Statistical Analysis of Stream Events

Victor L. Zitta
Associate Professor
Department of Civil Engineering
Bruce A. Mowry
Graduate Research Assistant
Department of Civil Engineering
Mississippi State University

INTRODUCTION

An innovative technique for disposal of lagoon or other treated wastewaters has been under study at Mississippi State University (3,1) utilizing the stochastic nature of stream flows to minimize the detrimental effects of wastewater discharge on stream quality and stream aesthetics. This technique has evolved from continuous wastewater discharge proportional to stream flow magnitude (3), to intermittent discharge during the rising portion of stream events (1). Thus waste discharge is made at optimum stream transport and assimilative capacities while zero waste discharge is made during the recession portion of the discharge time series. To implement such a discharge scheme a detailed study must be made of hydrologic time series to determine the properties of stream flow applicable to this discharge technique.

An annual time series for the Big Brown Creek near Booneville is shown in Figure 1-1. For this stream site and others, the flow is extremely stochastic in nature, continuously changing in magnitude with time. These changes in magnitude are in response to spatial and temporal inputs of rainfall on the drainage basin which produce hydrologic events.

For purposes of this study an event was defined as a maximum on the hydrograph irrespective of magnitude or time of occurrence. It is the intent of this study to define the stochastic nature of stream events as applied to the intermittent discharge of wastewater discharge during these events.

Before applications can be made of the intermittent discharge principle, a thorough understanding of the space - time distribution of stream events must be recognized. This paper is an attempt to set forth certain underlying principles peculiar to stream events. More specifically the objectives of the study are:

1. To determine the monthly distribution of time between stream events (spans) during the calendar year,
2. To determine the mean annual spans between events as related to drainage area,

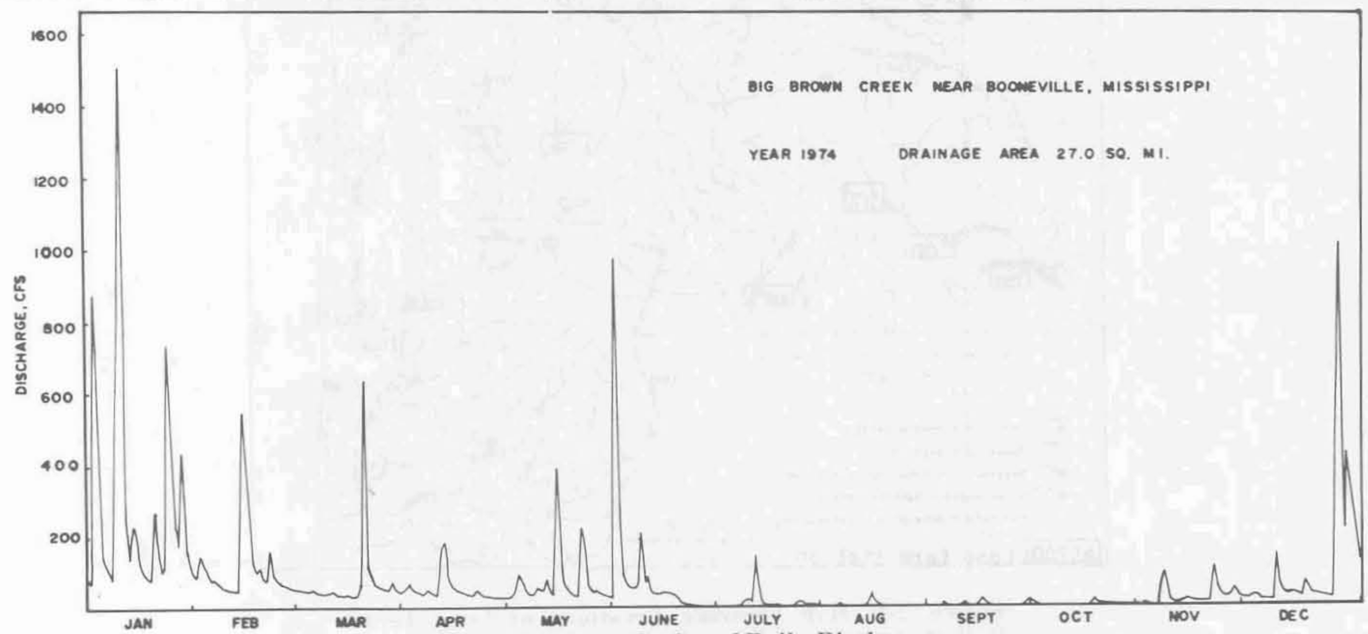

Figure 1-1. Time Series of Daily Discharge.

