A STATISTICAL ANALYSIS OF THE DAILY STREAMFLOW HYDROGRAPH AND A PROBABILITY DISTRIBUTION OF THE INTERARRIVAL TIMES OF PEAK FLOWS - APPLICATION TO INDIANA

by

M. L. Kavvas
J. W. Delleur

January 1983

PURDUE UNIVERSITY
WATER RESOURCES RESEARCH CENTER
WEST LAFAYETTE, INDIANA
A STATISTICAL ANALYSIS OF THE DAILY STREAMFLOW HYDROGRAPH AND
A PROBABILITY DISTRIBUTION OF THE INTERARRIVAL TIMES OF PEAK
FLOWS - APPLICATION TO INDIANA

by

M.L. Kavvas

and

J.W. Delleur

The work upon which this publication is based was supported in part by funds provided by the United States Department of the Interior, Office of Water Research and Technology of the Bureau of Reclamation, as authorized by the Water Research Act of 1978 (Public Law 95-467).

Period of Investigation: October 1979 - December 1982

Completion Report for Matching Fund Project
OWRT-B-112-IND
Agreement No. 14-34-0001-0219

Purdue University Water Resources Research Center
Technical Report No. 147

January, 1983
ACKNOWLEDGEMENTS

The authors wish to thank Dr. Luis L. Cote, of the Statistics Department of Purdue University, for his kind assistance and review of the mathematical developments. We are grateful to Dr. Dan Wiersma, retired, Director of the Purdue University Water Resources Research Center for his assistance in the administration of the project.
PREFACE

This report is the last of a series of three reports which summarize the work performed on Project OWRT-B-112-IND entitled, "Simulation of Short-Time Rainfall, Flood and Drought Sequences by Stochastic Point Processes and Application to the Ohio River Basin and Central United States". The development of a model of the daily rainfall sequences was the object of technical report 146 completed in June 1982. The development of a two-dimensional cluster model for flood analysis was described in technical report 145, completed in December 1982. Finally, the current report (147) is on a statistical analysis of the daily streamflow hydrograph and on the probability distribution of droughts or more precisely of the interarrival time between hydrograph peaks taking into account the time and discharge dependence.

Regarding the daily rainfall sequences two newly defined stochastic processes are proposed for the modeling of daily precipitation time series. The first uses the Binary Discrete Autoregressive Moving Average model for the wet-dry day sequence and it is mixed with an exponential distribution to express the magnitude of the precipitation (B-DARMA-E). The second uses the Multi-state Discrete Autoregressive Moving Average process (M-DARMA). Both processes are used to model the daily precipitation time series at several stations in Indiana. The binary run length distribution is obtained for the first model while the multi-state run length distributions are constructed for the latter. The model is identified and the parameters are estimated
making use of the autocorrelation function while the diagnostic checking and the criterion for the selection of the best model among competing ones are based on run length distributions. These procedures are very effective for five Indiana watersheds tested. A Transfer Discrete Autoregressive Moving Average Model was constructed to obtain daily streamflows by inputting the daily precipitation generated by a B-DARMA-E or a M-DARMA model.

Regarding flood analysis, a two-level and two-dimensional non-homogeneous point stochastic process was developed to model the flood peaks of a hydrograph. The model is a cluster process of the Neyman–Scott Type and presents the occurrence of flood generating mechanisms at the precipitation level as the triggers for clusters of flood peaks at the runoff level. The statistical properties of the flood cluster process are found in terms of the probability generating functional of the process. The theoretical rate of occurrence, the theoretical covariance density and the theoretical probability mass function of the process are compared to the respective empirical functions obtained for several stations located in the Ohio River Basin and a good fit was found for the analyzed stations.

The present report deals with the estimation of the mean function and of the covariance function of daily streamflows, the probability distribution of the interarrival times of peak flows and the probability distribution of the time to the peak discharge.
Previous reports and publications on this project are the following:

Reports


Theses


Conference Presentations


"DARMA Processes for Daily Precipitation Modeling", by J.W. Delleur, T.J. Chang and M.L. Kavvas, Paper presented at the Fall Meeting of the American Geophysical Union,


Scientific Journal Publications


ABSTRACT

In this study a periodic statistical analysis of the daily streamflow data at five streamgaging stations in Indiana was performed to gain insight into the stochastic structure which governs the daily streamflow process. This analysis was performed by the mean function and the covariance function of the daily streamflows, by the recession limb of the daily streamflow hydrograph, by the probability distribution of the hydrograph peak interarrival time, and by the probability distribution of the time to the peak discharge. New statistical estimators were developed and used in this study. In gross features this analysis has shown that a) the daily streamflow process is annually periodic in all of its statistical functions and b) if the daily streamflow process is modeled as the release from a linear watershed storage, this release should depend on the state of the storage and on the time of the release as the persistence properties and the recession limb decay rates were observed to change with the state of the storage and the time. Therefore, a time-varying reservoir system needs to be considered for a comprehensive model of the daily streamflow hydrograph.

The results of the statistical analysis served as a basis for the derivation of a theoretical probability distribution of the interarrival times between hydrograph peaks. These interarrival times, and their probability distributions are dependent on
the time of the year and on the discharge exceedance level. The theoretical derivation of the distribution is based on the stochastic trigger model of flood peaks. (Kavvas, 1982)

The derived distribution is an approximation to the complete distribution in that it only depends on the most recent peak from which the interarrival time to the next peak is measured. However, this approximation has had limited success, due mainly to the averaging of the distribution’s parameter functions over large time intervals on the order of 15 days. It is believed that by a finer estimation of the parameter functions it is possible to replicate the observed probability behavior of the hydrograph peak interarrival times which are used in the drought descriptions.
LIST OF TABLES

TABLE 1. The streamflow gaging stations which were used for the analysis of the daily streamflow sequences in Indiana. 10

TABLE 2. Approximate time boundaries of the seasons determined from the estimated covariance functions. 28

TABLE 3. A summary of the features of the estimated interarrival time probability distributions of the hydrograph peaks at five gaging stations in Indiana. 76

TABLE 4. Kolmogorov-Smirnov test results of the cumulative distribution function of the hydrograph peak interarrival times at station 35245. 109
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Watersheds used for analysis.</td>
<td>11</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Mean function of the daily streamflow.</td>
<td>17</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Covariance functions of the daily streamflows at the station 32750.</td>
<td>23</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Covariance functions of the daily streamflows at the station 33030.</td>
<td>24</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Covariance functions of the daily streamflows at the station 33245.</td>
<td>25</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Covariance functions of the daily streamflows at the station 33265.</td>
<td>26</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Covariance functions of the daily streamflows at the station 33655.</td>
<td>27</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Recession decay rate function at station 32750.</td>
<td>36</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Recession decay rate function at station 33030.</td>
<td>37</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Recession decay rate function at station 33245.</td>
<td>38</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Recession decay rate function at station 33265.</td>
<td>39</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Recession decay rate function at station 33655.</td>
<td>40</td>
</tr>
<tr>
<td>Figure 13.a</td>
<td>Relative frequency histogram of the interarrival times at station 32750</td>
<td></td>
</tr>
<tr>
<td>Figure 13.a.1</td>
<td>Exceedence level = 0.</td>
<td>46</td>
</tr>
<tr>
<td>Figure 13.a.2</td>
<td>Exceedence level = mean</td>
<td>47</td>
</tr>
<tr>
<td>Figure 13.a.3</td>
<td>Exceedence level = mean + std. deviation.</td>
<td>48</td>
</tr>
<tr>
<td>Figure 13.a.4</td>
<td>Exceedence level = mean + 2 std. deviations.</td>
<td>49</td>
</tr>
<tr>
<td>Figure 13.b</td>
<td>Cumulative distributions of the interarrival times at station 32750.</td>
<td>50</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Figure 14.a</td>
<td>Relative frequency histogram of the interarrival times at station 33030</td>
<td>51</td>
</tr>
<tr>
<td>Figure 14.a.1</td>
<td>Exceedence level = 0.</td>
<td>52</td>
</tr>
<tr>
<td>Figure 14.a.2</td>
<td>Exceedence level = mean</td>
<td>53</td>
</tr>
<tr>
<td>Figure 14.a.3</td>
<td>Exceedence level = mean + std. deviation.</td>
<td>54</td>
</tr>
<tr>
<td>Figure 14.a.4</td>
<td>Exceedence level = mean + 2 std. deviations.</td>
<td>55</td>
</tr>
<tr>
<td>Figure 14.b</td>
<td>Cumulative distributions of the interarrival times at station 33030.</td>
<td>56</td>
</tr>
<tr>
<td>Figure 15.a</td>
<td>Relative frequency histogram of the interarrival times at station 33245</td>
<td>57</td>
</tr>
<tr>
<td>Figure 15.a.1</td>
<td>Exceedence level = 0.</td>
<td>58</td>
</tr>
<tr>
<td>Figure 15.a.2</td>
<td>Exceedence level = mean</td>
<td>59</td>
</tr>
<tr>
<td>Figure 15.a.3</td>
<td>Exceedence level = mean + std. deviation.</td>
<td>60</td>
</tr>
<tr>
<td>Figure 15.a.4</td>
<td>Exceedence level = mean + 2 std. deviations.</td>
<td>61</td>
</tr>
<tr>
<td>Figure 15.b</td>
<td>Cumulative distributions of the interarrival times at station 33245.</td>
<td>62</td>
</tr>
<tr>
<td>Figure 16.a</td>
<td>Relative frequency distribution of the interarrival times at station 33265</td>
<td>63</td>
</tr>
<tr>
<td>Figure 16.a.1</td>
<td>Exceedence level = 0.</td>
<td>64</td>
</tr>
<tr>
<td>Figure 16.a.2</td>
<td>Exceedence level = mean</td>
<td>65</td>
</tr>
<tr>
<td>Figure 16.a.3</td>
<td>Exceedence level = mean + std. deviation.</td>
<td>66</td>
</tr>
<tr>
<td>Figure 16.a.4</td>
<td>Exceedence level = mean + 2 std. deviations.</td>
<td>67</td>
</tr>
<tr>
<td>Figure 16.b</td>
<td>Cumulative distributions of the interarrival times at station 33265.</td>
<td>68</td>
</tr>
<tr>
<td>Figure 17.a</td>
<td>Relative frequency histogram of the interarrival times at station 33655</td>
<td>69</td>
</tr>
<tr>
<td>Figure 17.a.1</td>
<td>Exceedence level = 0.</td>
<td>70</td>
</tr>
<tr>
<td>Figure 17.a.2</td>
<td>Exceedence level = mean</td>
<td>71</td>
</tr>
<tr>
<td>Figure 17.a.3</td>
<td>Exceedence level = mean + std. deviation.</td>
<td>72</td>
</tr>
<tr>
<td>Figure 17.a.4</td>
<td>Exceedence level = mean + 2 std. deviations.</td>
<td>73</td>
</tr>
<tr>
<td>Figure 17.b</td>
<td>Cumulative distribution of interarrival times at station 33655.</td>
<td>74</td>
</tr>
<tr>
<td>Figure 18.a</td>
<td>Relative frequency histogram of the interarrival times with 50-day sampling at station 33245</td>
<td>75</td>
</tr>
<tr>
<td>Figure 18.a.1</td>
<td>Exceedence level = 0.</td>
<td>76</td>
</tr>
<tr>
<td>Figure 18.a.2</td>
<td>Exceedence level = mean</td>
<td>77</td>
</tr>
<tr>
<td>Figure 18.a.3</td>
<td>Exceedence level = mean + std. deviation.</td>
<td>78</td>
</tr>
<tr>
<td>Figure 18.a.4</td>
<td>Exceedence level = mean + 2 std. deviations.</td>
<td>79</td>
</tr>
</tbody>
</table>
**LIST OF FIGURES (Continued)**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 19.a.</td>
<td>Relative frequency histogram of the times to peak at station 32750.</td>
</tr>
<tr>
<td>Figure 19.b.</td>
<td>Cumulative distribution of the times to peak at station 32750.</td>
</tr>
<tr>
<td>Figure 20.a.</td>
<td>Relative frequency histogram of the times to peak at station 33050.</td>
</tr>
<tr>
<td>Figure 20.b.</td>
<td>Cumulative distribution of the times to peak at station 33050.</td>
</tr>
<tr>
<td>Figure 21.a.</td>
<td>Relative frequency histogram of the times to peak at station 33245.</td>
</tr>
<tr>
<td>Figure 21.b.</td>
<td>Cumulative distribution of the times to peak at station 33245.</td>
</tr>
<tr>
<td>Figure 22.a.</td>
<td>Relative frequency histogram of the times to peak at station 33265.</td>
</tr>
<tr>
<td>Figure 22.b.</td>
<td>Cumulative distribution of the times to peak at station 33265.</td>
</tr>
<tr>
<td>Figure 23.a.</td>
<td>Relative frequency histogram of the times to peak at station 33655.</td>
</tr>
<tr>
<td>Figure 23.b.</td>
<td>Cumulative distribution of the times to peak at station 33655.</td>
</tr>
<tr>
<td>Figure 24.</td>
<td>Schematic of drought period.</td>
</tr>
<tr>
<td>Figure 25.</td>
<td>The sample and theoretical cumulative distributions of the peak interarrival time for the 15-day interval that ends on April 28 at station 33245.</td>
</tr>
<tr>
<td>Figure 26.</td>
<td>Sample and theoretical cumulative distributions of the peak interarrival time for the 15-day interval that ends on January 23 at station 33245.</td>
</tr>
<tr>
<td>Figure 27.</td>
<td>Sample and theoretical cumulative distributions of the peak interarrival time for the 15-day interval that ends on March 29 at station 33245.</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

1.1 THE PLACE OF THIS STUDY IN WATER RESOURCES PLANNING

In the planning of urban water supply systems the planner needs to have daily or smaller-time-increment streamflow data in order to perform effective operation studies on the system (Weiss, 1973). Meanwhile, an important engineering problem related to floods is the derivation of the optimal flood control policies for the reservoirs in a river basin so as to determine the storage allocations for each operation period. For this purpose reservoir operation studies are performed in six hour or smaller time increments. In the flood control operation studies the historical and the synthetic streamflow sequences are routed through the reservoirs in the river basin. Consequently, the time increments of these sequences should be of the order of six hours or smaller. Therefore, a stochastic model is needed to generate synthetic streamflow sequences with the time increments less than or equal to a day, to be used in the flood control and water supply operation studies. As will be seen in the next section, various authors have constructed daily and continuous-time stochastic streamflow models. However, all of these attempts have had limited success, basically due to the lack of a rigorous statistical analysis of the basic statistical features of the daily or smaller-time-increment streamflow process. In this study an attempt towards the development of a statistical methodology for the analysis of a small-time-increment streamflow process will be made through the development of various statistical
estimators for the various features of the daily streamflow process. It is believed that such a methodology would enable the hydrologists to better understand the statistical properties of the small-time-increment streamflow process.

