dealing with the complex hydrologic abstraction process by assuming that some simple pdf is known directly for the rainfall excess (Frind 1969, Klemes 1974, and Singh and Birsoy 1977), or by just ignoring the abstraction altogether and acknowledging the serious limitations this imposes on applying their theoretical work to practical stream flow synthesis problems (Tarboton, Bras, and Puente 1987).

Another technique employed to simulate basin response to allow theoretical development of pdf's for runoff volume is the application of kinematic wave theory for hydrograph forecasting as applied by Eagleson (1971). Rather than completely ignore rainfall abstraction, Eagleson assumed a very simple constant rainfall abstraction rate to make the theoretical calculations tractable. He assumed that rainfall event duration was exponentially distributed and that event rainfall volume possessed an exponential pdf conditional on duration. Time between storms was assumed to be independent of volume and duration and assumed to be exponentially distributed.

Eagleson (1972) used the same general approach to derive theoretical pdf's for peak flood flow rate. Here he assumed rainfall volume and duration to be jointly exponentially distributed to facilitate his calculations. Similar work was carried out by Bras and Chan (1978) to derive pdf's for runoff volume above some given threshold value.

Eagleson (1978) extended his earlier theoretical work by using the Philip infiltration equation to estimate rainfall abstraction coupled with kinematic wave theory to simulate basin lag, and simplifying assumptions about the pdf's of rainfall event random variables.

A conceptual model using the Horton infiltration equation to simulate hydrologic abstractions and single linear reservoirs to simulate storage effects for both surface flow and subsurface flow has been developed by Koch (1985). Because of the increased complexity over Eagleson's models (1971, 1972, and 1978), Koch was unable to derive pdf's but derived relatively complex theoretical equations for the statistical moments of the stream flow distribution. Koch assumed that time between rainfall events, average rainfall intensity, and rainfall event duration were independent random variables with each

possessing a simple exponential distribution in order to make possible his calculations.

Other work has been done on ungaged watersheds using the U.S. Soil Conservation Service (SCS) runoff curve number method and other simple conceptual models to simulate basin response to a stochastic rainfall input. Barton (1974) used a two-stage conceptual model of watershed response to generate monthly stream flows from a stochastic monthly rainfall input for British catchments. Hanson and Woohiser (1978) developed a rainfall-runoff Monte Carlo simulation model where daily rainfall volume was assumed to possess an exponential distribution conditional on the occurrence of rainfall. The occurrence of rainfall was simulated by a simple two-state Markov chain. Then, the runoff coefficient (i.e., the ratio of runoff volume to rainfall volume) was assumed to possess a beta distribution. The model was used to evaluate the potential impacts of weather modification on stream flows in western South Dakota. Thames (1984) used a stochastic model of daily precipitation coupled with the SCS runoff formula to convert rainfall to runoff and an autoregressive evaporation model to evaluate the performance of small water impoundments in Arizona.

Most stochastic modeling of the rainfall-runoff process has been based on discrete time intervals for annual (Frind 1969, Klemes 1974, Singh and Birsoy 1977), monthly (Barton 1974; Hipel, McLeod, and Noakes 1982; Sharma 1985), weekly (Thompstone, Hipel, and McLeod 1985), or daily (Hanson and Woolhiser 1978; Mimikou and Rao 1982; Morris, Hood, and Ferguson 1984; Thames 1984; Tong, Thanoon, and Gudmundsson 1985) runoff volumes. Even the considerable body of literature dealing with stochastic rainfall modeling has dealt almost exclusively with a time interval no shorter than one day (Austin and Houze 1972; Caskey 1963; Chang, Kavvas, and Delleur 1984; Feyerherm and

Bark 1967; Gabriel and Neumann 1962; Green 1965; Kavvas and Delleur 1981; Weiss 1964; Wiser 1965).

Some work has been done with discrete time divisions of less than one day primarily for short-term, real-time flood forecasting. Hourly time steps were used by Caroni, Mannocchi, and Ubertini (1982) and 6-hour time steps were used by Georgakakos (1986).

A continuous time scale is most appropriate for rainfall-runoff modeling for NPS pollution load estimation where rainfall data is characterized on an individual event basis. Characterization of individual rain events was the approach used by Eagleson (1971, 1972, 1978), Koch (1985), Bras and Chan (1978). Basically, a daily time step is too long for accurate NPS pollution load estimation because it is the characteristics of individual rainfall events, such as intensity and duration, that control soil erosion losses from a watershed. On the other hand, an hourly discrete time step is too short for efficient generation of long-term sequences of synthetic stream flows because of the great preponderance of zero values in hourly rainfall records.

This research employs a continuous time approach by characterizing rainfall on an individual event basis.

Objectives

Stochastic Model Development and Use

The first objective of this research is the structuring of a stochastic simulation model of the rainfall-runoff process as the driving force for generation of NPS pollutant loads. The magnitude of NPS loads is strongly dependent on the nature of individual rainfall-runoff events so that the model will be designed to conduct Monte Carlo simulations for the evaluation of runoff statistics on an event basis.

On the other hand, it is the cumulative NPS pollutant load, particularly nutrients, that develops over a long period of time, on the order of months or years, that is of particular concern to the protection of the water quality of surface water supplies (Dallas Water Utilities 1977, Herricks 1986, Huber 1986, Marselek 1986). This results from the integrating dynamics of large bodies of water. Sufficient information on the impacts of pollutant loads can often be provided by a statement on the probability that over a period of time the NPS load will lie within certain ranges. A probabilistic statement about such cumulative loads can, in fact, be of more utility because it provides not only an estimate of average loads but an appropriate measure of the variability in loads that can be expected. Therefore, the model will also be structured to accumulate statistics over monthly and annual time periods.

Development of Methodology for Use of Limited Data Bases

Site specific field data available for use in defining strongly empirically based descriptions of NPS pollution or traditional calibration and verification of deterministically oriented simulation models will, in a practical sense, always be limited and commonly viewed as insufficient because of practical limits on time and resources to accumulate site-specific data. Expansion of a data base should, however, be viewed in the context of the additional information that can result from such expansion. This present research examines the application of a procedure to maximize the information on runoff variability that can be obtained from a given set of field data and demonstrate how that information is impacted by increases in data availability.

General Modeling Methodology

Monte Carlo simulation coupled with a limited field measurement program has been used in this research. A lumped parameter stochastic simulation model has been structured to describe the generation of runoff from a watershed, assumed to be small and of sporadic runoff, as a function of random inputs of rainfall. The field data collection effort was undertaken to provide real data for demonstration of the stochastic simulation methodology and the technique proposed for dealing with limited data availability. The Monte Carlo simulation strategy has been used to deduce rainfall-runoff event statistics and probability distributions.

The lumped parameter event oriented simulation model is intended to be reasonably simple and parsimonious (i.e., minimal number of model parameters) since the research focuses upon the stochastic features of rainfall-runoff and is intended for use with limited data bases. Runoff volume is estimated as a function of random variables characterizing rainfall events; volume, duration, time between events. Antecedent rainfall volume also plays a role, but it is calculated by accumulating preceding rainfall volumes.

Regionally representative probabilistic descriptions of rainfall event volume, duration, and time period between events can be deduced from available precipitation station data (North Central Texas Council of Governments 1984, 1978; Court 1979). These random variables, because of the controlling effects of short term infiltration, longer term percolation to the groundwater, and evapotranspiration on runoff production, are crucial to NPS pollution simulation (Hemain 1986). These data are described by empirical or fitted theoretical distributions which are randomly sampled to generate event rainfall magnitudes, durations, and inter-event times for Monte Carlo experimentation.

The simulation model incorporates parameters quantifying the probability distributions of the random variables in the model. These parameters will be determined through a combination of regionalized estimates and site specific calibration. Identification of parameters is feasible since the number of parameters in the lumped model is limited. These distributions then become the basis for Monte Carlo simulation of rainfall-runoff events.

CHAPTER 2

FIELD DATA COLLECTION AND ANALYSIS

Field Methodology

Introduction

A field monitoring program was conducted for 13 months from June 1987 to June 1988 to collect rainfall, streamflow rate, and water quality data for a small watershed within the Lake Ray Hubbard drainage basin. An ideal monitoring site was found on a small tributary to Squabble Creek near the town of Rockwall, Texas which is on the east shore of Lake Ray Hubbard about 20 miles east of Dallas as shown in Figure 2.1. The watershed has an area of 298 acres. Land use was approximately 80 percent rural (predominately grassland) with the remaining area being residential. There were no known point source waste discharges within the study watershed. There was a U.S. Weather Service cooperative observer rain gage for daily rainfall measurement within the study area.

In order to well define the processes involved in the generation of runoff from rainfall, and to determine NPS pollutant loads as accurately as possible, it was decided early in the planning of this study that automatic and continuous measurement of rainfall, streamflow rate, and a suitable indicator water quality parameter was necessary. Turbidity was chosen as the indicator water quality parameter because it is directly related to the total suspended solids (TSS) concentration, and because relatively inexpensive instrumentation for continuous turbidity measurement is readily available from several

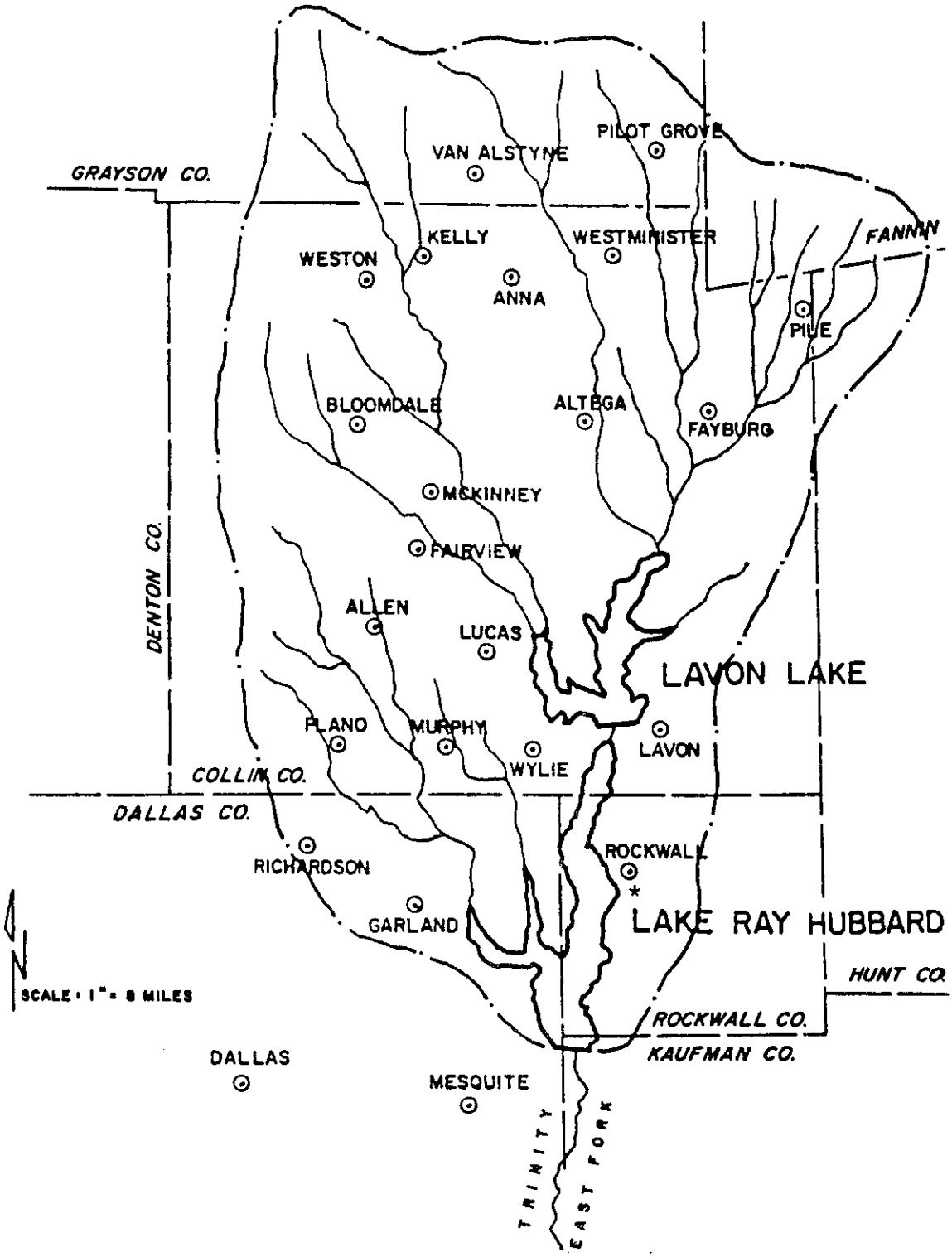


FIGURE 2.1
SITE LOCATION MAP

manufacturers. Suspended matter is of primary importance for the analysis of NPS pollutant loads since most of the nitrogen, phosphorus, and pesticides from nonpoint sources are associated with particulate matter as discussed in Chapter 1.

Automatic Monitoring Station

An innovative automatic stream monitoring station was designed and installed at the outlet of the study watershed on the property of the Rockwall County YMCA. Electrical power was readily available and only about 300 feet of direct burial electrical cable was required to bring power to the equipment enclosure. Power was then distributed to the various pieces of equipment through a breaker panel and 110 volt receptacles.

Overall system configuration is shown schematically in Figure 2.2. The primary components of the monitoring station were as follows:

1. Compound Weir - A sharp-crested compound weir was installed in a section of stream with almost vertical sides; the stream cross-section was roughly trapezoidal with a bottom width of approximately 7 feet, a top width of approximately 10 feet, and a height of approximately 8 feet. The weir consisted of a rectangular portion with a 7.25 foot top width for measurement of high flows and a 90° V-notch portion with a total height of 1 foot for low flow measurement.
2. Flow meter - An ISCO Model 2870 bubbler-type open channel flow meter was installed for continuous flow rate measurement. It included its own air compressor, chart recorded, and microprocessor to convert head measurements directly into flow rate. It also had a port for an ASCII digital output signal. Flexible 1/8-inch plastic tubing was used

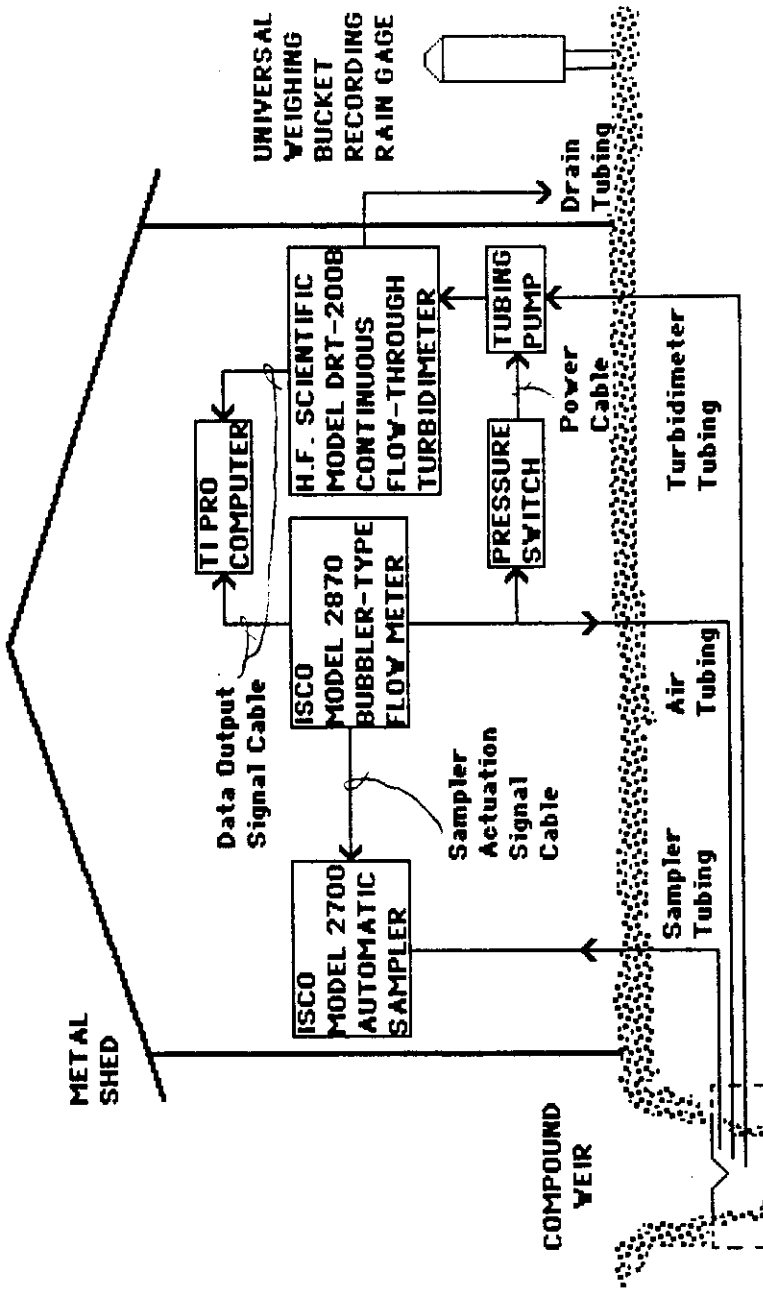


FIGURE 2.2
AUTOMATIC MONITORING STATION SCHEMATIC

to convey the air to the outlet point in the stream where it was discharged at the same elevation as the vertex of the V-notch weir.

3. Automatic Sampler - An ISCO Model 2700 automatic sequential sampler with 24 1-liter bottles was installed to collect water samples. The sampler was equipped with a self-priming tubing pump, a suction line of 1/4-inch flexible plastic tubing, and a programmable controller that provided flexibility in choosing sampling frequency. It could also interface with the flow meter to allow flow paced sampling. For this study, samples were collected on a sequential basis with the frequency of sample collection paced by the flow meter at a rate of one sample per 5000 cubic feet of runoff.
4. Turbidimeter - An H.F. Scientific Model DRT-200B continuous flow-through turbidimeter was installed for turbidity measurement. The unit had a flow-through module equipped with a glass cuvette, light source, and photocells to measure the light scattered at 90° from the light source. The instrument read directly in NTU and produced a 4-20 ma analog output signal.
5. Tubing Pump - A self-priming tubing pump was installed to deliver a continuous 1 gpm flow of sample from the stream to the turbidimeter. Flexible, 1/4-inch, plastic tubing was used with the pump. The pump was actuated when there was runoff in the stream by a low-pressure pressure switch installed on the flow meter bubbler line. The pump started when the depth above the vertex of the V-notch weir was about 4 inches, which corresponds to a flow rate of 0.156 cfs.
6. Computer - A Texas Instruments Professional Computer was installed to act as a sequential data logger using output signals from the flow meter and turbidimeter. An analog-to-digital converter was added to

the computer to handle the analog signal from the turbidimeter. The data were stored in the field on floppy disks. The data were then read directly from the field diskettes into a Lotus 1-2-3 spreadsheet file on a personal computer. The spreadsheet program was used as a convenient tool to store, analyze, and plot the observed data.

7. Recording Rain Gage - A Universal weighing bucket-type recording rain gage was installed at the monitoring site. Rainfall data were recorded continuously on a clock driven paper chart in order to provide detailed information on rainfall rates and rain event timing. The gage was not operational until July 2, 1987, consequently the first nine rainfall events of the study lack recording rain gage data.

A lockable metal building was erected adjacent to the weir installation to house the equipment.

The site was visited at least once weekly to calibrate and maintain the equipment. Water samples collected by the automatic sampler were analyzed by the Dallas Water Utilities Research Laboratory.

Data Analysis

Runoff Data

Raw flow rate data were input directly into a spreadsheet file on a personal computer for convenient manipulation and analysis. Hydrographs for all 45 runoff events monitored during the field study were plotted using the spreadsheet software as shown in the top graph of Figures A.1 through A.45 in Appendix A.

The first task in analysis of the hydrographs was to develop and apply a technique for separation of base flow from direct runoff. The point at which the flow rate first increased above previous base flow rates was easily identified as

the start of direct runoff. Then, based on a qualitative and quantitative review of all hydrographs in conjunction with the corresponding continuous turbidity measurements (turbidities tended to be high during direct runoff and low during base flow as expected), it was decided to define the end of direct runoff for all hydrographs as the point on the recession limb where the slope of the hydrograph first dropped below 0.10 cfs per hour. A straight line connecting the points at the start and end of direct runoff was then used for hydrograph separation.

Numerical integration of the area between the hydrograph and the base flow line yields the total direct runoff volume for the event. The area below the line is the total base flow volume during the event.

An analysis of all hydrographs produced the detailed runoff event data shown in Table A.1 in Appendix A. Table A.1 contains the following information:

(1) date

(2) total direct runoff volume, Q (inches of equivalent rainfall)

(3) peak runoff rate, q_p (cfs), defined as the difference between the hydrograph peak flow rate and the base flow rate at the time of the hydrograph peak

(4) time to peak, t_p (hours), defined as the time from the start of direct runoff to the peak runoff rate

(5) volume to peak, V_p (inches of equivalent rainfall), defined as the volume of direct runoff between the start of direct runoff and the peak runoff rate.

There are 44 runoff events listed in Table A.1, instead of 45, because the events of June 17, 1987 and June 18, 1987 (shown in Figures A.7 and A.8 in Appendix A) have been combined and treated as a single event starting on June 17. This was necessary because they occurred closely spaced in a way

that prevented resolution of the rainfall data from the daily gage into two separate events. The recording gage was not yet operational to assist in the resolution of the events. Notice, also, that the runoff volumes are missing for the three events of November 16, 1987; December 26, 1987; and December 27, 1987. These data were omitted because of erroneous flow rate measurements due to clogging of the weir by debris. This leaves a total of 41 events for which runoff volume data are listed.

An evaluation of hydrographs in Figures A.1 through A.45 indicates that events with a runoff volume less than 0.01 inch tend to possess unusual and highly variable hydrograph shapes. It is believed that this is a consequence of only a portion of the watershed contributing runoff to these events. Runoff events above the 0.01 inch cut-off generally possess classical hydrograph characteristics. It is believed that these events receive runoff from all parts of the watershed. Detailed statistical analyses performed in Chapter 4 confirm that events with volumes greater than or equal to 0.01 inch produce more consistent model results. Therefore, runoff events of 0.01 inch and larger are designated as significant events. There are 28 such events listed in Table A.1.

Rainfall Data

Daily rainfall data were available from the U.S. Weather Service gage at Rockwall spanning the entire field study. The recording gage became operational on July 2, 1987 so that rainfall intensity data are only available after that date. Recording rain gage data were compiled in hourly increments to simulate the form of historical hourly rainfall records. Historical hourly records are described in detail and definitions of pertinent rainfall event variables are given in Chapter 3.

Detailed rainfall event field data are shown in Table A.2 in Appendix A. Table A.2 contains the following information:

(1) date

(2) total direct runoff volume, Q (inches of equivalent rainfall)

(3) recording rain gage data

  (a.) total event rainfall volume, R_r (inches)

  (b.) duration of the event, D (clock hours)

  (c.) maximum hourly rainfall intensity, I_{max} (inches/hour)

  (d.) weighted hourly average rainfall intensity, I_w (inches/hour)

  (e.) time between rainfall events, T (clock hours)

(4) daily gage rainfall event volume, R_d (inches)

(5) area weighted average rainfall event volume, R_w (inches)

(6) elapsed time since last runoff event, T_l (clock hours), actually computed as the elapsed time since the end of the last rainfall event that produced runoff

(7) elapsed time since the last runoff event with $Q \geq 0.01$ inch, T_r (clock hours), actually computed as the elapsed time since the end of the last rainfall event that produced $Q \geq 0.01$ inch.

All rainfall data collected are included in Table A.2 regardless of whether both gages recorded rain for a given event and regardless of whether the rainfall event produced runoff. Runoff volumes were repeated in Table A.2 for easy identification of the rainfall events that did, in fact, produce runoff (i.e. $Q>0$). The units "clock hours" refers to the integer number of hours in a continuous sequence that begins on an exact clock hour (i.e. 01:00, 02:00, 03:00, etc.) and ends on an exact clock hour.

When recording rain gage data were not available, the elapsed time variables T, T_l, and T_r were estimated from an analysis of the runoff

hydrographs; the timing between rainfall events was approximated by the time from the peak of the previous hydrograph to the start of runoff for the current event. All or part of the recording rain gage data are missing for the eight events prior to July 2, 1987 and for December 13, 1987 when the total rain volume was measured but intensity data were missed because the rain gage clock malfunctioned. Also, elapsed time variables T_l and T_r are missing for the runoff event of February 17, 1988 because several preceding precipitation events were produced by snow and ice which made it impossible to relate precipitation to runoff. Fortunately, no further freezing weather occurred after February 17.

The area weighted average rainfall volume, R_w, was computed using weighting factors derived by the standard Thiessen polygon method:

(2.1) $$R_w = 0.393\, R_r + 0.607\, R_d\ .$$

Table A.2 contains a total of 82 rainfall events of which 40 possessed useable runoff data along with known values of R_w, T, T_l, and T_r. Of these 40 events, 27 produced significant runoff (i.e. $Q \geq 0.01$ inch). Of the 40 events, 9 lacked intensity data such that there were 31 events that possessed a complete set of known values of all rainfall and runoff variables. Of these later 31 events, 19 produced significant runoff.

Summing over the calendar year from July 1987 to June 1988, for which the complete set of rainfall-runoff event data were collected, yields a total weighted rainfall volume of 28.89 inches and a total runoff volume of approximately 1.49 inches.

Antecedent Rainfall Data

Values of nine different cumulative antecedent rainfall volume variables were computed (from the data in Table A.2 in Appendix A) for each rainfall event that produced runoff. The following general notation was used:

$_iR_j$ = rainfall volume occurring between the end of day i and the end of day j (inch).

The nine specific cumulative rainfall volume variables compiled were,

(1) $_0R_1$ = rainfall volume in the 24 hours (1 day) antecedent to the current event (inch)

(2) $_0R_2$ = rainfall volume in the 48 hours (2 days) antecedent to the current event (inch)

(3) $_0R_3$ = rainfall volume in the 72 hours (3 days) antecedent to the current event (inch)

(4) $_0R_4$ = rainfall volume in the 96 hours (4 days) antecedent to the current event (inch)

(5) $_0R_5$ = rainfall volume in the 120 hours (5 days) antecedent to the current event (inch)

(6) $_0R_6$ = rainfall volume in the 144 hours (6 days) antecedent to the current event (inch)

(7) $_0R_7$ = rainfall volume in the 168 hours (7 days) antecedent to the current event (inch)

(8) $_0R_{15}$ = rainfall volume in the 360 hours (15 days) antecedent to the current event (inch)

(9) $_0R_{31}$ = rainfall volume in the 744 hours (31 days) antecedent to the current event (inch)

These data are shown in Table A.3 in Appendix A. Only data for the 41 rainfall-runoff events for which the runoff volume was known (see Table A.2 in Appendix A) are shown.

Cumulative totals were employed so that a single index variable could represent the entire sequence of antecedent rainfall amounts. This approach was confirmed as acceptable by a preliminary analysis that demonstrated that an entire chronological series of daily antecedent rainfall volumes (i.e., ${}_0R_1$, ${}_1R_2$, ${}_2R_3$, ${}_3R_4$, ${}_4R_5$, ${}_5R_6$, ${}_6R_7$, ${}_7R_{15}$, and ${}_{15}R_{31}$) was less effective than a single cumulative rainfall total for prediction of runoff volume.

Water Quality Data

Liquid samples collected by the automatic sequential sampler were analyzed for the principal NPS pollutants: TSS, nitrogen, and phosphorus. Other miscellaneous chemical and biological laboratory tests were also performed. Raw laboratory results are not presented in this report since they were not explicitly used.

However, in order to provide some insight into the typical magnitude of NPS pollutant loads, the continuous field turbidity data and laboratory TSS results were used to estimate the TSS load for the rainfall-runoff events. To accomplish this, a correlation was developed between the TSS in the liquid samples to the corresponding field turbidity measurements. A linear regression analysis was conducted with TSS as the dependent variable and turbidity as the independent variable which produced the following relationship:

$$Y = 0.86\,X - 2.223 \tag{5.2}$$

where,

Y = TSS ($^{mg}/_{l}$)

X = turbidity (NTU).

The linear correlation coefficient was 0.97 which indicated an excellent linear relationship between TSS and turbidity as anticipated.

The regression equation of equation 5.2 was used, with the spreadsheet software, to convert the continuous field turbidity measurements into a continuous graph of TSS concentration versus time for each storm event monitored.

These are shown as the middle graphs in Figures A.1 through A.45 in Appendix A where field turbidity data were available.

Continuous graphs of TSS mass loading (in pounds of TSS per minute) were then computed from the hydrograph and TSS graph for each event. These TSS loading graphs are shown as the bottom graphs of Figures A.1 through A.45 where turbidity data were available. Numerical integration of the area under the TSS mass loading graphs yields the total TSS mass load (in pounds) for each event. These data are shown in Table A.4 in Appendix A. The maximum load from a single event was approximately 11,750 pounds for the event of April 17, 1988. The total load for the 13 month study (neglecting the events for which turbidity data were not available) was approximately 52,220 pounds. On an annual areal average basis this is approximately 162 pounds per acre per year.

The highly variable nature of the TSS concentration and mass loading rates shown in Figures A.1 through A.45 indicates that, for this watershed at least, continuous measurement of TSS concentrations is essential for obtaining accurate pollutant load estimates. A set of sequential samples would tend to

miss the high peaks of TSS concentration and mass loading and result in underestimation of the total load.

CHAPTER 3
HISTORICAL RAINFALL DATA ANALYSIS

Regional Rainfall Data

Runoff volumes and flow rates, and hence NPS pollutant loads, are related to specific rainfall event characteristics: rainfall volume, rainfall rate or intensity, event duration, and the spacing of sequential rainfall events in time. A statistical description of these pertinent rainfall event characteristics is indispensable for stochastic modeling of the rainfall-runoff process. Consequently, a long-term historical sequence of rainfall data for the study area is needed to facilitate estimation of the appropriate statistical distributions and parameters.

Rainfall data are readily available for numerous rain gage locations nationwide as published by the National Oceanic and Atmospheric Administration (NOAA) of the U.S. Department of Commerce. The vast majority of these data are obtained with rain gage instruments that are placed into two classifications based on the time interval over which rainfall volumes are totaled: daily gages and hourly gages. The data are published in summaries as calendar day totals for daily gages and as clock hour totals (i.e. midnight - 1:00 am, 1:00 am - 2:00 am, ..., 11:00 pm - midnight) by calendar day for hourly gages. The published data are obtained from rain gages operated by the U.S. Weather Service and the Federal Aviation Administration (FAA).

In order to statistically describe rainfall rates and event durations, information about the timing of rainfall volumes within individual events is necessary. This means that data obtained from hourly rain gages represent the

minimum level of detail that is generally applicable for stochastic modeling of the rainfall-runoff process. Daily gages yield only daily rainfall totals and, therefore do not define rainfall rate characteristics for events less than 24 hours in duration.

Published hourly rainfall records are available from NOAA through the National Climatic Data Center (NCDC) in Asheville, North Carolina. These records are available in paper copy by state, and in digital form on magnetic tape by state or on diskette by individual rain gage. It is also to be noted that rainfall data can be obtained from the NCDC for shorter totalizing time intervals in some cases. Intervals of 15 minutes are available for a very restricted number of rain gages. Intervals of 5 minutes can be obtained for an even smaller number of gages in the form of daily paper charts only. In order to make the methodology developed in this study as generally applicable as possible, it was decided to utilize hourly data records because hourly is the shortest time interval for which data are widely available.

The nearest hourly rain gage station, with a long-term record, to the study area is the FAA weather station at Love Field Airport in Dallas, Texas. The designation for this station in NOAA printed publications is "Dallas FAA AP." The distance from the rain gage at Love Field to the study area is only approximately 30 miles and there are no apparent differences in the geographical and climatological characteristics of the two sites. Therefore, it can be reasonably assumed that the FAA gage at Love Field adequately describes the rain event characteristics for the study area and the surrounding region.

Event Data From Hourly Records

Format of NOAA Hourly Records

Hourly precipitation records on diskette were obtained from the NCDC for the Love Field rain gage. A pair of diskettes with a total of 10 sequential data files contained the hourly data. A summary of the dates spanned by each file is shown in Table 3.1. Brief comments about any gaps in the data or changes in measurement technique are also shown in Table 3.1.

General information about the format of the data on the diskettes was also obtained (U.S. Department of Commerce 1985, 1986). A summary of the format is shown in Table 3.2. Each data line in the files contains a numeric code in field 1, called the line-type, which identifies the type of data contained in the remainder of the line. The first 5 lines of each file contain general descriptive information about the file and are identified with line-types 1 through 5, respectively. Lines 6 through the next-to-last line of each file are designated either line-type 8 or line-type 9 (there are no line-types 6 or 7). Line-type 8 identifies a line containing the date and data identification information. There is one line-type 8 for each day on which rainfall occurred during the record of each file. Days with zero rainfall are not included in the files. A line-type 8 is then followed by as many lines, each with line-type 9, as necessary to contain the hourly rainfall data for that day. There is one line-type 9 for each hour in which rain fell for that day, plus one additional line-type 9 for daily total rainfall. Hours with zero rainfall are not included in the files. The last line of each file, with line-type 10, identifies the end of the file.

An abbreviated sample of one of the files, HP12244D.PRN, is reproduced in Table 3.3. As described above, the first line identifies the record-type as hourly precipitation data, "HPD". The second line identifies the station name

TABLE 3.1

NCDC HOURLY PRECIPITATION DISKETTE FILES
LOVE FIELD IN DALLAS, TEXAS

File Name	Dates	Comments
HP12244A.PRN	11/1940 - 12/1944	Data recorded to the nearest 0.01 inch until noted otherwise.
HP12244B.PRN	1/1945 - 12/1949	Data is missing beginning 24:00 on 12/31/46; the last rain recorded before the gap occurred at 12:00 on 12/31/46.
HP12244C.PRN	1/1950 - 12/1954	Missing data ends at 01:00 on 10/1/47; the first rain recorded after the gap occurred at 02:00 on 10/8/47.
HP12244D.PRN	1/1955 - 12/1959	
HP12244E.PRN	1/1960 - 12/1964	
HP12244F.PRN	1/1965 - 12/1969	
HP12244G.PRN	1/1970 - 12/1974	Data is missing beginning 24:00 on 2/26/73; the last rain recorded before the gap occurred at 06:00 on 2/26/73.
HP12244H.PRN	1/1975 - 12/1979	Missing data ends at 01:00 on 3/1/75; the first rain recorded after the gap occurred at 16:00 on 7/2/75. Data recorded to the nearest 0.10 inch after 01:00 on 3/1/75 due to a change in instrumentation.
HP12244I.PRN	1/1980 - 12/1984	
HP12244J.PRN	1/1985 - 5/1987	

TABLE 3.2

NCDC HOURLY PRECIPITATION DATA FILE FORMAT

Line-Type	Line Format	Comments
1	1, "RECORD-TYPE"	RECORD-TYPE = "HPD" for hourly precipitation data.
2	2, "STATION NAME, STATE"	The city and state where the gage is located.
3	3, Begin STATION-ID #1, End STATION-ID #1, ...	Up to 5 sets of beginning and ending station-id numbers can be entered. Only one set is entered if the file contains data from one gage and the remaining 8 station-id numbers are zeros.
4	4, Begin YEAR, Begin MONTH, End YEAR, End MONTH	The beginning year and month and ending year and month for the data in the file.
5	5, "ELEMENT-TYPE", "ELEMENT-TYPE", ...	Up to 5 element-types can be entered, but when the file contains only hourly precipitation data, the first ELEMENT-TYPE = "HPCP" and the remaining four are blank characters (" ").
8	8, YEAR, MONTH, DAY, "ELEMENT-UNITS", NUM-VALUES	YEAR, MONTH, and DAY specify the date of the rainfall. ELEMENT-UNITS = "HI" for hundredths of inches, ELEMENT-UNITS = "TI" or "HT" for data reported in hundredths of inches but observed to the nearest tenth of an inch only. NUM-VALUES = number of hourly rainfall values, plus one value for the daily total, for this date.
9	9, HOUR, DATA-VALUE, "FLAG-1", "FLAG-2"	HOUR = military clock hour; DATA-VALUE = rainfall volume; FLAG-1 = "0" for acceptable data, any other character indicates a problem with the data; FLAG-2 = "0" is not used at this time and is always set equal to "0".

TABLE 3.3

NCDC HOURLY PRECIPITATION DATA FILE SAMPLE
File - HP12244D.PRN

Line No.	Listing of File	Comments
1	1, "HPD"	The first 5 lines indicate that this file contains hourly precipitation data for Station 412244 in Dallas, Texas from January 1955 through December 1959.
2	2, "Dallas, TX"	
3	3, 412244, 412244, 0, 0, 0, 0, 0, 0, 0, 0	
4	4, 1955, 01, 1959, 12	
5	5, "HPCP", " ", " ", " ", " ", " "	
6	8, 955, 1, 1, "HI", 2	January 1, 1955 is the first day of record and there was no precipitation on this date.
7	9, 1, 0, "0", "0"	
8	9, 25, 0, "0", "0"	
9	8, 955, 1, 5, "HI", 2	The first rain fell on January 5, 1955; 0.04 inch fell in the hour ending at 9:00 am. The daily total was 0.04 inch.
10	9, 9, 4, "0", "0"	
11	9, 25, 4, "0", "0"	
•	•	•
•	•	•
•	•	•
n-13	8, 959, 12, 31, "HI", 12	December 31, 1959 is the last day of record. Rain fell during 11 clock hours; the first hour ending at 1:00 am and the last hour ending at midnight. Hourly rain volumes ranged from 0.01 inch to 0.25 inch with a daily total of 1.13 inches.
n-12	9, 9, 1, "0", "0"	
n-11	9, 11, 6, "0", "0"	
n-10	9, 12, 23, "0", "0"	
n-9	9, 13, 25, "0", "0"	
n-8	9, 14, 15, "0", "0"	
n-7	9, 15, 25, "0", "0"	
n-6	9, 16, 13, "0", "0"	
n-5	9, 17, 1, "0", "0"	
n-4	9, 18, 2, "0", "0"	
n-3	9, 19, 1, "0", "0"	
n-2	9, 24, 1, "0", "0"	
n-1	9, 25, 113, "0", "0"	
n	10, "FILE-END"	End of file.

and state as Dallas, Texas. The third line indicates that all of the data in the file are from the same gage: beginning station-id #1 and ending station-id #1 are both 412244 (which identifies the gage at Love Field) and the remaining four sets of station-id numbers are zeros. These first three lines are identical for all 10 of the data files. The fourth line indicates that this particular file spans the time period from month 01 in 1955, that is January 1955, to month 12 in 1959, that is December 1959. This line is different for each of the 10 files to indicate the time period spanned by each individual file. The fifth line indicates that the first element-type present in the file is hourly precipitation data, coded as "HPCP", and the remaining 4 possible element types are not present in the file, coded as " ". This simply indicates that the file contains only hourly precipitation data. This line is identical for all 10 files.

The actual rainfall data begins with the sixth line of each file. There was no rainfall on the first day of record for the particular file shown in Table 3.3, but the first and last day of record is always included in each file regardless of the presence or absence of rain. The sixth line shown in Table 3.3 begins with a line-type designation of 8 indicating the start of a day. The next three elements of this line indicate the date. The 955 is the year with the leading 1 omitted (i.e., 1955) to save storage space, the second element, 1, indicates the month and the third element, 1, indicates the day (i.e., January 1). The next element in the line, "HI", indicates that the rainfall data for this date are in units of hundredths of inches. The final element, 2, indicates that there are 2 lines of line-type 9 immediately following which contain the hourly rainfall data for this date and the daily total rainfall.

In general, the last element of a line-type 8 ranges from 2 to 25. A value of 2 indicates that rain fell in only one clock hour of the current day and that the first line-type 9 that follows contains the data for that hour. A second line-type 9

then follows to give the daily total rainfall, which equals the hourly rainfall for a day with only one hour of rain. A value of 25 indicates that rain fell in all 24 hours of the current day and that 24 lines of line-type 9 follow with the hourly data. A 25th line-type 9 then follows with the daily total rainfall, which is just the sum of all hourly values. Obviously, a value between 2 and 25 indicates that rain fell in some given number of hours between the two possible extremes.

All lines with line-type 8 are in an identical format for all of the files.

The seventh line shown in Table 3.3 begins with a line-type designation of 9 indicating that it contains hourly rainfall data or total daily rainfall. The second element, 1, of this line indicates the end of the clock hour, in military clock time, covered by the current line. In this case, it is 01:00 or 1:00 am. The third element, 0, indicates the rain volume for this hour. As discussed previously, zero rainfalls are included only for the first and last days of the file. All others are omitted. The fourth element, "0", of this line is the flag indicating that the data is acceptable. Any other symbol in this location indicates that the data for the current hour is erroneous or missing. The fifth element, "0", is a flag not currently in use by the NCDC. Every line-type 9 of all 10 data files has the symbol "0" as the fifth element.

The eighth line of Table 3.3 begins with a line-type designation of 9. The second element of this line, 25, specifies the ending clock hour as 25:00, which is the flag used to signify that this line contains the daily total rainfall. The third element, 0, indicates that there was zero total rainfall on the current day. The fourth and fifth elements, both "0", are the data flags.

All lines of line-type 9 are in an identical format for all of the data files.

The remaining lines of the file, except for the last line, are line-types 8 and 9 and they contain the remainder of the hourly data and daily totals. As indicated in Table 3.3, the first rain of the file fell on February 5, 1955 and had a

duration of one clock hour with a rainfall amount of 0.04 inch. The last rain fell on the last day of the file: December 31, 1959. On this date, rain fell in 11 different clock hours between 1:00 am and midnight with hourly volumes ranging from 0.01 inch to 0.25 inch. The total for this date was 1.13 inches.

The last line of the file is of line-type 10 and contains only the character string "END-FILE" indicating the end of the current file. All 10 files end with this line.

Resolution of Hourly Data Into Separate Events .

In order to simulate the rainfall-runoff process, the NCDC data files must be resolved into individual storm events. The pertinent rainfall event variables are;

R = total event rainfall volume (inches)

T = time between events (clock hours)

D = duration of the event (clock hours)

I_{max} = maximum hourly rainfall intensity (inches/hour)

I_w = weighted hourly average rainfall intensity (inches/hour)

Total rainfall volume, R, can be obtained to the nearest 0.01 inch by simply summing the hourly values spanning the given event. Time between events, T, is an integer variable representing the number of clock hours with zero rainfall between two successive rainfall events. Duration, D, is an integer variable that represents the number of clock hours that the given event spans.

The maximum hourly rainfall intensity, I_{max}, and weighted hourly average rainfall intensity, I_w, are intended to characterize the event rainfall rate. Other intensity variables could be devised, but these two were selected because a maximum and average seem to be physically plausible indicators of intensity given the limitations of hourly rainfall data. Unfortunately, it is not possible to estimate instantaneous rainfall intensities from hourly records. Hourly data

represent a piecewise integration of the instantaneous rainfall intensity function to give rainfall volume in one hour increments. As such, the hourly rainfall amounts can be thought of as the average hourly rainfall intensity (inches/hour) over that hour. Thus, I_{max} is numerically equivalent to the maximum hourly volume of rain that fell during the storm event and is thought of as the maximum average hourly intensity for the event. The other intensity variable, I_w, is a simple weighted average intensity where each hourly intensity during an event is weighted by the fraction of the total event volume that fell during that hour. This can be stated as:

(3.1)
$$I_w = \sum_{j=1}^{D} i_j \left[\frac{r_j}{R} \right]$$

where,

i_j = average hourly intensity for hour j (inches/hour)

r_j = volume of rain that fell during hour j (inches)

$R = \sum_{j=1}^{D} r_j$.

Of course, R and D are as defined before. It should be noted that for hourly data, i_j and r_j are numerically equal so that this expression can be simplified to;

(3.2)
$$I_w = \sum_{j=1}^{D} \frac{\left[r_j / 1 \text{ hour} \right]^2}{R} \quad .$$

In order to separate the hourly data files into individual events, a continuous interval of time of fixed length with zero rainfall must first be specified as the separation criterion. A detailed evaluation of the hydrographs

and associated rainfall data presented in Chapter 2 revealed that hydrographs produced by closely spaced rainfall events could be resolved into separate runoff events only for intervening periods without rain of 3 clock hours or more. If the intervening dry period between occurrences of rain lasted for only 1 or 2 hours, the hydrographs did not fully recede to base flow conditions before the flow rate began to rise again. Therefore, a 3 clock hour intervening period between occurrences of rain was selected as the appropriate criterion for resolution of hourly rainfall data into separate events. Other researchers have also found 2 to 3 hours to be suitable for rainfall event separation (Eagleson, 1970).

As a consequence of the clockhour - calendar day - calendar month - calendar year format of the NCDC data files, it is somewhat difficult to separate the data into individual events. Difficulties arise because a single event can frequently span from one day to the next, fairly frequently from one month to the next, and occasionally even from one calendar year to the next. Also, erroneous or missing data due to equipment failures and recording errors can further complicate the separation process.

In order to overcome these problems, a carefully designed computer program for separation of the events was developed. The program, written in the Basic language, is listed in Table B.1 of Appendix B. The program requires an IBM compatible personal computer with a hard disk with Microsoft Basic installed. Upon execution of the program, the user interactively specifies the output sequential data file name for storage of the separated event data. The user then specifies the names of the NCDC input sequential data files, one at a time, in chronological order, when requested by the program. As the program is currently structured, the input data files and the program itself must be stored on the hard disk. The output file is also written to the hard disk.

As the program processes each of the input files, the date and time of the current hour of rain is compared to the date and time of the last hour of rain. If the intervening time is less than 3 hours, the current hour of rain belongs to the same rain event as the last hour of rain. If it is greater than 3 hours, then the current hour of rain represents the first hour of a new rain event.

In practice, the actual computation of R, D, I_{max}, and I_w are relatively simple such that most of the steps in the computer program deal with the task of determining the intervening time between each pair of successive hourly rainfall values. If the last hour of rain and the current hour of rain are in the same calendar day, then the intervening time is given by:

(3.3) $$t_{n+1} = (h_{n+1} - 1) - h_n$$

where,

- n = index variable indicating the total number of hourly rainfall values already included in the current event ($n > 0$)
- t_{n+1} = intervening number of hours between the ending clock hour of the last hour of rain and the beginning clock hour of the current hour of rain ($0 \leq t_{n+1} \leq 22$)
- h_n = ending clock hour of the last hour of rain ($1 \leq h_n \leq 23$)
- h_{n+1} = ending clock hour of the current hour of rain ($2 \leq h_{n+1} \leq 24$)
- $(h_{n+1} - 1)$ = beginning clock hour of the current hour of rain $[1 \leq (h_{n+1} - 1) \leq 23]$

Now, if the last hour of rain and the current hour of rain are in different calendar days, then t_{n+1} is given by:

(3.4) $$t_{n+1} = (h_{n+1} - 1) + c_{n+1} \text{ (24 hours/day)} + (24 - h_n)$$

where,

$(24-h_n)$ = number of hours from the ending clock hour of the last hour of rain to the end of the calendar day on which it occurred $[0 \leq (24-h_n) \leq 23]$

c_{n+1} = number of intervening days from the end of the last calendar day in which rain occurred to the beginning of the calendar day of the current rain ($c_{n+1} \geq 0$)

$(h_{n+1}-1)$ = number of hours from the beginning of the current calendar day to the beginning clock hour of the current hour of rain $[0 \leq (h_{n+1}-1) \leq 23]$.

The variable t_{n+1} is defined as before except that it no longer has an upper bound (i.e., $t_{n+1} \geq 0$) since the two successive hourly rains do not occur on the same calender day. The variables h_n and h_{n+1} are as defined before.

The number of intervening calendar days, c_{n+1}, is kept track of by computing the number of the day of the year (1 to 365, or 1 to 366 for a leap year) for each day of the record using subroutine YEARDAY of the program. Then, if the last day in which rain fell and the current day in which rain fell are in the same calendar year, c_{n+1} is given by:

(3.5) $$c_{n+1} = (d_{n+1} - 1) - d_n$$

where,

d_{n+1} = day of the year in which the current hour of rain fell ($1 \leq d_{n+1} \leq 365$ or 366 in a leap year)

d_n = day of the year in which the last hour of rain fell ($1 \leq d_n \leq 364$ or 365 in a leap year)

$(d_{n+1} - 1)$ = day of the year preceding the day in which the current hour of rain fell [$0 \leq (d_{n+1} - 1) \leq 364$ or 365 in a leap year].

The quantity $(d_{n+1} - 1)$ in this equation is necessary to satisfy the definition of c_{n+1} as the number of intervening days without rain.

Now, if the last hour of rain and the current hour of rain are in two successive calendar years, c_{n+1} is given by:

$$c_{n+1} = (d_{n+1} - 1) + (n_y - d_n) \tag{3.6}$$

where,

n_y = number of days in the last calendar year (n_y = 365 for a normal year, n_y = 366 for a leap year).

$(n_y - d_n)$ = number of days without rain from the end of the last day with rain to the end of the last calendar year [$0 \leq (n_y - d_n) \leq 364$ or 365 in a leap year].

$(d_{n+1} - 1)$ = number of days without rain from the beginning of the current calendar year to the beginning of the day with the current hourly rain [$0 \leq (d_{n+1} - 1) \leq 364$ or 365 in a leap year].

The definition d_n is the same as before except that it now has an upper bound of 365 or 366 since the current day of rain is in the next calendar year. The variable d_{n+1} is as defined before.

The first year of record, 1940, was a leap year. Therefore, since leap years occur once every 4 years, subsequent leap years can be identified as those years in which the following mathematical condition holds:

$$\text{INT}\left[\frac{(\text{YEAR} - 1940)}{4}\right] = \frac{(\text{YEAR} - 1940)}{4} \tag{3.7}$$

where,

YEAR = calendar year (e.g. 1944)

INT = a function that returns the largest integer that is less than or equal to the argument of the function.

Once the value of t_{n+1} is established, it is checked to see if it is less than 3 hours or greater than or equal to 3 hours. If it is less that 3 hours, then the current hourly rainfall is part of the current event and it is annexed to the set of all hourly rainfalls that make-up the current event. Data processing then proceeds to the next hourly rainfall for another iteration. If t_{n+1} is greater than or equal to 3 hours, the current hourly rainfall is the first hour of the next event. In this case, the values of the pertinent rainfall variables for the current event are computed and written to the output file. The variables are computed as follows:

(3.8) $$R = \sum_{j=1}^{n} r_j$$

(3.9) $$T = t_1$$

(3.10) $$D = n + \sum_{j=2}^{n} t_j$$

(3.11) $$I_{max} = \max (r_1, r_2, ..., r_j, ..., r_n)$$

(3.12) $$I_w = \frac{\sum_{j=1}^{n} (r_j / 1 \text{ hour})^2}{R}$$

where,

t_j = intervening number of hours between the ending clock hour of the (j-1)th hour of rain and the beginning clock hour of the j-th hour of rain of the current event.

t_1 = intervening number of hours between the ending clock hour of the last event and the beginning clock hour of the current event.

max = a function that returns the maximum value among the set of arguments.

All other variables are as defined previously. Obviously, the t_j are the set of (n-1) values of t_{n+1} previously computed for the current event. The last value of t_{n+1} computed for the current event was greater than 3 hours and, thus, it becomes t_1 for the next event.

When the NCDC files are processed in the manner described above, all data lines that include total daily rainfall (i.e., when the time is indicated as 25:00) or zero rainfall volume are appropriately skipped.

Other difficulties arise when there are erroneous or missing data in the NCDC files. Erroneous and missing data are flagged in the files as discussed previously. For an event with any erroneous or missing hourly data, the values of R, D, I_{max}, and I_w are indeterminate. In this situation, they are all set to a value of −1 as a flag to omit the event from later data analyses. The value of T is indeterminate if erroneous or missing data is present in the last hour of the last event and/or the first hour of the current event. In either case, T is set to −1 as a flag to omit the event from later data analyses. As a special case, the value of T is also set to −1 for the very first event of record since T cannot be determined for this event.

The computer program described above was applied to process the 10 data files for the Love Field rain gage received from the NCDC. A total of 3741 individual events occurred composed of 17,420 clock hour rainfall amounts

from 3,645 calendar days. The actual output sequential data file contained 3740 events because the computer program omits the last event of record from the data set. This was done because, in general, there is no way to insure that the last hour of record is actually the last hour of the last event of record.

For various reasons, not all of the 3740 events can be used for developing statistical descriptions of the pertinent rainfall event characteristics. First of all, some events have indeterminate values of R, T, D, I_{max}, and I_w as discussed in detail above. Omitting all events that have indeterminate values for any of these variables reduces the data set to 3604 events.

Second, two large gaps occurred in the data as shown in Table 3.1. Documentation received from the NCDC on the data files clearly identified these gaps. The first occurred from midnight at December 31, 1946 to 1:00 am on October 1, 1947 and the second from midnight on February 26, 1973 to 1:00 am on March 1, 1975. These gaps resulted from temporary termination of the precipitation measurement program at Love Field. Unfortunately, the data flags included in the files do not flag missing data due to termination of measurement. This caused the computer program to compute unrealistic values of T for the first event following each gap: 6733 hours for the first gap and 11,817 hours for the second. The two events following the gaps must be omitted, which reduces the data set to 3602 events.

Finally, when the precipitation measurement program at Love Field was re-instituted the second time, the rainfall measurement instrumentation was changed. The new equipment recorded rainfall volumes only in increments of 0.10 inch. Prior to the change, the data had been recorded in 0.01 inch increments. A sensitivity of only 0.10 inch is not adequate for the research purposes of this study. Therefore, all data collected after 1:00 am on March 1, 1975 were omitted from the data set. Only the data with a measurement

sensitivity of 0.01 inch is used herein. This spans from November 4, 1940 to the last event before the equipment change on February 26, 1973. This reduces the final data set to 2760 individual storm events.

Probability Distributions For Rainfall Event Variables

Sample Statistics

Sample statistics, computed using the set of 2760 individual storm events, for R, T, D, I_{max}, and I_w are listed in Table 3.4. Sample statistics were computed using the following definitions (SAS Institute, Inc. 1985a and 1985b):

(3.13) $$\text{Mean, } \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

(3.14) $$\text{Standard Deviation, } s_x = + \sqrt{\frac{\sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2}{n-1}}$$

(3.15) $$\text{Minimum Value, } x_{min} = \min \left(x_1, x_2, ..., x_i, ..., x_n \right)$$

(3.16) $$\text{Maximum Value, } x_{max} = \max \left(x_1, x_2, ..., x_i, ..., x_n \right)$$

(3.17) $$\text{Standard Error of the Mean, } e_{sx} = \frac{s_x}{\sqrt{n}}$$

(3.18) $$\text{Coefficient of Variation, } C_{vx} = \frac{s_x}{\bar{x}}$$

TABLE 3.4

SAMPLE STATISTICS FOR RAINFALL EVENT DATA

Statistic	R (in.)	T (hr.)	D (hr.)	I_{max} (in./hr.)	I_w (in./hr.)
Mean, $\bar{x}$	0.3933	92.8141	4.4370	0.1905	0.1461
Standard Deviation, s_x	0.6549	136.6676	4.9213	0.3005	0.2266
Minimum Value, x_{min}	0.01	3.	1.	0.01	0.01
Maximum Value, x_{max}	6.17	1186.	82.	3.08	2.27
Standard Error of the Mean, e_{sx}	0.0125	2.6014	0.0937	0.0057	0.0043
Coefficient of Variation, C_{vx}	1.6651	1.4725	1.1092	1.5776	1.5512
Skewness Coefficient, C_{sx}	3.3791	2.8574	3.7289	3.3448	3.3464
Kurtosis, k_x	15.8497	11.6859	31.8798	15.9970	15.9904

(3.19) Coefficient of Skewness, $$C_{sx} = \frac{n}{(n-1)(n-2)} \left[\frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^3}{s_x^3} \right]$$

(3.20) Kurtosis, $$k_x = \frac{n\,(n+1)}{(n-1)(n-2)(n-3)} \left[\frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^4}{s_x^4} \right] - \frac{3\,(n-1)^2}{(n-2)(n-3)}$$

Where X represents any of the five pertinent random variables in the set {R, T, D, I_{max}, I_w}, x_i represents any arbitrary value of X in the sample data set, and n is the total number of observed values of X in the data set.

The mean rainfall volume for the 2760 events was 0.3933 inches while it ranged from 0.01 inches to 6.17 inches. The mean time between storms was 92.8141 hours, or about 4 days. Time between storms ranged from 3 hours to 1186 hours, which is about 50 days. Event duration averaged 4.4370 hours and ranged from 1 hour to 82 hours, which is 3.4 days. Maximum hourly intensity averaged 0.1905 inch/hour and weighted intensity averaged 0.1461 inch/hour. These variables ranged from 0.01 to 3.08 inches/hour and 0.01 to 2.27 inches/hour, respectively.

All of the random variables are positively skewed, which is common for hydrologic variables, as indicated by the skewness coefficients of Table 3.4. The values of kurtosis shown in Table 3.4 imply relatively peaked probability density functions (pdf's) for all five of the random variables. These factors, and the fact that all of the variables have non-negative lower bounds and are unbounded in the direction of positive infinity, indicate that the gamma family of pdf's, the log-normal pdf, and the Weibull pdf may be applicable for describing the nature of the random variables. The low coefficient of variation for event

duration, 1.1092, implies that the exponential pdf, which is a special case of the general gamma pdf, may be useful for describing duration.

Further insight into the nature of the random variables can be achieved by ordering and ranking the observed values. A frequency analysis for R values is presented in Table B.2 of Appendix B. The observed values of R have been ranked and shown in ascending order along with the frequency of occurrence and percent of total occurrences with the given value. Also shown are the cumulative frequency of occurrence and the cumulative percent occurrence, which is the percent of total occurrences with a value less than or equal to the given value. Frequency analyses are presented for T, D, I_{max}, and I_w in Tables B.3, B.4, B.5, and B.6, respectively, in Appendix B.

In each case, the frequency analysis demonstrated that the most common frequency of occurrence was the minimum observed value. Specifically:

(1) 15.8 percent of the events had a value of R = 0.01 inch

(2) 8.4 percent of the events had a value of T = 3. hours

(3) 31.0 percent of the events had a value of D = 1. hours

(4) 21.4 percent of the events had a value of I_{max} = 0.01 inch/hour

(5) 22.5 percent of the events had a value of I_w = 0.01 inch/hour.

Furthermore, in each case the frequency decreased rapidly as the observed value increased. This implies that the gamma family of pdf's with a shape parameter, α, less than one may be particularly useful. The Weibull pdf with α less than one may also have some potential applicability.

The degree of correlation between the five random variables can be measured by the sample correlation coefficients for each pair of variables. Sample correlation coefficients are computed as follows:

$$r_{xy} = \frac{\sum_{i-1}^{n} \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{(n-1)\, s_x s_y} \qquad \left(-1 \le r_{xy} \le 1\right) \tag{3.21}$$

where X and Y are any arbitrary members of the set of 5 pertinent variables. All other symbols are as defined previously.

If r_{xy} is +1, then there is a perfect positive concordance between x and y, that is, as one increases the other increases in some specific predictable way. If r_{xy} is –1, then there is a perfect negative correspondence between x and y, that is, as one increases the other decreases in some specific predictable way. An r_{xy} value of zero implies no correlation, no predictability, between X and Y values. Although, by definition, two independent random variables have a zero correlation coefficient, the converse does not necessarily hold. However, the magnitude of r_{xy} gives a qualitative indication of the probability of showing independence by statistical procedures.

The symmetric matrix of correlation coefficients for R, T, D, I_{max}, and I_w is shown in Table 3.5. It is apparent from these statistics that T is poorly correlated with the other four random variables since the correlation coefficients are near zero. It is also apparent that I_{max} and I_w are highly correlated with R having correlation coefficients of 0.8603 and 0.7733, respectively. A relatively moderate correlation exists between D and R with a correlation coefficient of 0.6047.

In order to develop a stochastic model of the rainfall-runoff process in a direct and straightforward manner, independence of the important input random variables is a desirable property. Furthermore, D, I_{max}, and I_w are all essentially measures of the rainfall intensity (R/D is the overall average intensity), such that only one of the three would be sufficient for development of the rainfall-runoff

TABLE 3.5

MATRIX OF SAMPLE CORRELATION COEFFICIENTS

Variable	R	T	D	I_{max}	I_w
R	1.0000	−.0155	.6047	.8603	.7733
T	−.0155	1.0000	−.0193	−.0052	.0027
D	.6047	−.0193	1.0000	.2965	.1903
I_{max}	.8603	−.0052	.2965	1.0000	.9797
I_w	.7733	.0027	.1903	.9797	1.0000

model. Thus, the appropriate choice among the three is D based on the criterion of minimum degree of correlation with R. Fortunately, D is a better predictor of runoff from the study watershed than either I_{max} or I_w as demonstrated in the detailed analyses presented in Chapter 4. Therefore, I_{max} and I_w have been deleted from the set of pertinent random variables.

The remaining analyses for fitting pdf's to the observed data will deal exclusively with the random variables R, T, and D.

One final point must be emphasized before proceeding with the task of fitting pdf's to the observed data. Rainfall volume, time between events, and event duration are, in the strictest physical sense, continuous random variables. However, the nature of the instrumentation used for measuring precipitation, in concert with the convention of reporting the data in intervals of one clock hour, discretizes the data into fixed incremental classes. Rainfall volume is discretized into integer multiples of 0.01 inch, and time between events and event duration are discretized into integer numbers of clock hours. As is the common practice in hydrology, continuous pdf's and their associated cumulative distribution functions (cdf's) have been fit to these sets of discrete data, where possible, to approximate the actual underlying continuous functions.

Rainfall Volume

A 3-parameter gamma pdf was fit to the observed rainfall volume data. The gamma pdf has been commonly recommended for this purpose (Eagleson 1970, Haan 1977, Yevjevich 1972). This pdf is of the general form:

$$f(r) = \begin{cases} 0, & r \le \lambda \\ \frac{1}{\beta^{\alpha}} \frac{1}{\Gamma(\alpha)} (r-\lambda)^{\alpha-1} e^{\frac{-(r-\lambda)}{\beta}}, & r > \lambda \end{cases} \tag{3.22}$$

where,

r = any arbitrary value of the random variable R ($r > \lambda$)

α = shape parameter ($\alpha > 0$)

β = scale parameter ($\beta > 0$)

λ = location parameter ($0 \le \lambda < .01$)

$\Gamma(\alpha)$ = gamma function of α.

By definition, then, the cdf will be,

$$F(r) = \begin{cases} 0, & r \le \lambda \\ \frac{1}{\beta^{\alpha}} \frac{1}{\Gamma(\alpha)} \int_{\lambda}^{r} (u-\lambda)^{\alpha-1} e^{\frac{-(u-\lambda)}{\beta}} du, & r > \lambda \end{cases} \tag{3.23}$$

As discussed previously, reported values of R have a lower bound of 0.01 inch (i.e., $R \ge .01$). Since rainfall volume cannot be negative, λ must be in the range of $0 \le \lambda < .01$. The value of λ cannot be 0.01 inch because this implies $F(r = 0.01) = 0$ which is physically incompatible with the fact that 0.01 inch is the single most frequently observed value at 15.8 percent of the total observations. Therefore, by practical necessity, λ was assumed to have a known value of 0.005 inch so that any values of R in the range $0.005 < r < .01$ will round numerically to 0.01 inch. This assumption was based on physical reasoning solely for practical purposes and is, thus, independent of the sample data.

The pertinent population moments for parameter estimation for the 3-parameter gamma pdf are given by,

(3.24) Mean, $\mu = \alpha\beta + \lambda$

(3.25) Variance, $\sigma^2 = \alpha\beta^2$

(3.26) Skewness Coefficient, $\gamma = \frac{2}{\sqrt{\alpha}}$

(3.27) Kurtosis, $\kappa = 3 + \frac{6}{\alpha}$.

The parameters of the distribution were estimated by the Method of Moments (MOM) by equating the sample mean and variance to the population mean and variance in equations 3.24 and 3.25 as follows (Haan 1977, Yevjevich 1972):

(3.28) $\bar{r} = \hat{\alpha}\hat{\beta} + \lambda$

(3.29) $s_r^2 = \hat{\alpha}\hat{\beta}^2$

where,

$\bar{r}$ = sample mean for R

s_r^2 = sample variance for R

$\hat{\alpha}$ = estimator of α

$\hat{\beta}$ = estimator of β .

Equation 3.29 is not in strict conformance with the MOM since s_r^2 involves the denominator (n-1) instead of n as required by the definition of sample moments. However, the error is small, especially given the large sample size.

Solving equations 3.28 and 3.29 simultaneously yields the simple expressions for $\hat{\alpha}$ and $\hat{\beta}$:

(3.30) $$\hat{\alpha} = \frac{(\bar{r} - \lambda)^2}{s_r^2}$$

(3.31) $$\hat{\beta} = \frac{s_r^2}{\bar{r} - \lambda} \ .$$

Notice that $\hat{\alpha}$ is dimensionless and the scale parameter $\hat{\beta}$ has the same dimensions as R (inches) as required by the pdf.

Substituting the computed values of $\bar{r}$ and s_r from Table 3.4, and assuming λ is known to be 0.005 inch, yields the following:

$$\hat{\alpha} = 0.3515$$

$$\hat{\beta} = 1.1046 \text{ inch} \ .$$

Notice that $\hat{\alpha} < 1$ such that the gamma pdf is of the reverse J-shape form already expected from inspection of Table B.2.

It is also instructive to compare $\hat{\gamma}$ and $\hat{\kappa}$ to the sample skewness coefficient, C_{sr}, and the sample kurtosis, k_r, respectively. The estimators of γ and κ are,

(3.32) $$\hat{\gamma} = \frac{2}{\sqrt{\hat{\alpha}}}$$

(3.33) $$\hat{\kappa} = 3 + \frac{6}{\hat{\alpha}} \ .$$

Substituting the value of $\hat{\alpha}$ into these expressions yields the following comparisons to C_{sr} and k_r from Table 3.4:

skewness		kurtosis	
$\hat{\gamma}$	= 3.3734	$\hat{\kappa}$	= 20.0697
C_{sr}	= 3.3791	k_r	= 15.8497 .

Agreement is excellent between the skewness coefficient which inspires confidence in the parameter estimates. Agreement is only fair for kurtosis, but even this is deemed acceptable given that kurtosis involves fourth moments.

Considerable effort was also made to estimate the parameters of the 3-parameter gamma pdf by the method of Maximum Likelihood Estimation (MLE). A procedure described in detail by Haan (1977) and Yevjevich (1972) was applied. The MLE method generated estimates of the parameters α and β that produced relatively poor agreement between observed data and the resulting theoretical pdf. It is believed that this is a consequence of the shape parameter α being less than unity (Kottegoda 1980). As a result, the parameter estimates developed using the MOM have been used to fit the theoretical pdf.

A visual test for goodness of fit of the theoretical gamma cdf to the observed cumulative frequency data is shown on log-normal probability graph paper in Figure 3.1. The theoretical cdf is shown as a smooth solid curve and the observed cumulative frequencies taken from Table B.2 are shown as the plotted data points. Visual inspection readily demonstrates that there is excellent agreement between the theoretical gamma cdf and the observed empirical cdf.

In order to provide an objective evaluation of the fit, a chi-square test for goodness of fit was conducted. This test involves placing the observed data

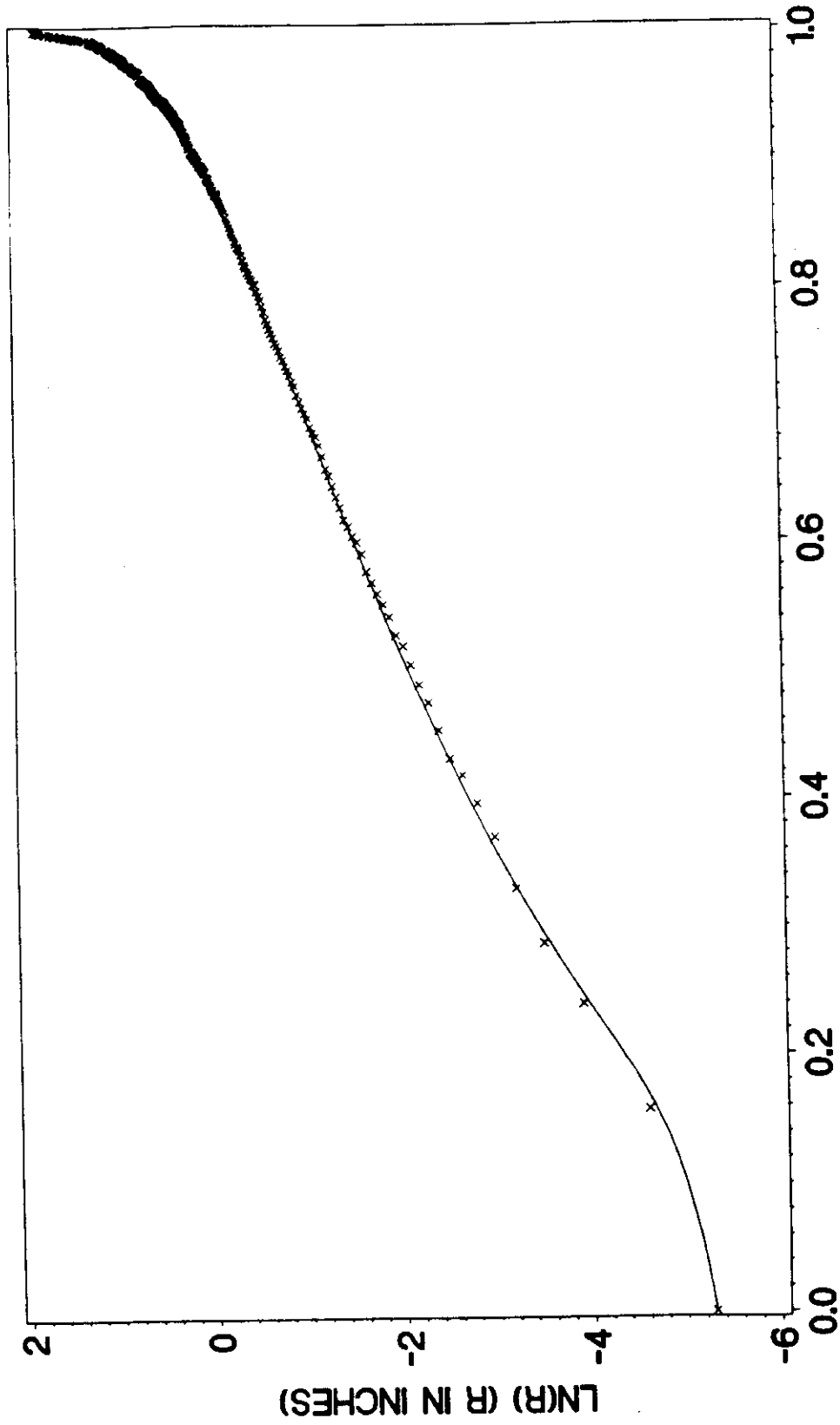


FIGURE 3.1
CDF FOR RAINFALL VOLUME

into class intervals and then computing the chi-square test statistic defined as follows:

(3.34) $$\chi^2 = \sum_{i=1}^{m} \frac{(n_i - np_i)^2}{np_i}$$

where,

χ^2 = chi-square test statistic

n_i = observed sample frequency of occurrence for the i-th class

n = total number of observed values

p_i = theoretical probability that the random variable will have a value in the i-th class interval obtained by integrating the pdf between the limits of the class interval

np_i = expected frequency in the i-th class interval for a random sample of size n from the theoretical pdf

m = number of class intervals used to categorize the data.

The chi-square statistic is approximately distributed, for a large sample ($n > 30$), as a chi-square random variable with degrees of freedom as follows:

(3.35) $$\nu = (m - 1) - q$$

where,

ν = degrees of freedom

q = number of parameters estimated for the theoretical pdf from the observed data.

The chi-square test for goodness of fit for the theoretical 3-parameter gamma pdf is presented in detail in Table 3.6. Common practice involves

TABLE 3.6

CHI-SQUARE TEST FOR RAINFALL VOLUME PDF

3-Prameter Gamma CDF
($\alpha = 0.3515$, $\beta = 1.1046$,
$\lambda = 0.005$)

Range	Observed Frequency	Cumulative Frequency	Expected Frequency	Chi-Square Statistic
0.01	435	464	464	1.812
0.02	225	681	217	0.295
0.03	131	813	132	0.008
0.04	119	913	100	3.610
0.05	110	995	82	9.561
0.06 - 0.10	286	1279	284	0.014
0.11 - 0.15	185	1467	188	0.048
0.15 - 0.20	135	1610	143	0.448
0.21 - 0.25	100	1726	116	2.207
0.26 - 0.30	110	1822	96	2.042
0.31 - 0.40	154	1976	154	0.000
0.41 - 0.50	106	2096	120	1.633
0.51 - 0.60	115	2191	95	4.211
0.61 - 0.70	74	2269	78	0.205
0.71 - 0.80	65	2334	65	0.000
0.81 - 0.90	55	2389	55	0.000
0.91 - 1.00	47	2435	46	0.022
1.01 - 1.50	145	2587	152	0.322
1.51 - 2.00	69	2664	77	0.831
2.01 - 2.50	38	2705	41	0.220
2.51 - 3.00	22	2728	23	0.043
3.01 - 3.50	17	2741	13	1.231
3.51 - 7.00	17	2760	19	0.211
	2760		2760	28.974

Null Hypothesis, H_o: The data represent a random sample from a population with a 3-parameter gamma pdf with parameters
$\alpha = 0.3515$, $\beta = 1.1046$, $\lambda = 0.005$.

Alternative Hypothesis, H_1: The data were not drawn from a population that follows the specified pdf.

$\nu = (23 - 1) - 2 = 20$ (the parameters α and β were estimated from the data: the parameter λ was assumed known).

Total Chi-Square Statistic: $\chi^2 = 28.974$.

Critical Region at 0.05 Level of Significance: $\chi_0^2 > 31.410$.

Decision: Do not reject H_o.

dividing the observed data into categories using class intervals of constant size or using class intervals of uneven size but so selected that the expected frequency is constant from one class to the next. Neither of these procedures was practical for the data in this study for two reasons. First, the data are so positively skewed (15.8 percent of the observations had the minimum possible value of 0.01 inch) that using uniform class sizes produces either too few class intervals in light of the large sample size ($m < 5$) or too many intervals with low expected cell frequencies ($np_i < 10$). Second, the latter approach cannot be used because the observed data have been discretized and, thus, are not truly continuous. This requires that the theoretical frequencies match the class interval boundaries exactly. Therefore, uneven class sizes were chosen such that the intervals increased as the R values become larger, as follows:

Class Size	Range
.01	.01 - .05
.05	.06 - .30
.10	.31 - 1.00
1.50	1.01 - 7.00 .

In this way, a sufficient number of class intervals was obtained with a large expected frequency in each class. The sizes of the intervals were increased in a simple rational progression that does not depend in any way on the observed data. Therefore, the fundamental assumption of independence among the observed class frequencies has not been violated. Overall, this approach is consistent with the general guidelines for practical application of the chi-square test (Steger 1971, Daniel 1978, Hoel 1984).

As shown in Table 3.6, the total chi-square statistic had a value of 28.974 which lies outside the critical region of the test ($\chi_0^2 > 31.410$) at a 0.05 level of significance. The decision rule suggests not to reject the null hypothesis that the observed data represent a random sample of size 2760 from a 3-parameter gamma pdf with the specified parameters.

Therefore, the final pdf to be used for simulation of rainfall volume is given by,

$$(3.36)\quad f(r) = \begin{cases} 0, & r \le .005 \\ \dfrac{1}{(1.1046)^{.3515}} \dfrac{1}{\Gamma(.3515)} (r - .005)^{.3515-1} e^{-\left[\frac{r-.005}{1.1046}\right]}, & r > .005 \end{cases}$$

and the corresponding cdf is given by,

$$(3.37)\quad F(r) = \begin{cases} 0, & r \le .005 \\ \dfrac{1}{(1.1046)^{.3515}} \dfrac{1}{\Gamma(.3515)} \displaystyle\int_{.005}^{r} (u - .005)^{.3515-1} e^{-\left[\frac{u-.005}{1.1046}\right]} du, & r > .005 . \end{cases}$$

Time Between Events

Fitting a pdf to the data for time between events, T, proved to be considerably more challenging than fitting a pdf to the R data. The high coefficient of variation, 1.4725 from Table 3.4, indicates that the exponential pdf has little applicability to this set of data. On the other hand, the relatively high skewness and kurtosis, 2.8574 and 11.6859 respectively, of the data suggest

that gamma or Weibull pdf's with $\alpha < 1$ may be useful. These pdf's are of a general reverse J-shape.