3. To determine a probability density function applicable to the distribution of maximum annual span between events.

It should be emphasized that the results are preliminary in scope and reflect the authors' investigation to this point in time. Additional work is forthcoming.

METHOD OF ANALYSIS

Data Available

It is well recognized that the longer the period of record the more confidence can be placed in the results of any statistical analysis (2,6,7). In establishing basic stochastic relationships the longest possible period of record is desired provided the data exhibits stationarity, homogeneity and consistency. For applications, a minimum of 30 years of record is desirable with less years acceptable but with reduced confidence in the results. For development of conceptual ideas 50 to 100 year records are desirable, preferably the latter.

Historical time series of average daily flow for 155 stations in Mississippi were obtained in tape form from the files of the U. S. Geological Survey (5). Although the magnitude of the list looks

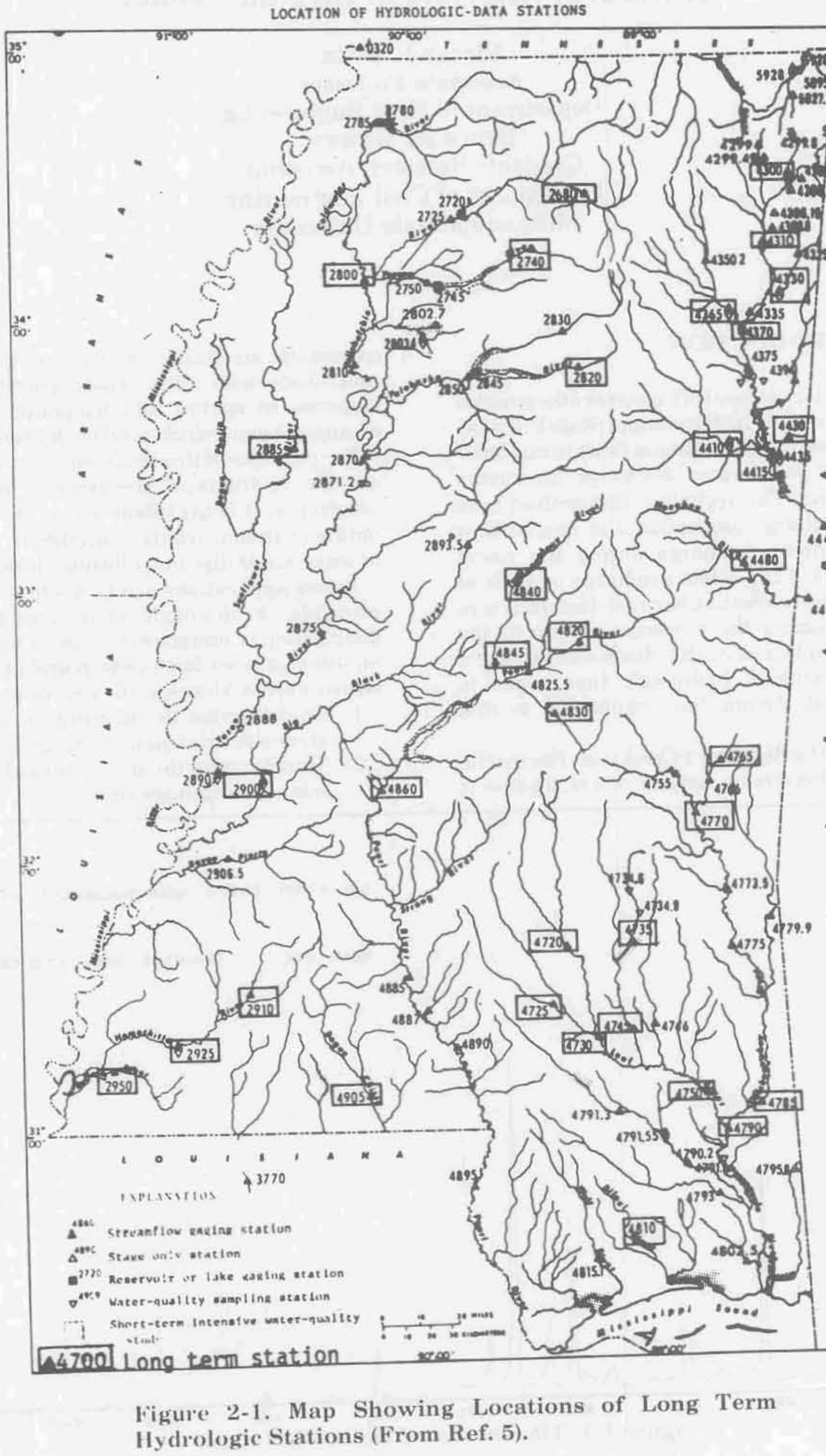

Figure 2-1. Map Showing Locations of Long Term Hydrologic Stations (From Ref. 5).