Any water utilization project in a geographical region must be based on the adequate knowledge of the drought characteristics in that region. One of the most important parameters of a drought event is the drought duration which is determined by the time locations of the hydrograph peaks immediately before and immediately after the drought period. Given the time location of the first hydrograph peak the interarrival time between the first and the second peak determines the time location of the second peak. Meanwhile, the droughts exhibit time-dependence in their stochastic structure. Therefore, using the results of the statistical analysis of the daily streamflow hydrograph, a time–discharge dependent theoretical distribution of the interarrival time between the hydrograph peak which starts the drought, and the peak which ends the drought will be developed and verified by the data.

1.2 A SURVEY OF LITERATURE

The first attempts to the stochastic modeling of the small-time-increment streamflow sequences were done using the autoregressive time series models based on the Gaussian distribution. These models were applied to the daily streamflow data (Quimpo, 1967; Payne et al. 1969) and yielded results that were not fully satisfactory.
They were based on a statistical analysis which only used the classical stationary covariance-spectrum estimators which can only detect periodicities in the mean (Kavvas and Delleur, 1975), and have ignored the statistical properties of the rising limb and the recession limb of the daily streamflow hydrograph. Weiss (1973) showed that the Gaussian autoregressive models cannot reproduce the rapid rises and slow recessions observed in the streamflow records of daily or smaller time intervals. Making an analogy to the theory of the unit hydrograph, he proposed the double shot-noise (filtered Poisson) process as a model for the continuous streamflow records. He used one shot-noise process for the surface runoff and another shot-noise process with different parameter values for the groundwater discharge so as to approximate the nonlinear behavior of the watershed system. His double shot-noise model was the summation of these two processes which he assumed to be independent of one another. Weiss then averaged the double shot-noise process in the daily intervals and applied the model to the daily streamflow records in England. The model could simulate the hydrograph peaks and recessions but could not simulate the group behavior of the hydrograph peaks and the low flow periods. Aside from being the first to develop a workable small-time-increment stochastic model, Weiss was also the first in developing a comprehensive model. Weiss was also the first in developing a comprehensive statistical analysis of the daily streamflow hydrograph. He estimated the mean, the standard deviation and the autocorrelation function (up to lag 10) of the daily flows in each month. Furthermore, he performed a
statistical analysis of the rising limb, of the recession limb and of the peak occurrence rate of the hydrograph, and of the interarrival time between the flood peaks. The important limitation of Weiss's work was that all of his statistical analysis was based on stationary estimators. This approach prevented him from obtaining a complete time-dependent statistical picture of the daily streamflow process.

Treiber and Plate (1975) assumed that the input rainfall pulses to the watershed system were a white-noise sequence and obtained the watershed system transfer function for the daily flows under this assumption. However, such a transfer function is simply the linear filter of the output (streamflow) series and cannot represent the input-output relation between an autocorrelated rainfall input series and an autocorrelated runoff output series. Later in their work Treiber and Plate (1975) assumed a Markov structure for the rainfall pulses and used these Markovian pulses as the input series. When the autocorrelated Markovian rainfall series are convoluted with the system transfer function which is only the linear filter of the output streamflow series, to obtain synthetic streamflow series, the autocorrelation function of the synthetic streamflow series will be different from the autocorrelation function of the historical streamflow series. The reason is that the system transfer function used in the convolution is only the linear filter of the output series and can only preserve the autocorrelation structure of the output streamflow series if the input is a white-noise sequence. The true
transfer function of the watershed can only be obtained by considering both the rainfall and the runoff series at the same time. Aside from these theoretical problems, in Treiber and Plate's work there is no unified statistical analysis to justify the fundamental assumptions concerning their model or to shed light to the statistical properties of the daily flows. There also seems to be a neglect of the periodicity of the daily flows since only one covariance function and only one transfer function are estimated for the whole process.

In constructing a daily flow model Kottegoda and Horder (1980) have used an alternating renewal process for the occurrence of rainy and dry days. To check their renewal process assumption they looked at the correlation coefficients of the lengths of adjacent wet and dry spells and have found that these coefficients are nearly zero to justify their assumption. Based on the chi-square tests they recommended the negative binomial distribution for the lengths of wet and dry spells. They made the fundamental assumption that the rainfall amounts on successive days are independent and gave no statistical justification of this crucial assumption. Based on the chi-square tests they recommended the use of a shifted gamma distribution for the rainfall amounts. Based on the independence assumption of the rainfall amounts and also assuming stationarity for the daily flows, they constructed a stationary system transfer function from a stationary covariance function estimate of the daily flows. Later, they generalized the system transfer function to a func-
tion which depends on the state of the flow when the rainfall pulse occurs. They then used regression equations to obtain effective rainfall amounts from the actual rainfall. Again this is only valid under the stationarity assumption. Finally, the daily flows were obtained from a convolution of the effective rainfalls with the variable transfer function. The daily flow model was compared to the historical record only in terms of the annual maxima of the observed flows and of the synthetic flows that are generated by the model. The results of this comparison are unsatisfactory. The work of Kottegoda and Horder (1980) suffers from lack of a comprehensive statistical analysis of the daily flows which in their case, led to the construction of an stochastic model which is not fully satisfactory.

Kelman (1980) has proposed another stochastic model for the daily flows. He assumed that the rising and the falling limbs of the hydrographs could be modeled separately "due to the fact that they translate different physical processes". For the rising limb he simply stated that it can be obtained by knowing the initial discharge at the start of the rise and the succession of discharges up to and including the peak discharge. On the other hand, he modeled the recession limb as the output from two parallel linear reservoirs, one for the groundwater storage and the other for the sum of surface detention, bank and channel storages. His model worked well for fast decreasing recession limbs with high peak discharge, but yielded unsatisfactory results when the peak discharge was low. The significant contri-
bution of Kelman (1980) with respect to the statistical analysis of daily flows was that he used time-varying mean, standard deviation and first-lag correlation coefficient estimators in order to draw inferences about the behavior of the first and second moments of the daily flow process. He also computed the recession curves of the observed daily flow series but only within the stationary domain. His use of constant decay coefficients resulted in unsatisfactory modeling of the recession limbs corresponding to the low peak discharges.

Assuming that the rainfall input to a watershed is a compound Poisson process, Pegram (1980) developed a state-space representation of the watershed in the form of a system of series and parallel reservoirs and showed that within a locally stationary time interval, the small-time-increment outflow from the watershed can be modeled in the form of an ARMA process. For the continuous hydrograph case, under the above assumptions, he naturally ended up with the filtered Poisson process. He then specified the response signal of the filtered Poisson process in terms of the parameters of the linear catchment model. All of the results of Pegram hold only for a time-invariant linear state-space representation of the watershed system (DeSoer, 1970). Recognizing this fact, Pegram proposed that all the time-nonhomogeneity of the outflow process be accounted for by a time-nonhomogeneous rainfall input process to a time-invariant deterministic linear watershed system, and applied his model to the hourly flow data within locally stationary seasons of the
year. Estimating his system model from the sample autocorrelation function of the hourly flows within the season he deconvoluted the outflow hydrograph with the system transfer function to obtain the effective rainfall record. From this record he calibrated the compound Poisson model and using this model as input to his watershed model, generated synthetic hourly runoff hydrographs. The results of this exercise as seen in Figure 9 and Table I of his paper are unsatisfactory. This is to be expected since the compound Poisson process assumption for the rainfall input is in conflict with the previous findings in hydrology about the rainfall process (see, for example, Kavvas and Delleur, 1975). Under the correlated rainfall input the continuous outflow process is no longer a filtered Poisson process and the whole application procedure of Pegram becomes redundant. In Pegram’s work there is no statistical analysis of the small-time-increment streamflow process and there is no hint as to how the locally stationary seasons within the year are determined.

From this survey of the hydrologic literature on the statistical analysis and modeling of the small-time-increment streamflow process one can conclude that a statistical methodology for the analysis of the small-time-increment streamflow process is needed to gain more insight into the various features of the process.

On the construction of the conditional probability distribution function of the interarrival time between the hydrograph peak that starts the drought and the peak that ends the drought
there is only the work of Kavvas and Delleur (1976). However, even this work is at the rainfall domain. Other works on this topic (e.g. Todorovic and Zelenhasic, 1970; Todorovic and Rousselle, 1971, and others) have all assumed independence among the interarrival times. However, one needs to use the information about the time location and magnitude of the peak that starts the drought, and about the properties of the interarrival times which have occurred in the most recent past to draw stronger inferences about the probability structure of the interarrival time pertaining to the drought period. Kavvas and Delleur (1976) have derived a conditional distribution of the interarrival time only for the stationary case. Although they had limited success in their application their results are encouraging (Yevjevich, et al. 1978).

1.3 OBJECTIVES OF THE REPORT

Following the discussion on sections 1.1 and 1.2, the objectives of this report may be stated as follows;

1. To develop a methodology for the statistical analysis of the small-time-increment runoff sequence and to apply it to the analysis of the daily runoff data in Indiana;

2. To develop and to test a theoretical time-discharge dependent probability distribution for the conditional probability distribution of the interarrival time between the hydrograph peak from which the drought
period starts and the peak where it ends.

1.4 THE DATA

The daily streamflow sequences were analyzed for five gaging stations in Indiana. These are the same stations as the ones which were analyzed by Cervantes et al. (1982) for the flood characteristics. The analyzed stations with their identification numbers, names, drainage areas, and record lengths are given in Table 1. The locations of these stations are shown in Figure 1.

<table>
<thead>
<tr>
<th>Watershed Identification Number</th>
<th>Station Name</th>
<th>Drainage Area (sq. mi.)</th>
<th>Record Length (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>32750</td>
<td>White River near Alpine, IN</td>
<td>529</td>
<td>52</td>
</tr>
<tr>
<td>33030</td>
<td>Blue River near White Cloude, IN</td>
<td>461</td>
<td>50</td>
</tr>
<tr>
<td>33243</td>
<td>Salamonia River at Dora, IN</td>
<td>557</td>
<td>54</td>
</tr>
<tr>
<td>33265</td>
<td>Mississinewa River at Marion, IN</td>
<td>682</td>
<td>57</td>
</tr>
<tr>
<td>33655</td>
<td>East Fork, White River at Seymour, IN</td>
<td>2341</td>
<td>53</td>
</tr>
</tbody>
</table>
Figure 1. Watersheds used for analysis.
II. THE STATISTICAL ANALYSIS OF THE DAILY STREAMFLOW HYDROGRAPH

II.1 MAIN FEATURES OF THE STATISTICAL ANALYSIS

The daily streamflow process will be described by the random variable $X(t)$ which denotes the volume of streamflow which passes a cross-section in the day $t$. Since there is an annual periodicity in the daily streamflow process, the observed daily streamflow data may be thought as a collection of sample records from a periodic stochastic process $[X(t); 0 < t < T]$ with an annual period (i.e. $T=365$ or 366 days). Due to the annual periodicity

$$P[X(t) < x] = P[X(t + T) < x], x > 0, 1 < t < T,$$  \hspace{1cm} (1)

and $T=365$ or 366 days depending on whether the year is a normal or a leap year. Consequently, all the moments and the marginal probability distributions of the daily streamflow process will be annually periodic. Due to this periodicity each of the statistical functions of the daily flow process within the year will depend both on the origin time from which the statistical function starts and on the time distance from this time origin. Therefore, the statistical analysis of the statistical functions of the daily flow process shall be a periodic statistical analysis with the period of one year. There may be time trends superimposed on the annual periodicity. However, these trends are neglected due to the shortness of the data which is of the order of 50 years. The periodic statistical analysis enables one
to construct a sample for each nonoverlapping time interval within the year by grouping the data of that interval from different years. Consequently, under the annual periodicity assumption, it is possible to make statistical inferences about the behavior of the daily streamflow process within a year.