The general form of a 3-parameter gamma pdf for T is given by,

$$f(t) = \begin{cases} 0, & t \le \lambda \\ \dfrac{1}{\beta^{\alpha}} \dfrac{1}{\Gamma(\alpha)} (t-\lambda)^{\alpha-1} e^{-\left[\frac{t-\lambda}{\beta}\right]}, & t > \lambda \end{cases} \tag{3.38}$$

where t is any arbitrary value of the random variable T ($t > \lambda$). All other parameters and functions are as defined in the previous section, except that the possible range of values for λ is now $0 \le \lambda < 3$. The corresponding cdf is as follows:

$$F(t) = \begin{cases} 0, & r \le \lambda \\ \dfrac{1}{\beta^{\alpha}} \dfrac{1}{\Gamma(\alpha)} \displaystyle\int_{\lambda}^{t} (u-\lambda)^{\alpha-1} e^{-\left[\frac{u-\lambda}{\beta}\right]} du, & t > \lambda . \end{cases} \tag{3.39}$$

Assuming that λ is known (or can be guessed) a priori, and using the MOM as discussed in detail in the previous section, the estimators $\hat{\alpha}$ and $\hat{\beta}$ are given by,

$$\hat{\alpha} = \frac{(\bar{t}-\lambda)^2}{s_t^2} \tag{3.40}$$

$$\hat{\beta} = \frac{s_t^2}{\bar{t}-\lambda} \tag{3.41}$$

where $\bar{t}$ and s_t are the sample mean and standard deviation of the observed data.

The general 3-parameter Weibull pdf has the form:

$$f(t) = \begin{cases} 0 \ , \ t \le \lambda \\ \alpha\,(t-\lambda)^{\alpha-1}\,(\beta-\lambda)^{-\alpha}\; e^{-\left[\frac{t-\lambda}{\beta-\lambda}\right]^{\alpha}} \ , \ t > \lambda \end{cases} \tag{3.42}$$

where,

α = shape parameter ($\alpha > 0$)

β = scale parameter ($\beta > 0$)

λ = location parameter ($0 \le \lambda < 3.0$).

The corresponding Weibull cdf is known to be,

$$F(t) = \begin{cases} 0 \ , \ t \le \lambda \\ 1 - e^{-\left[\frac{t-\lambda}{\beta-\lambda}\right]^{\alpha}} \ , \ t > \lambda \ . \end{cases} \tag{3.43}$$

The pertinent population moments of the Weibull pdf, in terms of the parameters of the distribution, are as follows (Haan 1977, Yevjevich 1972):

$$\text{Mean, } \mu = \lambda + (\beta - \lambda)\,\Gamma\left(1 + \frac{1}{\alpha}\right) \tag{3.44}$$

$$\text{Variance, } \sigma^2 = (\beta-\lambda)^2\left[\Gamma\left(1+\frac{2}{\alpha}\right) - \Gamma^2\left(1+\frac{1}{\alpha}\right)\right] \tag{3.45}$$

(3.46) Coefficient of Skewness,

$$\gamma = \frac{\Gamma(1+\frac{3}{\alpha}) - 3\Gamma(1+\frac{2}{\alpha})\,\Gamma(1+\frac{1}{\alpha}) + 2\Gamma^3(1+\frac{1}{\alpha})}{\left[\Gamma(1+\frac{2}{\alpha}) - \Gamma^2(1+\frac{1}{\alpha})\right]^{3/2}}$$

where

$\Gamma(\bullet)$ = gamma function of the argument

$\Gamma^a(\bullet)$ = a-th power of the gamma function of the argument.

The equations for μ and σ^2 can be rearranged as follows:

(3.47) $$\beta = \mu + \sigma A(\alpha)$$

(3.48) $$\lambda = \beta - \sigma B(\alpha)$$

where,

(3.49) $$A(\alpha) = \left[1 + \Gamma(1+\frac{1}{\alpha})\right] B(\alpha)$$

(3.50) $$B(\alpha) = \left[\Gamma(1+\frac{2}{\alpha}) - \Gamma^2(1+\frac{1}{\alpha})\right]^{-1/2}.$$

The parameters of the Weibull pdf can be estimated using the MOM by equating the sample mean and standard deviation to the population mean and standard deviation in equations 3.47 and 3.48 as follows:

(3.51) $$\hat{\beta} = \bar{t} + s_t A(\hat{\alpha})$$

(3.52) $$\lambda = \hat{\beta} + s_t B(\hat{\alpha}) .$$

Since λ is assumed known, these two equations can be solved simultaneously for $\hat{\alpha}$ and $\hat{\beta}$. However, a trial and error solution is necessary because of the extreme complexity of the interdependent functions $A(\alpha)$ and $B(\alpha)$.

A computer program was written to solve this problem using the Statistical Analysis System (SAS) software package (SAS Institute, Inc. 1985a and 1985b). The program is listed in Table B.7 of Appendix B. The program computes the values of the complicated A ($\hat{\alpha}$) and B ($\hat{\beta}$) functions for various trial values of $\hat{\alpha}$. A separate value of $\hat{\beta}$ is then computed from equations 3.51 and 3.52 using the trial values of $\hat{\alpha}$ as follows:

(3.53) $$\hat{\beta}_1 = \bar{t} + s_t A(\hat{\alpha})$$

(3.54) $$\hat{\beta}_2 = \lambda - s_t B(\hat{\alpha}) .$$

The analysis is complete upon determination of the specific trial value of $\hat{\alpha}$ that makes $\hat{\beta}_1 = \hat{\beta}_2$. The program also computes the value of $\hat{\gamma}$ for comparison to the sample skewness by substituting $\hat{\alpha}$ into equation 3.46 for the coefficient of skewness.

Values of T have a lower bound of 3 hours (i.e., $T \geq 3$). Since time between events cannot be negative, λ must be in the range $0 \leq \lambda < 3.0$. The value of λ cannot be 3.0 hours because this implies $F(t = 3.0) = 0$ for both the gamma and the Weibull pdf, which is physically incompatible with the fact that 3.0 hours is the single most frequently observed value at 8.4 percent of the total observations.

Considerable effort was expended on fitting the general 3-parameter gamma pdf and the 3-parameter Weibull pdf to the observed data for T. Several alternative guesses of the value of λ were tried in an attempt to improve the fit. The parameter estimates $\hat{\alpha}$ and $\hat{\beta}$ were computed using the assumed value of λ and values of $\bar{t}$ and s_t of 92.8141 hours and 136.6676 hours, respectively, from Table 3.4. In these trials, a graphical test of goodness of fit was employed. The theoretical cdf frequencies and the observed cumulative frequencies were plotted on log-normal probability paper and the fit was evaluated visually. This is, admittedly, a subjective test but it gives insight into the specific values of T where the greatest deviation from theoretical frequencies occur. In order to provide an objective evaluation of fit, the chi-square test for goodness of fit was also conducted in each case. The analyses are summarized in Table 3.7. No acceptable fit to the data was obtained.

In every case, the graphical analysis showed a poor fit to the data at values of T below approximately 50 hours and a good fit to the data above that point. Also, it was found that the gamma pdf produced a significantly better fit to the data over the entire range of T values than did the Weibull pdf. Therefore, no additional consideration was given to fitting the Weibull pdf.

The disparity between the goodness of fit between low and high values of T suggests that the lower values may represent a sample from a different pdf than that of the higher values. A possible explanation for this has been given at the end of this section.

As a result of the generally good fit to the data achieved with the gamma pdf for T values of 50 or above, it was decided to fit a 3-parameter gamma pdf to the upper tail of the data. The fitted function is a conditional pdf which is

TABLE 3.7
EVALUATION OF PDF'S FOR TIME BETWEEN EVENTS

Probability Density Function	Parameters $\hat{\alpha}$	Parameters $\hat{\beta}$	λ	Decision: Graphical CDF	Decision: Chi-Square Test
3-Parameter Gamma	.4367	206.8119	2.5	Do Not Accept H_0	Do Not Accept H_0
3-Parameter Gamma	.4415	205.6733	2.0	Do Not Accept H_0	Do Not Accept H_0
2-Parameter Gamma	.4612	201.2413	0.0	Do Not Accept H_0	Do Not Accept H_0
3-Parameter Weibull	.6796	71.8022	2.5	Do Not Accept H_0	Do Not Accept H_0
2-Parameter Weibull	.6959	72.8994	0.0	Do Not Accept H_0	Do Not Accept H_0

Null Hypothesis, H_0: The data are drawn from a pdf of the specified type with the specified parameters.

Alternative Hypothesis, H_1: The data are not drawn from the specified pdf.

Decision Rule, Graphical: If, by visual inspection of the graph on log-normal probability paper, the cdf does not fit the observed cumulative frequencies then do not accept H_0. Otherwise, do not reject H_0.

Decision Rule, Chi-Square: Using a 0.05 level of significance, do not accept H_0 if the chi-square statistic exceeds the critical value (χ_0^2) for the appropriate degrees of freedom. Otherwise, do not reject H_0.

conditional on the value of T being greater than or equal to some specified lower bound, t_l, defined as follows:

t_l = lowest integer value of t for which the fitted cdf has a non-zero value.

A trial and error approach was used to determine the lowest value of t_l for which an acceptable fit to the upper tail could be achieved. Each time a new trial value of t_l was used, a different subset of the observed data become applicable for which $\bar{t}$ and s_t had to be determined in order to compute the estimators $\hat{\alpha}$ and $\hat{\beta}$. These sample statistics are shown in Table 3.8 for the 6 values of t_l used in the analysis.

As before, a graphical test of goodness of fit was used as the primary method of selecting a suitable pdf. This analysis is summarized in Table 3.9. Again, the value of λ is assumed known and was chosen as either (t_l - 1) or (t_l - 0.5), whichever gave the best fit of the theoretical cdf at t_l by graphical inspection. A chi-square test of goodness of fit was conducted to provide a comparison to the graphical analysis.

As shown in Table 3.9, the subset of data comprised of T values of 12 hours or larger can be acceptably fit by a 3-parameter gamma pdf with parameters as follows:

$$\alpha = .6575$$

$$\beta = 183.3745 \text{ hours}$$

$$\lambda = 11.5 \text{ hours.}$$

A visual test for goodness of fit of the theoretical gamma cdf to the observed cumulative frequency data is shown on log-normal probability graph paper in Figure 3.2. The theoretical cdf is shown as a smooth solid curve and the

TABLE 3.8

SAMPLE STATISTICS FOR DATA SUBSETS FOR TIME BETWEEN EVENTS

Data ($T \geq t_l$)	$\bar{t}$ (hours)	s_t (hours)	Number of Observations
$T \geq 4$	101.0178	139.9298	2529
$T \geq 5$	107.3957	142.1561	2373
$T \geq 8$	120.8721	146.0827	2095
$T \geq 9$	124.4289	146.9648	2031
$T \geq 10$	127.2825	147.6319	1982
$T \geq 12$	132.0614	148.6872	1904

TABLE 3.9

UPPER TAIL PDF's FOR TIME BETWEEN EVENTS

Data Subset ($T \geq t_l$)	Gamma Probability Density Function Parameters $\hat{\alpha}$	$\hat{\beta}$	λ	Decision: Graphical CDF	Decision: Chi-Square Test
$T \geq 4$	.4907	199.7632	3.0	Do Not Accept H_0	Do Not Accept H_0
$T \geq 5$	.5290	195.4468	4.0	Do Not Accept H_0	Do Not Accept H_0
$T \geq 8$	.6023	188.2311	7.5	Do Not Accept H_0	Do Not Accept H_0
$T \geq 9$	.6222	186.3095	8.5	Do Not Accept H_0	Do Not Accept H_0
$T \geq 10$	.6370	185.0460	9.5	Do Not Accept H_0	Do Not Accept H_0
$T \geq 12$	.6575	183.3745	11.5	Do Not Reject H_0	Do Not Reject H_0

Null Hypothesis, H_0: The specified subset of data is drawn from a 3-parameter gamma pdf with the specified parameters.

Alternative Hypothesis, H_1: The specified subset of data is not drawn from the hypothesized pdf.

Decision Rule, Graphical: If, by visual inspection of the graph on log-normal probability paper, the cdf does not fit the observed cumulative frequencies of the subset of data then do not accept H_0. Otherwise, do not reject H_0.

Decision Rule, Chi-Square: Using a 0.05 level of significance, do not accept H_0 if the chi-square statistic exceeds the critical value (χ_0^2) for the appropriate degrees of freedom. Otherwise, do not reject H_0.

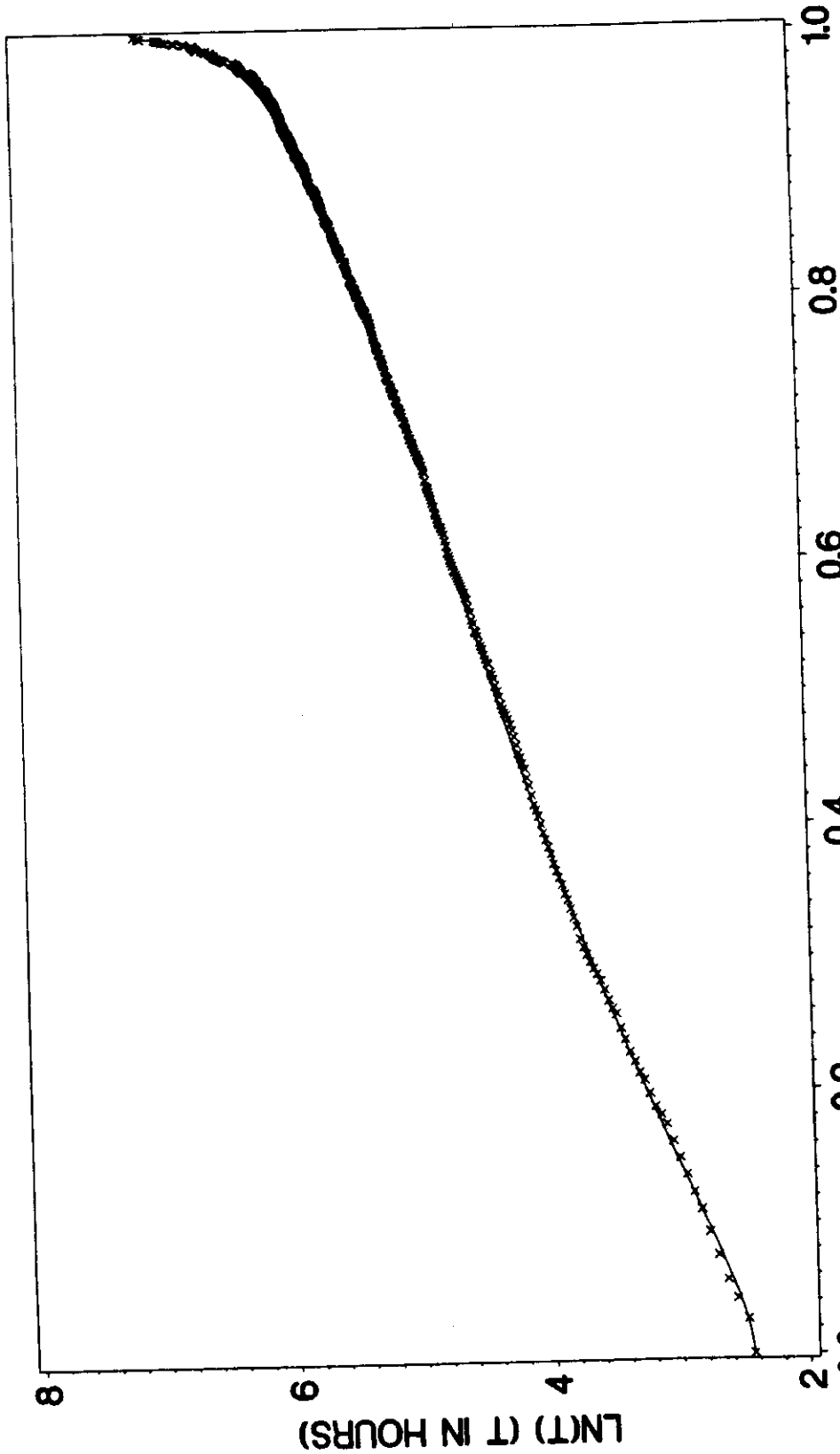


FIGURE 3.2
UPPER TAIL CDF FOR TIME BETWEEN EVENTS

observed cumulative frequencies taken from Table B.3 in Appendix B are shown as the plotted data points. Visual inspection clearly demonstrates that there is excellent agreement between the theoretical cdf and the observed empirical cdf.

The chi-square test for goodness of fit for the fitted pdf is shown in detail in Table 3.10. As was done for rainfall volume previously, uneven class sizes have been used so that a sufficient number of class intervals could be obtained with large expected frequencies (at least 10) in each class. As shown, the total chi-square statistic was 34.618 which is outside the critical region of the test ($\chi_0^2 > 35.172$) at a 0.05 level of significance and 23 degrees of freedom. The decision rule suggests not to reject the null hypothesis that the subset of the observed data represents a random sample of size 1904 from a 3-parameter gamma pdf with the specified parameters.

The entire data set, that is lower and upper tails, for time between events can now be represented as a composite pdf. The upper tail of the distribution is characterized by the continuous 3-parameter gamma pdf just developed, and the lower tail (i.e., $3 \leq T < 12$) is characterized by a discrete probability function. The composite pdf takes the form:

$$f(t) = \begin{cases} 0 \ , \ t < 3 \\ p_i \ , \ t = i \ \ (i = 3, 4, ..., 11) \\ \left(1 - \sum_{i=3}^{11} p_i\right) \left[\frac{1}{\beta^{\alpha}} \frac{1}{\Gamma(\alpha)} \ (t-\lambda)^{\alpha} e^{-\left[\frac{t-\lambda}{\beta}\right]}\right], \ t \geq 12 \end{cases} \tag{3.55}$$

TABLE 3.10

CHI-SQUARE TEST FOR UPPER TAIL PDF FOR TIME BETWEEN EVENTS

3-Parameter Gamma PDF ($\alpha = 0.6575$, $\beta = 183.3745$, $\lambda = 11.5$)

Range	Observed Frequency	Cumulative Frequency	Expected Frequency	Chi-Square Statistic
12	49	43	43	0.837
13	30	89	46	5.565
14	27	125	36	2.250
15	36	155	30	1.200
16	34	183	28	1.286
17	33	208	25	2.560
18	25	232	24	0.042
19	26	254	22	0.727
20	24	275	21	0.429
21-25	92	369	94	0.043
26-30	78	449	80	0.050
31-35	71	520	71	0.000
36-40	51	585	65	3.015
41-50	121	698	113	0.566
51-60	108	796	98	1.020
61-70	101	882	86	2.616
71-80	72	959	77	0.325
81-90	62	1029	70	0.914
91-100	67	1092	63	0.254
101-200	389	1497	405	0.632
201-300	183	1690	193	0.518
301-400	123	1789	99	5.818
401-500	48	1841	52	0.308
501-600	19	1869	28	2.893
601-700	14	1885	16	0.250
701-1300	21	1904	19	0.211
	1904		1904	34.329

Null Hypothesis, H_0: The data represent a random sample from a population with a conditional ($T \geq 12$) 3-parameter gamma pdf with parameters $\alpha = .6575$, $\beta = 183.3745$, $\lambda = 11.5$.

Alternative Hypothesis, H_1: The data were not drawn from a population that follows the specified pdf.

$\nu = (26-1) - 2 = 23$ (the parameters α and β were estimated from the data, the parameter λ was assumed known).

Total Chi-Square Statistic: $\chi^2 = 34.329$. Critical Region at 0.05 Level of Significance: $\chi^2_0 > 35.172$.

Decision: Do not reject H_0.

where,

p_i = theoretical probability that the random variable T has the value i.

The accompanying cdf is of the form:

(3.56)

$$F(t) = \begin{cases} 0 \ , \ t < 3 \\ \sum_{i=3}^{j} p_i \ , \ t = j \ \ (j = 3, 4, \ldots, 11) \\ \sum_{i=3}^{11} p_i + \left(1 - \sum_{i=3}^{11} p_i\right)\left[\frac{1}{\beta^{\alpha}} \frac{1}{\Gamma(\alpha)} \int_{\lambda}^{t} (u-\lambda)^{\alpha} e^{-\left[\frac{u-\lambda}{\beta}\right]} du\right] \ , \ t \geq 12 \ . \end{cases}$$

The estimators of the p_i by either the MOM or MLE are

(3.57)
$$\hat{p}_i = \frac{n_i}{n}$$

where,

n_i = number of observations with T = i

n = total number of observations.

The p_i values were computed using the frequency data of Table B.3 in Appendix B. The resulting final pdf is then:

(3.58)

$$f(t) = \begin{cases} 0, t < 3 \\ .0837, t = 3 \\ .0565, t = 4 \\ .0428, t = 5 \\ .0315, t = 6 \\ .0264, t = 7 \\ .0232, t = 8 \\ .0178, t = 9 \\ .0105, t = 10 \\ .0178, t = 11 \\ 0.6898 \left[\dfrac{1}{(183.3745)^{.6575}} \dfrac{1}{\Gamma(.6575)} (t - 11.5)^{.6575-1} e^{-\left[\frac{t-11.5}{183.3745}\right]} \right], t \geq 12 . \end{cases}$$

The associated final cdf is then given by:

(3.59)

$$F(t) = \begin{cases} 0, t < 3 \\ .0837, t = 3 \\ .1402, t = 4 \\ .1830, t = 5 \\ .2145, t = 6 \\ .2409, t = 7 \\ .2641, t = 8 \\ .2819, t = 9 \\ .2924, t = 10 \\ .3102, t = 11 \\ .3102 + .6898 \left[\dfrac{1}{(183.3745)^{.6575}} \dfrac{1}{\Gamma(.6575)} \displaystyle\int_{11.5}^{t} (u - 11.5)^{.6575-1} e^{-\left[\frac{u-11.5}{183.3745}\right]} du \right], t \geq 12 . \end{cases}$$

As a final check, a chi-square test of goodness of fit was conducted for the final composite pdf. This analysis is shown in detail in Table 3.11. The total chi-square statistic had a value of 33.182 which lies outside the critical region of the test ($\chi_0^2 > 35.172$) at the 0.05 level of significance and 23 degrees of freedom. The decision rule suggests not to reject the null hypothesis that the data are drawn from the specified composite pdf. Notice that 11 parameters were estimated from the data: α and β for the upper tail continuous pdf and p_3 through p_{11} for the lower tail discrete probabilities.

The final composite cdf is shown graphically in Figure 3.3 along with the observed empirical cdf. The agreement is excellent.

At this point, it seems appropriate to address the question as to why the values of T below 12 hours appear to represent a different pdf than the values above. This probably occurs because of the way T is defined and reported as an integer number of clock hours. This means that for any given reported value of T, the actual sequential time elapsed, T_e, can be as much as two hours longer. This can be stated mathematically as:

$$T = i \quad \text{implies} \quad i \le T_e \le i + 2 \quad (i = 3, 4, \ldots) \; . \tag{3.60}$$

This has been demonstrated graphically in Figure 3.4 for the specific case of T = 3. As shown, the only way that the actual sequential time elapsed between events can be exactly 3 hours is when the first rain event stops exactly on a clock hour and then the next rain starts on the clock hour exactly 3 hours later. At the other extreme, if 0.01 inch of rain fell 1 minute after a clock hour and then a second 0.01 inch fell 1 minute before the clock hour 5 hours later, then the actual sequential elapsed time would be 4 hours and 58 minutes but still T = 3

TABLE 3.11

CHI-SQUARE TEST FOR FINAL COMPOSITE PDF
FOR TIME BETWEEN EVENTS

		Final Composite CDF for T		
Range	Observed Frequency	Cumulative Frequency	Expected Frequency	Chi-Square Statistic
3	231	231	231	0.000
4	156	387	156	0.000
5	118	505	118	0.000
6	87	592	87	0.000
7	73	665	73	0.000
8	64	729	64	0.000
9	49	778	49	0.000
10	29	807	29	0.000
11	49	856	49	0.000
12	49	900	44	0.568
13	30	945	45	5.000
14	27	981	36	2.250
15	36	1011	30	1.200
16	34	1039	28	1.286
17	33	1064	25	2.560
18	25	1088	24	0.042
19	26	1110	22	0.727
20	24	1131	21	0.429
21-25	92	1225	94	0.043
26-30	78	1306	81	0.111
31-35	71	1377	71	0.000
36-40	51	1441	64	2.641
41-50	121	1554	113	0.566
51-60	108	1652	98	1.020
61-70	101	1738	86	2.616
71-80	72	1815	77	0.325
81-90	62	1885	70	0.914
91-100	67	1948	63	0.254
101-200	389	2353	405	0.632
201-300	183	2546	193	0.518
301-400	123	2645	99	5.818
401-500	48	2697	52	0.308
501-600	19	2725	28	2.893
601-700	14	2741	16	0.250
701-1300	21	2760	19	0.211
	2760		2760	33.182

TABLE 3.11 (Continued)
CHI-SQUARE TEST FOR FINAL COMPOSITE PDF
FOR TIME BETWEEN EVENTS

Null Hypothesis, H_o: The data represents a random sample from a population with a composite discrete-continuous 3-parameter gamma pdf with parameters
$p_3 = .0837$, $p_4 = .0565$, $p_5 = .0428$, $p_6 = .0315$, $p_7 = .0264$,
$p_8 = .0232$, $p_9 = .0178$, $p_{10} = .0105$, $p_{11} = .0178$,
$\alpha = .6575$, $\beta = 183.3745$, λ - 11.5.

Alternative Hypothesis, H_1: The data were not drawn from a population that follows the specified pdf.

$\nu = (35 - 1) - 11 = 23$ (the parameters p_3 through p_{11}, α, and β were estimated from the data: the parameter λ was assumed known).

Total Chi-Square Statistic : $\chi^2 = 33.182$.

Critical Region at 0.05 Level of Significance: $\chi_0^2 > 35.172$.

Decision: Do no reject H_o.

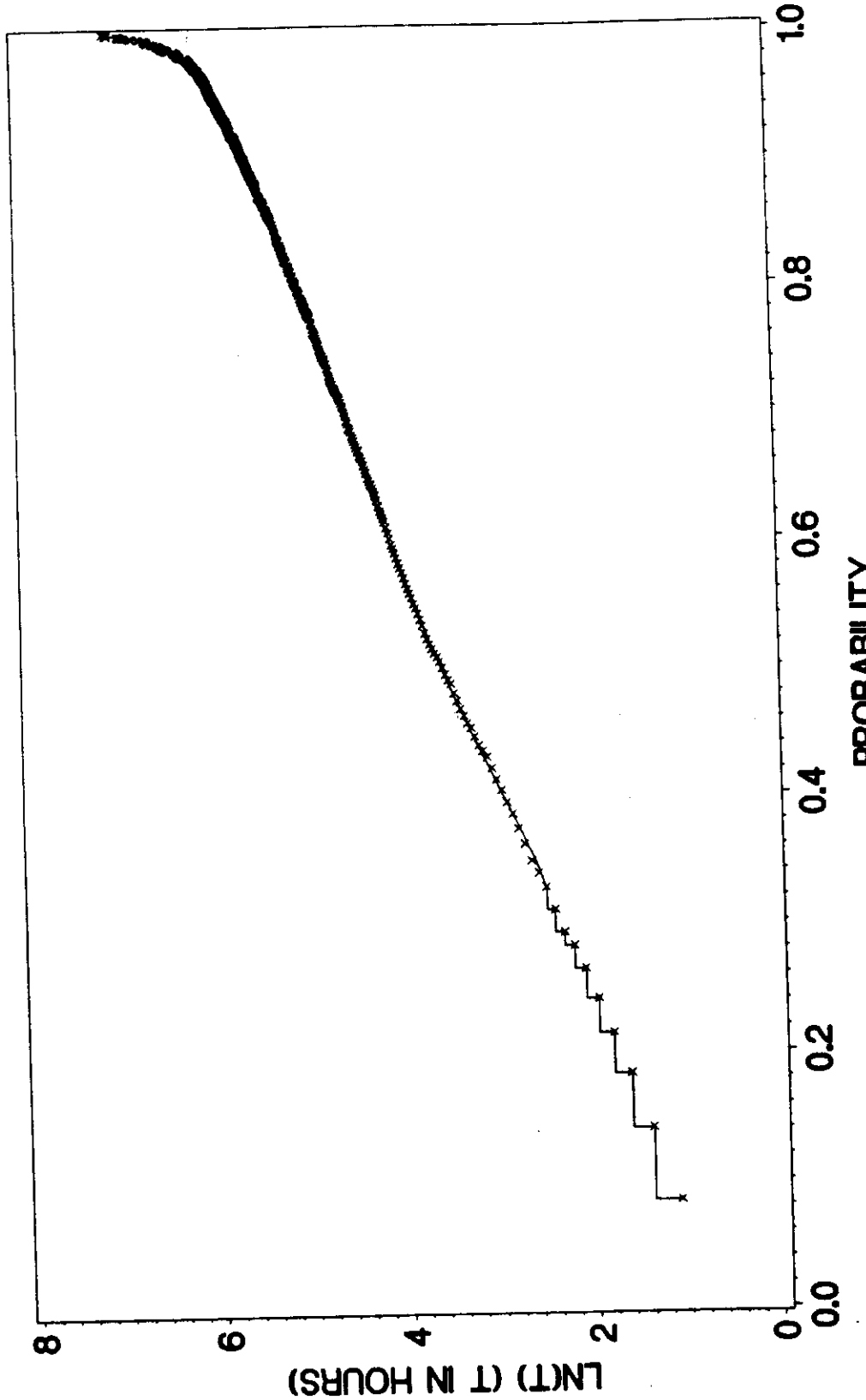


FIGURE 3.3
FINAL CDF FOR TIME BETWEEN EVENTS

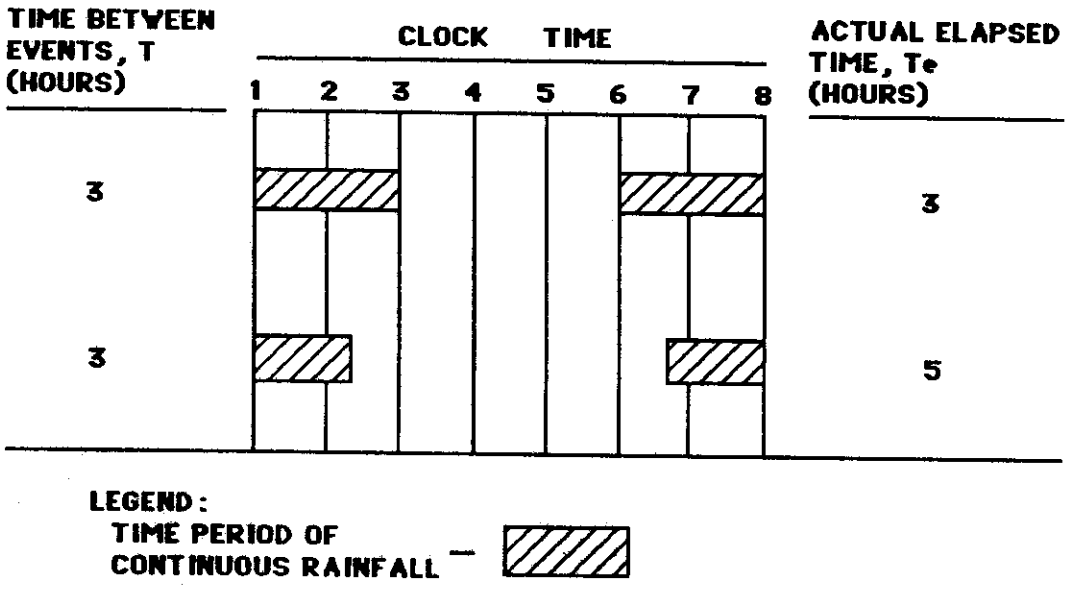


FIGURE 3.4
SAMPLE RAINFALL PATTERNS FOR T

because T is defined as the number of clock hours with zero recorded rainfall between the successive events.

The 2 hour range of difference between T and T_e is relatively large for small values of T: for T = 3 the actual elapsed time can be as much as 67 percent longer. However, this 2 hour difference is small compared to large values of T: for T = 12 the actual elapsed time can only be as much as 17 percent larger. At the mean observed value of T, 92.8141, the 2 hour difference is a mere 2 percent larger. Basically, the larger the value of T, the closer it is to the true elapsed time between events. This phenomena is exacerbated by the fact that the lower the value of T the higher the frequency of occurrence. The net result is that values of T below 12 hours effectively comprise a random sample from a different population from the values above.

Event Duration

The sample statistics of Table 3.4 show that D had a coefficient of variation of 1.1092, a skewness of 3.7289, and a very high kurtosis of 31.8798. The coefficient of variation near unity indicates that the 2-parameter exponential pdf might be particularly applicable. The exponential distribution has sometimes been recommended for fitting duration data (Eagleson 1970). The more general 3-parameter gamma pdf or the 3-parameter Weibull pdf, each with $\alpha < 1$, may also have some applicability. All 3 of these functions are of the reverse J-shape form that is suggested by the results of the frequency analysis for D shown in Table B.4 in Appendix B.

The general form of the 2-parameter exponential pdf for D is given by,

$$f(d) = \begin{cases} 0, & d \le \lambda \\ \frac{1}{\beta} e^{\frac{-(d-\lambda)}{\beta}}, & d > \lambda \end{cases} \tag{3.61}$$

where,

d = any arbitrary value of the random variable D ($d > \lambda$)

β = scale parameter ($\beta > 0$)

λ = location parameter ($0 \le \lambda < 1$).

The accompanying cdf is given by,

$$F(d) = \begin{cases} 0, & d \le \lambda \\ 1 - e^{\frac{-(d-\lambda)}{\beta}}, & d > \lambda \end{cases} . \tag{3.62}$$

Reported values of D have a lower bound of 1 hour (i.e., $D \ge 1$). Since event duration cannot be negative, λ must be in the range of $0 \le \lambda < 1$. The value of λ cannot be 1 hour because this implies $F(d = 1) = 0$ which is physically incompatible with the fact that 1 hour is the single most frequently observed value at 31.0 percent of the total observations as shown in Table B.4 in Appendix B. This same argument holds for the gamma and Weibull pdf's.

The pertinent population moments for estimation of parameters for the 2-parameter exponential pdf are given by,

$$\text{Mean, } \mu = \beta + \lambda \tag{3.63}$$

$$\text{Variance, } \sigma^2 = \beta^2 \tag{3.64}$$

Assuming that λ is known a priori, and using either the MOM or MLE, the estimator $\hat{\beta}$ is given by

$$\hat{\beta} = \bar{d} - \lambda \tag{3.65}$$

or by,

$$\hat{\beta} = s_d \tag{3.66}$$

where $\bar{d}$ and s_d are the sample mean and standard deviation of the observed data.

The general form, the population moments, and the parameter estimators for the 3-parameter gamma pdf and the 3-parameter Weibull pdf are as shown in the previous section. The only difference is that the range of possible values for λ is now $0 \leq \lambda < 1$. The computer program listed in Table B.7 of Appendix B was again applied to compute $\hat{\alpha}$ and $\hat{\beta}$ for the Weibull pdf.

Considerable effort was devoted to fitting the 2-parameter exponential pdf, 3-parameter gamma pdf, and 3-parameter Weibull pdf to the observed data for D. Several alternative values of λ were tried in an attempt to improve the fits. The parameter estimates were computed using the assumed value of λ and values of $\bar{d}$ and s_d of 4.4370 and 4.9213, respectively, from Table 3.4. In these trials, a graphical test of goodness of fit was used, as in the previous section, in order to give insight into the specific values of D where the greatest deviation from theoretical frequencies occur. In order to provide a more objective education of fit, the chi-square test of goodness of fit was also conducted in

each case. These analyses are summarized in Table 3.12. No acceptable fit to the data was obtained.

In every case, the graphical analysis showed that the theoretical cdf frequencies were lower than the observed frequencies over virtually the entire range of observed data. It was also found that the fit was extremely poor below about 8 hours and somewhat better, although still only fair, above that point. None of the 3 pdf's demonstrated an obviously superior ability to fit the data than the others.

The disparity between the goodness of fit between low and high values of D suggests that the lower values may represent a sample from a different pdf from that of the higher values, just as it did for the T data. A possible explanation is given for this at the end of this section. However, this does not explain why the theoretical cdf's are offset from the observed data toward lower frequencies.

A clue to the reason for the offset was obtained by inspection of the frequency data in Table B.4 in Appendix B. The values of D form an almost continuous sequence of integers from 1 hour to 35 hours: only missing values of 29 hours and 34 hours interrupt the sequence. Then, suddenly, the next value jumps to 42 hours, then a bigger leap to 56 hours, and finally all the way to 82 hours. These last 3 values are so far out of line with the remaining 2657 values of D that they have a highly disproportionate affect on the sample statistics. When taken together, the 3 outliers make-up only about one-tenth of one percent of the data. Omitting these 3 outliers from the data set solved the offset problem as demonstrated by the quality of the fit eventually achieved as shown below. Although it should be emphasized that these data are not erroneous nor does their omission necessarily imply that they are drawn from a different population than the rest of the data. They are simply extremely rare

TABLE 3.12
EVALUATION OF PDF'S FOR EVENT DURATION

Probability Density Function	Parameter Estimates $\hat{\alpha}$	$\hat{\beta}$	λ	Decision: Graphical CDF	Decision: Chi-Square Test
2-Parameter Exponential	–	3.4370	1.0	Do Not Accept H_0	Do Not Accept H_0
1-Parameter Exponential	–	4.4370	0.0	Do Not Accept H_0	Do Not Accept H_0
3-Parameter Gamma	.6400	6.1517	.50	Do Not Accept H_0	Do Not Accept H_0
2-Parameter Gamma	.8128	5.4585	0.0	Do Not Accept H_0	Do Not Accept H_0
3-Parameter Weibull	.7131	3.7628	1.0	Do Not Accept H_0	Do Not Accept H_0
3-Parameter Weibull	.8062	3.9939	.50	Do Not Accept H_0	Do Not Accept H_0
2-Parameter Weibull	.9030	4.2240	0.0	Do Not Accept H_0	Do Not Accept H_0

Null Hypothesis, H_0: The data are drawn from a pdf of the specified type with the specific parameters.

RAlternative Hypothesis, H_1: The data are not drawn from the specified pdf.

Decision Rule, Graphical: If, by visual inspection of the graph on log-normal probability paper, the cdf does not fit the observed frequencies then do not accept H_0. Otherwise, do not reject H_0.

Decision Rule, Chi-Square: Using a 0.05 level of significance, do not accept H_0 if the chi-square statistic exceeds the critical value (χ_0^2) for the appropriate degrees of freedom. Otherwise, do not reject Ho.

events which, by random chance, occurred during the period of record and thus distorted the sample statistics. The sample statistics for the random variable D were particularly susceptible to distortion by a few extremely large values since almost 50 percent of the observed values were either 1 hour or 2 hours.

Since the lowest values of D can effectively be considered drawn from a different population from the higher values, it was decided to utilize a composite pdf as was done for the T data. A continuous pdf was fit to the upper tail of the data and a discrete probability function was used for the lower values. The function fit to the upper tail is a conditional pdf which is conditional on the value of D being greater than or equal to some specified lower bound, d_l, which is as defined in the previous section.

A trial and error approach was used to determine the lowest value of d_l for which an acceptable fit to the upper tail could be achieved. Each time a new trial value of d_l was used, a different subset of the observed data became applicable for which $\bar{d}$ and s_d had to be determined in order to compute the applicable parameter estimates. Only two trial values of d_l, 2 hours and 3 hours, were required before getting an acceptable fit to the data using the simple 2-parameter exponential distribution. The sample statistics for these two values of d_l were as follows:

d_l (hours)	$\bar{t}$ (hours)	S_d (hours)	Number of Observations
2	5.8943	4.7336	1902
3	5.1581	4.8121	1436.

Since the exponential pdf is much simpler mathematically than the gamma and Weibull pdf's, and it fit the data acceptably, it was decided not to proceed with any further analyses for the more complex functions.

The single unknown parameter, β, of the 2-parameter exponential pdf was estimated three ways: using $\bar{d}$, then using s_d, and then computing the average of the first two. This can be stated mathematically as:

(3.67) $$\hat{\beta}_1 = \bar{d} - \lambda$$

(3.68) $$\hat{\beta}_2 = s_d$$

(3.69) $$\hat{\beta}_3 = (\hat{\beta}_1 + \hat{\beta}_2) / 2.$$

This was done for both trial values of d_l. As before, a graphical test of goodness of fit was used as the primary method of selecting the most suitable pdf. This analysis is summarized in Table 3.13. The value of λ was assumed known and was chosen to have a value equal to $(d_l - 1)$. A chi-square test for goodness of fit was conducted in each case to provide an objective alternative to the graphical analysis.

As shown in Table 3.13, the subset of data comprised of D values of 3 hours or larger can be acceptably fit by a 2-parameter exponential pdf with parameters,

$$\beta = 4.8121 \text{ hours}$$

$$\lambda = 2.0 \text{ hours.}$$

A visual test for goodness of fit of the theoretical exponential cdf to the observed cumulative frequencies is shown in Figure 3.5. The theoretical cdf is

TABLE 3.13
UPPER TAIL PDF's FOR EVENT DURATION

Data Subset ($D \geq d_l$)	Exponential Probability Density Function Parameters				Decision	
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	λ	Graphical CDF	Chi-Square Test
$D \geq 2$	4.8943	–	–	1.0	Do Not Accept H_o	Do Not Accept H_o
$D \geq 2$	–	4.7336	–	1.0	Do Not Accept H_o	Do Not Accept H_o
$D \geq 2$	–	–	4.8140	1.0	Do Not Accept H_o	Do Not Accept H_o
$D \geq 3$	5.1581	–	–	2.0	Do Not Accept H_o	Do Not Accept H_o
$D \geq 3$	–	4.8121	–	2.0	Do Not Reject H_0	Do Not Reject H_0
$D \geq 3$	–	–	4.9851	2.0	Do Not Accept H_o	Do Not Accept H_o

Null Hypothesis, H_0: The specified subset of data is drawn from a 2-parameter exponential pdf with the specified parameters.

Alternative Hypothesis, H_1: The specified subset of data is not drawn from the hypothesized pdf.

Decision Rule, Graphical: If, by visual inspection of the graph on log-normal probability paper, the cdf does not fit the observed cumulative frequencies of the subset of data then do not accept H_0. Otherwise, do not reject H_0.

Decision Rule, Chi-Square: Using a 0.05 level of significance, do not accept H_0 if the chi-square statistic exceeds the critical value (χ_0^2) for the appropriate degrees of freedom. Otherwise, do not reject H_0.

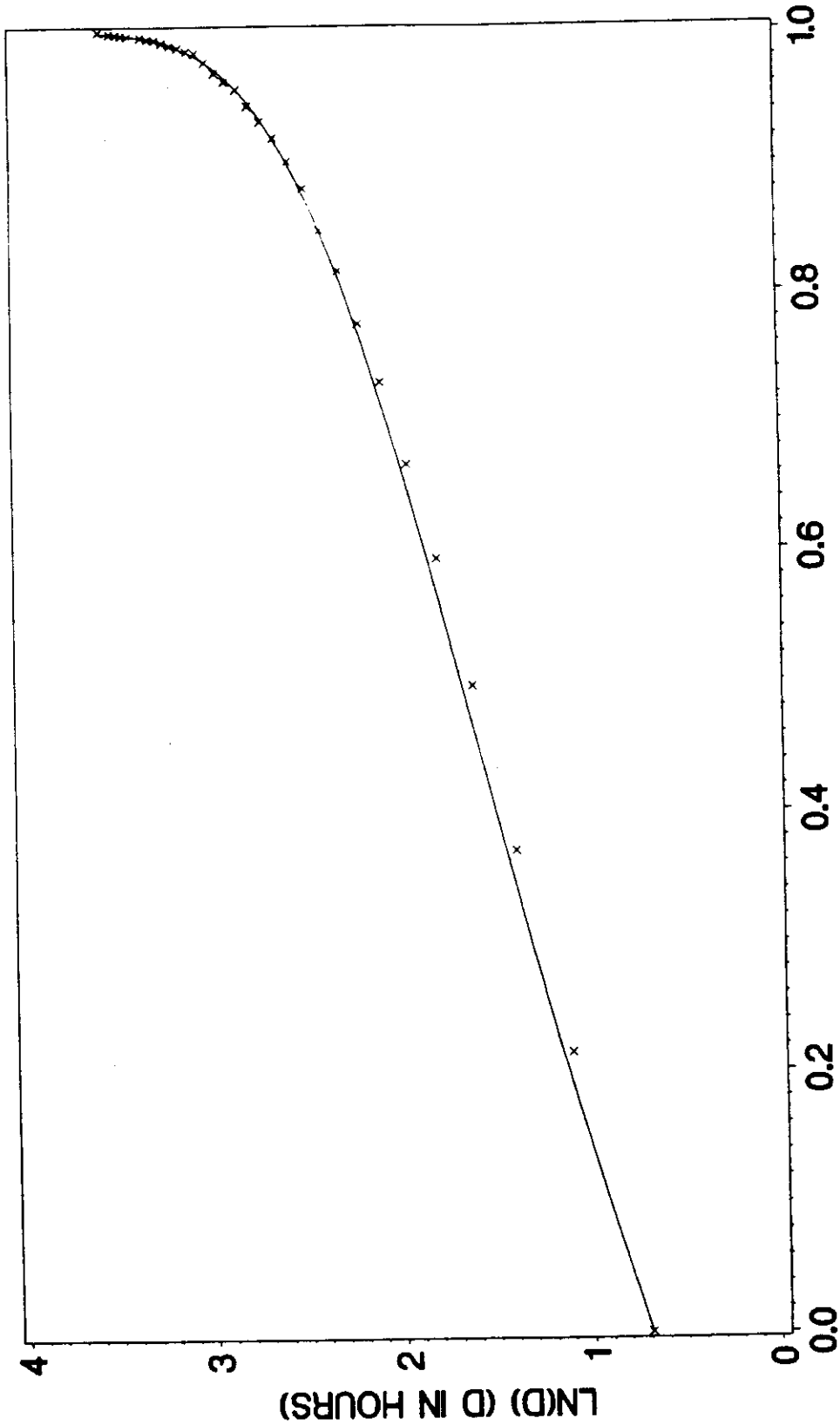


FIGURE 3.5
UPPER TAIL CDF FOR EVENT DURATION

shown as a smooth solid curve and the observed cumulative frequencies taken from Table B.4 in Appendix B are shown as the plotted data points. Visual inspection clearly demonstrates that there is good agreement between the theoretical cdf and the observed empirical cdf.

The chi-square test for goodness of fit for the fitted pdf is shown in detail in Table 3.14. Above a value of 19 hours, the data were grouped into classes so that an expected frequency of a least 10 was obtained for each class interval. As shown, the total chi-square statistic was 20.998 which lies outside the critical region of the test ($\chi_0^2 > 28.869$) at a 0.05 level of significance and 18 degrees of freedom. The decision rule suggests not to reject the null hypothesis that the subset of the observed data represent a random sample of size 1436 from a 2-parameter exponential pdf with the specified parameters.

The entire data set, that is lower and upper tails, for event duration can now be represented as a composite pdf. The upper tail of the distribution is characterized by the continuous 2-parameter exponential pdf just developed, and the lower tail (i.e., $1 \leq D < 3$) is characterized by a discrete probability function. The composite pdf takes the form:

$$f(d) = \begin{cases} 0, & d < 1 \\ p_1 & d = 1 \\ p_2 & d = 2 \\ \left(1 - p_1 - p_2\right) \left[\frac{1}{\beta} e^{-\left[\frac{d-\lambda}{\beta}\right]}\right], & d \geq 3 \end{cases} \tag{3.70}$$

The accompanying cdf is of the form:

TABLE 3.14

CHI-SQUARE TEST FOR UPPER TAIL PDF FOR EVENT DURATION

Exponential CDF

$\beta = 4.8121; \lambda = 2.0$

Value	Observed Frequency	Cumulative Frequency	Expected Frequency	Chi-Square Statistic
3	307	269	269	5.368
4	223	488	219	0.073
5	182	666	178	0.090
6	139	811	145	0.248
7	105	928	117	1.231
8	91	1023	95	0.168
9	65	1101	78	2.167
10	59	1164	63	0.254
11	45	1215	51	0.706
12	47	1256	41	0.878
13	30	1290	34	0.471
14	26	1317	27	0.037
15	18	1340	23	1.087
16	17	1358	18	0.056
17	18	1372	14	1.143
18	9	1384	12	0.750
19	9	1394	10	0.100
20-21	22	1408	14	4.571
22-23	6	1418	10	1.600
24-41	18	1436	18	0.000
	1436		1436	20.998

Null Hypothesis, H_0: The data represent a random sample from a population with a conditional ($D \geq 3$) 2-parameter exponential pdf with parameters $\beta = 4.8121$, $\lambda = 2.0$.

Alternative Hypothesis, H_1: The data were not drawn from a population that follows the specified pdf.

$\nu = (20 - 1) - 1 = 18$ (the parameter β was estimated from the data: the parameter λ was assumed known).

Total Chi-Square Statistic: $\chi^2 = 20.998$.

Critical Region at 0.05 Level of Significance: $\chi_0^2 > 28.869$.

Do not reject H_0.

$$(3.71)\quad F(d) = \begin{cases} 0, & d < 1 \\ p_1 & d = 1 \\ (p_1 + p_2) & d = 2 \\ (p_1 + p_2) + [1 - p_1 - p_2]\left[1 - e^{-\left[\frac{d-\lambda}{\beta}\right]}\right] & , \quad d \geq 3 \end{cases}$$

The estimators of p_1 and p_2 by either the MOM or MLE are:

$$(3.72)\qquad \hat{p}_1 = \frac{n_1}{n} \quad \text{and} \quad \hat{p}_2 = \frac{n_2}{n}$$

where,

n_1 = number of observations with D = 1

n_2 = number of observations with D = 2

n = total number of observations.

The estimates of p_1 and p_2 were computed using the frequency data of Table B.4 in Appendix B. The resulting final pdf is then:

$$(3.73)\qquad f(d) = \begin{cases} 0, & d < 1 \\ .3098, & d = 1 \\ .1688, & d = 2 \\ .5214\left[\frac{1}{4.8121}\, e^{-\left[\frac{d-2.0}{4.8121}\right]}\right] & , \quad d \geq 3\,. \end{cases}$$

The associated final cdf is then given by:

$$(3.74)\qquad F(d) = \begin{cases} 0, & d < 1 \\ .3098, & d = 1 \\ .4786, & d = 2 \\ .4786 + .5214\left[1 - e^{-\left[\frac{d-2.0}{4.8121}\right]}\right], & d \geq 3 . \end{cases}$$

As a final check, a chi-square test of goodness of fit was conducted for the final composite pdf. This analysis is shown in detail in Table 3.15. The total chi-square statistic had a value of 21.125 which lies outside the critical region of the the test ($\chi_0^2 > 28.869$) at the 0.05 level of significance and 18 degrees of freedom. The decision rule suggests not to reject the null hypothesis that the data are drawn from the specified composite pdf. Notice that 3-parameters were estimated from the data: β for the upper tail continuous pdf and p_1 and p_2 for the lower tail discrete probabilities.

The final composite cdf is shown graphically in Figure 3.6 along with the observed empirical cdf. The agreement is excellent.

It now seems appropriate to address the question as to why the values of D below 3 hours appear to represent a different pdf from the values above. This occurs because of the way D is defined and reported as an integer number of clock hours. This means that for any given reported value of D above 1 hour, the actual sequential time elapsed, D_e, can be as much as 2 hours shorter. When D = 1, D_e can be as much as 1 hour shorter. This can be stated mathematically as,

TABLE 3.15

CHI-SQUARE TEST FOR FINAL COMPOSITE PDF FOR EVENT DURATION

Value	Observed Frequency	Final CompositeCDF for D Cumulative Frequency	Expected Frequency	Chi-Square Statistic
1	855	855	855	0.000
2	466	1321	466	0.000
3	307	1590	269	5.368
4	223	1809	219	0.073
5	182	1987	178	0.090
6	139	2131	144	0.174
7	105	2249	118	1.432
8	91	2344	95	0.168
9	65	2422	78	2.167
10	59	2485	63	0.254
11	45	2536	51	0.706
12	47	2577	41	0.878
13	30	2611	34	0.471
14	26	2638	27	0.037
15	18	2661	23	1.087
16	17	2679	18	0.056
17	18	2693	14	1.143
18	9	2705	12	0.750
19	9	2715	10	0.100
20-21	22	2729	14	4.571
22-23	6	2739	10	1.600
24-41	18	2757	18	0.000
	2757		2757	21.125

Null Hypothesis, H_0: The data represent a random sample from a population with a composite discrete-continuous 2-parameter exponential pdf with parameters $p_1 = .3098$, $p_2 = .1688$, $\beta = 4.8121$, $\lambda = 2.0$.

Alternative Hypothesis, H_1: The data were not drawn from a population that follows the specified pdf.

$\nu = (22 - 1) - 3 = 18$ (the parameters p_1, p_2, and β were estimated from the data: the parameter λ was assumed known).

Total Chi-Square Statistic: $\chi^2 = 21.125$.

Critical Region at 0.05 Level of Significance: $\chi_0^2 > 28.869$.

Decision: Do no reject H_0.

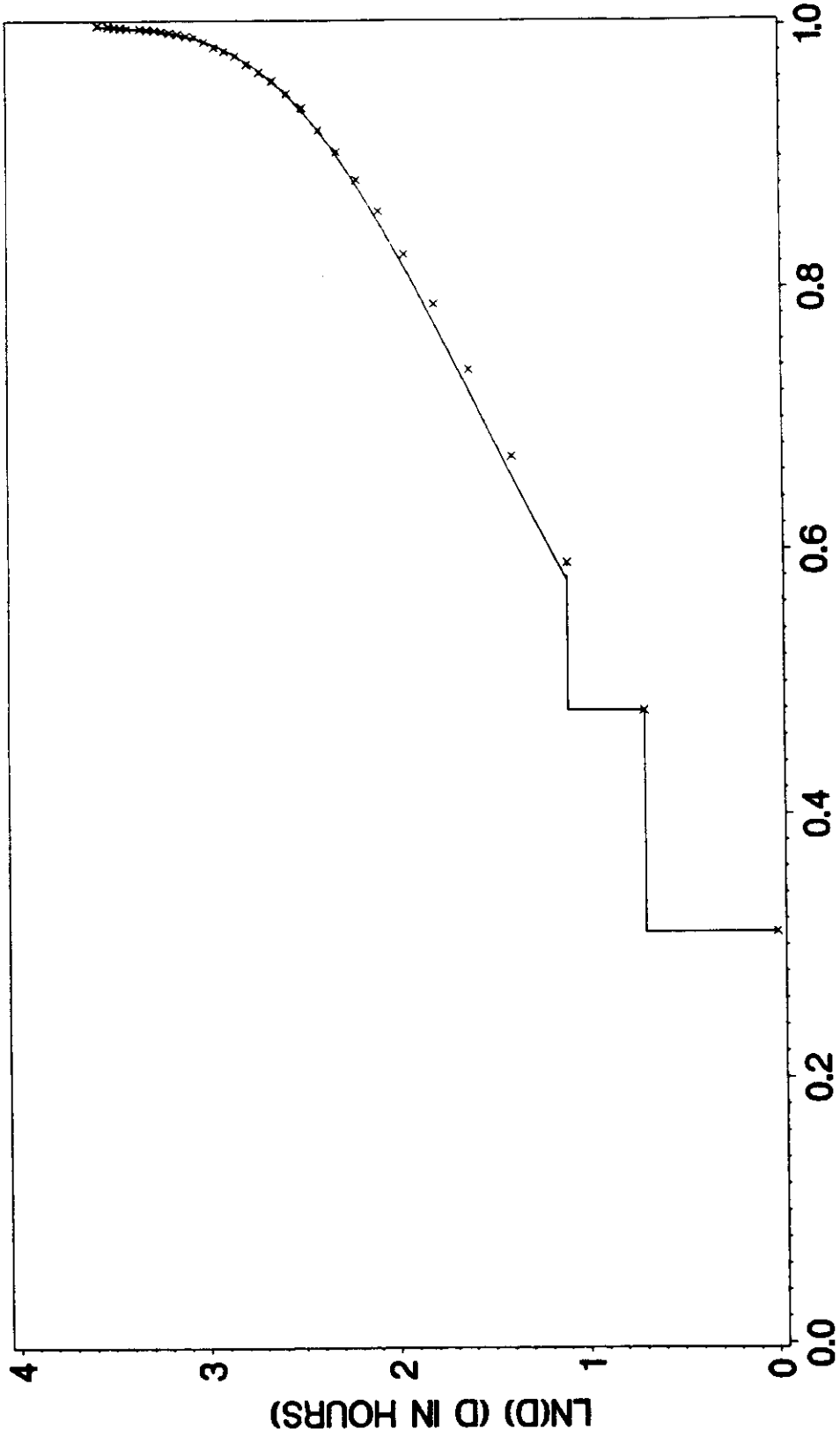


FIGURE 3.6
FINAL CDF FOR EVENT DURATION

(3.75) $D = 1$ implies $0 \leq D_e \leq 1$

$D = i$ implies $(i-2) \leq D_e \leq i$ $(i = 2,3,...)$.

This has been demonstrated graphically in Figure 3.7 for the specific cases of D = 2 and D = 5. As shown, the only way for the actual sequential elapsed time of a continuous period of rainfall to equal the event duration is for the event to start exactly on a clock hour and then end exactly on some later clock hour. Otherwise, the elapsed time will be less than the event duration, and can be as much as 2 hours less.

The 2 hour range of difference between D and D_e is relatively large for small values of D. At D = 1 and D = 2 the actual elapsed time can be as much as 100 percent less than D. In fact, the observed data reveal such an extreme tendency toward high frequencies at the lowest values of duration that it is very likely that the great majority of events with D = 1 and D = 2 have an actual elapsed time that is only a tiny fraction of D. On the other hand, the relative effect of the 2 hour difference declines rapidly as D increases: for D = 5 the actual elapsed time can be at most only 20 percent less. Basically, the larger the value of D, the closer it is to the true elapsed time duration of the event. The net result is that values of D below 3 hours effectively comprise a random sample from a different population than the values above.

Preliminary Simulation of Rainfall Event Variables

The purpose of fitting theoretical pdf's to the historical rainfall event data was to allow Monte Carlo simulation techniques to be used to generate multiple synthetic sequences of rainfall data. To accomplish this, a SAS computer program was written to randomly sample the fitted theoretical distributions for R, T, and D. The program is listed in Table B.8 in Appendix B.

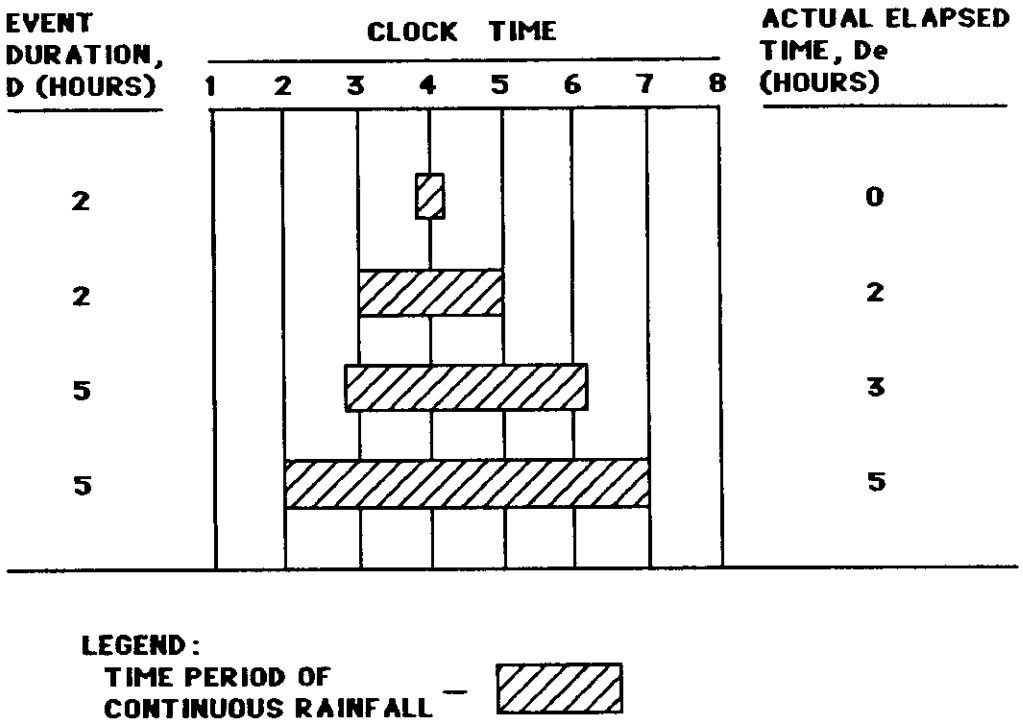


FIGURE 3.7
SAMPLE RAINFALL PATTERNS FOR D

The program uses the SAS internal function for generating gamma deviates to randomly sample the theoretical pdf for R. Once an R value is generated, it is rounded upwards to the nearest integer multiple of 0.01 inch in order to mimic the step function behavior of the empirical cdf of the observed R data.

The SAS internal function for sampling a uniform cdf on the interval 0 to 1 was used to map into the discrete portion of the composite cdf for T. For example, if a uniform random number was generated with a value at 0.2000, then $T = 6$. This is true because the probability that $T \leq 5$ is 0.1830 and the probability that $T \leq 6$ is 0.2145, therefore, the generated value of T lies in the interval $5 < T \leq 6$, but T has only integer values so that it must be that $T = 6$. If the generated uniform random number exceeds 0.3102, then the upper tail continuous gamma pdf applies and it is subsequently sampled using the SAS internal function for generating gamma deviates. Once an upper tail T value is generated, it is rounded upwards to the nearest integer to mimic the step function behavior of the empirical cdf for the observed T data.

For sampling D values, the SAS internal function for sampling a uniform cdf on the interval 0 to 1 was used to map into the discrete portion of the composite pdf in the exact manner used for T values. If the generated uniform deviate exceeds 0.4786, then the upper tail continuous exponential pdf applies and it is subsequently sampled using the SAS internal function for generating exponential deviates. Once an upper tail D value is generated, it is rounded upwards to the nearest integer to mimic the step function behavior of the empirical cdf for the observed D data.

More detail on the nature and application of SAS random number functions is given in the chapter dealing with simulations.

The SAS program was used to generate a sequence of 2760 values for each of the rainfall event variables R, T, and D. If the fitting of theoretical pdf's has truly been successful, then these generated sets of data should have the same statistical characteristics as the observed data and, hence, the same as the fundamental stochastic process that produced the observed data. This can be tested by application of the chi-square test of homogeneity.

Generally, the chi-square test of homogeneity tests the null hypothesis that multiple sequences of sample data are random samples from the same population without specifying the underlying distribution. This hypothesis can be written mathematically as follows:

$$H_o: p_{i1} = p_{i2} = \ldots = p_{ij} = \ldots = p_{ic} \qquad (i = 1, 2, \ldots, m) \quad (j = 1, 2, \ldots, c) \tag{3.76}$$

where,

P_{ij} = theoretical probability that the random variable has the i-th class value of the distribution in the j-th realization

m = number of class values used to categorize the data

c = number of realizations being compared.

Then, the chi-square statistic is computed from (Daniel 1978, Hoel 1984),

$$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{c} \left[\frac{\left(n_{ij} - n\, p_{ij}\right)^2}{n\, p_{ij}} \right] \tag{3.77}$$

where,

n_{ij} = number of observations with the i-th class value in the j-th realization

n = total number of observations in the c realizations

$n\,p_{ij}$ = expected frequency in the i-th class interval in the j-th realization for a random sample of size n.

The estimators of the p_{ij} by MLE are given by,

(3.78) $$\hat{p}_{ij} = \frac{n_{i\cdot}\, n_{\cdot j}}{n^2}$$

where,

$n_{i\cdot}$ = total number of observations with the i-th class value for all realizations

$n_{\cdot j}$ = total number of observations in the j-th realization for all class values.

Substituting, the chi-square statistic is then given by,

(3.79) $$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{c} \left[\frac{\left(n_{ij} - \frac{n_{i\cdot}\, n_{\cdot j}}{n} \right)^2}{\frac{n_{i\cdot}\, n_{\cdot j}}{n}} \right].$$

Since there are (mxc) values of $\hat{p}_{ij}$ to be estimated from the data, the number of degrees of freedom is $\nu = (m - 1)\,(c - 1)$.

In this analysis, the observed data will be compared to a single synthetically generated sequence of data of the same sample size. This means that c = 2, and the preceding equation for the chi-square statistic can be re-written as,

$$(3.80) \qquad \chi^2 = \sum_{i=1}^{m} \frac{\left(n_{i1} - \frac{n_{i.}\, n_{.1}}{n}\right)^2}{\frac{n_{i.}\, n_{.1}}{n}} + \sum_{i=1}^{m} \frac{\left(n_{i2} - \frac{n_{i.}\, n_{.2}}{n}\right)^2}{\frac{n_{i.}\, n_{.2}}{n}}$$

or, more briefly as,

$$\chi^2 = \chi^2_{.1} + \chi^2_{.2}$$

where,

$\chi^2_{.1}$ = sum of the chi-square statistics for all class values for the first realization

$\chi^2_{.2}$ = sum of the chi-square statistics for all class values for the second realization.

Since, in this case, both realizations have the same number of observations (2760) then,

$$(3.81) \qquad n_{.1} = n_{.2} = 1/2\, n$$

so that,

$$(3.82) \qquad \frac{n_{i.}\, n_{.1}}{n} = \frac{n_{i.}}{2} = \frac{n_{i1} + n_{i2}}{2}$$

and that,

$$(3.83) \qquad \frac{n_{i.}\, n_{.2}}{n} = \frac{n_{i.}}{2} = \frac{n_{i1} + n_{i2}}{2} \quad .$$

This simply means that the estimated theoretical frequency of occurrence for the i-th class value in either realization is just the average of the two observed frequencies. This means that when only two realizations are compared,

(3.84) $$\chi^2_{.1} = \chi^2_{.2} \quad .$$

The chi-square test for homogeneity between the observed data and simulated data for R is shown in detail in Table 3.16. The total chi-square statistic has a value of 12.764 which lies well outside the critical region of the test ($\chi^2_0 > 33.924$) at the 0.05 level of significance and 22 degrees of freedom. In fact, the test statistic would even lie outside the critical region at more than the 0.90 level of significance. Therefore, do not reject the null hypothesis that the two sets of data are homogeneous, that is, that they are random samples from the same population.

The chi-square test for homogeneity between the observed data and simulated data for T is shown in Table 3.17. The total chi-square statistic was 17.716 which lies well outside the critical region of the test ($\chi^2_0 > 41.337$) at the 0.05 level of significance and 28 degrees of freedom. The test statistic would even lie outside the critical region at more than the 0.90 level of significance. Therefore, do not reject the null hypothesis that the two sets of data are drawn from the same population.

The chi-square test for homogeneity between the observed data and simulated data for D is shown in detail in Table 3.18. The total chi-square statistic was 10.638 which lies well outside the critical region of the test ($\chi^2_0 > 33.924$) at the 0.05 level of significance and 22 degrees of freedom. In fact, the test statistic would even lie outside the critical region at more than the

TABLE 3.16

CHI-SQUARE TEST FOR HOMOGENEITY FOR RAINFALL VOLUME

		Column (j = 1)	Column (j = 2)		Chi-Square Statistic	
Range	Row	Observed Frequency	Simulated Frequency	$n_{i.}$	χ^2_{i1}	χ^2_{i2}
0.01	i = 1	435	464	899	0.468	0.468
0.02	i = 2	225	230	455	0.027	0.027
0.03	i = 3	131	136	267	0.047	0.047
0.04	i = 4	119	104	223	0.504	0.504
0.05	i = 5	110	84	194	1.742	1.742
0.06-0.10	i = 6	286	276	562	0.089	0.089
0.11-0.15	i = 7	185	190	375	0.033	0.033
0.16-0.20	i = 8	135	138	273	0.016	0.016
0.21-0.25	i = 9	100	112	212	0.340	0.340
0.26-0.30	i = 10	110	106	216	0.037	0.037
0.31-0.40	i = 11	154	148	302	0.060	0.060
0.41-0.50	i = 12	106	116	222	0.225	0.225
0.51-0.60	i = 13	115	95	210	0.952	0.952
0.61-0.70	i = 14	74	76	150	0.013	0.013
0.71-0.80	i = 15	65	67	132	0.015	0.015
0.81-0.90	i = 16	55	53	108	0.019	0.019
0.91-1.00	i = 17	47	41	88	0.205	0.205
1.01-1.50	i = 18	145	144	289	0.002	0.002
1.51-2.00	i = 19	69	86	155	0.932	0.932
2.01-2.50	I = 20	38	37	75	0.007	0.007
2.51-3.00	I = 21	22	25	47	0.096	0.096
3.01-3.50	i = 22	17	12	29	0.431	0.431
3.51-10.53	i = 23	17	20	37	0.122	0.122
		$n_{.1}$=2760	$n_{.2}$=2760	n=5520	$\chi^2_{.1}$=6.382	$\chi^2_{.2}$=6.382

Null Hypothesis, H_0: The synthetic data and observed data populations are homogeneous.

Alternative Hypothesis, H_1: The populations are not homogeneous.

$\nu = (23-1)(2-1) = 22$ (the 46 p_{ij} parameters were estimated from the data).

Total Chi-Square Statistic: $\chi^2 = \chi^2_{.1} + \chi^2_{.2} = 12.764$.

Critical Region at 0.05 Level of significance: $\chi^2_0 > 33.924$.

Decision: Do not reject H_0.

TABLE 3.17

CHI-SQUARE TEST FOR HOMOGENEITY FOR TIME BETWEEN EVENTS

		Column (j = 1) Observed	Column (j = 2) Simulated		Chi-Square Statistic	
Range	Row	Frequency	Frequency	$n_{.1}$	χ^2_{i1}	χ^2_{i2}
3	i = 1	231	231	462	0.000	0.000
4	i = 2	156	157	313	0.002	0.002
5	i = 3	118	120	238	0.008	0.008
6	i = 4	87	82	169	0.074	0.074
7	i = 5	73	64	137	0.296	0.296
8	i = 6	64	69	133	0.094	0.094
9	i = 7	49	57	106	0.302	0.302
10	i = 8	29	31	60	0.033	0.033
11	i = 9	49	45	94	0.085	0.085
12	i = 10	49	41	90	0.356	0.356
13	i = 11	30	47	77	1.877	1.877
14	i = 12	27	32	59	0.212	0.212
15	i = 13	36	34	70	0.029	0.029
16-20	i = 14	142	125	267	0.541	0.541
21-25	i = 15	92	101	193	0.210	0.210
26-30	i = 16	78	81	159	0.028	0.028
31-35	i = 17	71	80	151	0.268	0.268
36-40	i = 18	51	54	105	0.043	0.043
41-50	i = 19	121	120	241	0.002	0.002
51-60	i = 20	108	98	206	0.243	0.243
61-70	i = 21	101	82	183	0.986	0.986
71-80	i = 22	72	77	149	0.084	0.084
81-90	i = 23	62	78	140	0.914	0.914
91-100	i = 24	67	69	136	0.015	0.015
101-200	i = 25	389	414	803	0.389	0.389
201-300	i = 26	183	182	365	0.001	0.001
301-400	i = 27	123	98	221	1.414	1.414
401-500	i = 28	48	41	89	0.275	0.275
501-1200	i = 29	54	50	104	0.077	0.077
		$n_{.1}$=2760	$n_{.2}$=2760	5520	$\chi^2_{.1}$ =8.858	$\chi^2_{.2}$=8.858

Null Hypothesis, H_0: The synthetic data and observed data populations are homogeneous.

Alternative Hypothesis, H_1: The populations are not homogeneous.

$\nu = (29-1)(2-1) = 28$ (the 58 p_{ij} parameters were estimated from the data).

Total Chi-Square Statistic: $\chi^2 = \chi^2_{.1} + \chi^2_{.2} = 17.716$.

Critical Region at 0.05 Level of Significance: $\chi^2_0 > 41.337$.

Decision: Do not reject H_0.

TABLE 3.18

CHI-SQUARE TEST FOR HOMOGENEITY FOR EVENT DURATION

Range	Row	Column (j = 1) Observed Frequency	Column (j = 2) Simulated Frequency	$n_{i.}$	Chi-Square Statistic χ^2_{i1}	Chi-Square Statistic χ^2_{i2}
1	i = 1	855	837	1692	0.096	0.096
2	i = 2	466	465	931	0.001	0.001
3	i = 3	307	274	581	0.937	0.937
4	i = 4	223	234	457	0.132	0.132
5	i = 5	182	189	371	0.066	0.066
6	i = 6	139	145	284	0.063	0.063
7	i = 7	105	103	208	0.010	0.010
8	i = 8	91	93	184	0.011	0.011
9	i = 9	65	83	148	1.095	1.095
10	i = 10	59	67	126	0.254	0.254
11	i = 11	45	48	93	0.048	0.048
12	i = 12	47	37	84	0.595	0.595
13	i = 13	30	33	63	0.071	0.071
14	i = 14	26	24	50	0.040	0.040
15	i = 15	18	23	41	0.305	0.305
16	i = 16	17	18	35	0.014	0.014
17	i = 17	18	16	34	0.059	0.059
18	i = 18	9	8	17	0.029	0.029
19-20	i = 19	21	18	39	0.115	0.115
21-22	i = 20	12	19	31	0.790	0.790
23-24	i = 21	7	11	18	0.444	0.444
25-30	i = 22	10	8	18	0.111	0.111
31-82	i = 23	8	7	15	0.033	0.033
		$n_{.1}$=2760	$n_{.2}$=2760	n=5520	$\chi^2_{.1}$=5.319	$\chi^2_{.2}$=5.319

Null Hypothesis, H_0: The synthetic data and observed data populations are homogeneous.

Alternative Hypothesis, H_1: The populations are not homogeneous.

$\nu = (23-1)(2-1) = 22$ (the 46 p_{ij} parameter were estimated from the data).

Total Chi-Square Statistic: $\chi^2 = \chi^2_{.1} + \chi^2_{.2} = 10.638$.

Critical Region at 0.05 Level of Significance: $\chi^2_0 > 33.924$.

Decision: Do not reject H_0.

0.95 level of significance. Notice that the 3 outliers identified in the previous section have been re-inserted into the observed data set.

It can be concluded that the program written to simulate R, T, and D using the fitted theoretical pdf's does, in fact, generate randomly sampled sequences of rainfall event data that are drawn from the same populations as the actual observed data.

Independence of Rainfall Event Variables

In order to make the development of a stochastic model of the rainfall-runoff process straightforward, and to make the resulting model useful in engineering practice, it is important to demonstrate, or at least reasonably assume, independence between the rainfall event variables. Random variables can be tested for independence using the chi-square test of independence.

Generally, the chi-square test of independence uses a two-way contingency table. The data in a contingency table consist of n pairs of observations. Each pair of observations must represent two different measurements on the same object or event. The null hypothesis is that the two sets of observations represent random samples from populations of two independent random variables. Using the definition of independent random variables, the null hypothesis can be written mathematically as follows:

$$H_0: \; p_{ij} = p_{i\cdot}\, p_{\cdot j} \qquad \begin{matrix} (i = 1,2,\ldots,m) \\ (j = 1,2,\ldots,c) \end{matrix} \tag{3.85}$$

where,

p_{ij} = theoretical probability than an (X,Y) data pair lies in the i-th row and j-th column (i.e., the ij-th cell) of the contingency table

$p_{i\cdot}$ = theoretical probability that an (X,Y) data pair lies in the i-th row

$p_{\cdot j}$ = theoretical probability that on (X,Y) data pair lies in the j-th column

m = number of rows in the contingency table

c = number of columns in the contingency table

X,Y = any two arbitrary numbers of the set of pertinent random variables.

In this specific application, X and Y represent members of the set {R,T,D}.

The chi-square test statistic is given by (Daniel 1978, Hoel 1984),

$$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{c} \left[\frac{\left(n_{ij} - np_{ij}\right)^2}{np_{ij}} \right] \tag{3.86}$$

where,

n_{ij} = observed frequency of occurrence for the ij-th cell of the table

n = total number of observations (i.e., the number of X,Y pairs)

np_{ij} = expected frequency in the ij-th cell for a random sample of size n given that X and Y are independent.

Under H_0, it is known that,

$$p_{ij} = p_{i\cdot}\, p_{\cdot j} \quad . \tag{3.87}$$

Using MLE, the estimators of $p_{i\cdot}$ and $p_{\cdot j}$ are as follows:

$$\hat{p}_{i\cdot} = \frac{n_{i\cdot}}{n} \text{ and } \hat{p}_{\cdot j} = \frac{n_{\cdot j}}{n} \tag{3.88}$$

where,

$n_{i\cdot}$ = total number of observations in the i-th row

$n_{\cdot j}$ = total number of observations in the j-th column.