impressive, a more detailed examination of the stations shown in Figure 2-1, currently in daily operation, indicates wide spatial scattering of the gaging sites. For example, within the Big Black Basin only two continuous stations exist and within the Big Sunflower only one. In addition, analysis of the data at some stations indicates considerable non-stationarity in the time series, some stations show gaps in the time series, some having only monthly data at random times and others show considerable non-homogeneity due to reservoir construction or changes in the hydrologic environment (5). Most of these errors are indicated in the description of the records by the U. S. Geological Survey although the basic data has not, in most cases, been adjusted to eliminate or adjust for such conditions.

Methodological Considerations

Much work has recently been done to predict the magnitude and frequency of flood events as well as the merits of partial versus annual flood series (4). In flood analysis the maximum event in an interval of time, with specified probability is desired. The model in this study is similar but with some differences.

Consider the conceptual streamflow hydrograph shown in Figure 2-2 in which by definition an event is every maximum in the hydrologic time series irrespective of magnitude or time of occurrence. Following a similar notation as Todorovic (4), denote by Q_i the magnitude of any hydrologic event occurring in the time interval [τ]. Let Q_b be a base level below which Q_i will not be recognized as an event. Denote by k the number of events, Q_i, in the interval [τ]. It is obvious that there exists a sequence of random events in the interval [τ] having magnitudes

$$Q_1, Q_2 ---, Q_k \qquad 2\text{-}1$$

This series over n years contains all of the events in the hydrologic time series.

Let T_j be the time (span) between events. Let j be the number of spans in the interval [τ]. Thus there also exists a sequence of random spans in the interval [τ] having magnitudes

$$T_1, T_2, --- T_j \qquad 2\text{-}2$$

This series over n years contains all of the spans between events in a hydrologic series.

Figure 2-2. Conceptual Hydrograph of Stream Discharge.

Denote the maximum span in each interval of time τ as

$$T_S = \sup_{\tau} [T_j] \qquad S = 1, 2, --- n \qquad 2\text{-}3$$

If τ is one year then there exists an annual series of random events T_S for which a histogram can be constructed, a probability density function determined, a probability attached and a return period calculated.

Frequency histograms were developed for various levels of Q_b. Plotting positions were determined from Weibull's relation

$$p = \frac{m}{n + 1} \qquad 2\text{-}4$$

and a continuous probability density function fitted to T_S. By visual inspection of the histograms, the normal and lognormal distributions seemed most appropriate to attempt to fit the data. In each case the Kolmogorov-Smirnov test was preferred over the Chi-Square test (2,6) to determine the goodness of fit.

RESULTS OF DATA ANALYSIS

General

The data on tape furnished by the U. S. Geological Survey for all gaging stations in Mississippi was searched for the longest period of record exhibiting stationarity, homogeneity, and consistency. Of the stations investigated, the Pearl River Station at Edinburg seemed most nearly to meet these basic statistical criteria. This station was extremely complete having daily discharge values from October 1928 to date with little major dam construction or environmental change within the basin during this period. As a check on the results obtained for this station, the results from several stations within the Pearl and Pascagoula Basins with shorter records were used for comparison, but are not included as part of this study.

To obtain an intuitive feel for the span between events during the year, the mean monthly spans and standard deviation for each month of the year was determined and related to drainage area. Following this analysis, frequency distributions were determined for an annual series of maximum spans (T_S) with various levels of Q_b as multiples of $Q_{10,7}$. Multiples of $Q_{10,7}$ were deemed most appropriate since most water quality personnel are familiar with and can readily relate to this statistical parameter. The Kolmogorov-Smirnov (K-S) test of goodness of fit was preferred over the Chi-Square test (2,6). Empirical plotting positions were based on Weibull's plotting position formula.

Mean Monthly Span

In Figure 3-1 is shown the variation of mean monthly span during the year for the Pearl River at Edinburg. The values tend to be cyclic with minimum spans in November and December and again in July and August. Maximum spans occur in October and in February or March.