The main features of the daily streamflow hydrograph are a) the peak discharge, b) the recession limb, c) the time to the peak, and d) the overall features such as the first and the second moments of the process. The first and the second moments of the process are analyzed respectively in terms of the time-dependent mean function and of the time origin-time lag dependent covariance function of the observed daily flows. The recession limbs are statistically analyzed in terms of an estimator which depends both on the occurrence time of the peak discharge from which the recession starts and of the time distance from the peak occurrence time since it is believed that the hydrograph recession is dependent on the storage state of the watershed. The peak discharges were already extensively analyzed and modeled by Cervantes et al. (1982) in terms of their flood characteristics. Here, in order to have an insight into the drought characteristics of the daily streamflow process the interarrival time between the consecutive peaks, as a function of the peak occurrence time-discharge exceedence level, is analyzed. The time to the peak is analyzed by its probability distribution function within four 90-day nonoverlapping intervals through the year.
II.2 THE MEAN FUNCTION OF THE DAILY STREAMFLOW PROCESS

Due to the annual periodicity in the daily streamflow process, the behavior of the first moment of the process is investigated by means of the mean function $M(t) = E(X(t))$ which varies with day $t$ within the year. Let the observed daily streamflow volume of day $t$ of the $i$-th year be denoted by $X_i(t)$ and let there be $Y$ years of observation. Under the annual periodicity assumption, since the flow values at day $t$ in different years are coming from the same population, the statistical estimator of the mean function $M(t)$ may be defined as

$$\bar{M}(t) = \frac{1}{Y} \sum_{i=1}^{Y} X_i(t). \quad (2)$$

When the expectation of both sides of (2) is taken

$$E[\bar{M}(t)] = \frac{1}{Y} \sum_{i=1}^{Y} E[X_i(t)] = M(t), \quad (3)$$

is obtained. Therefore, $\bar{M}(t)$ is an unbiased estimator of $M(t)$ for all $t$. When the variances of both sides of (2) are taken,

$$\text{Var}(\bar{M}(t)) = \frac{\text{Var}(\sum_{i=1}^{Y} X_i(t))}{Y^2}$$

$$= \left[ \frac{1}{Y} \text{Var}(X_1(t)) \right]$$

$$+ \frac{1}{Y} \sum_{k=1}^{Y-1} \text{Cov}(X_1(t), X_{1+k}(t)) / Y^2 \quad (4)$$
is obtained. Since all \( X_i(t) \) come from the population of the day \( t \), \([X_i(t), i=1,2,...]\) may be considered as a stationary sequence. Furthermore, it may be assumed that

\[
\lim_{k \to \infty} \text{Cov}[X_i(t), X_{i+k}(t)] = 0 \tag{5}
\]

or

\[
\sum_{k=1}^{\infty} \text{Cov}[X_i(t), X_{i+k}(t)] < \infty
\]

since the correlation among the streamflow volumes at day \( t \) in different years may be assumed to decrease as the difference between the years increases. Then

\[
\lim_{Y \to \infty} \frac{1}{Y} \sum_{i=1}^{Y} k \cdot \text{Cov}[X_i(t), X_{i+k}(t)] = 0 \tag{6}
\]

is obtained. Since the variance of a stationary sequence is finite, combining with (4),

\[
\lim_{Y \to \infty} \text{Var}(\overline{M}(t)) = 0 \tag{7}
\]

is obtained. It follows from (3) and (7) that \( \overline{M}(t) \) is a consistent estimator of \( M(t) \). Thus

\[
\lim_{Y \to \infty} \text{P}[|\overline{M}(t) - M(t)| > \varepsilon] = 0 \quad \text{for all } \varepsilon > 0. \tag{8}
\]
As can be seen from (2) the estimation of the mean function is not affected by the time origin of the estimation. In the mean function estimation the time origin was taken at March 15 to be consistent with the previous statistical work on the flood peaks data of the same stations that are investigated in this study (see Cervantes et al. 1982). The estimated mean functions for the daily streamflow time series at the stations 32750, 33030, 33245, 33265 and 33655 are seen in Figure 2. In this figure the annual periodicity in the first moment of the daily streamflow process is apparent. There is, however, considerable variation in the estimated daily flow means although the streamflow records are sufficiently long (on the order of 50 years).

II.3 THE COVARIANCE FUNCTION OF THE DAILY STREAMFLOW PROCESS

The theoretical covariance between the daily streamflow volumes at a day \( t \) and some other day \( t+k \) within a year is defined as

\[
C(t;k) = E[(X(t) - E(X(t)))[X(t+k) - E(X(t+k))]. \tag{9}
\]

As seen from (9) the covariance function \( C(t;k) \) is a function of both the time origin \( t \) and the time lag \( k \) from the origin \( t \). Thus it describes the second-moment behavior of a time-nonhomogeneous stochastic process. Due to the annual periodicity in the daily streamflow process, \( C(t;k) \) is expected to be a periodic function with annual periodicity both in the time origin.
Figure 2. Mean function of the daily streamflows.
and in the lag $k$. Consequently, an estimator $\bar{C}(t,k)$ of $C(t;k)$ may be defined as

$$
\bar{C}(t;k) = \frac{1}{Y} \sum_{i=1}^{Y} [X_i(t) - \bar{M}(t)] [X_i(t+k) - \bar{M}(t+k)]
$$

(10)

where $1 \leq t \leq 364$, and $k=1,2,...(365-t)$. In (10) $Y$ denotes the years of observation, $X_i(t)$ denotes the observed streamflow volume at day $t$ of the $i$-th year and $\bar{M}(t)$ is defined by (2). When the expectations at both sides of (10) are taken, the equality,

$$
E[C(t;k)] = \frac{1}{Y} E\left( \sum_{i=1}^{Y} [X_i(t) - \bar{M}(t)] [X_i(t+k) - \bar{M}(t+k)] \right).
$$

(11)

is obtained. After introducing the population means $M(t)$ and $M(t+k)$ into (11), and after some simple statistical operations, the equality

$$
E[C(t;k)] = C(t;k) - E([\bar{M}(t) - M(t)] [\bar{M}(t+k) - M(t+k)]).
$$

(12)

is obtained. Combining expression (8) with (12) and taking the limit over the number of observation years,
\[ \lim_{y \to \infty} E[\bar{C}(t;k)] = C(t;k) \quad (13) \]

is obtained. Therefore, as the number of observation years increases, it follows from (13) that \( \bar{C}(t;k) \) asymptotically becomes an unbiased estimator for \( C(t;k) \). Since the estimator \( \bar{C}(t;k) \) is based on the average of the single day products the covariance estimates are quite unstable. In order to obtain more stable covariance estimates, the year is divided into nonoverlapping 15-day intervals and the daily streamflow process is assumed to be locally stationary within each one of these 15-day intervals. Under this assumption the covariance function within a certain 15-day interval becomes independent of the origin times which fall into that interval, and, thus may be thought to depend on the whole interval as its time origin. Consequently, the covariance function estimator for each one of these 15-day intervals may be simply obtained by averaging \( \bar{C}(t;k) \) over the time origins \( t \) falling into that interval. That is, the 15-day smoothed covariance estimator \( \bar{C}(L;k) \) is defined as

\[ \bar{C}(L;k) = \frac{1}{15} \sum_{t=15(L-1)+1}^{15L} \bar{C}(t;k) \quad (14) \]

for \( L=1,2, \ldots [(365 \text{ or } 366)/15] \). Thus, beginning with the March 15-March 29 period, \( L \) denotes the position of the 15-day interval within a year. Taking the expectation of both sides in (14),
\[ E[\bar{C}(L;k)] = \frac{1}{15} \sum_{t} E[\bar{C}(t;k)] \] (15)

is obtained. Using (13), one obtains the equality

\[ \lim_{Y \to \infty} E[\bar{C}(L;k)] = \frac{1}{15} \sum_{t} C(t;k) \] (16)

From the assumption of local stationarity within each 15-day interval of the year, the equality

\[ C(t_1;k) = C(t_2;k), \text{ for } 15(L-1)+1 < t_1 \neq t_2 < 15L \] (17)

is obtained. Combining (16) and (17), one finally obtains

\[ \lim_{Y \to \infty} E[\bar{C}(L;k)] = C(t;k), \text{ for } 15(L-1)+1 < t < 15L. \] (18)

Therefore, as the number of observation years increases, \( \bar{C}(L;k) \) becomes an asymptotically unbiased estimator of \( C(t;k) \) within the time interval \([15(L-1) + 1, 15L]\) for \( L=1,2,\ldots,[(365 \text{ or } 366)/15] \).

Using the annual periodicity assumption for the daily streamflow process, the covariance function of the process was estimated by the 15-day smoothed covariance estimator \( \bar{C}(L;k) \) for the 15-day interval origins \( L=1,2,\ldots,24 \) and for the time lags \( k=1,2,\ldots,40 \). Consequently, a two-dimensional covariance estimation was obtained for the annually periodic covariance function.
of the daily streamflow process. The first origin-interval for estimation was taken as the March 15-March 29 period to be consistent with the previous work on the same data (see Cervantes et al., 1982). The estimated covariance functions for the daily streamflow time series at the stations 32750, 33030, 33245, 33265 and 33655 are respectively shown in Figures 3, 4, 5, 6 and 7. As may be seen from these figures, the covariance function of the daily streamflows has a very strong annual periodicity with respect to its time origin. Although it seems from these figures that the covariance function decays with respect to the time lag, this is deceiving since for a given 15-day interval time origin the covariance function was estimated up to a time lag of 40 days rather than covering the whole annual span of 365-366 days. Since the daily streamflow process is annually periodic in the time-dimension it is expected to have an annual periodicity not only in its time origin but also in its time lag. The whole annual span was not covered with respect to the time lag since a complete picture of the covariance function in both the time origin and the time lag directions would require the computation of approximately 365x365 covariance functions. This requires too large of a computer storage memory. Consequently, the covariance function was estimated up to a lag of 40 days in the time lag direction. When Figures 3, 4, 5, 6, and 7 are compared with Figure 2 for the mean function it is seen that the periodic behavior of the covariance function and the mean function are quite similar. This is an expected result since all the moments of a periodic stochastic process will be periodic with the same
period. From Figures 3, 4, 5, 6, and 7 it is seen that there is a significant variation among the estimated covariance functions in the time-origin direction. Therefore, a comprehensive model for the daily streamflow process needs to be a time-nonhomogeneous one within the period of a year. For a modeling approach which would use seasonally stationary stochastic models for the daily flows, from Figures 3, 4, 5, 6, and 7 it may be seen that the annually periodic daily streamflow process can be roughly separated into 3 seasons whose time boundaries vary for different gaging stations. This variation is to be expected since these stations are on locations with different soil properties. As the soil properties change the groundwater conditions also change and thus yield different persistence properties for the streamflow process at different locations. The approximate seasonal separation of the daily streamflow process in terms of the estimated covariance functions is given in Table 2.
Figure 3. Covariance functions of the daily streamflows at station 32750.
Figure 4. Covariance functions of the daily streamflows at station 33030.
Figure 5. Covariance functions of the daily streamflows at station 33245.
Figure 6. Covariance functions of the daily streamflows at station 33265.
Figure 7. Covariance function of the daily streamflows at station 33655.
TABLE 2. Approximate Time Boundaries of the Seasons Determined from the Estimated Covariance Functions.

<table>
<thead>
<tr>
<th>Station</th>
<th>Streamflow Seasons</th>
</tr>
</thead>
<tbody>
<tr>
<td>32750</td>
<td>May 29-November 9</td>
</tr>
<tr>
<td></td>
<td>November 10-December 24</td>
</tr>
<tr>
<td></td>
<td>December 25-May 28</td>
</tr>
<tr>
<td>33030</td>
<td>June 28-November 9</td>
</tr>
<tr>
<td></td>
<td>November 10-January 8</td>
</tr>
<tr>
<td></td>
<td>January 9-June 27</td>
</tr>
<tr>
<td>33245</td>
<td>June 28-November 9</td>
</tr>
<tr>
<td></td>
<td>November 10-December 24</td>
</tr>
<tr>
<td></td>
<td>December 25-June 27</td>
</tr>
<tr>
<td>33265</td>
<td>June 28-October 25</td>
</tr>
<tr>
<td></td>
<td>October 26-December 9</td>
</tr>
<tr>
<td></td>
<td>December 10-June 27</td>
</tr>
<tr>
<td>33655</td>
<td>May 29-October 10</td>
</tr>
<tr>
<td></td>
<td>October 11-December 9</td>
</tr>
<tr>
<td></td>
<td>December 10-May 28</td>
</tr>
</tbody>
</table>

II.4 THE HYDROGRAPH RECESSION LINES

The runoff from a water basin may be conceptualized as the release from a cascade of linear reservoirs which is fed at the upstream by a precipitation input (Nash, 1957; Kraighenhoff van de Leur, 1966; Keiman, 1980, etc.). In the simplest case of a time-invariant single reservoir the system equations are

\[ x(t) - y(t) = \frac{ds(t)}{dt} \]  

(19a)

for the continuity equation and
\[ y(t) = bs(t) \]  

(19b)

for the dynamic equation. In (19a) and (19b) \( x \) is the instantaneous precipitation rate, \( y \) is the instantaneous runoff rate and \( s \) is the basin storage volume. It is assumed that runoff is proportional to the level of storage \( s(t) \) with a proportionality constant \( b \) which may be called as the basin storage coefficient. Combining (19a) with (19b) yields the differential equation

\[
\frac{dy(t)}{dt} + by(t) = bx(t).
\]  

(20)

The solution of this linear, first-order, constant coefficient, nonhomogeneous differential equation is

\[
y(t) = y(t_0) e^{-b(t-t_0)} + \int_{t_0}^{t} be^{-b(t-\tau)} x(\tau)d\tau.
\]  

(21)

On the recession limb of the hydrograph if it is assumed that the runoff is only due to the release from the basin storage (Kelman, 1980), then on this limb the runoff may be represented as

\[
y(t) = y(t_0) e^{-b(t-t_0)}
\]  