Substituting these estimators into the equation for the test statistic yields,

$$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{c} \left[\frac{\left(n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n} \right)^2}{\frac{n_{i\bullet} n_{\bullet j}}{n}} \right] . \tag{3.89}$$

The number of degrees of freedom is $\nu = (m-1)(c-1)$.

Contingency tables were created for (R,T), (T,D), and (R,D) as shown in Tables B.9, B.10, and B.11 in Appendix B. The format of the tables is such that each cell contains the observed frequency, theoretical expected frequency, chi-square statistic, percent of total observations located in the cell, percent of the row total of observations located in the cell, and percent of column total of observations located in the cell. Also, at the end of each row and column is shown the total number of observations in that row or column, and the percent of total observations in that row or column. The last page of each table contains summary test statistics including the total chi-square statistic.

The contingency tables were constructed with uneven class intervals for both the rows and column. This was done to maintain a minimum expected frequency of about 5 in every cell and yet to provide a sufficient number of cells for the large sample size.

The contingency table for (R,T), of dimension 11 by 8, is shown in Table B.9 of Appendix B. The chi-square test of independence for R and T can be summarized as follows:

Null Hypothesis, H_0: The random variables R and T are independent.

Alternative Hypothesis, H_1: The random variables R and T are not independent.

Degrees of Freedom: $\nu = (11-1)(8-1) = 70$.

Total Chi-Square Statistic: $\chi^2 = 77.725$.

Critical Region at 0.05 Level of Significance: $\chi_0^2 > 90.531$.

Do Not Reject H_0.

The chi-square test statistic lies well outside the critical region of the test. In fact, the test statistic would lie outside the critical region even at the 0.246 level of significance. It can be concluded, with a high degree of confidence, that R and T can be treated as independent random variables in the modeling analysis.

The contingency table for (T,D), of dimension 8 by 10, is shown in Table B.10 of Appendix B. The chi-square test of independence for T and D can be summarized as follows:

Null Hypothesis, H_0: The random variables T and D are independent.

Alternative Hypothesis, H_1: The random variables T and D are not independent.

Degrees of Freedom: $\nu = (8-1)(10-1) = 63$.

Total Chi-Square Statistic: $\chi^2 = 75.672$.

Critical Region at 0.05 Level of Significance: $\chi_0^2 > 82.529$.

Decision: Do Not Reject H_0.

The chi-square test statistic lies well outside the critical region of the test. The null hypothesis would not be rejected even at the 0.131 level of significance. It can be concluded, with a high degree of confidence, that T and D can be treated as independent random variable in the modeling analysis.

The contingency table for (R,D), of dimension 11 by 10, is shown in Table B.11 in Appendix B. The chi-square test of independence for R and D can be summarized as follows:

Null Hypothesis, H_0: The random variables R and D are independent.

Alternative Hypothesis, H_1: The random variables R and D are not independent.

Degrees of Freedom: $\nu = (11-1)(10-1) = 90$.

Total Chi-Square Statistic: $\chi^2 = 1739.126$.

Critical Region at 0.05 Level of Significance: $\chi_0^2 > 113.145$.

Do Not Accept H_0.

It cannot be concluded from this test that R and D are independent random variables.

A visual inspection of Table B.11 shows that events with small values of R tend to have small values of D and large values of R tend to have large values of D. At first glance, it seems intuitively obvious that the longer it rains the larger the amount of rain received. Unfortunately, the physics of rainfall is not that simplistic. Rainfall intensity, that is rate of rainfall, is also an important consideration. Physically, it is known that very high intensity rainfalls tend to be of relatively short duration (e.g. thunderstorms), which means that relatively large volumes can sometimes be measured in a short time. On the other hand, very long duration events tend to be of fairly low intensity (e.g. weather front storms). This means that modest volumes can sometimes be measured over a fairly extended period of time. These factors, among others, account for the poor correlation between R and D previously discussed and demonstrated in Table 3.5.

A detailed evaluation of the correlation between R and D is imperative. Several sample correlation coefficient matrices for various subsets of the observed data are shown in Table 3.19. The variable T is included only for comparison and completeness. The sample mean and standard deviation for

TABLE 3.19

CORRELATION COEFFICIENT MATRICES FOR DATA SUBSETS

	Correlation Coefficient Matrix			Statistics	
Row	R	T	D	Mean	Standard Deviation
	ALL RAIN VOLUME VALUES (n=2757)				
R	1.0000	-.0119	.5829	.3889	.6405
T	-.0119	1.0000	-.0132	92.899	136.717
D	.5829	-.0132	1.000	4.3765	4.5368
	EVENTS WITH VOLUME > .01 INCH (n=2322)				
R	1.000	-.0224	.5468	.4599	.6747
T	-.0224	1.0000	-.0268	94.828	138.815
D	.5468	-.0268	1.0000	5.0090	4.6801
	EVENTS WITH VOLUME > .10 INCH (n=1451)				
R	1.0000	-.0440	.4334	.7075	.7515
T	-.0440	1.0000	-.0433	96.641	139.624
D	.4334	-.0433	1.0000	6.6389	5.1565
	EVENTS WITH VOLUME > .18 INCH (n=1192)				
R	1.0000	-.0462	.4010	.8307	.7761
T	-.0462	1.0000	-.0334	95.674	139.483
D	.4010	-.0334	1.0000	7.1804	5.3664
	EVENTS WITH VOLUME > .28 INCH (n=962)				
R	1.0000	-.0314	.3740	.9736	.8002
T	-.0314	1.0000	.0043	91.619	127.059
D	.3740	.0043	1.0000	7.7422	5.5256

each variable are also shown to emphasize the overall effects of omitting selected subsets of the data.

The top matrix in Table 3.19 simply omits the three storm events for which D > 35 because they have already been shown to be outliers for computation of sample statistics involving the variable D. At best, the correlation between R and D can be classified as weak with a correlation coefficient of 0.5829. When events with rainfall volumes of 0.01 inch are omitted from the analysis, the correlation coefficient drops to 0.5468. These events unfairly bias the correlation between R and D because every event with a volume of 0.01 inch is constrained to have a duration of exactly one hour by the measuring apparatus and method of reporting the data. When events with volumes of 0.10 inch or less are omitted, the correlation coefficient drops farther to 0.4334. Then, as Table 3.19 shows, the correlation coefficient is a poor 0.4010 when events of 0.18 inch or less are omitted. This specific volume of 0.18 inch was chosen because no rain event with a volume of less than 0.19 inch produced runoff during the field study as discussed in Chapter 2. In fact, the three rains observed during the field study that had volumes near 0.20 inch produced only negligible runoff. This points out that the correlation between R and D is very weak in the range of R values which produce runoff of even the smallest amount.

Finally, the correlation coefficient is a very poor 0.3740 when events of 0.28 inch or less are omitted. This last volume of 0.28 inch was chosen as a data separation point because no rain event with a volume of less than 0.29 inch produced significant runoff (i.e., runoff equalling or exceeding 0.01 inch in volume) during the field study as discussed in Chapter 2. This clearly emphasizes that the correlation between R and D is weakest in the range of R

values which produce significant runoff. It is precisely these events which are the focus of this study.

The conclusion to be reached from this analysis is that R and D are dependent random variables, but that the correlation between them is extremely poor in the range of R values of most importance. Of course, poor correlation means that there is little predictability between R and D and that the dependence relationship between them is very weak.

In hydrologic research, two approaches have been commonly used to deal with the dependence of R and D. First, approximate conditional pdf's have been developed for R conditional on selected ranges of D (Eagleson 1971, 1972, and 1978). This approach suffers from the serious flaw that accurately fitting theoretical conditional pdf's to the small subsets of data is virtually impossible. Often, in order to provide a convenient and practical way to proceed, a mathematically simple conditional pdf type is assumed (usually an exponential pdf to facilitate analytical solutions). Invariably, the appropriate parameters are then estimated from the data subsets without regard to the goodness of fit. Obviously, this is a simplification which could introduce significant error.

The second approach is to simply assume that R and D are independent while recognizing this as a potential source of error in simulations (Koch 1985; Tarboton, Bras, and Puente 1987).

In this study, the latter approach has been used and it has been assumed that R and D are independent random variables. It is believed that this approach will introduce the least amount of error because of the excellent fit obtained with theoretical marginal pdf's for both R and D, and because the dependence between R and D is very weak.

Furthermore, the runoff variable that has the greatest influence on nonpoint source pollutant loads is runoff volume. The function derived for runoff volume in Chapter 4 does not include D as a random variable, thus, the independence assumption introduces little or no error into the estimation of runoff volume. The variable D enters the direct analysis of runoff only in evaluation of peak runoff rate.

In conclusion, it is believed that the assumption of independence between R and D introduces on acceptably small level of error into the rainfall-runoff modeling analysis for the purpose of estimating nonpoint source pollutant loads.

CHAPTER 4

BASIN RAINFALL - RUNOFF TRANSFORMATION FUNCTIONS

Derivation of Fundamental Runoff Equations

Dependent Runoff Variables

Simulation of runoff volumes and flow rates is essential for the estimation of NPS pollutant loads generated by rainfall-runoff events. Ideally, an entire runoff hydrograph, which provides a detailed time history of volumes and flow rates throughout a runoff event, would be generated for each simulated rainfall-runoff event. This would provide the greatest flexibility for estimation of NPS loads.

Generation of a complete runoff hydrograph for each simulated rainfall-runoff event has been accomplished in this study by simulation of total runoff volume and peak flow rate, which are the two runoff variables with obvious significant influence on NPS pollutant loads, coupled with a modification of the standard U.S. Soil Conservation Service (SCS) unit hydrograph shape. Consequently, two principal basin rainfall-runoff transformation functions have been derived for the study watershed with one equation having runoff volume as the dependent, or simulated, variable and the other with peak flow rate as the dependent variable.

Runoff Volume

The hydrologic budget equation developed around a watershed for a

single rainfall-runoff event can be written in its fundamental form as:

(4.1) $$Q = R - \Delta S$$

where,

Q = runoff volume for the event (inch)

R = rainfall volume for the event (inch)

ΔS = change in basin storage (inch).

This equation is based on the assumptions that evapotranspiration and groundwater seepage are negligible over the duration of a single rainfall event.

Basin storage can be divided into two components: initial abstraction and soil storage. Initial abstraction is that fraction of the rainfall volume that does not come in contact with the soil because it adheres to surfaces on vegetation and structures (i.e., interception) or is retained in depressions in impervious surfaces (e.g. pavements). Let A_i be the initial abstraction and modify equation 4.1 as follows:

$$Q = R - \Delta S + A_i - A_i$$

or,

$$Q = (R-A_i) - (\Delta S-A_i).$$

Now define,

$$\Delta S_s = \Delta S - A_i$$

where,

ΔS_s = change in soil storage

which upon substitution yields,

(4.2) $$Q = (R-A_i) - \Delta S_s.$$

The volume added to the soil, ΔS_s, for any given rainfall event that produces runoff is a function of the volume of rain available to be stored in the soil, $(R\text{-}A_i)$, the amount of water lost from the soil since the previous rainfall-runoff event by evapotranspiration and accretion to the deep groundwater, the amount of water gained by the soil from non-runoff producing rains occurring between the previous and the current rainfall-runoff events, and the rainfall intensity relative to the limiting soil infiltration rate. The loss of soil moisture by evapotranspiration and by accretion to the groundwater are, in turn, related to the elapsed time since the previous runoff event. All of this can be stated in a concise mathematical form as:

$$\Delta S_s = f\left[(R-A_i), E, A, (I - f_c)\right] \tag{4.3}$$

where,

E = elapsed time variable which is an index of evapotranspiration and groundwater accretion losses

A = antecedent rainfall variable which is an index of the water gained by the soil between consecutive rainfall-runoff events

I = intensity of rainfall variable

f_c = critical, or limiting, soil infiltration rate.

All other symbols are as defined previously. The critical infiltration rate for a particular soil, f_c, is a constant which represents the saturated soil permeability of the limiting layer of the subsurface soil column. Observed surface infiltration rates commonly exceed f_c during the first minutes of storm events, but this is due to soil storage effects which are quantified in the other three terms of the functional relationship.

Using a linear transformation function to describe the process yields,

$$\Delta S_s = \alpha\,(R - A_i) + \beta E - \gamma A - \delta\,(I - f_c) - \varepsilon_v \tag{4.4}$$

where,

$\alpha, \beta, \gamma, \delta$ = parameters of the linear model

ε_v = residual (i.e., random error or white noise component).

Algebraic signs for the parameters in this equation are dictated by the physical hydrologic principals that the volume of water entering the soil during a rainfall-runoff event increases with increasing available rainfall volume, $(R\text{-}A_i)$; increases with increasing elapsed time between events, E; decreases with increasing antecedent rainfall amounts, A; and decreases with increasing rainfall intensity over and above the limiting infiltration rate, $(I\text{-}f_c)$. Now, substituting the expression for ΔS_s into equation 4.2, yields,

$$Q = (R - A_i) - \left[\alpha\,(R - A_i) + \beta E - \gamma A - \delta\,(I - f_c) - \varepsilon_v\right] \tag{4.5}$$

which upon combining like terms yields,

$$Q = \left[(\alpha - 1)A_i - \delta f_c\right] + (1 - \alpha)R - \beta E + \gamma A + \delta I + \varepsilon_v \tag{4.6}$$

Assuming that A_i is constant from one event to the next, as is the custom in most hydrologic studies, allows this equation to be written in a more general form:

$$Q = \beta_0 + \beta_1 R + \beta_2 E + \beta_3 A + \beta_4 I + \varepsilon_v \tag{4.7}$$

where,

$$\beta_0 = (\alpha-1)A_i - \delta f_c$$

$$\beta_1 = (1-\alpha)$$

$$\beta_2 = -\beta$$

$$\beta_3 = \gamma$$

$$\beta_4 = \delta$$

This is the fundamental functional form that transforms rainfall event variables R, E, A, and I into runoff volume Q. Equation 4.7 emphasizes that runoff volume can be described as a linear function of no more than four independent variables, one of which must be the corresponding rainfall volume.

The β_j parameters (j = 0, 1, 2, 3, 4) and the residual term ε_V, which is a normally distributed random variable with zero mean, all depend on the hydrologic characteristics of the given watershed. These hydrologic characteristics, and hence, the parameters and residual term, are assumed to be time invariant.

The linear form of the transformation function, and the assumptions associated with its derivation, have been verified by the high correlation coefficient achieved with the final model and by objective statistical tests performed on the parameters and residuals of the model as shown in the "Runoff Volume Model" section below.

Peak Flow Rate

Unit hydrograph theory states that for a rainfall-runoff event resulting from a single rainfall pulse of uniform intensity and areal distribution, the direct runoff

flow rate at any time on the observed hydrograph is a linear function of total runoff volume. Specifically, then, the flow rate corresponding to the direct runoff hydrograph peak, q_p, can be written as,

(4.8) $$q_p = \lambda Q$$

where,

λ = parameter of the model.

However, this simple proportionality does not encompass the more general and common situation where the underlying assumptions of unit hydrograph theory (e.g. spatially and temporally uniform rainfall) are not precisely satisfied. In this case, the value of q_p could be considered to have a random error or noise component that results from the interaction of the random variables characterizing the rainfall event over and above the effects already incorporated into the observed value of Q. Assuming that the linear nature of the flow rate generation function of equation 4.8 still holds, this can be written as,

(4.9) $$q_p = \lambda Q + \varepsilon_f$$

where,

ε_f = residual or random error component.

Substituting the expression for Q from equation 4.6 and rearranging yields,

(4.10) $$q_p = \left[\lambda(\alpha-1)A_i - \lambda\delta f_c\right] + \lambda(1-\alpha)R - \lambda\beta E + \lambda\gamma A + \lambda\delta I + (\lambda\varepsilon_v+\varepsilon_f).$$

This equation can be written in a more general form as:

$$q_p = \beta_0 + \beta_1 R + \beta_2 E + \beta_3 A + \beta_4 I + \varepsilon_q \tag{4.11}$$

where,

$$\beta_0 = \lambda(\alpha-1)A_i - \delta f_c$$

$$\beta_1 = \lambda(1-\alpha)$$

$$\beta_2 = -\lambda\beta$$

$$\beta_3 = \lambda\gamma$$

$$\beta_4 = \lambda\delta$$

$$\varepsilon_q = \lambda\varepsilon_v + \varepsilon_f$$

This linear model for peak flow rate is of an identical form to that for total runoff volume. Of course, the β_j and the residual component are different for the two models.

Equation 4.11 emphasizes that q_p can be represented by a linear model involving four independent variables which are measures or indices of the effects of rainfall-runoff event volume, elapsed time between events, antecedent moisture, and rainfall intensity plus some random effect.

The five β_j parameters and the residual term ε_q, which is normally distributed with zero mean, are all functions of the hydrologic characteristics of the watershed and are assumed to be time invariant.

The linear form of the transformation function for peak flow rate, and the assumptions underlying the derivation, have been verified by the high correlation coefficient achieved with the final model and by objective statistical tests performed on the parameters and residuals of the model as shown in the "Peak Flow Rate Model" section below.

Independent Variables

Potentially applicable independent variables for inclusion in the basin rainfall-runoff transformation functions for runoff volume and peak flow rate can be placed into one of four classifications or sets:

(1) a set of variables that measure the volume of water involved in the event

(2) a set of elapsed time variables which are indices of evapotranspiration and groundwater accretion losses from the soil between consecutive rainfall events

(3) a set of antecedent rainfall variables which are indices of the water gained by the soil of the watershed between consecutive rainfall-runoff events

(4) a set of rainfall intensity variables.

No more than one variable from each of these four sets can be included in a model because all of the variables in a given set are measures or indices of the same physical phenomenon.

The only pertinent element of the set of volume variables is the area weighted rainfall volume R_w. Values of R_w were computed using the Thiessen polygon method applied to rainfall measurements from the two rain gages in the study area as explained in Chapter 2.

The three elapsed time variables compiled for evaluation were:

(1) T = elapsed time between rainfall events (hours)

(2) T_l = elapsed time since the last runoff event (hours)

(3) T_r = elapsed time since the last runoff event with $Q \geq 0.01$ inch (hours).

Values of 9 different cumulative antecedent rainfall volume variables were compiled using field data from the study watershed using the following notation:

${}_iR_j$ = rainfall volume occurring between the end of day i and the end of day j (inch).

The 9 specific cumulative rainfall volume variables compiled were,

(1) ${}_0R_1$ = rainfall volume in the 24 hours (1 day) antecedent to the current event (inch)

(2) ${}_0R_2$ = rainfall volume in the 48 hours (2 days) antecedent to the current event (inch)

(3) ${}_0R_3$ = rainfall volume in the 72 hours (3 days) antecedent to the current event (inch)

(4) ${}_0R_4$ = rainfall volume in the 96 hours (4 days) antecedent to the current event (inch)

(5) ${}_0R_5$ = rainfall volume in the 120 hours (5 days) antecedent to the current event (inch)

(6) ${}_0R_6$ = rainfall volume in the 144 hours (6 days) antecedent to the current event (inch)

(7) ${}_0R_7$ = rainfall volume in the 168 hours (7 days) antecedent to the current event (inch)

(8) ${}_0R_{15}$ = rainfall volume in the 360 hours (15 days) antecedent to the current event (inch)

(9) ${}_0R_{31}$ = rainfall volume in the 744 hours (31 days) antecedent to the current event (inch).

Cumulative totals were employed so that a single index variable could represent the entire sequence of antecedent rainfall amounts. This approach

was confirmed as acceptable by a preliminary analysis that demonstrated that an entire chronological series of daily antecedent rainfall volumes (i.e., ${}_0R_1$, ${}_1R_2$, ${}_2R_3$, ${}_3R_4$, ${}_4R_5$, ${}_5R_6$, ${}_6R_7$, ${}_7R_{15}$, and ${}_{15}R_{31}$) was less effective than a single cumulative rainfall total for prediction of runoff volume.

Values were compiled for three rainfall intensity variables as follows:

(1) I_{max} = maximum hourly intensity during the rainfall event (inch/hour)

(2) I_w = weighted average intensity (inch/hour)

(3) I_a = average intensity, R_w/D (inch/hour).

Definitions of I_{max} and I_w were given in detail in Chapter 3.

Selection of the optimum predictor variable from each of the four sets of independent variables, and simultaneous optimization of the corresponding parameter estimates, is the objective of the methodology derived in the following section.

Variable Seclection and Parameter Optimization Methodology

Either of the two pertinent rainfall-runoff transformation functions can be represented mathematically in the form of a general linear model as follows:

$$Y = \beta_0 + \beta_1 V + \beta_2 E + \beta_3 A + \beta_4 I + \varepsilon \tag{4.12}$$

where,

Y = dependent variable

V = volume variable

E = elapsed time variable

A = antecedent rainfall variable

I = intensity variable

ε = residual (i.e., random error or noise component)

β_j = parameters of the model ($j = 0, 1, 2, 3, 4$).

Using set notation,

$$Y \in \{Q, q_p\}$$

$$V \in \{R_w\}$$

$$E \in \{T, T_l, T_r\}$$

$$A \in \{{}_0R_1, {}_0R_2, {}_0R_3, {}_0R_4, {}_0R_5, {}_0R_6, {}_0R_7, {}_0R_{15}, {}_0R_{31}\}$$

$$I \in \{I_{max}, I_w, I_a\}.$$

A value of each of these dependent and independent variables, among others, was measured for each rainfall-runoff event monitored in the field. For any given model, the data can be presented as follows:

$$Y_1 = \beta_0 + \beta_1 V_1 + \beta_2 E_1 + \beta_3 A_1 + \beta_4 I_1 + \varepsilon_1$$

$$Y_2 = \beta_0 + \beta_1 V_2 + \beta_2 E_2 + \beta_3 A_2 + \beta_4 I_2 + \varepsilon_2$$

$$\vdots$$

$$Y_n = \beta_0 + \beta_1 V_n + \beta_2 E_n + \beta_3 A_n + \beta_4 I_n + \varepsilon_n$$

where n is the number of rainfall-runoff events monitored. In matrix form, this can be shown as,

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & V_1 & E_1 & A_1 & I_1 \\ 1 & V_2 & E_2 & A_2 & I_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & V_n & E_n & A_n & I_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}.$$

This can be written more concisely in matrix notation as,

$$\underset{(n \times 1)}{\mathbf{Y}} = \underset{(n \times 5)}{\mathbf{X}} \quad \underset{(5 \times 1)}{\boldsymbol{\beta}} + \underset{(n \times 1)}{\boldsymbol{\varepsilon}} \tag{4.13}$$

where boldface type is used to distinguish a matrix from a real-valued variable. The dimension of each matrix is shown in parenthesis. This can be rearranged in terms of the column vector of residuals as,

$$\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{X}\,\boldsymbol{\beta} \tag{4.14}$$

In order to compute the optimum values of the estimators $\hat{\beta}_j$, the least squares optimization criterion can be used. That is, choose the $\hat{\beta}_j$ such that the sum of the squares of the ε_i is minimized. The sum of the squares of the ε_i is a real-valued linear function of the $\hat{\beta}_j$ and it can be represented as an inner product of vectors as follows:

$$\sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \tag{4.15}$$

where the prime denotes the transpose matrix. Substituting from the previous equation and using the estimator matrix $\hat{\boldsymbol{\beta}}$ in place of $\boldsymbol{\beta}$ yields,

$$\sum_{i=1}^{n} \varepsilon_i^2 = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \tag{4.16}$$

In order to minimize, take the partial derivative of both sides of this equation with respect to each of the $\hat{\beta}_j$ and set to zero. This produces five equations in terms of the five $\hat{\beta}_j$ unknowns. This can be written in matrix notation as,

$$\mathbf{0} = \frac{\partial}{\partial \hat{\boldsymbol{\beta}}} \left[(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}) \right] \tag{4.17}$$

where $\mathbf{0}$ is the null matrix of dimension (5x1) and the differential operator is defined as,

$$\frac{\partial}{\partial \hat{\boldsymbol{\beta}}} = \begin{bmatrix} \frac{\partial}{\partial \beta_0} \\ \frac{\partial}{\partial \beta_1} \\ \frac{\partial}{\partial \beta_2} \\ \frac{\partial}{\partial \beta_3} \\ \frac{\partial}{\partial \beta_4} \end{bmatrix}.$$

Differentiating the right-hand side yields,

$$\mathbf{0} = -2\mathbf{X}' (\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}).$$

Multiplying both sides of the equation by the scalar $-\frac{1}{2}$ and rearranging produces the normal equations,

(4.18) $$\mathbf{X'\ X}\hat{\boldsymbol{\beta}} = \mathbf{X'\ Y}.$$

Premultiplying both sides by the inverse matrix $(\mathbf{X'\ X})^{-1}$, assuming that it exists, yields,

(4.19) $$\hat{\boldsymbol{\beta}} = (\mathbf{X'\ X})^{-1}\ \mathbf{X'\ Y}.$$

This shows that the computation of the five optimum $\hat{\beta}_j$ values involves a simple matrix inversion of a (5x5) matrix and elementary matrix multiplication.

For future reference, notice that the matrix $(\mathbf{X'\ X})$ is a (5x5) symmetric matrix made up of the sums of squares and sums of cross products of the observed values of the independent variables. The sums of squares lie along the principal diagonal with the cross products symmetric about the diagonal as follows:

(4.20) $$\mathbf{X'X} = \begin{bmatrix} n & \Sigma V_i & \Sigma E_i & \Sigma A_i & \Sigma I_i \\ \Sigma V_i & \Sigma V_i^2 & \Sigma V_i E_i & \Sigma V_i A_i & \Sigma V_i I_i \\ \Sigma E_i & \Sigma E_i V_i & \Sigma E_i^2 & \Sigma E_i A_i & \Sigma V_i I_i \\ \Sigma A_i & \Sigma A_i V_i & \Sigma A_i E_i & \Sigma A_i^2 & \Sigma A_i I_i \\ \Sigma I_i & \Sigma I_i V_i & \Sigma I_i E_i & \Sigma I_i A_i & \Sigma I_i^2 \end{bmatrix}$$

where all summations are from i = 1 to i = n.

For any given model, it is not necessary that all of the β_j be non-zero. Stated another way, a model may contain less than the maximum of four independent variables. Under these circumstances, the matrix inversion may involve only a (4x4) matrix for three independent variables, a (3x3) matrix for two independent variables, or a (2x2) matrix for one independent variable.

Although at most four independent variables are to be included in any given model, each of the four can be chosen from an entire set of possible variables. These sets are as defined above. Choices must be made among one volume variable, three elapsed time variables, nine antecedent rainfall variables, and three rainfall intensity variables. This is a total of 16 independent variables from which to choose.

A stepwise least squares optimization procedure was employed to determine how many and which specific independent variables to include in a given model, and to simultaneously compute the optimum values of the pertinent parameter estimators $\hat{\beta}_j$ using equation 4.19. The optimization criterion employed for variable selection was to maximize the coefficient of determination, r^2, which was defined as,

$$r^2 = \frac{s^2_{\hat{y}}}{s^2_y} \tag{4.21}$$

where,

$s^2_{\hat{y}}$ = Sample variance of the model predicted values of the dependent variable

s^2_y = Sample variance of the observed values of the dependent variable.

Maximization of the r^2 value was chosen as the optimization criterion because r^2 represents the fraction of the variation in the dependent variable that is accounted for by the model. The multiple correlation coefficient, r, is defined simply as the positive square root of r^2.

The stepwise optimization procedure for choosing the best four independent variables from among a total of m independent variables involves the following sequence of steps:

(1) Begin by finding the model with one independent variable that has the largest r^2 value. All m possible models are evaluated and compared.

(2) While maintaining the first variable in the model, the specific variable that produces the greatest increase in r^2 is added to the model as the second independent variable. All of the possible choices for the second variable are evaluated and compared. There are at most (m-1) such choices. As a check, each of the two variables now in the model is compared to the remaining, at most (m-2), variables not in the model. For each comparison, the procedure determines if removing one variable and replacing it with the other increases the value of r^2. After comparing all, at most 2(m-2), possible switches, the switch that produces the largest increase in r^2 is made. Once the

switch is made, comparisons begin again and the process continues until no switch increases r^2. Thus, the best model with two independent variables is obtained.

(3) While maintaining the first two variables in the model, the specific variable that produces the greatest increase in r^2 is added to the model as the third independent variable. All of the possible, at most (m-2), choices for the third variable are evaluated and compared. As a check, each of the three variables in the model is compared to the remaining, at most (m-3), variables not in the model. The comparing and switching process described in Step (2) is repeated, there are at most 3(m-3) comparisons made in each iteration, until no switch increases r^2. This yields the best three variable model.

(4) Finally, a fourth independent variable, the one that increases r^2 the greatest, is added to the model. All of the possible, at most (m-3), choices are evaluated and compared. Again, as a check, each of the four variables now in the model is compared to the remaining, at most (m-4), variables not in the model. The comparing and switching process is repeated, there are at most 4(m-4) comparisons made in each iteration, until no switch increases r^2. The resulting model is the best four variable model.

(5) If the final four variable model contains only one variable from each of the four sets of independent variables, then the model is optimal. Otherwise, the j variables that were entered into the model prior to the specific choice that repeated a set are retained in the model. The

unselected variables contained in the sets from which the j selected variables were obtained are eliminated from further consideration. Then, the procedure returns to Step (j+1) with a new total of m' independent variables from which to choose where,

$$m' = m - \sum_{i=1}^{j} m_i \quad (j=1,2,3)$$

and where,

m_1 = number of variables in the set corresponding to the first variable entered into the model

m_2 = number of variables in the set corresponding to the second variable entered into the model

m_3 = number of variables in the set corresponding to the third variable entered into the model.

For example, suppose that the first three variables selected were R_w, T_r, and I_a and then the fourth variable selected was I_w. The fourth choice repeated a selection from the set of intensity variables. In this case, the first three variables selected are retained in the model (i.e., j=3) and all other variables in the sets from which these three were obtained are eliminated from further consideration. The procedure returns to Step 4 (i.e., j+1=4) for the selection of the fourth variable from the set of nine (i.e., m'=9) antecedent rainfall variables.

This optimization procedure used equation 4.19 and numerical matrix inversion and matrix multiplication techniques to compute the many sets of $\hat{\beta}_j$

values needed. Equation 4.21 was used for computation of r^2. All computations and data manipulation operations were made with computer programs employing the linear modeling procedures from the SAS software package (SAS Institute, Inc. 1985a, 1985b).

The fundamental assumptions inherent in the linear modeling procedures discussed above are as follows:

(1) the relationship between the dependent variable and each of the independent variables is fixed

(2) the linear form of the model is correct

(3) the expected value of the residuals is zero

(4) the variance of the residuals is a constant across all observations

(5) the residuals are uncorrelated

(6) the residuals are normally distributed.

Given these assumptions, it is possible to construct a statistical test to establish that the specific variables selected by the stepwise least squares optimization procedure do, in fact, significantly influence the value of the dependent variable. This is accomplished by testing the null hypothesis that the β_j associated with a given variable is equal to zero against the alternative hypothesis that it is different from zero. If the null hypothesis is not rejected, then it is concluded that the associated independent variable does not belong in the model for the specified level of significance. On the other hand, if the null hypothesis is not accepted, it is concluded that the associated independent variable does belong in the model.

Under the null hypothesis that $\beta_j = 0$, and given that the basic

assumptions hold, the ratio of the $\hat{\beta}_j$ to the standard deviation of $\hat{\beta}_j$ is distributed as Student's t (Haan 1977):

(4.22) $$t_{j,\upsilon} = \frac{\hat{\beta}_j}{s_j}$$

where,

$t_{j,\upsilon}$ = Student's t statistic for $\hat{\beta}_j$ with ν degrees of freedom

s_j = Sample standard deviation of the $\hat{\beta}_j$ which is called the standard error of the estimate.

Now, ν is given by,

(4.23) $$\upsilon = n - p$$

where,

n = number of observations in the data set

p = number of $\hat{\beta}_j$ parameters estimated for the given model (p = 2, 3, 4, or 5).

The standard deviation of $\hat{\beta}_j$ is given by,

(4.24) $$s_j = +\sqrt{(\mathbf{X}'\ \mathbf{X})^{-1}_{jj}\ s^2_\varepsilon}$$

where,

$(\mathbf{X}'\ \mathbf{X})^{-1}_{jj}$ = j-th diagonal component of the matrix $(\mathbf{X}'\ \mathbf{X})^{-1}$, which is known from the solution for $\hat{\beta}_j$

s_ε^2 = sample variance of the residuals, ε_i, which is called the mean square error (MSE).

In turn, s_ε^2 is given by,

$$s_\varepsilon^2 = \frac{\sum_{i=1}^{n} \varepsilon_i^2}{(n-p)} \tag{4.25}$$

This technique allows a Student's t test to be conducted for each $\hat{\beta}_j$ for each model developed to confirm that the variables included in the model are appropriate for estimation of the dependent variable.

Runoff Volume Model

Variable Selection and Parameter Optimization

The optimization procedure outlined in the previous section was applied to develop a model with runoff volume, Q, as the dependent variable. The model can be written as,

$$Q = \beta_0 + \beta_1 R_w + \beta_2 E + \beta_3 A + \beta_4 I + \varepsilon_v \tag{4.26}$$

where the variable R_w and the sets of independent variables E, A, and I are as defined previously and ε_v is the residual component.

Since a rainfall intensity variable is to be included in the model, only the field data gathered after July 2, 1987 can be used. This was the date of installation of a recording rain gage, from which intensity data is generated, in the study area.

The stepwise least squares optimization analysis for the runoff volume model is summarized in Table 4.1. Analyses were completed for four different subsets of the data. In Part (A.) of Table 4.1, the analysis was conducted for all rainfall-runoff events, that is all events with nonzero runoff volume, for which intensity data were available. There were 31 such events. The variable of greatest influence on Q, that is the first entered into the model, was rainfall volume as expected. Then the elapsed time variable T_r was entered followed by antecedent moisture and intensity variables ${}_0R_7$ and I_a, respectively. However, the fact that the last variable added, I_a, resulted in an extremely slight increase in r^2 (from 0.7890 to 0.7899) implies that rainfall intensity has a negligible influence on runoff volume. This conclusion was confirmed by the statistical tests of significance conducted for the parameters of the model as shown in detail in the next section. Therefore, the three-variable model is the best achievable model based on the data of Part (A.). The r^2 value of 0.7890 indicates that this model accounts for approximately 79 percent of the variation in Q.

In the second part of Table 4.1, only the 19 events with significant runoff (i.e., $Q \geq 0.01$ inch) were included in the analysis. As before, the variables R_w

TABLE 4.1

STEPWISE OPTIMIZATION ANALYSIS FOR RUNOFF VOLUME MODEL

Variable Entered	Model	Coefficient of Determination, r^2	Correlation Coefficient, r
	(A.) DATA FROM JULY 2, 1987 TO JUNE 27, 1988: Q > 0.0 INCH (n = 31)		
R_w	$Q = -.02349 + .09338 (R_w)$	.5394	.7344
T_r	$Q = -.02884 + .1491 (R_w) - .00009820 (T_r)$	.7784	.8823
${}_0R_7$	$Q = -.03547 + .1493 (R_w) - .00009784 (T_r) + .009966 ({}_0R_7)$	.7890	.8883
I_a	$Q = -.03702 + .1476 (R_w) - .00009933 (T_r) + .01031 ({}_0R_7) + .01360 (I_a)$	.7899	.8887
	(B.) DATA FROM JULY 2, 1987 TO JUNE 27, 1988: Q ≥ 0.01 INCH (n = 19)		
R_w	$Q = -.02229 + .1033 (R_w)$	.5693	.7545
T_r	$Q = -.03883 + .1671 (R_w) - .0001134 (T_r)$	.7925	.8902
${}_0R_6$	$Q = -.05017 + .1671 (R_w) - .0001263 (T_r) + .02976 ({}_0R_6)$	.8476	.9207
I_a	$Q = -.05366 + .1645 (R_w) - .0001326 (T_r) + .03241 ({}_0R_6) + .02630 (I_a)$	.8500	.9219
	(C.) DATA FROM JUNE 9, 1987 TO JUNE 27, 1988: Q ≥ 0.0 INCH (n = 40)		
R_w	$Q = -.01908 + .09736 (R_w)$	.4861	.6972
T_r	$Q = -.02928 + .1583 (R_w) - .0001101 (T_r)$	.7641	.8741
${}_0R_3$	$Q = -.03970 + .1593 (R_w) - .0001048 (T_r) + .01790 ({}_0R_3)$	.7909	.8893
	(D.) DATA FROM JUNE 9, 1987 TO JUNE 27, 1988: Q ≥ 0.01 INCH (n = 27)		
R_w	$Q = -.01613 + .1054 (R_w)$	.4990	.7064
T_r	$Q = -.04020 + .1776 (R_w) - .0001286 (T_r)$	.7683	.8765
${}_0R_3$	$Q = -.05325 + .1753 (R_w) - .0001201 (T_r) + .03102 ({}_0R_3)$	.8291	.9105

and T_r were the first variables entered into the model. An antecedent rainfall variable was again entered third, but this time the variable chosen was ${}_0R_6$. The fourth variable entered was the intensity variable I_a. As before, the negligible increase in r^2 from 0.8476 to 0.8500 with the addition of I_a to the model, in conjunction with the statistical tests of the next section, indicate that the three-variable model is optimal based on the data of Part (B.). The final three-variable model accounts for about 85 percent of the variation in Q. Obviously, the use of rainfall-runoff events with Q equalling or exceeding 0.01 inch measurably enhances the model.

The analyses of Part (A.) and Part (B.) of Table 4.1 conclusively demonstrate that a three-variable model (i.e., $\beta_4 = 0$) is appropriate for estimation of runoff volume. This allows inclusion into the data set of the nine rainfall-runoff events for which intensity data are not available. Eight of these events were monitored prior to the installation of the recording gage on July 2, 1987 and the ninth occurred on December 13, 1987 when the recording gage malfunctioned. Part (C.) of Table 4.1 summarizes the optimization analysis for a three-variable model using all 40 rainfall-runoff events monitored during this study regardless of the magnitude of the runoff volume. Again, the two most important variables were R_w and T_r. The antecedent rainfall variable ${}_0R_3$ was the third variable added. This three-variable model accounted for 79 percent of the variation in Q.

Finally, in Part (D.) of Table 4.1, only the 27 events with significant runoff volume were included in the analysis. As before, the exclusive use of events with significant runoff improved the model. The same three variables were

included as for the previous model. The model accounted for about 83 percent of the variation in Q.

The models fit to the data subsets of Parts (B.) and (D.), which both include only significant runoff events, have very similar parameter values throughout. The only difference between the models is that the antecedent rainfall variables are different: ${}_0R_6$ for the former and ${}_0R_3$ for the latter. This difference is fairly unimportant since the antecedent rainfall variable has a relatively modest influence on the r^2 value compared to R_w and T_r. A plot of the model predicted values of runoff volume, $\hat{Q}$, versus the observed values, Q, is shown for both models in Figure 4.1. A 45 degree line with zero intercept is shown for reference. It is apparent that both models describe the data well, although with a relatively high degree of random error. That is, the ε_v values are of a significant magnitude.

The greater values of r^2 achieved in each step of the optimization procedure for the first model, the model of Part (B.), was probably due to the greater accuracy of the rainfall data achieved with use of the recording rain gage. The area weighted rainfall volume, R_w, was more accurate because two rain gages were used to determine R_w rather than only one. The elapsed time variable, T_r, was more accurate because the second rain gage was a recording instrument (the other is only a daily gage) which allowed accurate measurement of timing between rainfall events. Without the use of the recording gage, elapsed time between rainfall events was approximated from an analysis of the associated runoff hydrographs. Finally, the antecedent rainfall variable was more accurate because of the improved ability to account for the timing of consecutive rainfall events using recording rain gage data.

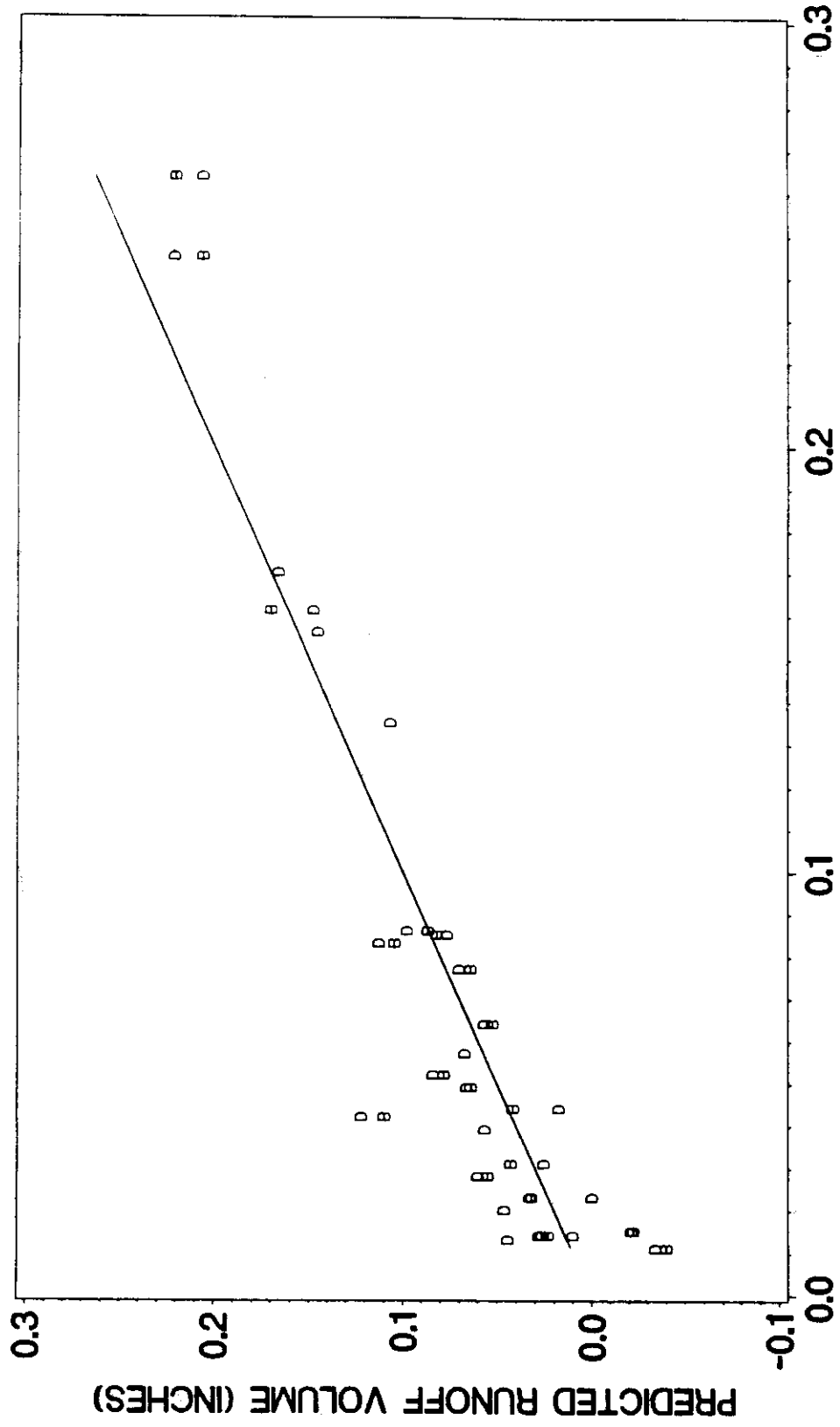


FIGURE 4.1
PREDICTED VS. OBSERVED RUNOFF VOLUMES

Based on the greater accuracy of the data, and the larger r^2 value, the three-variable model of Part (B.) of Table 4.1 has been chosen as the final basin transformation model for runoff volume:

$$(4.27) \qquad Q = -.05017 + .1671\ (R_w) - .0001263\ (T_r) + .02976\ ({}_0R_6) + \varepsilon_v$$

with $r^2 = 0.8476$ and, thus, $r = 0.9207$.

There are several additional important points to emphasize about the analyses summarized in Table 4.1. First, the elapsed time variable of most significance, without exception, was time between runoff events with $Q \geq 0.01$ inch, T_r. This was true regardless of whether all runoff data ($Q > 0.0$ inch) were included or only significant events ($Q \geq 0.01$ inch) were included. This reinforces the conclusion that the entire basin becomes saturated to the point of contributing runoff only for events with $Q \geq 0.01$ inch, and it emphasizes that T_r is the best indicator of basinwide areal average soil drainage and evapotranspiration moisture losses among the variables considered. Second, the model parameters were estimated using data collected from all seasons of the year. Therefore, the success of the final model in accounting for 85 percent of the variation in the runoff volume indicates that any seasonal factors that may influence runoff have been adequately quantified by the model. This depends, of course, on the fact that the remaining 15 percent of the variability is due to normally distributed random error, which is successfully demonstrated in the "Analysis of Residuals" section below. Third, the parameters of the derived models have algebraic signs and magnitudes that are consistent with the

physics of the rainfall-runoff process:

(1) Q increases with increasing R_w which implies $\beta_1>0$. Furthermore, Q cannot exceed R_w such that $0 < \beta_1 \leq 1$. These conditions were satisfied by the values of $\hat{\beta}_1$ throughout the analysis.

(2) Q decreases with increasing T_r which implies $\beta_2 < 0$. This condition on the algebraic sign of β_2 was satisfied by all values of $\hat{\beta}_2$.

(3) Q increases with increasing antecedent rainfall volume which implies $\beta_3 > 0$. The parameter estimator $\hat{\beta}_3$ satisfied this condition throughout the analysis.

(4) Modifying equation 4.5 to omit the negligible rainfall intensity term, $(I - f_c)$, and then rearranging, yields

$$Q = (\alpha-1)A_i + (1-\alpha)R - \beta E + \gamma A + \varepsilon_v .$$

Clearly, then, $\beta_0 = (\alpha-1)A_i$ and $\beta_1 = (1-\alpha)$. Solving these simultaneously for A_i,

$$A_i = -\left[\frac{\beta_0}{\beta_1}\right].$$

This indicates that $\beta_0 < 0$ since $A_i > 0$ and $\beta_1 > 0$. This condition on the sign of β_0 was satisfied by the values of $\hat{\beta}_0$ throughout the analysis. Also, for the final model,

$$A_i = -\left[\frac{-0.05017 \text{ inch}}{0.1674 \text{ inch/inch}}\right] = 0.3002 \text{ inch},$$

which is a physically reasonable value.

Obtaining physically realistic values for the model parameters inspires confidence in the methodology and the predictive ability of the final model.

Objective statistical tests to confirm the applicability of the model parameters are presented in the following section.

Statistical Tests of Significance for Model Parameters

The Student's t test described in detail in the methodology section was employed to verify the applicability of the parameters, and thus, the independent variables included in the runoff volume model. Computations and results of the test of significance for each parameter of the four-variable models developed in the previous section are shown in detail in Table 4.2. Analyses for the model developed using all runoff volume data ($Q > 0.0$ inch, $n = 31$) are shown in Part (A.) of Table 4.2. Analyses for the model developed using only significant runoff events ($Q \geq 0.01$ inch, $n = 19$) are shown in Part (B.). The critical values of the test statistic were chosen based on a 0.05 level of significance using a two-sided test.

Test statistics for β_4 lie well outside the critical region of the test for both models such that the null hypothesis that $\beta_4 = 0$ is not rejected in either case. In fact, the null hypothesis would not be rejected even at the 0.741 level of significance for Part (A.) nor at the 0.647 level of significance for Part (B.). Therefore, it can be concluded that β_4 is not significantly different from zero and that rainfall intensity is not pertinent to estimation of Q.

Test statistics for β_3 are borderline. In Part (A.), the test statistic for β_3 lies just outside the critical region of the test and for Part (B.) it lies just inside.

TABLE 4.2

TESTS OF SIGNIFICANCE FOR PARAMETER ESTIMATES OF RUNOFF VOLUME MODELS WITH FOUR INDEPENDENT VARIABLES

Parameter	Parameter Estimate, $\hat{\beta}_j$	Mean Squared Error, s_ε^2	$(\mathbf{x}'\mathbf{x})_{ij}^{-1}$	Standard Error of the Estimate, s_j	Student's t Statistics: Test Statistic $t_{j,\upsilon}$	Student's t Statistics: Critical Value, $t_{c,\upsilon}$	Decision
	(A.) DATA FROM JULY 2, 1987 TO JUNE 27, 1988: Q > 0.0 INCH (n=31; υ=n-5=26)						
β_0	-.03702	.001071	.1527	.01279	-2.89	2.06	Do Not Accept H_0
β_1	.1476	.001071	.2420	.01610	9.17	2.06	Do Not Accept H_0
β_2	-.00009933	.001071	3.234×10^{-7}	.00001861	-5.34	2.06	Do Not Accept H_0
β_3	.01031	.001071	.07168	.008762	1.18	2.06	Do Not Reject H_0
β_4	.01360	.001071	1.5482	.04072	0.33	2.06	Do Not Reject H_0
	(B.) DATA FROM JULY 2, 1987 TO JUNE 27, 1988: Q ≥ 0.01 INCH (n=19; υ=n-5=14)						
β_0	-.05366	.001046	.3060	.01789	-3.00	2.14	Do Not Accept H_0
β_1	.1645	.001046	.4081	.02066	7.96	2.14	Do Not Accept H_0
β_2	-.0001326	.001046	7.984×10^{-7}	.00002890	-4.59	2.14	Do Not Accept H_0
β_3	.03241	.001046	.1952	.01429	2.27	2.14	Do Not Accept H_0
β_4	.02630	.001046	3.0260	.05626	0.47	2.14	Do Not Reject H_0

Null Hypotheses, H_0: The specified parameter is equal to zero; $\beta_j = 0$.

Alternative Hypothesis, H_1: The specified parameter is not equal to zero; $\beta_j \neq 0$.

Degrees of Freedom: υ = n-5 (the 5 parameters β_0 through β_4 were estimated from the data).

Decision Rule: Do not accept H_0 if the absolute value of the Student's t test statistic, $| t_{j,\upsilon} |$, exceeds the critical value, $t_{c,\upsilon}$, for a two-sided test at the 0.05 level of significance with υ degrees of freedom. Otherwise, do not reject H_0.

Further evaluation of the significance of β_3 is required via a similar analysis for three-variable models.

Computations and results of the test of significance for each parameter for three-variable runoff models are shown in detail in Table 4.3. Only the two models involving significant runoff events ($Q \geq 0.01$ inch) are shown for the sake of clarity and conciseness. The absolute values of the Student's t test statistic for all parameters of both models exceed the critical value. Therefore, the conclusion in statistical terminology is, do not accept the null hypotheses that the β_j are zero at the 0.05 level of significance. Based on these results, it can be concluded that all of the parameters and variables in the models are pertinent to estimation of Q.

Model parameters evaluated in Part (A.) of Table 4.3 are from the final model selected in the previous section. It is relevant to emphasize several important points about the analyses for this model. First, test statistics for the parameters β_1 and β_2 lie far within the critical region of the test. This emphasizes the major influence that R_w and T_r have on the observed values of Q. Second, the test statistic for the parameter β_0 is a relatively moderate distance inside the critical region, but the null hypothesis that $\beta_0 = 0$ would still not be accepted even at the 0.01 level of significance. Third, the test statistic for β_3, 2.33, is borderline. The null hypothesis that $\beta_3 = 0$ would not be rejected at a level of significance of 0.0342. This emphasizes the relatively small impact that ${}_0R_6$ has on the observed values of Q.

Analysis of Residuals

A critical step in verifying the applicability of the derived runoff volume model is to demonstrate that the model residuals are normally distributed with

TABLE 4.3

TESTS OF SIGNIFICANCE FOR PARAMETER ESTIMATES OF RUNOFF VOLUME MODELS WITH THREE INDEPENDENT VARIABLES

Parameter	Parameter Estimate, $\hat{\beta}_j$	Mean Squared Error, s_ε^2	$(\mathbf{x}'\mathbf{x})^{-1}_{jj}$	Standard Error of the Estimate, s_j	Student's t Statistics: Test Statistic $t_{j,\upsilon}$	Student's t Statistics: Critical Value, $t_{c,\upsilon}$	Decision
	(A.) DATA FROM JULY 2, 1987 TO JUNE 27, 1988: Q > 0.0 INCH (n=19; υ=n-4=15)						
β_0	-.05017	.0009913	.2528	.01583	-3.17	2.13	Do Not Accept H_0
β_1	.1671	.0009913	.3789	.01938	8.62	2.13	Do Not Accept H_0
β_2	-.0001263	.0009913	6.204×10^{-7}	.00002480	-5.09	2.13	Do Not Accept H_0
β_3	.02976	.0009913	.1645	.01277	2.33	2.13	Do Not Accept H_0
β_4	0	-----	-----	-----	-----	-----	-----
	(B.) DATA FROM JULY 9, 1987 TO JUNE 27, 1988: Q ≥ 0.01 INCH (n=27; υ=n-4=14)						
β_0	-.05325	.0009472	.2040	.01390	-3.83	2.07	Do Not Accept H_0
β_1	.1753	.0009472	.3274	.01761	9.95	2.07	Do Not Accept H_0
β_2	-.0001201	.0009472	4.912×10^{-7}	.00002157	-5.57	2.07	Do Not Accept H_0
β_3	.03102	.0009472	.1241	.01084	2.86	2.07	Do Not Accept H_0
β_4	0	-----	-----	-----	-----	-----	-----

Null Hypotheses, H_0: The specified parameter is equal to zero; $\beta_j = 0$.

Alternative Hypothesis, H_1: The specified parameter is not equal to zero; $\beta_j \neq 0$.

Degrees of Freedom: υ = n-4 (the 4 parameters β_0 through β_3 were estimated from the data; $\beta_4 = 0$ was known a priori).

Decision Rule: Do not accept H_0 if the absolute value of the Student's t test statistic, $|t_{j,\upsilon}|$, exceeds the critical value, $t_{c,\upsilon}$, for a two-sided test at the 0.05 level of significance with υ degrees of freedom. Otherwise, do not reject H_0.

zero mean and standard deviation σ_ε. If the residuals can be shown to possess this distribution, it can be concluded that all deterministic sources of variation in the dependent variable Q have been accounted for and only random error, or noise, remains in the residual component of the model.

The population mean of the residuals is known a priori to be zero (i.e., $\mu_\varepsilon = 0$) because the least squares optimization procedure requires this condition regardless of the pdf of the residuals. The population standard deviation, σ_ε, must be estimated from the sample data using equation 4.25: $s_\varepsilon = 0.03148$ inch. Plots of the observed empirical cfd and the theoretical normal cdf with $\mu_\varepsilon = 0$ and $\sigma_\varepsilon = 0.03148$ are shown in Figure 4.2. The empirical cdf is the step function and the theoretical cdf is shown as the smooth curve. The agreement between them is generally acceptable.

The next step was to conduct an objective test for goodness of fit. As a result of the small sample size, n = 19, the chi-square test for goodness of fit used extensively in Chapter 3 was not practical. Consequently, the Kolmogorov-Smirnov one-sample test was employed. This is a computationally simple nonparametric statistical test for goodness of fit that involves the maximum deviation of the empirical cdf, $E(\varepsilon)$, from the hypothesized theoretical cdf, $F(\varepsilon)$, as the test statistic:

$$\Delta_{max} = \max_{\varepsilon_i} \left\{ \max \left[\left| E(\varepsilon_i) - F(\varepsilon_i) \right| , \left| E(\varepsilon_{i-1}) - F(\varepsilon_i) \right| \right] \right\} \tag{4.28}$$

where,

Δ_{max} = absolute value of the maximum difference between the empirical and theoretical cdf's

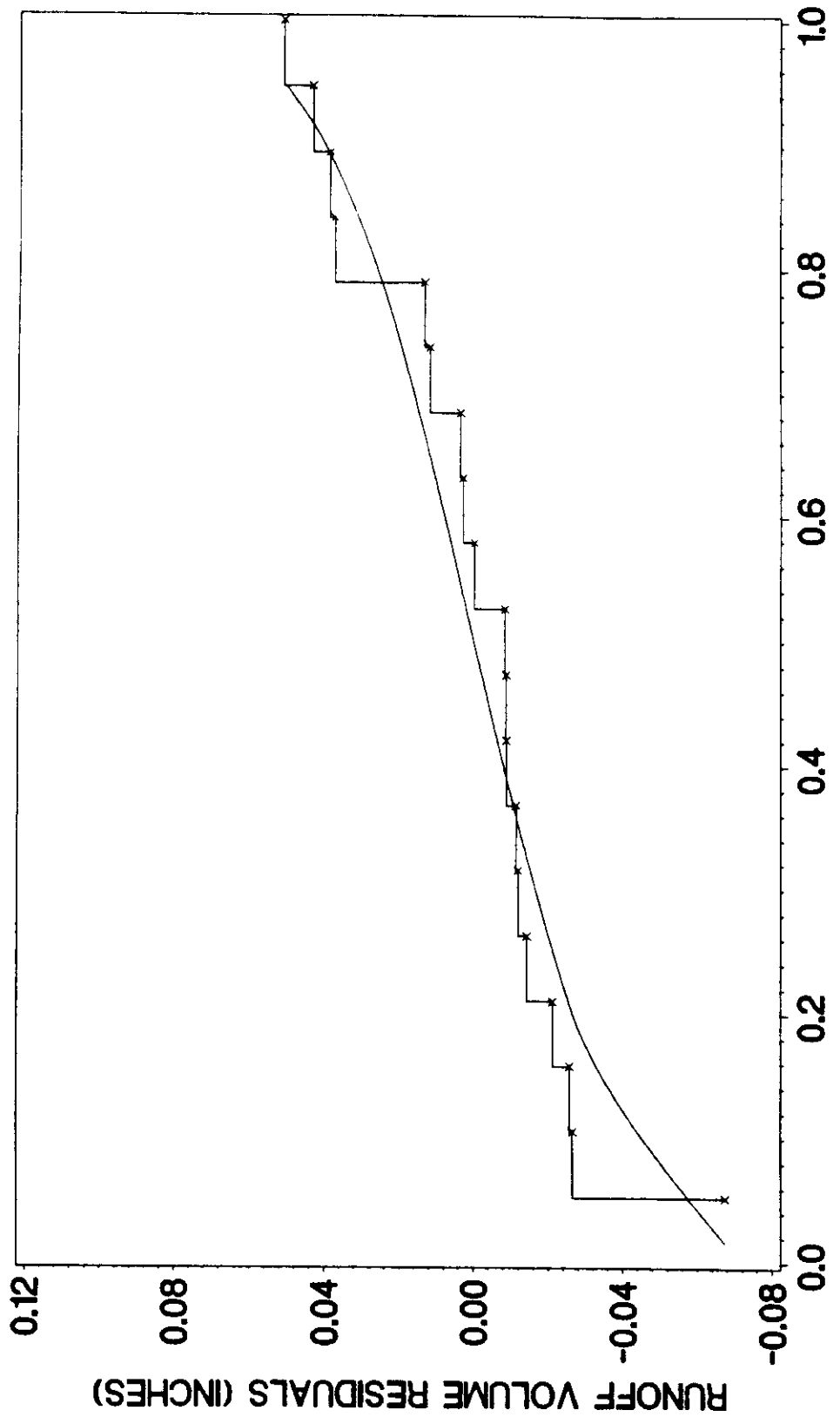


FIGURE 4.2
CDF FOR RUNOFF VOLUME RESIDUALS

$\max_{\varepsilon_i}\{\ \}$ = a function that returns the maximum value of the argument over all ε_i values; i = 1, 2, . . ., n.

The second term in the function argument, | E (ε_{i-1}) - F (ε_i) |, is necessary because the maximum deviation of E (ε) from F (ε) does not necessarily occur at an observed value of ε_i, but could occur at some other point along the empirical step function. Close scrutiny of Figure 4.2 illustrates that the maximum difference between the empirical step function and the smooth theoretical curve must occur at one of the corner points of the empirical cdf. For any given value of F (ε_i), there are two corresponding corner point values: E (ε_i) and E (ε_{i-1}). Therefore, the maximum deviation must be compared for all E (ε_i) and E (ε_{i-1}) in order to establish the true Δ_{max}.

The Kolmogorov-Smirnov one-sample test for the normal cdf fit to the residuals of the runoff volume model is shown in detail in Table 4.4. The maximum deviation was 0.147. The critical value of the test statistic, $\Delta_0 = 0.301$, was obtained from standard statistical tables for a two-sided test with n = 19 (Daniel 1978). The test statistic lies well outside the critical region of the test at the 0.05 level of significance. Therefore, do no reject the null hypothesis that the residuals of the runoff volume model represent a random sample of size 19 from a normal pdf with $\mu_\varepsilon = 0.0$ and $\sigma_\varepsilon = 0.03148$.

The result of this analysis verifies the applicability of the runoff volume model.

TABLE 4.4

KOLMOGOROV-SMIRNOV TEST FOR RESIDUALS CDF FOR RUNOFF VOLUME MODEL

	Observed Empirical CDF		Normal CDF ($\mu_\varepsilon = 0.0$, $\sigma_\varepsilon = 0.03148$)			
Residuals, ε_i (inches)	Cumulative Frequency	Cumulative Probability, $E(\varepsilon_i)$	$Z_i = \frac{\varepsilon_i}{s_\varepsilon}$	Cumulative Probability, $F(\varepsilon_i)$	Deviations, Δ: $\lvert E(\varepsilon_i) - F(\varepsilon_i) \rvert$	Deviations, Δ: $\lvert E(\varepsilon_{i-1}) - F(\varepsilon_i) \rvert$
-0.06735	1	0.053	-2.14	0.016	0.037	—
-0.02647	2	0.105	-0.84	0.200	0.095	0.147
-0.02564	3	0.158	-0.81	0.209	0.051	0.104
-0.02107	4	0.211	-0.67	0.251	0.040	0.003
-0.01422	5	0.263	-0.45	0.326	0.063	0.115
-0.01194	6	0.316	-0.38	0.352	0.036	0.089
-0.01134	7	0.368	-0.36	0.359	0.009	0.043
-0.00872	8	0.421	-0.28	0.390	0.031	0.022
-0.00864	9	0.474	-0.27	0.394	0.080	0.027
-0.00837	10	0.526	-0.26	0.397	0.129	0.077
-0.00014	11	0.579	-0.00	0.500	0.079	0.026
0.00290	12	0.632	0.09	0.536	0.096	0.043
0.00373	13	0.684	0.12	0.548	0.136	0.084
0.01199	14	0.737	0.38	0.648	0.089	0.036
0.01353	15	0.789	0.43	0.666	0.123	0.071
0.03769	16	0.842	1.20	0.885	0.043	0.096
0.03905	17	0.895	1.24	0.892	0.003	0.050
0.04361	18	0.947	1.39	0.918	0.029	0.023
0.05139	19	1.000	1.63	0.948	0.052	0.001

Null Hypothesis, H_0: The residuals for the runoff volume model are normally distributed with parameters $\mu_\varepsilon = 0.0$ and $\sigma_\varepsilon = 0.03148$.
Alternative Hypothesis, H_1:The residuals are not normally distributed with the specified parameters.
Kolmogorov-Smirnov Test Statistic: $\Delta_{max} = 0.147$.
Critical Region at 0.05 Level of Significance for n=19 (Two-Sided Test): $\Delta_0 > 0.301$.
Decision: Do Not Reject H_0.

Peak Flow Rate Model

Variable Selection and Parameter Optimization

The stepwise least squares optimization procedure was applied to develop a model for peak flow rate, q_p, as the dependent variable. The model can be written as:

$$q_p = \beta_0 + \beta_1 R_w + \beta_2 E + \beta_3 A + \beta_4 I + \varepsilon \tag{4.29}$$

where the variable R_w and the sets of independent variables E, A, and I are as defined in the methodology section.

Peak flow rate was expected to be highly dependent on rainfall intensity. Therefore, as for the runoff volume model, only the field data collected after installation of the recording rain gage could be used.

The stepwise least squares optimization procedure for the peak flow rate model is summarized in Table 4.5. Analyses were completed for two subsets of the observed data. In part (A.) of Table 4.5, the analysis was conducted for all rainfall-runoff events, that is, all events with nonzero runoff volume for which intensity data were available. There were 31 events meeting this requirement. The variable of greatest influence on q_p was rainfall volume, R_w. Then, the elapsed time variable, T_l, was entered followed by the intensity variable, I_a. The fourth variable entered was the antecedent rainfall variable ${}_0R_{31}$.

The fact that the addition of the last variable ${}_0R_{31}$ resulted in only a slight increase in r^2 (from 0.8337 to 0.8425) implies that antecedent rainfall has a relatively insignificant influence on peak flow rate. This conclusion was confirmed by the statistical tests of significance conducted for the parameters of the model as shown in detail in the following section. Therefore, the three-

TABLE 4.5

STEPWISE OPTIMIZATION ANALYSIS FOR PEAK FLOW RATE MODEL

Variable Entered	Model	Coefficient of Determination, r^2	Correlation Coefficient, r
	(A) DATA FROM JULY 2, 1987 TO JUNE 27, 1988: Q > 0.0 INCH (n = 31)		
R_w	$q_p = -4.3272 + 15.5179\ (R_w)$	.6081	.7798
T_I	$q_p = -3.2181 + 17.8173(R_w) - .01194\ (T_I)$	.7242	.8510
I_a	$q_p = -5.1775 + 13.8537(R_w) - .01290(T_I) + 22.6846(I_a)$	.8337	.9131
${}_0R_{31}$	$q_p = -7.7168 + 14.1909(R_w) - .01175(T_I) + 22.9901(I_a) + .7141({}_0R_{31})$	.8425	.9179
	(B.) DATA FROM JULY 2, 1987 TO JUNE 27, 1988: Q ≥ 0.01 INCH (n =19)		
R_w	$q_p = -4.7804 + 17.4903\ (R_w)$	.6567	.8104
I_a	$q_p = -6.4318 + 12.1630(R_w) + 25.2760(I_a)$	.7775	.8818
T_I	$q_p = -4.6053 + 12.3698(R_w) + 32.4367(I_a) - .02144(T_I)$	.8444	.9189
${}_0R_{31}$	$q_p = -6.9086 + 12.5783(R_w) + 32.6351(I_a) - .02084(T_I) + .7344\ ({}_0R_{31})$	.8533	.9237

variable model is the best achievable model based on the data of Part (A.). The r^2 value of 0.8337 indicates that this model accounts for approximately 83 percent of the variation in q_p.

In the second part of Table 4.5, only the 19 events with significant runoff volume were included in the analysis. The best three-variable model involved R_w, I_a, and T_I as before. Although, in this case, the intensity variable was selected second in preference to the elapsed time variable T_I. The fourth variable entered was, again, the antecedent rainfall variable, ${}_0R_{31}$. As before, the slight increase in r^2 from 0.8444 to 0.8533, in conjunction with the statistical tests of the next section, indicate that the three-variable model is optimal based on the data of Part (B.). The final three-variable model accounts for about 84 percent of the variation in q_p. Use of rainfall-runoff events with $Q \geq 0.01$ inch improves the predictive performance of the model, especially considering that only 19 of the 31 runoff events were used.

It has been conclusively demonstrated that a three-variable model (i.e., $\beta_3 = 0$) is appropriate for estimation of q_p. Based on the larger r^2 value, and for the sake of consistency with the final runoff volume model with respect to the data subset used, the model of Part (B.) of Table 4.5 has been selected as the final basin transformation model for peak flow rate:

$$(4.30) \qquad q_p = -4.6053 + 12.3698\,(R_w) - .02144\,(T_I) + 32.4367\,(I_a)$$

with $r^2 = 0.8444$ and, thus, $r = 0.9189$. A plot of model predicted values of peak flow rate, $\hat{q}_p$, versus the observed values, q_p, is shown in Figure 4.3. A 45

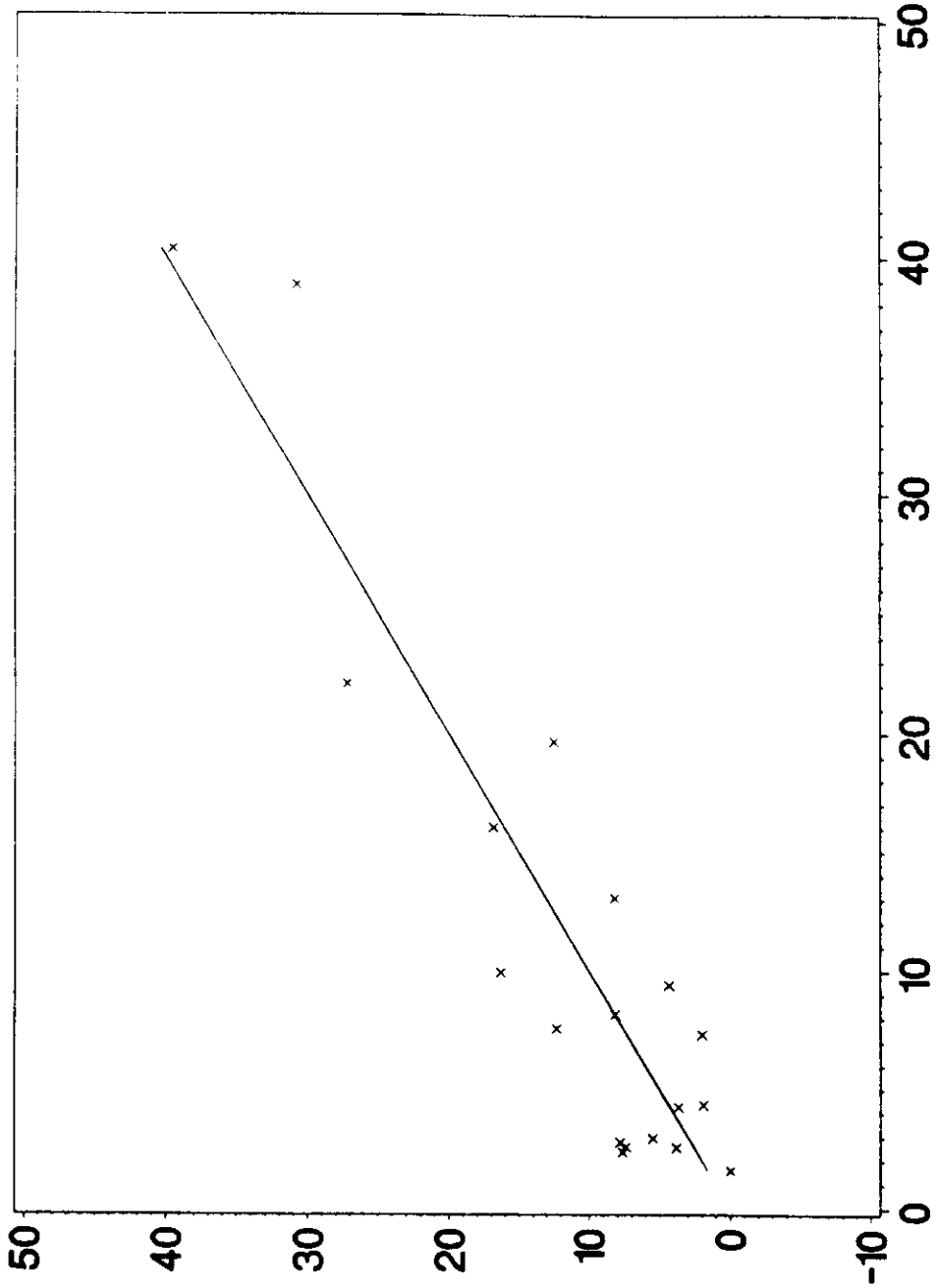


FIGURE 4.3
PREDICTED VS. OBSERVED PEAK RUNOFF RATE

degree line with zero intercept is shown for reference. It is apparent that the model describes the data well, although with a relatively high degree of random error. That is, the ε_q values are of a relatively large magnitude.

There are several additional important points to emphasize about the analyses summarized in Table 4.5. First, the fact that I_a, which is just the ratio of R_w to D, is the preferred intensity variable means that event duration is the only rainfall variable related to intensity (other than, of course, rainfall volume) that must be simulated. This is significant, and fortunate, since D is the only intensity variable that can be reasonably assumed to be independent of R_w, as discussed in detail in Chapter 3. Second, the parameters of the final model have algebraic signs that are consistent with the physics of the rainfall-runoff process:

(1) q_p increases with increasing R_w, which implies $\beta_1 > 0$. This condition on the algebraic sign of β_1 was satisfied by the values of $\hat{\beta}_1$ throughout the analysis.

(2) q_p decreases with increasing elapsed time between runoff events, which implies $\beta_2 < 0$. This condition on the sign of β_2 was satisfied by all values of $\hat{\beta}_2$.

(3) q_p increases with increasing rainfall intensity, which implies $\beta_4 > 0$. The parameter estimator $\hat{\beta}_4$ satisfied this condition on algebraic sign throughout the analysis.

Third, in order to independently estimate the parameter f_c and to evaluate its magnitude for physical realism, it is possible to write an equation with two independent variables where q_p is a function of runoff volume, Q, and an intensity variable as follows:

(4.31) $$q_p = \lambda_1 Q + \lambda_2 (I - f_c) .$$

A two-variable model of this form is expected because the runoff volume model of equation 4.27 expresses Q as a function of the three independent variables R_w, T_r, and ${}_0R_6$ but does not include an intensity variable. Therefore, equation 4.31, by incorporating an intensity term, expresses q_p as a function of the four general classes of independent variables in a manner analogous to equation 4.29. Rearranging equation 4.31,

(4.32) $$q_p = -\lambda_2 f_c + \lambda_1 Q + \lambda_2 I .$$

Applying the least squares optimization procedure using the observed data subset of Part (B.) of Table 4.5 yields,

(4.33) $$q_p = -4.2851 + 83.8863\ (Q) + 36.6401\ (I_a) .$$

Notice that the intensity variable I_a was again chosen and that the intercept parameter and coefficient of I_a are very near those of the final model of equation 4.30. This two-variable model has a coefficient of determination of $r^2 = 0.8313$ and, thus, $r = 0.9118$. These values are only slightly lower than the values of the final model. Finally, notice that,

$$f_c = -\left(\frac{-\lambda_2 f_c}{\lambda_2}\right) = -\left(\frac{-4.2851 \text{ cfs}}{36.6401 \text{ cfs/inch/hour}}\right)$$

$$f_c = 0.1170 \text{ inch/hour}$$

which is a physically realistic value for the clayey soils found in the study watershed. Obtaining physically realistic algebraic signs and values for the

model parameters promotes confidence in the methodology and the predictive ability of the final model.

Objective statistical tests to confirm the applicability of the model parameters are presented in the following section.

Statistical Tests of Significance for Model Parameters

The Student's t test was employed to verify the applicability of the parameters, and thus the independent variables, included in the peak flow rate model. Computations and results of the test of significance for each parameter of the four-variable models derived in the previous section are shown in detail in Table 4.6. Analyses for the model which was developed using all runoff volume data (i.e., $Q > 0.0$ inch, $n = 31$) are shown in Part (A.) of Table 4.6. Analyses for the model developed using only significant runoff events (i.e., $Q \geq 0.01$ inch, $n = 19$), are shown in Part (B.). The critical values of the test statistic were chosen based on a 0.05 level of significance using a two-sided test.

Test statistics for β_3 lie well outside the critical region of the test for both models such that the null hypothesis that $\beta_3 = 0$ is not rejected in either case. In fact, the null hypothesis would not be rejected even at the 0.241 level of significance for Part (A.) nor at the 0.370 level of significance for Part (B.). Therefore, it can be concluded that β_3 is not significantly different zero and that antecedent rainfall volume is not pertinent to estimation of q_p.

Test statistics for β_0 are borderline. In Part (A.), the test statistic for β_0 lies a moderate distance inside the critical region of the test and for Part (B.) it lies just outside. Further evaluation of the significance of β_0 is required via a similar analysis for the three-variable models.

TABLE 4.6

TESTS OF SIGNIFICANCE FOR PARAMETER ESTIMATES OF PEAK FLOW RATE MODELS WITH FOUR INDEPENDENT VARIABLES

Parameter	Parameter Estimate, $\hat{\beta}_j$	Mean Squared Error, s_ε^2	$(\mathbf{x'x})^{-1}_{jj}$	Standard Error of the Estimate, s_j	Student's t Statistics: Test Statistic $t_{j,\upsilon}$	Student's t Statistics: Critical Value, $t_{c,\upsilon}$	Decision
	(A.) DATA FROM JULY 2, 1987 TO JUNE 27, 1988: Q > 0.0 INCH (n=31; υ= n-5=26)						
β_0	-7.7168	19.6615	.3427	2.5958	-2.97	2.06	Do Not Accept H_0
β_1	14.1909	19.6615	.1843	1.9036	7.45	2.06	Do Not Accept H_0
β_2	-.01175	19.6615	4.2744×10^{-7}	.002899	-4.05	2.06	Do Not Accept H_0
β_3	.7141	19.6615	.01803	.5954	1.20	2.06	Do Not Reject H_0
β_4	22.9901	19.6615	1.4515	5.3422	4.30	2.06	Do Not AcceptH_0
	(B.) DATA FROM JULY 2, 1987 TO JUNE 27, 1988: Q ≥ 0.01 INCH (n=19; υ= n-5 = 14)						
β_0	-6.9086	25.4181	.4892	3.5263	-1.96	2.14	Do Not RejectH_0
β_1	12.5783	25.4181	.2910	2.7197	4.62	2.14	Do Not Accept H_0
β_2	-.02084	25.4181	2.8505×10^{-6}	.008512	-2.45	2.14	Do Not Accept H_0
β_3	.7344	25.4181	.02472	.7927	0.93	2.14	Do Not Reject H_0
β_4	32.6351	25.4181	2.4973	7.9672	4.10	2.14	Do Not Accept H_0

Null Hypotheses, H_0: The specified parameter is equal to zero; $\beta_j = 0$.