The mean monthly spans between events is 9.3 days and the mean monthly standard deviation is 5.8 days. During the dry period of October, the monthly spans exceed the mean and during the wet season from January through June the monthly spans exceed or equal the mean.

The standard deviation follows the same general cyclic pattern with high means and standard deviations occurring in the same month. Conversely low means and low standard deviations occur in the same month.

In Figure 3-2 the dramatic effects of $Q_b = 0$, $Q_{10,7}$, 5 $Q_{10,7}$ and 10 $Q_{10,7}$ are depicted. No change in the curve occurs by elimination of events of magnitude equal to or less than $Q_{10,7}$. As events of magnitude less than or equal to 5 $Q_{10,7}$ are eliminated, there is an increase in the span between events, from June through December, the drier months. During the wet months, January through May, no significant increase in the span is noted. This result was expected as the time series shows larger magnitude

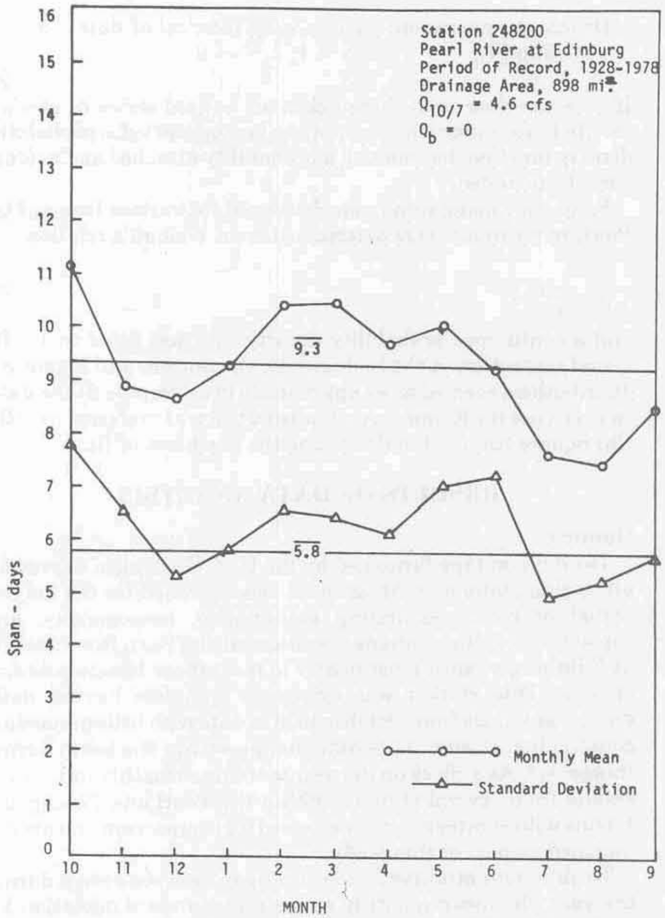

Figure 3-1. Mean Monthly Span Between Events.

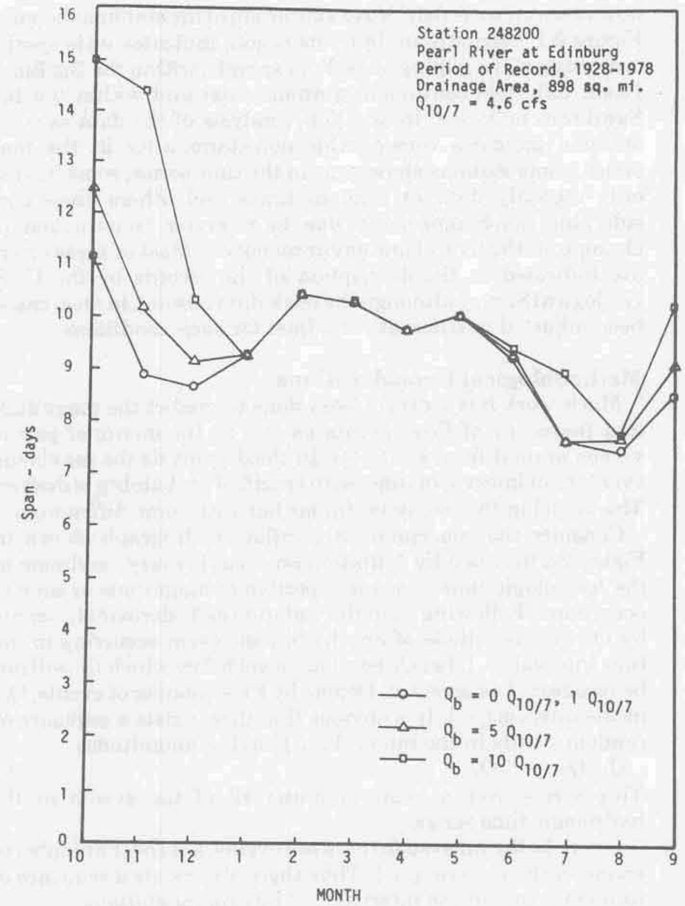

Figure 3-2. Mean Monthly Span Between Events as a Function of Q_b, Days.

events during the wet season and small magnitude events during the dry season.