(22)

where \( y(t_0) \) is the peak discharge at time \( t_0 \) from which the hydrograph recession starts. Due to this representation the
basin storage coefficient b may also be called the hydrograph recession decay rate. Various hydrologists (Nash, 1957; Weiss, 1973, Kelman, 1980) have used a constant basin storage coefficient in the deterministic and in the stochastic modeling of the continuous streamflow hydrograph. However, the stochastic models which have used the constant basin storage coefficient have yielded unsatisfactory results in the modeling of the recession limb of the continuous streamflow hydrograph (Weiss, 1973; Kelman, 1980). Weiss (1973) reported that the "recession decay rates are far from constant" and that "positive correlation exists between the decay rates in each interval, and ... the (magnitude of the) rise at the beginning of the interval". In other words, Weiss (1973) has found from his data that the recession decay rates are dependent both on time and on the magnitude of the peak discharge. Kelman's (1980) application has shown that the constant decay rate works well for the fast decreasing recession limbs with a high peak discharge y(t_o). However, from Figure 6 in Kelman (1980) it may be seen that when the peak discharge is low it is not possible to model the recession limb with a constant decay rate. Using these findings, it may be hypothesized that the basin storage coefficient should be a function of both the time and the occurrence time-magnitude of the peak discharge which in turn is a function of the state of the basin storage volume due to relation (19b). Under this generalization the linear reservoir equations become
\[ x(t) - y(t) = \frac{ds(t)}{dt} \] \hspace{1cm} (23a)

and

\[ y(t) = b(t; t_o, y_o) s(t) \] \hspace{1cm} (23b)

where the peak discharge \( y(t_o) \) is denoted as \( y_o \) for notational convenience. Expressions (23a) and (23b) yield the time varying linear differential equation

\[ \frac{dy(t)}{dt} + b(t; t_o, y_o) y(t) = x(t)' b(t; t_o, y_o). \] \hspace{1cm} (24)

The solution of this differential equation is

\[ y(t) = y_o e^{-\int_{t_o}^{t} b(t; t_o, y_o) dt} \] \hspace{1cm} (25)

\[ y(t) = y_o e^{t - \int_{t_o}^{t} b(t; t_o, y_o) dt} + \int_{t_o}^{t} e^{t - \int_{\tau}^{t} b(\xi; t_o, y_o) d\xi} x(\tau) d\tau. \]

If any hydrograph recession limb is defined as a non-increasing portion of the hydrograph, then it is reasonable to assume that on the recession limb the runoff is only due to the release from the time-varying basin storage. Using this definition for the recession limb and combining it with the above assumption and expression (25), one obtains
\[ y(t) = y_o \exp\left[ - \int_{t_0}^{t} b(\tau; t_0, y_o) d\tau \right] \]  \hspace{1cm} (26)

for the representation of the continuous runoff on the recession limb of the hydrograph.

Using the definition that each hydrograph recession limb is a non-increasing portion of the streamflow hydrograph which starts from the occurrence time of a peak discharge and ends at the time when the discharge starts to increase, one can describe the hydrograph recession limb by (26) where the variable \((t_0, y_o)\) may be considered either as deterministic or as a two-dimensional random point in the time-discharge plane (Kavvas, 1982). The basic parameter function of (26) is the linear time-varying basin storage coefficient \(b(\tau; t_0, y_o)\). Since this coefficient is a function of \((t_0, y_o)\) it is a random function in general. However, once \((t_0, y_o)\) is given \(b(\tau; t_0, y_o)\) becomes a deterministic function of the time \(\tau (\tau > t_0)\) and of the given peak discharge occurrence time-magnitude \((t_0, y_o)\). The parameter \(b(\tau; t_0, y_o)\) is computed from the daily streamflow data by representing the integral in (26) by the Riemann-Stieltjes sum (Kolmogorov and Fomin, 1970) as

\[ \int_{t_0}^{t} b(\tau; t_0, y_o) d\tau \approx \sum_{i=1}^{n} b(\varepsilon_i; t_0, y_o) [\tau_i - \tau_{i-1}] \]  \hspace{1cm} (27)

for a certain partition \(\delta \tau: t_0 < \tau_1 < \ldots < \tau_n = t\) of the interval \([t_0, t]\) and where \(\varepsilon_i\) is \(\tau_{i-1} < \varepsilon_i < \tau_i\). Assuming that \(\delta \tau = 1\)
day yields an adequately refined partition of the interval 
\([t_o, t]\) so as to approximate the Riemann integral in (26). Taking
the intermediate value \( \varepsilon_i = \tau_i, \ i=1,2,...,n \) and observing
that \( \delta\tau = \tau_i - \tau_{i-1} \), expression (27) becomes

\[
\int_{t_o}^{t} b(\tau; t_o, y_o) d\tau \approx \sum_{i=1}^{n} b(\tau_i; t_o, y_o) \delta\tau.
\]  

(28)

Combining (28) with (26) and taking \( t = \tau_n \) yields

\[
y(\tau_n) \equiv y_o \exp[-\sum_{i=1}^{n} b(\tau_i; t_o, y_o) \delta\tau],
\]  

(29)

and taking \( t = \tau_{n-1} \) yields

\[
y(\tau_{n-1}) \equiv y_o \exp[-\sum_{i=1}^{n-1} b(\tau_i; t_o, y_o) \delta\tau].
\]  

(30)

Thus \( y(\tau_n) \) may be expressed as

\[
y(\tau_n) \equiv y(\tau_{n-1}) \exp[-b(\tau_n; t_o, y_o) \delta\tau].
\]  

(31)

Therefore, by starting at \( \tau_0 = t_o \) and taking time intervals of
\( \delta\tau = 1 \) day,

\[
y(t_o + J) \equiv y(t_o + J - 1) \exp[-b(t_o, J; t_o, y_o)]
\]  

(32)
J = 1, 2, ..., days,

and \( b(t_o + J; t_o, y_o) \) may be expressed by

\[
b(t_o + J; t_o, y_o) = \ln[y(t_o + J - 1)] - \ln[y(t_o + J)],
\]

\( J = 1, 2, ... \).

For each hydrograph peak occurrence time-discharge magnitude point \((t_o, y_o)\) the basin storage coefficient (hydrograph recession decay rate) \( b(t_o + J; t_o, y_o) \) is estimated as a function of the time distance \( J \) in days from the occurrence time \( t_o \) of the peak discharge using the expression (33). In order to obtain stable estimates, the peak occurrence time-discharge magnitude points are grouped into finite time-discharge rectangles and the recession decay rate functions whose hydrograph peaks fall into a particular time-discharge rectangle are averaged for each time lag to obtain the estimate of the recession decay rate function for that time-discharge rectangle. In other words, if the hydrograph peak points \((t_o, y_o)^k, k=1,...,m\) fall into a particular time-discharge rectangle \( D \), then the hydrograph recession decay rate function (the basin storage coefficient) for that rectangle is estimated by

\[
b(J; D) = \frac{1}{m} \sum_{k=1}^{m} b(t_o^k + J; t_o, y_o^k), \quad J = 1, 2, ...
\]

(34)
where $D$ is a time-discharge rectangle such that $(t_0, \gamma_0, D) \forall k = 1, \ldots, m$ and $J$ is the time lag in days from the occurrence time of a hydrograph peak. In the computation of the decay rate function it is assumed that if there is no observed decay rate after a time distance, the decay rate will stay the same as the last computed decay rate. This assumption is made for the purpose of increasing the sample points of estimation. In accord with the previous report on the flood peaks (Cervantes et al. 1982) the size of each time-discharge rectangle was taken as 15 days in the time direction and one daily streamflow standard deviation $\sigma$ in the discharge magnitude direction and only the peaks above the mean flow, $m$, were considered. Thus starting with the March 15-March 29 period and spanning the whole year, the hydrograph recession decay rate function was estimated up to a time lag of 30 days from the hydrograph for the $m \rightarrow m + \sigma$, $m + \sigma \rightarrow m + 2\sigma$, $m + 2\sigma \rightarrow m + 3\sigma$, and $m + 3\sigma \rightarrow m + 4\sigma$ peak magnitude intervals for the Indiana precipitation gauging stations 32750, 33030, 33245, 33265, and 33655. The estimated decay rates are shown in Figures 8, 9, 10, 11, and 12. From the examination of these figures the following observations are made: a) the recession decay rate functions vary with the origin times and with the discharge magnitudes of the hydrograph peaks; b) the variation of the decay rate with the peak discharge magnitude depends both on the time of the year and on the station location which points to a dependence on the basin storage conditions and on the soil conditions; c) there are inverse relations between the mean function and the recession decay rate function, and
Figure 8. Recession decay rate function at station 32750.
Figure 9. Recession decay rate function at station 33030.
Figure 10. Recession decay function at station 33245.
Figure 11. Recession decay rate function at station 33265.
Figure 12. Recession decay rate function at station 33655.
between the covariance function and the recession decay rate function. When the mean function and the covariance function values are high (low) the decay rate is low (high). This observation points to an important relation between the basin storage volume and the streamflow quantity. It follows that for a given high (low) discharge $y(t)$ and low (high) decay rate $b(.)$, the basin storage $s(t)$ should be high (low) in order to satisfy the relation (23b). In the meantime, since the low decay rates correspond to high covariance values in the flow sequence, the high basin storage $s(t)$ appears to be one of the major causes of high persistence among the flow values. The observations a), b), and c) together indicate that if the basin is modeled as a linear reservoir, this reservoir should definitely have a storage coefficient which is dependent on the initial state of the reservoir (i.e. the peak discharge), the time origin of the release from the reservoir (i.e. the occurrence time of the peak discharge), and on the time-lag from the initial time of release from the reservoir.

II.5 THE PROBABILITY DISTRIBUTION OF THE TIME INTERVAL BETWEEN CONSECUTIVE HYDROGRAPH PEAKS

One of the important parameters in the description of droughts for the irrigation-water supply purposes or for the design of flood-control-purpose reservoirs is the time interval between consecutive hydrograph peaks considered in a time period of planning interest. To be consistent with the statistical terminology this time interval will be called as the interarrival time in the following. Since the flood trigger mechanism charac-
teristics and the basin storage levels change with time through the year it is believed that the probability distribution of the interarrival times of the hydrograph peaks should depend on time. Furthermore, since it is observed from the flood peaks data that the mean number of peaks above a discharge exceedence level (DEL) decreases as the DEL increases (Kavvas, 1980; Cervantes et al., 1982) the mean interarrival time between consecutive flood peaks increases as the DEL increases. Therefore, the probability distribution of the hydrograph peak interarrival time is also dependent on the magnitude of the DEL considered for a particular design alternative. Consequently, a comprehensive probability description of the hydrograph peak interarrival time should be both time and discharge dependent.

In order to obtain such a description the time-discharge dependent behavior of the interarrival time probability distribution needs to be statistically investigated by means of the observed records. For this purpose the hydrograph peaks data for the Indiana stream gaging stations 32750, 33030, 33245, 33265 and 33655 were constructed from the recorded daily streamflow data in these stations. The peaks data were formed as a two-dimensional sequence \([t_i, m_i], i = 1, 2, \ldots\) where \(t_i\) and \(m_i\) respectively denote the occurrence time and the magnitude of the \(i\)-th hydrograph peak in a chronological order.

Taking account of the annual periodicity of the daily streamflow process and assuming that the peaks process is locally
stationary within the nonoverlapping 15-day intervals through the year, in a given 15-day interval the hydrograph peaks which have occurred in this interval at any year \( i \) come from the same population as the peaks which have occurred in this same interval at any other year \( J (J \neq i) \). Therefore, in accord with the previous work on the flood peaks (Cervantes et al., 1982), the year was started with the March 15-29 period and was divided into nonoverlapping 15-day intervals. In a given 15-day interval, for each observed peak which fell into this interval the interarrival time from this peak to the next peak was recorded. Under the local stationarity assumption, all of the interarrival times recorded for the peaks in this 15-day interval for all the years of record were gathered together to form a sample of interarrival times for this interval. Let us denote this sample by \( (x_1, x_2, \ldots, x_n) \) where \( n \) is the number of observed peaks in this 15-day interval and \( x_j \) is the observed interarrival time from the \( J \)-th peak to the next peak within or outside the 15-day interval. Since the data are the daily streamflows, the interarrival times are recorded in days. Consequently, the relative frequency histogram estimator \( f(k), k = 1, 2, \ldots \) days may be defined as

\[
f(k) = \frac{\text{number of } X \text{ which are equal } k \text{ days}}{n}, \; k = 1, 2, \ldots (35)
\]

In order to estimate the sample cumulative distribution function (cdf), first the \( n \) observed interarrival times \( x_i, i = 1, 2, \ldots, n \)
are ordered such that

$$0 < X_{(1)} < X_{(2)} < X_{(3)} \ldots < X_{(n)}.$$  \hspace{1cm} (36)

From this ordered sample the sample cdf estimator $F_n(k)$ is constructed as

$$F_n(k) = \begin{cases} 
0 & \text{if } k < X_{(1)} \\
\frac{i}{n} & \text{if } X_{(i)} \leq k < X_{(i+1)} \\
1 & \text{if } X_{(n)} \leq k 
\end{cases}$$  \hspace{1cm} (37)

It can be shown (Mood et al., 1974) that if $(X_1, \ldots, X_n)$ is a random sample, then $F_n(k)$ is Binomially distributed with the parameters $n$ and $F(k)$ where $F(k)$ is the population cdf. Unfortunately, in the case of the peak interarrival times there is no reason why these times should be independent. Therefore, in general, the sample which is formed from the peak interarrival times is not a random sample and the result of Mood et al., (1974) can not be used to deduce the statistical properties of this cdf estimator.