Alternative Hypothesis, H_1: The specified parameter is not equal to zero; $\beta_j \neq 0$.

Degrees of Freedom: υ = n-5 (the 5 parameters β_0 through β_4 were estimated from the data).

Decision Rule: Do not accept H_0 if the absolute value of the Student's t test statistic, $| t_{j,\upsilon} |$, exceeds the critical value, $t_{c,\upsilon}$, for a two-sided test at the 0.05 level of significance with υ degrees of freedom. Otherwise, do not reject H_0.

The values of the test statistics for all remaining parameters of both models exceed the critical value. Therefore, do not accept the null hypotheses that the remaining β_j are zero at the 0.05 level of significance. Based on these results, it can be concluded that the parameters β_1, β_2, and β_4 and their associated variables are pertinent to estimation at q_p. However, this must be confirmed by re-evaluation of the significance of the β_j using a similar analysis for three-variable models where it is known a priori that $\beta_3 = 0$.

Computations and results of the test of significance for each parameter of the three-variable peak flow rate models are shown in detail in Table 4.7. For both models, the absolute values of the Student's t test statistic for β_1, β_2, and β_4 exceed the critical value. Therefore, do not accept the null hypotheses that these parameters are zero. Based on these results it can be concluded that β_1, β_2, and β_4 and their associated variables are pertinent to estimation of q_p. In fact, the test statistics for β_1 and β_4 lie well within the critical region of test, which emphasizes the major influence that R_w and I_a have on the observed values of q_p.

Test statistics for β_0 are again contradictory. In Part (A.) of Table 4.7, the test statistic for β_0, - 3.42, lies well inside the critical region of the test. The null hypothesis that $\beta_0 = 0$ would not be accepted at a level of significance of 0.0020. In Part (B.), the test statistic for β_0, - 1.85, lies just outside the critical region of the test. The null hypothesis that $\beta_0 = 0$ would not be accepted at a level of significance of 0.0841. Since the analysis of Part (A.) strongly suggests a high degree of importance for the intercept parameter and the analysis of Part (B.) yields a very borderline value of the Student's t test statistic, it has been decided to retain the intercept parameter β_0 = -4.6053 cfs in the final model.

TABLE 4.7

TESTS OF SIGNIFICANCE FOR PARAMETER ESTIMATES OF PEAK FLOW RATE MODELS WITH THREE INDEPENDENT VARIABLES

Parameter	Parameter Estimate, $\hat{\beta}_j$	Mean Squared Error, s_ε^2	$(\mathbf{x'x})^{-1}_{jj}$	Standard Error of the Estimate, s_j	Student's t Statistics: Test Statistic $t_{j,\upsilon}$	Student's t Statistics: Critical Value, $t_{c,\upsilon}$	Decision
	(A.) DATA FROM JULY 2, 1987 TO JUNE 27, 1988: Q > 0.0 INCH (n=31; υ= n-4 =27)						
β_0	-5.1775	19.9808	.1147	1.5139	-3.42	2.05	Do Not Accept H_0
β_1	13.8537	19.9808	.1803	1.8980	7.30	2.05	Do Not Accept H_0
β_2	-.01290	19.9808	3.8097×10^{-7}	.002759	-4.68	2.05	Do Not Accept H_0
β_3	0	------	------	------	------	----	----
β_4	22.6846	19.9808	1.4482	5.3792	4.22	2.05	Do Not AcceptH_0
	(B.) DATA FROM JULY 2, 1987 TO JUNE 27, 1988: Q ≥ 0.01 INCH (n=19; υ= n-4 = 15)						
β_0	-4.6053	25.1782	.2461	2.4892	-1.85	2.13	Do Not RejectH_0
β_1	12.3698	25.1782	.2891	2.6979	4.59	2.13	Do Not Accept H_0
β_2	-.02144	25.1782	2.8339×10^{-6}	.008447	-2.54	2.13	Do Not Accept H_0
β_3	0	------	------	------	------	----	----
β_4	32.4367	25.1782	2.4955	7.9267	4.09	2.13	Do Not Accept H_0

Null Hypotheses, H_0: The specified parameter is equal to zero; $\beta_j = 0$.

Alternative Hypothesis, H_1: The specified parameter is not equal to zero; $\beta_j \neq 0$.

Degrees of Freedom: υ=n-4 (the parameters β_0, β_1, β_2, and β_4 were estimated from the data; $\beta_3 = 0$ was known a priori)).

Decision Rule: Do not accept H_0 if the absolute value of the Student's t test statistic, $|t_{j,\upsilon}|$, exceeds the critical value, $t_{c,\upsilon}$, for a two-sided test at the 0.05 level of significance with υ degrees of freedom. Otherwise, do not reject H_0.

The decision on the final form of the peak flow rate model has been verified in the next section where the resulting residuals are successfully shown to be normally distributed.

Analysis of Residuals

The population mean of the residuals for the peak flow rate model is known a priori to be zero ($\mu_\varepsilon = 0$). The population standard deviation, σ_ε, must be estimated from the sample data using equation 4.25 : $s_\varepsilon = 4.5806$ cfs. A plot of the observed empirical cdf and the theoretical normal cdf with $\mu_\varepsilon = 0$ and $\sigma_\varepsilon = 4.5806$ cfs is shown in Figure 4.4. The empirical cdf is the step function and the theoretical cdf is shown as the smooth curve. The agreement between them is generally acceptable.

An objective evaluation of the goodness of fit was obtained by application of the Kolmogorov-Smirnov one-sample test. This analysis for the normal cdf fit to the residuals of the peak flow rate model is shown in detail in Table 4.8. The maximum deviation was 0.167. The critical value of the test statistic was $\Delta_0 = 0.301$ for a two-sided test with n = 19 at a 0.05 level of significance. The test statistic lies well outside the critical region of the test. Therefore, do not reject the null hypothesis that the residuals of the peak flow rate model represent a random sample of size 19 from a normal pdf with $\mu_\varepsilon = 0.0$ and $\sigma_\varepsilon = 4.5806$.

The result of this analysis verifies the applicability of the peak flow rate model.

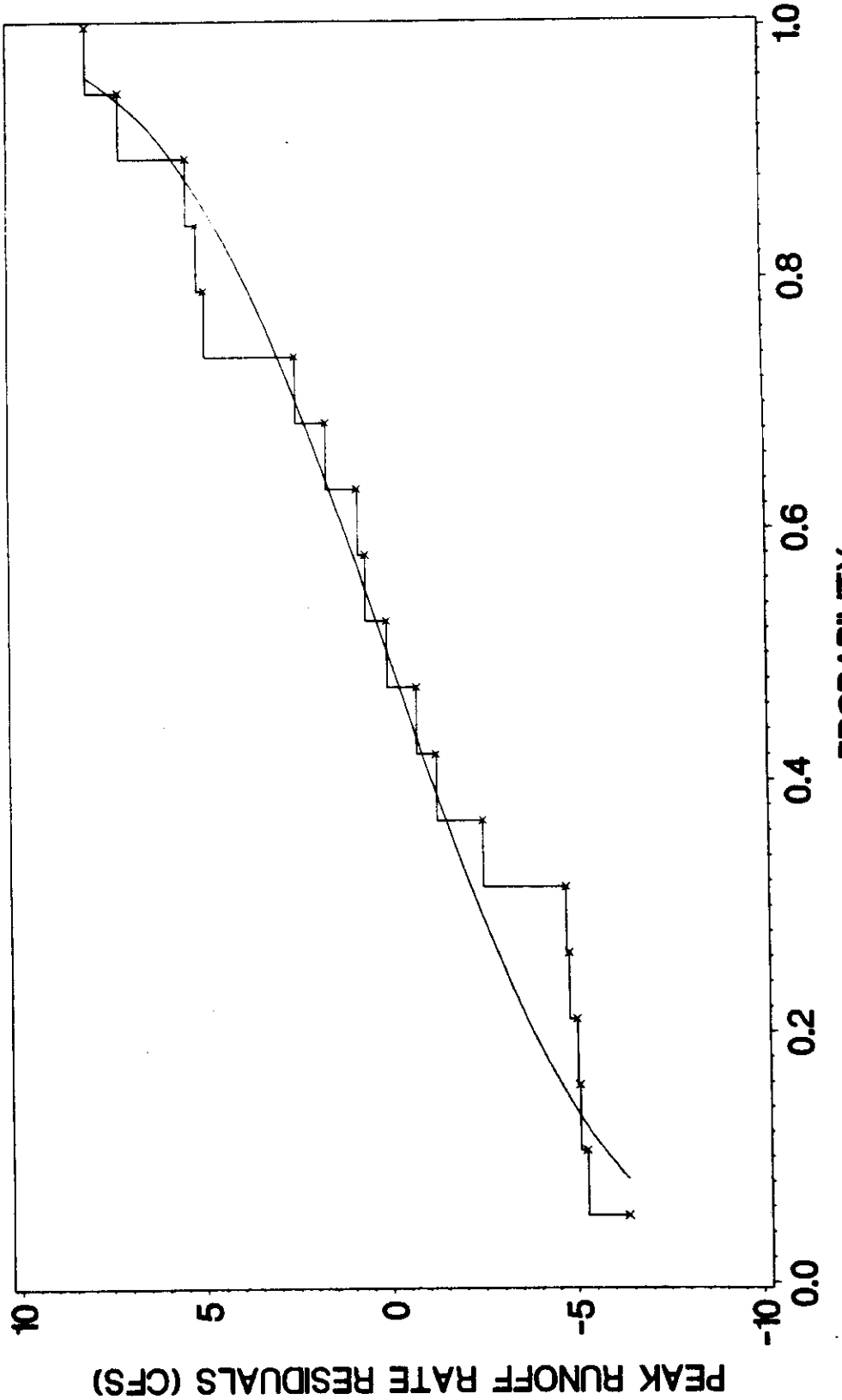


FIGURE 4.4
CDF FOR PEAK RUNOFF RATE RESIDUALS

TABLE 4.8

KOLMOGOROV-SMIRNOV TEST FOR RESIDUALS CDF FOR PEAK FLOW RATE MODEL

Residuals, ε_i (cfs)	Observe Empirical CDF: Cumulative Frequency	Observe Empirical CDF: Cumulative Probability, E (ε_i)	Normal CDF ($\mu_\varepsilon = 0.0, \sigma_\varepsilon = 4.5806$): $Z_i = \frac{\varepsilon_i}{s_\varepsilon}$	Normal CDF: Cumulative Probability, F(ε_i)	Deviations, Δ: | E (ε_i) – F (ε_i) |	Deviations, Δ: | E (ε_{i-1}) – F (ε_i) |
-6.3765	1	0.053	-1.39	0.082	0.029	—
-5.2736	2	0.105	-1.15	0.125	0.020	0.072
-5.0872	3	0.158	-1.11	0.134	0.024	0.029
-5.0323	4	0.211	-1.10	0.136	0.075	0.022
-4.8289	5	0.263	-1.05	0.147	0.116	0.064
-4.7618	6	0.316	-1.04	0.149	0.167	0.114
-2.5283	7	0.368	-0.55	0.291	0.077	0.025
-1.2871	8	0.421	-0.28	0.390	0.031	0.022
-.7631	9	0.474	-0.17	0.432	0.042	0.011
.02537	10	0.526	0.01	0.504	0.022	0.030
.5891	11	0.579	0.13	0.552	0.027	0.026
.7866	12	0.632	0.17	0.568	0.064	0.011
1.6172	13	0.684	0.35	0.637	0.047	0.005
2.4345	14	0.737	0.53	0.702	0.035	0.018
4.8788	15	0.789	1.07	0.858	0.069	0.121
5.0710	16	0.842	1.11	0.866	0.024	0.077
5.3446	17	0.895	1.17	0.879	0.016	0.037
7.1537	18	0.947	1.56	0.941	0.006	0.046
8.0380	19	1.000	1.75	0.960	0.040	0.013

Null Hypothesis, H_0: The residuals for the peak flow rate model are normally distributed with parameters $\mu_\varepsilon = 0.0$ and $\sigma_\varepsilon = 4.5806$.

Alternative Hypothesis, H_1: The residuals do not possess a normal distribution with the specified parameters.

Kolmogorov-Smirnov Test Statistic: $\Delta_{max} = 0.167$.

Critical Region at 0.05 Level of Significance for n=19: $\Delta_0 > 0.301$.

Decision: Do Not Reject H_0.

Hydrograph Shape

Modified SCS Standard Shape

The synthetic unit hydrograph method developed by the U.S. SCS uses a standard dimensionless hydrograph shape derived from an analysis of a large number of natural unit hydrographs for watersheds varying widely in size and geographic location (U.S. Department of Agriculture 1971). This standard unit hydrograph has been applied frequently throughout the world for design of hydrologic structures and systems.

The SCS standard unit hydrograph is shown graphically in Figure 4.5. The dimensionless ordinates are expressed as (h/h_p) where h is the flow rate per inch of runoff volume at any time t, and h_p is the peak flow rate per inch of runoff volume. The dimensionless values along the abscissa are expressed as (t/t_p) where t_p is the time-to-peak, which is the time from the beginning of direct runoff to the peak of the hydrograph.

Derivation, and application, of the SCS hydrograph is based upon the volume-to-peak, time-to-peak, and peak flow rate being interrelated as they would be for a triangular hydrograph shape. For any arbitrary hydrograph with a triangular shape,

$$Q_p = \frac{1/2\, t_p\, q_p}{A} \tag{4.34}$$

where,

Q_p = volume-to-peak, which is the volume between the start of runoff and the peak flow rate, (inches)

t_p = time-to-peak, (hours)

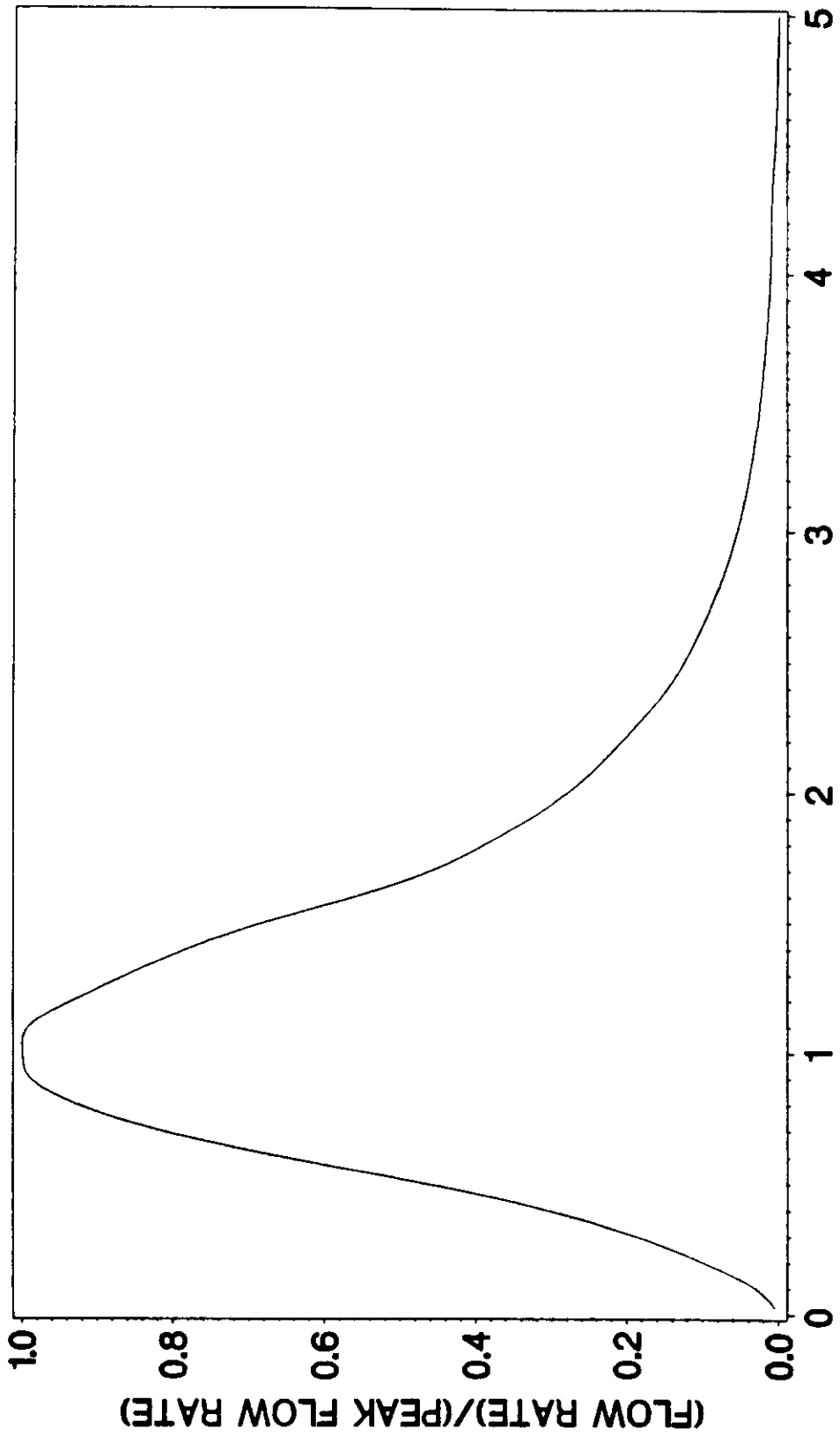


FIGURE 4.5
SCS DIMENSIONLESS UNIT HYDROGRAPH

q_p = peak flow rate, (acre-inch/hour)

A = watershed surface area, (acres).

Rearranging this in terms of q_p and dividing both sides by the total runoff volume, Q, yields

$$\frac{q_p}{Q} = \frac{2\,(Q_p / Q)A}{t_p} .$$

Now, define

$$U = Q_p / Q$$

which is just the ratio of volume-to-peak to total volume. The ratio U is a parameter that describes the shape of the hydrograph. Also, recall that,

$$h_p = q_p / Q$$

Substituting yields,

(4.35) $$h_p = \frac{2UA}{t_p} .$$

Equation 4.35 defines the relationship between the peak flow rate h_p, the time lag parameter t_p, and the shape parameter U for any triangular unit hydrograph.

In a similar fashion, the volume under the recession limb of a triangular hydrograph is given by,

(4.36) $$Q_r = \frac{1/2\, t_r q_p}{A}$$

where,

Q_r = recession limb volume, which is the volume between the peak flow rate and the end of runoff, (inches)

t_r = recession time, which is the time between the peak flow rate and the end of runoff, (hours).

But,

$$Q_r = Q - Q_p$$

Substituting, rearranging in terms of q_p, and dividing both sides by the total runoff volume, Q, yields,

$$\frac{q_p}{Q} = \frac{2\,[Q - Q_p)\,/\,Q\,]\,A}{t_r}.$$

But, recall that,

$$h_p = \frac{q_p}{Q} \text{ and } U = \frac{Q_p}{Q}.$$

Substituting yields,

(4.37) $$h_p = \frac{2(1-U)A}{t_r}$$

Equation 4.37 defines the relationship between the peak flow rate h_p, the recession time parameter t_r, and the shape parameter U for any triangular unit hydrograph.

The SCS method is based upon the assumptions that the shape parameter U is a constant, U = 3/8, and that the relationships for triangular unit hydrographs of equations 4.35 and 4.37 also hold for the standard shape defined in Figure 4.5. Therefore, the basic relationships between h_p, t_p, and t_r for the standard SCS dimensionless unit hydrograph are given by:

(1) $$h_p = \frac{3/4\,A}{t_p}$$

(2) $$h_p = \frac{5/4\ A}{t_r}$$

(3) $$t_r = 5/3\ t_p.$$

Unfortunately, the observed hydrographs in this study were found to have a highly variable ratio of volume-to-peak to total volume as discussed in Chapter 2. In fact, the parameter U of equations 4.35 and 4.37 has been shown to be an independent random variable possessing a log-normal pdf as demonstrated in the following section. Therefore, the SCS procedure has been modified in this study to allow for a variable value of the parameter U which acts to modify the standard shape of Figure 4.5.

The modified SCS procedure for generating a synthetic hydrograph involves using the observed, or simulated, values of q_p, Q_p, and Q to compute $h_p = (q_p/Q)$ and $U = (Q_p/Q)$. Equations 4.35 and 4.37 are then used to compute the appropriate values of t_p and t_r. Then, the flow rates for the synthetic hydrograph are obtained, as they are in the standard procedure, by multiplying the ordinates of Figure 4.5 by the product of h_p and Q. The values on the time scale that correspond to the flow rate values are then computed as follows:

(4.38)
$$t = \begin{cases} \left(\frac{t}{t_p}\right) t_p, & t \le t_p \\ t_p + \left(\frac{t'}{t_r}\right) t_r, & t > t_p . \end{cases}$$

The term (t/t_p) represents the standard SCS abscissas on the rising limb of the dimensionless unit hydrograph from Figure 4.5. The term (t'/t_r) represents the standard SCS abscissas on the recession limb of the hydrograph from Figure

4.5, but reported in terms of t_r rather than t_p and with the origin of the time scale shifted to correspond to the peak of the hydrograph. This shift is accomplished by defining the new time variable t' as,

$$t' = t - t_p$$

that is, t' is just the time after the hydrograph peak. Any abscissa value, (t/t_p), from Figure 4.5 can be written as

$$\left(\frac{t}{t_p}\right) = \left(\frac{t' + t_p}{t_p}\right)$$

or,

$$\left(\frac{t}{t_p}\right) = \left(\frac{t'}{t_p}\right) + 1$$

by simple substitution. But, for the standard shape, we already know that $t_p = 3/5\ t_r$. Substituting this on the right-hand side yields,

$$\left(\frac{t}{t_p}\right) = \left(\frac{t'}{3/5\ t_r}\right) + 1$$

Rearranging,

(4.39)
$$\left(\frac{t'}{t_r}\right) = 3/5 \left[\left(\frac{t}{t_p}\right) - 1\right].$$

Equation 4.39 is used to accomplish the transformation from the abscissas of Figure 4.5 to the equivalent form (t'/t_r) for use in Equation 4.38.

Application of equation 4.38, in conjunction with the computed values of t_p and t_r based on the shape parameter U from equations 4.35 and 4.37, yields a synthetic hydrograph with a modified shape from that produced by the standard SCS procedure. The modified and standard SCS synthetic hydrographs have

identical total volumes and peak flow rates; but the distribution of flow rates in time, time-to-peak, recession time, and overall time base of the hydrographs are different.

A visual comparison of several observed direct runoff hydrographs to synthetic hydrographs are shown in Figures 4.6, 4.7, and 4.8. Synthetic hydrographs were generated using the modified SCS shape based on the observed values of q_p, Q_p, and Q for each event. Therefore, the total volume, volume-to-peak, recession limb volume, and peak flow rate of the synthetic hydrographs are identical to those of the observed hydrographs. The overall agreement between observed and computed hydrograph shapes is generally good. The events selected for presentation in Figures 4.6, 4.7, and 4.8 were chosen to represent small, medium, and large runoff events, respectively.

Probability Distribution for the Shape Parameter U

As defined in the previous section, the parameter U is the ratio of volume-to-peak to total volume for a direct runoff hydrograph. In order to generate a complete hydrograph for each stochastically simulated rainfall-runoff event, it is necessary to obtain a simulated value of U for each event. Therefore, a 2-parameter log-normal pdf was fit to the observed data for U. This was actually accomplished by fitting a normal pdf to the log-transformed variable V, defined as follows:

$$V = \ln(U).$$

The normal pfd for V is of the familiar form,

$$f(v) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-1/2\left[\frac{v-\mu_v}{\sigma_v}\right]^2}, \quad -\infty < v < \infty \tag{4.40}$$

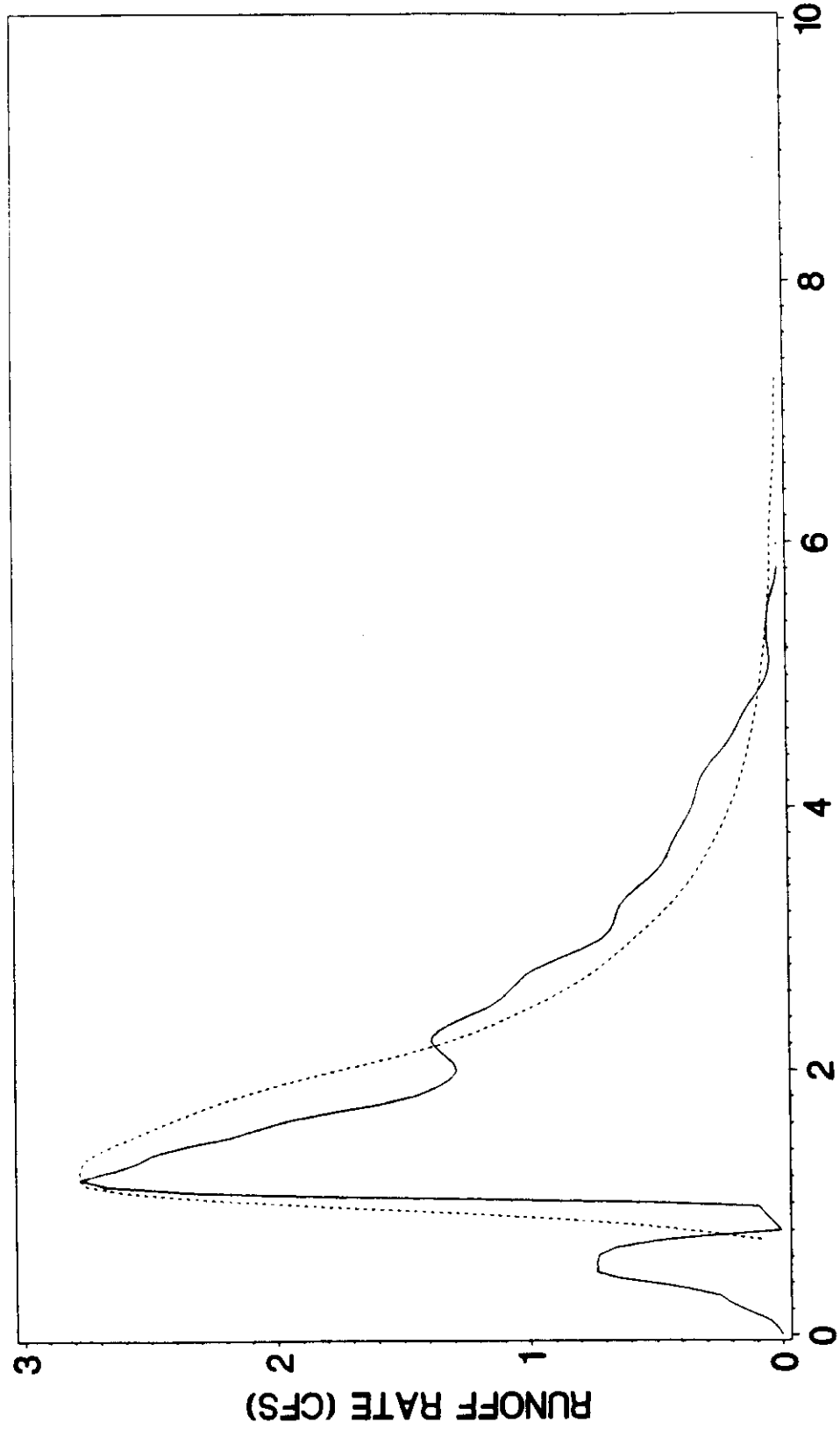


FIGURE 4.6
HYDROGRAPHS FOR EVENT OF 07/17/87

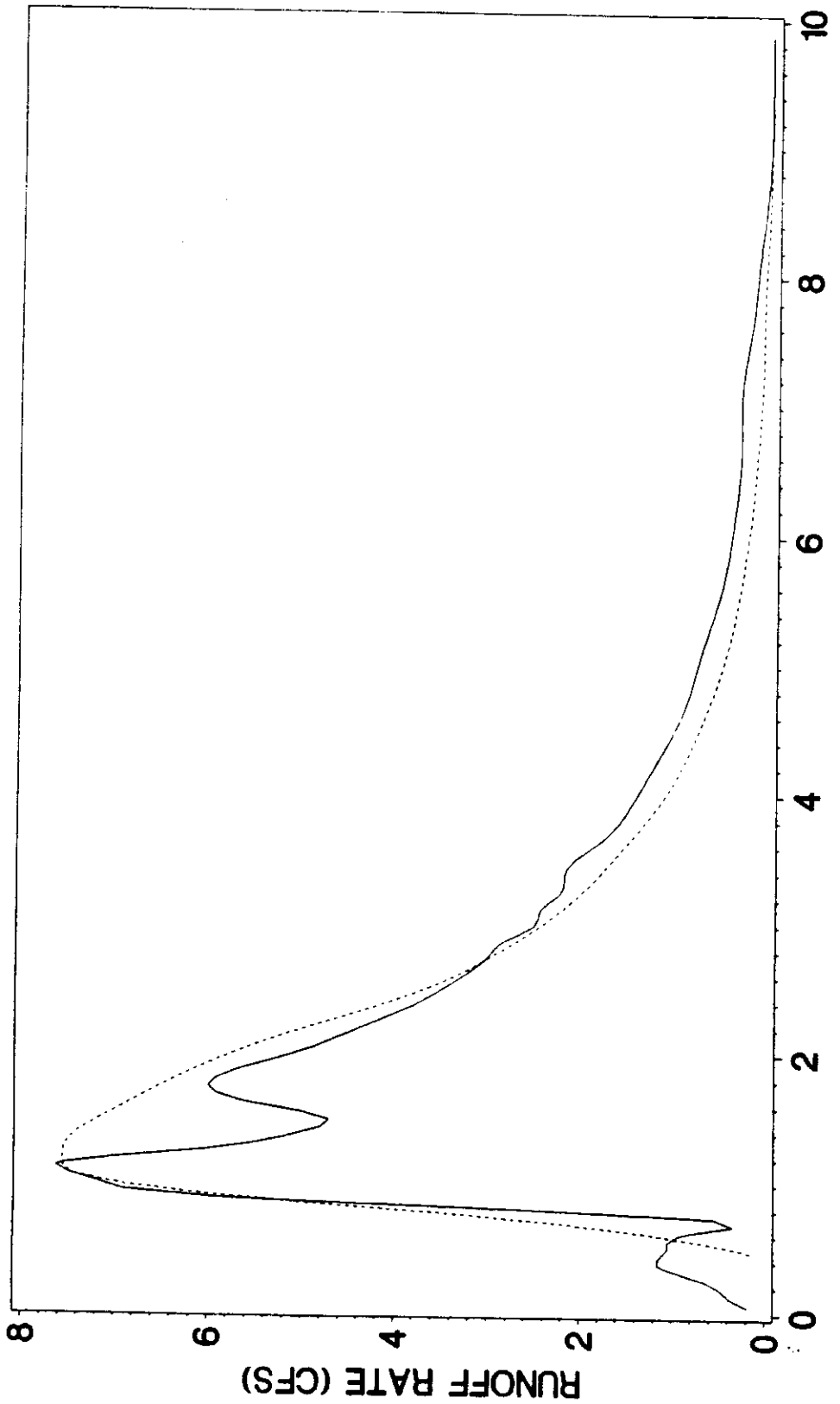


FIGURE 4.7
HYDROGRAPHS FOR EVENT OF 11/25/87

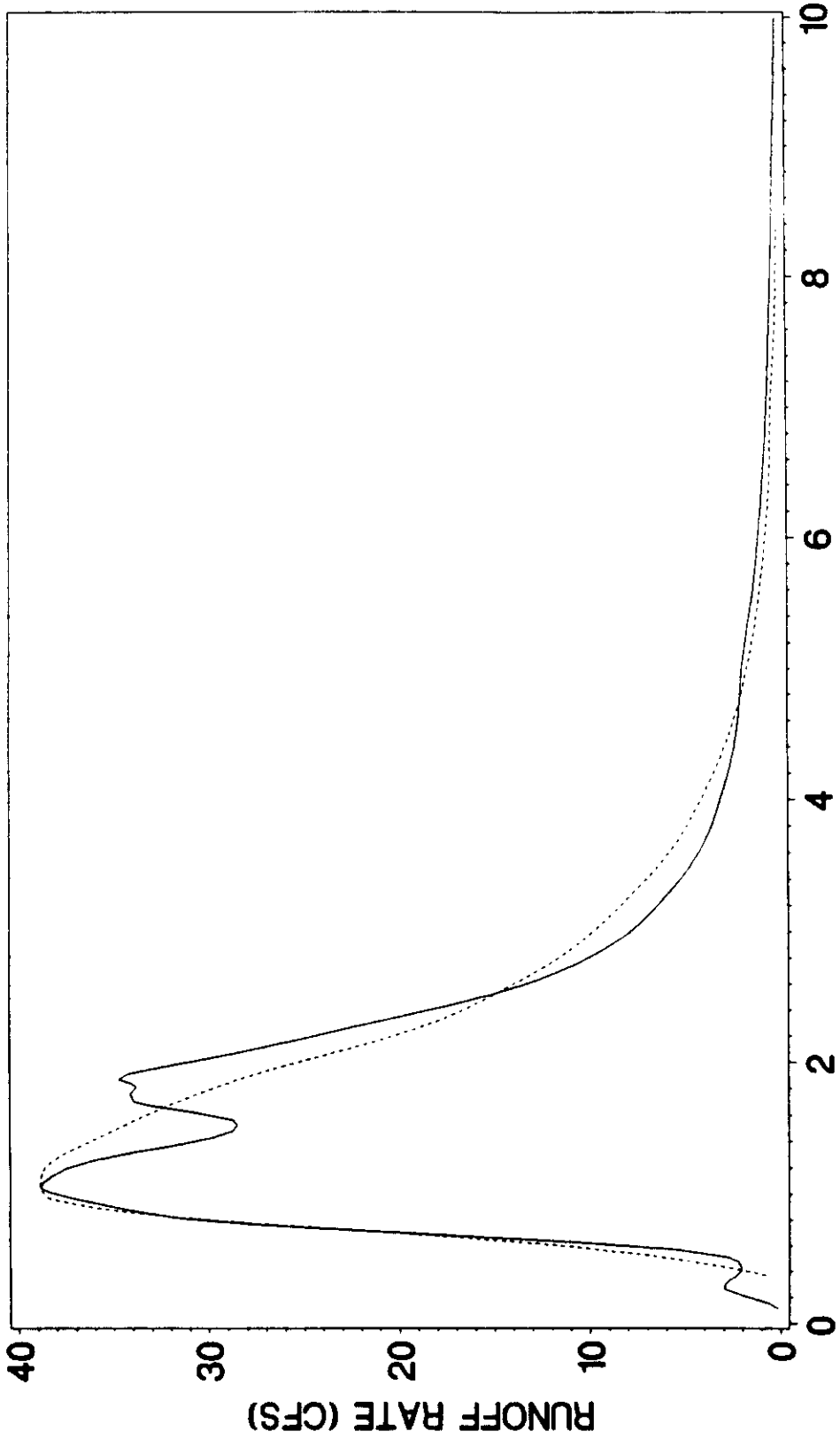


FIGURE 4.8
HYDROGRAPHS FOR EVENT OF 04/17/88

where,

v = any arbitrary value of the random variable V ($-\infty<v<\infty$)

μ_v = population mean ($-\infty<\mu_v<\infty$)

σ_v = population standard deviation ($\sigma_v>0$).

The population mean and standard deviation must be estimated from the sample data using the MOM as follows:

$$\hat{\mu}_v = \bar{v} = \frac{\sum_{i=1}^{n} v_i}{n} \tag{4.41}$$

$$\hat{\sigma}_v = s_v = + \sqrt{\frac{\sum_{i=1}^{n} (v_i - \bar{v})^2}{n-1}} \; . \tag{4.42}$$

Where, of course, $\bar{v}$ and s_v are the sample mean and sample standard deviation of the observed data. Equation 4.42 is not in strict conformance with the MOM since s_v involves the denominator (n-1) instead of n as required by the definition of sample moments. However, the error is acceptably small in this case.

Since observed hydrographs with total runoff volumes of less than 0.01 inch were of extremely unusual and highly variable shape, as discussed in Chapter 2, it was decided to use only significant runoff events with total volume greater than or equal to 0.01 inch to characterize U. Using the data for the 27 events meeting this specification monitored during the field study, the sample moments for the log-transformed variable V were computed as,

$$\bar{v} = -1.8316$$

$$s_v = 0.5332.$$

Plots of the observed empirical cdf and the theoretical normal cdf with μ_V = - 1.8316 and σ_V = 0.5332 are shown on normal probability paper in Figure 4.9. The agreement between them is very good.

An objective test of the goodness of fit was obtained by application of the Kolmogorov - Smirnov one-sample test. This analysis is shown in detail in Table 4.9. The maximum deviation between empirical and theoretical cdf's was Δ_{max} = 0.107. The critical value was Δ_0 = 0.254 for a two-sided test with n = 27 at a 0.05 level of significance. The test statistic lies well outside the crucial region of the test. Therefore, do not reject the null hypothesis that the observed values of the log-transformed variable V represent a random sample of size 27 from a normal pdf with the specified mean and standard deviation.

Before the fitted cdf for V can be legitimately used in simulations to generate randomly sampled values U to allow development of synthetic runoff hydrographs, it must first be shown that U is statistically independent of runoff volume, Q. Because of the small sample size, n = 27, the chi-square test of independence used extensively in Chapter 3 was not practical here. Consequently, Kendall's tau test of independence was employed. This is a computationally simple nonparametric statistical test of independence between two random variables. The data required to conduct the test consists of n pairs of observations of the two variables. Each pair of observations must represent two different measurements on the same object or event.

Kendall's tau, τ, is a population parameter defined as the probability of concordance minus the probability of discordance between two arbitrary pairs of observations. Pairs of observations (x_i, y_i) and (x_j, y_j) are defined as concordant if the algebraic sign of the difference $(x_i - x_j)$ is the same as the sign

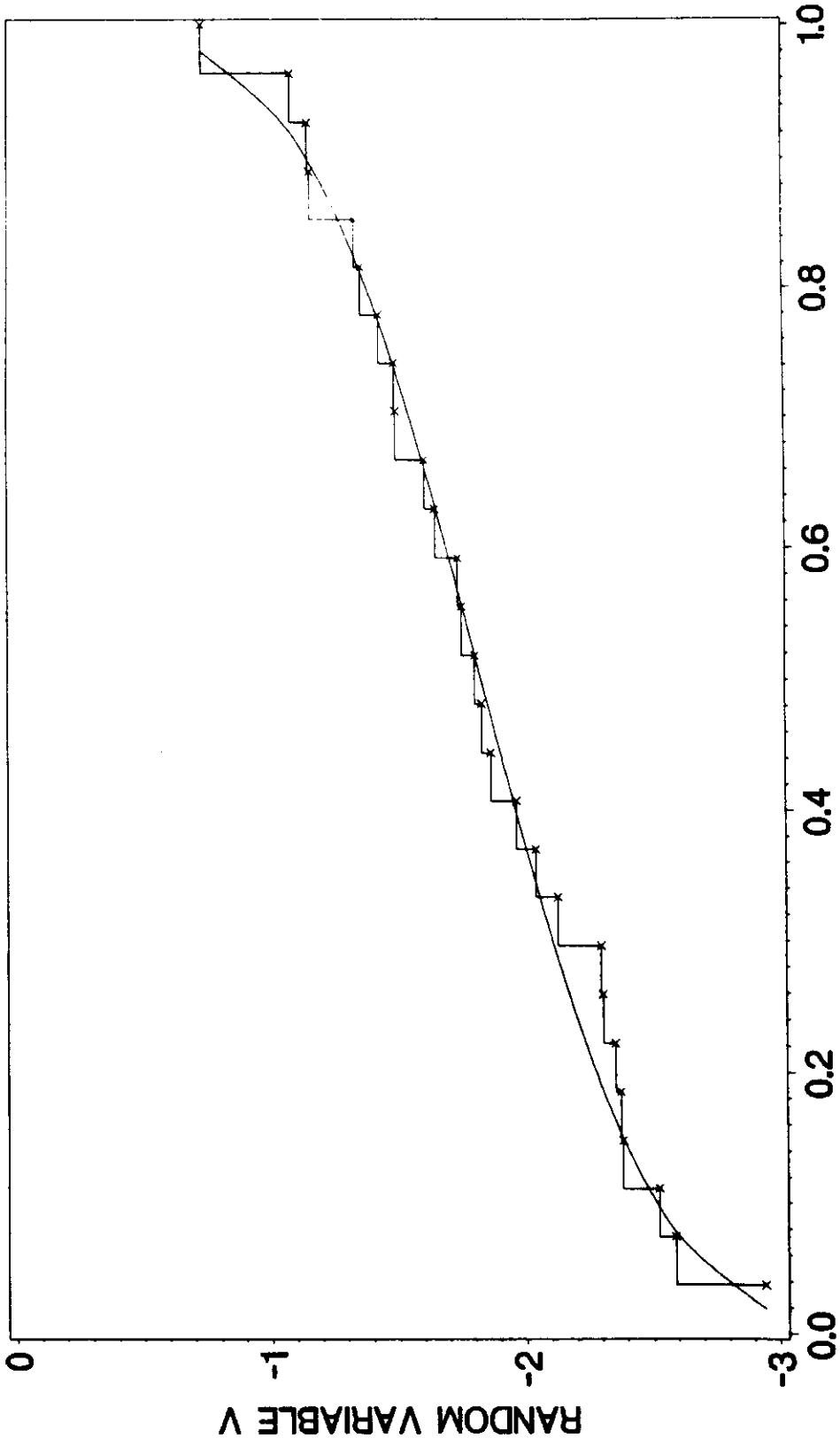


FIGURE 4.9
CDF FOR RANDOM VARIABLE V

TABLE 4.9

KOLMOGOROV-SMIRNOV TEST FOR RESIDUALS CDF FOR THE RANDOM VARIABLE V

	Observed Empirical CDF		Lognormal CDF ($\mu_v = -1.8316$, $\sigma_v = 0.5332$)		Deviations, Δ	
$v_i = \ln\left(\frac{V_p}{Q}\right)$	Cumulative Frequency	Cumulative Probability, E (v_i)	$Z_i = \frac{v_i - \bar{v}}{s_v}$	Cumulative Probability, F(v_i)	$\lvert E(v_i) - F(v_i) \rvert$	$\lvert E(v_{i-1}) - F(v_i) \rvert$
-2.9426	1	.037	-2.08	.019	.020	---
-2.5903	2	.074	-1.42	.078	.004	.041
-2.5257	3	.111	-1.30	.097	.014	.023
-2.3839	4	.148	-1.04	.149	.001	.038
-2.3776	5	.185	-1.02	.154	.031	.006
-2.3571	6	.222	-0.99	.161	.061	.024
-2.3088	7	.259	-0.90	.184	.075	.038
-2.3026	8	.296	-0.88	.189	.107	.070
-2.1327	9	.333	-0.56	.288	.045	.008
-2.0448	10	.370	-0.40	.345	.025	.012
-1.9694	11	.407	-0.26	.397	.010	.027
-1.8683	12	.444	-0.07	.472	.028	.065
-1.8326	13	.481	-0.00	.500	.019	.056
-1.8062	14	.519	0.05	.520	.001	.039
-1.7540	15	.556	0.15	.560	.004	.041
-1.7401	16	.593	0.17	.568	.025	.012
-1.6511	17	.630	0.34	.633	.003	.040
-1.6094	18	.667	0.42	.663	.004	.033

TABLE 4.9 (Continued)

KOLMOGOROV-SMIRNOV TEST FOR RESIDUALS CDF FOR THE RANDOM VARIABLE V

	Observed Empirical CDF		Lognormal CDF ($\mu_v = -1.8316$, $\sigma_v = 0.5332$)		Deviations, Δ	
$v_i = \ln\left(\frac{V_p}{Q}\right)$	Cumulative Frequency	Cumulative Probability, E (v_i)	$Z_i = \frac{v_i - \bar{v}}{s_v}$	Cumulative Probability, F(v_i)	E (v_i) – F (v_i) |	| E (v_{i-1}) – F (v_i) |
-1.4970	19	.704	0.63	.736	.032	.069
-1.4916	20	.741	0.64	.739	.002	.035
-1.4308	21	.778	0.75	.773	.005	.032
-1.3610	22	.815	0.88	.811	.004	.033
-1.3363	23	.852	0.93	.824	.028	.009
-1.1619	24	.889	1.26	.896	.007	.044
-1.1540	25	.926	1.27	.898	.028	.009
-1.0857	26	.963	1.40	.919	.044	.007
-0.7370	27	1.000	2.05	.980	.020	.017

Null Hypothesis, H_0: The residuals for the peak flow rate model are normally distributed with parameters $\mu_v = -1.8316$ and $\sigma_v = 0.5332$.
Alternative Hypothesis, H_1: The residuals do not possess a normal distribution with the specified parameters.
Kolmogorov-Smirnov Test Statistic: $\Delta_{max} = 0.107$.
Critical Region at 0.05 Level of Significance for n=27 (Two-Sided Test): $\Delta_0 > 0.254$.
Decision: Do Not Reject H_0.

of the difference $(y_i - y_j)$. Pairs of observations are said to be discordant if the algebraic signs differ.

The null hypothesis in Kendall's τ test is that the random variables X and Y are independent which, in turn, requires that $\tau = 0$. For the two-sided test used here, the alternative hypothesis is that X and Y are not independent such that $\tau \neq 0$. The test statistic is the sample estimator $\hat{\tau}$ which is evaluated using the following procedure:

(1) Arrange the n pairs of (X, Y) values in columns ranked by increasing values of X.

(2) Then, for each of the n pairs, count and tabulate the number of concordances and the number of discordances for all subsequent pairs in the table.

(3) Determine N_c, which is the sum of the number of concordances for all pairs. Determine N_d, which is the sum of the number of discordances for all pairs.

(4) Then, the test statistic is computed from,

$$\hat{\tau} = \frac{2\left(N_c - N_d\right)}{n\,(n-1)} \quad , \quad -1 \leq \hat{\tau} \leq 1 \, . \tag{4.43}$$

The decision rule for the test is based on comparing the absolute value of $\hat{\tau}$ to tabulated critical values, τ_c, from standard statistical tables. Then, do not accept H_0 if $|\hat{\tau}| > \tau_c$. Otherwise, do not reject H_0.

The Kendall's τ test of independence for the random variables U and Q is shown in detail in Table 4.10. Using equation 4.43, $\hat{\tau} = -0.191$. The critical value of the test statistic, $\tau_c = 0.271$, was obtained from standard tables for a

TABLE 4.10
KENDALL'S τ TEST OF INDEPENDENCE FOR
THE RANDOM VARIABLES U AND Q

Runoff Volume, Q		$U = (Q_p / Q)$		
Value (inch)	Rank	Value (inch/inch)	Subsequent Concordances	Subsequent Discordances
.011	1	.053	26	0
.013	2	.315	2	23
.014	3	.164	12	12
.014	4	.479	0	23
.015	5	.080	21	1
.020	6	.160	11	10
.023	7	.200	7	13
.023	8	.239	4	15
.028	9	.225	4	14
.031	10	.313	1	16
.039	11	.256	2	14
.042	12	.224	2	13
.044	13	.075	14	0
.049	14	.176	3	10
.052	15	.173	3	9
.057	16	.154	3	8
.064	17	.092	10	0
.077	18	.338	0	9
.083	19	.093	8	0
.086	20	.140	2	5
.085	21	.129	2	4
.135	22	.119	2	3
.156	23	.263	0	4
.161	24	.099	2	1
.170	25	.100	1	1
.245	26	.192	0	1
.264	27	.095	0	0
			$N_c = 142$	$N_d = 209$

Null Hypothesis, H_0: The random variables Q and U are independent; $\tau = 0$.

Alternative Hypothesis, H_i: The random variables are not independent; $\tau \neq 0$.

Test Statistic from Equation 4.43 for n=27: $\hat{\tau} = -0.191$.

Critical Value of the Test Statistic at 0.05 Level of Significance (Two-Sided Test): $\tau_c = 0.271$.

Decision Rule: Do not accept H_0 if the absolute value of Kendall's τ test statistic, $|\hat{\tau}|$, exceeds the critical value, τ_c, for a two-sided test at the 0.05 Level of Significance. Otherwise, do not reject H_0.

Decision: Do Not Reject H_0.

two-sided test with n = 27 (Daniel 1978). Therefore, since $|\hat{\tau}| < \tau_c$, do not reject the null hypothesis that U and Q are independent random variables.

This analysis confirms that the hydrograph shape parameter U can be simulated as an independent random variable using the log-transformed variable V and it's associated pdf.

CHAPTER 5

STOCHASTIC MODEL DEVELOPMENT AND SIMULATION

Stochastic Rainfall Runoff Model

Model Development

A stochastic simulation model was developed to generate synthetic sequences of rainfall-runoff event data for the study watershed. The simulation model was developed in the form of a SAS computer program which is listed in Table C.1 in Appendix C. There are numerous comment statements in the program listing to make the program as self-explanatory as possible. A detailed discussion of the operations performed by the program is presented in the following paragraphs. The variable names used in the computer program are shown in parenthesis in the discussion as a cross-reference to the listing in Table C.1 in Appendix C.

The stochastic simulation model performs the following general procedures in sequence:

(1) An initialization period in years is specified by the user (STARTYRS). All events in the initialization period are omitted from the final results of the simulation in order to eliminate any bias introduced by the initial conditions. A minimum initialization period of 5 years is recommended.

(2) The number of years for simulation output (NYEARS) is set by the user. Thus, the total number of years simulated for a given execution of the program is the sum of the initialization period and the simulation output period (i.e., STARTYRS + NYEARS).

(3) Variables which accumulate the elapsed time since the last runoff event (TL) and the elapsed time since the last runoff event with a volume of 0.01 inch or more (TR) are initialized to zero.

(4) All counter variables are initialized to zero. Four counters are maintained by the model; the total cumulative time in hours since the beginning of the simulation (CUMHOURS), the calendar year since the beginning of the simulation (YEAR), the total number of rainfall events simulated (NRAINS), and the total number of runoff events simulated (NRUNOFFS).

(5) Theoretical probability distributions are randomly sampled for rainfall event volume (R), time between events (T), and event duration (D). Derivation of the theoretical distributions and development of the computer program for random sampling of these distributions were presented in detail in Chapter 3. During simulation of the initial rainfall event, the initial value of the time between events (T) is chosen to exceed 144 hours, or 6 days, so that the initial condition for the antecedent rainfall variable (R0T6) will be zero (i.e., R0T6 = 0 initially).

(6) The accumulated time elapsed since the last runoff event (TL) is incremented either by the time between rainfall events (T) for the current event if the prior event produced runoff, or by the time between rainfall events (T) for the current event plus the event duration (D) for the prior event if the prior event did not produce runoff. An upper bound of 1513 hours is set for the time elapsed since the last runoff event (TL). The reason for this upper bound is explained below.

(7) The accumulated time elapsed since the last runoff event with a volume of 0.01 inch or greater (TR) is incremented either by the time between rainfall events (T) for the current event if the prior event produced 0.01 inch of runoff or more, or by the time between rainfall events (T) for the current

event plus the event duration (D) for the prior event if the prior event produced less than 0.01 inch of runoff. An upper bound of 1513 hours is set for the time elapsed since the last runoff event with a volume of 0.01 inch or greater (TR). The reason for this upper bound is explained below.

(8) All rainfall event volumes in the 6 days preceding the current rainfall event are summed to yield the antecedent rainfall variable (R0T6). This requires maintaining, in computer memory, the values of all rainfall event variables (R, T, and D) for the 36 storms immediately preceding the current event. No more than 36 storms can occur in 6 days because of the way hourly rainfall data is resolved into individual events as explained below.

(9) Theoretical normal pdf's are randomly sampled for the runoff volume white noise component (QEPS) and for the peak runoff rate white noise component (QPEPS). The program uses the SAS internal function for generating normal deviates. The white noise component of the runoff volume model has a mean of zero and a standard deviation of 0.03148 inch and the white noise component of the peak runoff rate model has a mean of zero and a standard deviation of 4.5806 cfs as shown in Chapter 4.

(10) The areal weighted rainfall volume (RW) is assumed equal to the simulated point rainfall volume (R), and the average rainfall intensity (IA) is computed (i.e., IA = RW/D).

(11) The direct runoff volume (Q) and the peak runoff rate (QP) are then computed using equations 4.27 and 4.30, respectively. In the computer model, these equations assume the form,

(5.1) $$Q = -.05017+.1671*RW-.0001263*TR+.02976*ROT6+QEPS.$$

and,

(5.2) $$QP = -4.6053+12.3678*RW-.02144*TL+32.4367*IA+QPEPS.$$

(12) The results of the simulation are checked for consistency. If either the simulated runoff volume (Q) or peak runoff rate (QP), or both, is non-positive then the rainfall event is treated as a non-runoff producing event (i.e., Q=0 and QP=0). If the random components of the runoff equations produce a rainfall-runoff event with a runoff volume that exceeds the rainfall volume (i.e., Q>RW), then the rainfall event is discarded, the appropriate variables are reset to their previous values, and the iteration is restarted at Step (5.). Preliminary simulations indicated that this occurs with negligible frequency.

(13) The counters for number of rainfall events (NRAINS) and cumulative simulation hours (CUMHOURS) are incremented and the antecedent rainfall variables are reset.

(14) If the current rainfall event does not produce runoff (i.e., Q and/or QP is non-positive), then the program returns to Step (5.) for simulation of the next rainfall-runoff event.

(15) If the given rainfall event produces runoff (i.e., Q>0 and QP>0), then the counter variables for number of runoff events (NRUNOFFS) and current calendar year of the simulation (YEAR) are incremented.

(16) If the rainfall-runoff event occurs in a calendar year after the initialization period (STARTYRS) and before the end of the simulation period (i.e., STARTYRS + NYEARS), then the simulation results are written to the output file.

(17) The elapsed time between runoff events (TL) is reset to zero and the elapsed time between runoff events with a volume of 0.01 inch or greater (TR) is reset to zero, if appropriate. The program then returns to Step (5.) for another iteration until the simulation is completed (i.e., until YEAR > STARTYRS + NYEARS).

The initial conditions specified in the program correspond to assuming that the rainfall event that occurred immediately prior to the start of the simulation produced a runoff volume of 0.01 inch or more and that it occurred more than 6 days before.

As a practical hydrologic matter, an upper bound was placed on the values of the two elapsed time between runoff event variables (i.e., TL and TR). These elapsed time variables are measures of the evapotranspiration and gravity drainage moisture losses from the soil between runoff events as discussed in Chapter 4. The larger the numerical magnitude of these variables, the more moisture that has been lost from the soil. However, since there is obviously a finite amount of moisture stored in the soil of a given watershed, there must be an upper bound on the values of the elapsed time variables above which there is no additional moisture loss. The best estimate of this upper bound is the maximum observed values of the elapsed time variables from the field study. The maximum observed value in the field data was 1513 hours, or 63 days, for both the time elapsed since the last runoff event (TL) and the time elapsed since the last runoff event with a volume of 0.01 inch or more (TR). These upper bounds were incorporated into the simulation model.

An explanation is needed for the statement that no more than 36 rainfall events can occur in 6 days. The absolute maximum number of events for any period of time must involve a sequence of events all having a duration of one hour, because this is the minimum possible duration obtained with hourly data.

Since 3 hours is the time used to resolve hourly data into individual events, the sequence of 1-hour events must be separated from each other by a minimum of 3 hours with no rainfall and the last event must be followed by a minimum of 3 hours without rain to separate it from the current event. Therefore, for 36 consecutive 1-hour events, the minimum possible total elapsed time is 36(1 hour)+36(3 hours)=144 hours, which is 6 days. Conversely, no more than 36 rainfall events can occur in 6 days.

The output file created by the simulation model is a permanent, machine language SAS file stored on disk. This file can be readily accessed by easily constructed programs using SAS procedures for detailed analysis of the simulation results. For example, it is a simple matter to obtain summary statistics and frequency analyses of simulated runoff volumes and peak runoff rates or to generate hydrographs using the synthetic hydrograph procedure derived in detail in Chapter 4. The variable values stored in the output file for each event simulated, in the actual sequence in which they appear in the file, are as follows:

(1) cumulative number of rainfall events (NRAINS)

(2) cumulative number of runoff events (NRUNOFFS)

(3) cumulative time since the beginning of the simulation in hours (CUMHOURS)

(4) the calendar year since the beginning of the simulation (YEAR)

(5) runoff volume in inches (Q)

(6) peak runoff rate in cfs (QP)

(7) rainfall volume in inches (R)

(8) elapsed time since the last rainfall event in hours (T)

(9) duration of the rainfall event in hours (D)

(10) elapsed time since the last runoff event in hours (TL)

(11) elapsed time since the last runoff event with a volume of 0.01 inch or more in hours (TR)

(12) antecedent rainfall volume in the preceding 6 days (R0T6).

Simulation Results

The stochastic simulation model was used to generate long-term runoff event data for the study watershed. An initialization period of 5 years was used and the simulation was set at 100 years. This means that the model simulated runoff events for 105 years and the first 5 years were omitted from the results and then years 6 through 105 were written to the output file as the simulation results. A series of relatively elementary SAS programs were then used to read the output file and to perform the desired statistical analyses on selected variables.

The 100 years of simulated data which followed the 5 year initialization period involved 8800 rainfall events which resulted in 1570 runoff events being written to the output file. Simple statistics were computed and a frequency analysis was performed for runoff event volume and for peak runoff rate for the 1570 simulated runoff events. The results of these analyses are summarized in Table 5.1. The average runoff event volume was 0.160 inch and the average peak runoff rate was 27.3 cfs. These arithmetic mean values have been measurably influenced by the few extreme flood events as revealed by the significantly lower median values of 0.115 inch and 16.7 cfs, respectively. Empirical cdf's for runoff volume and peak runoff rate, developed using simulation output data, are shown graphically in Figures 5.1 and 5.2, respectively.

These final analyses of the simulation results satisfy the primary objective of this study, which was to develop a stochastic simulation model capable of

TABLE 5.1

STATISTICS FOR SIMULATED RUNOFF VOLUME AND PEAK RUNOFF RATE

Statistic	Runoff Volume Q (inches)	Peak Runoff Rate, q_p (cfs)
Number of Observations, n	1570	1570
Mean, $\bar{x}$	.160	27.3
Standard Deviation, s_x	.154	32.3
Minimum Value, x_{min}	.000268	.00110
5 Percentile Value, x_5	.0111	1.30
10 Percentile Value, x_{10}	.0216	2.57
25 Percentile Value, x_{25}	.0536	6.54
50 Percentile Value, x_{50}	.115	16.7
75 Percentile Value, x_{75}	.216	35.7
90 Percentile Value, x_{90}	.348	62.9
95 Percentile Value, x_{95}	.464	86.3
Maximum Value, x_{max}	1.25	316.

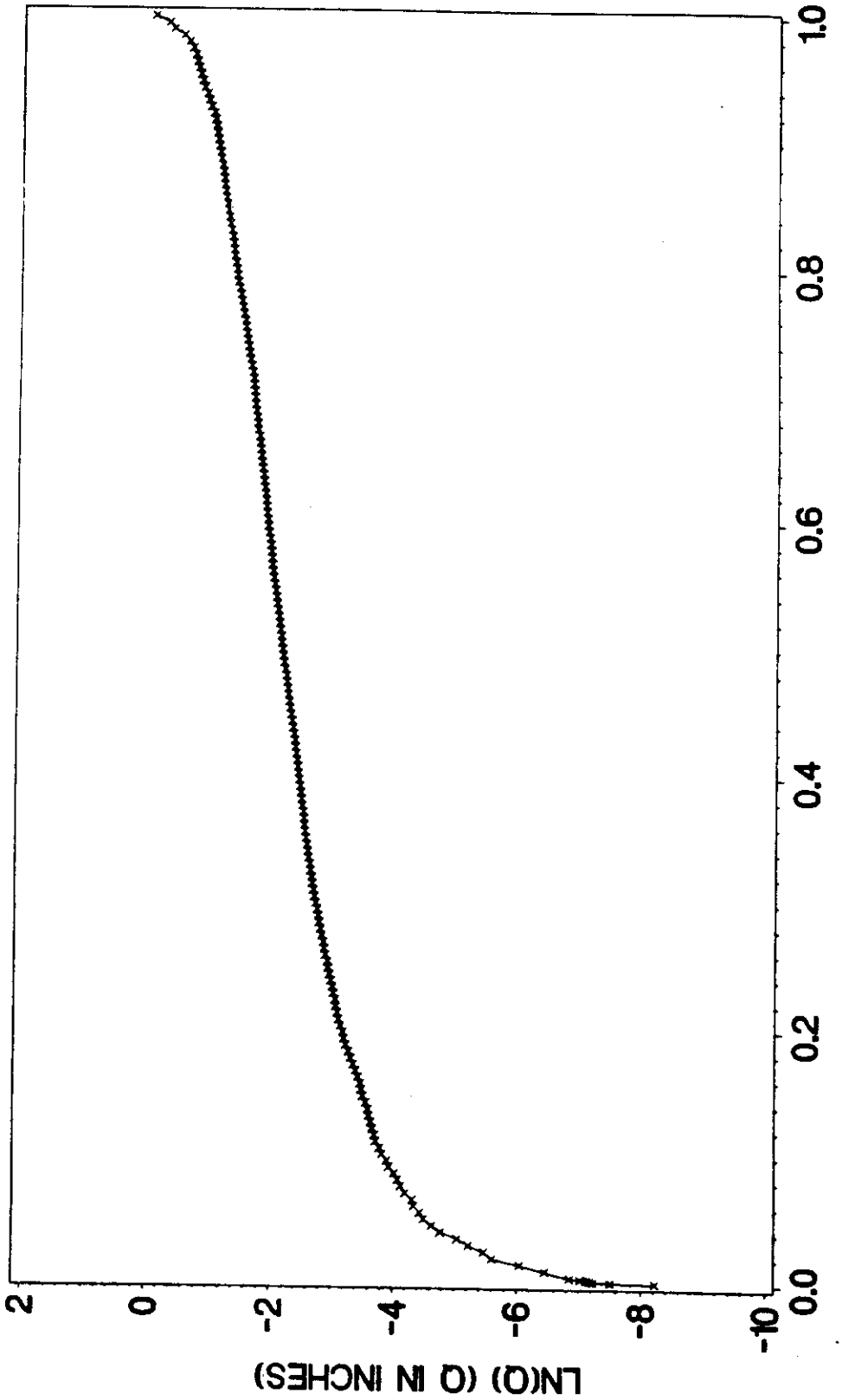


FIGURE 5.1
CDF FOR SIMULATED RUNOFF VOLUME

FIGURE 5.2
CDF FOR SIMULATED PEAK RUNOFF RATE

defining the detailed statistical characteristics of the rainfall-runoff process for the study watershed. These results, and the simulation model itself, are ideally suited for application to NPS pollutant load estimation for the study watershed.

Verification of Simulation Results

It is appropriate to verify that the simulation results are consistent with observed field data. The calendar year from July 1987 to June 1988 over which the complete set of rainfall-runoff event data were collected happened to be an unusually dry year. A total weighted rainfall volume of only 28.89 inches was recorded for the study watershed over this period as shown in Chapter 2. A frequency analysis for total annual rainfall volume for the 50 years from 1939 to 1988 at the rain gage at Love Field (NOAA 1939-1988) is shown in Table C.2 in Appendix C. Only 34 percent of the annual rainfall amounts are less than or equal to the 28.89 inches that fell on the study watershed during the calendar year monitored.

The total annual runoff volume for each calendar year of the simulation was computed and then a frequency analysis was performed on these 100 annual totals as shown in Table C.3 of Appendix C. The total annual runoff volume for the study watershed over the year monitored was approximately 1.49 inches as shown in Chapter 2. Only 27 percent of the simulated annual runoff amounts are less than or equal to 1.49 inches. This is remarkably consistent with the 34 percentile value for the corresponding rainfall volume. This verifies that the simulation model produces long-term results that are consistent with physical reality.

Another way of verifying simulation results is to demonstrate that the observed sequence of runoff event volumes could have been generated by the simulation model. This was done by selecting a calendar year from the

simulation results that had a total annual runoff volume approximately the same as the observed year and then performing a Kolmogovov-Smirnov two-sample test to test the null hypothesis that the two independent annual sequences of runoff volume were obtained from the same population. The Kolmogorov-Smirnov two-sample test employs as the test statistic the maximum deviation between the observed empirical cdf, $E(Q_0)$, and the simulated empirical cdf, $E(Q_s)$, for runoff event volume spanning one calendar year:

$$\Delta_{max} = \max_i \left\{ \left| E_i(Q_0) - E_i(Q_s) \right| \right\} \tag{5.3}$$

where,

Δ_{max} = absolute value of the maximum difference between the two empirical cdf's

i = index variable; $i = 1,2,\ldots, (m+n)$

m = number of data values in the observed empirical cdf, $E(Q_0)$

n = number of data values in the simulated empirical cdf, $E(Q_s)$.

The Kolmogorov-Smirnov two-sample test is shown in detail in Table 5.2 for the observed sequence of 21 runoff event volumes and the simulated sequence of 15 runoff event volumes for simulation year 36. Empirical cdf's for the test were developed using only significant runoff events with total runoff volumes of 0.01 inch or greater since these were the events with accurate observed runoff volume data. The total annual runoff volume for events of 0.01

TABLE 5.2

KOLMOGOROV-SMIRNOV TWO-SAMPLE TEST FOR OBSERVED AND SIMULATED RUNOFF VOLUMES

Row	Observed Empirical cdf		Simulated Empirical cdf		Deviations, Δ_i
	Q_0 (in.)	$E_i(Q_0)$	Q_s (in.)	$E_i(Q_s)$	$\lvert E_i(Q_0)-E_i(Q_s)\rvert$
i = 1	.011	.048	--	0	.048
i = 2	.014	.095	--	0	.095
i = 3	.014	.143	--	0	.143
i = 4	.015	.190	--	0	.190
i = 5	--	.190	.0169	.067	.123
i = 6	--	.190	.0172	.133	.057
i = 7	--	.190	.0175	.200	.010
i = 8	--	.190	.0180	.267	.077
i = 9	.023	.238	--	.267	.029
i = 10	.023	.286	--	.267	.019
i = 11	.028	.333	--	.267	.066
i = 12	.030	.381	--	.267	.114
i = 13	.031	.429	--	.267	.162
i = 14	--	.429	.0344	.333	.096
i = 15	.042	.476	--	.333	.143
i = 16	.044	.524	--	.333	.191
i = 17	.049	.571	--	.333	.238
i = 18	.052	.619	--	.333	.286
i = 19	--	.619	.0607	.400	.219
i = 20	.064	.667	--	.400	.267
i = 21	.077	.714	--	.400	.314
i = 22	--	.714	.0796	.467	.247
i = 23	.083	.762	--	.467	.295
i = 24	.085	.810	--	.467	.343
i = 25	.086	.857	--	.467	.390
i = 26	--	.857	.0979	.533	.324
i = 27	--	.857	.0983	.600	.257
i = 28	--	.857	.1047	.667	.190
i = 29	--	.857	.1271	.733	.124
i = 30	--	.857	.1288	.800	.057
i = 31	--	.857	.1488	.867	.010
i = 32	.161	.905	--	.867	.038
i = 33	.245	.952	--	.867	.085
i = 34	.264	1.000	--	.867	.133
i = 35	--	1.000	.2681	.933	.067
i = 36	--	1.000	.3172	1.000	0

TABLE 5.2 (Continued)

KOLMOGOROV-SMIRNOV TWO-SAMPLE TEST FOR OBSERVED AND SIMULATED RUNOFF VOLUMES

Null Hypothesis, H_0: The observed sequence of runoff event volumes and the simulated sequence of runoff event volumes for simulation year 36 were drawn from the same population.

Alternative Hypothesis, H_1: The observed and simulated sequences were not drawn from the same population.

Kolmogorov-Smirnov Test Statistic: $\Delta_{max} = 0.390$.

Critical Region at 0.05 Level of Significance for m = 21 and n = 15 (Two-Sided Test): $\Delta_0 > 0.486$.

Decision: Do Not Reject H_0 .

inch or greater for the observed year was 1.441 inches and for simulation year 36 was approximately 1.535 inches.

Notice in Table 5.2 that the empirical cdf's are step functions which have a maximum deviation of Δ_{max} = 0.390. The critical value of the test statistic, Δ_0 = 0.486, was obtained from standard statistical tables for a two-sided test with m=21 and n=15 (Harter and Owen 1970). The test statistic lies well outside the critical region of the test at the 0.05 level of significance. Therefore, do not reject the null hypothesis that the observed and simulated sequences of runoff event volumes were drawn from the same population.

It can, therefore, be concluded that the simulation model produces sequences of runoff event volumes that are consistent with field observations. In other words, the observed annual sequence of runoff event volumes is one possible realization that could have been produced by the random processes described by the simulation model.

The results of the analyses in this sub-section verify the applicability of the model for simulation of rainfall-runoff events for the study watershed.

Ironically, the rare nature of the particular year monitored during the field study emphasizes the inherent superiority of the stochastic simulation model developed here over deterministic models for rainfall-runoff and NPS pollutant load modeling. As discussed above, only 27 percent of annual runoff volumes are expected to be less than or equal to that of the observed year. Predictions of long-term annual runoff volumes and associated NPS pollutant loads would, therefore, be grossly underestimated if only the observed data were utilized for making deterministic estimates. On the other hand, it was just as likely that the observed year would be unusually wet which would result in gross overestimation of long-term magnitudes using deterministic methods. This emphasizes that the inherent stochastic nature of rainfall event characteristics

as the driving force behind generation of runoff events and NPS pollutant loads requires application of stochastic simulation methods for reliable estimation of long-term trends. Stochastic modeling provides not only estimates of numerical magnitudes for pertinent variables, but also probabilities of occurrence of these magnitudes which is vital for decision making in real world applications.

Evaluation of Data Base Sample Size

The second major objective of this work was to address the real-world problem of evaluating the minimum acceptable data base sample size for development of a stochastic simulation model of the rainfall-runoff process. Sample size, in turn, reflects the relative cost of collecting the field data needed for model development.

The 100-year sequence of simulation results can now be used as the given set of long-term data that is needed to evaluate the probabilities of successfully deriving models from samples of various sizes. That is, in this evaluation, the simulation results were treated as a long-term observed record of rainfall-runoff events for the study watershed.

Runoff volume, Q, was selected as the indicator variable for the analysis of sample size. The form of the linear model derived for runoff volume in Chapter 4 was assumed known:

$$Q = \beta_0+\beta_1(R_w)+\beta_2(T_r)+\beta_3({}_0R_6)+\varepsilon_v. \tag{5.3}$$

Using this form of model and the long-term simulated record, runoff volume models can be derived from various size samples of rainfall-runoff event data.

Sample sizes of 5, 10, 15, 19, 25, 30 and 40 were utilized. The unusual sample size of 19 was chosen because if corresponded to the actual size used

to develop the original simulation model. A minimum sample size of 5 is required to develop a linear model with four parameters as discussed in Chapter 4.

The long-term simulation results contain 1499 rainfall-runoff events with a runoff volume of 0.01 inch or more. These events were used for the sample size analysis because the original model was derived for runoff events of this magnitude.

The approach used was to fit the linear model of equation 5.3 to all possible subsets of the 1499 events for each chosen sample size, n. The total number of such subsets is given by,

$$N = (1499-n) + 1 \tag{5.4}$$

where,

n = sample size

N = total number of subsets of sample size n from a set of 1499 events.

For example, for samples of size 10, a total of 1490 different samples can be drawn from the 100-year record. The first sample of size 10 includes events 1 through 10, the second includes events 2 through 11, and so on until the 1490th sample of size 10 includes events 1490 through 1499.

The linear model of equation 5.3 was fit to the various samples of size n using the method of least squares. A SAS computer program was written to carry out the necessary computations using as the input file the simulation model output file which was previously described in detail. The program listing is shown in Table C.4 in Appendix C. Cramer's Rule was used to solve the four normal equations simultaneously for the parameters β_0, β_1, β_2, and β_3 for each linear model.

Once the value of the coefficients was determined for a specific sample of size n, the mean square error was then computed and used as the goodness of fit parameter for comparisons between the derived models. For each derived model, the mean square error was computed from the difference between the observed values and model predicated values for all 1499 events of the long-term record. That is, for this analysis, the mean square error is defined as,

$$s_\varepsilon^2 = \frac{\sum_{i=1}^{1499}\left(Q_i - \hat{Q}_i\right)^2}{(1499 - 4)} \qquad (5.5)$$

where,

s_ε^2 = mean square error

Q_i = observed runoff volume for event i of the long-term simulated record

$\hat{Q}_i$ = model predicted runoff volume for event i.

Of course, the model predicted values of runoff volume, $\hat{Q}_i$, were computed using,

$$\hat{Q} = \beta_0+\beta_1(R_w)+\beta_2(T_r)+\beta_3({}_0R_6) \qquad (5.6)$$

where all coefficients and variables are as defined in Chapter 4. The denominator (1499-4) was used because four parameters were estimated from the data.

The SAS program steps through the entire input file fitting the linear model and computing s_ε^2 for the first set of data of sample size n, then the second, and

so on until all N possible subsets of sample size n have been analyzed. The program then creates an output file containing the square root of s_ε^2, which is the sample standard deviation of the errors, for each model derived from samples of size n. Thus, there was a permanent, machine language SAS file created on disk containing the [(1499-n)+1] values of s_ε for each of the seven sample sizes evaluated. These files can be readily accessed by easily constructed programs using SAS procedures for detailed statistical analyses.

Simple statistics were computed and a frequency analysis was performed for s_ε values using SAS procedures for samples of size 5, 10, 15, 19, 25, 30, and 40 as summarized in Table 5.3. As expected, the value of s_ε decreases (i.e. the goodness of fit improves) as the sample size increases. For example, the median value of s_ε was 0.06747 inch for samples of size 5 while the median value was only 0.03181 inch for samples of size 40.

The results are summarized in a more readily understandable graphical manner using the results of the frequency analysis as shown in Figure 5.3. A family of curves are presented for various probabilities of success (i.e. non-exceedence probabilities or percentile values for s_ε) with s_ε shown on the vertical axis and sample size shown on the horizontal axis. Curves for probabilities of success in obtaining given values of s_ε or lower are shown for probabilities of 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, and 0.95. All curves asymptotically approach a value of 0.03032 inch, which was the sample standard deviation of the white noise component for the 100-year simulation record. This value represents the estimate of the population standard deviation of the model residuals, $\hat{\sigma}_\varepsilon$. The practical implications of this asymptotic behavior will be discussed in more detail later.