To verify the cyclic nature of mean monthly spans, data from other gaging stations in Mississippi were compared with the Pearl at Edinburg. In all cases the same cyclic trend was exhibited with increasing spans in the dry season with increases in Q_b from 0 through 10 $Q_{10/7}$.

Drainage Area Considerations

Drainage area seems to have a dramatic effect upon the average monthly span between events. In Figure 3-3 is shown the mean annual span as related to drainage area. This graph clearly indicates that as drainage area increases the mean annual span increases in an apparent linear fashion. This indicates that small drainage areas have more events with shorter spans between events than do large basins which have less events with longer spans between events. It would seem logical that large basins should have shorter spans between events since more rainfall events should occur on larger basins in a given time. This is not the case. Obviously events on tributary streams must be masked by storage effects as they propagate downstream into the larger river channels or they appear as part of larger events downstream.

One outlier point appears in Figure 3-3. This outlier is from data for the Big Sunflower River at Sunflower where the mean annual span for a drainage area of 767 sq. mi is 14 days. The graphical line of best fit is approximately 8.5 days. This indicates that the level terrain of the Mississippi Delta streams must have considerable effect upon the span between events, even to the point of classifying the Delta streams as a separate hydrologic region for this analysis. Unfortunately the station at Sunflower is the only long term station in the Mississippi Delta.

Histograms

Histograms of maximum annual span between events are shown in Figure 3-4 for the hydrologic series of the Pearl River at Edinburg. Three histograms are shown corresponding to different levels of Q_b. For levels of Q_b = 0 and $Q_{10/7}$ the histograms were identical conforming to a normal type distribution. With levels of Q_b = 5 $Q_{10/7}$ and 10 $Q_{10/7}$ there is a noticeable skew in the distributions with the more pronounced skew associated with Q_b = 10 $Q_{10/7}$. Thus the distributions lose their normality as events of small magnitude are eliminated from the series.

In each histogram the mode of the distribution lies between 25 and 30 days. But there is a marked decrease in the magnitude of the mode as small magnitude events are eliminated from the series by Q_b equal to 5 $Q_{10/7}$ and 10 $Q_{10/7}$. This decrease in the magnitude of the mode is associated with increasing magnitude of the spans. It is of interest to note that the maximum span is 48 days when all events, irrespective of magnitude are included, but this value increases to 116 days when Q_b equals 10 $Q_{10/7}$.

Probability Density Functions

Basic to this project was the establishment of a probability density function which would be adequate to describe the histograms of the annual maximum spans irrespective of the value of Q_b. It was decided that since this research was deemed to be a pioneering type effort, the distribution chosen should be as simple in concept as possible. It was also felt that conceptual

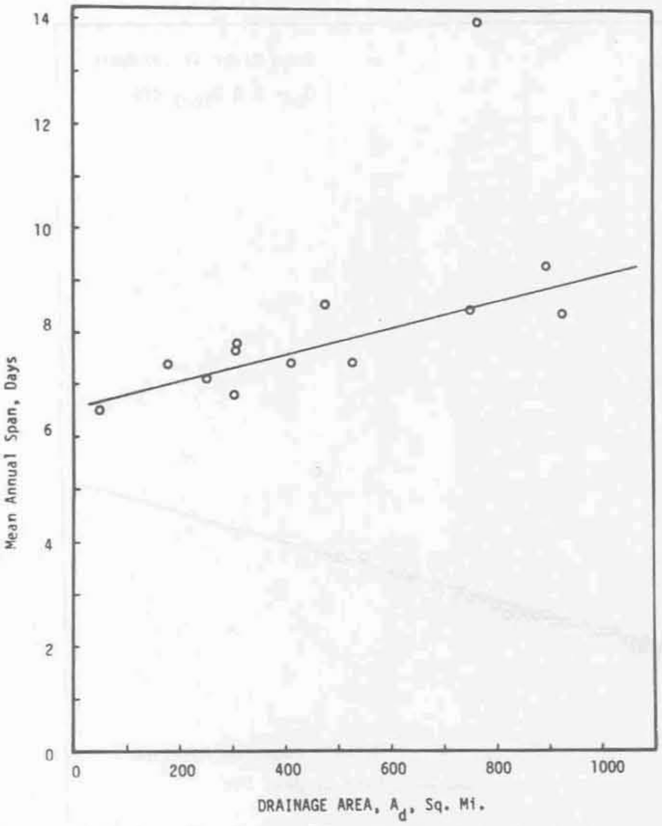

Figure 3-3. Average Annual Spans as Related to Drainage Area.