In order to investigate the behavior of the interarrival time distribution on the discharge dimension the DELs $0$, $m$, $m+\sigma$, and $m+2\sigma$, where $m$ and $\sigma$ respectively denote the mean and the standard deviation of the daily streamflows, were considered. Then the relative frequency histogram $f_n(k)$ and the sample cdf $F_n(k)$ were estimated by (1) and (3) for each DEL in each nonoverlapping 15-day interval of the year by considering only the peaks
in that 15-day interval that are above the given DEL. By this procedure 24x4 distributions were computed to describe the behavior of the probability distribution of the peak interarrival times as a function of the DEL in the discharge dimension. The relative frequency histogram and the sample cdf estimations are respectively shown in Figures 13a, 13b, 14a, 14b, 15a, 15b, 16a, 16b, 17a, and 17b for the respective Indiana stations 32750, 33030, 33245, 33265 and 33655. The main features of the interarrival time probability distribution for each station as a function of the exceedence level and of the origin time (origin interval) is given in Table 3. From this table it may be inferred that the distribution of the hydrograph peak interarrival time depends both on the occurrence time of the peak from which it originates (origin time) and on the discharge exceedence level. As the level gets higher the number of sample data points gets fewer and the sample probability distributions get less pronounced. Also, as expected, the observed maximum interarrival time length increases as the DEL gets higher. At the DELs 0 and m the probability distribution looks predominantly like a gamma-type with very few exponentially distributed exceptions. From the DEL m + σ upwards no apparent probability structures can be inferred due to the scarcity of the sample data points.

In order to investigate the effect of the sampling interval size on the shape of the probability distribution for the peak interarrival time the sampling interval was extended from 15 days to 30 days and the relative frequency histogram f(k) was
Figure 13.a.1. Relative frequency histogram of the interarrival time at station 32750. Exceedence level = 0.
Station 32750

Inter-arrival times considered but not plotted:

Mar 15 to Mar 29  46   38   93   70
Mar 30 to Apr 13  44   43   86   359
Apr 14 to Apr 28  41   42   86   361 121 158 248 250 390
Apr 29 to May 13  47   54  146  149 247 296
May 14 to May 28  43   47  76  121 176 225 261 285
May 29 to Jun 12  43   120 189 192 198 358
Jun 13 to Jun 27  94  108 193 175 192 248
Jun 28 to Jul 12  45   70   82  85 125 138 156 158 185 201 204 225
Jul 13 to Jul 27  106 122 178
Jul 28 to Aug 11  80  119 202
Aug 12 to Aug 26  90  147
Aug 27 to Sep 10  43   70   83  88  98
Sep 11 to Sep 25  47   61   62  63  101
Sep 26 to Oct 10  47   61   60
Oct 11 to Oct 25  60
Oct 26 to Nov 9   41   46   57
Nov 10 to Nov 24  43   51   67
Nov 25 to Dec 9   48   48   66   85
Dec 10 to Dec 24  46   72
Dec 25 to Jan 8   88
Jan 9 to Jan 23  64
Jan 24 to Feb 7   41   48   61
Feb 8 to Feb 22  42  46  46  69

Figure 13.a.2. Relative frequency histogram of the interarrival times at station 32750. Exceedence level = mean.
Figure 13.a.3. Relative frequency histogram of the interarrival times at station 32750. Exceedence level = mean + std. deviation.
Figure 13.a.4. Relative frequency histogram of the interarrival times at station 32750. Exceedence level = mean + 2 std. deviations.
Figure 13.b. Cumulative distribution of the interarrival times at station 32750.
Figure 14a.1. Relative frequency histogram of the interarrival times at station 33030. Exceedence level = 0.
Station 33039

Inter-arrival times considered but not plotted:

<table>
<thead>
<tr>
<th>Period</th>
<th>Times Considered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar 15 to Mar 29</td>
<td>09 28</td>
</tr>
<tr>
<td>Mar 30 to Apr 13</td>
<td>44 47 49 69 86 114 179</td>
</tr>
<tr>
<td>Apr 14 to Apr 29</td>
<td>42 54 61 87 138 176 206</td>
</tr>
<tr>
<td>Apr 29 to May 13</td>
<td>44 47 61 58 80 162 177 198 244</td>
</tr>
<tr>
<td>May 14 to May 28</td>
<td>48 62 65 175 241</td>
</tr>
<tr>
<td>May 29 to Jun 12</td>
<td>66 80 148</td>
</tr>
<tr>
<td>Jun 13 to Jun 27</td>
<td>44 48 49 33 86 172 172 198</td>
</tr>
<tr>
<td>Jun 28 to Jul 12</td>
<td>41 48 60 58 81 86 131 230 237</td>
</tr>
<tr>
<td>Jul 13 to Jul 27</td>
<td>49 65 105 105 119 141 166 224</td>
</tr>
<tr>
<td>Jul 28 to Aug 11</td>
<td>91 108 114 117 153 169 214</td>
</tr>
<tr>
<td>Aug 12 to Aug 26</td>
<td>89 91 92 100 109 162</td>
</tr>
<tr>
<td>Aug 27 to Sep 10</td>
<td>78 92 111</td>
</tr>
<tr>
<td>Sep 11 to Sep 25</td>
<td>97 120</td>
</tr>
<tr>
<td>Sep 26 to Oct 10</td>
<td>42 51 68 85</td>
</tr>
<tr>
<td>Oct 11 to Oct 25</td>
<td>59 99 74</td>
</tr>
<tr>
<td>Oct 26 to Nov 9</td>
<td>60 104</td>
</tr>
<tr>
<td>Nov 10 to Nov 24</td>
<td>51 66 78</td>
</tr>
<tr>
<td>Nov 25 to Dec 9</td>
<td>68</td>
</tr>
<tr>
<td>Dec 10 to Dec 24</td>
<td>68</td>
</tr>
<tr>
<td>Dec 25 to Jan 9</td>
<td>42</td>
</tr>
<tr>
<td>Jan 10 to Jan 23</td>
<td>48 48</td>
</tr>
<tr>
<td>Jan 24 to Feb 7</td>
<td>48 47 51 52</td>
</tr>
<tr>
<td>Feb 8 to Feb 22</td>
<td>43 47</td>
</tr>
<tr>
<td>Feb 23 to Mar 9</td>
<td>45 46</td>
</tr>
</tbody>
</table>

Figure 14.a.2. Relative frequency histogram of the interarrival times at station 33039. Exceedence level = mean.
Station 33030

Inter-arrival times considered but not plotted:

<table>
<thead>
<tr>
<th>Month 1 to Month 2</th>
<th>42 48 53 34 89 300 344 382</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar 15 to Mar 29</td>
<td>44 47 51 182 287 337</td>
</tr>
<tr>
<td>Apr 14 to Apr 28</td>
<td>51 51 51 90 121 178 222 237</td>
</tr>
<tr>
<td>Apr 29 to May 13</td>
<td>233</td>
</tr>
<tr>
<td>May 14 to May 25</td>
<td>48 54 55 84 77 171 200 269</td>
</tr>
<tr>
<td>May 25 to Jun 12</td>
<td>41 52 167 207 206 261</td>
</tr>
<tr>
<td>Jun 13 to Jun 27</td>
<td>95 95 145 190 209 232 234 236 241</td>
</tr>
<tr>
<td>Jun 22 to Jul 12</td>
<td>44 92 240 240 440</td>
</tr>
<tr>
<td>Jul 13 to Jul 27</td>
<td>81 77 90 105 127 170 182 223</td>
</tr>
<tr>
<td>Jul 22 to Aug 11</td>
<td>123 143 177</td>
</tr>
<tr>
<td>Aug 12 to Aug 28</td>
<td>47 117</td>
</tr>
<tr>
<td>Aug 27 to Sep 10</td>
<td>78 116</td>
</tr>
<tr>
<td>Sep 11 to Sep 25</td>
<td>99 107</td>
</tr>
<tr>
<td>Sep 25 to Oct 9</td>
<td>45 56 54</td>
</tr>
<tr>
<td>Oct 11 to Oct 23</td>
<td>83</td>
</tr>
<tr>
<td>Oct 23 to Nov 9</td>
<td>56 56 57</td>
</tr>
<tr>
<td>Nov 10 to Nov 24</td>
<td>56</td>
</tr>
<tr>
<td>Nov 24 to Dec 9</td>
<td>44</td>
</tr>
<tr>
<td>Dec 10 to Dec 24</td>
<td>43 54 57</td>
</tr>
<tr>
<td>Dec 25 to Jan 8</td>
<td>42 45 51 70</td>
</tr>
<tr>
<td>Jan 9 to Jan 23</td>
<td>43 81</td>
</tr>
<tr>
<td>Jan 24 to Feb 7</td>
<td>42 47 57 91 96</td>
</tr>
<tr>
<td>Feb 8 to Feb 22</td>
<td>41 49 52 64 98 123</td>
</tr>
<tr>
<td>Feb 23 to Mar 9</td>
<td>48 122</td>
</tr>
</tbody>
</table>

Figure 14.a.3. Relative frequency histogram of the interarrival times at station 33030. Exceedence level = mean + std. deviation.
Figure 14.a.4. Relative frequency histogram of the interarrival times at station 33030. Exceedence level = mean + 2 std. deviations.
Figure 14b. Cumulative distributions of the interarrival times at station 33030.
Figure 15.a.1. Relative frequency histogram of the interarrival times at station 33245. Exceedence level = 0.
Figure 15.a.2. Relative frequency histogram of the interarrival times at station 33245. Exceedence level = mean.
Figure 15.a.3. Relative frequency histogram of the interarrival times at station 33245. Exceedence = mean + std. deviation.
Figure 15.a.4. Relative frequency histogram of interarrival times at station 33245. Exceedence level = mean + 2 std. deviation.
Figure 15.b. Cumulative distributions of the interarrival times at station 33245.
Figure 16.a.1. Relative frequency distribution of the interarrival times at station 33265. Exceedence level = 0.
Figure 16.a.2. Relative frequency histogram of interarrival times at station 33265. Exceedence level = mean.
Station 33265

Inter-arrival times considered but not plotted:

<table>
<thead>
<tr>
<th>Mar 15</th>
<th>Mar 29</th>
<th>52 53 70 96 114 207 288 374</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar 30</td>
<td>Apr 13</td>
<td>46 49 53 63 67 106 199 120 257 261</td>
</tr>
<tr>
<td>Apr 14</td>
<td>Apr 28</td>
<td>49 44 47 80 149 199 250 285 311 315</td>
</tr>
<tr>
<td>Apr 23</td>
<td>May 13</td>
<td>46 47 47 70 154 197 206 207 294</td>
</tr>
<tr>
<td>May 14</td>
<td>May 22</td>
<td>46 49 120 106 202 211 279</td>
</tr>
<tr>
<td>May 22</td>
<td>Jun 12</td>
<td>45 51 68 56</td>
</tr>
<tr>
<td>Jun 13</td>
<td>Jun 27</td>
<td>57 92 104 104 152 154 184 188 216 217 236</td>
</tr>
<tr>
<td>Jun 26</td>
<td>Jul 12</td>
<td>49 78 109 161 182 194 216 218 226</td>
</tr>
<tr>
<td>Jul 13</td>
<td>Jul 27</td>
<td>48 54 64 111 183 193 196 228 227</td>
</tr>
<tr>
<td>Aug 12</td>
<td>Aug 26</td>
<td>92 102 128 157 147</td>
</tr>
<tr>
<td>Sep 11</td>
<td>Sep 28</td>
<td>86 90 81 98 182</td>
</tr>
<tr>
<td>Sep 28</td>
<td>Oct 10</td>
<td>85 109 181</td>
</tr>
<tr>
<td>Oct 11</td>
<td>Oct 23</td>
<td>44</td>
</tr>
<tr>
<td>Oct 20</td>
<td>Nov 9</td>
<td>56 94</td>
</tr>
<tr>
<td>Nov 10</td>
<td>Nov 24</td>
<td>70 93</td>
</tr>
<tr>
<td>Nov 23</td>
<td>Dec 9</td>
<td>43 57</td>
</tr>
<tr>
<td>Dec 10</td>
<td>Dec 24</td>
<td>48 55 64 30</td>
</tr>
<tr>
<td>Dec 23</td>
<td>Jan 3</td>
<td>44 59 53 92 82</td>
</tr>
<tr>
<td>Jan 9</td>
<td>Jan 23</td>
<td>53</td>
</tr>
<tr>
<td>Jan 24</td>
<td>Feb 7</td>
<td>42 82 95</td>
</tr>
<tr>
<td>Feb 8</td>
<td>Feb 22</td>
<td>43 44 47 301</td>
</tr>
<tr>
<td>Feb 23</td>
<td>Mar 9</td>
<td>43 60 55 333 358</td>
</tr>
</tbody>
</table>