TABLE 5.3

SAMPLE STANDARD DEVIATION OF ERRORS FOR RUNOFF VOLUME MODELS DERIVED FROM RANDOM SAMPLES OF VARIOUS SIZES

Statistic	Sample Size 5	10	15	19	25	30	40
Number of Observations, n	1495	1490	1485	1481	1475	1470	1460
Mean, $\bar{x}$	.11694	.04396	.03752	.03536	.03370	.03285	.03208
Standard Deviation, s_x	.17604	.01872	.00972	.00667	.00369	.00223	.00131
Minimum Value, x_{min}	.03059	.03042	.03038	.03043	.03033	.03032	.03033
5 Percentile Value, x_5	.03540	.03198	.03112	.03093	.03069	.03064	.03062
10 Percentile Value, x_{10}	.03778	.03289	.03156	.03136	.03102	.03084	.03074
25 Percentile Value, x_{25}	.04529	.03482	.03277	.03217	.03176	.03149	.03115
50 Percentile Value, x_{50}	.06747	.03849	.03493	.03384	.03278	.03227	.03181
75 Percentile Value, x_{75}	.12160	.04578	.03854	.03625	.03443	.03349	.03265
90 Percentile Value, x_{90}	.22463	.05833	.04510	.04054	.03717	.03526	.03365
95 Percentile Value, x_{95}	.34024	.07302	.04952	.04346	.03984	.03690	.03458
Maximum Value, x_{max}	3.59268	.25613	.14015	.12376	.08017	.05232	.03929

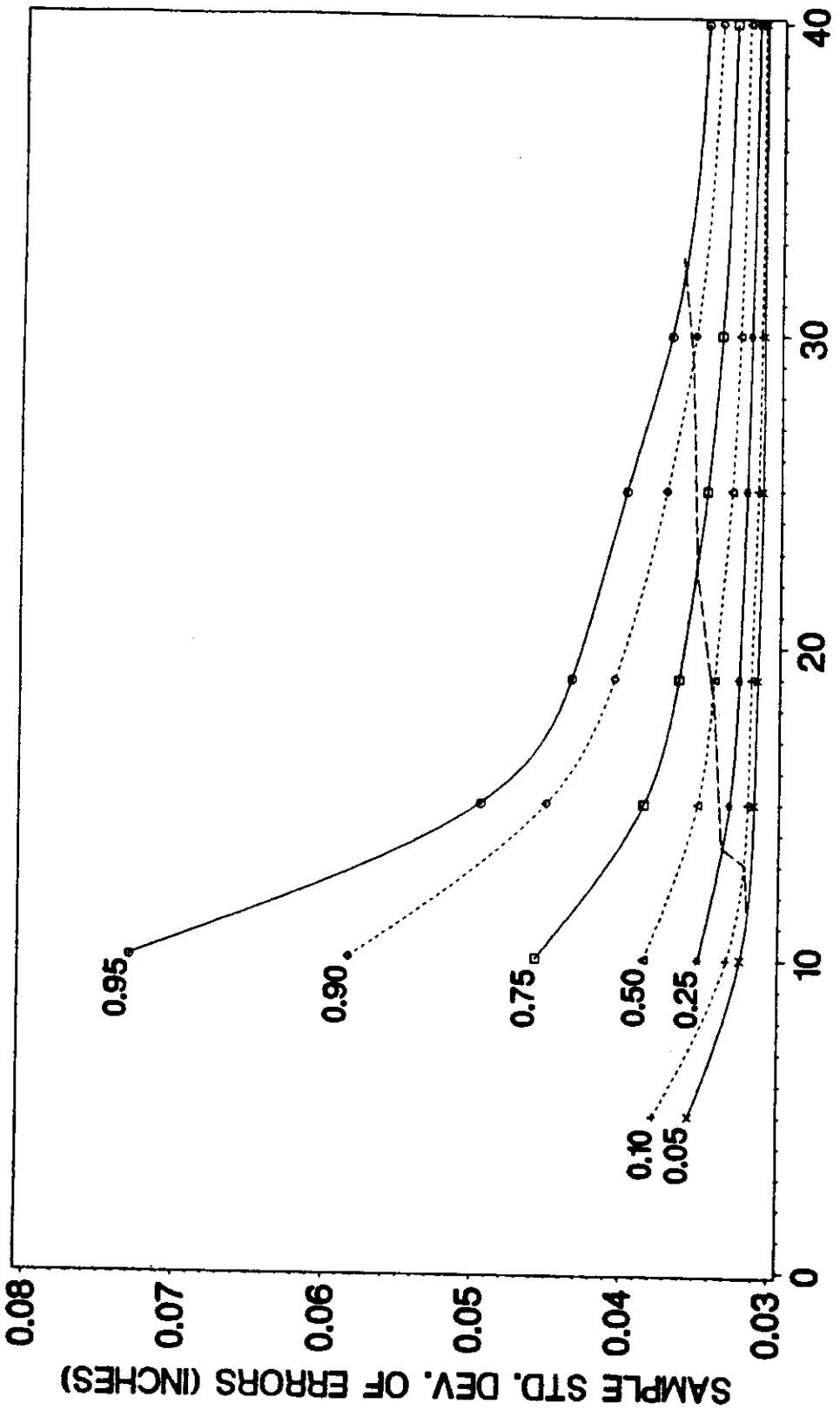


FIGURE 5.3
SAMPLE STANDARD DEVIATION OF ERRORS

For now, Figure 5.3 allows statements about the probability of obtaining runoff volume models with certain s_ε values for samples of various sizes. As a preliminary example, if it was desired to develop a linear model of runoff volume using the methdology developed in Chapter 4, and the standard deviation of the white noise component is to be no more than 1.25 times the actual value (i.e., $s_\varepsilon \leq 1.25\ (0.03032) = .03790$), then the approximate number of samples that must be collected for selected probabilities of success in obtaining the target value of s_ε or lower as estimated from Figure 5.3 are:

Probability of Success	Risk of Failure	Sample Size
0.10	0.90	5
0.25	0.75	8
0.50	0.50	11
0.75	0.25	17
0.90	0.10	23
0.95	0.05	28

Thus, if a field study were planned to collect the data to compute the necessary runoff volume model parameters, and at least a 90 percent chance were wanted that the standard deviation of the white noise component would be no more than 1.25 times the actual value, then a schedule and budget should be allotted to sample 23 or more rainfall-runoff events.

The cost to perform such a field study would be approximately directly proportional to the length of time required to obtain a sample of the desired size. To provide a means of estimating the length of time required, the long-term simulation results were used to develop empirical cdf's for the time to collect samples of various sizes. This was done by computing the length of time

between rainfall-runoff events with a runoff volume of 0.01 inch or greater for all possible samples of the desired size n. Again, the number of such samples can be obtained from equation 5.4. Empirical cdf's for the time to collect samples of sizes 5, 10, 20, 30, and 40 are shown graphically in Figure 5.4.

For example, if a set of data with 20 samples is to be collected, the median length of time a sampling program would need to last would be approximately 10,500 hours, which is 438 days or 1.2 years. If a field study to collect these 20 samples were to be completed with a specified 90 percent chance of being within the allotted budget, the length of time to be budgeted would be about 19,500 hours, which is 813 days or 2.2 years.

Practically speaking, the asymptotic behavior of the curves in Figure 5.3 as the sample increases clearly emphasizes that a sort of "law of diminishing returns" applies. Very little improvement in accuracy results from relatively large increases in the sample size (and cost) when the sample size is already large. A practical criterion is needed to decide at what sample size and at what corresponding level of accuracy the diminishing improvement in accuracy no longer warrants collecting additional samples.

The criterion recommended here involves the accuracy of the rainfall measuring instrument since rainfall volume was found to be the single most important independent variable in the rainfall-runoff transformation process. The best rain gages are accurate to 0.01 inch of rainfall. The coefficient of rainfall volume in the runoff volume transformation model, equation 4.27 in Chapter 4, is 0.1671 which means that the accuracy of the rain gage when transformed to runoff volume is 0.001671 inch. Now, since a minimum of 5 samples is needed to develop a runoff volume model of the form of equation 4.27 with 4 parameters, and since the median time to collect 5 samples is approximately a calendar quarter (actually 82 days from Figure 5.4) which is a

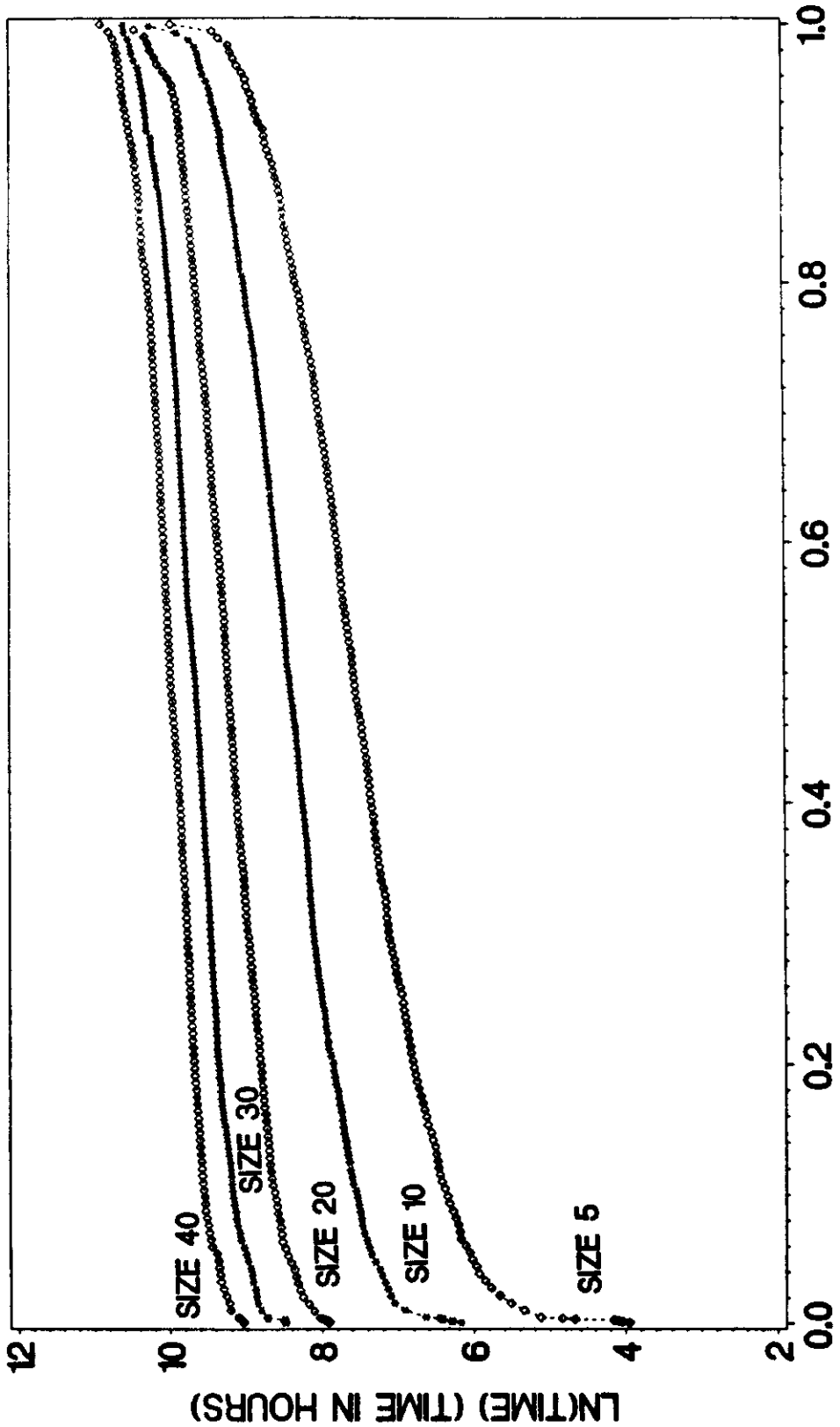


FIGURE 5.4
CDF'S FOR TIME TO COLLECT DATA SETS

common budgeting period, a rational approach would be to plan and budget a field study to collect samples based on increments of 5. Therefore, it is recommended that field studies not be continued beyond the point at which collecting 5 more samples would not be expected to reduce s_ε by 0.001671 inch. In other words, a field study should be terminated when the slope of the selected curve in Figure 5.3 (selected based on the acceptable level of risk) is approximately equal to -0.0003342 inch per sample (i.e., -0.001671/5).

These points were computed for the seven risk curves of Figure 5.3 by using quadratic interpolation of selected points from each curve. A quadratic polynomial, $y = a+bx+cx^2$, was fit to the points on each curve in the neighborhood of the critical slope value. The critical point on each curve was then obtained by differentiating the polynomial to obtain the slope function, setting the slope function equal to -0.0003342 inch per sample, and solving for the critical value of the sample size; $x_c = (-0.0003342-b)/2c$. Results were as follows:

Probability of Success	Risk of Failure	Critical Point	
		Sample Size	s_ε(inch)
0.95	0.05	32.53	.03617
0.90	0.10	29.29	.03548
0.75	0.25	22.37	.03507
0.50	0.50	18.19	.03385
0.25	0.75	13.64	.03332
0.10	0.90	13.06	.03168
0.05	0.95	11.60	.03159

These critical points were plotted on Figure 5.3 and a smooth curve was drawn through them as shown.

At a level of acceptable risk of 0.10 (i.e. 0.90 probability of success), the field study should be planned to collect 30 samples since the critical sample

size was 29.29. Collecting an additional 5 samples would not reduce s_ε by the transformed accuracy of the rain gage, 0.001671 inch of runoff. Hence, collection of additional samples would not be economically justified. With 30 samples, s_ε would be less than or equal to 0.03526 inch (from Table 5.3) or 1.16 times $\hat{\sigma}_\varepsilon$ with a probability of 0.90. Using Figure 5.4, and again choosing an acceptable risk of 0.10 (i.e. the 90 percentile value), the time to collect 30 samples should be less than or equal to 27,500 hours, or 3.1 years. Therefore, the field study to collect 30 samples should be scheduled and budgeted for 3 or 3-1/4 years.

CHAPTER 6
CONCLUSIONS AND RECOMMENDATIONS

Conclusions

A new and innovative field monitoring station was designed and constructed to continuously and automatically measure and record in computer memory rainfall, runoff, and water quality data. Data were collected over a period of 13 months from June 1987 to June 1988. During this time, 45 rainfall-runoff events occurred. Of these, 19 events were of significant runoff volume ($Q \geq 0.01$ inch) with a complete and detailed set of rainfall-runoff data.

Long-term historical hourly rainfall data from the FAA rain gage at Love Field Airport in Dallas, Texas were used to establish pertinent rainfall event statistical characteristics. To make fullest use of this massive amount of data in a practical and manageable way, a computer program was developed to accept input hourly rainfall data using a format compatible with readily available diskette data files from the National Climatic Data Center (NCDC). The program reads the NCDC diskette files; resolves the hourly rainfall totals into individual rainfall events; and then stores the resulting data on rainfall event volume, event duration, and time between events to a disk file. The program was written in the BASIC language for use on personal computers so that it could be easily applied by practicing engineers. Event data for Love Field were used to derive regionally applicable marginal pdf's for event volume, duration, and time between events.

A new stepwise variable selection and parameter optimization methodology was developed for derivation of linear transformation functions for the rainfall-runoff process from field data. This innovative approach is practical and readily applied to real-world engineering problems. Using this methodology, transformation functions, with white noise components, were derived for runoff volume and peak runoff rate as functions of rainfall event variables:

$$(6.1) \qquad Q = -.05017+.1671(R_w)-.0001263(T_r)+.02976({}_0R_6)+\varepsilon_v$$

where,

- Q = runoff volume (inches)
- R_w = rainfall volume (inches)
- T_r = time since the last runoff event with a volume of 0.01 inch or greater (hours)
- ${}_0R_6$ = total rainfall volume in the six days antecedent to the current event (inches)
- ε_v = normally distributed white noise component with $\mu = 0$ and $\sigma = .03148$ inch (inches)

and,

$$(6.2) \qquad q_p = -4.6053+12.3698(R_w)-.02144(T_l)+32.4367(I_a)+\varepsilon_q$$

where,

- q_p = peak runoff rate (cfs)
- T_l = time since the last runoff event (hours)
- I_a = average rainfall intensity, R_w/D (inches/hour)

ε_q = normally distributed white noise component with $\mu = 0$ and $\sigma = 4.5806$ cfs

Also, a new modification of the standard U.S. SCS dimensionless hydrograph procedure was developed. The modification involves the use of a random variable describing hydrograph shape rather than using a constant shape as in the standard method. This random shape variable, in conjunction with known or simulated values of Q and q_p, can be used to generate complete runoff hydrographs for rainfall-runoff events.

A stochastic simulation model was developed for generation of synthetic sequences of rainfall-runoff event data. A Monte Carlo simulation technique was used to randomly sample pdf's for rainfall volume, time between rainfall events, event duration, and white noise components of transformation functions. Various tracking and summation routines were then used to compute values of variables that were dependent upon the time sequencing of rainfall-runoff events, such as, time between runoff events and antecedent rainfall volumes. The simulation model then used the derived rainfall-runoff transformation functions for Q and q_p to create a disk file of long-term synthetic rainfall-runoff event data. In order to make the simulation model as widely applicable as possible in engineering practice, the model was developed in the form of a SAS computer program. The SAS statistical software package is readily available to most practicing engineers.

The simulation model was used to generate a 100-year sequence of rainfall-runoff data for the study watershed. Results of the simulation model were verified using statistical tests for comparing observed field data to simulated data. Ironically, successful verification of the model under the unusually dry conditions of the particular year monitored during the field study emphasizes the inherent superiority of the stochastic simulation model

developed here over deterministic models for rainfall-runoff and NPS pollutant modeling. Predictions of long-term annual runoff volumes and associated NPS pollutant loads would be grossly underestimated if only the field data from this unusually dry year were utilized for making deterministic estimates. The inherent stochastic nature of rainfall event characteristics as the driving force behind generation of runoff events and NPS pollutant loads requires application of stochastic simulation methods for reliable estimation of long-term trends. Stochastic modeling provides not only estimates of numerical magnitudes for pertinent variables but also probabilities of occurrence of these magnitudes which is vital for decision-making in real world applications.

Simulation results were then used to satisfy the first of the two major objectives of this study. Empirical cdf's were developed for runoff volume and peak runoff rate for the study watershed. These cdf's can now be applied in on-going research to establish the statistical characteristics and assess the impacts of NPS pollutant loads within the geographical region of the watershed.

An evaluation was performed of the minimum acceptable field data base sample size required for application of the practical stochastic simulation methodology developed here in order to achieve the second major objective of this study. Based on this evaluation, and using only a 10 percent risk of failure, it is recommended that a field monitoring program be scheduled and budgeted for 3 to 3-1/4 years in order to collect 30 samples. Additional sampling would not be economically justified for the resulting moderate improvement in model accuracy. Accepting higher levels of risk would, of course, allow planning for shorter and less costly field monitoring programs.

Recommendations for Further Research

Recommendations for further research include:

(1) Continue with work to develop a stochastic NPS water quality simulation model to couple with the rainfall-runoff model developed here. NPS modeling should allow for development of individual event loads and for the accumulation of event loads to generate annual load estimates. Stochastic modeling of NPS loads provides not only numerical magnitudes but also probabilities of occurrence for these loads which is vital for proper allocation of scarce public funds for NPS pollution control.

(2) Modify and extend the methodology developed here for application to large watersheds. As a minimum, this will require development of a method to account for spatial variability in event rainfall amounts.

(3) The fact that the hydrographs for the watershed possessed a random shape parameter with a well defined pdf suggests the potential for a probabilitistic based method to address hydrograph variability for a given watershed.

APPENDIX A

FIELD DATA

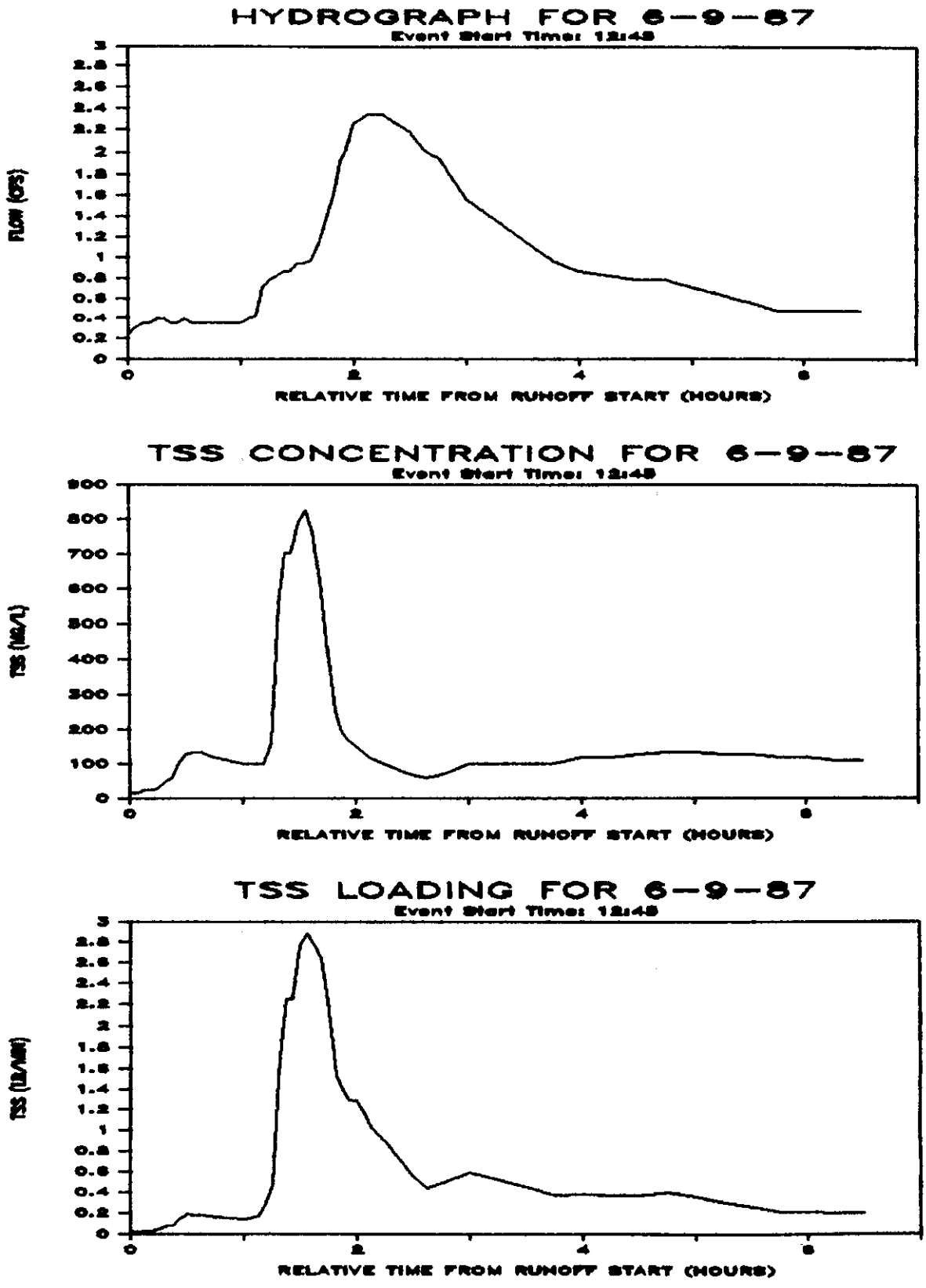


FIGURE A.1
FIELD DATA FOR EVENT OF 6-9-87

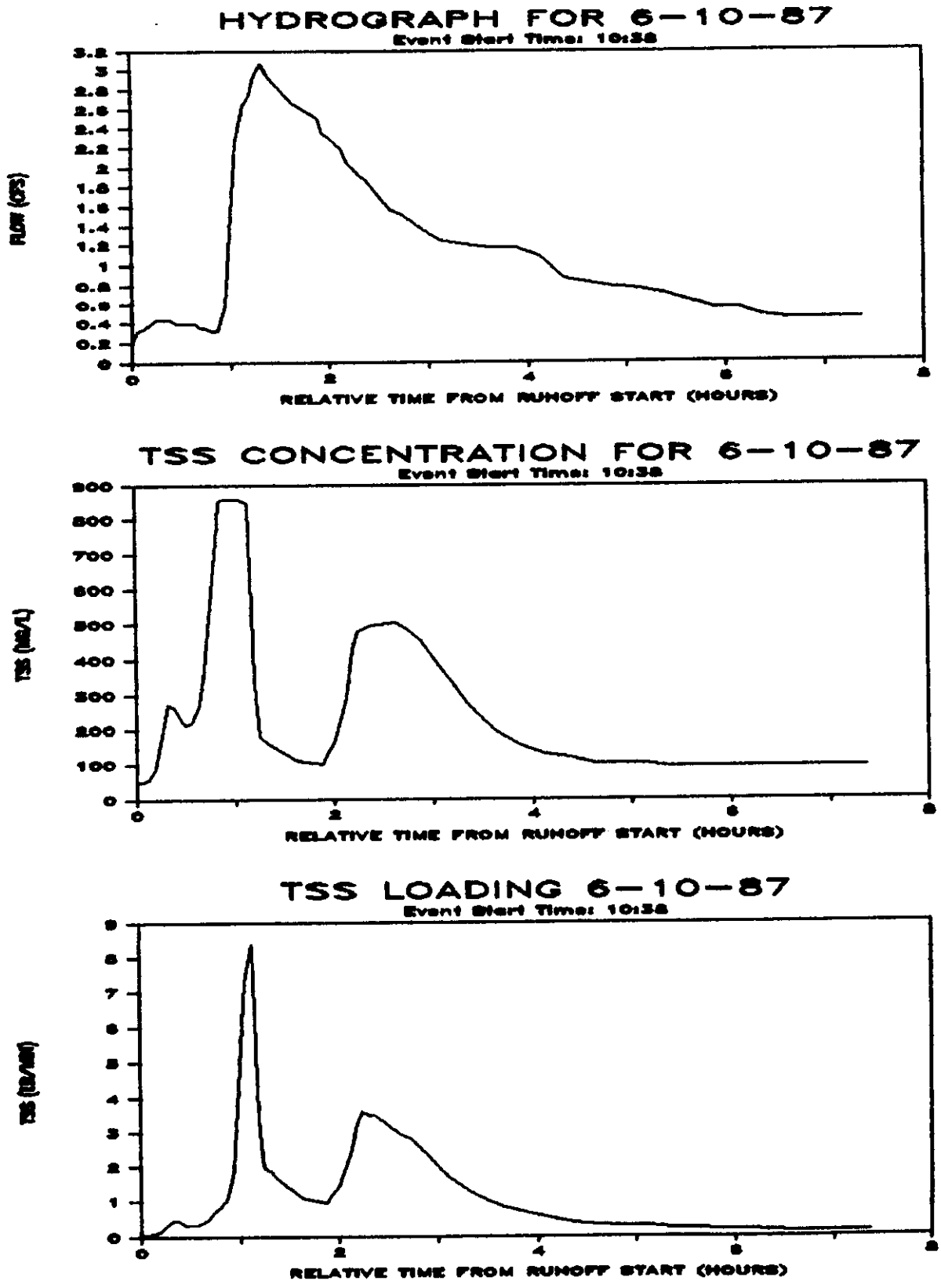


FIGURE A.2
FIELD DATA FOR EVENT OF 6-10-87

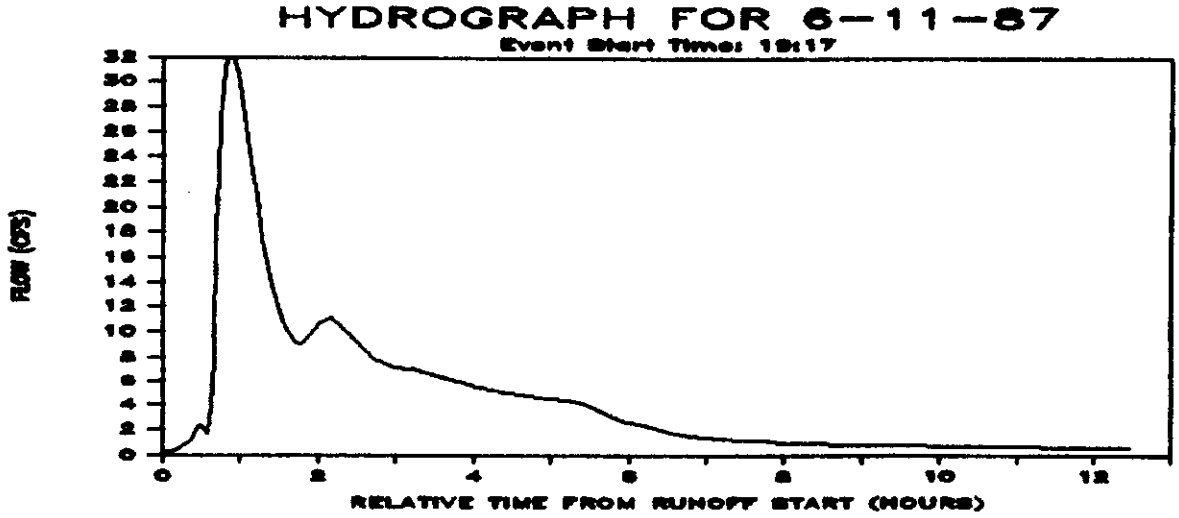


FIGURE A.3
FIELD DATA FOR EVENT OF 6-11-87

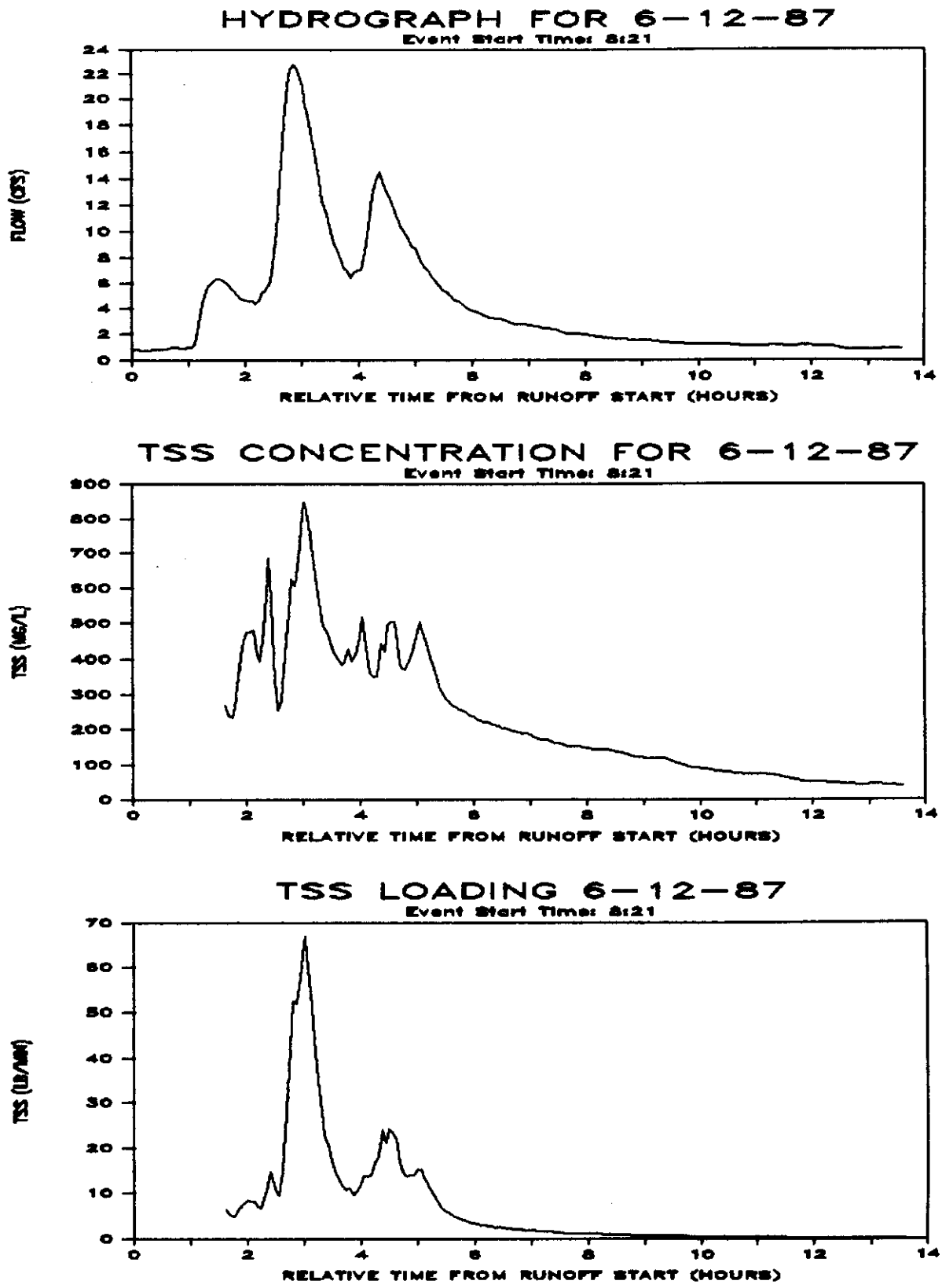


FIGURE A.4
FIELD DATA FOR EVENT OF 6-12-87

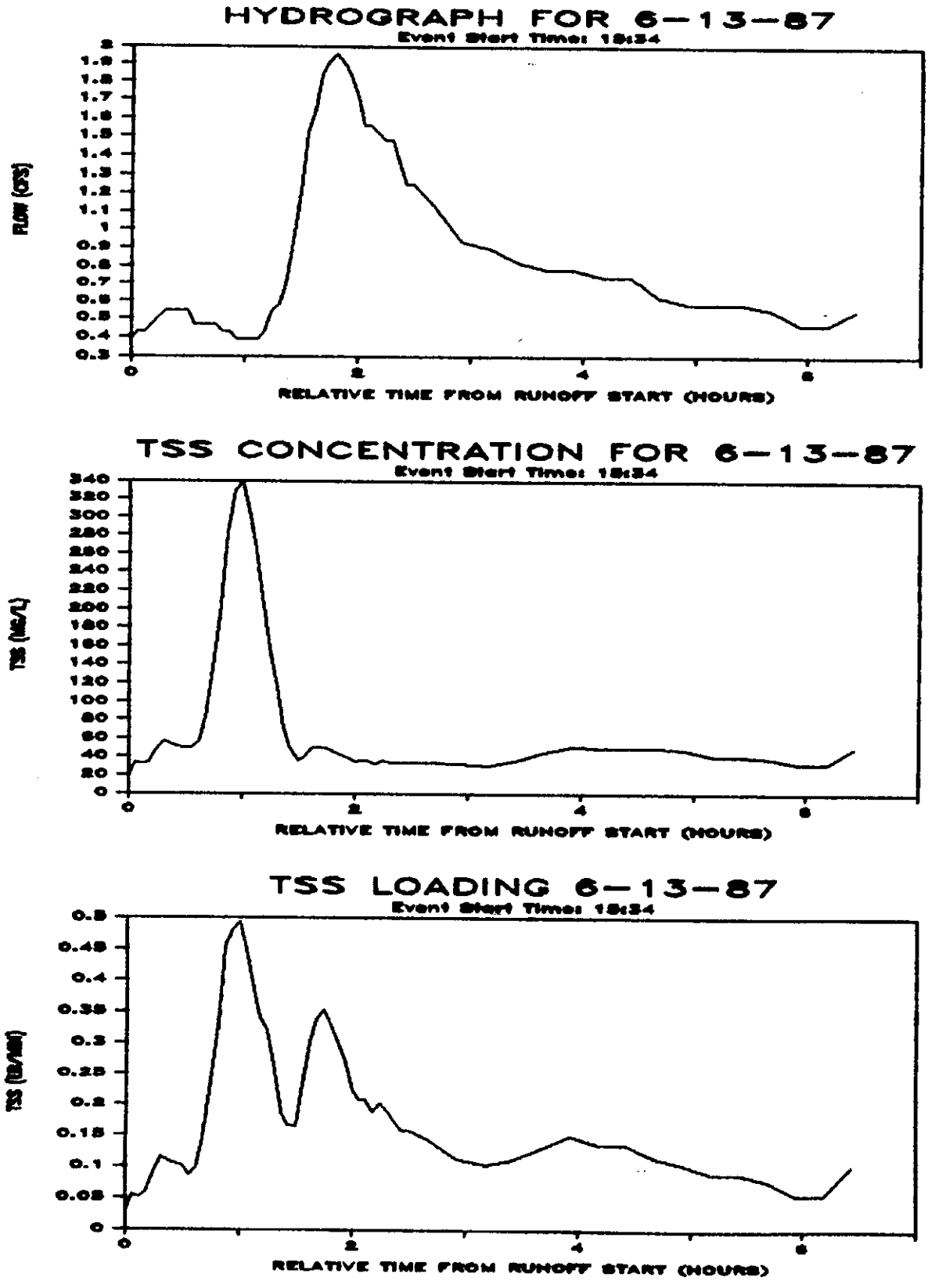


FIGURE A.5
FIELD DATA FOR EVENT OF 6-13-87

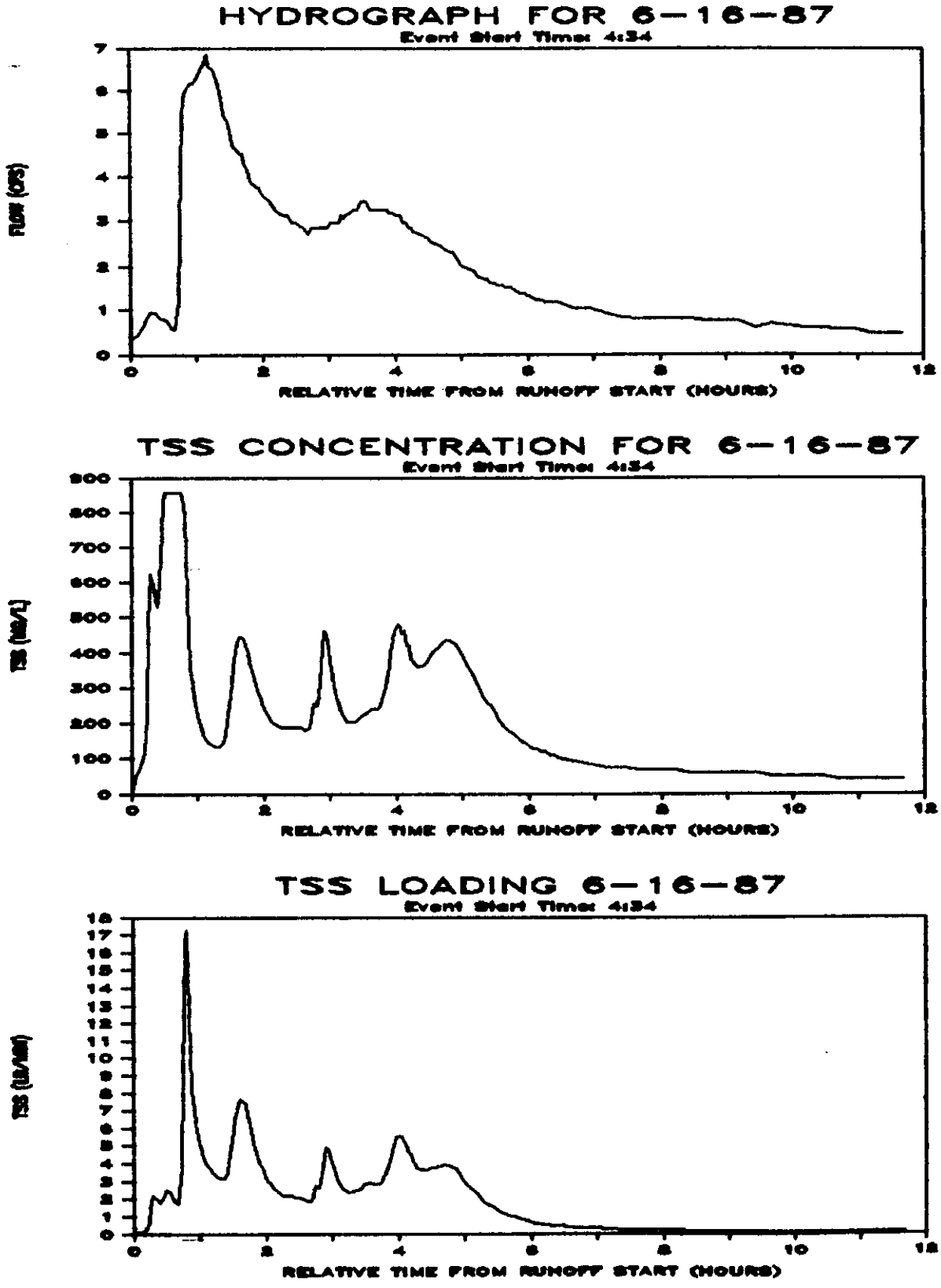


FIGURE A.6
FIELD DATA FOR EVENT OF 6-16-87

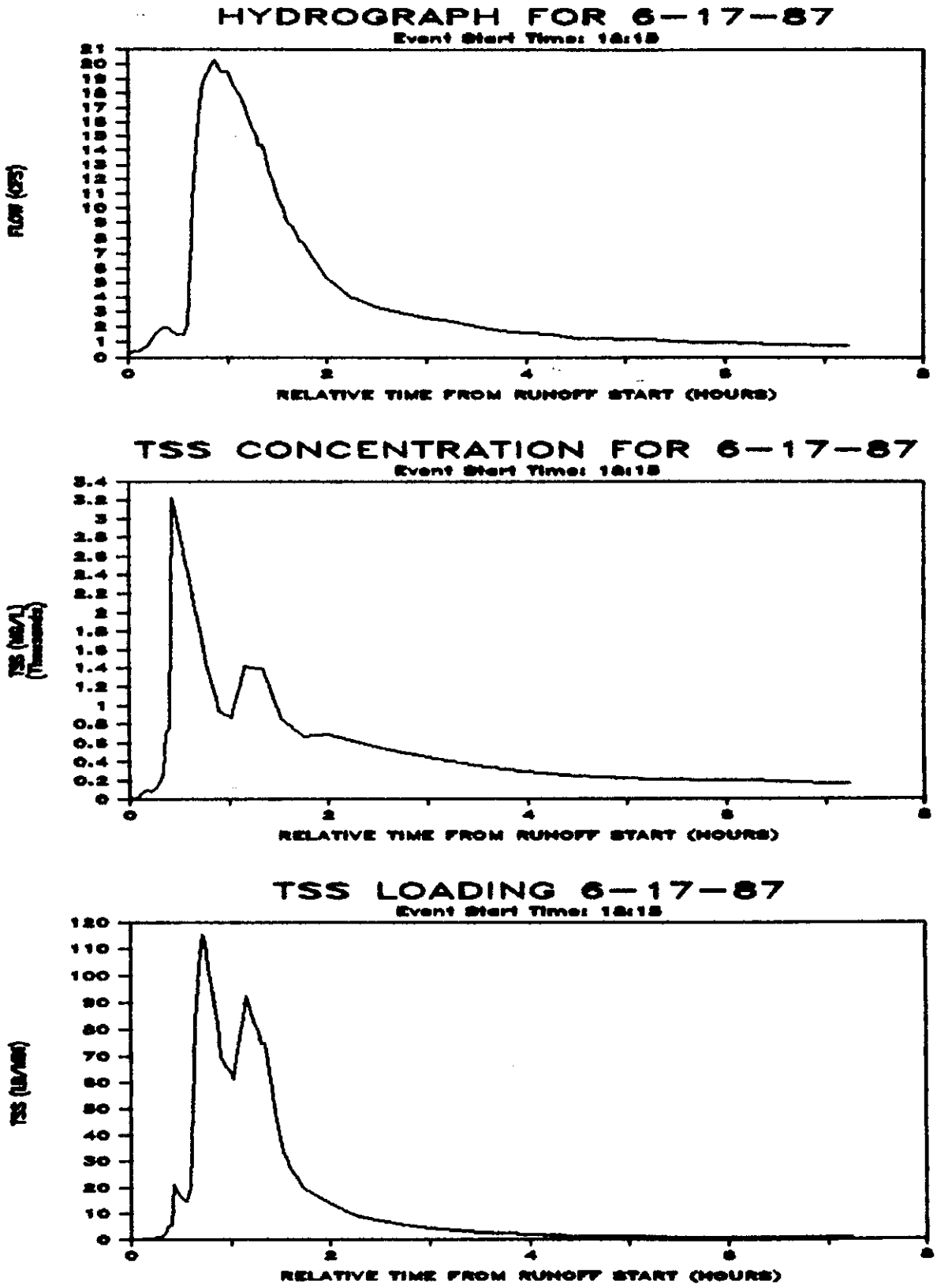


FIGURE A.7
FIELD DATA FOR EVENT OF 6-17-87

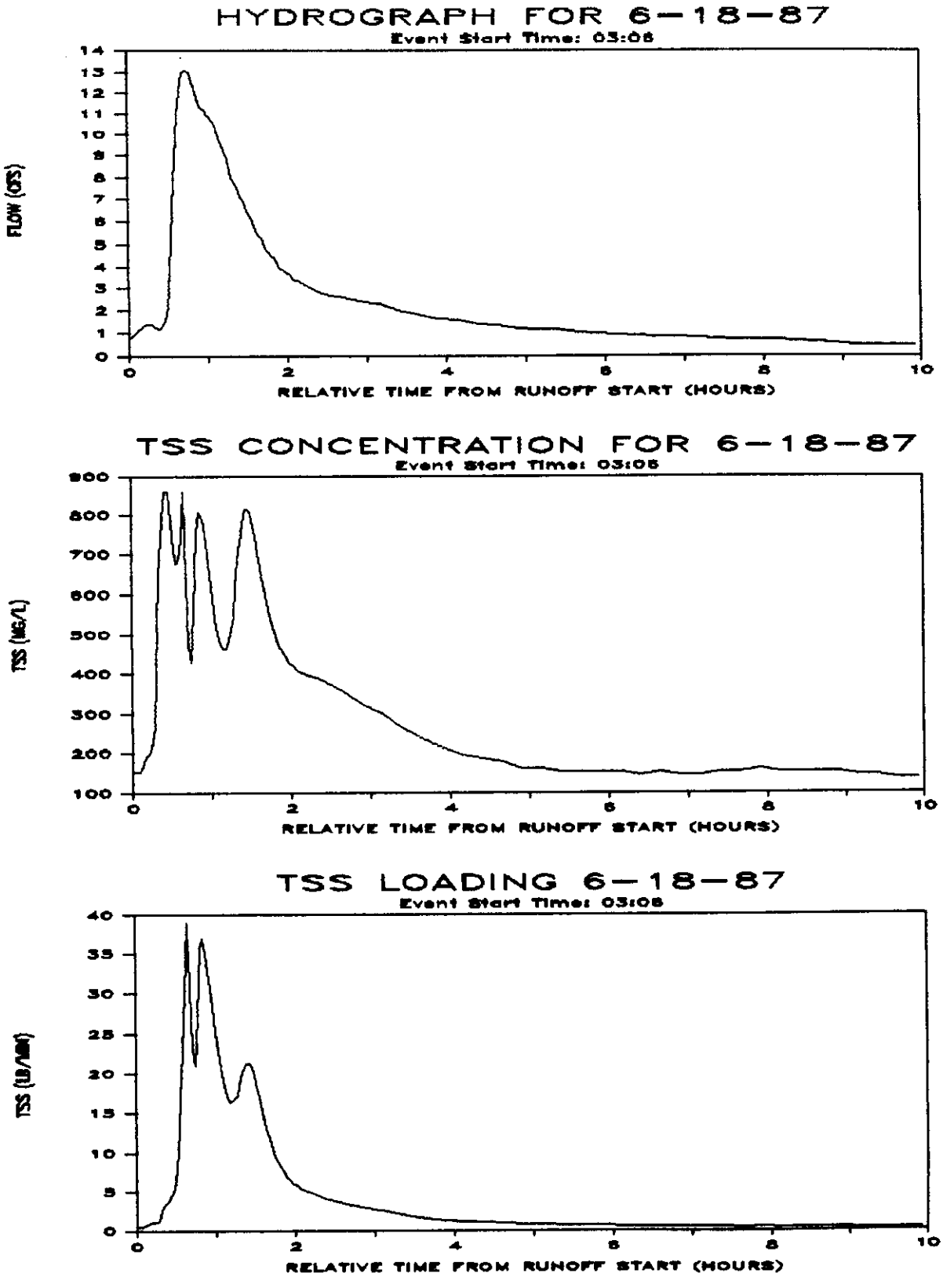
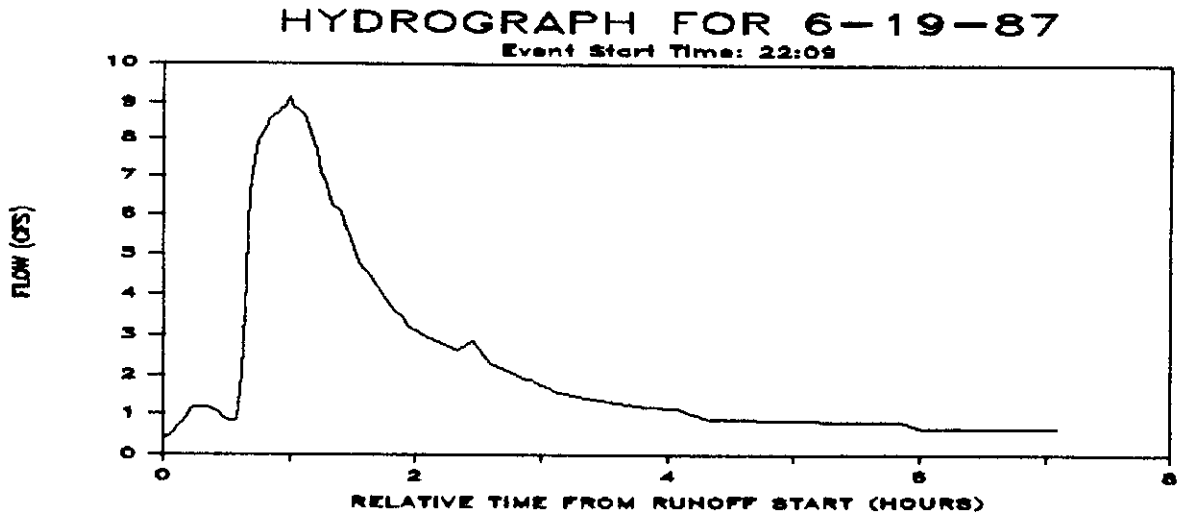
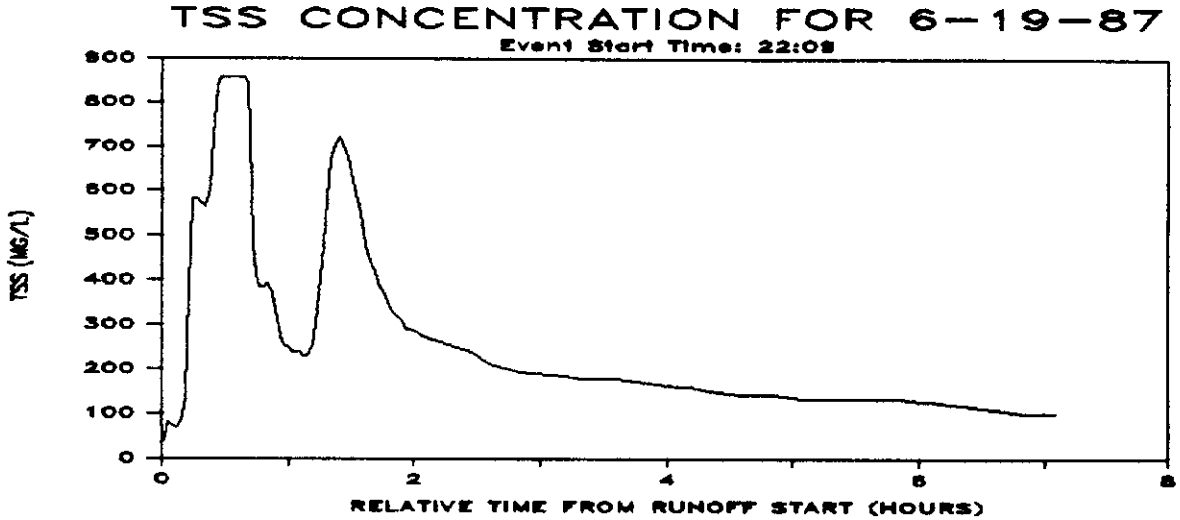