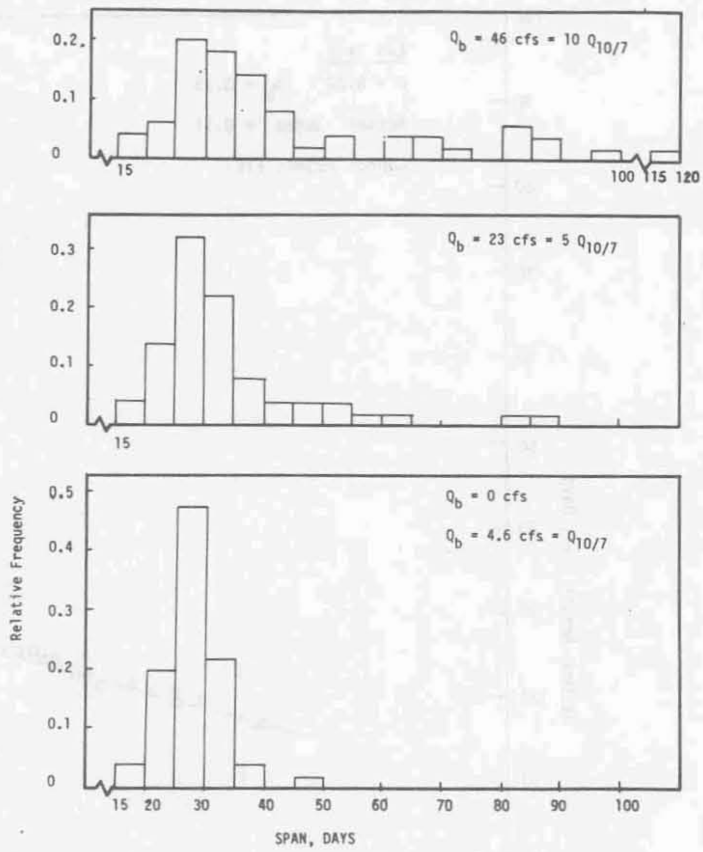

Figure 3-4. Histograms of Maximum Annual Span, T , Days. Pearl River at Edinburg.

simplicity should not be held above the adequacy of the distribution to describe the data.

Maximum annual spans for the Pearl River at Edinburg were ranked from the largest to the smallest with the largest having a rank of one. Weibull's plotting position formula (Eqn. 2-4) was used as the basis to assign values of exceedance to each span in the maximum annual series due to its simplicity and due to its fully satisfying Gumbel's conditions (6). Values of Weibull's plotting positions are illustrated in Figures 3-5 through 3-10 for the Pearl River at Edinburg. Three sets of plots are included, one for maximum annual spans with $Q_b = 0$ and $Q_{10/7}$, one for $Q_b = 5$ $Q_{10/7}$ and one for $Q_b = 10 Q_{10/7}$. Each set contains a plot on normal probability scale and one on lognormal scale.

Inspection of the histograms indicates a normal type distribution for the data in Figure 3-5 but the data seems to plot well on lognormal scale in Figure 3-6 also. Haan (2) indicates that the two parameter lognormal distribution has found wide application in hydrology since much hydrologic data are bounded by zero. This is precisely the case here. No annual maximum span can be less than zero nor can it exceed 365 days. Thus by description and by inspection of Figures 3-5 through 3-10 it was deemed appropriate to test the data against the normal and lognormal distribution for "goodness of fit."

Two tests are available for testing the goodness of fit of probability density functions. One is the Chi-Square test, the other the Kolmorgorov-Smirnov test (6,7,2). It is well recognized that both tests lack the rigor desired but statistical academe has not devised to date a more adequate test than the K-S test. Its inadequacies are well recognized.