Figure 16.a.3. Relative frequency histogram of the interarrival times at station 33265. Exceedence level = mean + std. deviation.
Figure 16.a.4. Relative frequency histogram of the interarrival times at station 33265. Exceedence level = mean + 2 std. deviations.
Figure 16.b. Cumulative distributions of the interarrival times at station 33265.
Figure 17.a.1. Relative frequency histogram of the interarrival times at station 33655. Exceedence level = 0.
Figure 17.a.2. Relative frequency histogram of the interarrival times at station 33655. Exceedence level = mean.
Station 33655
Inter-arrival times considered but not plotted:

| Mar 13 to Mar 29 | 41 52 64 58 59 76 114 241 272 403 |
| Mar 30 to Apr 13 | 43 47 49 55 63 185 253 271 357 |
| Apr 14 to Apr 23 | 44 46 207 220 231 233 251 270 307 683 |
| Apr 24 to May 13 | 45 49 149 212 231 249 252 261 287 |
| May 14 to May 29 | 62 64 71 210 220 247 247 286 690 |
| May 30 to Jun 12 | 69 250 |
| Jun 13 to Jun 27 | 67 92 103 149 152 294 |
| Jun 28 to Jul 12 | 65 65 90 154 181 211 217 239 250 |
| Jul 13 to Jul 27 | 129 124 171 210 253 |
| Jul 28 to Aug 11 | 122 147 170 |
| Aug 12 to Sep 10 | 94 97 |
| Sep 11 to Sep 25 | 45 |
| Sep 26 to Oct 10 | 48 63 83 |
| Oct 11 to Oct 26 | 97 |
| Oct 27 to Nov 9 | 55 94 |
| Nov 10 to Nov 24 | 41 52 70 94 |
| Nov 25 to Dec 9 | 52 |
| Dec 10 to Dec 24 | 42 64 63 77 85 89 99 |
| Dec 25 to Jan 8 | 41 44 52 |
| Jan 9 to Jan 23 | 105 157 |
| Jan 24 to Feb 7 | 60 83 186 |
| Feb 8 to Feb 22 | 41 52 67 77 132 262 |
| Feb 23 to Mar 9 | 45 46 81 231 323 400 |

Figure 17.a.3. Relative frequency histogram of the interarrival times at station 33655. Exceedence level = mean + 1 std. deviation.
Figure 17.a.4. Relative frequency histogram of the interarrival times at station 33655. Exceedence level = mean $+ 2$ std. deviations.
Figure 17.b. Cumulative distribution of interarrival times at station 33655.
estimated for the station 33245 by (35) for each DEL in each nonoverlapping 30-day interval of the year by considering the peaks above the given DEL. By this procedure $12 \times 4$ distributions were computed for the station 33245 and are shown in Figure 18. Comparing Figure 18 by Figure 15a, one can see that while the relative frequency histograms have more pronounced shapes for the 30-day interval size, the main features of the peak interarrival time probability structure have not changed. Still the probability structure depends both on the origin time and on the DEL, still there is no apparent structure above the $m + \sigma$ DEL and still the dominant structure is a gamma-like distribution.
Figure 18.a.1. Relative frequency histogram of the interarrival times with 30-day sampling at station 33245. Exceedence level = 0.
Figure 18.a.2. Relative frequency histogram of the interarrival times with 30-day sampling at station 33245. Exceedence level = mean.
Figure 18.a.3. Relative frequency histogram of the interarrival times with 30-day sampling at station 33245. Exceedance level = mean + std. deviation.
Figure 18.a.4. Relative frequency histogram of the interarrival times with 30-day sampling at station 33245. Exceedence level = mean + 2 std. deviations.
<table>
<thead>
<tr>
<th>Station</th>
<th>Record Length (years)</th>
<th>Exceedence Level</th>
<th>No. of Sample Data Points</th>
<th>Main Features of the Interarrival Time Probability Distribution at the Given Exceedence Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>32750*</td>
<td>52</td>
<td>0</td>
<td>139 171</td>
<td>The distribution changes from a gamma-like behavior at the beginning one-third of the defined year, to an exponential-like behavior in the middle to back to gamma-like behavior at the last one-third of the year.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>6 102</td>
<td>The beginning one-third and the last one-third of the defined year have gamma-like probability behavior. In the middle of the year no apparent structure due to the scarcity of sample data points.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean + 1 std</td>
<td>2 38</td>
<td>Except in a few intervals at the beginning and at the end of the year, hardly any probability structure can be observed.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean + 2 std</td>
<td>0 26</td>
<td>No apparent probability structure can be observed due to the scarcity of the sample data points.</td>
</tr>
<tr>
<td>33030*</td>
<td>50</td>
<td>0</td>
<td>121 146</td>
<td>Gamma-like probability behavior dominant throughout the defined year. However, the distribution changes with origin intervals through the year.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>6 100</td>
<td>Gamma-like probability behavior at the beginning one-third and last one-third of the defined year. In the middle part of the year no apparent structure due to the scarcity of data points.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean + 1 std</td>
<td>2 45</td>
<td>Although the data is too scarce for a definitive inference, a somewhat gamma-like behavior at the beginning and end of the year in a few intervals. No structure in the rest of the intervals.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean + 2 std</td>
<td>0 26</td>
<td>No apparent structure can be observed due to the scarcity of the sample data points.</td>
</tr>
</tbody>
</table>

*General Features of the Interarrival Time Probability Distribution at the given Station: There is a noticeable change in the probability behavior both as a function of origin-interval within the year, and as a function of exceedence level. In general, the behavior is gamma-like at the zero and mean exceedence levels. A gamma-like structure at the beginning and end of the year at the mean + 1st level is seen. No structure can be observed above the mean + 2 std level due to the scarcity of data.
<table>
<thead>
<tr>
<th>Station</th>
<th>Record Length (years)</th>
<th>Exceedence Level</th>
<th>No. of Sample Data Points</th>
<th>Main Features of the Interarrival Time Probability Distribution at the Given Exceedence Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>33245**</td>
<td>54</td>
<td>0</td>
<td>121 159</td>
<td>March 15 - August 12 period has a gamma-like distribution, transition from gamma to exponential in August 12-26 period, August 27 - September 25 period has exponential distribution, September 26 - March 9 period has a gamma-like distribution.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean 17 94 At the beginning one-third and last one-third of the defined year the probability behavior is gamma-like. In the middle of the year not much structure due to the scarcity of sample data points.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean + 1 std 6 54 Except in a few intervals at the beginning and at the end of the defined year hardly any probability structure can be observed.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean + 2 std 1 35 Except the gamma-like distributed first two intervals no apparent structure can be observed due to the scarcity of the data points. The maximum interarrival time at this level is longer than the ones at lower levels.</td>
</tr>
<tr>
<td>33265**</td>
<td>57</td>
<td>None</td>
<td>134 174</td>
<td>Except the August 12-26 period which has an exponential-like distribution, all the other periods have gamma-like distributions.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean 12 116 At the beginning one-third and last one-third of the defined year the probability behavior is gamma-like. In the middle one-third of the year no apparent structure due to the scarcity of the sample data points.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean + 1 std 2 61 Except in a few intervals at the beginning of the year no apparent structure can be observed.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean + 2 std 2 36 No apparent structure can be observed due to the scarcity of data points.</td>
</tr>
<tr>
<td>33655**</td>
<td>53</td>
<td>0</td>
<td>116 156</td>
<td>Except the August 27 - September 10 and September 26 - October 25 periods which have exponential-like distributions, the probability structure at the rest of the year is gamma-like.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean 8 99 At a few intervals in the beginning and at the last one-third periods of the defined year the distribution looks like gamma. In the rest of the defined year no apparent structure can be observed due to the scarcity of the sample data points.</td>
</tr>
</tbody>
</table>
General Features of the Interarrival Time Probability Distribution at the Given Station: The probability behavior changes both as a function of origin time (origin interval within the year) and as a function of exceedance level. As the exceedence level increases, the maximum observed interarrival time increases, but the probability structures get less pronounced. Whenever observed, the probability structure looks generally like gamma rather than exponential.
II.6 THE PROBABILITY DISTRIBUTION OF THE TIME TO PEAK

One of the most important features of the flood hydrograph is the time that it takes to reach the peak discharge from the start of the flood. The most comprehensive marginal estimator of the time to peak is its probability distribution. Taking account of the seasonality of the daily streamflow process, the year is divided into four 90-day intervals, in each of these intervals the daily flow process is assumed to be locally stationary, and the probability distribution of the time to peak is estimated in these four intervals. These 90-day intervals are March 15–June 12, June 13–September 10, September 11–December 9, and December 10–March 14, where the last interval is extended to March 14 to cover the whole year.

For the estimation of the probability distribution of the times to peak, the hydrograph peak data of the Indiana stream gaging stations 32750, 33030, 33245, 33265, and 33655 were constructed from the daily streamflow data at these stations. Then in each 90-day interval a sample of the times to peaks was formed. Denoting this sample by \((T_1, T_2, \ldots, T_n)\) where \(n\) is the number of observed peaks in a 90-day interval, the relative frequency histogram estimator \(f_n(k)\) and the cumulative distribution function estimator \(F_n(k), k = 1, 2, \ldots\) days, are defined by (35), (36) and (37) by replacing \(X\) by \(T\).

The relative frequency histograms and the cumulative distribution functions of the time to peak in each of the stations 32750, 33030, 33245, 33265 and 33655 are shown in figures 19, 20,
21, 22 and 23. As can be seen from these figures, the distribution of the time to peak is approximately exponential. The distribution function changes with the seasons although the variation is not too significant.
Figure 19.a(top). Relative frequency histogram of the times to peak at station 32750.

b(bottom). Cumulative distribution of the times to peak at station 32750.
Figure 20.a. (top). Relative frequency histogram of the times to peak at station 33030.

b. (bottom). Cumulative distribution of the times to peak at station 33030.
b. (bottom) Cumulative distribution of the times to peak at station 33245.
Figure 22.a. (top). Relative frequency histogram of the time to peak at station 33265.

b. (bottom). Cumulative distribution of the time to peak at station 33265.
Figure 23.a. (top). Relative frequency histogram of the time to peak at station 33655.

b. (bottom). Cumulative distribution of the time to peak at station 33655.
II.7 DISCUSSION OF THE STATISTICAL ANALYSIS RESULTS, AND CONCLUSIONS

A periodic statistical analysis of the daily streamflow data at five streamgaging stations in Indiana was performed to gain an insight into the stochastic structure which governs the daily streamflow process. This analysis was carried out by the mean function and the covariance function of the daily streamflows, by the recession limb of the daily streamflow hydrograph, by the probability distribution of the hydrograph peak interarrival time, and by the probability distribution of the time to peak.

The estimated mean function of the daily flows in Indiana has annual periodicity and demonstrates the well known fact that the daily streamflow process is annually periodic. The covariance function was estimated to infer the persistence properties of the daily flows. In section II.3 it is observed that the covariance function of the daily flows depends both on the time origin from which it starts and on the time lag. The covariance function is observed to have an annual periodicity with respect to the time origin. From this observation it may be inferred that the persistence properties of the daily flows depend on the storage state of the basin at the time origin. As the time origin changes, the storage state changes and different covariance structures are obtained throughout the year. Since the covariance function is time-origin dependent, the daily streamflow process, in its complete form, is time irreversible since the second moments of the process about a time origin in the opposite time directions are not equivalent.
The analysis of the daily streamflow hydrograph recession limbs for the five stations in Indiana showed that the recessions are dependent on the location of the gaging station, on the occurrence time-magnitude of the peak from which the recession starts, and on the time distance from the occurrence time of the peak. These observations together indicate that if the basin is modeled as a linear reservoir, the release should depend on the initial state of the reservoir, on the time from which the release starts and on the time lag from the time when the release starts. In earlier works in hydrology the watershed was modeled in terms of the time-invariant, single or series linear reservoirs to obtain expressions for the continuous runoff (Nash, 1957; O'Donnell, 1966). From the statistical analysis of the daily streamflow recessions it is seen that the time-invariant reservoir assumption is not valid for the daily streamflow series and the time-varying reservoirs need to be considered for a proper modeling of the daily streamflow hydrograph.

The analysis of the interarrival time between consecutive hydrograph peaks was carried out by the time-DEL dependent relative frequency histogram and the sample cdf of the observed peak interarrival times at the five gaging stations in Indiana. The estimated time-DEL dependent relative frequency histograms showed a) that the probability structure of the interarrival time between the consecutive hydrograph peaks depends on the origin time from which the interarrival time starts, b) that the probability structure is dependent on the magnitude of the DEL, and c)
that at low DELs (0 and mean) the probability structure predominantly looks like gamma rather than the popularly assumed exponential distribution. These results demonstrate that the efforts which defined the probability distribution of the interarrival time between consecutive peaks by taking a fixed DEL (Todorovic and Zelenhasic, 1970; Todorovic and Rousselle, 1971; etc.) have oversimplified the real situation since the probability distribution is highly dependent on a given DEL. A very important result is the time-origin dependence of the interarrival time probability structure. This result shows that the stationary descriptions of the interarrival times are approximate and, in certain time intervals of the year, may be erroneous since the probability distribution changes through the year. A first approximation in the direction of accounting for the time variability would be the separation of the year into time intervals within which the probability structure remains relatively homogeneous. However, for a complete description of the time-dependence one needs to derive the time-dependent probability distribution of the interarrival time among the consecutive peaks. Ultimately, a time-DEL dependent description of the probability structure of the peak interarrival time would enable the hydrologists to better describe the stochastic structures of the flood and the drought realizations.