FIGURE A.8
FIELD DATA FOR EVENT OF 6-18-87





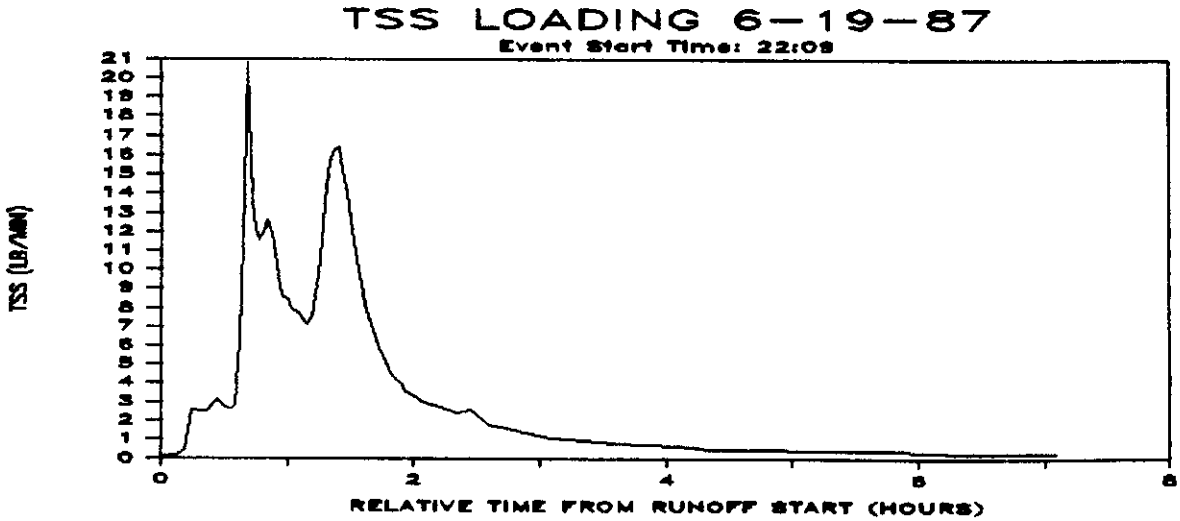


FIGURE A.9
FIELD DATA FOR EVENT OF 6-19-87

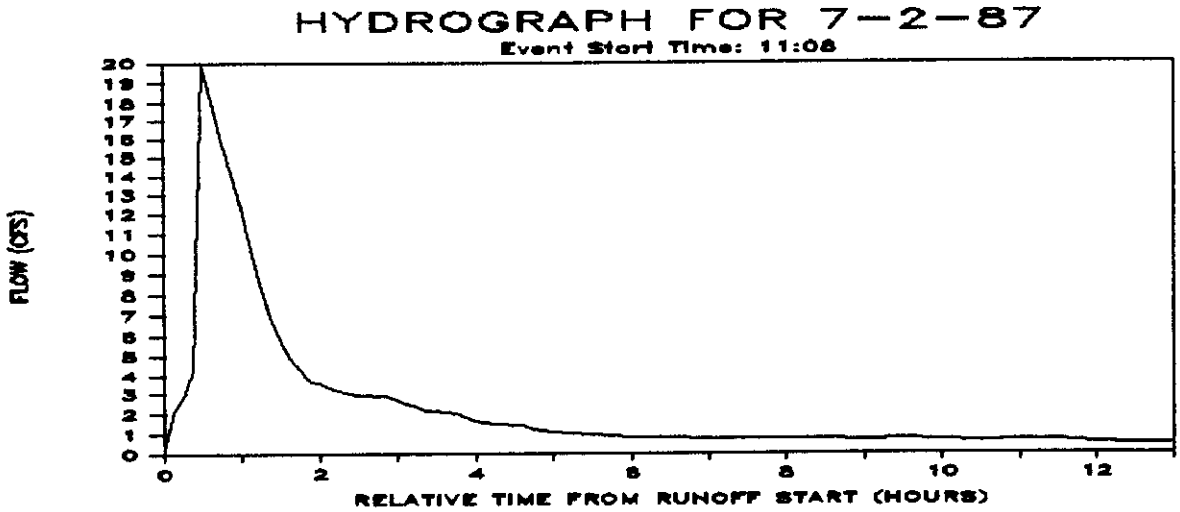


FIGURE A.10
FIELD DATA FOR EVENT OF 7-2-87

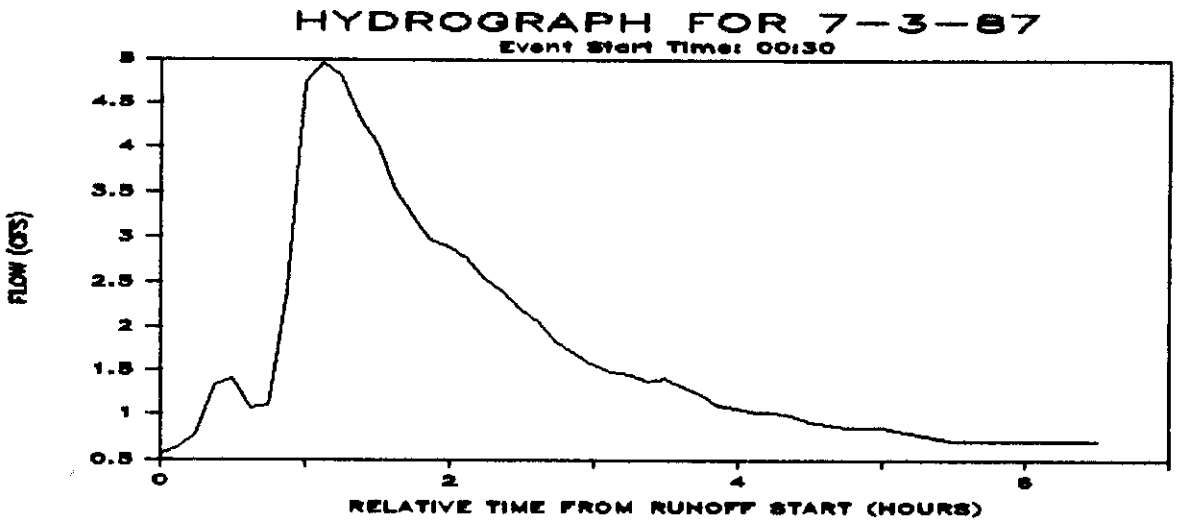


FIGURE A.11
FIELD DATA FOR EVENT OF 7-3-87

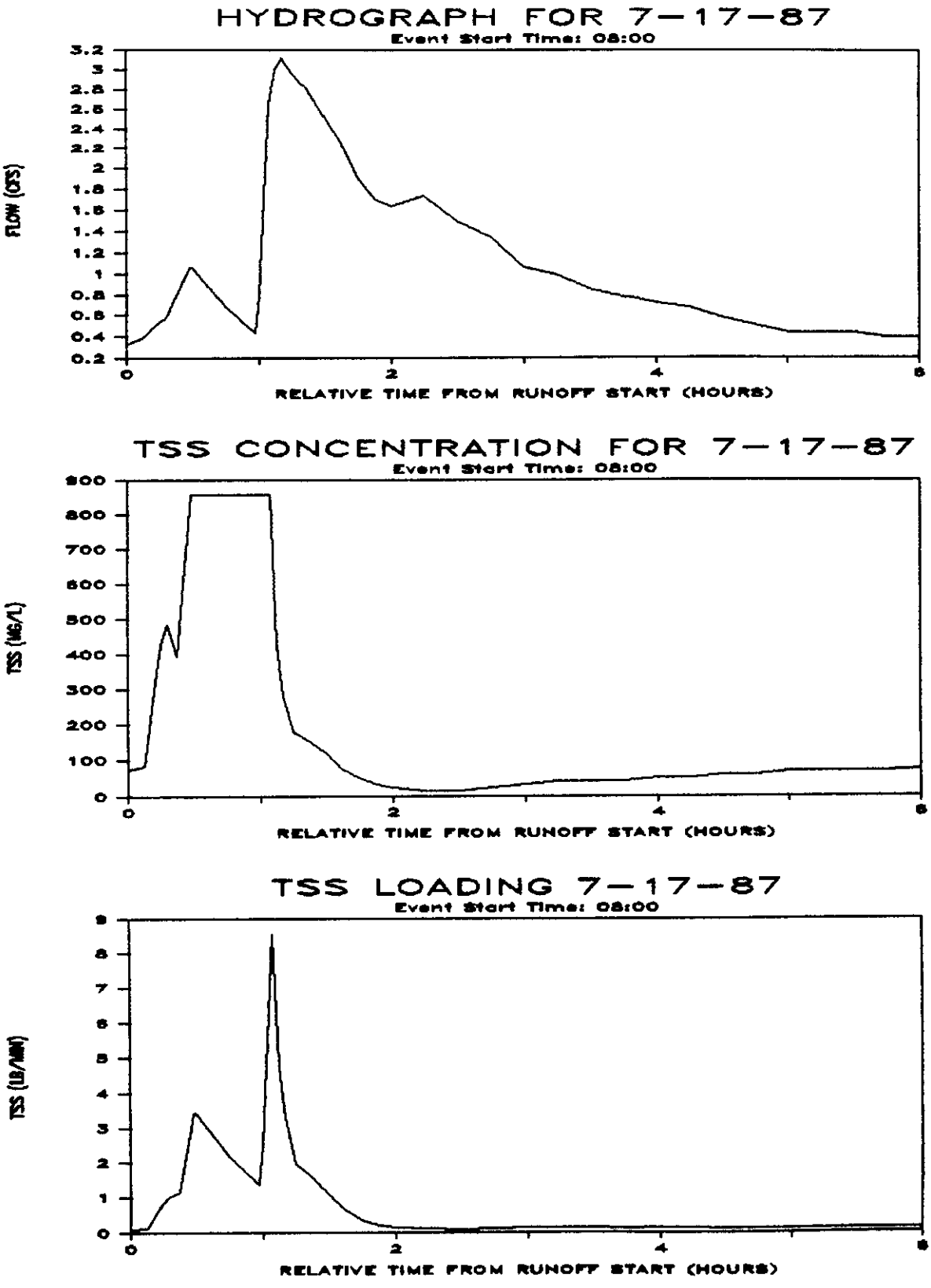


FIGURE A.12
FIELD DATA FOR EVENT OF 7-17-87

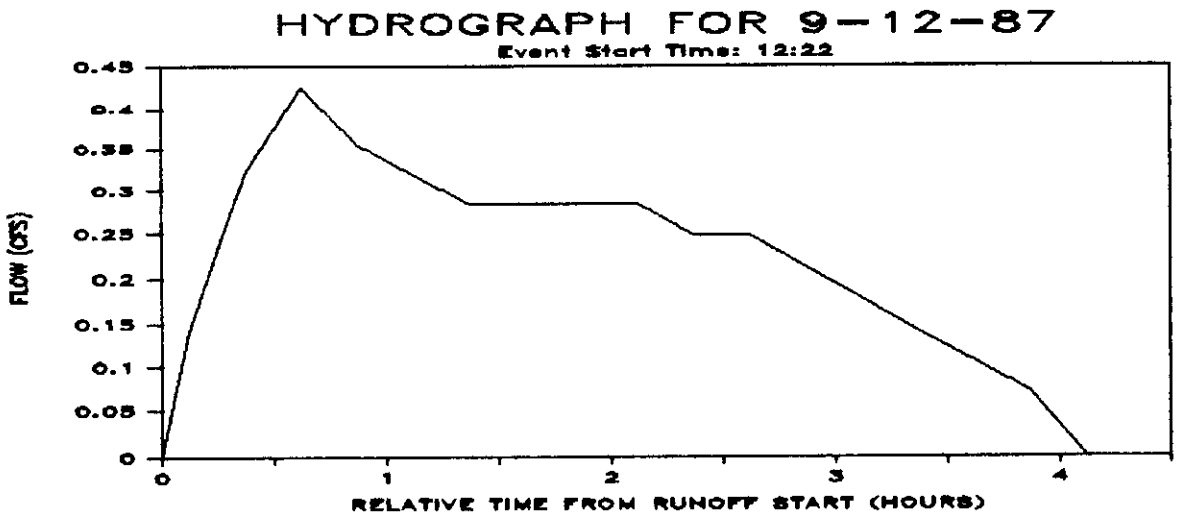


FIGURE A.13
FIELD DATA FOR EVENT OF 9-12-87

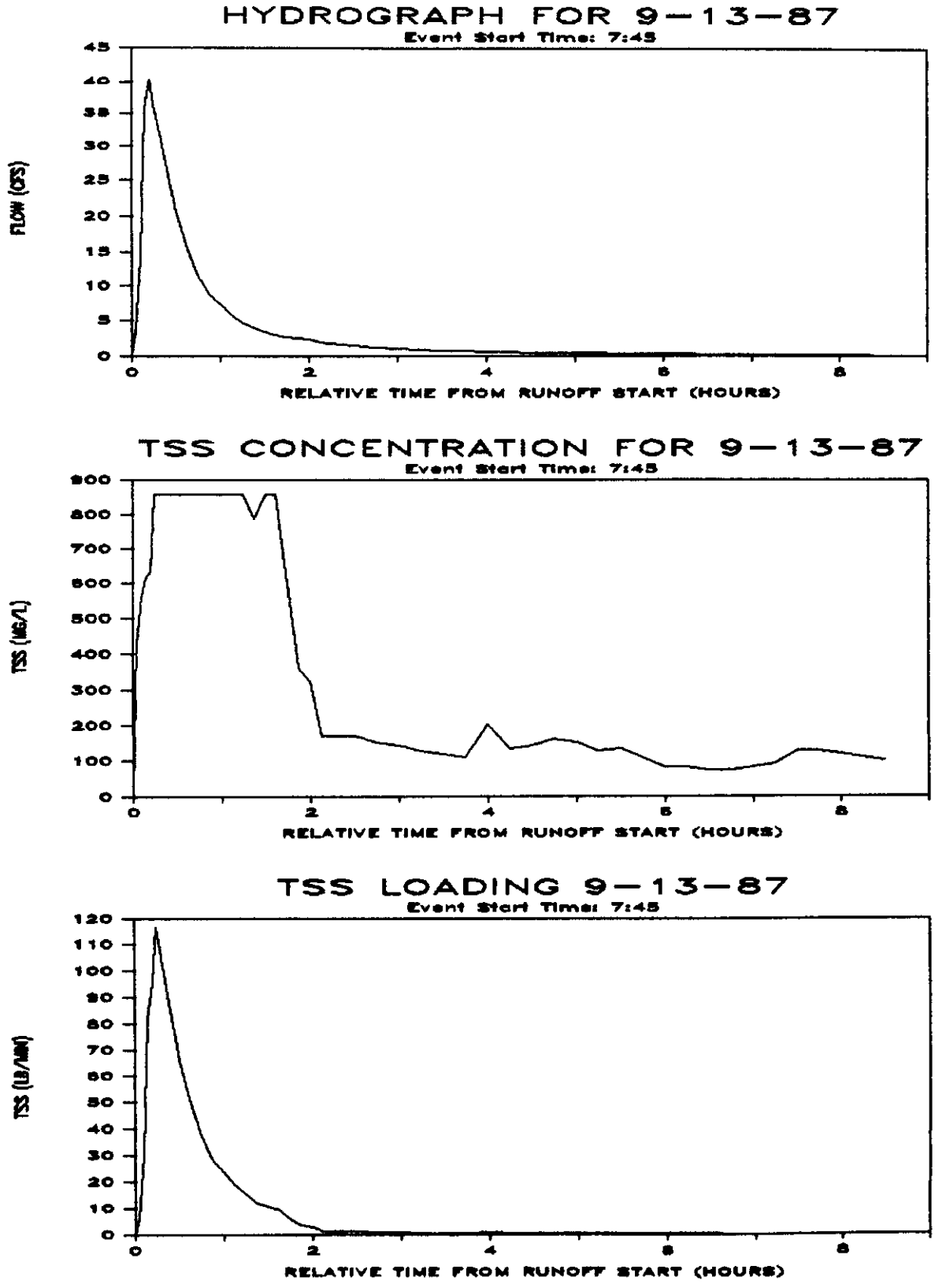


FIGURE A.14
FIELD DATA FOR EVENT OF 9-13-87

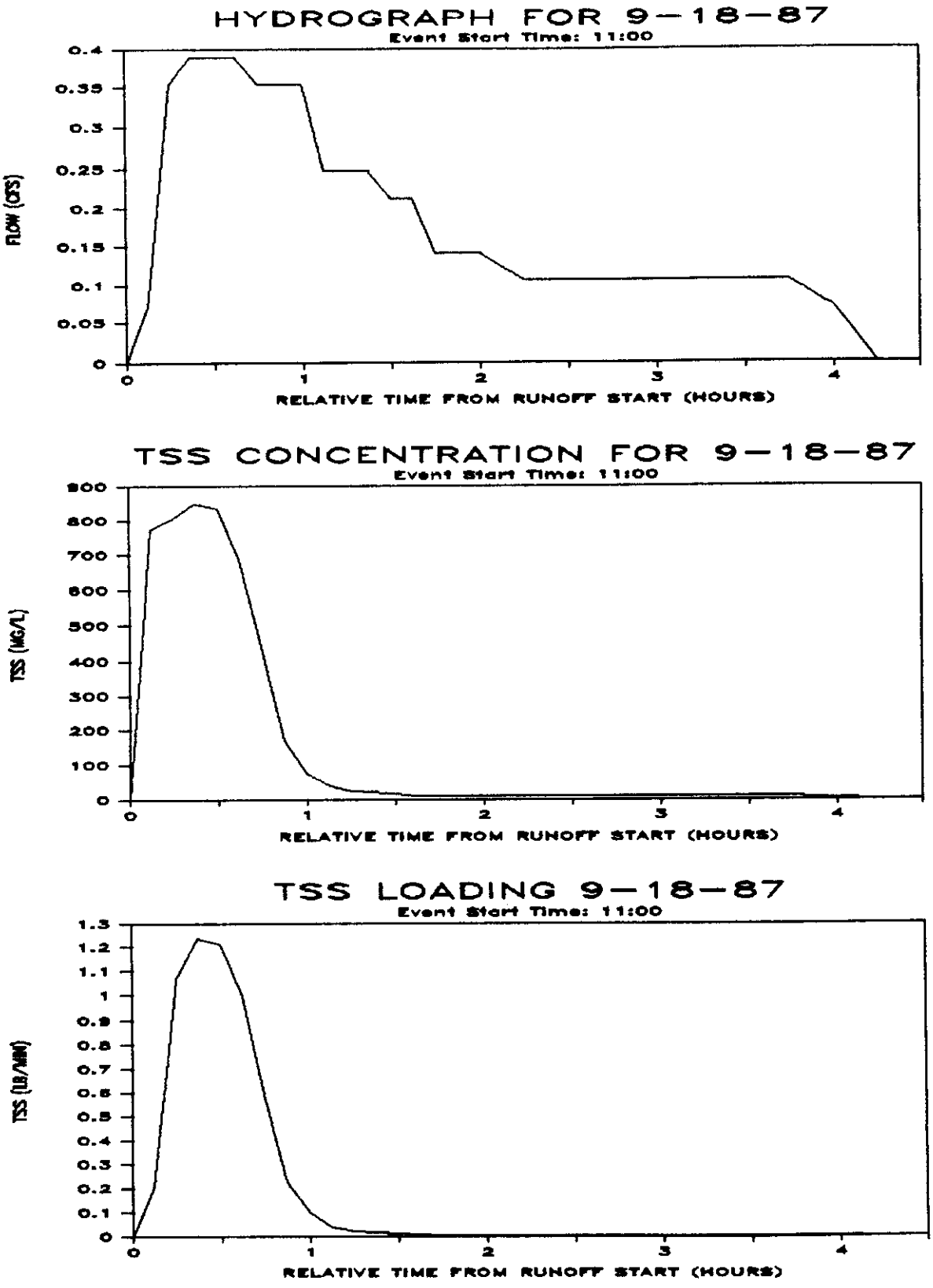


FIGURE A.15
FIELD DATA FOR EVENT OF 9-18-87

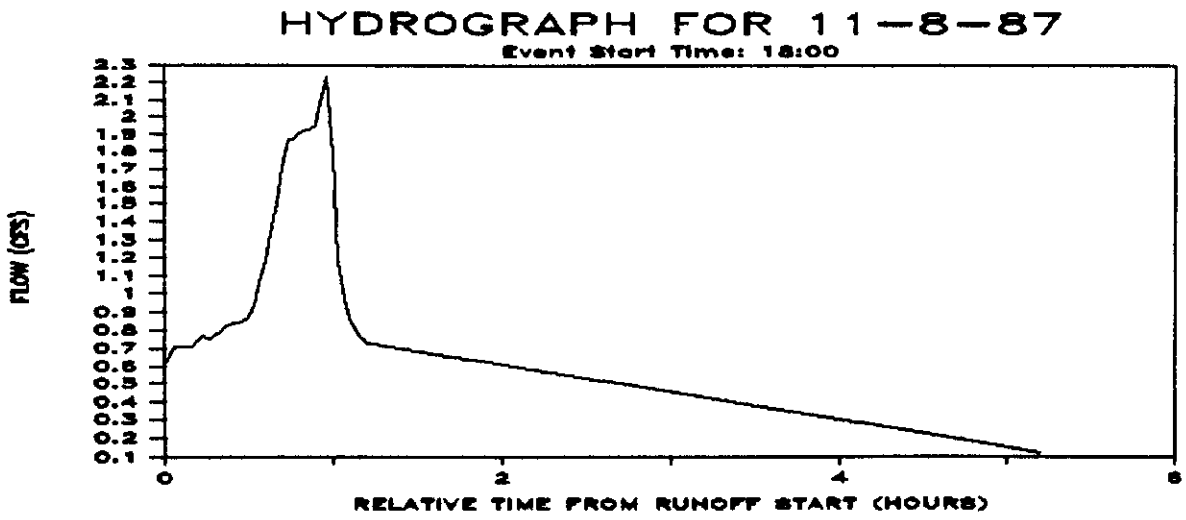


FIGURE A.16
FIELD DATA FOR EVENT OF 11-8-87

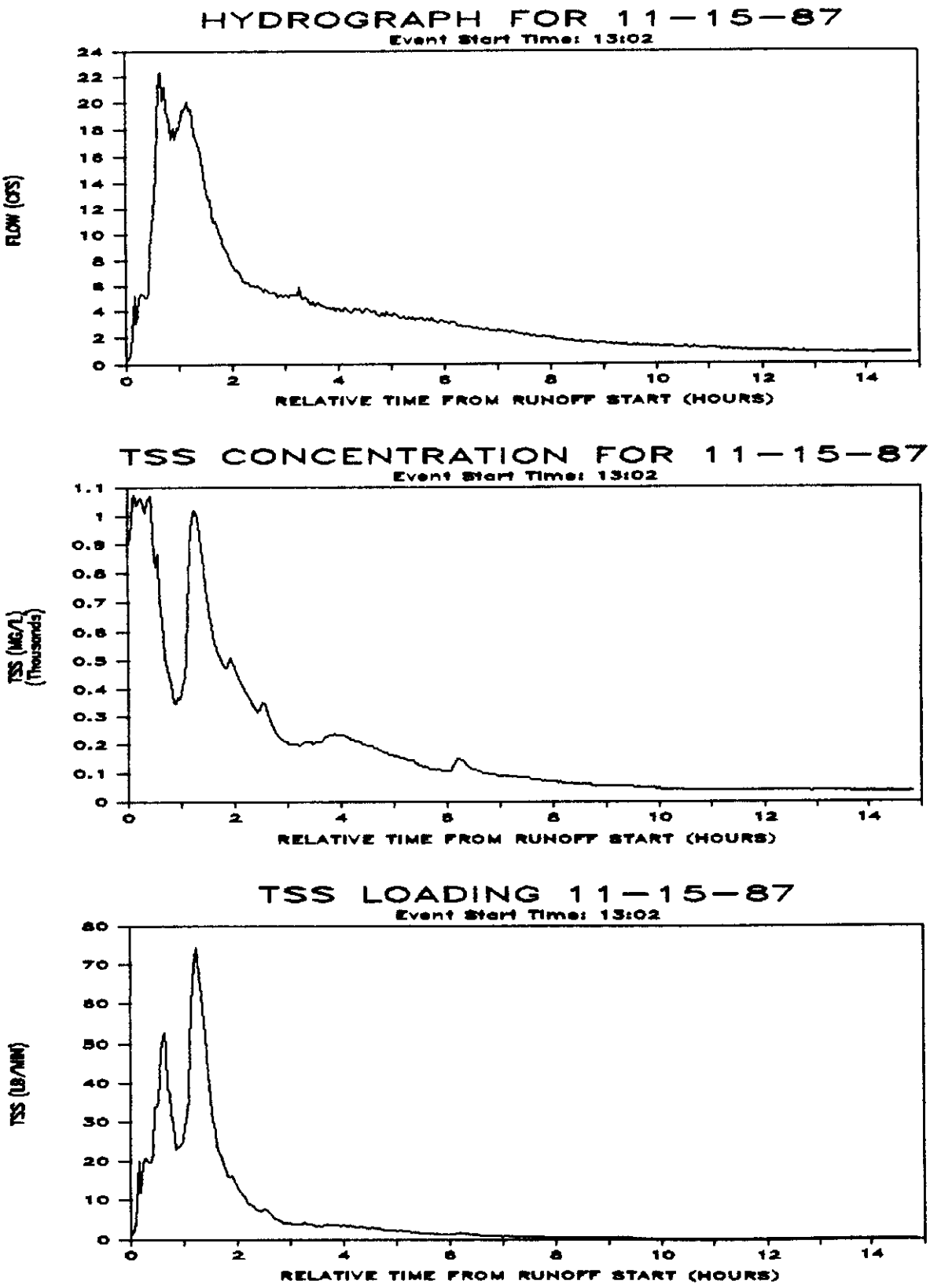


FIGURE A.17
FIELD DATA FOR EVENT OF 11-15-87

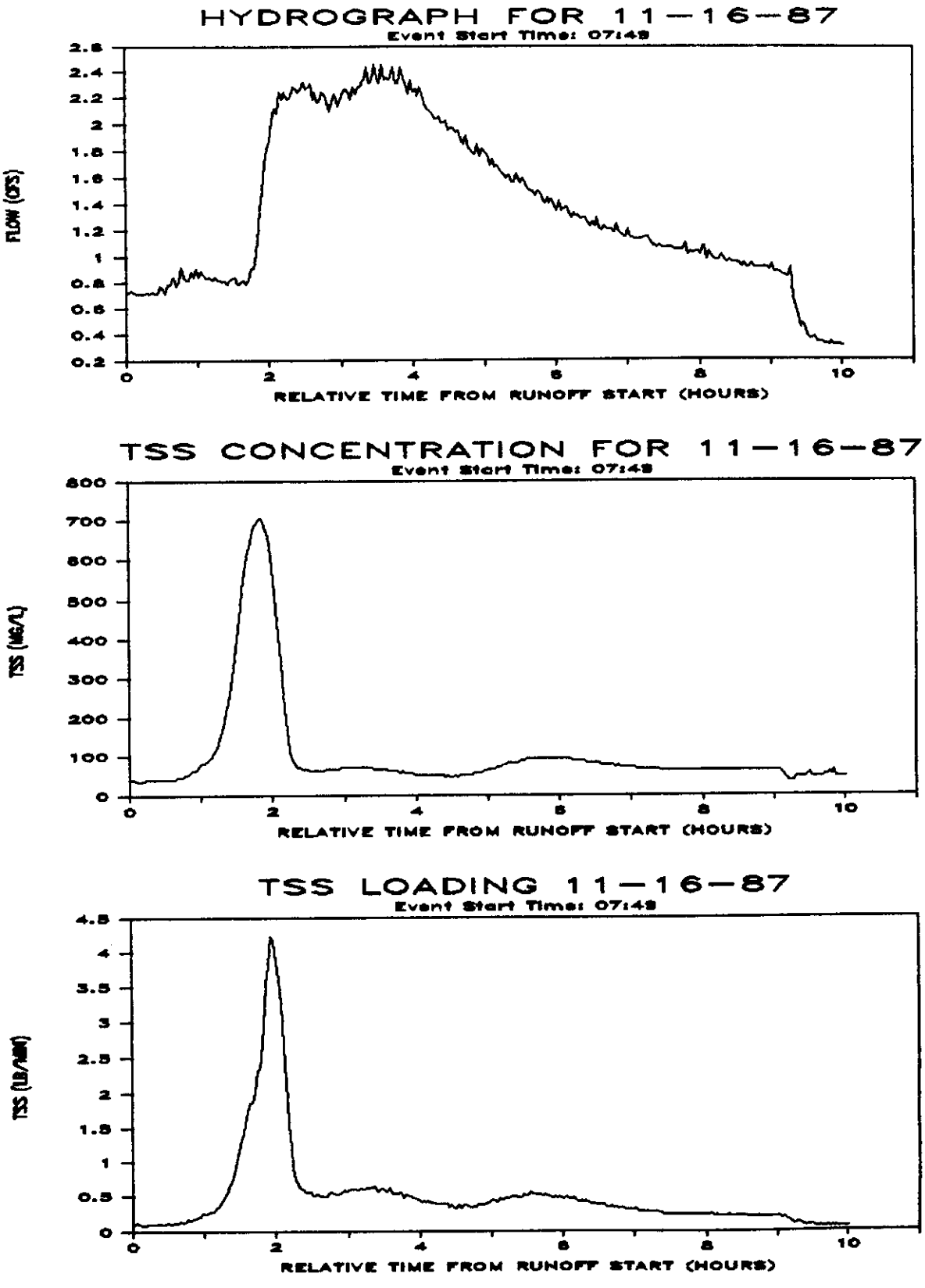


FIGURE A.18
FIELD DATA FOR EVENT OF 11-16-87

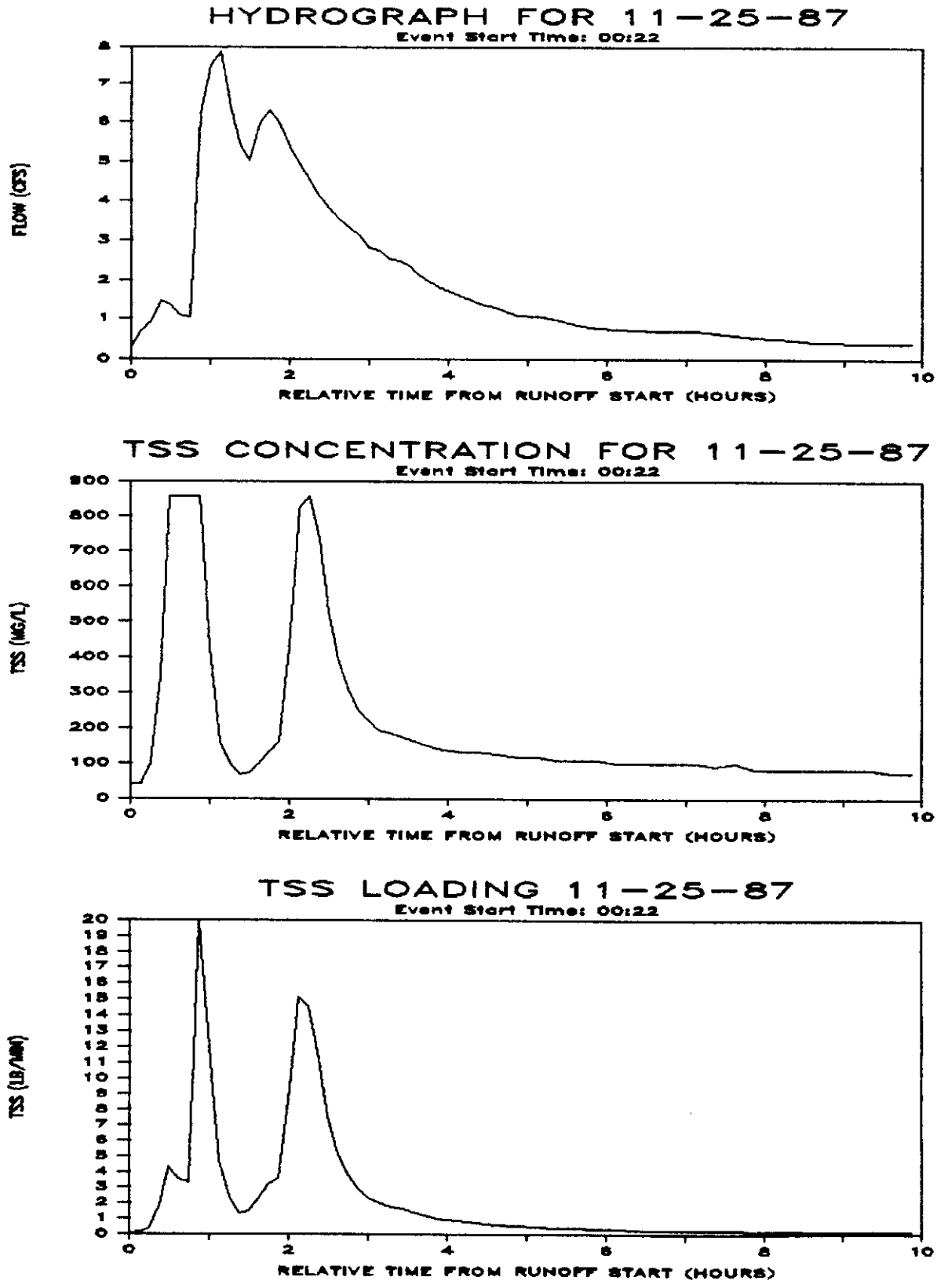


FIGURE A.19
FIELD DATA FOR EVENT OF 11-25-87

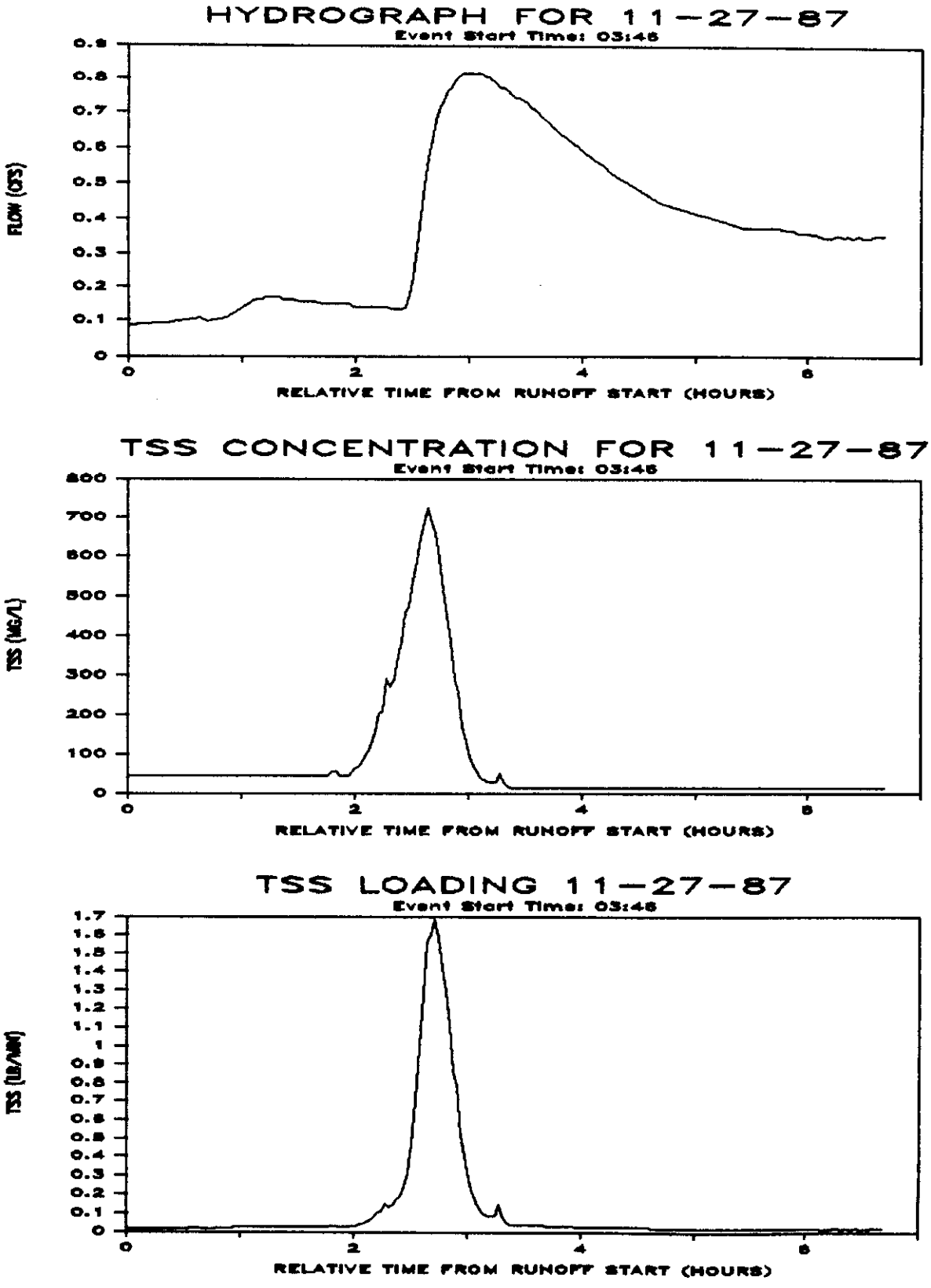


FIGURE A.20
FIELD DATA FOR EVENT OF 11-27-87

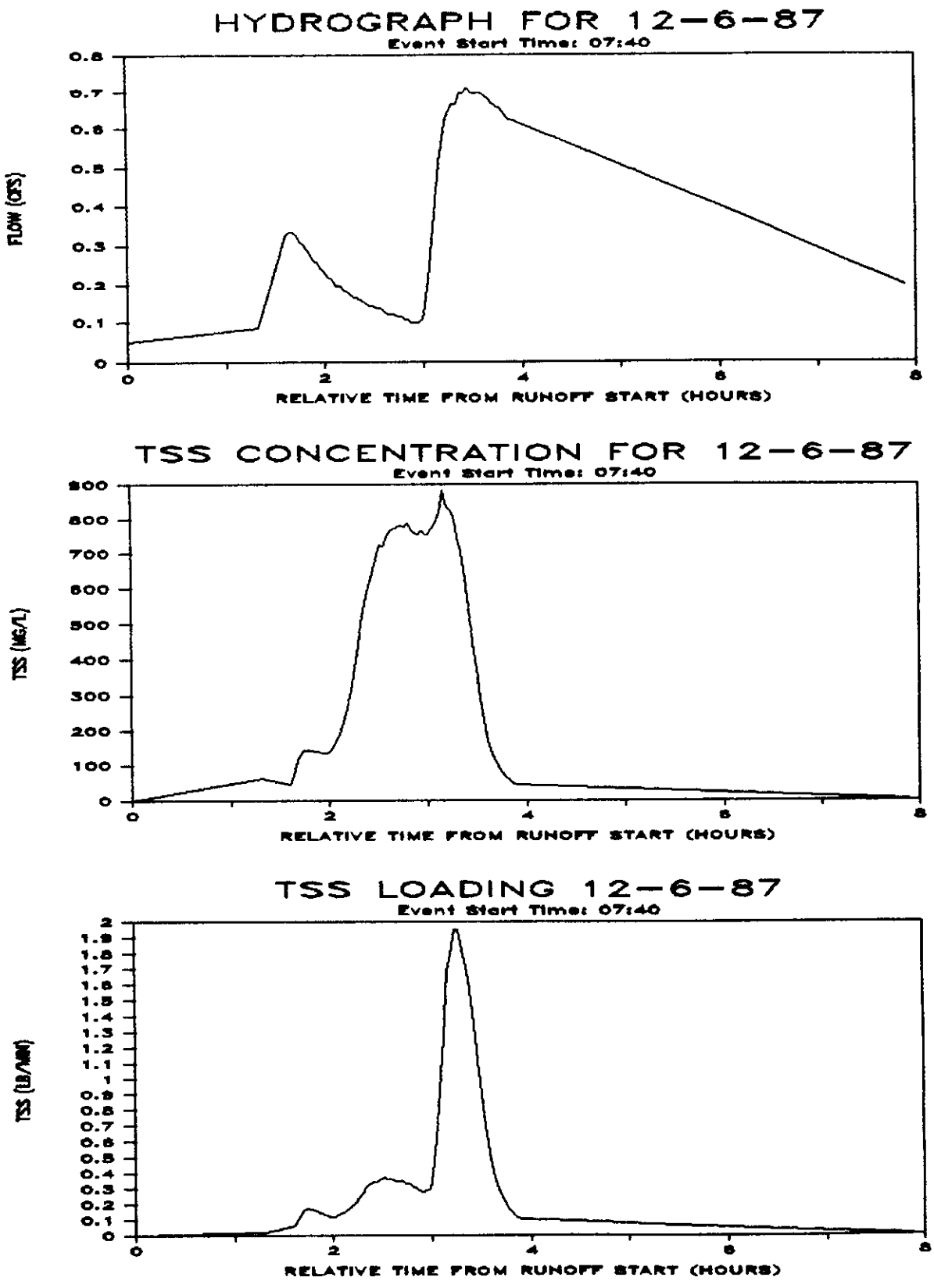


FIGURE A.21
FIELD DATA FOR EVENT OF 12-6-87

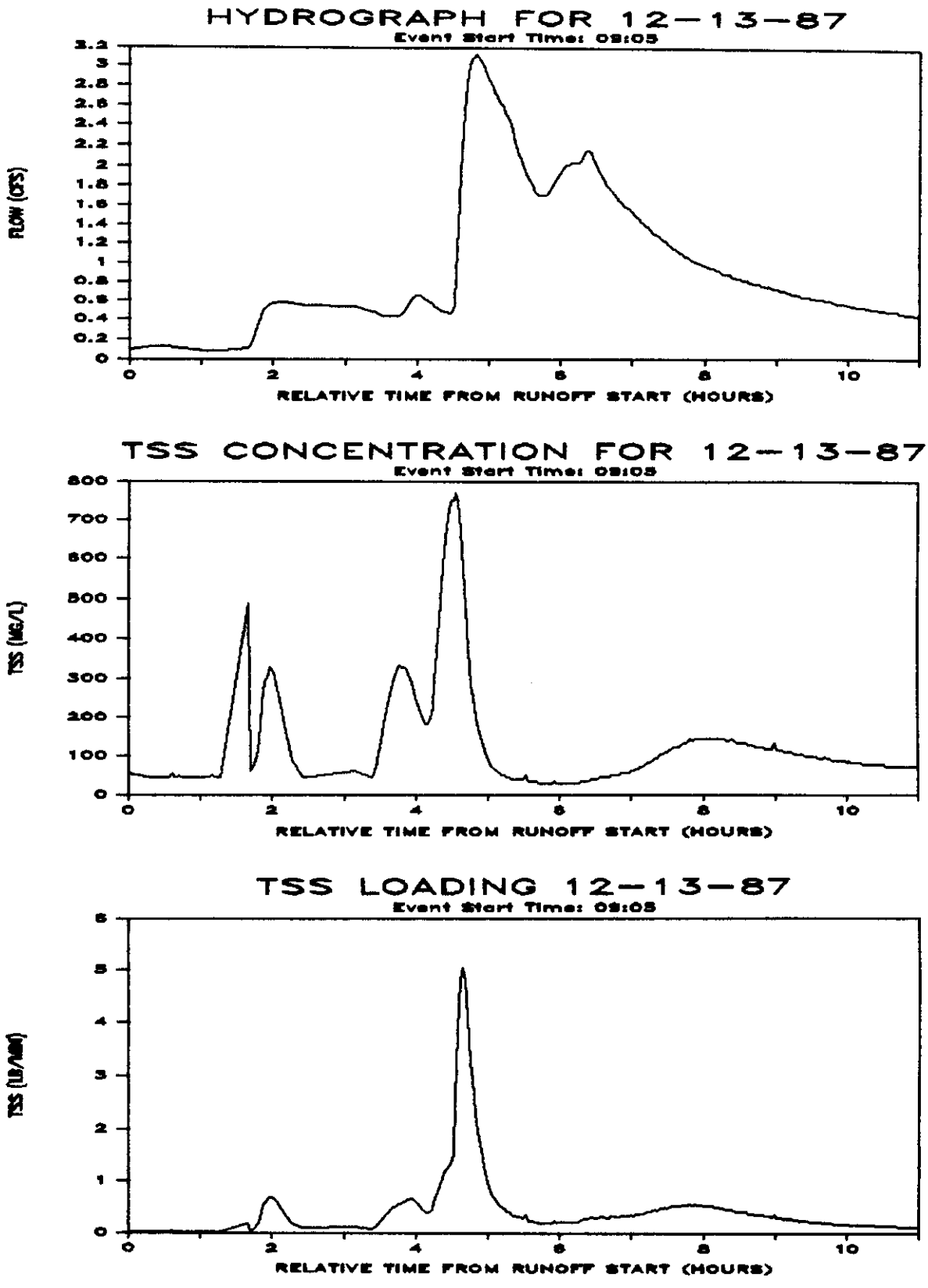


FIGURE A.22
FIELD DATA FOR EVENT OF 12-13-87

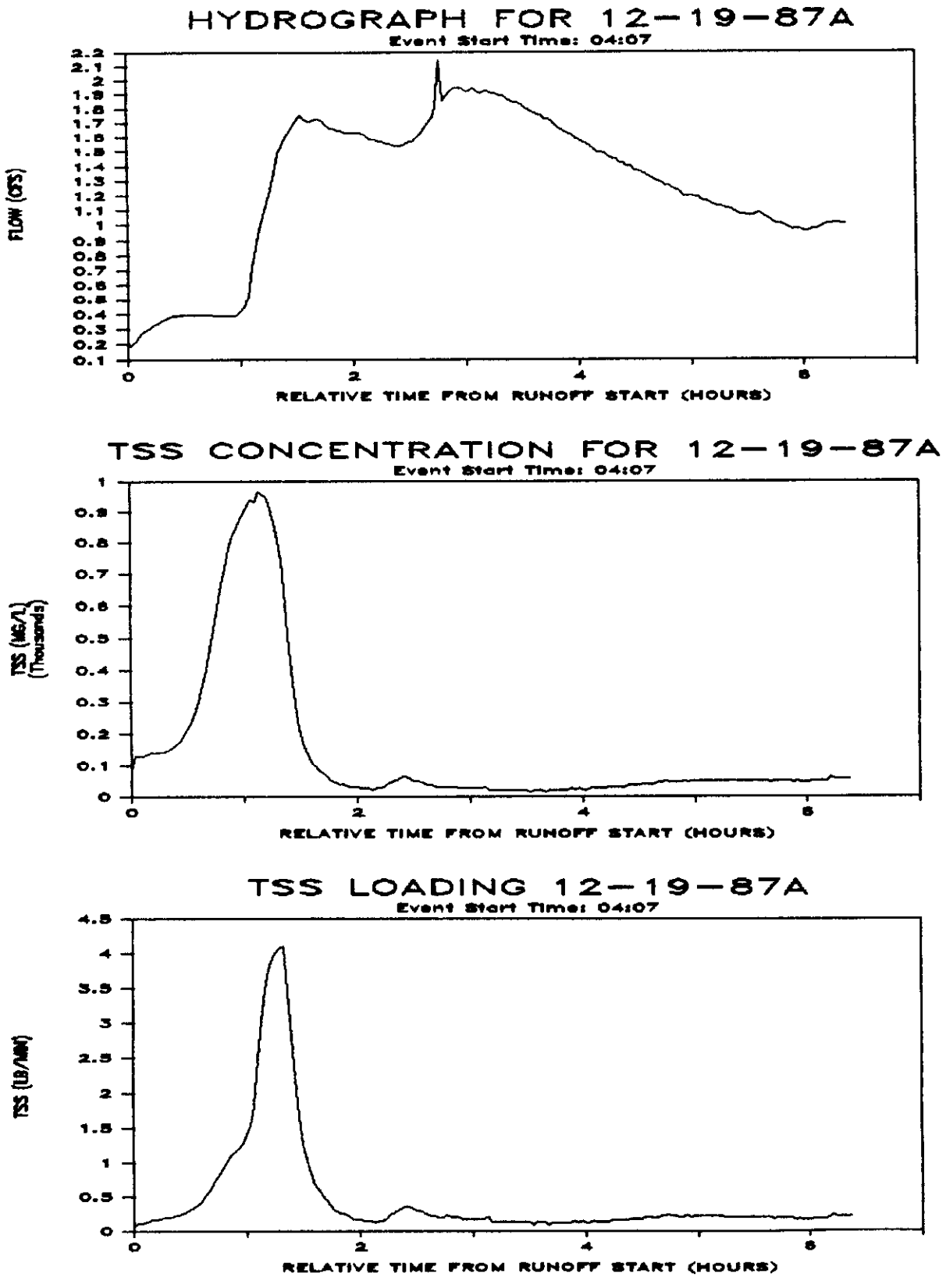


FIGURE A.23
FIELD DATA FOR EVENT OF 12-19-87A

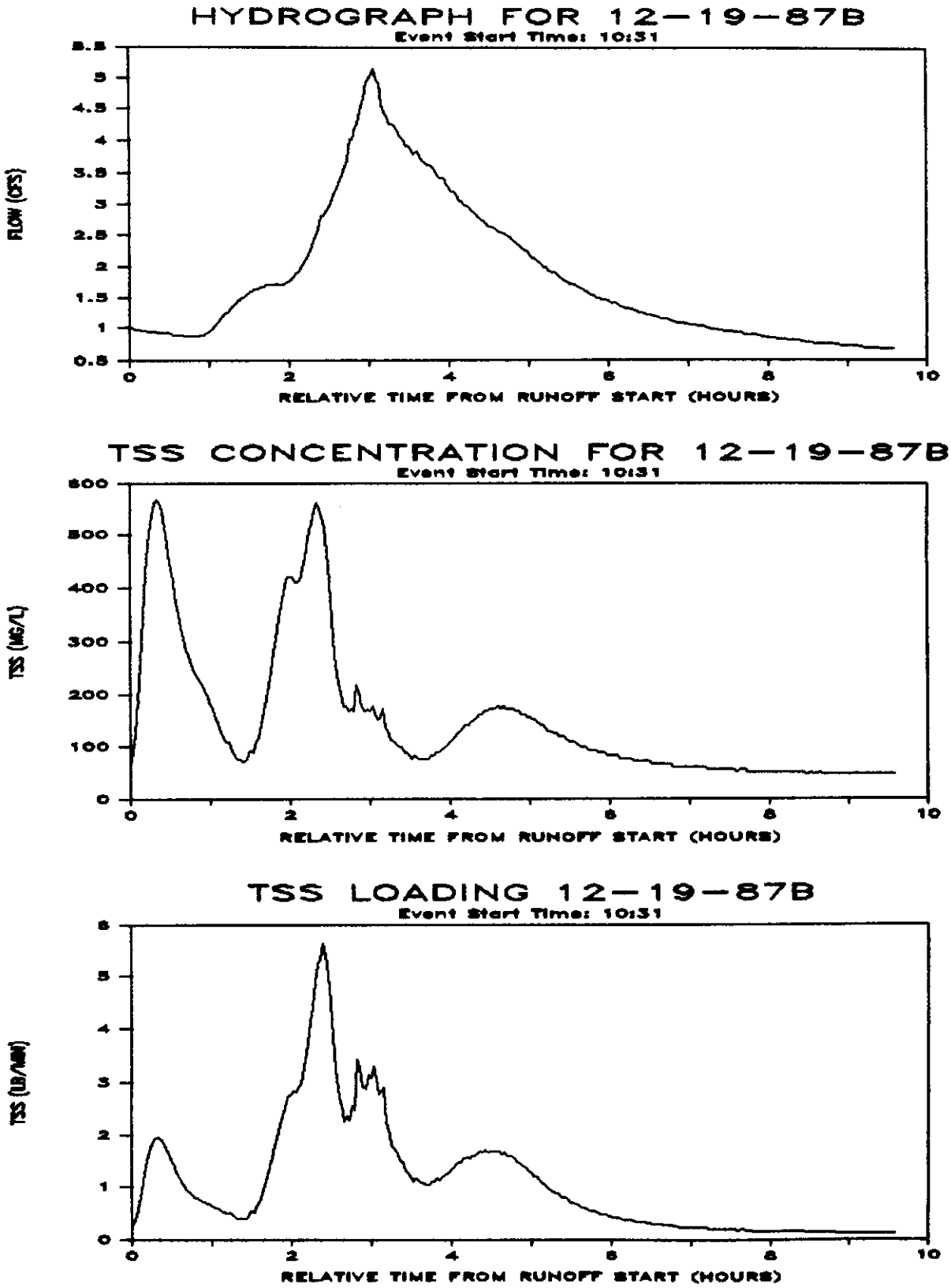


FIGURE A.24
FIELD DATA FOR EVENT OF 12-19-87B

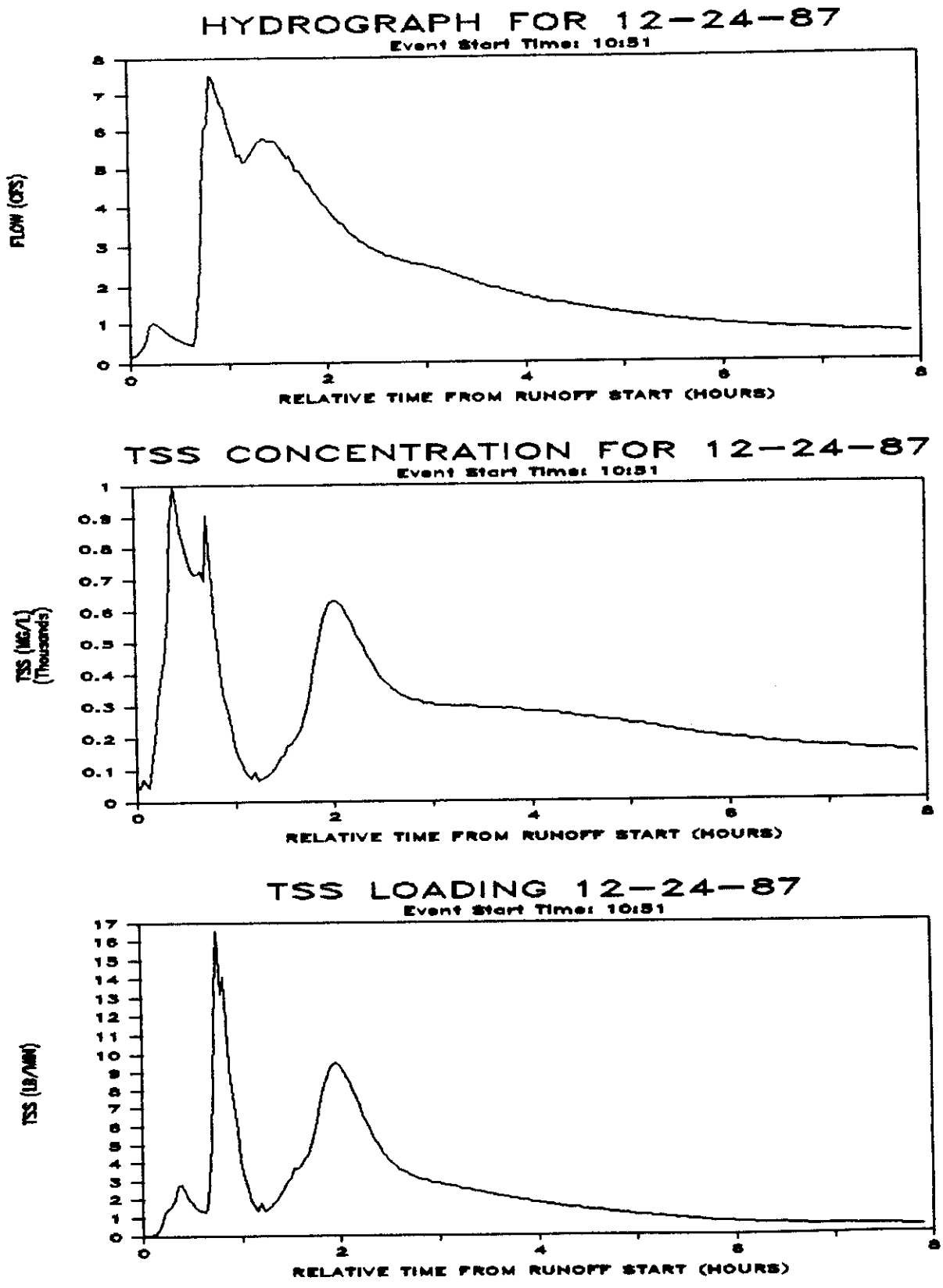


FIGURE A.25
FIELD DATA FOR EVENT OF 12-24-87

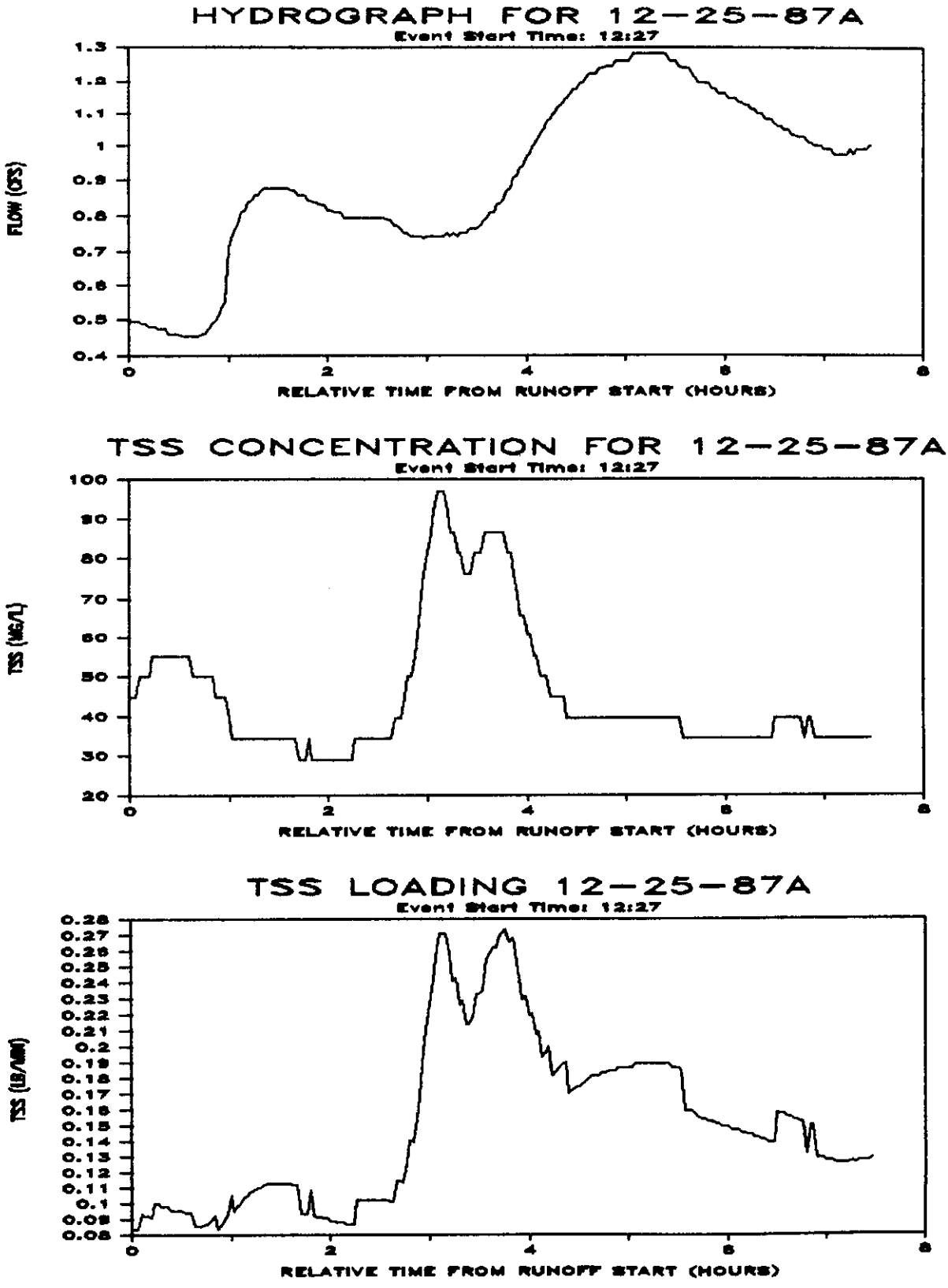


FIGURE A.26
FIELD DATA FOR EVENT OF 12-25-87A

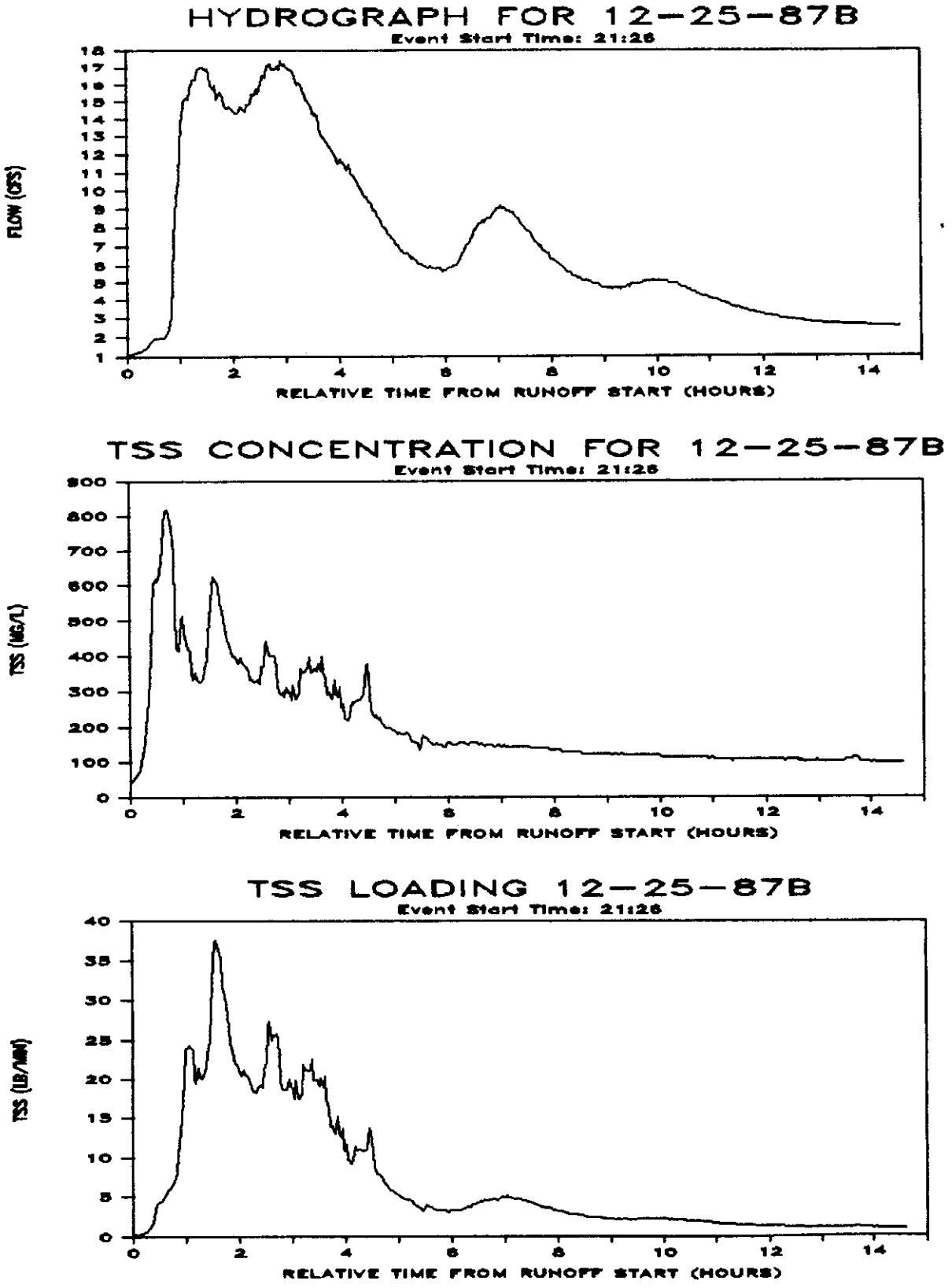


FIGURE A.27
FIELD DATA FOR EVENT OF 12-25-87B

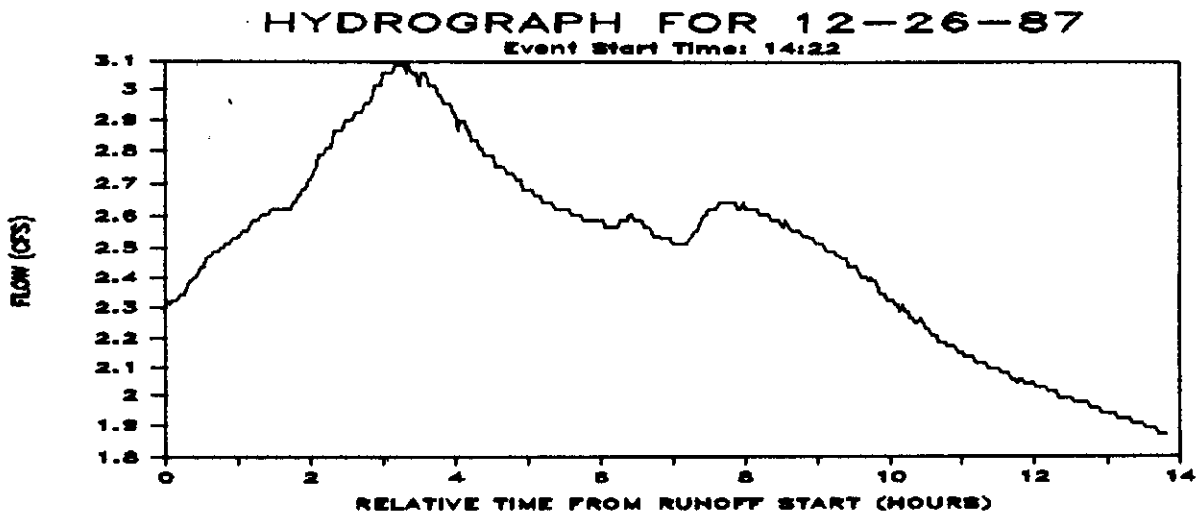


FIGURE A.28
FIELD DATA FOR EVENT OF 12-26-87

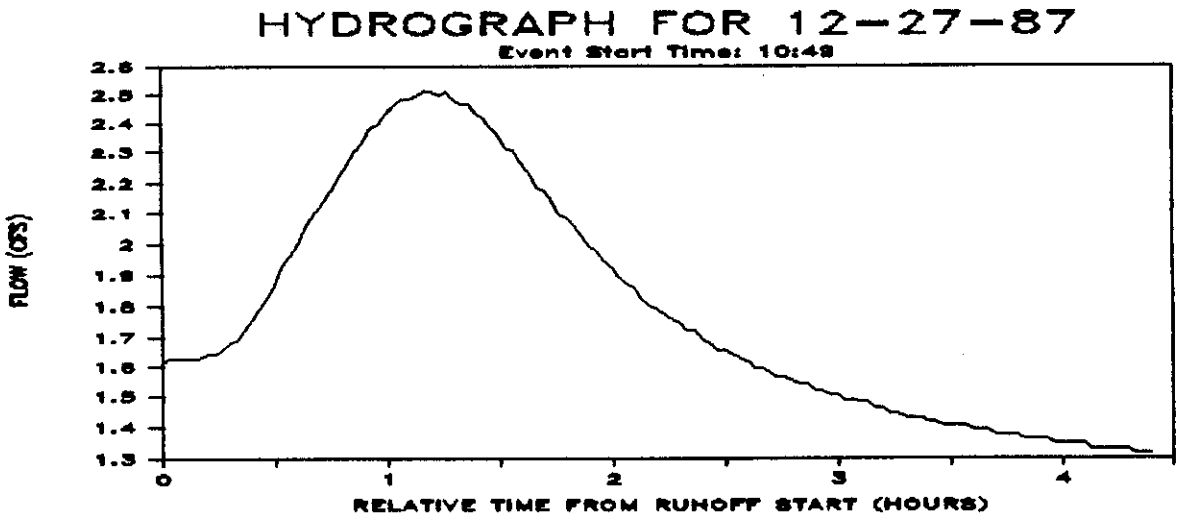


FIGURE A.29
FIELD DATA FOR EVENT OF 12-27-87

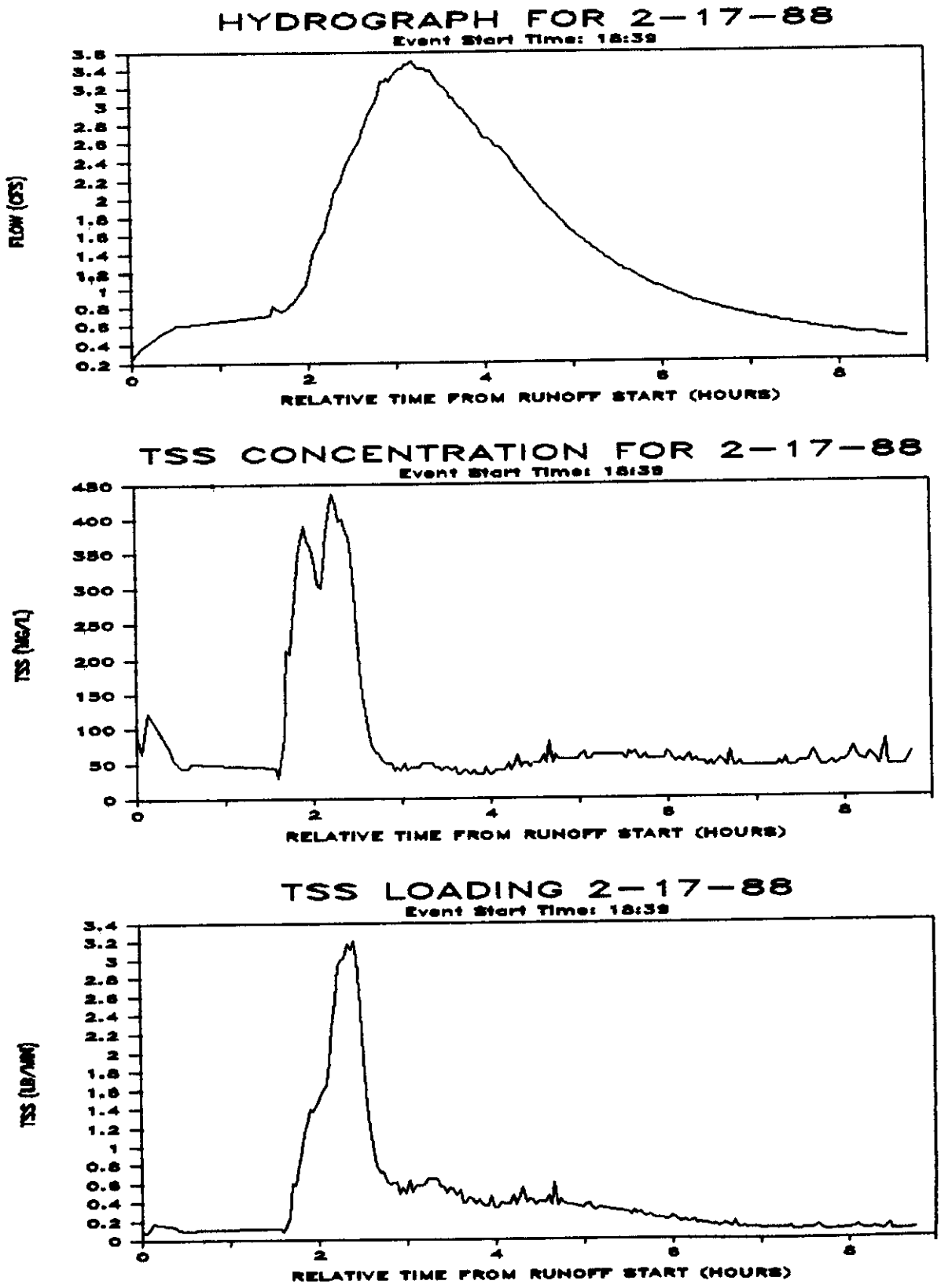


FIGURE A.30
FIELD DATA FOR EVENT OF 2-17-88

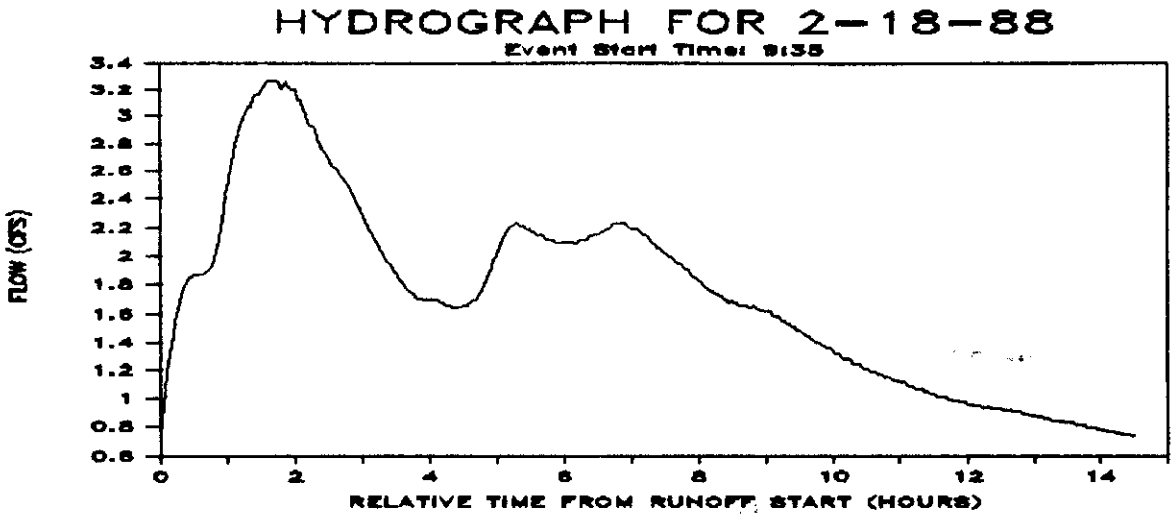


FIGURE A.31
FIELD DATA FOR EVENT OF 2-18-88

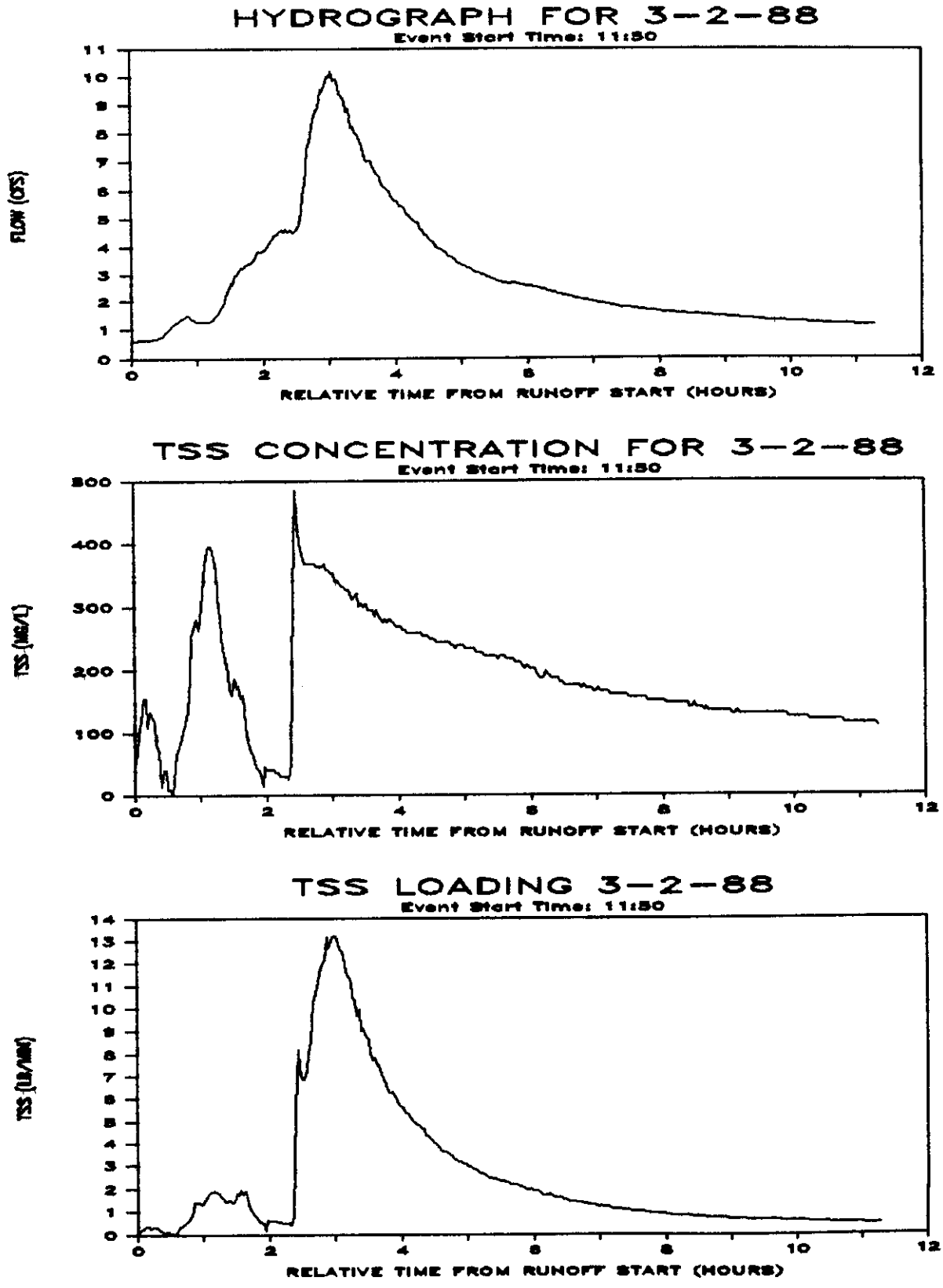


FIGURE A.32
FIELD DATA FOR EVENT OF 3-2-88

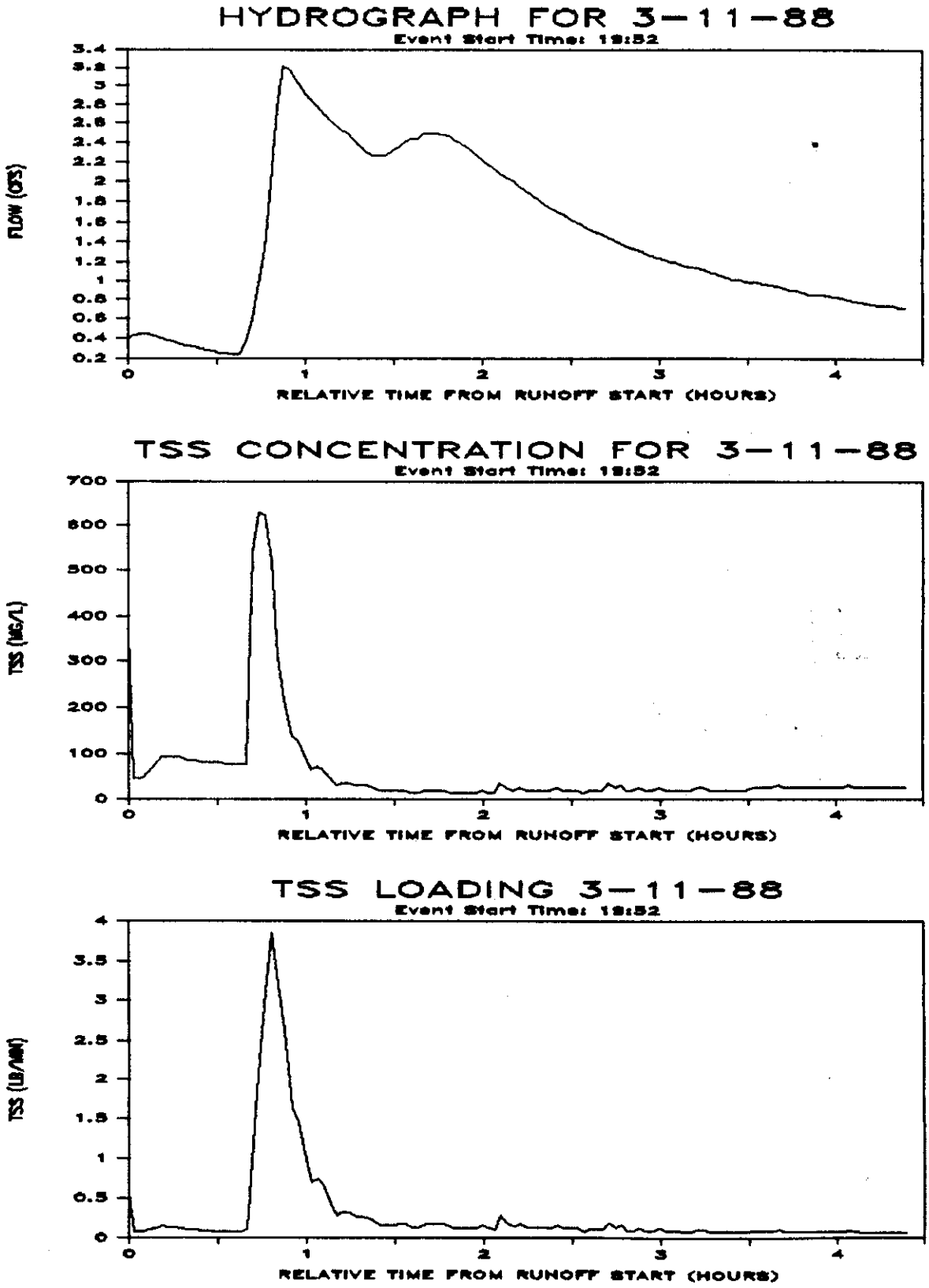


FIGURE A.33
FIELD DATA FOR EVENT OF 3-11-88

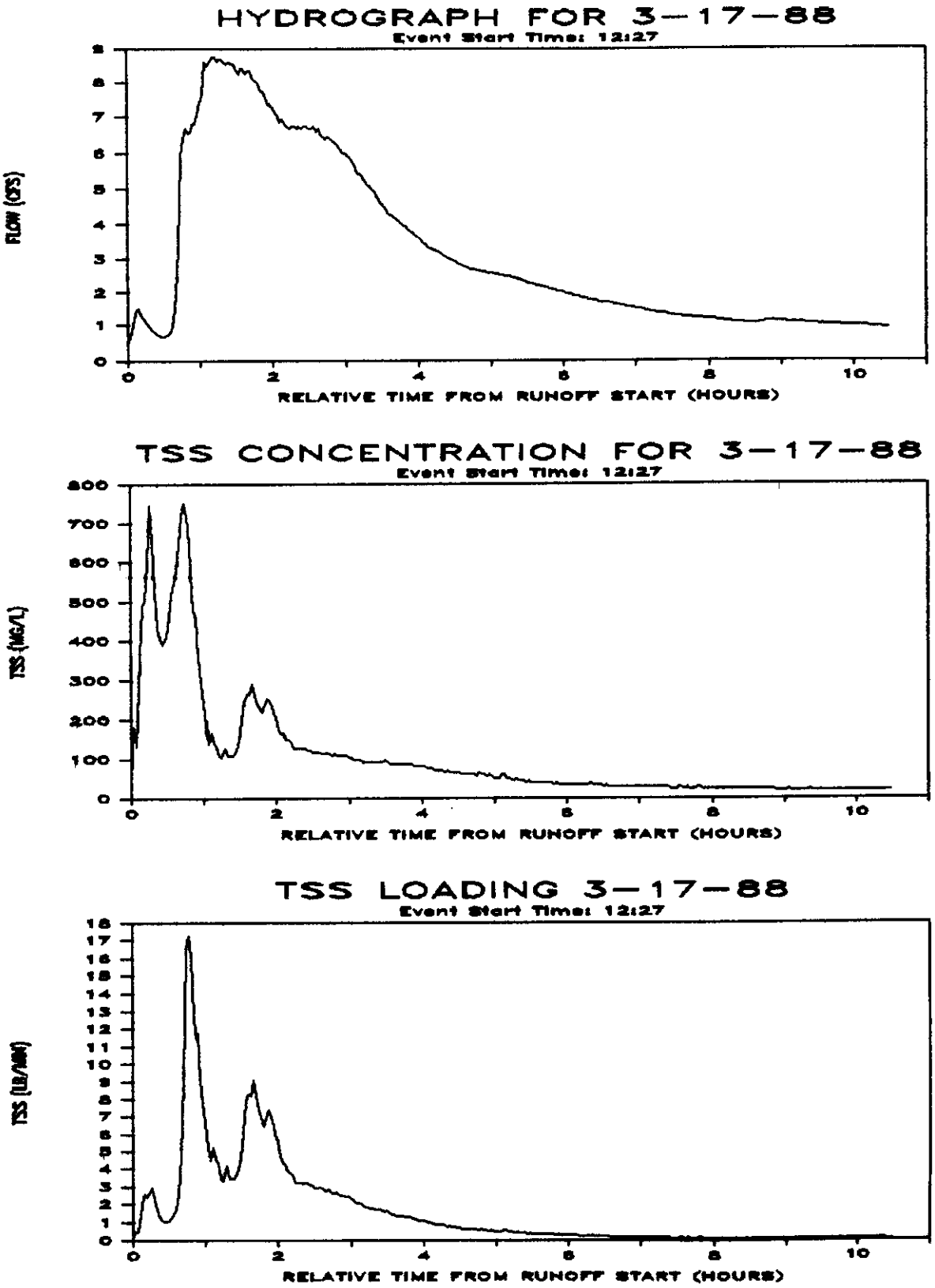


FIGURE A.34
FIELD DATA FOR EVENT OF 3-17-88

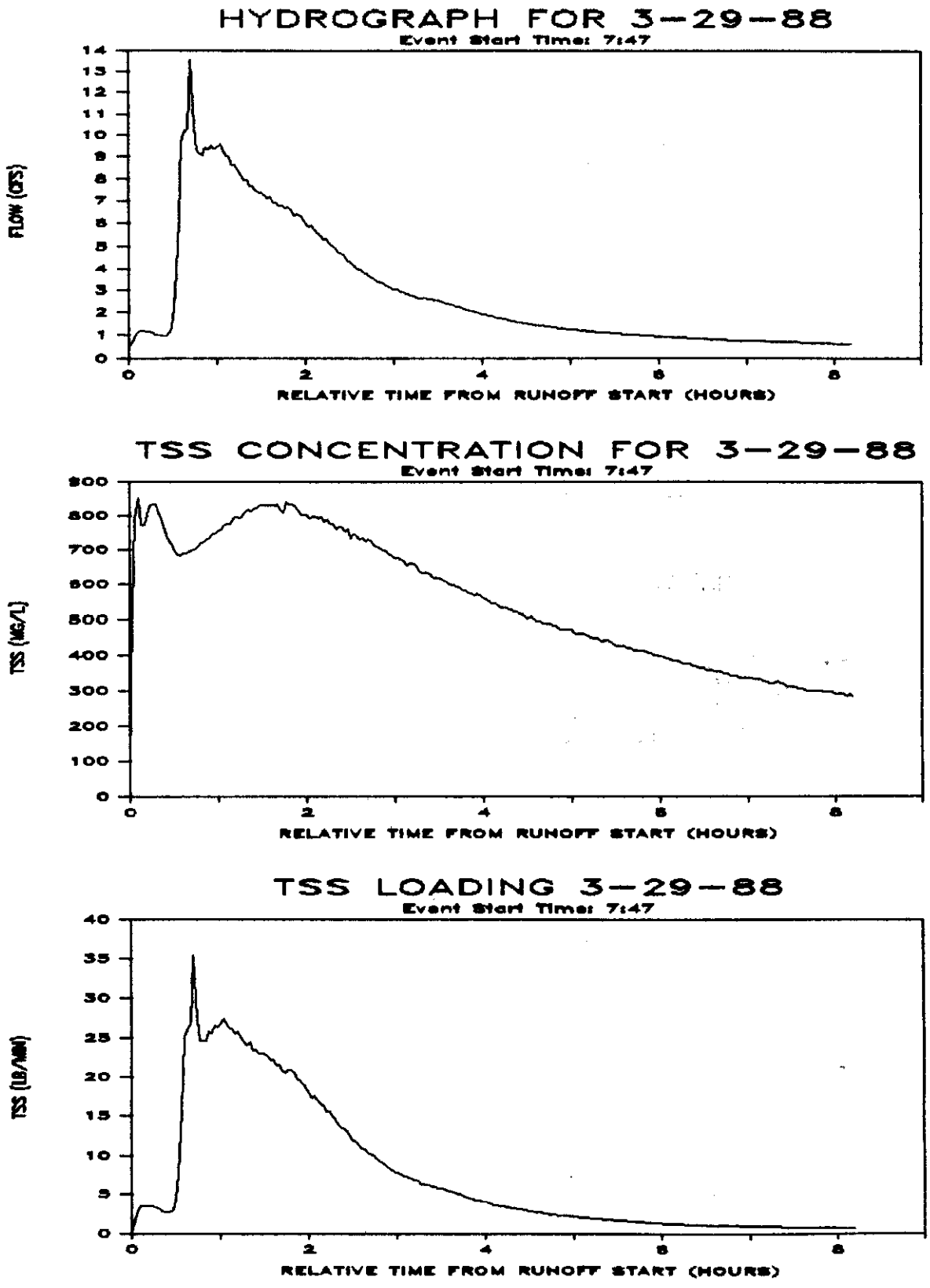


FIGURE A.35
FIELD DATA FOR EVENT OF 3-29-88

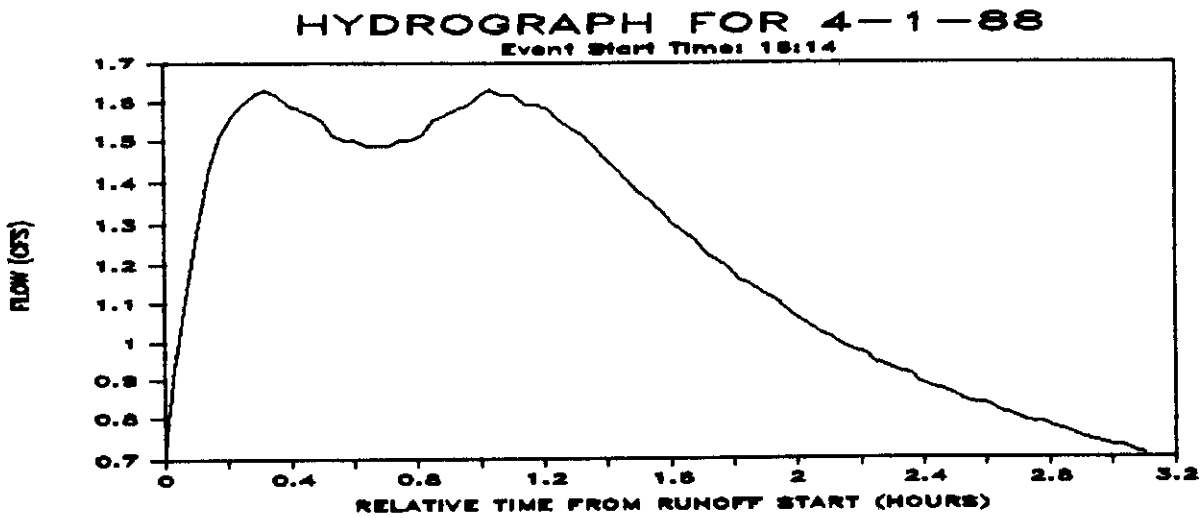


FIGURE A.36
FIELD DATA FOR EVENT OF 4-1-88

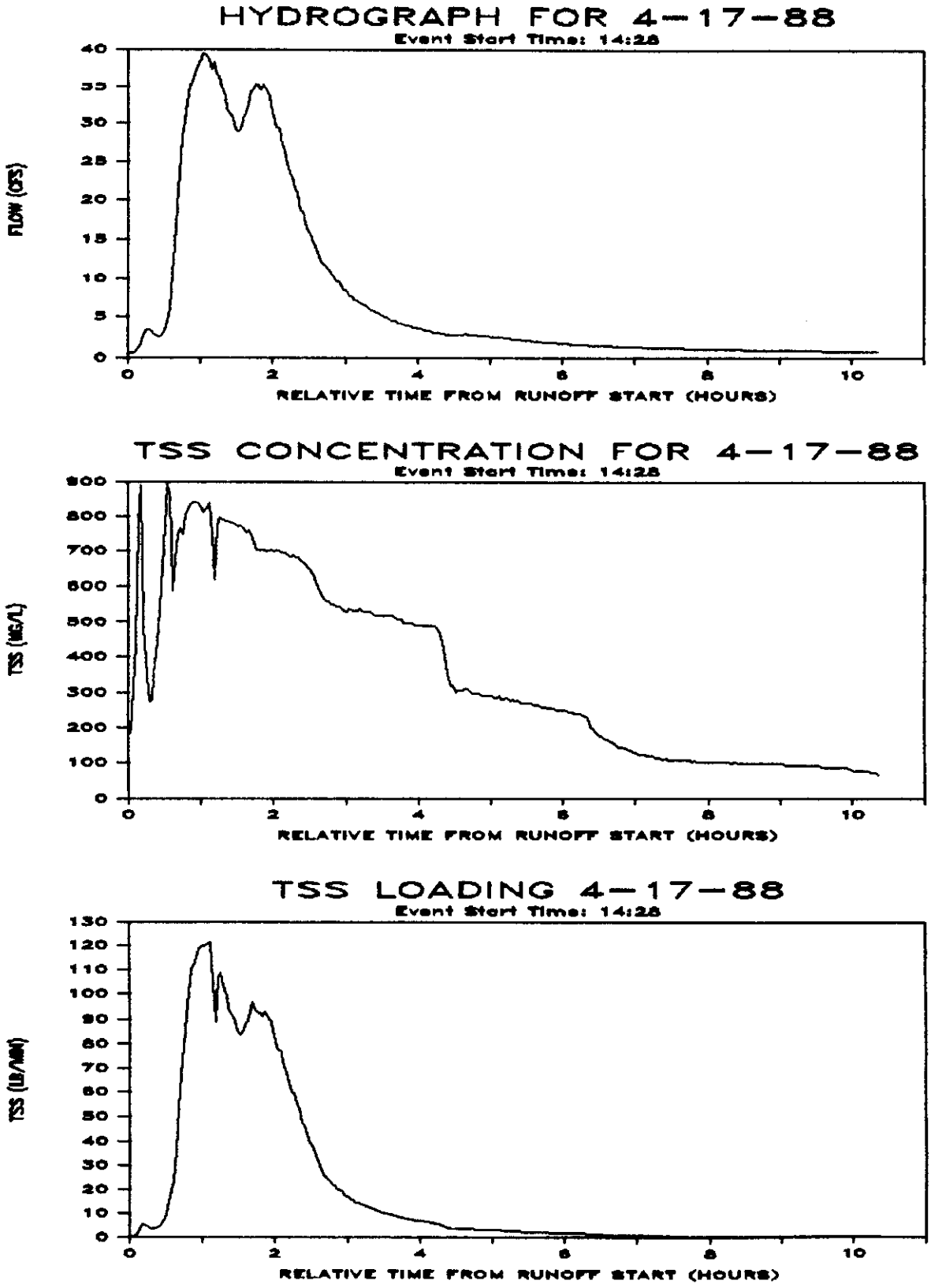


FIGURE A.37
FIELD DATA FOR EVENT OF 4-17-88

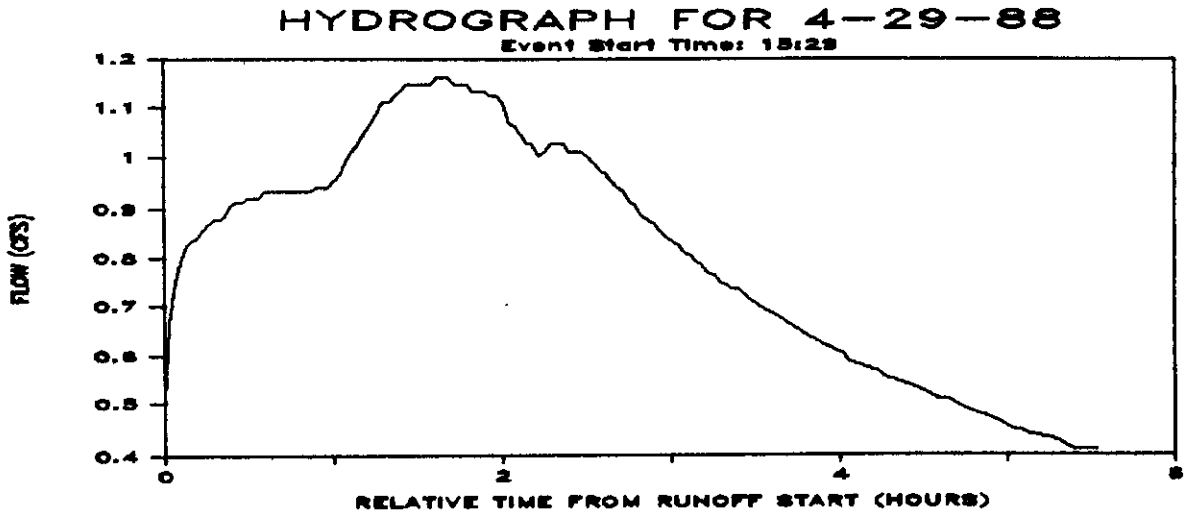


FIGURE A.38
FIELD DATA FOR EVENT OF 4-29-88

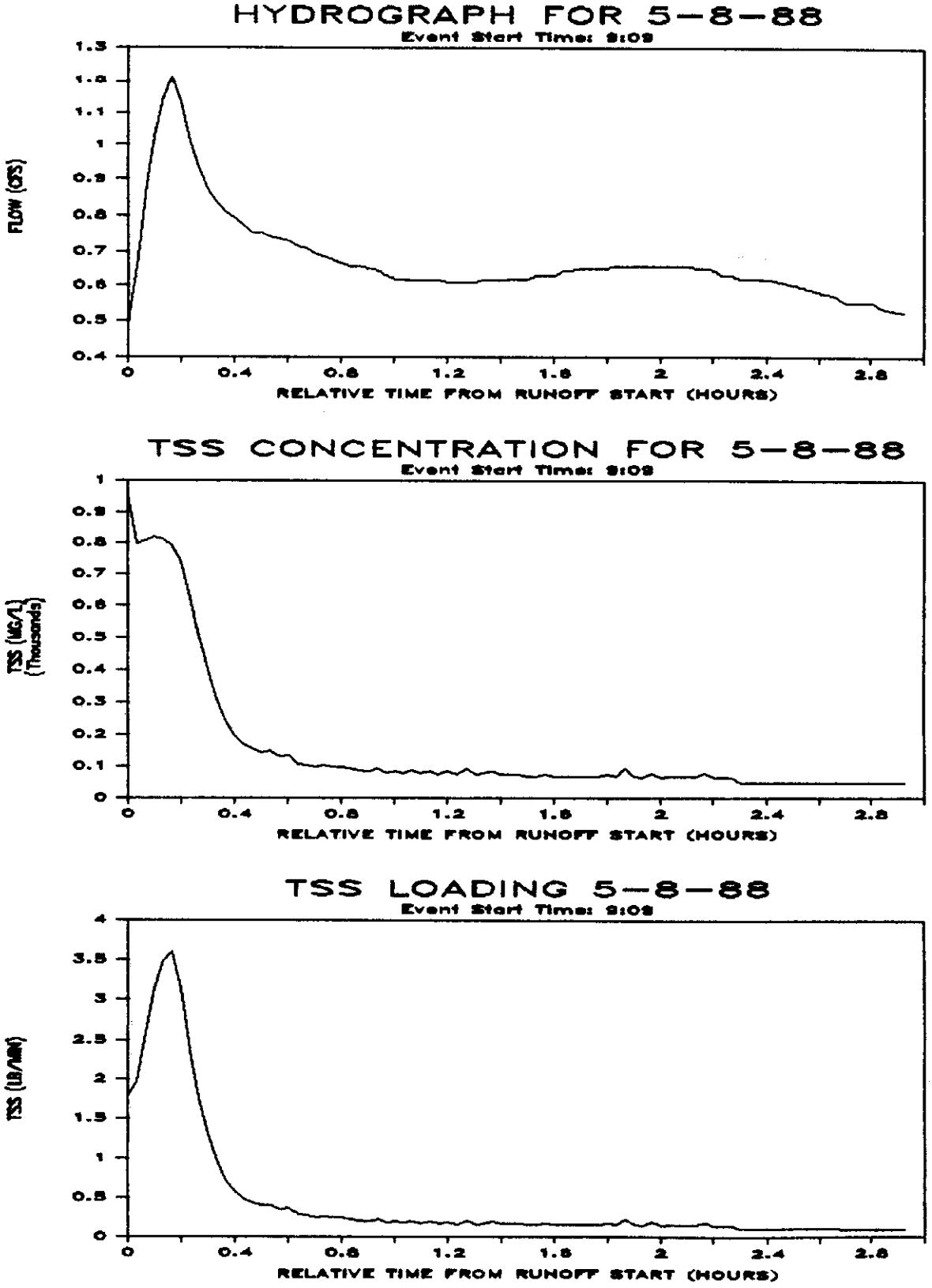


FIGURE A.39
FIELD DATA FOR EVENT OF 5-8-88

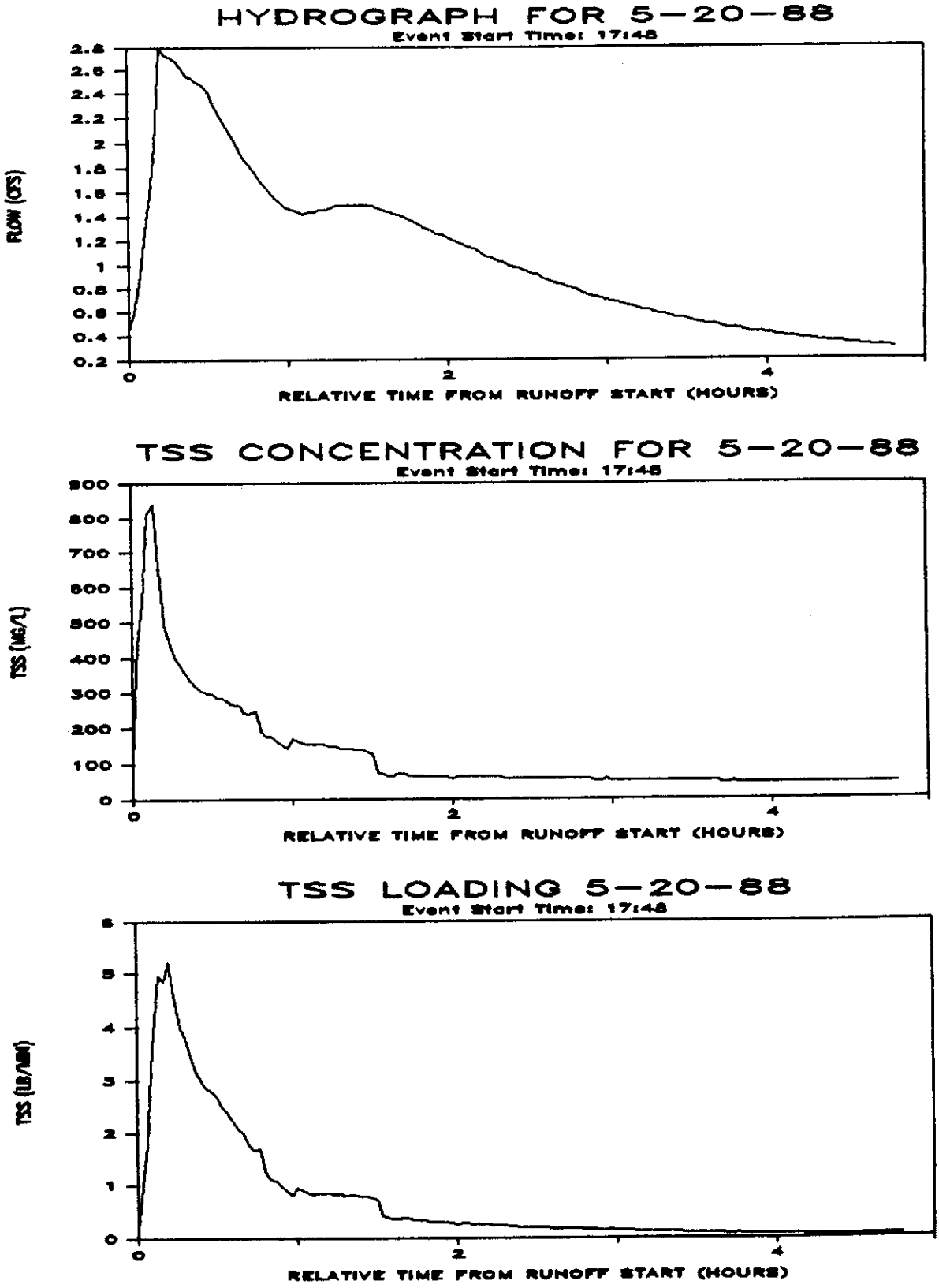


FIGURE A.40
FIELD DATA FOR EVENT OF 5-20-88

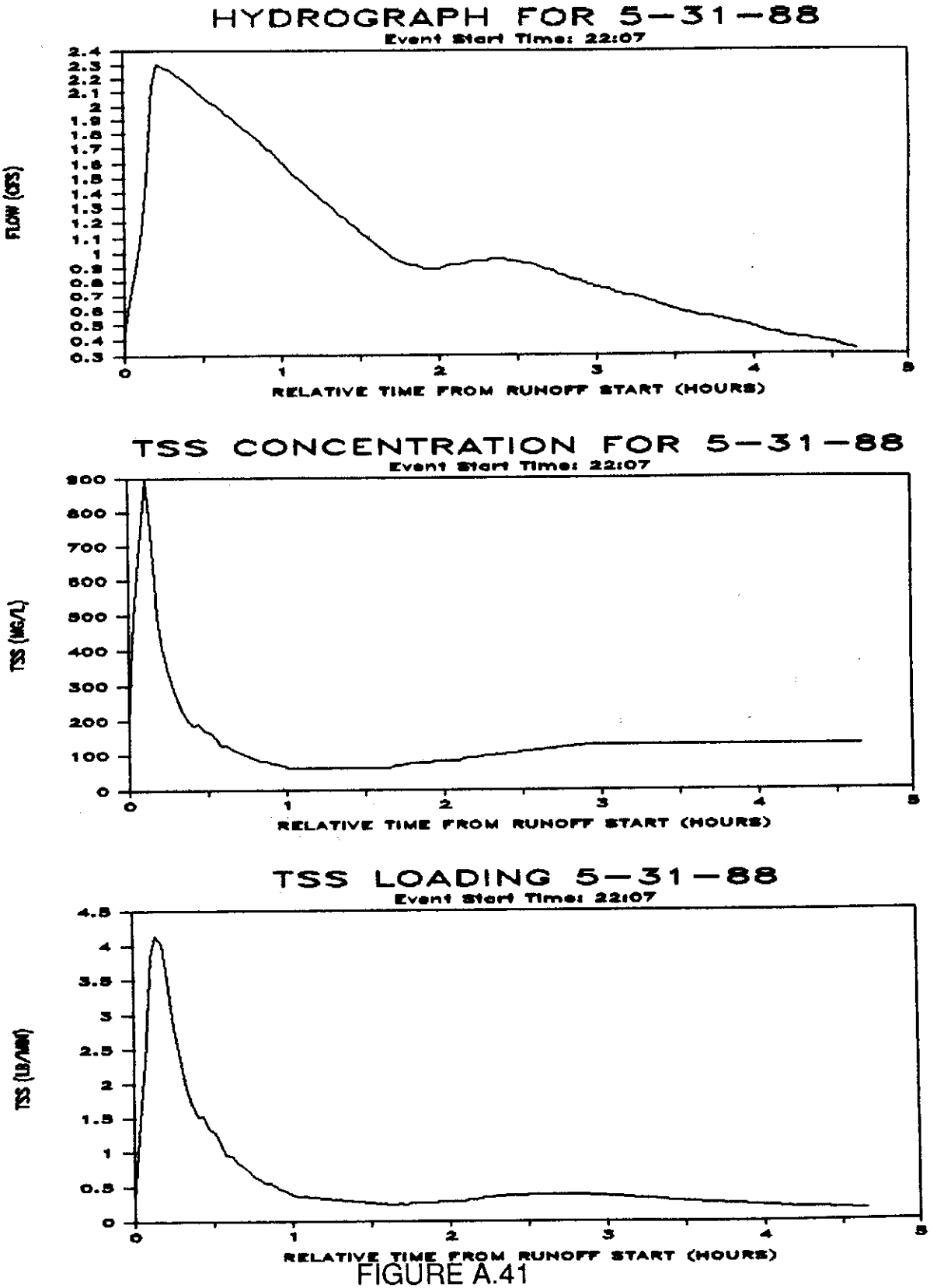


FIGURE A.41
FIELD DATA FOR EVENT OF 5-31-88

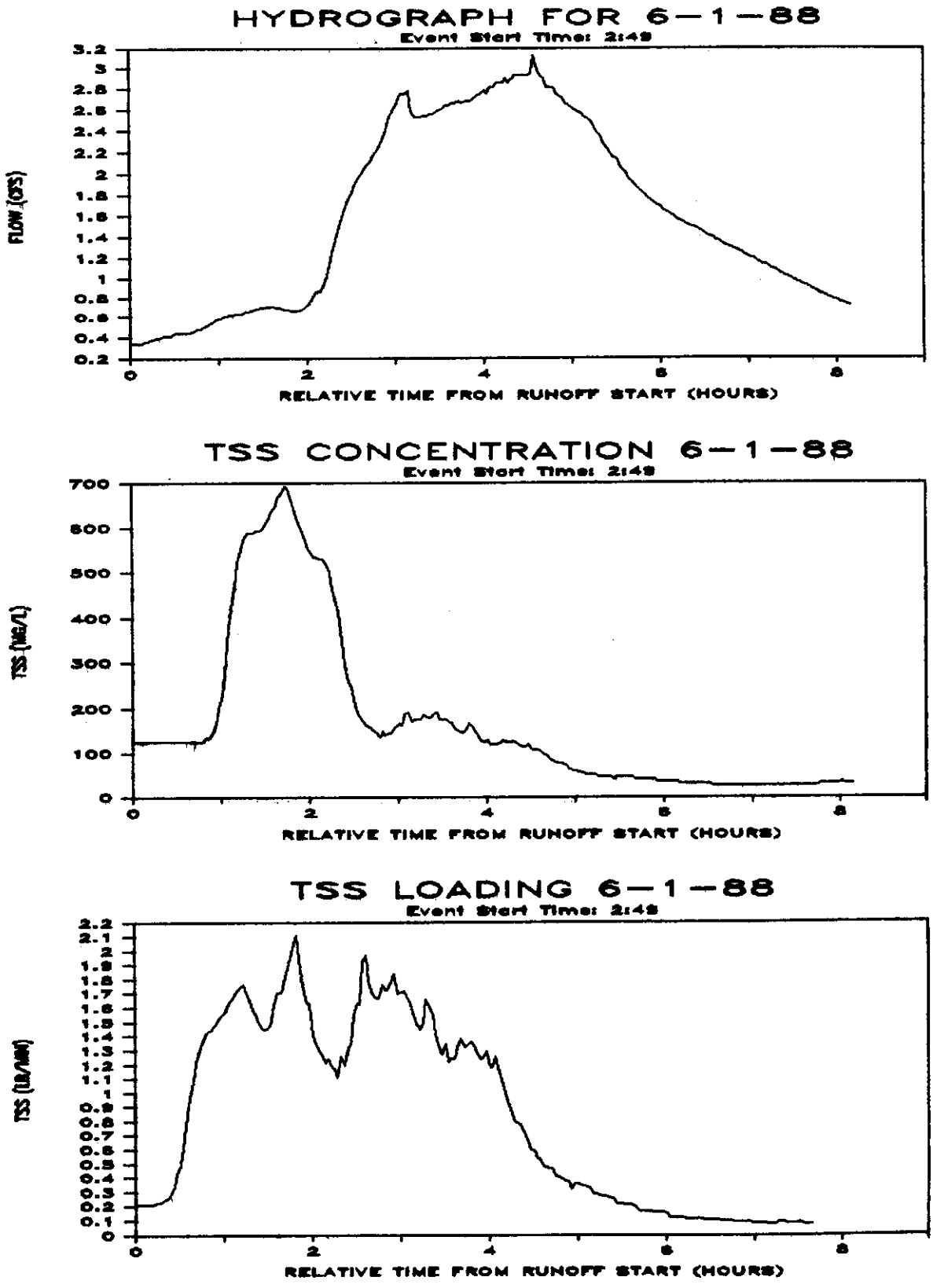


FIGURE A.42
FIELD DATA FOR EVENT OF 6-1-88

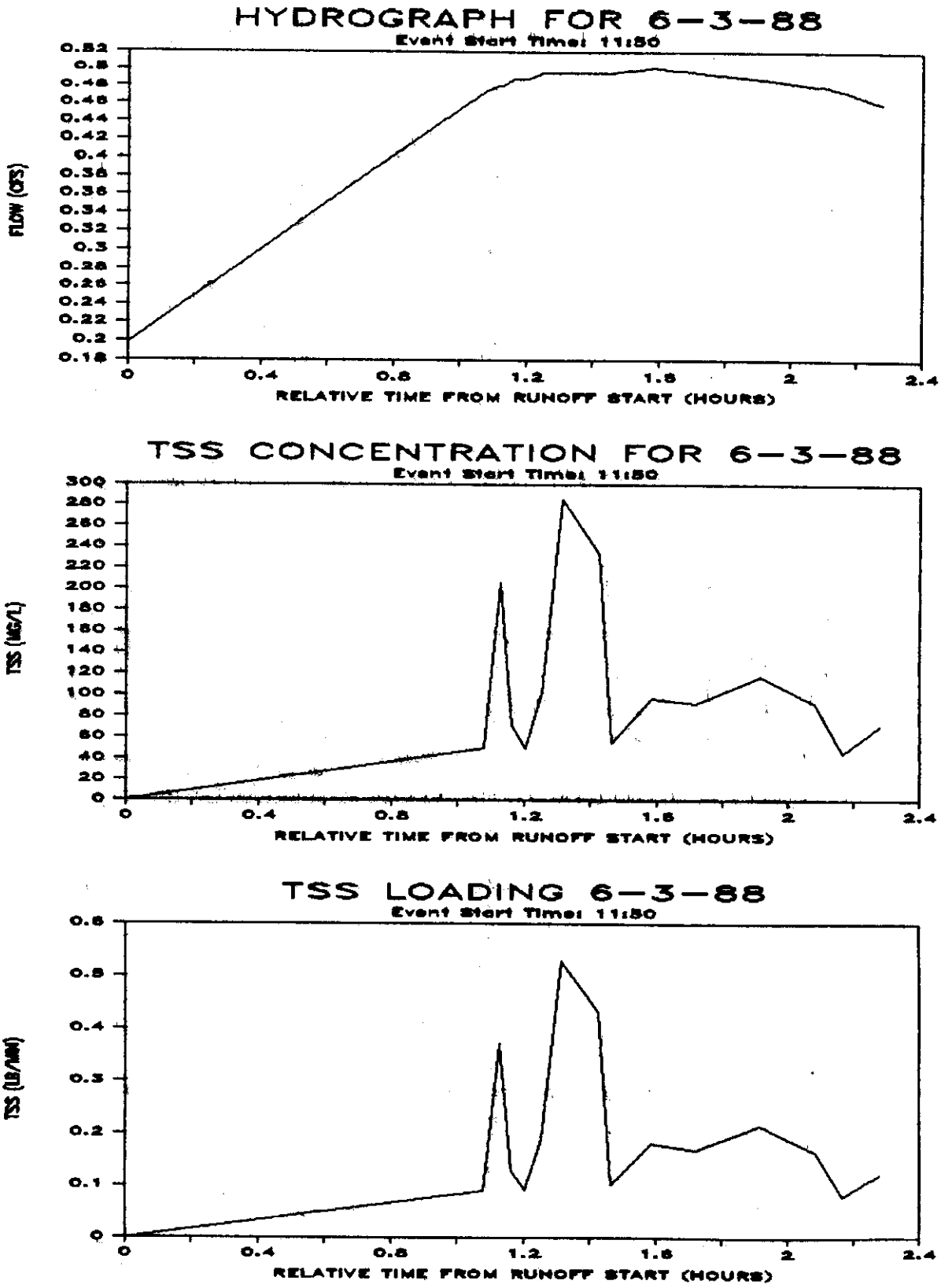


FIGURE A.43
FIELD DATA FOR EVENT OF 6-3-88

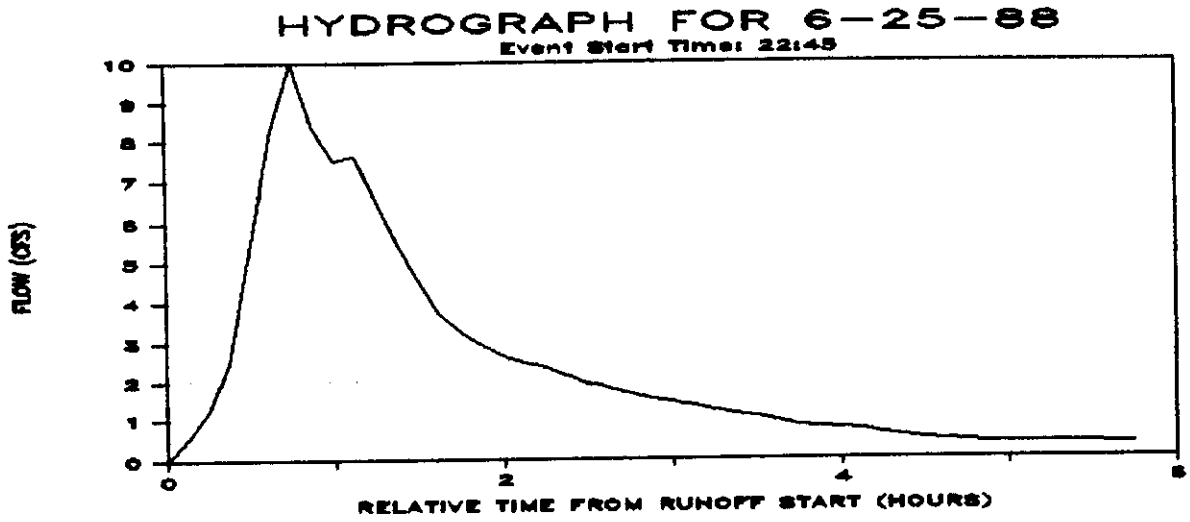


FIGURE A.44
FIELD DATA FOR EVENT OF 6-25-88

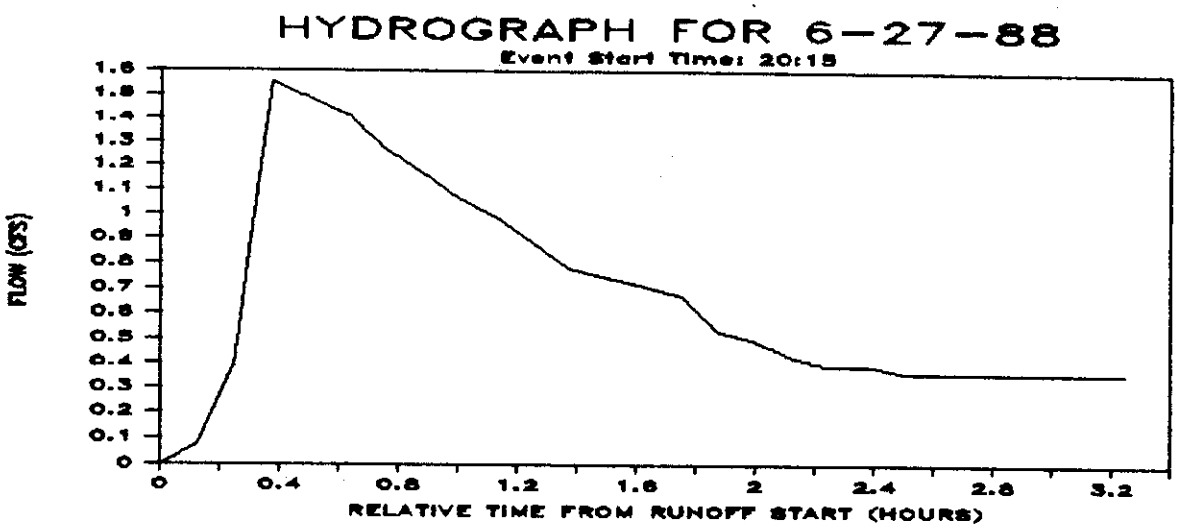


FIGURE A.45
FIELD DATA FOR EVENT OF 6-27-88

TABLE A.1
RUNOFF EVENT DATA

DATE	RUNOFF VOLUME (IN)	PEAK RUNOFF RATE (CFS)	TIME TO PEAK (HRS)	VOLUME TO PEAK (IN)
06/09/87	0.0130	2.027	2.19	0.0041
06/10/87	0.0200	2.823	1.32	0.0032
06/11/87	0.1700	31.742	0.84	0.0170
06/12/87	0.1560	22.063	2.87	0.0410
06/13/87	0.0073	1.516	1.81	0.0018
06/16/87	0.0570	6.477	1.16	0.0088
06/17/87	0.1350	19.942	0.88	0.0160
06/19/87	0.0390	8.737	1.00	0.0100
07/02/87	0.0830	19.667	0.50	0.0077
07/03/87	0.0230	4.359	1.13	0.0046
07/17/87	0.0140	2.777	1.18	0.0023
09/12/87	0.0032	0.420	0.63	0.0005
09/13/87	0.0860	40.350	0.20	0.0120
09/18/87	0.0025	0.390	0.58	0.0004
11/08/87	0.0024	1.717	0.96	0.0020
11/15/87	0.1610	22.086	0.67	0.0160
11/25/87	0.0520	7.565	1.13	0.0090
11/27/87	0.0039	0.643	3.04	0.0011
12/06/87	0.0043	0.681	3.46	0.0012
12/13/87	0.0230	2.910	4.82	0.0055
12/19/87	0.0140	1.599	2.77	0.0067
12/19/87	0.0310	4.265	3.07	0.0097
12/24/87	0.0440	7.325	0.83	0.0033
12/25/87	0.0056	0.457	5.23	0.0045
12/25/87	0.2640	15.681	1.41	0.0250
02/17/88	0.0300	3.168	3.20	0.0110
02/18/88	0.0490	2.549	1.67	0.0086
03/02/88	0.0770	9.421	3.04	0.0260
03/11/88	0.0150	2.957	0.88	0.0012
03/17/88	0.0850	8.177	1.17	0.0110
03/29/88	0.0640	13.869	0.71	0.0059
04/01/88	0.0053	0.916	0.32	0.0007
04/17/88	0.2450	38.846	1.04	0.0470
04/29/88	0.0063	0.684	1.65	0.0026
05/08/88	0.0016	0.713	0.17	0.0002
05/20/88	0.0110	2.357	0.28	0.0006
05/31/88	0.0099	1.876	0.22	0.0006
06/01/88	0.0280	2.244	3.15	0.0063
06/03/88	0.0006	0.138	1.43	0.0004
06/25/88	0.0420	9.938	0.75	0.0094
06/27/88	0.0055	1.519	0.38	0.0005

TABLE A.2
RAINFALL EVENT DATA

DATE	RUNOFF VOLUME (IN)	RECORDING GAGE RAIN VOLUME (IN)	DURATION (HRS)	MAX HOURLY INTENSITY (IN/HR)	WEIGHTED AVG INTENSITY (IN/HR)	TIME BETWEEN RAIN EVENTS (HRS)	DAILY GAGE RAIN VOLUME (IN)	WEIGHTED AVG RAIN VOLUME (IN)	TIME SINCE Q > 0 IN (HRS)	TIME SINCE Q >= .01 IN (HRS)
06/09/87	0.0130	.	.	.	.	168	0.67	0.67	168	168
06/10/87	0.0280	.	.	.	.	20	0.46	0.46	20	20
06/11/87	0.1700	.	.	.	.	31	1.07	1.07	31	31
06/12/87	0.1860	.	.	.	.	12	0.75	0.75	12	12
06/13/87	0.0973	.	.	.	.	27	0.20	0.20	27	27
06/16/87	0.0570	.	.	.	.	60	0.71	0.71	60	88
06/17/87	0.1350	.	.	.	.	36	0.81	0.81	36	36
06/18/87	.	.	.	.	.	.	.	.	.	.
06/19/87	0.0390	.	.	.	.	42	0.51	0.51	42	42
07/02/87	0.0830	1.68	4	1.48	1.81	300	0.81	1.15	300	300
07/03/87	0.0230	0.42	3	0.30	0.24	9	0.20	0.29	9	9
07/16/87	0.0090	0.05	1	0.05	0.05	331	0.00	0.02	331	331
07/17/87	0.0140	0.70	2	0.40	0.36	9	0.68	0.69	341	341
07/23/87	0.0000	0.00	.	.	.	.	0.32	0.19	.	.
08/26/87	0.0000	0.20	8	0.08	0.04	958	0.24	0.22	958	958
08/30/87	0.0000	0.00	.	.	.	.	0.01	0.01	.	.
08/31/87	0.0000	0.00	.	.	.	.	0.06	0.04	.	.
09/07/87	0.0000	0.20	4	0.13	0.11	285	0.37	0.30	1251	1251
09/10/87	0.0000	0.60	5	0.34	0.23	62	0.45	0.51	1317	1317
09/12/87	0.0032	0.93	2	0.86	0.80	46	0.89	0.91	1368	1368
09/13/87	0.0860	1.57	2	0.87	0.79	20	1.56	1.56	20	1390
09/15/87	0.0000	0.08	1	0.08	0.08	53	0.00	0.03	53	53
09/18/87	0.0025	0.49	3	0.24	0.18	63	0.43	0.45	117	117
10/18/87	0.0000	0.35	6	0.16	0.12	726	0.27	0.30	726	846
10/19/87	0.0000	0.05	2	0.03	0.03	15	0.07	0.06	747	867
10/25/87	0.0000	0.00	.	.	.	.	0.04	0.08	.	.
11/08/87	0.0024	1.50	29	0.30	0.14	473	1.58	1.43	1222	1342
11/09/87	0.0000	0.15	3	0.07	0.05	6	0.04	0.14	6	1377
11/15/87	0.1610	2.17	9	0.67	0.43	133	2.39	2.18	142	1513
11/16/87	.	0.33	5	0.17	0.11	11	0.33	0.33	11	11
11/18/87	0.0000	0.10	4	0.05	0.04	54	0.12	0.11	54	54
11/25/87	0.0520	0.93	3	0.80	0.70	147	0.91	0.92	205	205
11/26/87	0.0000	0.00	.	.	.	.	0.09	0.05	.	.
11/27/87	0.0039	0.31	7	0.13	0.09	48	0.33	0.32	48	48
12/06/87	0.0043	0.38	3	0.24	0.18	213	0.30	0.33	213	268
12/13/87	0.0230	0.62	.	.	.	167	0.59	0.60	167	438
12/18/87	0.0000	0.05	2	0.04	0.03	120	0.06	0.06	120	120
12/19/87	0.0140	0.40	7	0.10	0.09	13	0.46	0.44	135	135
12/19/87	0.0310	0.34	3	0.16	0.13	3	0.37	0.36	3	3
12/24/87	0.0440	0.45	5	0.32	0.24	115	0.50	0.48	115	115

TABLE A.2
RAINFALL EVENT DATA

DATE	RUNOFF VOLUME (IN)	RECORDING GAGE RAIN VOLUME (IN)	DURATION (HRS)	MAX HOURLY INTENSITY (IN/HR)	WEIGHTED AVG INTENSITY (IN/HR)	TIME BETWEEN RAIN EVENTS (HRS)	DAILY GAGE RAIN VOLUME (IN)	WEIGHTED AVG RAIN VOLUME (IN)	TIME SINCE Q > 0 IN (HRS)	TIME SINCE Q >= .01 IN (HRS)
12/25/87	0.0056	0.20	6	0.07	0.05	21	0.20	0.20	21	21
12/25/87	0.2640	1.36	10	0.37	0.21	5	1.39	1.38	5	32
12/26/87	.	0.19	8	0.05	0.04	6	0.21	0.20	6	.
12/27/87	.	0.00	.	.	.	.	0.02	0.01	.	.
01/06/88	0.0000	0.75	.	.	.	.	0.82	0.79	.	.
01/16/88	0.0000	0.00	.	.	.	.	0.02	0.01	.	.
02/10/88	0.0000	0.00	.	.	.	.	0.07	0.04	.	.
02/17/88	0.0300	0.62	9	0.20	0.12	1007	0.66	0.64	.	.
02/18/88	0.0490	0.57	12	0.15	0.08	7	0.57	0.57	7	7
03/02/88	0.0770	0.87	7	0.22	0.17	304	0.94	0.91	304	304
03/03/88	0.0000	0.10	2	0.07	0.06	20	0.00	0.04	20	20
03/04/88	0.0000	0.00	.	.	.	.	0.14	0.06	.	.
03/11/88	0.0150	0.29	1	0.29	0.29	197	0.35	0.33	219	219
03/17/88	0.0850	0.87	5	0.56	0.40	135	0.81	0.83	135	135
03/24/88	0.0000	0.05	1	0.05	0.05	174	0.01	0.03	174	174
03/28/88	0.0000	0.05	1	0.05	0.05	85	0.04	0.04	260	260
03/29/88	0.0640	0.81	3	0.57	0.45	18	0.81	0.81	279	279
03/31/88	0.0000	0.08	1	0.08	0.08	65	0.01	0.04	65	65
04/01/88	0.0053	0.18	2	0.10	0.09	10	0.20	0.19	76	76
04/09/88	0.0000	0.00	.	.	.	.	0.02	0.01	.	.
04/10/88	0.0000	0.00	.	.	.	.	0.01	0.01	.	.
04/16/88	0.0000	0.00	.	.	.	.	0.01	0.01	.	.
04/17/88	0.2450	2.01	3	1.13	0.89	381	1.80	1.88	381	459
04/22/88	0.0000	0.00	.	.	.	.	0.05	0.03	.	.
04/24/88	0.0000	0.00	.	.	.	.	0.07	0.04	.	.
04/29/88	0.0063	0.27	2	0.19	0.16	285	0.31	0.29	285	285
04/30/88	0.0000	0.00	.	.	.	.	0.03	0.02	.	.
05/07/88	0.0000	0.00	.	.	.	.	0.03	0.02	.	.
05/08/88	0.0016	0.25	1	0.25	0.25	208	0.25	0.25	208	495
05/16/88	0.0000	0.05	1	0.05	0.05	188	0.04	0.04	188	684
05/20/88	0.0110	0.55	2	0.51	0.48	107	0.72	0.65	296	792
05/20/88	0.0000	0.05	2	0.04	0.03	3	0.07	0.06	3	3
05/29/88	0.0000	0.00	.	.	.	.	0.04	0.02	.	.
05/31/88	0.0099	0.67	2	0.46	0.38	261	0.68	0.68	266	266
06/01/88	0.0280	0.70	7	0.17	0.13	3	0.72	0.71	3	271
06/03/88	0.0006	0.18	4	0.09	0.06	47	0.29	0.25	47	47
06/04/88	0.0000	0.13	2	0.11	0.10	16	0.21	0.18	16	67
06/14/88	0.0000	0.15	2	0.14	0.13	254	0.20	0.18	272	323
06/15/88	0.0000	0.00	.	.	.	.	0.03	0.02	.	.
06/16/88	0.0000	0.05	2	0.03	0.03	48	0.01	0.03	322	373

TABLE A.2
RAINFALL EVENT DATA

DATE	RUNOFF VOLUME (IN)	RECORDING GAGE RAIN VOLUME (IN)	DURATION (HRS)	MAX HOURLY INTENSITY (IN/HR)	WEIGHTED AVG INTENSITY (IN/HR)	TIME BETWEEN RAIN EVENTS (HRS)	DAILY GAGE RAIN VOLUME (IN)	WEIGHTED AVG RAIN VOLUME (IN)	TIME SINCE Q > 0 IN (HRS)	TIME SINCE Q >= .01 IN (HRS)
06/17/88	0.0000	0.17	2	0.16	0.15	13	0.00	0.07	337	388
06/25/88	0.0420	1.44	3	0.78	0.57	199	1.37	1.40	538	589
06/27/88	0.0055	0.33	1	0.33	0.33	43	0.39	0.37	43	43

TABLE A.3
CUMULATIVE WEIGHTED AVERAGE ANTECEDENT RAINFALL VOLUMES

DATE	PREVIOUS 1 DAY (IN)	PREVIOUS 2 DAYS (IN)	PREVIOUS 3 DAYS (IN)	PREVIOUS 4 DAYS (IN)	PREVIOUS 5 DAYS (IN)	PREVIOUS 6 DAYS (IN)	PREVIOUS 7 DAYS (IN)	PREVIOUS 15 DAYS (IN)	PREVIOUS 31 DAYS (IN)
06/09/87	0.00	0.00	0.00	0.00	0.00	0.00	0.28	2.94	4.60
06/10/87	0.67	0.67	0.67	0.67	0.67	0.67	0.67	3.61	5.27
06/11/87	0.00	0.46	1.13	1.13	1.13	1.13	1.13	4.07	5.73
06/12/87	1.07	1.53	2.20	2.20	2.20	2.20	2.20	5.14	6.80
06/13/87	0.00	1.82	2.28	2.95	2.95	2.95	2.95	3.27	7.55
06/16/87	0.00	0.00	0.20	0.95	2.02	2.48	3.15	3.43	7.52
06/17/87	0.00	0.71	0.71	0.91	0.91	2.73	3.19	4.14	7.56
06/19/87	0.00	0.81	0.81	1.52	1.52	1.72	2.47	4.67	8.37
07/02/87	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.32	5.46
07/03/87	1.15	1.15	1.15	1.15	1.15	1.15	1.15	2.47	6.61
07/17/87	0.02	0.02	0.02	0.02	0.02	0.02	0.02	1.46	3.49
09/12/87	0.00	0.51	0.51	0.51	0.81	0.81	0.81	0.81	1.03
09/13/87	0.91	0.91	1.42	1.42	1.42	1.72	1.72	1.72	1.94
09/18/87	0.00	0.00	0.03	0.03	1.59	2.50	2.50	3.31	3.53
11/08/87	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.36
11/15/87	0.00	0.00	0.00	0.00	0.00	1.57	1.57	1.57	1.93
11/25/87	0.00	0.00	0.00	0.00	0.00	0.00	0.11	4.19	4.19
11/27/87	0.00	0.92	0.92	0.92	0.92	0.92	0.92	3.54	5.11
12/06/87	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.24	5.43
12/13/87	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.33	4.19
12/19/87	0.06	0.06	0.06	0.06	0.06	0.66	0.66	0.99	2.34
12/19/87	0.50	0.50	0.50	0.50	0.50	1.10	1.10	1.43	2.78
12/24/87	0.00	0.00	0.00	0.00	0.80	0.86	0.86	1.46	3.03
12/25/87	0.48	0.48	0.48	0.48	0.48	1.28	1.34	1.94	3.51
12/25/87	0.20	0.68	0.68	0.68	0.68	1.48	1.54	2.14	3.71
02/17/88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
02/18/88	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64
03/02/88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.21	1.21
03/11/88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	2.16
03/17/88	0.00	0.00	0.00	0.00	0.00	0.33	0.33	1.28	2.49
03/29/88	0.04	0.04	0.04	0.04	0.07	0.07	0.07	0.90	2.18
04/01/88	0.04	0.04	0.04	0.89	0.89	0.89	0.89	1.75	3.03
04/17/88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.94
04/29/88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.88	2.92
05/08/88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.29	2.17
05/20/88	0.00	0.00	0.00	0.00	0.04	0.04	0.04	0.29	0.58
05/31/88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.71	1.00
06/01/88	0.68	0.68	0.68	0.68	0.68	0.68	0.68	1.39	1.68
06/03/88	0.00	0.71	1.39	1.39	1.39	1.39	1.39	2.10	2.39
06/25/88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.28	2.10
06/27/88	0.00	1.40	1.40	1.40	1.40	1.40	1.40	1.68	3.50

TABLE A.4
EVENT TOTAL TSS LOAD DATA

DATE	RUNOFF VOLUME (IN)	TSS LOAD (LBS)
06/09/87	0.0130	215.05
06/10/87	0.0200	468.25
06/11/87	0.1700	.
06/12/87	0.1560	4908.50
06/13/87	0.0073	59.52
06/16/87	0.0570	1266.48
06/17/87	0.1350	5828.80
06/19/87	0.0390	1125.60
07/02/87	0.0830	.
07/03/87	0.0230	.
07/17/87	0.0140	233.43
09/12/87	0.0032	.
09/13/87	0.0860	4159.26
09/18/87	0.0025	43.29
11/08/87	0.0024	.
11/15/87	0.1610	4892.56
11/25/87	0.0520	1207.73
11/27/87	0.0039	44.80
12/06/87	0.0043	84.32
12/13/87	0.0230	262.02
12/19/87	0.0140	179.10
12/19/87	0.0310	620.95
12/24/87	0.0440	1125.67
12/25/87	0.0056	67.35
12/25/87	0.2640	6129.79
02/17/88	0.0300	225.37
02/18/88	0.0490	.
03/02/88	0.0770	1690.17
03/11/88	0.0150	78.20
03/17/88	0.0850	1031.82
03/29/88	0.0640	3676.72
04/01/88	0.0053	.
04/17/88	0.2450	11748.90
04/29/88	0.0063	.
05/08/88	0.0016	80.14
05/20/88	0.0110	204.52
05/31/88	0.0099	157.40
06/01/88	0.0280	390.70
06/03/88	0.0006	17.83
06/25/88	0.0420	.
06/27/88	0.0055	.

APPENDIX B

HISTORICAL RAINFALL DATA

TABLE B.1
EVENT SEPARATION COMPUTER PROGRAM LISTING

```
1000 CLS
1010 CLOSE
1020 REM*** INITIALIZE COUNTERS AND DATA FLAGS
1030 TOTDAYS%=1
1040 TOTVAL%=0
1050 NUMRAIN%=0
1060 REM*** T% IS UNKNOWN FOR FIRST RAIN, SET GOODLAST$="NO"
1070 GOODLAST$="NO"
1080 GOODDATA$="YES"
1090 REM*** OPEN OUTPUT FILE.
1100 PRINT "ENTER SEQUENTIAL DATA FILE NAME"
1110 INPUT "FOR STORAGE OF PROGRAM OUTPUT";NAMEOUTF$
1120 OPEN NAMEOUTF$ FOR OUTPUT AS #2
1130 REM*** NEXTFILE:
1140 CLOSE #1
1150 INPUT "ENTER NEXT INPUT SEQUENTIAL DATA FILE NAME (QUIT TO STOP)";NAMEF$
1160 REM*** IF NAMEF$<>"QUIT" THEN GOTO NOQUIT.
1170 IF NAMEF$<>"QUIT" THEN GOTO 1200
1180 CLOSE #2
1190 END
1200 REM*** NOQUIT:
1210 OPEN NAMEF$ FOR INPUT AS #1
1220 INPUT #1,LINE1%,RECTYP$
1230 INPUT #1,LINE2%,STNLOC$
1240 INPUT #1,LINE3%,STNID1,STNID2,Z1,Z2,Z3,Z4,Z5,Z6,Z7,Z8
1250 INPUT #1,LINE4%,BEGINYR%,BEGINMTH%,ENDYR%,ENDMTH%
1260 INPUT #1,LINE5%,ELMTYP$,B1$,B2$,B3$,B4$
1270 REM*** CHECK FOR FIRST DAY OF RECORD.
1280 REM*** IF TOTDAYS%>1 THEN GOTO NEXTDAY.
1290 IF TOTDAYS%>1 THEN GOTO 1410
1300 REM*** INPUT FIRST DAY OF RECORD
1310 INPUT #1,FLAG3%,YEAR%,MONTH%,DAY%,UNITS$,NUMVAL%
1320 REM*** INITIALIZE VARIABLES
1330 REM*** GOSUB YEARDAY.
1340 GOSUB 3120
1350 YEARLAST%=YEAR%
1360 MNTHLAST%=MONTH%
1370 DAYLAST%=DAY%
1380 HOURLAST%=0
1390 REM*** GOTO FIRSTDAY.
1400 GOTO 1650
1410 REM*** NEXTDAY:
1420 INPUT #1,FLAG3%
1430 REM*** FLAG3%=10 INDICATES END OF FILE, GOTO FINISH.
1440 IF FLAG3%=10 THEN GOTO 2530
1450 INPUT #1,YEAR%,MONTH%,DAY%,UNITS$,NUMVAL%
1460 TOTDAYS%=TOTDAYS%+1
1470 REM*** GOSUB YEARDAY.
1480 GOSUB 3120
1490 REM*** IF YEAR%=YEARLAST% THEN GOTO SAMEYEAR.
1500 IF YEAR%=YEARLAST% THEN GOTO 1610
1510 REM*** GOSUB LASTYEAR.
1520 GOSUB 2600
1530 REM** HRS. ELAPSED SINCE END OF LAST RAIN IN LAST YEAR UP TO 24:00 DEC. 31
1540 HRLASTYR%=DLASTYR%*24+24-HOURLAST%
1550 REM*** HRS. ELAPSED SINCE 00:00 JAN. 1 THIS YR. UP TO 00:00 TODAY.
1560 HRTHISYR%=(YRDAY%-1)*24
1570 REM** TOTAL HRS. ELAPSED SINCE END OF LAST RAIN LAST YR. UP TO 00:00 TODAY
1580 HRPASTD%=HRLASTYR%+HRTHISYR%
1590 REM*** GOTO TODAY.
```

TABLE B.1
EVENT SEPARATION COMPUTER PROGRAM LISTING

```
1600 GOTO 1660
1610 REM*** SAMEYEAR:
1620 REM*** HRS. ELAPSED SINCE THE LAST HOUR OF RAIN IN A PREVIOUS DAY
1630 REM*** THIS YEAR UP TO 00:00 TODAY.
1640 HRPASTD%=(YRDAY%-YRDLAST%-1)*24+24-HOURLAST%
1650 REM*** FIRSTDAY:
1660 REM*** TODAY:
1670 REM***
1680 REM*** HOURLY DATA LOOP
1690 REM***
1700 REM*** NOTE:  TIME OF END OF CURRENT HR. OF RAIN IS HOUR% AND
1710 REM*** TIME OF START OF CURRENT HR. OF RAIN IS (HOUR%-1).
1720 NUMWETHR%=0
1730 FOR I=1 TO NUMVAL%
1740 INPUT #1,FLAG3%,HOUR%,RAIN,FLAG1$,FLAG2$
1750 TOTVAL%=TOTVAL%+1
1760 REM*** CHECK FOR FIRST HOUR OF RECORD.
1770 REM*** IF TOTVAL%=1 THEN GOTO FIRSTHOUR.
1780 IF TOTVAL%=1 THEN GOTO 2150
1790 REM*** SKIP DAILY TOTAL VALUES.
1800 REM*** IF HOUR%=25 THEN GOTO SKIPDATA.
1810 IF HOUR%=25 THEN GOTO 2470
1820 REM*** SKIP HOURS WITH GOOD DATA (I.E. FLAG1$="0") WHEN RAIN=0.
1830 REM*** BUT DO NOT SKIP HRS. WITH BAD DATA SINCE T% IS AFFECTED.
1840 REM*** IF FLAG1$<>"0" THEN GOTO NOSKIP.
1850 IF FLAG1$<>"0" THEN GOTO 1880
1860 REM*** IF RAIN=0 THEN GOTO SKIPDATA.
1870 IF RAIN=0 THEN GOTO 2470
1880 REM*** NOSKIP:
1890 REM*** NUMBER OF HRS. TODAY WITH RAIN>0 OR BAD DATA.
1900 NUMWETHR%=NUMWETHR%+1
1910 REM*** HRS. SINCE THE END OF THE LAST HR. OF RAIN IN A PREVIOUS DAY TO THE
1920 REM*** START OF THE FIRST HR. OF RAIN TODAY.
1930 IF NUMWETHR%=1 THEN HRSINCE%=HRPASTD%+(HOUR%-1)
1940 REM*** HRS. SINCE THE END OF THE PREVIOUS HR. OF RAIN TODAY TO THE
1950 REM*** START OF THE CURRENT HR. OF RAIN TODAY.
1960 IF NUMWETHR%>1 THEN HRSINCE%=(HOUR%-1)-HOURLAST%
1970 REM*** IF HRSINCE%<3 THEN GOTO SAMERAIN.
1980 IF HRSINCE%<3 THEN GOTO 2300
1990 REM*** NEW RAIN STORM.
2000 REM*** PRINT RESULTS OF LAST RAIN BEFORE RESETTING VARIABLES.
2010 REM*** CHECK FOR BAD DATA.
2020 REM*** IF GOODDATA$="YES" THEN GOTO PRNTDATA.
2030 IF GOODDATA$="YES" THEN GOTO 2090
2040 REM*** R, D%, IMH, IW ARE UNCERTAIN BECAUSE GOODDATA$="NO".
2050 R=-1
2060 D%=-1
2070 IMH=-1
2080 IW=-1
2090 REM*** PRNTDATA:
2100 WRITE #2,NUMRAIN%,T%,R,D%,IMH,IW
2110 REM*** RESET GOODDATA$ FOR NEXT RAIN.
2120 GOODDATA$="YES"
2130 REM*** RESET VARIABLES FOR NEW RAIN.
2140 T%=HRSINCE%
2150 REM*** FIRSTHOUR:
2160 REM*** T% IS UNCERTAIN IF THE LAST HR. OF THE PREVIOUS RAIN HAS
2170 REM*** BAD DATA (I.E. GOODLAST$="NO").
2180 IF GOODLAST$="NO" THEN T%=-1
2190 REM*** T% IS UNCERTAIN IF THE FIRST HR. OF THE CURRENT
```

TABLE B.1
EVENT SEPARATION COMPUTER PROGRAM LISTING

```
2200 REM*** RAIN HAS BAD DATA (I.E. FLAG1$<>"0").
2210 IF FLAG1$<>"0" THEN T%=-1
2220 R=RAIN/100
2230 D%=1
2240 IMH=R
2250 IW=R
2260 RSQRSUM=R^2
2270 NUMRAIN%=NUMRAIN%+1
2280 REM*** GOTO NEXTHOUR
2290 GOTO 2360
2300 REM*** SAMERAIN:
2310 R=R+RAIN/100
2320 D%=D%+HRSINCE%+1
2330 IF RAIN/100>IMH THEN IMH=RAIN/100
2340 RSQRSUM=RSQRSUM+(RAIN/100)^2
2350 IF R=0 THEN IW=0 ELSE IW=RSQRSUM/R
2360 REM*** NEXTHOUR:
2370 REM*** RESET VARIABLES FOR PREVIOUS HOUR.
2380 YEARLAST%=YEAR%
2390 MNTHLAST%=MONTH%
2400 DAYLAST%=DAY%