In each of Figures 3-5 through 3-10, critical values for the K-S test statistic are given for the most critical significance level provided by Yevjevich (6) and Haan (2). If the difference between any of the plotted values and the line of best fit exceed this critical value the fit is rejected. Lines of best fit for the normal distribution was determined from the well established normal probability relation

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad 3\text{-}1$$

where μ is the population mean and σ is the population standard deviation. Using sample estimates of μ and σ, the normal distribution is plotted in Figures 3-5, 3-7, and 3-9 for levels of Q_b equal to 0 and $Q_{10/7}$, 5 $Q_{10/7}$ and 10 $Q_{10/7}$. Only in the case of $Q_b = 0$ and $Q_{10/7}$ was the normal distribution adequate to describe the distribution of maximum annual spans. The normal distribution was rejected based on the K-S test for the data in Figures 3-7 and 3-9.

The lognormal distribution describes any range of variable values from 0 to $+\infty$ Although the actual data is bounded by 0 it is not bounded by $+\infty$ but this is of no consequence since none of the histograms have spans exceeding 116 days. Even then, the relative frequency of such large spans is quite small. Based on simplicity and availability of tables, the two parameter lognormal distribution was fitted to the data. This function is

$$f(x) = \frac{1}{\sigma_y\sqrt{2\pi}} e^{-(y-\mu_y)^2/2\sigma_y^2} \qquad y = \text{Ln } x \qquad 3\text{-}2$$

In all cases, the lognormal distribution could not be rejected based on the K-S test at a significance level of 0.20. This is indicated in Figures 3-6, 3-8, and 3-10, where μ_y is the mean of the logarithms of x and σ_y is the standard deviation of the logarithms.

Lines of best fit of the lognormal distribution are plotted in Figures 3-6, 3-8 and 3-10. In all three figures the distribution