The analysis of the probability distribution of the time to peak in Indiana showed that the distribution is nearly exponential and there is not much variation of the distribution through
the year.
III. A THEORETICAL PROBABILITY DISTRIBUTION OF THE INTERARRIVAL TIMES BETWEEN HYDROGRAPH PEAKS (TAKING INTO ACCOUNT THE TIME-DISCHARGE DEPENDENCE) AND APPLICATION TO DROUGHT DESCRIPTION

A drought period at a given geographical location may be defined as a time interval during which the water supply consistently falls short of the water demand at that location. If it is assumed that the water supply of this geographical location is solely obtained from a river, the drought period for the streamflow process in this river may be defined as the time interval during which the discharge is consistently below the water demand. As seen from Figure 24, in a possible streamflow realization at the river cross-section of interest if the water demand is \( d \) ft\(^3\)/sec, then the drought period becomes \((Z_1, Z_2)\). Given the water demand \( d \), the random length \((Z_2 - Z_1)\) of this drought period is determined by the time locations of the hydrograph peaks immediately before and immediately after the drought period, by the nature of flow recession from the first peak, and by the nature of the rising limb to the second peak. Given the occurrence time of the first peak, the interarrival time between the first and the second peak determines the time location of the second peak. Thus the interarrival time between the first and the second hydrograph peaks, conditioned on the time location of the first peak, is one of the most important parameters which describe a drought realization.

In this section a theoretical expression for the probability distribution of the conditional interarrival time is developed. This distribution depends both on the time location of the first
Figure 24. Schematic of drought period.
hydrograph peak and on the variable water demand or equivalently the DEL (discharge exceedence level) $\text{ft}^3/\text{sec}$. The reason for the development of a DEL dependent probability distribution is that in the planning phase of a water supply project there may be various alternative water demands which may be plausible. For each demand alternative the drought realization will be different. This statement may be easily verified by lowering the variable DEL in Figure 19 to obtain a different critical drought realization. Obviously, the probability characteristics of the drought realization corresponding to a low DEL will be different from the one which corresponds to a high DEL. Furthermore, in Section II.5 the statistical analysis of the interarrival time probability distribution showed that different distributions are obtained at different exceedence levels. Consequently, both from the engineering and from the statistical points of view a comprehensive probability description of the hydrograph peak interarrival time should be DEL (or water demand) dependent. The statistical analysis results in Section II.5 also showed that the probability distribution of the interarrival time heavily depends on the time location of the hydrograph peak from which the interarrival time starts. Consequently, the theoretical probability distribution of the interarrival time is also a function of the time location of the hydrograph peak from which the interarrival time starts. Thus the theoretical probability distribution of the hydrograph peak interarrival time is developed as time-DEL dependent. This time-DEL dependent distribution is derived from the stochastic trigger model for the hydrograph
peaks which was developed by Kavvas (1980, 1982) and later extended by Cervantes et al. (1982) so as to explicitly include the properties of the rainfall process. The stochastic trigger model seems quite suitable for this purpose since a) it describes the time-DEL dependent behavior of the hydrograph peaks and b) since it explicitly describes the simultaneous behavior of the precipitation and the hydrograph peak processes. The second property is especially important for a comprehensive drought description since historically the droughts were either treated in the precipitation context (Palmer, 1965) or in the streamflow context (Dracup, et al., 1980).

The theoretical distribution is compared to its empirical counterparts which were obtained in Section 11.5 for the station 33245 for several time origins within the year.

III.1 THE THEORETICAL DEVELOPMENT OF THE INTERARRIVAL TIME PROBABILITY DISTRIBUTION

Given that the observation of the streamflows has started at some arbitrary time \( t_0 \) and that the first hydrograph peak above a DEL \( 3 \text{ ft}^3/\text{sec} \) after \( t_0 \) has occurred at time \( T_1 \), the interarrival time \( X_2 \) from \( T_1 \) to the time \( T_2 \) of the next peak above the DEL \( d \text{ ft}^3/\text{sec} \) describes the interarrival time which is conditioned on \( T_1 \) and \( d \) (see Figure 24). Denoting the conditional cumulative distribution function (CDF) of \( X_2 \) given \( T_1 \), \( d \) by \( F_{X_2 | T_1} (x_2; t_1, d) \)

and the conditional CDF of \( T_2 \) given \( T_1 \), \( d \) by \( F_{T_2 | T_1} (t_2; t_1, d) \).
since \( X_2 = T_2 - T_1 \), it follows that

\[
F_{X_2 \mid T_1} (x_2; t_1, d) = F_{T_2 \mid T_1} (t_2 + x_2; t_1, d) = F_{T_2 \mid T_1} (t_2; t_1, d),
\]

\[
t_2 = t_1 + x_2, \quad x_2 > 0.
\]  
(38)

Denoting the corresponding probability density functions (pdf) by the letter \( f \), and assuming that \( T_1, T_2 \) and \( X_2 \) have everywhere differentiable CDF's, it follows that

\[
f_{X_2 \mid T_1} (x_2; t_1, d) = f_{T_2 \mid T_1} (t_2 + x_2; t_1, d) = f_{T_2 \mid T_1} (t_2; t_1, d),
\]

\[
t_2 = t_1 + x_2, \quad x_2 > 0.
\]  
(39)

From (38) and (39) it follows that the conditional distribution of \( X_2 \) can be obtained in terms of the conditional distribution of \( T_2 \) conditioned on \( T_1 \) and \( d \).

The conditional pdf \( f_{T_2 \mid T_1} (t_2; t_1, d) \) may be expressed as

\[
f_{T_2 \mid T_1} (t_2; t_1, d) = f_{T_1, T_2} (t_1, t_2; d) / f_{T_1} (t_1; d),
\]

\[
t_2 > t_1 > t_0,
\]  
(40)

where \( f_{T_1} (t_1; d) > 0 \) for all \( t_1 > t_0 \). Therefore, in order to obtain \( f_{X_2 \mid T_1} (x_2; t_1, d) \) the pdfs \( f_{T_1, T_2} (t_1, t_2; d) \) and \( f_{T_1} (t_1; d) \)
need to be derived.

First the pdf \( f_{T_1}(t_1; d) \) is derived. The bivariate probability generating function (PGF) of the stochastic trigger model for the number of peaks above the DEL \( d \) ft\(^3\)/sec in the time intervals \((t_0, t_1)\) and \((t_1, t_1 + \Delta t_1)\) may be expressed as (Kavvas, 1982),

\[
C_{N_1, N_2}(z_1, z_2) = \exp\left[- \int_{-\infty}^{t_0} \int_{\nu}^{\tau} \lambda(\tau) \, d\tau \right] \\
\)

\[
\left(1 - \exp\left[ \sum_{i=1}^{2} (z_i - 1) I(D_i) \right] \right) d\tau \\
\)

\[
\left(1 - \exp\left( z_1 - 1 \right) I(D_1; \tau) + (z_2 - 1) I(D_2) \right) d\tau \\
\)

\[
\left(1 - \exp\left( z_2 - 1 \right) I(D_2; \tau) \right) d\tau \\
\]

\[
(41) \\
\]

where
\[ I(D_1) = \int \int h(x, y; \tau) dy dx, \]

\[ I(D_2) = \int \int h(x, y, \tau) dy dx, \]

\[ I(D_1, \tau) = \int \int h(x, y, \tau) dy dx, \]

\[ I(D_2, \tau) = \int \int h(x, y, \tau) dy dx, \]

(41a)

and where \( \lambda(\tau) \) is the rate of occurrence of the precipitation clusters at the time-volume point \( \tau = (\tau, \nu) \), and \( h(x, y, \tau) \) is the conditional rate of occurrence of the hydrograph peaks which are triggered by the precipitation cluster at \( \tau = (\tau, \nu) \), at the time-discharge point \( (x, y) \). The pdf \( f_{T_1}(t_1) \) is expressed in terms of the bivariate pgf \( G_{N_1, N_2}(z_1, z_2) \) of the stochastic trigger model as

\[ f_{T_1}(t_1) = \lim_{\Delta t_1 \to 0} \frac{\partial^2 G_{N_1, N_2}(z_1, z_2)}{\partial z_1 \partial z_2} \bigg|_{z_1 = z_2 = 0} \Delta t_1. \]  

(42)

Taking the derivatives and the limit on \( G_{N_1, N_2}(z_1, z_2) \), and using
(42) one obtains the expression

\[
\int_{t_1}^{t_0} \int_{v} \lambda(\tau) H(t_1; \tau)e^{-\Lambda(t_0, t_1; \tau)} d\tau + \\
\int_{v} \lambda(\tau) H(t_1; \tau)e^{-\Lambda(t_1, t; \tau)} d\tau |
\]

\[
x \exp(-\int_{v} \lambda(\tau) \cdot (1-e^{-\Lambda(t_0, t_1; \tau)}) d\tau)
\]

\[
-\int_{v} \lambda(\tau) \cdot (1-e^{-\Lambda(t_1, t; \tau)}) d\tau),
\]

(43)

where \( t_1 > t_0 \),

\[
H(t_1; \tau) = \int_{d} h(t_1, y, \tau) dy,
\]

\[
\Lambda(t_0, t_1; \tau) = I(D_1),
\]

\[
\Lambda(t_1, \tau; \tau) = I(D_1; \tau).
\]

(43a)

Defining
\[ \Lambda(t_1; \tau) = \Lambda(t_0, t_1; \tau) \quad \text{for} \quad \tau < t_0 \]
\[ = \Lambda(t, t_1; \tau) \quad \text{for} \quad t_0 < \tau < t_1, \]

and combining (43) with (44), the pdf \( f_{T_1}(t_1; d) \) is obtained as

\[ f_{T_1}(t_1; d) = \left| \int \int \Lambda(\tau)^* H(t_1, \tau) e^{-\Lambda(t_1; \tau)} d\tau \right|. \]

\[ t_1 > t_0. \]  

Next, the derivation of \( f_{T_1, T_2}(t_1, t_2; d) \) is considered. The 4-variate PGF of the number of peaks above the DEL \( d \) ft \(^3\)/sec in the time intervals \((t_0, t_1), (t_1, t_1 + \Delta t_1), (t_1 + \Delta t_1, t_2), \) and \((t_2, t_2 + \Delta t_2)\) may be expressed as (Kavvas, 1982),

\[ G_{N_1, N_2, N_3, N_4}(z_1, z_2, z_3, z_4) = \exp\left\{ - \int \int \Lambda(\tau)^* (1 - \exp[\Sigma (z_i - 1)I(D_i)]) d\tau \right\}, \]

\[ t_1 \]
\[ - \int \int \Lambda(\tau)^* (1 - \exp[\Sigma (z_1 - 1)I(D_1) + (z_2 - 1)I(D_2; \tau)]) d\tau \]

\[ t_1 + \Delta t_1 \]
\[ - \int \int \Lambda(\tau)^* (1 - \exp[\Sigma (z_1 - 1)I(D_1) + (z_2 - 1)I(D_2; \tau)]) d\tau \]
\[
\int_{t_1 + \Delta t_1}^{t_2} \int_{\tau}^{t_2 + \Delta t_2} \lambda(\tau) \cdot (1 - \exp[(z_4 - 1)I(D_4) + (z_3 - 1)I(D_3, \tau)]) \, d\tau \, dt \\
\]

where the same definitions in (41a) are used for \( I(D_1), I(D_2), I(D_1, \tau) \) and \( I(D_2, \tau) \), and \( I(D_3), I(D_4), I(D_3, \tau), I(D_4, \tau) \) are defined as

\[
I(D_3) = \int_{t_1 + \Delta t_1}^{t_2} h(x, y, \tau) \, dx \, dy, \\
I(D_4) = \int_{t_2}^{t_2 + \Delta t_2} h(x, y, \tau) \, dx \, dy, \\
I(D_3, \tau) = \int_{\tau}^{t_2} h(x, y, \tau) \, dx \, dy, \\
I(D_4, \tau) = \int_{\tau}^{t_2 + \Delta t_2} h(x, y, \tau) \, dx \, dy. \\
\] (46a)

The bivariate pdf \( f_{T_1, T_2}(t_1, t_2; d) \) is expressed in terms of the 4-variate pgf \( G_{N_1, N_2, N_3, N_4}(z_1, z_2, z_3, z_4) \) of the stochastic trigger
model as

\[ f_{T_1, T_2}(t_1, t_2; d) = \lim_{\max(\Delta t_1, \Delta t_2) \to 0} \partial^2 G_{N_1, N_2, N_3, N_4}(0, 0, 0, 0) / \partial x_2 \partial x_4 / \Delta t_1 \Delta t_2. \]

Using (47), taking the derivatives and the limits on \( G_{N_1, N_2, N_3, N_4} \), and defining

\[ \Lambda(t_2; \tau) = \Lambda(t_0, t_2; \tau) \quad \text{for } \tau < t_0 \]

\[ \Lambda(t, t_2; \tau) \quad \text{for } t_0 < \tau < t_2 \]

(48)

where \( \Lambda(t_0, t_2; \tau) \) and \( \Lambda(t, t_2; \tau) \) are respectively the expected number of flood peaks in \((t_0, t_2)\) and \((\tau, t_2)\) above the DEL d ft³/sec that are due to the FGM at \( \tau \), one obtains the bivariate pdf \( f_{T_1, T_2}(t_1, t_2; d) \) as

\[
\begin{align*}
  f_{T_1, T_2}(t_1, t_2; d) &= \int_{-\infty}^{t_1} \int_v^{t_2} \lambda(\tau) H(t_1; \tau) H(t_2; \tau) \exp[-\Lambda(t_2; \tau)] d\tau \\
  &\quad + \int_{-\infty}^{t_2} \int_v^{t_2} \lambda(\tau) H(t_2; \tau) \exp[-\Lambda(t_2; \tau)] d\tau \\
  &\quad \cdot \left( \int_{-\infty}^{t_1} \lambda(\tau) H(t_1; \tau) \exp[-\Lambda(t_2; \tau)] d\tau \right) 
\end{align*}
\]
Combining (39), (45) and (49) one obtains the pdf of the interarrival time $X_2$, conditioned on the occurrence time of the first peak $T_1$ and the DEL $d ft^3/sec$, as

\[
\begin{align*}
  f_{X_2|T_1}(x_2; t_1, d) &= (\int_{-\infty}^{x_1} \lambda(\tau) H(t_1 + x_2, \tau) e^{-A(t_1 + x_2; \tau)} d\tau) \\
  &+ (\int_{-\infty}^{x_1} \lambda(\tau) H(t_1 + x_2, \tau) e^{-A(t_1 + x_2; \tau)} d\tau)'
\end{align*}
\]

\[
\begin{align*}
  &\cdot \left(\int_{-\infty}^{x_1} \lambda(\tau) H(t_1 + x_2, \tau) e^{-A(t_1 + x_2; \tau)} d\tau\right) \\
  \cdot \exp\left(-\int_{-\infty}^{x_1} \lambda(\tau)' (1 - e^{-A(t_1 + x_2; \tau)}) d\tau\right)
\end{align*}
\]

\[
\begin{align*}
  &\cdot \left(\int_{-\infty}^{x_1} \lambda(\tau) H(t_1 + x_2, \tau) e^{-A(t_1 + x_2; \tau)} d\tau\right) \\
  \cdot \exp\left(-\int_{-\infty}^{x_1} \lambda(\tau)' (1 - e^{-A(t_1 + x_2; \tau)}) d\tau\right)
\end{align*}
\]

\[x_2 > 0.\]
From the definition of the CDF it follows that the CDF of $X_2$, conditioned on $T_1$, is expressed as

$$F_{X_2|T_1}(x^0; t_1, d) = \frac{\int_{0}^{x^0} f_{X_2|T_1}(x_2; t_1, d) dx_2}{\int_{X_2|T_1} dx_2}, \quad x_2 > 0, \quad (51)$$

where $f_{X_2|T_1}(x_2; t_1, d)$ is expressed by (50).