2410 HOURLAST%=HOUR%
2420 YRDLAST%=YRDAY%
2430 REM*** GOODDATA$="NO" IF ANY HR. DURING THE RAIN HAS FLAG1$<>"0".
2440 IF FLAG1$<>"0" THEN GOODDATA$="NO"
2450 REM*** GOODLAST$="YES" IF CURRENT FLAG1$="0" AND GOODLAST$="NO" OTHERWISE.
2460 IF FLAG1$="0" THEN GOODLAST$="YES" ELSE GOODLAST$="NO"
2470 REM*** SKIPDATA:
2480 NEXT I
2490 REM*** IF EOF(1) THEN GOTO FINISH.
2500 IF EOF(1) THEN GOTO 2530
2510 REM*** GOTO NEXTDAY
2520 GOTO 1410
2530 REM*** FINISH:
2540 PRINT TOTDAYS%,TOTVAL%,NUMRAIN%
2550 REM*** GOTO NEXTFILE
2560 GOTO 1130
2570 END
2580 REM***
2590 REM*** SUBROUTINE LASTYEAR.
2600 REM*** LASTYEAR:
2610 IF MNTHLAST%<12 THEN GOTO 2640
2620 DLASTYR%=31-DAYLAST%
2630 RETURN
2640 REM*** NOV:
2650 IF MNTHLAST%<11 THEN GOTO 2680
2660 DLASTYR%=61-DAYLAST%
2670 RETURN
2680 REM*** OCT:
2690 IF MNTHLAST%<10 THEN GOTO 2720
2700 DLASTYR%=92-DAYLAST%
2710 RETURN
2720 REM*** SEP:
2730 IF MNTHLAST%<9 THEN GOTO 2760
2740 DLASTYR%=122-DAYLAST%
2750 RETURN
2760 REM*** AUG:
2770 IF MNTHLAST%<8 THEN GOTO 2800
2780 DLASTYR%=153-DAYLAST%
2790 RETURN
```

TABLE B.1
EVENT SEPARATION COMPUTER PROGRAM LISTING

```
2800 REM*** JUL:
2810 IF MNTHLAST%<7 THEN GOTO 2840
2820 DLASTYR%=184-DAYLAST%
2830 RETURN
2840 REM*** JUN:
2850 IF MNTHLAST%<6 THEN GOTO 2880
2860 DLASTYR%=214-DAYLAST%
2870 RETURN
2880 REM*** MAY:
2890 IF MNTHLAST%<5 GOTO 2920
2900 DLASTYR%=245-DAYLAST%
2910 RETURN
2920 REM*** APR:
2930 IF MNTHLAST%<4 THEN GOTO 2960
2940 DLASTYR%=275-DAYLAST%
2950 RETURN
2960 REM*** MAR:
2970 IF MNTHLAST%<3 THEN GOTO 3000
2980 DLASTYR%=306-DAYLAST%
2990 RETURN
3000 REM*** FEB:
3010 FEBDAYS%=28
3020 IF INT((YEARLAST%-1940)/4)=(YEARLAST%-1940)/4 THEN FEBDAYS%=29
3030 IF MNTHLAST%<2 THEN GOTO 3060
3040 DLASTYR%=306+FEBDAYS%-DAYLAST%
3050 RETURN
3060 REM*** JAN:
3070 DLASTYR%=337+FEBDAYS%-DAYLAST%
3080 RETURN
3090 END
3100 REM***
3110 REM*** SUBROUTINE YEARDAY
3120 REM*** YEARDAY:
3130 IF MONTH%>1 THEN GOTO 3160
3140 YRDAY%=DAY%
3150 RETURN
3160 REM*** FEB2:
3170 IF MONTH%>2 THEN GOTO 3200
3180 YRDAY%=31+DAY%
3190 RETURN
3200 REM*** MAR2:
3210 FEBDAYS%=28
3220 IF INT((YEAR%-1940)/4)=(YEAR%-1940)/4 THEN FEBDAYS%=29
3230 IF MONTH%>3 THEN GOTO 3260
3240 YRDAY%=31+FEBDAYS%+DAY%
3250 RETURN
3260 REM*** APR2:
3270 IF MONTH%>4 THEN GOTO 3300
3280 YRDAY%=62+FEBDAYS%+DAY%
3290 RETURN
3300 REM*** MAY2:
3310 IF MONTH%>5 THEN GOTO 3340
3320 YRDAY%=92+FEBDAYS%+DAY%
3330 RETURN
3340 REM*** JUN2:
3350 IF MONTH%>6 THEN GOTO 3380
3360 YRDAY%=123+FEBDAYS%+DAY%
3370 RETURN
3380 REM*** JUL2:
3390 IF MONTH%>7 THEN GOTO 3420
```

TABLE B.1
EVENT SEPARATION COMPUTER PROGRAM LISTING

```
3400 YRDAY%=153+FEBDAYS%+DAY%
3410 RETURN
3420 REM*** AUG2:
3430 IF MONTH%>8 THEN GOTO 3460
3440 YRDAY%=184+FEBDAYS%+DAY%
3450 RETURN
3460 REM*** SEP2:
3470 IF MONTH%>9 THEN GOTO 3500
3480 YRDAY%=215+FEBDAYS%+DAY%
3490 RETURN
3500 REM*** OCT2:
3510 IF MONTH%>10 THEN GOTO 3540
3520 YRDAY%=245+FEBDAYS%+DAY%
3530 RETURN
3540 REM*** NOV2:
3550 IF MONTH%>11 THEN GOTO 3580
3560 YRDAY%=276+FEBDAYS%+DAY%
3570 RETURN
3580 REM*** DEC2:
3590 YRDAY%=306+FEBDAYS%+DAY%
3600 RETURN
3610 END
```

TABLE B.2
FREQUENCY ANALYSIS FOR RAINFALL VOLUME

R	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
0.01	435	15.8	435	15.8
0.02	225	8.2	660	23.9
0.03	131	4.7	791	28.7
0.04	119	4.3	910	33.0
0.05	110	4.0	1020	37.0
0.06	72	2.6	1092	39.6
0.07	60	2.2	1152	41.7
0.08	35	1.3	1187	43.0
0.09	60	2.2	1247	45.2
0.10	59	2.1	1306	47.3
0.11	39	1.4	1345	48.7
0.12	42	1.5	1387	50.3
0.13	41	1.5	1428	51.7
0.14	23	0.8	1451	52.6
0.15	40	1.4	1491	54.0
0.16	27	1.0	1518	55.0
0.17	23	0.8	1541	55.8
0.18	24	0.9	1565	56.7
0.19	24	0.9	1589	57.6
0.20	37	1.3	1626	58.9
0.21	26	0.9	1652	59.9
0.22	12	0.4	1664	60.3
0.23	21	0.8	1685	61.1
0.24	15	0.5	1700	61.6
0.25	26	0.9	1726	62.5
0.26	23	0.8	1749	63.4
0.27	22	0.8	1771	64.2
0.28	24	0.9	1795	65.0
0.29	14	0.5	1809	65.5
0.30	27	1.0	1836	66.5
0.31	24	0.9	1860	67.4
0.32	18	0.7	1878	68.0
0.33	10	0.4	1888	68.4
0.34	11	0.4	1899	68.8
0.35	20	0.7	1919	69.5
0.36	10	0.4	1929	69.9
0.37	12	0.4	1941	70.3
0.38	14	0.5	1955	70.8
0.39	16	0.6	1971	71.4
0.40	19	0.7	1990	72.1
0.41	9	0.3	1999	72.4
0.42	15	0.5	2014	73.0
0.43	11	0.4	2025	73.4
0.44	8	0.3	2033	73.7
0.45	14	0.5	2047	74.2
0.46	7	0.3	2054	74.4

TABLE B.2
FREQUENCY ANALYSIS FOR RAINFALL VOLUME

R	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
0.47	14	0.5	2068	74.9
0.48	10	0.4	2078	75.3
0.49	6	0.2	2084	75.5
0.50	12	0.4	2096	75.9
0.51	10	0.4	2106	76.3
0.52	9	0.3	2115	76.6
0.53	11	0.4	2126	77.0
0.54	9	0.3	2135	77.4
0.55	18	0.7	2153	78.0
0.56	11	0.4	2164	78.4
0.57	14	0.5	2178	78.9
0.58	9	0.3	2187	79.2
0.59	10	0.4	2197	79.6
0.60	14	0.5	2211	80.1
0.61	1	0.0	2212	80.1
0.62	10	0.4	2222	80.5
0.63	4	0.1	2226	80.7
0.64	6	0.2	2232	80.9
0.65	7	0.3	2239	81.1
0.66	9	0.3	2248	81.4
0.67	6	0.2	2254	81.7
0.68	10	0.4	2264	82.0
0.69	12	0.4	2276	82.5
0.70	9	0.3	2285	82.8
0.71	4	0.1	2289	82.9
0.72	3	0.1	2292	83.0
0.73	7	0.3	2299	83.3
0.74	5	0.2	2304	83.5
0.75	9	0.3	2313	83.8
0.76	7	0.3	2320	84.1
0.77	5	0.2	2325	84.2
0.78	9	0.3	2334	84.6
0.79	8	0.3	2342	84.9
0.80	8	0.3	2350	85.1
0.81	6	0.2	2356	85.4
0.82	11	0.4	2367	85.8
0.83	5	0.2	2372	85.9
0.84	6	0.2	2378	86.2
0.85	5	0.2	2383	86.3
0.86	5	0.2	2388	86.5
0.87	6	0.2	2394	86.7
0.88	6	0.2	2400	87.0
0.89	3	0.1	2403	87.1
0.90	2	0.1	2405	87.1
0.91	1	0.0	2406	87.2
0.92	6	0.2	2412	87.4

TABLE B.2
FREQUENCY ANALYSIS FOR RAINFALL VOLUME

R	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
0.93	3	0.1	2415	87.5
0.94	6	0.2	2421	87.7
0.95	5	0.2	2426	87.9
0.96	6	0.2	2432	88.1
0.97	5	0.2	2437	88.3
0.98	2	0.1	2439	88.4
0.99	3	0.1	2442	88.5
1.00	10	0.4	2452	88.8
1.01	2	0.1	2454	88.9
1.02	4	0.1	2458	89.1
1.03	2	0.1	2460	89.1
1.04	4	0.1	2464	89.3
1.05	1	0.0	2465	89.3
1.06	2	0.1	2467	89.4
1.07	3	0.1	2470	89.5
1.08	5	0.2	2475	89.7
1.09	1	0.0	2476	89.7
1.10	5	0.2	2481	89.9
1.11	2	0.1	2483	90.0
1.12	4	0.1	2487	90.1
1.13	2	0.1	2489	90.2
1.14	3	0.1	2492	90.3
1.15	2	0.1	2494	90.4
1.16	2	0.1	2496	90.4
1.17	2	0.1	2498	90.5
1.18	7	0.3	2505	90.8
1.19	2	0.1	2507	90.8
1.20	7	0.3	2514	91.1
1.21	6	0.2	2520	91.3
1.22	3	0.1	2523	91.4
1.23	5	0.2	2528	91.6
1.24	1	0.0	2529	91.6
1.25	4	0.1	2533	91.8
1.26	6	0.2	2539	92.0
1.27	2	0.1	2541	92.1
1.28	4	0.1	2545	92.2
1.29	4	0.1	2549	92.4
1.30	5	0.2	2554	92.5
1.31	2	0.1	2556	92.6
1.32	3	0.1	2559	92.7
1.33	2	0.1	2561	92.8
1.34	1	0.0	2562	92.8
1.35	2	0.1	2564	92.9
1.36	3	0.1	2567	93.0
1.37	4	0.1	2571	93.2
1.38	1	0.0	2572	93.2

TABLE B.2
FREQUENCY ANALYSIS FOR RAINFALL VOLUME

R	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
1.39	3	0.1	2575	93.3
1.40	3	0.1	2578	93.4
1.41	2	0.1	2580	93.5
1.42	1	0.0	2581	93.5
1.43	1	0.0	2582	93.6
1.44	3	0.1	2585	93.7
1.45	2	0.1	2587	93.7
1.46	2	0.1	2589	93.8
1.47	3	0.1	2592	93.9
1.48	1	0.0	2593	93.9
1.49	1	0.0	2594	94.0
1.50	3	0.1	2597	94.1
1.51	1	0.0	2598	94.1
1.52	1	0.0	2599	94.2
1.53	3	0.1	2602	94.3
1.54	1	0.0	2603	94.3
1.55	1	0.0	2604	94.3
1.56	3	0.1	2607	94.5
1.59	1	0.0	2608	94.5
1.60	1	0.0	2609	94.5
1.61	1	0.0	2610	94.6
1.64	2	0.1	2612	94.6
1.65	2	0.1	2614	94.7
1.66	1	0.0	2615	94.7
1.67	2	0.1	2617	94.8
1.68	3	0.1	2620	94.9
1.69	1	0.0	2621	95.0
1.70	1	0.0	2622	95.0
1.71	1	0.0	2623	95.0
1.72	2	0.1	2625	95.1
1.73	2	0.1	2627	95.2
1.74	1	0.0	2628	95.2
1.75	2	0.1	2630	95.3
1.76	2	0.1	2632	95.4
1.77	4	0.1	2636	95.5
1.78	2	0.1	2638	95.6
1.79	1	0.0	2639	95.6
1.80	3	0.1	2642	95.7
1.81	1	0.0	2643	95.8
1.82	1	0.0	2644	95.8
1.83	1	0.0	2645	95.8
1.86	1	0.0	2646	95.9
1.87	1	0.0	2647	95.9
1.90	2	0.1	2649	96.0
1.91	1	0.0	2650	96.0
1.92	1	0.0	2651	96.1

TABLE B.2
FREQUENCY ANALYSIS FOR RAINFALL VOLUME

R	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
1.94	2	0.1	2653	96.1
1.95	1	0.0	2654	96.2
1.96	3	0.1	2657	96.3
1.97	1	0.0	2658	96.3
1.98	1	0.0	2659	96.3
2.00	7	0.3	2666	96.6
2.01	1	0.0	2667	96.6
2.02	1	0.0	2668	96.7
2.03	1	0.0	2669	96.7
2.04	1	0.0	2670	96.7
2.05	2	0.1	2672	96.8
2.06	1	0.0	2673	96.8
2.09	1	0.0	2674	96.9
2.13	1	0.0	2675	96.9
2.14	3	0.1	2678	97.0
2.15	1	0.0	2679	97.1
2.18	3	0.1	2682	97.2
2.20	1	0.0	2683	97.2
2.23	1	0.0	2684	97.2
2.25	1	0.0	2685	97.3
2.27	2	0.1	2687	97.4
2.32	1	0.0	2688	97.4
2.33	1	0.0	2689	97.4
2.34	2	0.1	2691	97.5
2.35	2	0.1	2693	97.6
2.36	1	0.0	2694	97.6
2.38	2	0.1	2696	97.7
2.39	2	0.1	2698	97.8
2.42	1	0.0	2699	97.8
2.43	1	0.0	2700	97.8
2.44	1	0.0	2701	97.9
2.46	1	0.0	2702	97.9
2.48	1	0.0	2703	97.9
2.50	1	0.0	2704	98.0
2.52	1	0.0	2705	98.0
2.57	1	0.0	2706	98.0
2.58	1	0.0	2707	98.1
2.60	1	0.0	2708	98.1
2.61	1	0.0	2709	98.2
2.63	1	0.0	2710	98.2
2.65	2	0.1	2712	98.3
2.67	1	0.0	2713	98.3
2.68	1	0.0	2714	98.3
2.71	1	0.0	2715	98.4
2.74	1	0.0	2716	98.4
2.75	1	0.0	2717	98.4

TABLE B.2
FREQUENCY ANALYSIS FOR RAINFALL VOLUME

R	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
2.78	1	0.0	2718	98.5
2.79	1	0.0	2719	98.5
2.80	1	0.0	2720	98.6
2.84	1	0.0	2721	98.6
2.85	2	0.1	2723	98.7
2.93	2	0.1	2725	98.7
2.95	1	0.0	2726	98.8
3.01	1	0.0	2727	98.8
3.02	2	0.1	2729	98.9
3.04	1	0.0	2730	98.9
3.05	1	0.0	2731	98.9
3.07	1	0.0	2732	99.0
3.10	2	0.1	2734	99.1
3.18	1	0.0	2735	99.1
3.22	1	0.0	2736	99.1
3.25	1	0.0	2737	99.2
3.29	1	0.0	2738	99.2
3.35	1	0.0	2739	99.2
3.36	1	0.0	2740	99.3
3.38	1	0.0	2741	99.3
3.40	2	0.1	2743	99.4
3.55	1	0.0	2744	99.4
3.79	1	0.0	2745	99.5
3.92	1	0.0	2746	99.5
3.96	1	0.0	2747	99.5
4.04	1	0.0	2748	99.6
4.14	1	0.0	2749	99.6
4.31	1	0.0	2750	99.6
4.41	1	0.0	2751	99.7
4.54	1	0.0	2752	99.7
4.63	1	0.0	2753	99.7
4.89	1	0.0	2754	99.8
4.96	1	0.0	2755	99.8
5.16	1	0.0	2756	99.9
5.42	1	0.0	2757	99.9
5.82	1	0.0	2758	99.9
5.96	1	0.0	2759	100.0
6.17	1	0.0	2760	100.0

TABLE B.3
FREQUENCY ANALYSIS FOR TIME BETWEEN EVENTS

T	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
3	231	8.4	231	8.4
4	156	5.7	387	14.0
5	118	4.3	505	18.3
6	87	3.2	592	21.4
7	73	2.6	665	24.1
8	64	2.3	729	26.4
9	49	1.8	778	28.2
10	29	1.1	807	29.2
11	49	1.8	856	31.0
12	49	1.8	905	32.8
13	30	1.1	935	33.9
14	27	1.0	962	34.9
15	36	1.3	998	36.2
16	34	1.2	1032	37.4
17	33	1.2	1065	38.6
18	25	0.9	1090	39.5
19	26	0.9	1116	40.4
20	24	0.9	1140	41.3
21	24	0.9	1164	42.2
22	24	0.9	1188	43.0
23	14	0.5	1202	43.6
24	11	0.4	1213	43.9
25	19	0.7	1232	44.6
26	19	0.7	1251	45.3
27	11	0.4	1262	45.7
28	16	0.6	1278	46.3
29	14	0.5	1292	46.8
30	18	0.7	1310	47.5
31	16	0.6	1326	48.0
32	20	0.7	1346	48.8
33	8	0.3	1354	49.1
34	12	0.4	1366	49.5
35	15	0.5	1381	50.0
36	13	0.5	1394	50.5
37	10	0.4	1404	50.9
38	8	0.3	1412	51.2
39	10	0.4	1422	51.5
40	10	0.4	1432	51.9
41	11	0.4	1443	52.3
42	12	0.4	1455	52.7
43	18	0.7	1473	53.4
44	13	0.5	1486	53.8
45	13	0.5	1499	54.3
46	12	0.4	1511	54.7
47	9	0.3	1520	55.1
48	13	0.5	1533	55.5

TABLE B.3
FREQUENCY ANALYSIS FOR TIME BETWEEN EVENTS

T	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
49	10	0.4	1543	55.9
50	10	0.4	1553	56.3
51	10	0.4	1563	56.6
52	14	0.5	1577	57.1
53	10	0.4	1587	57.5
54	11	0.4	1598	57.9
55	9	0.3	1607	58.2
56	13	0.5	1620	58.7
57	11	0.4	1631	59.1
58	9	0.3	1640	59.4
59	7	0.3	1647	59.7
60	14	0.5	1661	60.2
61	13	0.5	1674	60.7
62	11	0.4	1685	61.1
63	13	0.5	1698	61.5
64	8	0.3	1706	61.8
65	9	0.3	1715	62.1
66	6	0.2	1721	62.4
67	11	0.4	1732	62.8
68	12	0.4	1744	63.2
69	10	0.4	1754	63.6
70	8	0.3	1762	63.8
71	8	0.3	1770	64.1
72	6	0.2	1776	64.3
73	3	0.1	1779	64.5
74	6	0.2	1785	64.7
75	6	0.2	1791	64.9
76	11	0.4	1802	65.3
77	8	0.3	1810	65.6
78	8	0.3	1818	65.9
79	10	0.4	1828	66.2
80	6	0.2	1834	66.4
81	4	0.1	1838	66.6
82	11	0.4	1849	67.0
83	2	0.1	1851	67.1
84	6	0.2	1857	67.3
85	6	0.2	1863	67.5
86	8	0.3	1871	67.8
87	5	0.2	1876	68.0
88	6	0.2	1882	68.2
89	9	0.3	1891	68.5
90	5	0.2	1896	68.7
91	2	0.1	1898	68.8
92	10	0.4	1908	69.1
93	6	0.2	1914	69.3
94	11	0.4	1925	69.7

TABLE B.3
FREQUENCY ANALYSIS FOR TIME BETWEEN EVENTS

T	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
95	8	0.3	1933	70.0
96	10	0.4	1943	70.4
97	3	0.1	1946	70.5
98	7	0.3	1953	70.8
99	6	0.2	1959	71.0
100	4	0.1	1963	71.1
101	4	0.1	1967	71.3
102	4	0.1	1971	71.4
103	4	0.1	1975	71.6
104	4	0.1	1979	71.7
105	4	0.1	1983	71.8
106	3	0.1	1986	72.0
107	6	0.2	1992	72.2
108	5	0.2	1997	72.4
109	4	0.1	2001	72.5
110	5	0.2	2006	72.7
111	5	0.2	2011	72.9
112	10	0.4	2021	73.2
113	7	0.3	2028	73.5
114	9	0.3	2037	73.8
115	5	0.2	2042	74.0
116	4	0.1	2046	74.1
117	2	0.1	2048	74.2
118	4	0.1	2052	74.3
119	3	0.1	2055	74.5
120	5	0.2	2060	74.6
121	5	0.2	2065	74.8
122	4	0.1	2069	75.0
123	8	0.3	2077	75.3
124	5	0.2	2082	75.4
125	5	0.2	2087	75.6
126	6	0.2	2093	75.8
127	5	0.2	2098	76.0
128	4	0.1	2102	76.2
129	2	0.1	2104	76.2
130	3	0.1	2107	76.3
131	7	0.3	2114	76.6
132	11	0.4	2125	77.0
133	4	0.1	2129	77.1
134	7	0.3	2136	77.4
135	5	0.2	2141	77.6
136	3	0.1	2144	77.7
137	2	0.1	2146	77.8
138	3	0.1	2149	77.9
139	3	0.1	2152	78.0
140	4	0.1	2156	78.1

TABLE B.3
FREQUENCY ANALYSIS FOR TIME BETWEEN EVENTS

T	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
141	4	0.1	2160	78.3
142	3	0.1	2163	78.4
143	4	0.1	2167	78.5
144	2	0.1	2169	78.6
145	5	0.2	2174	78.8
146	1	0.0	2175	78.8
147	5	0.2	2180	79.0
148	3	0.1	2183	79.1
149	6	0.2	2189	79.3
150	3	0.1	2192	79.4
151	4	0.1	2196	79.6
152	3	0.1	2199	79.7
153	1	0.0	2200	79.7
154	7	0.3	2207	80.0
155	4	0.1	2211	80.1
156	2	0.1	2213	80.2
157	3	0.1	2216	80.3
158	2	0.1	2218	80.4
159	2	0.1	2220	80.4
160	6	0.2	2226	80.7
161	3	0.1	2229	80.8
162	2	0.1	2231	80.8
163	7	0.3	2238	81.1
164	2	0.1	2240	81.2
165	1	0.0	2241	81.2
166	3	0.1	2244	81.3
167	4	0.1	2248	81.4
168	2	0.1	2250	81.5
169	4	0.1	2254	81.7
170	6	0.2	2260	81.9
171	1	0.0	2261	81.9
172	2	0.1	2263	82.0
173	1	0.0	2264	82.0
174	2	0.1	2266	82.1
175	6	0.2	2272	82.3
176	4	0.1	2276	82.5
177	5	0.2	2281	82.6
178	3	0.1	2284	82.8
179	2	0.1	2286	82.8
180	4	0.1	2290	83.0
181	1	0.0	2291	83.0
182	5	0.2	2296	83.2
183	2	0.1	2298	83.3
184	1	0.0	2299	83.3
185	6	0.2	2305	83.5
186	2	0.1	2307	83.6

TABLE B.3
FREQUENCY ANALYSIS FOR TIME BETWEEN EVENTS

T	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
187	2	0.1	2309	83.7
188	3	0.1	2312	83.8
189	4	0.1	2316	83.9
190	5	0.2	2321	84.1
191	5	0.2	2326	84.3
193	6	0.2	2332	84.5
194	5	0.2	2337	84.7
195	6	0.2	2343	84.9
196	1	0.0	2344	84.9
197	4	0.1	2348	85.1
199	3	0.1	2351	85.2
200	1	0.0	2352	85.2
201	4	0.1	2356	85.4
202	2	0.1	2358	85.4
203	2	0.1	2360	85.5
204	1	0.0	2361	85.5
205	1	0.0	2362	85.6
206	1	0.0	2363	85.6
207	3	0.1	2366	85.7
208	2	0.1	2368	85.8
209	2	0.1	2370	85.9
210	2	0.1	2372	85.9
211	3	0.1	2375	86.1
212	2	0.1	2377	86.1
214	5	0.2	2382	86.3
215	1	0.0	2383	86.3
216	2	0.1	2385	86.4
217	3	0.1	2388	86.5
218	2	0.1	2390	86.6
219	2	0.1	2392	86.7
221	2	0.1	2394	86.7
222	4	0.1	2398	86.9
223	1	0.0	2399	86.9
224	1	0.0	2400	87.0
226	3	0.1	2403	87.1
227	1	0.0	2404	87.1
228	3	0.1	2407	87.2
229	7	0.3	2414	87.5
231	3	0.1	2417	87.6
232	4	0.1	2421	87.7
233	4	0.1	2425	87.9
234	3	0.1	2428	88.0
235	1	0.0	2429	88.0
236	2	0.1	2431	88.1
237	1	0.0	2432	88.1
240	5	0.2	2437	88.3

TABLE B.3
FREQUENCY ANALYSIS FOR TIME BETWEEN EVENTS

T	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
241	1	0.0	2438	88.3
242	1	0.0	2439	88.4
243	3	0.1	2442	88.5
244	2	0.1	2444	88.6
245	2	0.1	2446	88.6
246	2	0.1	2448	88.7
248	1	0.0	2449	88.7
249	2	0.1	2451	88.8
250	1	0.0	2452	88.8
251	3	0.1	2455	88.9
252	2	0.1	2457	89.0
253	3	0.1	2460	89.1
254	1	0.0	2461	89.2
255	3	0.1	2464	89.3
256	2	0.1	2466	89.3
257	1	0.0	2467	89.4
258	2	0.1	2469	89.5
259	1	0.0	2470	89.5
260	3	0.1	2473	89.6
261	3	0.1	2476	89.7
262	1	0.0	2477	89.7
263	4	0.1	2481	89.9
264	1	0.0	2482	89.9
265	2	0.1	2484	90.0
266	1	0.0	2485	90.0
267	2	0.1	2487	90.1
269	1	0.0	2488	90.1
271	2	0.1	2490	90.2
272	1	0.0	2491	90.3
273	2	0.1	2493	90.3
274	1	0.0	2494	90.4
275	4	0.1	2498	90.5
276	2	0.1	2500	90.6
277	4	0.1	2504	90.7
278	3	0.1	2507	90.8
280	2	0.1	2509	90.9
281	2	0.1	2511	91.0
283	1	0.0	2512	91.0
284	3	0.1	2515	91.1
285	2	0.1	2517	91.2
286	1	0.0	2518	91.2
287	2	0.1	2520	91.3
288	3	0.1	2523	91.4
289	3	0.1	2526	91.5
290	1	0.0	2527	91.6
291	1	0.0	2528	91.6

TABLE B.3
FREQUENCY ANALYSIS FOR TIME BETWEEN EVENTS

T	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
292	1	0.0	2529	91.6
293	1	0.0	2530	91.7
294	2	0.1	2532	91.7
295	2	0.1	2534	91.8
299	1	0.0	2535	91.8
301	1	0.0	2536	91.9
304	3	0.1	2539	92.0
305	5	0.2	2544	92.2
306	1	0.0	2545	92.2
308	3	0.1	2548	92.3
309	1	0.0	2549	92.4
310	3	0.1	2552	92.5
311	2	0.1	2554	92.5
312	3	0.1	2557	92.6
314	2	0.1	2559	92.7
315	2	0.1	2561	92.8
316	2	0.1	2563	92.9
317	2	0.1	2565	92.9
318	2	0.1	2567	93.0
320	1	0.0	2568	93.0
323	1	0.0	2569	93.1
324	2	0.1	2571	93.2
325	1	0.0	2572	93.2
326	1	0.0	2573	93.2
327	1	0.0	2574	93.3
328	1	0.0	2575	93.3
329	2	0.1	2577	93.4
330	2	0.1	2579	93.4
331	2	0.1	2581	93.5
332	1	0.0	2582	93.6
335	1	0.0	2583	93.6
336	1	0.0	2584	93.6
338	1	0.0	2585	93.7
339	1	0.0	2586	93.7
340	2	0.1	2588	93.8
341	2	0.1	2590	93.8
342	4	0.1	2594	94.0
345	1	0.0	2595	94.0
346	1	0.0	2596	94.1
348	1	0.0	2597	94.1
349	2	0.1	2599	94.2
350	2	0.1	2601	94.2
351	1	0.0	2602	94.3
352	2	0.1	2604	94.3
354	1	0.0	2605	94.4
356	3	0.1	2608	94.5

TABLE B.3
FREQUENCY ANALYSIS FOR TIME BETWEEN EVENTS

T	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
357	3	0.1	2611	94.6
359	1	0.0	2612	94.6
360	1	0.0	2613	94.7
364	2	0.1	2615	94.7
365	2	0.1	2617	94.8
367	1	0.0	2618	94.9
369	2	0.1	2620	94.9
370	1	0.0	2621	95.0
371	1	0.0	2622	95.0
372	2	0.1	2624	95.1
373	1	0.0	2625	95.1
374	1	0.0	2626	95.1
376	1	0.0	2627	95.2
377	3	0.1	2630	95.3
378	2	0.1	2632	95.4
379	2	0.1	2634	95.4
380	2	0.1	2636	95.5
381	1	0.0	2637	95.5
382	1	0.0	2638	95.6
383	2	0.1	2640	95.7
386	1	0.0	2641	95.7
387	4	0.1	2645	95.8
388	1	0.0	2646	95.9
389	1	0.0	2647	95.9
391	3	0.1	2650	96.0
393	1	0.0	2651	96.1
395	1	0.0	2652	96.1
396	2	0.1	2654	96.2
397	1	0.0	2655	96.2
398	1	0.0	2656	96.2
399	2	0.1	2658	96.3
402	1	0.0	2659	96.3
403	1	0.0	2660	96.4
406	1	0.0	2661	96.4
407	1	0.0	2662	96.4
408	2	0.1	2664	96.5
410	1	0.0	2665	96.6
414	1	0.0	2666	96.6
415	2	0.1	2668	96.7
416	1	0.0	2669	96.7
418	1	0.0	2670	96.7
419	2	0.1	2672	96.8
420	3	0.1	2675	96.9
423	1	0.0	2676	97.0
424	1	0.0	2677	97.0
425	1	0.0	2678	97.0

TABLE B.3
FREQUENCY ANALYSIS FOR TIME BETWEEN EVENTS

T	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
426	1	0.0	2679	97.1
431	1	0.0	2680	97.1
435	1	0.0	2681	97.1
441	1	0.0	2682	97.2
442	3	0.1	2685	97.3
443	1	0.0	2686	97.3
447	1	0.0	2687	97.4
448	2	0.1	2689	97.4
450	1	0.0	2690	97.5
452	1	0.0	2691	97.5
453	1	0.0	2692	97.5
454	1	0.0	2693	97.6
458	1	0.0	2694	97.6
461	1	0.0	2695	97.6
462	1	0.0	2696	97.7
464	2	0.1	2698	97.8
465	1	0.0	2699	97.8
468	1	0.0	2700	97.8
470	1	0.0	2701	97.9
473	1	0.0	2702	97.9
474	1	0.0	2703	97.9
479	1	0.0	2704	98.0
483	1	0.0	2705	98.0
488	1	0.0	2706	98.0
510	1	0.0	2707	98.1
511	1	0.0	2708	98.1
512	1	0.0	2709	98.2
514	1	0.0	2710	98.2
519	1	0.0	2711	98.2
522	1	0.0	2712	98.3
523	1	0.0	2713	98.3
525	1	0.0	2714	98.3
526	1	0.0	2715	98.4
528	1	0.0	2716	98.4
531	1	0.0	2717	98.4
536	1	0.0	2718	98.5
538	1	0.0	2719	98.5
544	1	0.0	2720	98.6
549	1	0.0	2721	98.6
580	1	0.0	2722	98.6
584	1	0.0	2723	98.7
592	1	0.0	2724	98.7
597	1	0.0	2725	98.7
605	1	0.0	2726	98.8
616	1	0.0	2727	98.8
617	1	0.0	2728	98.8

TABLE B.3
FREQUENCY ANALYSIS FOR TIME BETWEEN EVENTS

T	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
628	1	0.0	2729	98.9
631	2	0.1	2731	98.9
632	1	0.0	2732	99.0
641	1	0.0	2733	99.0
648	2	0.1	2735	99.1
655	1	0.0	2736	99.1
674	1	0.0	2737	99.2
678	1	0.0	2738	99.2
697	1	0.0	2739	99.2
723	1	0.0	2740	99.3
728	1	0.0	2741	99.3
729	1	0.0	2742	99.3
731	1	0.0	2743	99.4
742	1	0.0	2744	99.4
743	1	0.0	2745	99.5
749	1	0.0	2746	99.5
753	1	0.0	2747	99.5
794	1	0.0	2748	99.6
804	1	0.0	2749	99.6
866	1	0.0	2750	99.6
876	1	0.0	2751	99.7
927	1	0.0	2752	99.7
951	1	0.0	2753	99.7
974	1	0.0	2754	99.8
991	1	0.0	2755	99.8
1008	1	0.0	2756	99.9
1129	1	0.0	2757	99.9
1132	1	0.0	2758	99.9
1146	1	0.0	2759	100.0
1186	1	0.0	2760	100.0

TABLE B.4
FREQUENCY ANALYSIS FOR DURATION

D	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
1	855	31.0	855	31.0
2	466	16.9	1321	47.9
3	307	11.1	1628	59.0
4	223	8.1	1851	67.1
5	182	6.6	2033	73.7
6	139	5.0	2172	78.7
7	105	3.8	2277	82.5
8	91	3.3	2368	85.8
9	65	2.4	2433	88.2
10	59	2.1	2492	90.3
11	45	1.6	2537	91.9
12	47	1.7	2584	93.6
13	30	1.1	2614	94.7
14	26	0.9	2640	95.7
15	18	0.7	2658	96.3
16	17	0.6	2675	96.9
17	18	0.7	2693	97.6
18	9	0.3	2702	97.9
19	9	0.3	2711	98.2
20	12	0.4	2723	98.7
21	10	0.4	2733	99.0
22	2	0.1	2735	99.1
23	4	0.1	2739	99.2
24	3	0.1	2742	99.3
25	3	0.1	2745	99.5
26	3	0.1	2748	99.6
27	1	0.0	2749	99.6
28	2	0.1	2751	99.7
30	1	0.0	2752	99.7
31	1	0.0	2753	99.7
32	1	0.0	2754	99.8
33	1	0.0	2755	99.8
35	2	0.1	2757	99.9
42	1	0.0	2758	99.9
56	1	0.0	2759	100.0
82	1	0.0	2760	100.0

TABLE B.5
FREQUENCY ANALYSIS FOR MAXIMUM HOURLY INTENSITY

IMH	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
0.01	591	21.4	591	21.4
0.02	228	8.3	819	29.7
0.03	174	6.3	993	36.0
0.04	147	5.3	1140	41.3
0.05	124	4.5	1264	45.8
0.06	75	2.7	1339	48.5
0.07	70	2.5	1409	51.1
0.08	80	2.9	1489	53.9
0.09	57	2.1	1546	56.0
0.10	71	2.6	1617	58.6
0.11	43	1.6	1660	60.1
0.12	53	1.9	1713	62.1
0.13	53	1.9	1766	64.0
0.14	38	1.4	1804	65.4
0.15	41	1.5	1845	66.8
0.16	24	0.9	1869	67.7
0.17	34	1.2	1903	68.9
0.18	26	0.9	1929	69.9
0.19	34	1.2	1963	71.1
0.20	41	1.5	2004	72.6
0.21	34	1.2	2038	73.8
0.22	21	0.8	2059	74.6
0.23	26	0.9	2085	75.5
0.24	26	0.9	2111	76.5
0.25	24	0.9	2135	77.4
0.26	15	0.5	2150	77.9
0.27	26	0.9	2176	78.8
0.28	18	0.7	2194	79.5
0.29	16	0.6	2210	80.1
0.30	25	0.9	2235	81.0
0.31	16	0.6	2251	81.6
0.32	20	0.7	2271	82.3
0.33	17	0.6	2288	82.9
0.34	11	0.4	2299	83.3
0.35	15	0.5	2314	83.8
0.36	10	0.4	2324	84.2
0.37	9	0.3	2333	84.5
0.38	10	0.4	2343	84.9
0.39	14	0.5	2357	85.4
0.40	11	0.4	2368	85.8
0.41	12	0.4	2380	86.2
0.42	11	0.4	2391	86.6
0.43	12	0.4	2403	87.1
0.44	8	0.3	2411	87.4
0.45	10	0.4	2421	87.7
0.46	8	0.3	2429	88.0

TABLE B.5
FREQUENCY ANALYSIS FOR MAXIMUM HOURLY INTENSITY

IMH	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
0.47	10	0.4	2439	88.4
0.48	6	0.2	2445	88.6
0.49	8	0.3	2453	88.9
0.50	16	0.6	2469	89.5
0.51	6	0.2	2475	89.7
0.52	9	0.3	2484	90.0
0.53	7	0.3	2491	90.3
0.54	5	0.2	2496	90.4
0.55	14	0.5	2510	90.9
0.56	8	0.3	2518	91.2
0.57	6	0.2	2524	91.4
0.58	5	0.2	2529	91.6
0.59	12	0.4	2541	92.1
0.60	11	0.4	2552	92.5
0.61	1	0.0	2553	92.5
0.62	6	0.2	2559	92.7
0.63	2	0.1	2561	92.8
0.64	2	0.1	2563	92.9
0.65	5	0.2	2568	93.0
0.66	6	0.2	2574	93.3
0.67	2	0.1	2576	93.3
0.68	8	0.3	2584	93.6
0.69	2	0.1	2586	93.7
0.70	5	0.2	2591	93.9
0.71	3	0.1	2594	94.0
0.72	3	0.1	2597	94.1
0.73	3	0.1	2600	94.2
0.74	1	0.0	2601	94.2
0.75	6	0.2	2607	94.5
0.77	5	0.2	2612	94.6
0.78	3	0.1	2615	94.7
0.79	6	0.2	2621	95.0
0.80	3	0.1	2624	95.1
0.81	4	0.1	2628	95.2
0.82	1	0.0	2629	95.3
0.83	2	0.1	2631	95.3
0.84	4	0.1	2635	95.5
0.85	6	0.2	2641	95.7
0.86	2	0.1	2643	95.8
0.87	7	0.3	2650	96.0
0.88	3	0.1	2653	96.1
0.89	1	0.0	2654	96.2
0.90	7	0.3	2661	96.4
0.92	6	0.2	2667	96.6
0.93	1	0.0	2668	96.7
0.94	2	0.1	2670	96.7

TABLE B.5
FREQUENCY ANALYSIS FOR MAXIMUM HOURLY INTENSITY

IMH	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
0.95	4	0.1	2674	96.9
0.96	1	0.0	2675	96.9
0.97	4	0.1	2679	97.1
1.00	3	0.1	2682	97.2
1.01	2	0.1	2684	97.2
1.02	4	0.1	2688	97.4
1.03	4	0.1	2692	97.5
1.04	2	0.1	2694	97.6
1.05	1	0.0	2695	97.6
1.06	1	0.0	2696	97.7
1.08	1	0.0	2697	97.7
1.09	1	0.0	2698	97.8
1.10	1	0.0	2699	97.8
1.11	1	0.0	2700	97.8
1.14	3	0.1	2703	97.9
1.15	2	0.1	2705	98.0
1.16	3	0.1	2708	98.1
1.17	2	0.1	2710	98.2
1.18	1	0.0	2711	98.2
1.19	1	0.0	2712	98.3
1.20	2	0.1	2714	98.3
1.21	1	0.0	2715	98.4
1.23	1	0.0	2716	98.4
1.25	2	0.1	2718	98.5
1.28	1	0.0	2719	98.5
1.30	3	0.1	2722	98.6
1.31	1	0.0	2723	98.7
1.32	1	0.0	2724	98.7
1.33	1	0.0	2725	98.7
1.35	1	0.0	2726	98.8
1.37	2	0.1	2728	98.8
1.40	1	0.0	2729	98.9
1.41	1	0.0	2730	98.9
1.42	1	0.0	2731	98.9
1.43	2	0.1	2733	99.0
1.44	1	0.0	2734	99.1
1.45	1	0.0	2735	99.1
1.47	1	0.0	2736	99.1
1.49	1	0.0	2737	99.2
1.50	1	0.0	2738	99.2
1.55	1	0.0	2739	99.2
1.56	1	0.0	2740	99.3
1.57	1	0.0	2741	99.3
1.66	1	0.0	2742	99.3
1.74	1	0.0	2743	99.4
1.75	1	0.0	2744	99.4

TABLE B.5
FREQUENCY ANALYSIS FOR MAXIMUM HOURLY INTENSITY

IMH	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
1.76	2	0.1	2746	99.5
1.77	1	0.0	2747	99.5
1.81	1	0.0	2748	99.6
1.86	2	0.1	2750	99.6
1.95	1	0.0	2751	99.7
1.98	1	0.0	2752	99.7
2.13	1	0.0	2753	99.7
2.20	1	0.0	2754	99.8
2.22	1	0.0	2755	99.8
2.25	1	0.0	2756	99.9
2.42	1	0.0	2757	99.9
2.69	1	0.0	2758	99.9
2.83	1	0.0	2759	100.0
3.08	1	0.0	2760	100.0

TABLE B.6
FREQUENCY ANALYSIS FOR WEIGHTED INTENSITY

IW	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
0.01	622	22.5	622	22.5
0.02	288	10.4	910	33.0
0.03	206	7.5	1116	40.4
0.04	141	5.1	1257	45.5
0.05	127	4.6	1384	50.1
0.06	88	3.2	1472	53.3
0.07	90	3.3	1562	56.6
0.08	75	2.7	1637	59.3
0.09	59	2.1	1696	61.4
0.10	70	2.5	1766	64.0
0.11	55	2.0	1821	66.0
0.12	48	1.7	1869	67.7
0.13	60	2.2	1929	69.9
0.14	27	1.0	1956	70.9
0.15	33	1.2	1989	72.1
0.16	40	1.4	2029	73.5
0.17	32	1.2	2061	74.7
0.18	36	1.3	2097	76.0
0.19	37	1.3	2134	77.3
0.20	38	1.4	2172	78.7
0.21	21	0.8	2193	79.5
0.22	21	0.8	2214	80.2
0.23	21	0.8	2235	81.0
0.24	29	1.1	2264	82.0
0.25	27	1.0	2291	83.0
0.26	18	0.7	2309	83.7
0.27	20	0.7	2329	84.4
0.28	11	0.4	2340	84.8
0.29	10	0.4	2350	85.1
0.30	17	0.6	2367	85.8
0.31	12	0.4	2379	86.2
0.32	13	0.5	2392	86.7
0.33	12	0.4	2404	87.1
0.34	19	0.7	2423	87.8
0.35	11	0.4	2434	88.2
0.36	13	0.5	2447	88.7
0.37	12	0.4	2459	89.1
0.38	15	0.5	2474	89.6
0.39	4	0.1	2478	89.8
0.40	11	0.4	2489	90.2
0.41	8	0.3	2497	90.5
0.42	14	0.5	2511	91.0
0.43	8	0.3	2519	91.3
0.44	11	0.4	2530	91.7
0.45	9	0.3	2539	92.0
0.46	9	0.3	2548	92.3

TABLE B.6
FREQUENCY ANALYSIS FOR WEIGHTED INTENSITY

IW	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
0.47	5	0.2	2553	92.5
0.48	9	0.3	2562	92.8
0.49	5	0.2	2567	93.0
0.50	6	0.2	2573	93.2
0.51	3	0.1	2576	93.3
0.52	6	0.2	2582	93.6
0.53	10	0.4	2592	93.9
0.54	5	0.2	2597	94.1
0.55	10	0.4	2607	94.5
0.56	6	0.2	2613	94.7
0.57	2	0.1	2615	94.7
0.58	4	0.1	2619	94.9
0.59	3	0.1	2622	95.0
0.60	4	0.1	2626	95.1
0.61	2	0.1	2628	95.2
0.62	7	0.3	2635	95.5
0.63	5	0.2	2640	95.7
0.64	1	0.0	2641	95.7
0.65	2	0.1	2643	95.8
0.66	4	0.1	2647	95.9
0.67	4	0.1	2651	96.1
0.68	4	0.1	2655	96.2
0.69	3	0.1	2658	96.3
0.70	3	0.1	2661	96.4
0.71	6	0.2	2667	96.6
0.72	3	0.1	2670	96.7
0.73	4	0.1	2674	96.9
0.74	5	0.2	2679	97.1
0.75	1	0.0	2680	97.1
0.76	4	0.1	2684	97.2
0.77	4	0.1	2688	97.4
0.78	2	0.1	2690	97.5
0.79	2	0.1	2692	97.5
0.80	4	0.1	2696	97.7
0.82	1	0.0	2697	97.7
0.83	2	0.1	2699	97.8
0.85	1	0.0	2700	97.8
0.86	1	0.0	2701	97.9
0.87	1	0.0	2702	97.9
0.88	3	0.1	2705	98.0
0.89	2	0.1	2707	98.1
0.90	4	0.1	2711	98.2
0.91	1	0.0	2712	98.3
0.92	1	0.0	2713	98.3
0.93	2	0.1	2715	98.4
0.94	3	0.1	2718	98.5

TABLE B.6
FREQUENCY ANALYSIS FOR WEIGHTED INTENSITY

IW	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
0.95	1	0.0	2719	98.5
0.96	2	0.1	2721	98.6
0.97	1	0.0	2722	98.6
0.98	1	0.0	2723	98.7
1.00	2	0.1	2725	98.7
1.01	1	0.0	2726	98.8
1.03	1	0.0	2727	98.8
1.06	3	0.1	2730	98.9
1.07	1	0.0	2731	98.9
1.08	1	0.0	2732	99.0
1.10	1	0.0	2733	99.0
1.13	2	0.1	2735	99.1
1.14	2	0.1	2737	99.2
1.15	2	0.1	2739	99.2
1.16	1	0.0	2740	99.3
1.20	2	0.1	2742	99.3
1.23	1	0.0	2743	99.4
1.25	1	0.0	2744	99.4
1.30	1	0.0	2745	99.5
1.32	1	0.0	2746	99.5
1.37	1	0.0	2747	99.5
1.38	1	0.0	2748	99.6
1.40	1	0.0	2749	99.6
1.43	2	0.1	2751	99.7
1.48	1	0.0	2752	99.7
1.56	1	0.0	2753	99.7
1.57	1	0.0	2754	99.8
1.59	1	0.0	2755	99.8
1.85	1	0.0	2756	99.9
1.96	1	0.0	2757	99.9
1.98	1	0.0	2758	99.9
2.18	1	0.0	2759	100.0
2.27	1	0.0	2760	100.0

TABLE B.7
SAS PROGRAM LISTING FOR COMPUTING THE PARAMETERS OF THE WEIBULL PDF

```
DATA WEIBPARA;
INPUT ALPHA;
MU=92.8141;
SIGMA=136.6676;
LAMBDA=0.0;
SKEWNESS=2.8574;
G1=GAMMA(1+1/ALPHA);
G2=GAMMA(1+2/ALPHA);
G3=GAMMA(1+3/ALPHA);
BALPHA=1/SQRT(G2-G1**2);
AALPHA=(1-G1)*BALPHA;
BETA1=MU+SIGMA*AALPHA;
BETA2=LAMBDA+SIGMA*BALPHA;
SKEWCHEK=(G3-3*G2*G1+2*G1**3)*BALPHA**3;
CARDS;
0.6940
0.6941
0.6942
0.6943
0.6944
0.6945
0.6946
0.6947
0.6948
0.6949
0.6950
0.6951
0.6952
0.6953
0.6954
0.6955
0.6956
0.6957
0.6958
0.6959
;
PROC PRINT;
VAR ALPHA BETA1 BETA2 SKEWNESS SKEWCHEK MU SIGMA LAMBDA;
```

TABLE B.8
SAS PROGRAM LISTING FOR MONTE CARLO SIMULATION OF RAINFALL EVENTS

```
CMS FI WORK DISK TESTSIM WORK B;
DATA WORK.TESTSIM;
RETAIN ITS 0;
DO UNTIL (ITS=10) /* (ITS=2760) */ ;
*** SAMPLE R PDF. ***;
R= .005+1.1046*RANGAM(0,.3515);
*** SET R TO THE SMALLEST MULTIPLE OF .01 INCH >= R. ***;
R=CEIL(R*100)/100;
*** SAMPLE T PDF. ***;
TUNIFORM=RANUNI(0);
T=3;
IF TUNIFORM <= .0837 THEN GOTO TSET;
T=4;
IF TUNIFORM <= .1402 THEN GOTO TSET;
T=5;
IF TUNIFORM <= .1830 THEN GOTO TSET;
T=6;
IF TUNIFORM <= .2145 THEN GOTO TSET;
T=7;
IF TUNIFORM <= .2409 THEN GOTO TSET;
T=8;
IF TUNIFORM <= .2641 THEN GOTO TSET;
T=9;
IF TUNIFORM <= .2819 THEN GOTO TSET;
T=10;
IF TUNIFORM <= .2924 THEN GOTO TSET;
T=11;
IF TUNIFORM <= .3102 THEN GOTO TSET;
T=11.5+183.3745*RANGAM(0,.6575);
*** SET T TO THE SMALLEST INTEGER >= T. ***;
T=CEIL(T);
*** THE VALUE OF T IS NOW SET. ***;
*** SAMPLE D PDF. ***;
TSET: DUNIFORM=RANUNI(0);
D=1;
IF DUNIFORM <= .3098 THEN GOTO DSET;
D=2;
IF DUNIFORM <= .4786 THEN GOTO DSET;
D=2.0+4.8121*RANEXP(0);
*** SET D TO THE SMALLEST INTEGER >= D. ***;
D=CEIL(D);
*** THE VALUE OF D IS NOW SET. ***;
DSET: ITS= ITS+1;
OUTPUT;
END;
 /*
TITLE1 'TEST SIMULATION FOR R, T, AND D';
TITLE2 'USING DERIVED PROBABILITY DENSITY FUNCTIONS';
TITLE3 'WITH 2760 TRIALS';
PROC FREQ;
TABLES R;
 */
```

TABLE B.9
CONTINGENCY TABLE OF R BY T

TABLE OF R BY T

R T

FREQUENCY / EXPECTED / CELL CHI2 / PERCENT / ROW PCT / COL PCT	5	10	15	25	50	100	200	1186	TOTAL
0.10	267 239.0 3.29023 9.67 20.44 52.87	145 142.9 .030775 5.25 11.10 48.01	87 90.4 0.12633 3.15 6.66 45.55	109 110.7 .026908 3.95 8.35 46.58	135 151.9 1.87888 4.89 10.34 42.06	197 194.0 .046166 7.14 15.08 48.05	181 184.1 .051212 6.56 13.86 46.53	185 193.1 .336565 6.70 14.17 45.34	1306 47.32
0.15	35 33.8 .039094 1.27 18.92 6.93	21 20.2 .028327 0.76 11.35 6.95	14 12.8 .112003 0.51 7.57 7.33	18 15.7 .341747 0.65 9.73 7.69	21 21.5 .012389 0.76 11.35 6.54	29 27.5 .083862 1.05 15.68 7.07	22 26.1 .636632 0.80 11.89 5.66	25 27.3 .201562 0.91 13.51 6.13	185 6.70
0.20	15 24.7 3.81 0.54 11.11 2.97	11 14.8 .963056 0.40 8.15 3.64	9 9.3 .012548 0.33 6.67 4.71	10 11.4 .182594 0.36 7.41 4.27	16 15.7 .005691 0.58 11.85 4.98	29 20.1 3.99039 1.05 21.48 7.07	17 19.0 .215977 0.62 12.59 4.37	28 20.0 3.24192 1.01 20.74 6.86	135 4.89
0.30	33 38.4 .765639 1.20 15.71 6.53	22 23.0 .041648 0.80 10.48 7.28	11 14.5 .858712 0.40 5.24 5.76	9 17.8 4.3538 0.33 4.29 3.85	31 24.4 1.7706 1.12 14.76 9.66	30 31.2 .045826 1.09 14.29 7.32	38 29.6 2.38519 1.38 18.10 9.77	36 31.0 .791377 1.30 17.14 8.82	210 7.61
TOTAL	505 18.30	302 10.94	191 6.92	234 8.48	321 11.63	410 14.86	389 14.09	408 14.78	2760 100.00

(CONTINUED)

TABLE B.9
CONTINGENCY TABLE OF R BY T

TABLE OF R BY T

R T

FREQUENCY / EXPECTED / CELL CHI2 / PERCENT / ROW PCT / COL PCT	5	10	15	25	50	100	200	1186	TOTAL
0.40	17	16	10	15	23	28	18	27	154
	28.2	16.9	10.7	13.1	17.9	22.9	21.7	22.8	
	4.43393	0.04295	.040533	.289289	1.44601	1.14732	.632459	.787754	
	0.62	0.58	0.36	0.54	0.83	1.01	0.65	0.98	5.58
	11.04	10.39	6.49	9.74	14.94	18.18	11.69	17.53	
	3.37	5.30	5.24	6.41	7.17	6.83	4.63	6.62	
0.50	16	8	8	9	16	15	18	16	106
	19.4	11.6	7.3	9.0	12.3	15.7	14.9	15.7	
	.594255	1.11648	.060194	2E-05	1.09356	.035378	.626812	.006968	
	0.58	0.29	0.29	0.33	0.58	0.54	0.65	0.58	3.84
	15.09	7.55	7.55	8.49	15.09	14.15	16.98	15.09	
	3.17	2.65	4.19	3.85	4.98	3.66	4.63	3.92	
0.60	19	11	12	18	11	16	12	16	115
	21.0	12.6	8.0	9.8	13.4	17.1	16.2	17.0	
	.198102	.199227	2.05257	6.98077	.421729	.068699	1.09265	.058824	
	0.69	0.40	0.43	0.65	0.40	0.58	0.43	0.58	4.17
	16.52	9.57	10.43	15.65	9.57	13.91	10.43	13.91	
	3.76	3.64	6.28	7.69	3.43	3.90	3.08	3.92	
0.80	21	21	9	7	20	16	26	18	138
	25.3	15.1	9.5	11.7	16.0	20.5	19.4	20.4	
	.715347	2.3053	.031675	1.88803	.972118	.987805	2.20578	.282353	
	0.76	0.76	0.33	0.25	0.72	0.58	0.94	0.65	5.00
	15.22	15.22	6.52	5.07	14.49	11.59	18.84	13.04	
	4.16	6.95	4.71	2.99	6.23	3.90	6.68	4.41	
TOTAL	505	302	191	234	321	410	389	408	2760
	18.30	10.94	6.92	8.48	11.63	14.86	14.09	14.78	100.00

(CONTINUED)

TABLE B.9
CONTINGENCY TABLE OF R BY T

TABLE OF R BY T

R T

FREQUENCY / EXPECTED / CELL CHI2 / PERCENT / ROW PCT / COL PCT	5	10	15	25	50	100	200	1186	TOTAL
1.00	27	12	8	7	10	13	14	12	103
	18.8	11.3	7.1	8.7	12.0	15.3	14.5	15.2	
	3.52793	.047246	.106702	.343761	.327048	.345953	.018414	0.68354	
	0.98	0.43	0.29	0.25	0.36	0.47	0.51	0.43	3.73
	26.21	11.65	7.77	6.80	9.71	12.62	13.59	11.65	
	5.35	3.97	4.19	2.99	3.12	3.17	3.60	2.94	
1.50	21	13	12	17	20	20	22	20	145
	26.5	15.9	10.0	12.3	16.9	21.5	20.4	21.4	
	1.15299	.517689	.385025	1.80188	.583112	.110082	.119601	0.09604	
	0.76	0.47	0.43	0.62	0.72	0.72	0.80	0.72	5.25
	14.48	8.97	8.28	11.72	13.79	13.79	15.17	13.79	
	4.16	4.30	6.28	7.26	6.23	4.88	5.66	4.90	
6.17	34	22	11	15	18	17	21	25	163
	29.8	17.8	11.3	13.8	19.0	24.2	23.0	24.1	
	.584647	.972386	.006954	0.10083	.048372	2.14913	.169539	.033942	
	1.23	0.80	0.40	0.54	0.65	0.62	0.76	0.91	5.91
	20.86	13.50	6.75	9.20	11.04	10.43	12.88	15.34	
	6.73	7.28	5.76	6.41	5.61	4.15	5.40	6.13	
TOTAL	505	302	191	234	321	410	389	408	2760
	18.30	10.94	6.92	8.48	11.63	14.86	14.09	14.78	100.00

TABLE B.9
CONTINGENCY TABLE OF R BY T

STATISTICS FOR TABLE OF R BY T

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	70	77.725	0.246
LIKELIHOOD RATIO CHI-SQUARE	70	76.992	0.265
MANTEL-HAENSZEL CHI-SQUARE	1	0.015	0.902
PHI		0.168	
CONTINGENCY COEFFICIENT		0.165	
CRAMER'S V		0.063	

SAMPLE SIZE = 2760

TABLE B.10
CONTINGENCY TABLE OF T BY D

TABLE OF T BY D

T D

FREQUENCY / EXPECTED / CELL CHI2 / PERCENT / ROW PCT / COL PCT	1	2	3	4	5	6	8	10	13	82	TOTAL
5	163 156.4 .275062 5.91 32.28 19.06	85 85.3 8E-04 3.08 16.83 18.24	50 56.2 .678181 1.81 9.90 16.29	35 40.8 0.82518 1.27 6.93 15.70	25 33.3 2.06908 0.91 4.95 13.74	24 25.4 .080738 0.87 4.75 17.27	35 35.9 .020735 1.27 6.93 17.86	24 22.7 .075822 0.87 4.75 19.35	34 22.3 6.10886 1.23 6.73 27.87	30 26.7 0.40426 1.09 5.94 20.55	505 18.30
10	92 93.6 .025825 3.33 30.46 10.76	55 51.0 .315382 1.99 18.21 11.80	23 33.6 3.33981 0.83 7.62 7.49	18 24.4 1.67902 0.65 5.96 8.07	24 19.9 .838152 0.87 7.95 13.19	14 15.2 0.09617 0.51 4.64 10.07	28 21.4 2.00267 1.01 9.27 14.29	13 13.6 .023788 0.47 4.30 10.48	15 13.3 .204123 0.54 4.97 12.30	20 16.0 1.01392 0.72 6.62 13.70	302 10.94
15	61 59.2 .056694 2.21 31.94 7.13	31 32.2 0.04834 1.12 16.23 6.65	21 21.2 .002832 0.76 10.99 6.84	11 15.4 1.27297 0.40 5.76 4.93	14 12.6 .156748 0.51 7.33 7.69	14 9.6 1.99511 0.51 7.33 10.07	9 13.6 1.53556 0.33 4.71 4.59	10 8.6 .234596 0.36 5.24 8.06	9 8.4 0.03678 0.33 4.71 7.38	11 10.1 .079525 0.40 5.76 7.53	191 6.92
25	76 72.5 .170042 2.75 32.48 8.89	46 39.5 1.06653 1.67 19.66 9.87	30 26.0 .606061 1.09 12.82 9.77	12 18.9 2.52294 0.43 5.13 5.38	16 15.4 .021024 0.58 6.84 8.79	12 11.8 0.00393 0.43 5.13 8.63	18 16.6 .115037 0.65 7.69 9.18	8 10.5 .600719 0.29 3.42 6.45	7 10.3 1.08076 0.25 2.99 5.74	9 12.4 .921991 0.33 3.85 6.16	234 8.48
TOTAL	855 30.98	466 16.88	307 11.12	223 8.08	182 6.59	139 5.04	196 7.10	124 4.49	122 4.42	146 5.29	2760 100.00

(CONTINUED)

TABLE B.10
CONTINGENCY TABLE OF T BY D

TABLE OF T BY D

T D

FREQUENCY / EXPECTED / CELL CHI2 / PERCENT / ROW PCT / COL PCT	1	2	3	4	5	6	8	10	13	82	TOTAL
50	87	52	43	37	18	12	19	22	15	16	321
	99.4	54.2	35.7	25.9	21.2	16.2	22.8	14.4	14.2	17.0	
	1.5563	.089126	1.49027	4.71991	.473954	1.07372	.632005	3.98219	.046339	.056609	
	3.15	1.88	1.56	1.34	0.65	0.43	0.69	0.80	0.54	0.58	11.63
	27.10	16.20	13.40	11.53	5.61	3.74	5.92	6.85	4.67	4.98	
	10.18	11.16	14.01	16.59	9.89	8.63	9.69	17.74	12.30	10.96	
100	126	72	41	43	24	19	32	21	19	13	410
	127.0	69.2	45.6	33.1	27.0	20.6	29.1	18.4	18.1	21.7	
	.008045	0.11127	.465007	2.94263	.340976	.131618	.285678	.361281	.042421	3.48059	
	4.57	2.61	1.49	1.56	0.87	0.69	1.16	0.76	0.69	0.47	14.86
	30.73	17.56	10.00	10.49	5.85	4.63	7.80	5.12	4.63	3.17	
	14.74	15.45	13.36	19.28	13.19	13.67	16.33	16.94	15.57	8.90	
200	113	67	54	30	29	21	30	14	14	17	389
	120.5	65.7	43.3	31.4	25.7	19.6	27.6	17.5	17.2	20.6	
	.467461	0.02657	2.66125	.065068	.437121	.101345	.204251	.691672	.593638	.621978	
	4.09	2.43	1.96	1.09	1.05	0.76	1.09	0.51	0.51	0.62	14.09
	29.05	17.22	13.88	7.71	7.46	5.40	7.71	3.60	3.60	4.37	
	13.22	14.38	17.59	13.45	15.93	15.11	15.31	11.29	11.48	11.64	
1186	137	58	45	37	32	23	25	12	9	30	408
	126.4	68.9	45.4	33.0	26.9	20.5	29.0	18.3	18.0	21.6	
	.890444	1.72058	.003226	.493838	.965111	.292642	.545041	2.18622	4.5261	3.28285	
	4.96	2.10	1.63	1.34	1.16	0.83	0.91	0.43	0.33	1.09	14.78
	33.58	14.22	11.03	9.07	7.84	5.64	6.13	2.94	2.21	7.35	
	16.02	12.45	14.66	16.59	17.58	16.55	12.76	9.68	7.38	20.55	
TOTAL	855	466	307	223	182	139	196	124	122	146	2760
	30.98	16.88	11.12	8.08	6.59	5.04	7.10	4.49	4.42	5.29	100.00

TABLE B.10
CONTINGENCY TABLE OF T BY D

STATISTICS FOR TABLE OF T BY D

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	63	75.672	0.131
LIKELIHOOD RATIO CHI-SQUARE	63	76.295	0.121
MANTEL-HAENSZEL CHI-SQUARE	1	1.689	0.194
PHI		0.166	
CONTINGENCY COEFFICIENT		0.163	
CRAMER'S V		0.063	

SAMPLE SIZE = 2760

TABLE B.11
CONTINGENCY TABLE OF R BY D

TABLE OF R BY D

R D

FREQUENCY / EXPECTED / CELL CHI2 / PERCENT / ROW PCT / COL PCT	1	2	3	4	5	6	8	10	13	82	TOTAL
0.10	751	283	124	74	33	22	14	4	1	0	1306
	404.6	220.5	145.3	105.5	86.1	65.8	92.7	58.7	57.7	69.1	
	296.63	17.7117	3.11398	9.41589	32.7654	29.1318	66.8583	50.948	55.7463	69.0855	
	27.21	10.25	4.49	2.68	1.20	0.80	0.51	0.14	0.04	0.00	47.32
	57.50	21.67	9.49	5.67	2.53	1.68	1.07	0.31	0.08	0.00	
	87.84	60.73	40.39	33.18	18.13	15.83	7.14	3.23	0.82	0.00	
0.15	30	39	24	23	24	13	20	4	7	1	185
	57.3	31.2	20.6	14.9	12.2	9.3	13.1	8.3	8.2	9.8	
	13.0139	1.93009	.569095	4.33808	11.4152	1.45586	3.58445	2.23662	.169561	7.88842	
	1.09	1.41	0.87	0.83	0.87	0.47	0.72	0.14	0.25	0.04	6.70
	16.22	21.08	12.97	12.43	12.97	7.03	10.81	2.16	3.78	0.54	
	3.51	8.37	7.82	10.31	13.19	9.35	10.20	3.23	5.74	0.68	
0.20	15	29	26	16	8	8	13	10	5	5	135
	41.8	22.8	15.0	10.9	8.9	6.8	9.6	6.1	6.0	7.1	
	17.2008	1.69	8.03404	2.37746	.091429	.212182	1.21507	2.55267	.156827	.642065	
	0.54	1.05	0.94	0.58	0.29	0.29	0.47	0.36	0.18	0.18	4.89
	11.11	21.48	19.26	11.85	5.93	5.93	9.63	7.41	3.70	3.70	
	1.75	6.22	8.47	7.17	4.40	5.76	6.63	8.06	4.10	3.42	
0.30	28	38	28	28	14	16	29	12	8	9	210
	65.1	35.5	23.4	17.0	13.8	10.6	14.9	9.4	9.3	11.1	
	21.1058	.182457	.922214	7.17367	.001672	2.78164	13.3066	.697455	.177222	.400281	
	1.01	1.38	1.01	1.01	0.51	0.58	1.05	0.43	0.29	0.33	7.61
	13.33	18.10	13.33	13.33	6.67	7.62	13.81	5.71	3.81	4.29	
	3.27	8.15	9.12	12.56	7.69	11.51	14.80	9.68	6.56	6.16	
TOTAL	855	466	307	223	182	139	196	124	122	146	2760
	30.98	16.88	11.12	8.08	6.59	5.04	7.10	4.49	4.42	5.29	100.00

(CONTINUED)

TABLE B.11
CONTINGENCY TABLE OF R BY D

TABLE OF R BY D

R D

FREQUENCY / EXPECTED / CELL CHI2 / PERCENT / ROW PCT / COL PCT	1	2	3	4	5	6	8	10	13	82	TOTAL
0.40	12	25	25	21	15	9	19	9	12	7	154
	47.7	26.0	17.1	12.4	10.2	7.8	10.9	6.9	6.8	8.1	
	26.725	.038571	3.61603	5.88507	2.31149	.199598	5.94577	.626004	3.96117	.161321	
	0.43	0.91	0.91	0.76	0.54	0.33	0.69	0.33	0.43	0.25	5.58
	7.79	16.23	16.23	13.64	9.74	5.84	12.34	5.84	7.79	4.55	
	1.40	5.36	8.14	9.42	8.24	6.47	9.69	7.26	9.84	4.79	
0.50	4	13	15	8	19	12	9	9	7	10	106
	32.8	17.9	11.8	8.6	7.0	5.3	7.5	4.8	4.7	5.6	
	25.3242	1.33997	.873611	.037206	20.6361	8.31275	.288029	3.77084	1.14329	3.44131	
	0.14	0.47	0.54	0.29	0.69	0.43	0.33	0.33	0.25	0.36	3.84
	3.77	12.26	14.15	7.55	17.92	11.32	8.49	8.49	6.60	9.43	
	0.47	2.79	4.89	3.59	10.44	8.63	4.59	7.26	5.74	6.85	
0.60	9	8	16	8	15	9	23	11	13	3	115
	35.6	19.4	12.8	9.3	7.6	5.8	8.2	5.2	5.1	6.1	
	19.8987	6.7128	.804696	.179559	7.25366	1.77728	26.9422	6.58602	12.3292	1.56279	
	0.33	0.29	0.58	0.29	0.54	0.33	0.83	0.40	0.47	0.11	4.17
	7.83	6.96	13.91	6.96	13.04	7.83	20.00	9.57	11.30	2.61	
	1.05	1.72	5.21	3.59	8.24	6.47	11.73	8.87	10.66	2.05	
0.80	3	22	18	13	16	13	19	10	9	15	138
	42.8	23.3	15.3	11.1	9.1	6.9	9.8	6.2	6.1	7.3	
	36.9605	.072532	.457492	.306951	5.23187	5.26655	8.63673	2.32903	1.37869	8.12192	
	0.11	0.80	0.65	0.47	0.58	0.47	0.69	0.36	0.33	0.54	5.00
	2.17	15.94	13.04	9.42	11.59	9.42	13.77	7.25	6.52	10.87	
	0.35	4.72	5.86	5.83	8.79	9.35	9.69	8.06	7.38	10.27	
TOTAL	855	466	307	223	182	139	196	124	122	146	2760
	30.98	16.88	11.12	8.08	6.59	5.04	7.10	4.49	4.42	5.29	100.00

(CONTINUED)

TABLE B.11
CONTINGENCY TABLE OF R BY D

TABLE OF R BY D

R D

FREQUENCY EXPECTED CELL CHI2 PERCENT ROW PCT COL PCT	1	2	3	4	5	6	8	10	13	82	TOTAL
1.00	0	4	7	11	11	8	12	16	17	17	103
	31.9	17.4	11.5	8.3	6.8	5.2	7.3	4.6	4.6	5.4	
	31.9076	10.3106	1.73379	.861698	2.60703	1.5251	3.00144	27.9486	34.0289	24.4902	
	0.00	0.14	0.25	0.40	0.40	0.29	0.43	0.58	0.62	0.62	3.73
	0.00	3.88	6.80	10.68	10.68	7.77	11.65	15.53	16.50	16.50	
	0.00	0.86	2.28	4.93	6.04	5.76	6.12	12.90	13.93	11.64	
1.50	3	5	19	14	10	13	17	24	13	27	145
	44.9	24.5	16.1	11.7	9.6	7.3	10.3	6.5	6.4	7.7	
	39.1188	15.503	.511191	.445439	.020101	4.44518	4.36325	46.9327	6.77686	48.7123	
	0.11	0.18	0.69	0.51	0.36	0.47	0.62	0.87	0.47	0.98	5.25
	2.07	3.45	13.10	9.66	6.90	8.97	11.72	16.55	8.97	18.62	
	0.35	1.07	6.19	6.28	5.49	9.35	8.67	19.35	10.66	18.49	
6.17	0	0	5	7	17	16	21	15	30	52	163
	50.5	27.5	18.1	13.2	10.7	8.2	11.6	7.3	7.2	8.6	
	50.4946	27.521	9.50967	2.89053	3.6359	7.39412	7.67352	8.04751	72.1171	218.222	
	0.00	0.00	0.18	0.25	0.62	0.58	0.76	0.54	1.09	1.88	5.91
	0.00	0.00	3.07	4.29	10.43	9.82	12.88	9.20	18.40	31.90	
	0.00	0.00	1.63	3.14	9.34	11.51	10.71	12.10	24.59	35.62	
TOTAL	855	466	307	223	182	139	196	124	122	146	2760
	30.98	16.88	11.12	8.08	6.59	5.04	7.10	4.49	4.42	5.29	100.00

TABLE B.11
CONTINGENCY TABLE OF R BY D

STATISTICS FOR TABLE OF R BY D

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	90	1739.126	0.000
LIKELIHOOD RATIO CHI-SQUARE	90	1851.503	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	451.222	0.000
PHI		0.794	
CONTINGENCY COEFFICIENT		0.622	
CRAMER'S V		0.265	

SAMPLE SIZE = 2760

APPENDIX C

SIMULATION RESULTS

TABLE C.1
SAS PROGRAM LISTING FOR RAINFALL-RUNOFF STOCHASTIC SIMULATION MODEL

```
CMS FI DAT DISK EVENTSIM DAT A;
DATA DAT.EVENTSIM;
KEEP NRAINS NRUNOFFS CUMHOURS YEAR Q QP R T D TL TR R0T6;
ARRAY RLAST(36);
ARRAY TLAST(36);
ARRAY DLAST(36);
*** SET NUMBER OF YEARS FOR THE INITIALIZATION PERIOD (STARTYRS).  *** ;
  * ALL EVENTS IN THE INITIALIZATION PERIOD ARE OMITTED IN ORDER TO
    ELIMINATE ANY BIAS INTRODUCED BY THE INITIAL CONDITIONS.
    A MINIMUM OF STARTYRS=5 IS RECOMMENDED.  * ;
STARTYRS=5;
*** SET NUMBER OF YEARS TO BE SIMULATED BEYOND THE
    INITIALIZATION PERIOD (NYEARS).  *** ;
NYEARS=100;
*** INITIALIZE TIME BETWEEN EVENT VARIABLES.  ASSUME THAT THE LAST EVENT
    PRIOR TO THE INITIALIZATION PERIOD HAD Q>=0.01 INCH.  THEN,
    SET TL=0 AS A FLAG THAT THE LAST EVENT HAD Q>0 AND
    SET TR=0 AS A FLAG THAT THE LAST EVENT HAD Q>=0.01. *** ;
TL=0;
TR=0;
*** INITIALIZE THE CUMULATIVE TIME SINCE THE BEGINNING OF THE SIMULATION
    (HOURS), THE CURRENT CALENDAR YEAR SINCE THE BEGINNING OF THE
    SIMULATION, THE COUNTER FOR NUMBER OF RAIN EVENTS SIMULATED,
    AND THE COUNTER FOR NUMBER OF RUNOFF EVENTS (I.E., Q>0)
    SIMULATED.   LEAP YEARS ARE NOT ACCOUNTED FOR IN ASSIGNING THE
    NUMBER OF CALENDAR YEARS.  *** ;
CUMHOURS=0;
YEAR=0;
NRAINS=0;
NRUNOFFS=0;
*** SIMULATE RUNOFF EVENTS UNTIL THE FIRST EVENT FOR
    WHICH YEAR>NYEARS+STARTYRS.  *** ;
DO UNTIL (YEAR>NYEARS+STARTYRS);
*** INCLUDE ONLY RUNOFF EVENTS FOR WHICH Q>0 AND QP>0.  EVENTS WITH
    Q<=0 AND/OR QP<=0 ARE TREATED AS NON-RUNOFF PRODUCING
    RAIN EVENTS.  *** ;
DO UNTIL (Q>0);
DO UNTIL (QP>0);
*** SAMPLE R PDF. ***;
RESTART:  R= .005+1.1046*RANGAM(0,.3515);
  * SET R TO THE SMALLEST MULTIPLE OF .01 INCH >= R. * ;
R=CEIL(R*100)/100;
*** SAMPLE T PDF. ***;
TRESET:  TUNIFORM=RANUNI(0);
T=3;
IF TUNIFORM <= .0837 THEN GOTO TSET;
T=4;
IF TUNIFORM <= .1402 THEN GOTO TSET;
T=5;
IF TUNIFORM <= .1830 THEN GOTO TSET;
T=6;
IF TUNIFORM <= .2145 THEN GOTO TSET;
T=7;
IF TUNIFORM <= .2409 THEN GOTO TSET;
```

TABLE C.1
SAS PROGRAM LISTING FOR RAINFALL-RUNOFF STOCHASTIC SIMULATION MODEL

```
T=8;
IF TUNIFORM <= .2641 THEN GOTO TSET;
T=9;
IF TUNIFORM <= .2819 THEN GOTO TSET;
T=10;
IF TUNIFORM <= .2924 THEN GOTO TSET;
T=11;
IF TUNIFORM <= .3102 THEN GOTO TSET;
T=11.5+183.3745*RANGAM(0,.6575);
  * SET T TO THE SMALLEST INTEGER >= T. *;
T=CEIL(T);
  * INSURE THAT THE INITIAL (I.E., NRAINS=0) VALUE OF T>=144 HOURS
    (I.E., 6 DAYS) SO THAT THE INITIAL VALUE OF ROT6=0.  * ;
TSET:  IF NRAINS=0 AND T<144 THEN GOTO TRESET;
  * THE VALUE OF T IS NOW SET. *;
*** SAMPLE D PDF. ***;
DUNIFORM=RANUNI(0);
D=1;
IF DUNIFORM <= .3098 THEN GOTO DSET;
D=2;
IF DUNIFORM <= .4786 THEN GOTO DSET;
D=2.0+4.8121*RANEXP(0);
  * SET D TO THE SMALLEST INTEGER >= D. *;
D=CEIL(D);
  * THE VALUE OF D IS NOW SET. *;
*** COMPUTE TL.  *** ;
  * IF THE CURRENT VALUE OF TL=0, THEN THE LAST STORM PRODUCED RUNOFF,
    SET TL=T.  IF TL>0, THEN THE LAST STORM DID NOT PRODUCE RUNOFF, SO
    INCREASE TL BY D FOR THE LAST EVENT PLUS T FOR THE CURRENT
    EVENT.  * ;
DSET:  I=1;
TLLAST=TL;
IF TL=0 THEN TL=T;
ELSE TL=TL+DLAST(I)+T;
IF TL>1513 THEN TL=1513;
*** COMPUTE TR.  *** ;
  * IF THE CURRENT VALUE OF TR=0, THEN THE LAST STORM HAD Q>=0.01 INCH,
    SET TR=T.  IF TR>0, THEN THE LAST EVENT HAD Q<0.01 INCH, SO
    INCREASE TR BY D FOR THE LAST EVENT PLUS T FOR THE CURRENT
    EVENT.  * ;
TRLAST=TR;
IF TR=0 THEN TR=T;
ELSE TR=TR+DLAST(I)+T;
IF TR>1513 THEN TR=1513;
*** COMPUTE ROT6.  *** ;
  * THE MAXIMUM NUMBER OF STORMS THAT CAN OCCUR IN 6 DAYS IS 36,
    I.E., 36*(1-HOUR STORMS + 3-HOUR DRY PERIODS)=144 HOURS=6 DAYS.  * ;
  * INITIALIZE ROT6 AND CUMULATIVE TIME BETWEEN RAIN EVENTS (CUMT).  * ;
ROT6=0;
CUMT=T;
  * IF T>144 THERE ARE NO RAINS IN THE LAST 144 HOURS, ELSE THERE ARE
    BETWEEN 1 AND 36 RAINS.  * ;
IF T>144 THEN GOTO COMPUTEQ;
DO I=1 TO 36;
ROT6=ROT6+RLAST(I);
CUMT=CUMT+DLAST(I)+TLAST(I);
  * IF CUMT>144 THEN THE CURRENT RAIN, I, IS THE LAST RAIN TO END IN THE
    LAST 144 HOURS, ELSE THERE ARE BETWEEN 1 AND 35 MORE.  * ;
IF CUMT>144 THEN GOTO COMPUTEQ;
END;
*** COMPUTE Q AND QP.  *** ;
COMPUTEQ:  RW=R;
IA=RW/D;
QEPS=.03148*RANNOR(0);
QPEPS=4.5806*RANNOR(0);
Q=-.05017+.1671*RW-.0001263*TR+.02976*ROT6+QEPS;
```

TABLE C.1
SAS PROGRAM LISTING FOR RAINFALL-RUNOFF STOCHASTIC SIMULATION MODEL

```
QP=-4.6053+12.3698*RW-.02144*TL+32.4367*IA+QPEPS;
*** CHECK SIMULATION RESULTS FOR CONSISTENCY.  *** ;
  * IF Q>RW, THEN IGNORE THE EVENT AND RESTART THE SIMULATION FOR THE
    CURRENT EVENT.  EVENTS WITH Q<0 AND QP>0, OR Q>0 AND QP<0 ARE
    TREATED AS NON-RUNOFF PRODUCING EVENTS.  * ;
IF Q<RW THEN GOTO CONSIST;
TL=TLLAST;
TR=TRLAST;
GOTO RESTART;
*** INCREMENT RAIN EVENT COUNTER.  *** ;
CONSIST:  NRAINS=NRAINS+1;
*** COMPUTE CUMULATIVE HOURS FROM THE BEGINNING OF THE SIMULATION
    TO THE BEGINNING OF THE CURRENT EVENT.  *** ;
  * IF THE CURRENT VALUE OF CUMHOURS=0, THEN THIS IS THE FIRST
    SIMULATED EVENT, SET CUMHOURS=T.  IF CUMHOURS>0, THEN SET
    CUMHOURS=CUMHOURS+DLAST(I)+T.  * ;
I=1;
IF CUMHOURS=0 THEN CUMHOURS=T;
ELSE CUMHOURS=CUMHOURS+DLAST(I)+T;
*** RESET RLAST(I), TLAST(I), AND DLAST(I).  *** ;
  * DELETE THE CURRENT DATA FOR I=36, SLIDE ALL REMAINING DATA DOWN
    ONE STEP, AND SET THE VALUES FOR I=1 TO CURRENT STORM VALUES.  * ;
DO I=36 TO 2 BY -1;
RLAST(I)=RLAST(I-1);
TLAST(I)=TLAST(I-1);
DLAST(I)=DLAST(I-1);
END;
I=1;
RLAST(I)=R;
TLAST(I)=T;
DLAST(I)=D;
END;
END;
*** INCREMENT RUNOFF EVENT COUNTERS SINCE BOTH Q>0 AND QP>0.  *** ;
NRUNOFFS=NRUNOFFS+1;
YEAR=1+INT(CUMHOURS/(365*24));
*** ADD EVENT TO SIMULATED DATA SET.  *** ;
  * OMIT ALL EVENTS OCCURRING DURING THE INITIALIZATION PERIOD AND THE
    LAST EVENT, WHICH OCCURS AFTER THE FINAL SIMULATION YEAR ENDS.    * ;
IF STARTYRS < YEAR <= NYEARS+STARTYRS THEN OUTPUT;
*** RESET TL=0 SINCE Q>0.  *** ;
TL=0;
*** RESET TR=0 IF Q>=0.01 INCH.  *** ;
IF Q>=0.01 THEN TR=0;
END;
```

TABLE C.2
FREQUENCY ANALYSIS FOR TOTAL ANNUAL RAINFALL VOLUME

RANNUAL	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
17.52	1	2.0	1	2.0
21.75	1	2.0	2	4.0
22.41	1	2.0	3	6.0
22.67	1	2.0	4	8.0
22.94	1	2.0	5	10.0
24.36	1	2.0	6	12.0
25.31	1	2.0	7	14.0
25.59	1	2.0	8	16.0
26.43	1	2.0	9	18.0
27.64	1	2.0	10	20.0
27.76	1	2.0	11	22.0
28.01	1	2.0	12	24.0
28.03	1	2.0	13	26.0
28.07	1	2.0	14	28.0
28.41	1	2.0	15	30.0
28.51	1	2.0	16	32.0
28.58	1	2.0	17	34.0
29.95	1	2.0	18	36.0
30.2	1	2.0	19	38.0
30.76	1	2.0	20	40.0
32.4	1	2.0	21	42.0
32.88	1	2.0	22	44.0
34.34	1	2.0	23	46.0
35.5	1	2.0	24	48.0
36.19	1	2.0	25	50.0
37.47	1	2.0	26	52.0
37.8	1	2.0	27	54.0
37.99	1	2.0	28	56.0
38.55	1	2.0	29	58.0
38.78	1	2.0	30	60.0
38.9	1	2.0	31	62.0
39.33	1	2.0	32	64.0
39.37	1	2.0	33	66.0
39.62	1	2.0	34	68.0
39.99	1	2.0	35	70.0
40.75	1	2.0	36	72.0
41.74	1	2.0	37	74.0
42.57	1	2.0	38	76.0
42.58	1	2.0	39	78.0
42.68	1	2.0	40	80.0
42.71	1	2.0	41	82.0
42.97	1	2.0	42	84.0
44.05	1	2.0	43	86.0
44.57	1	2.0	44	88.0
44.93	1	2.0	45	90.0
47.96	1	2.0	46	92.0

TABLE C.2
FREQUENCY ANALYSIS FOR TOTAL ANNUAL RAINFALL VOLUME

RANNUAL	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
49.19	1	2.0	47	94.0
54.74	1	2.0	48	96.0
55.14	1	2.0	49	98.0
55.31	1	2.0	50	100.0

TABLE C.3
FREQUENCY ANALYSIS FOR SIMULATED TOTAL ANNUAL RUNOFF VOLUME

QSUM	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
0	1	1.0	1	1.0
0.05222254	1	1.0	2	2.0
0.114502	1	1.0	3	3.0
0.1591444	1	1.0	4	4.0
0.2796553	1	1.0	5	5.0
0.3431481	1	1.0	6	6.0
0.4081132	1	1.0	7	7.0
0.4103254	1	1.0	8	8.0
0.4758122	1	1.0	9	9.0
0.5096861	1	1.0	10	10.0
0.6354305	1	1.0	11	11.0
0.6987544	1	1.0	12	12.0
0.8836587	1	1.0	13	13.0
0.9410385	1	1.0	14	14.0
1.024767	1	1.0	15	15.0
1.079096	1	1.0	16	16.0
1.080932	1	1.0	17	17.0
1.153002	1	1.0	18	18.0
1.160326	1	1.0	19	19.0
1.183184	1	1.0	20	20.0
1.201762	1	1.0	21	21.0
1.203904	1	1.0	22	22.0
1.225803	1	1.0	23	23.0
1.285116	1	1.0	24	24.0
1.388023	1	1.0	25	25.0
1.433112	1	1.0	26	26.0
1.480747	1	1.0	27	27.0
1.53505	1	1.0	28	28.0
1.560076	1	1.0	29	29.0
1.655565	1	1.0	30	30.0
1.737681	1	1.0	31	31.0
1.742664	1	1.0	32	32.0
1.77271	1	1.0	33	33.0
1.776907	1	1.0	34	34.0
1.83502	1	1.0	35	35.0
1.866404	1	1.0	36	36.0
1.881162	1	1.0	37	37.0
1.906198	1	1.0	38	38.0
1.929005	1	1.0	39	39.0
1.929025	1	1.0	40	40.0
1.99013	1	1.0	41	41.0
2.161733	1	1.0	42	42.0
2.208659	1	1.0	43	43.0
2.23919	1	1.0	44	44.0
2.329871	1	1.0	45	45.0
2.339984	1	1.0	46	46.0

TABLE C.3
FREQUENCY ANALYSIS FOR SIMULATED TOTAL ANNUAL RUNOFF VOLUME

QSUM	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
2.382021	1	1.0	47	47.0
2.385732	1	1.0	48	48.0
2.389577	1	1.0	49	49.0
2.471421	1	1.0	50	50.0
2.523395	1	1.0	51	51.0
2.534713	1	1.0	52	52.0
2.571516	1	1.0	53	53.0
2.63216	1	1.0	54	54.0
2.63344	1	1.0	55	55.0
2.67889	1	1.0	56	56.0
2.688887	1	1.0	57	57.0
2.702663	1	1.0	58	58.0
2.800876	1	1.0	59	59.0
2.818818	1	1.0	60	60.0
2.841864	1	1.0	61	61.0
2.878732	1	1.0	62	62.0
2.929791	1	1.0	63	63.0
2.936071	1	1.0	64	64.0
2.937665	1	1.0	65	65.0
2.950182	1	1.0	66	66.0
3.021785	1	1.0	67	67.0
3.06038	1	1.0	68	68.0
3.07702	1	1.0	69	69.0
3.113839	1	1.0	70	70.0
3.119133	1	1.0	71	71.0
3.280916	1	1.0	72	72.0
3.363168	1	1.0	73	73.0
3.367551	1	1.0	74	74.0
3.396425	1	1.0	75	75.0
3.42031	1	1.0	76	76.0
3.499966	1	1.0	77	77.0
3.506609	1	1.0	78	78.0
3.557334	1	1.0	79	79.0
3.591218	1	1.0	80	80.0
3.6938	1	1.0	81	81.0
3.706633	1	1.0	82	82.0
3.710822	1	1.0	83	83.0
3.88329	1	1.0	84	84.0
3.959517	1	1.0	85	85.0
4.098741	1	1.0	86	86.0
4.183774	1	1.0	87	87.0
4.230102	1	1.0	88	88.0
4.421627	1	1.0	89	89.0
4.431908	1	1.0	90	90.0
4.580828	1	1.0	91	91.0
4.752908	1	1.0	92	92.0

TABLE C.3
FREQUENCY ANALYSIS FOR SIMULATED TOTAL ANNUAL RUNOFF VOLUME

QSUM	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
4.818077	1	1.0	93	93.0
4.906061	1	1.0	94	94.0
5.077805	1	1.0	95	95.0
5.133651	1	1.0	96	96.0
5.259998	1	1.0	97	97.0
5.749385	1	1.0	98	98.0
6.057894	1	1.0	99	99.0
6.203568	1	1.0	100	100.0

TABLE C.4
SAS PROGRAM LISTING FOR EVALUATION OF SAMPLE SIZE

```
OPTIONS NONUMBER NODATE NOSOURCE NONOTES;
CMS FI 12 DISK HRSAMP LISTING C;
***;
***  BEGIN USER INPUT.;
***;
*;
***  DEFINE THE INPUT PERMANENT SAS DATA SET NAME STORED ON DISK AND THEN
     DEFINE SEPARATLY THE FILENAME FILETYPE FILEMODE OF THIS FILE.;
*;
%LET DATANAME=DAT.EVENTSIM;
%LET FILENAME=EVENTSIM;
%LET FILETYPE=DAT;
%LET FILEMODE=B;
CMS FI &FILETYPE DISK &FILENAME &FILETYPE &FILEMODE;
*;
***  SET THE SAMPLE SIZE FOR LINEAR MODELING.;
*;
%LET SAMPSIZE=40;
*;
***  SET THE TOTAL NUMBER OF RUNOFF EVENTS IN THE SUBSET OF THE INPUT
     DISK FILE TO BE USED.;
*;
%LET TEVENTS=1499;
*;
***  SUBSET THE INPUT DATA FILE AS DESIRED.;
*;
DATA SIMULATE;
SET &DATANAME;
RETAIN NEVENTS 0;
KEEP NEVENTS CUMHOURS Q D TR;
*  USE ONLY EVENTS WITH Q=.01 INCH OR GREATER AND
   KEEP A CUMULATIVE TOTAL OF THESE EVENTS.;
IF Q<.01 THEN DELETE;
NEVENTS=NEVENTS+1;
RUN;
***;
***  END USER INPUT;
***;
*;
***  THE FINAL OUTPUT DISK FILE NAME IS AUTOMATICALLY
     DEFINED AS HRSAMP&SAMPSIZE.;
*;
CMS FI DAT DISK HRSAMP&SAMPSIZE DAT A;
*;
***  DEFINE THE MACRO FOR CREATING THE DATA SET WITH HOURS REQUIRED
     TO COLLECT SAMPLES OF SIZE "&SAMPSIZE";
*;
%MACRO TIME;
%*;
%***  THIS MACRO READS ALL POSSIBLE SUBSETS OF DATA OF
      SAMPLE SIZE "SAMPSIZE", IN SEQUENCE, FROM THE INPUT
      DATA FILE AND THEN CREATES AN OUTPUT DATA FILE CONTAINING
      THE TIME (IN HOURS) REQUIRED TO COLLECT SAMPLES OF THAT SIZE.
      THE TOTAL NUMBER OF SUCH SUBSETS IS (TEVENTS-SAMPSIZE)+1.;
```

TABLE C.4
SAS PROGRAM LISTING FOR EVALUATION OF SAMPLE SIZE

```
%*;
%DO K=&SAMPSIZE %TO &TEVENTS;
DATA HRSAMP;
SET SIMULATE;
KEEP HRSAMPLE;
RETAIN START 0;
IF NEVENTS<=(&K-&SAMPSIZE) THEN DELETE;
IF NEVENTS>&K THEN DELETE;
*  COMPUTE THE STARTING VALUE OF CUMHOURS AT THE END OF THE LAST
   RUNOFF EVENT WITH Q>=0.01 INCH BEFORE THE CURRENT SUBSET;
IF NEVENTS=((&K-&SAMPSIZE)+1) THEN START=(CUMHOURS-TR);
*  COMPUTE THE ENDING VALUE OF CUMHOURS AT THE END OF THE LAST
   EVENT IN THE SUBSET;
END=(CUMHOURS+D);
*  COMPUTE THE TOTAL TIME TO COLLECT THE CURRENT
   SAMPLE OF SIZE "SAMPSIZE";
HRSAMPLE=END-START;
IF NEVENTS=&K THEN OUTPUT;
PROC APPEND OUT=DAT.HRSAMP&SAMPSIZE DATA=HRSAMP;
%END;
%MEND;
*;
***  INVOKE THE MACRO TIME;
*;
%TIME
 /*
PROC FREQ DATA=HRSAMP&SAMPSIZE;
TABLES HRSAMPLE;
TITLE1 'FREQUENCY DISTRIBUTION FOR';
TITLE2 'HOURS TO COLLECT A SET OF RUNOFF VOLUME DATA';
TITLE3 "SAMPLES OF SIZE &SAMPSIZE";
 */
```

REFERENCE LIST

Arnold, J.G., and J.R. Williams. 1985. SWRRB, A Simulator for Water Resources in Rural Basins; Volume I, Model Documentation. Temple, Texas: U.S. Department of Agriculture, Agricultural Research Service.

Arnold, J.G., et al. 1986. Modeling sediment yields into Stamford Lake, Texas. Proceedings Fourth Federal Interagency Sedimentation Conference. Washington, D.C.: U.S. Department of Agriculture, Agricultural Research Service.

Austin, P.M., and R.A. Houze, Jr. 1972. Analysis of the structure of precipitation patterns in New England. Journal of Applied Meteorology 11: 926-935.

Barton, B.M.J. 1974. The generation of synthetic monthly runoff records for ungaged British catchments. Mathematical Models in Hydrology: Proceedings of the Warsaw Symposium 1: 121-133. International Association of Hydrological Sciences Publication No. 100.

Bhat, U.N. 1984. Time series analysis. Chapter 21 of Elements of Applied Stochastic Processes. New York: John Wiley & Sons.

Box, E.P., and G.M. Jenkins. 1970. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day.

Bras, R.L., and S. Chan. 1978. Theoretical models of hydrologic parameters using derived distribution techniques. Proceedings International Symposium on Risk and Reliability in Water Resources, eds. E.A. McBean, K.W. Hipel, and T.E. Unny, 2: 500-520. Waterloo, Ontario: University of Waterloo.

Brown, J.A.H. 1975. Data needs, acquisition, and availability for hydrologic models. Prediction in Catchment Hydrology, eds. T.G. Chapman and F.X. Dunin, 429-455. Canberra, Australia: Australian Academy of Science.

Caroni, E., F. Mannocchi, and L. Ubertini. 1982. Rainfall-flow relationship in some Italian rivers by multiple stochastic models. Time Series Methods in Hydrosciences, eds. A.H. El-Shaarawi and S.R. Esterby, 455-464. Amsterdam: Elsevier.

Caskey, I.E., Jr. 1963. A Markov chain model for the rainfall occurrence in intervals at various lengths. Monthly Weather Review 91: 289-301.

Chang, T.J., M.L. Kavvas, and J.W. Delleur. 1982. Stochastic Daily Precipitation Modeling and Daily Streamflow Transfer Processes. Lafayette, Indiana: Water Resources Research Center, Purdue University. NTIS PB83-108191.

Chang, T.J., M.L. Kavvas, and J.W. Delleur. 1984. Daily precipitation modeling by discrete autoregressive moving average processes. Water Resources Research 20(5): 565-580.

Chow, V.T. 1964. Handbook of Applied Hydrology. New York: McGraw Hill.

Court, A. 1979. Precipitation research, 1975-1978. Papers in Hydrology: U.S. National Report to the IUGG, 1975-1978, 1165-1175. Washington, D.C.: American Geophysical Union.

Daniel, W.W. 1978. Applied Nonparametric Statistics. Boston: Houghton Mifflin.

Driver, N.E., and D.J. Lystrom. 1986. Estimation of urban storm-runoff loads. Urban Runoff Quality - Impact and Quality Enhancement Technology, eds. B. Urbonas and L.A. Roesner, 122-132. New York: American Society of Civil Engineers.

Eagleson, P.S. 1970. Precipitation. Chapter 11 of Dynamic Hydrology: 184-185. New York: McGraw-Hill.

Eagleson, P.S. 1971. The stochastic kinematic wave. Systems Approach to Hydrology: Proceedings of the First Bilateral U.S. - Japan Seminar in Hydrology. Fort Collins, Colorado: Water Resources Publications.

Eagleson, P.S. 1972. Dynamics of flood frequency. Water Resources Research 8(4): 878-898.

Eagleson, P.S. 1978. Climate, soil, and vegetation, 5: A derived distribution of storm surface runoff. Water Resources Research 14(5): 741-748.

Ellis, J.B. 1986. Physical and chemical data needs for design. Urban Runoff Quality - Impact and Quality Enhancement Technology, eds. B. Urbonas and L.A. Roesner, 49-59. New York: American Society of Civil Engineers.

El-Shaarawi, A.H., and S.R. Esterby, ed. 1982. Time Series Methods in Hydrosciences. Amsterdam: Elsevier.

Farnworth, E.G., et al. 1979. Impacts of sediment on biota in surface waters. Chapter 9 of Impacts of Sediment and Nutrients on Biota in Surface Waters of the United States, 183-230. Athens, Georgia: U.S. Environmental Protection Agency, Environmental Research Laboratory. Report EPA-600/3-79-105.

Feyerherm, A.M., and L.D. Bark. 1967. Goodness of fit of a Markov chain model for sequences of wet and dry days. Journal of Applied Meteorology 6: 770-773.

Fiering, M.B., and B.B. Jackson. 1971. Synthetic Streamflows. Washington, D.C.: American Geophysical Union. Water Resources Monograph 1.

Frind, E.O. 1969. Rainfall-runoff relationships expressed by distribution parameters. Journal of Hydrology 9: 405-426.

Gabriel, K.R., and J. Neumann. 1962. A Markov chain model for daily rainfall occurrence at Tel Aviv. Quarterly Journal of the Royal Meteorological Society 88: 90-95.

Georgakakos, K.P. 1986. Generalized stochastic hydrometeorological model for flood and flash-flood forecasting: 2. Case studies. Water Resources Research 22(13): 2096-2106.

Green, J.R. 1965. Two probability models for sequences of wet and dry days. Monthly Weather Review 93: 155-156.

Haan, C.T. 1977. Statistical Methods in Hydrology. Ames, Iowa: The Iowa State University Press.

Hanson, C.L., and D.A. Woolhiser. 1978. Probable effect of summer weather modification on runoff. Journal of the Irrigation and Drainage Division, American Society of Civil Engineers 104 (IR1): 1-11.

Harter, H.L., and D.B. Owen, ed. 1970. Selected Tables in Mathematical Statistics 1: 79-170. Chicago: Markham Publishing Company.

Hemain, J.C. 1986. Statistically based modeling of urban quality. Proceedings of the NATO Workshop on Urban Runoff Pollution, 277-304. Heidelberg, Federal Republic of Germany: NATO.

Hipel, K.W., A.I. McLeod, and D.J. Noakes. 1982. Fitting dynamic models to hydrological time series. Time Series Methods in Hydrosciences, eds. A.H. El-Shaarawi and S.R. Esterby, 110-129. Amsterdam: Elsevier.

Hipel, K.W., ed. 1985. Time Series Analysis in Water Resources. Bethesda, Maryland: American Water Resources Association.

Hoel, P.G. 1984. Introduction to Mathematical Statistics. New York: John Wiley and Sons.

Huber, W.C., et al. 1977. Interim Documentation, November 1977 Release of EPA SWMM. Cincinnati, Ohio: U.S. Environmental Protection Agency, National Environmental Research Center.

Huber, W.C., et al. 1979. Urban Rainfall-Runoff-Quality Data Base. Washington, D.C.: U.S. Environmental Protection Agency. Report EPA-600/8-79-004.

Huber, W.C.. 1986. Modeling urban runoff quality: state-of-the-art. Urban Runoff Quality - Impact and Quality Enhancement Technology, eds. B. Urbonas and L.A. Roesner, 34-48. New York: American Society of Civil Engineers.

Jennings, M.E. 1982. Data collection and instrumentation. Chapter 7 of Urban Stormwater Hydrology, ed. D.F. Kibler, 189-217. Washington, D.C.: American Geophysical. Union Water Resources Monograph 7.

Kavvas, M.L., and J.W. Delleur. 1981. A stochastic cluster model of daily rainfall sequences. Water Resources Research 17(4): 1151-1160.

Klemes, V. 1974. Probability distribution of outflow from a linear reservoir. Journal of Hydrology 21: 305-314.

Koch, R.W. 1985. A stochastic streamflow model based on physical principles. Water Resources Research 21(4): 545-553.

Kottegoda, N.T. 1980. Stochastic Water Resources Technology. New York: John Wiley & Sons.

Loucks, D.P., J.R. Stedinger, and D.A. Haith. 1981. Synthetic streamflow generation. Chapter 6 of Water Resource Systems Planning and Analysis. Englewood Cliffs, New Jersey.: Prentice-Hall.

Mancini, J.L., and A.H. Plummer. 1986. Urban runoff and quality criteria. Urban Runoff Quality - Impact and Quality Enhancement Technology, eds. B. Urbonas and L.A. Roesner, 133-149. New York: American Society of Civil Engineers.

Marselek, J. 1986. Report on NATO workshop on urban runoff quality. Urban Runoff Quality - Impact and Quality Enhancement Technology, eds. B. Urbonas and L.A. Roesner, 15-28. New York: American Society of Civil Engineers.

Matalas, N.C. 1967. Mathematical assessment of synthetic hydrology. Water Resources Research 3(4): 937-945.

McCuen, R.H., and W.M. Snyder. 1986. Time series analysis and synthesis. Chapter 9 of Hydrologic Modeling: Statistical Methods and Applications. Englewood Cliffs, New Jersey: Prentice-Hall.

Medina, M.A., Jr. 1979. Level III: Receiving Water Quality Modeling for Urban Stormwater Management. Washington, D.C.: U.S. Environmental Protection Agency. Report EPA-600/2-79-100.

Mimikou, M., and A.R. Rao. 1982. A rainfall-runoff model for daily flow synthesis. Time Series Methods in Hydrosciences, eds. A.H. El-Shaarawi and S.R. Esterby. Amsterdam: Elsevier.

Morris, C.D., J.W. Hood, and J.A. Ferguson. 1984. Quantitative Aspects of Water Resource Management. Rolla, Missouri: Dept. of Civil Engineering, Missouri University-Rolla. NTIS, PB85-230092/AS.

Niedzialkowski, D., and D. Athayde. 1985. Water quality data and urban nonpoint source pollution: the nationwide urban runoff program. Proceedings Perspectives on Nonpoint Source Pollution. Washington, D.C.: U.S. Environmental Protection Agency.

North Central Texas Council of Governments. 1978. An Assessment of Pollutant Loadings to Lakes and Rivers in North Central Texas. Arlington, Texas: North Central Texas Council of Governments.

North Central Texas Council of Governments. 1984. Stormwater Management Handbook: Stormwater Estimation and Management Techniques for North Central Texas. Arlington, Texas: North Central Texas Council of Governments.

Roesner, L.A. 1982. Quality of urban runoff. Chapter 6 of Urban Stormwater Hydrology, ed. D.F. Kibler, 161-187. Washington, D.C.: American Geophysical Union. Water Resources Monograph 7.

Roles, J.M., and P.J. Jonker. 1985. Representation and accuracy of measurement of soil loss from runoff plots. Transactions of the American Society of Agricultural Engineers 28:1458.

SAS Institute, Inc. 1985a. SAS User's Guide: Basics, Version 5 Edition. Cary, North Carolina: SAS Institute Inc.

SAS Institute, Inc. 1985b. SAS User's Guide: Statistics, Version 5 Edition. Cary, North Carolina: SAS Institute Inc.

Sharma, T.C. 1985. Stochastic models applied to evaluating hydrologic changes. Journal of Hydrology 78: 61-81.

Singh, V.P., and Y.K. Birsoy. 1977. Some statistical relationships between rainfall and runoff. Journal of Hydrology 34: 251-268.

Singh, V.P. 1988. Hydrologic Systems, Volume I: Rainfall-Runoff Modeling. Englewood Cliffs, New Jersey: Prentice-Hall.

Sonnen, M.B. 1986. Review of data needs and collection technology. Urban Runoff Quality - Impact and Quality Enhancement Technology, eds. B. Urbonas and L.A. Roesner, 95-103. New York: American Society of Civil Engineers.

Steger, J.A., ed. 1971. Chi-square. Chapter 2 in Readings in Statistics: For the Behavioral Scientist. New York: Holt, Reinhart, and Winston.

Tarboton, D.G., R.L. Bras, and C.E. Puente. 1987. Combined hydrologic sampling criteria for rainfall and streamflow. Journal of Hydrology 95: 323-339.

Terstriep, M.L., D.C. Noel, and G.M. Bender. 1986. Sources of urban pollution - do we know enough? Urban Runoff Quality - Impact and Quality Enhancement Technology, eds. B. Urbonas and L.A. Roesner, 107-121. New York: American Society of Civil Engineers.

Thames, J.L. 1984. Effect of Stockponds on Downstream Surface Water Yields. Tucson, Arizona: Arizona Water Resources Research Center, Arizona University. NTIS, PB85-233005/AS.

Thomas, H.A., Jr., and M.B. Fiering. 1962. Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. Design of Water Resources Systems, ed. A. Maass, et al. Cambridge, Massachusetts: Harvard University Press.

Thompstone, R.M., K.W. Hipel, and A.I. McLeod. 1985. Forecasting quarter-monthly riverflow. Time Series Analysis in Water Resources, ed. K.W. Hipel, 731-741. Bethesda, Maryland: American Water Resources Association.

Tong, H., B. Thanoon, and G. Gudmundsson. 1985. Threshold modelling of two Icelandic riverflow systems. Time Series Analysis in Water Resources, ed. K.W. Hipel, 651-661. Bethesda, Maryland: American Water Resources Association.

U.S. Army Corps of Engineers. Hydrologic Engineering Center. 1974. Urban Storm Water Runoff: Storage, Treatment, Overflow, Runoff Model. Davis, California: U.S. Army Corps of Engineers.

U.S. Department of Agriculture. 1984. CREAMS: A Field Scale Model for Chemicals, Runoff, and Erosion from Agricultural Management Systems. Washington, D.C.: U.S. Department of Agriculture. Conservation Report 26.

U.S. Department of Agriculture. Agricultural Research Service. 1985. Proceedings of the Natural Resources Modeling Symposium, ed. D. G. DeCoursey. Fort Collins, Colorado: U.S. Department of Agriculture.

U.S. Department of Agriculture. Soil Conservation Service. 1971. Hydrology. Section 4 of National Engineering Handbook. Washington, D.C.: U.S. Department of Agriculture.

U.S. Department of Commerce. National Oceanic and Atmospheric Administration. 1939-1988. Climatological Summary-Texas. Asheville, North Carolina: National Climatic Data Center.

U.S. Department of Commerce. National Oceanic and Atmospheric Administration. 1985. Diskette Documentation for NCDC's Element Archive Format Release B - Condensed. Asheville, North Carolina: National Climatic Data Center.

U.S. Department of Commerce. National Oceanic and Atmospheric Administration. 1986. Hourly Precipitation (TD-3240). Asheville, North Carolina: National Climatic Data Center.

U.S. Environmental Protection Agency. 1971. Storm Water Management Model, Final Report. Washington, D.C.: U.S. Environmental Protection Agency. Report 11024DOC07/71.

U.S. Environmental Protection Agency. 1973. Methods and Practices for Controlling Water Pollution from Agricultural Nonpoint Sources. Washington, D.C.: U.S. Environmental Protection Agency. Report EPA-430/9-73-015.

U.S. Environmental Protection Agency. Water Planning Division. 1983. Results of the National Urban Runoff Program. Washington, D.C.: U.S. Environmental Protections Agency. Report WH-554.

Urbonas, B., and L.A. Roesner. 1986. Conference overview. Urban Runoff Quality - Impact and Quality Enhancement Technology, eds. B. Urbonas and L.A. Roesner, 1-9. New York: American Society of Civil Engineers.

Weiss, L.L. 1964. Sequences of wet and dry days described by a Markov chain model. Monthly Weather Review 92: 169-176.

Wiser, E.H. 1965. Modified Markov probability models of sequences of precipitation events. Monthly Weather Review 93: 511-516.

Yevjevich, V. 1972. Probability and Statistics in Hydrology. Fort Collins, Colorado: Water Resources Publications.