III.2 APPLICATION OF THE THEORETICAL INTERARRIVAL TIME PROBABILITY DISTRIBUTION TO THE HYDROGRAPH PEAKS DATA AND DISCUSSION OF RESULTS

In order to statistically verify the suitability of the derived theoretical conditional CDF of the hydrograph peak interarrival times, as expressed by (50) and (51), it is applied to the hydrograph peaks data at the streamflow gaging station 33245 in Indiana.

The theoretical interarrival CDF has the parameter functions $\lambda(\tau)$ and $h(x, y; \tau)$. As explained earlier, $\lambda(\tau)$ is the rate of occurrence of the precipitation clusters at the time-volume coordinate $\tau=(\tau, v)$ and $h(x, y; \tau)$ is the conditional rate of occurrence of the hydrograph peaks which are triggered by the precipitation cluster at $\tau$, at the time-discharge coordinate $(x, y)$. In order to compute the theoretical CDF one first needs to estimate the parameter functions $\lambda(\tau)$ and $h(x, y; \tau)$. For the hydrograph peaks above the mean daily streamflow discharge exceedence level (DEL) these parameter functions were already estimated by Cervantes et
al. (1982) for the gaging station 33245 at 30-day intervals for \[ \lambda(\tau) \] and at 15-day intervals for \[ h(x, y; \tau) \]. These parameter estimates were used in the theoretical CDF computation. Consequently, the theoretical CDF is computed above the mean, mean + one standard deviation and mean + two standard deviations DELs.

From expressions (50) and (51) it is seen that the CDF of the peak interarrival times is conditioned on the time location of the initial peak, \( t_1 \), and on the DEL, \( d \). In the theoretical CDF computation one also needs to fix a time origin \( t_0 \) from which the observation starts and where the first peak occurs at \( t_1 > t_0 \) (see Figure 24). Once \( t_0, t_1 \) and \( d \) are fixed, one can then compute the theoretical CDF in (50) and (51). However, because the parameter functions \( \lambda(\tau) \) and \( h(x, y; \tau) \), rather than being represented in functional form, were estimated numerically in nonparametric form, the computation of the theoretical CDF as a function of the interarrival time \( x_2 \) had to be carried out numerically. For this purpose the trapezoidal rule of integration was used to numerically compute the various integrations in (50) and (51). The theoretical CDF was computed for 3 arbitrary initial-peak time locations within the year. These initial-peak time locations \( t_1 \) are the time locations of the occurrence of the first hyrdograph peak from which the interarrival time \( x_2 \) is measured. The initial-peak time locations are taken in such a way that they fall into different 15-day time-origin intervals of the year, in order to be consistent with the sample CDFs which were estimated in Section II.5. The selected initial-peak time
locations are March 29, April 28 and January 23. For each initial peak time location \( t_1 \) two time origins \( t_0 \) were considered in order to see the effect of varying \( t_0 \) on the CDF of the peak interarrival time. For \( t_1 = \text{March 29}, \ t_0 = \text{March 14 and March 26}, \) for \( t_1 = \text{January 23}, \ t_0 = \text{January 8 and 20} \) were considered. The computed theoretical CDFs are seen in Figures 25a, 25b, 26a, 26b, and 27a, 27b together with the empirical counterparts.

When Figures 25a through 27b are analyzed, it may be seen that varying \( t_0 \) does not have a significant influence on the shape of the numerically computed CDF with the important exception that as \( t_0 \) gets near to \( t_1 \) the numerically computed CDF values become unstable and may overshoot 100% by about 5%. The major conclusion which may be drawn from a visual inspection of Figures 25a through 27b is that the numerically computed theoretical CDFs do not replicate their empirical counterparts satisfactorily. The goodness-of-fit of the theoretical peak interarrival CDF was statistically tested by the Kolmogorov-Smirnov test and the results are given in Table 4. As can be seen from this table, out of the 18 cases considered only 4 have passed the Kolmogorov-Smirnov test at the 1% significance level, and in 4 cases the test statistic was within 2% of the Kolmogorov-Smirnov 1% significance point. The test results agree with the conclusion drawn from the visual inspection that the numerically computed theoretical CDF of the hydrograph peak interarrival times does not satisfactorily replicate its empirical counterpart computed from the observed data. However, there are some important
Figure 25. Sample and theoretical cumulative distributions of the peak interarrival time for the 15-day interval that ends on April 28 at station 33245, a(top) $t_0 = April 13$, b(bottom) $t_0 = April 23$. 
Figure 26. Sample and theoretical cumulative distributions of the peak interarrival time for the 15-day interval that ends on January 23 at station 33245, a(top) $t_0 = \text{January } 8$, b(bottom) $t_0 = \text{January } 20$. 
Figure 27. Sample and theoretical cumulative distributions of the peak interarrival time for the 15-day interval that ends on March 29 at station 33245, a(top) $t_o = March 14$, b(bottom) $t_o = March 26$. 
reasons for the failure of the derived theoretical CDF in replicating the sample CDF estimated from observations. The most important reason is that in the computation of the theoretical CDF both $\lambda(\tau)$ and $h(x,y;\tau)$ values are constant within each non-overlapping 15-day interval due to the estimation procedure of Cervantes et al. (1982). However, from Figures 25a through 27b it may be seen that most of the increase in the sample CDFs takes place during the first 15 days, indicating that using an average value for the parameter functions within this 15-day interval eliminates the opportunity of catching the important changes in the probability distribution within the 15 days. This is believed to be the main reason for the inadequate fits. The second reason is the errors that are being introduced by the trapezoidal numerical integration of the complicated convolution-type integrals in expression (50). Finally, the theoretical conditional CDF which was derived in this study is only an approximation to the complete conditional CDF which depends on the location times and magnitudes of any n-number of previously observed hydrograph peaks. It is believed that once these points are accounted for it will be possible to obtain an adequate theoretical representation for the probability distribution of the hydrograph peak interarrival times.
<table>
<thead>
<tr>
<th>Exceedence Level</th>
<th>Time-Origin Interval</th>
<th>t_o</th>
<th>t_1</th>
<th>Sample Size</th>
<th>Kolm.-Smir. Statistic From Test</th>
<th>Kolm.-Smir. 1% Significance Point</th>
<th>Kolm.-Smir 5% Significance Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>March 15-29</td>
<td>March 14</td>
<td>March 29</td>
<td>91</td>
<td>.27</td>
<td>.17</td>
<td>.14</td>
</tr>
<tr>
<td>Mean + 1 St. Dev.</td>
<td></td>
<td></td>
<td></td>
<td>54</td>
<td>.24</td>
<td>.22</td>
<td>.19</td>
</tr>
<tr>
<td>Mean + 2 St. Dev.</td>
<td></td>
<td></td>
<td></td>
<td>35</td>
<td>.27</td>
<td>.28</td>
<td>.23</td>
</tr>
<tr>
<td>Mean</td>
<td>March 26</td>
<td>March 29</td>
<td></td>
<td>91</td>
<td>.24</td>
<td>.17</td>
<td>.14</td>
</tr>
<tr>
<td>Mean + 1 St. Dev.</td>
<td></td>
<td></td>
<td></td>
<td>54</td>
<td>.23</td>
<td>.22</td>
<td>.19</td>
</tr>
<tr>
<td>Mean + 2 St. Dev.</td>
<td></td>
<td></td>
<td></td>
<td>35</td>
<td>.27</td>
<td>.28</td>
<td>.23</td>
</tr>
<tr>
<td>Mean</td>
<td>April 14-28</td>
<td>April 13</td>
<td>April 28</td>
<td>72</td>
<td>.17</td>
<td>.19</td>
<td>.16</td>
</tr>
<tr>
<td>Mean + 1 St. Dev.</td>
<td></td>
<td></td>
<td></td>
<td>47</td>
<td>.34</td>
<td>.24</td>
<td>.20</td>
</tr>
<tr>
<td>Mean + 2 St. Dev.</td>
<td></td>
<td></td>
<td></td>
<td>31</td>
<td>.36</td>
<td>.29</td>
<td>.24</td>
</tr>
<tr>
<td>Mean</td>
<td>April 25</td>
<td>April 28</td>
<td></td>
<td>72</td>
<td>.17</td>
<td>.19</td>
<td>.16</td>
</tr>
<tr>
<td>Mean + 1 St. Dev.</td>
<td></td>
<td></td>
<td></td>
<td>47</td>
<td>.34</td>
<td>.24</td>
<td>.20</td>
</tr>
<tr>
<td>Mean + 2 St. Dev.</td>
<td></td>
<td></td>
<td></td>
<td>31</td>
<td>.36</td>
<td>.29</td>
<td>.24</td>
</tr>
<tr>
<td>Mean</td>
<td>January 9-23</td>
<td>January 8</td>
<td>January 23</td>
<td>59</td>
<td>.32</td>
<td>.21</td>
<td>.18</td>
</tr>
<tr>
<td>Mean + 1 St. Dev.</td>
<td></td>
<td></td>
<td></td>
<td>35</td>
<td>.30</td>
<td>.28</td>
<td>.23</td>
</tr>
<tr>
<td>Mean + 2 St. Dev.</td>
<td></td>
<td></td>
<td></td>
<td>23</td>
<td>.38</td>
<td>.34</td>
<td>.28</td>
</tr>
<tr>
<td>Mean</td>
<td>January 20</td>
<td>January 23</td>
<td></td>
<td>59</td>
<td>.31</td>
<td>.21</td>
<td>.18</td>
</tr>
<tr>
<td>Mean + 1 St. Dev.</td>
<td></td>
<td></td>
<td></td>
<td>35</td>
<td>.30</td>
<td>.28</td>
<td>.23</td>
</tr>
<tr>
<td>Mean + 2 St. Dev.</td>
<td></td>
<td></td>
<td></td>
<td>23</td>
<td>.38</td>
<td>.34</td>
<td>.28</td>
</tr>
</tbody>
</table>
IV. SUMMARY AND CONCLUSIONS

In this study a fairly detailed statistical analysis of the daily streamflow hydrograph was performed, and using the findings from this analysis, a time-varying probability distribution for the hydrograph peak interarrival times was developed and tested.

A periodic statistical analysis of the daily streamflow data at five streamgaging stations in Indiana was performed to gain an insight into the stochastic structure which governs the daily streamflow process. This analysis was carried by the mean function and the covariance function of the daily streamflow, by the recession limb of the daily streamflow hydrograph, by the probability distribution of the hydrograph peak interarrival time, and by the probability distribution of the time to peak. For the estimation of these statistical functions new statistical estimators were developed and used in this study. A detailed discussion of the results of this statistical analysis may be found in Section II.6. In gross features this analysis has shown that a) the daily streamflow process is annually periodic in all of its statistical functions and b) if the daily streamflow process is modeled as the release from a linear watershed storage, this release should depend on the state of the watershed storage and on the time of the release since the persistence properties and the recession limb decay rates were observed to change with the state of the basin storage and the time of the year. Therefore, a time-varying linear reservoir system need to be considered for a comprehensive model of the daily streamflow hydrograph.

Using the results of the statistical analysis of the hydrograph peak interarrival times a time-discharge exceedence level dependent theoretical distribution was developed for these interarrival times. This distribution was developed from the stochastic trigger model of the flood peaks which was successfully applied to the flood peaks data in Indiana (Cervantes et. al., 1982). The derived distribution is only an approximate form of the complete distribution which would depend on the entire history of the hydrograph peaks, in that it only depends on the most recent peak from which the interarrival time to the next peak is measured. However, as explained in Section III.2, this approximation has had limited success, mainly due to the averaging in the estimation of the distribution's parameter functions over large time intervals on the order of 15 days. It is believed that by a finer estimation of the parameter functions and by a better approximation of the complete distribution it is possible to replicate the observed probability behavior of the hydrograph peak interarrival times.
V. REFERENCES