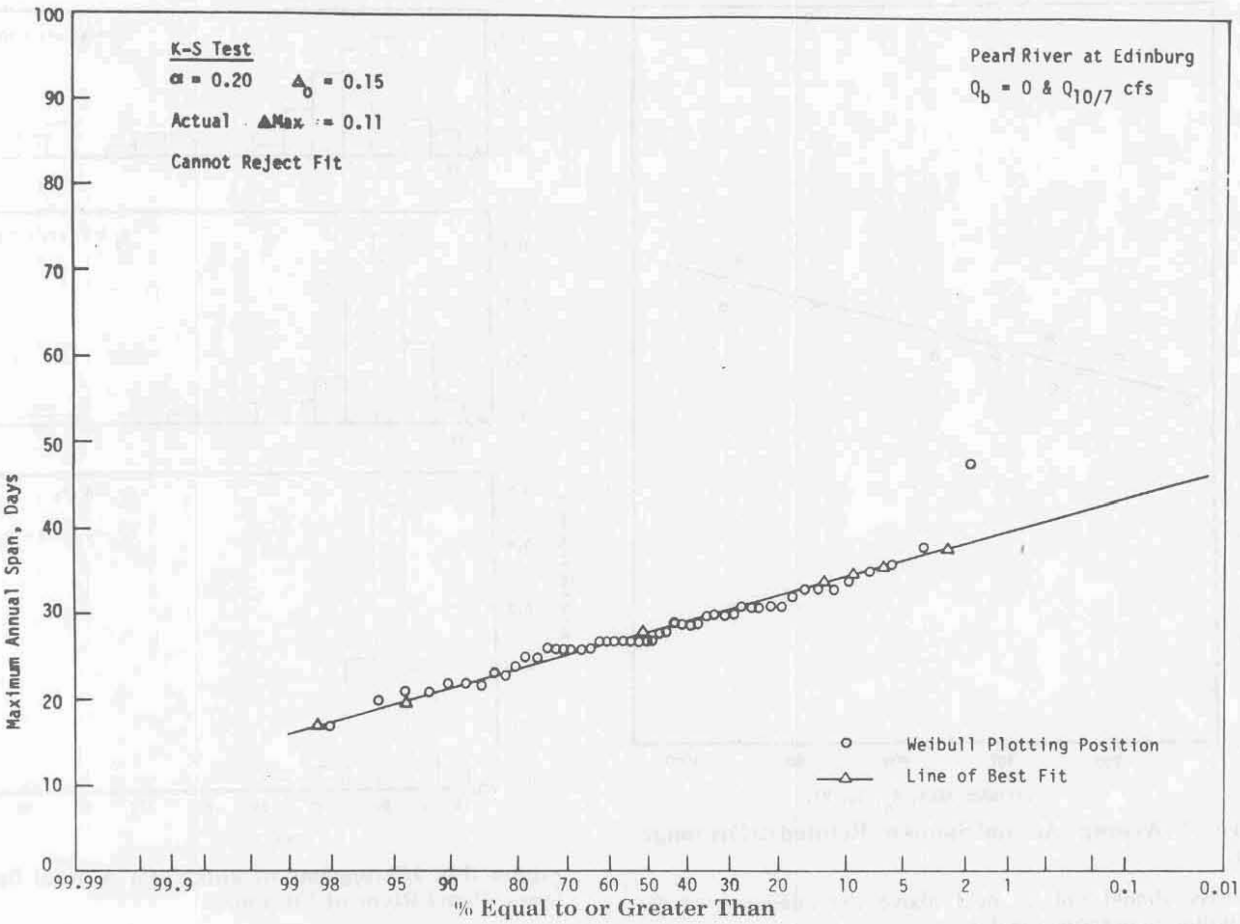

Figure 3-5. Normal Distribution of Best Fit

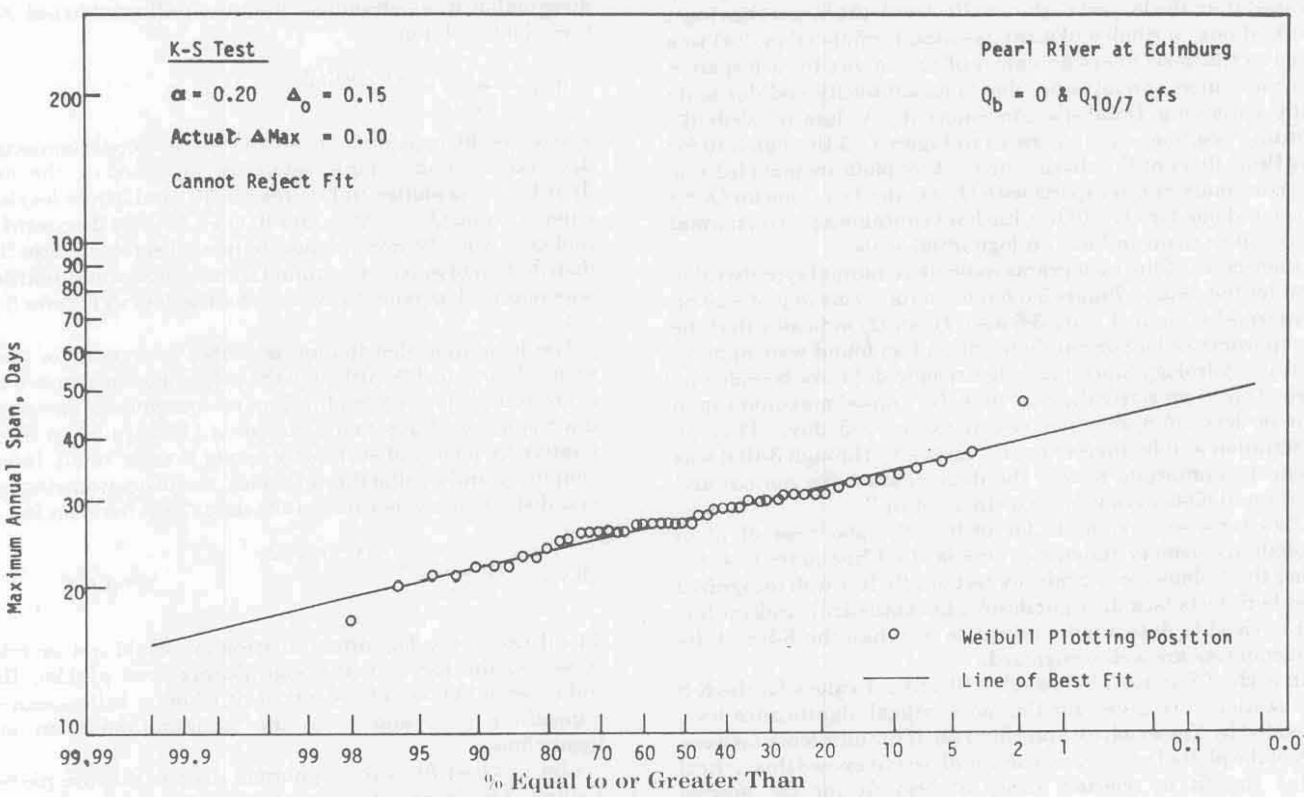

Figure 3-6. Log-Normal Distribution of Best Fit

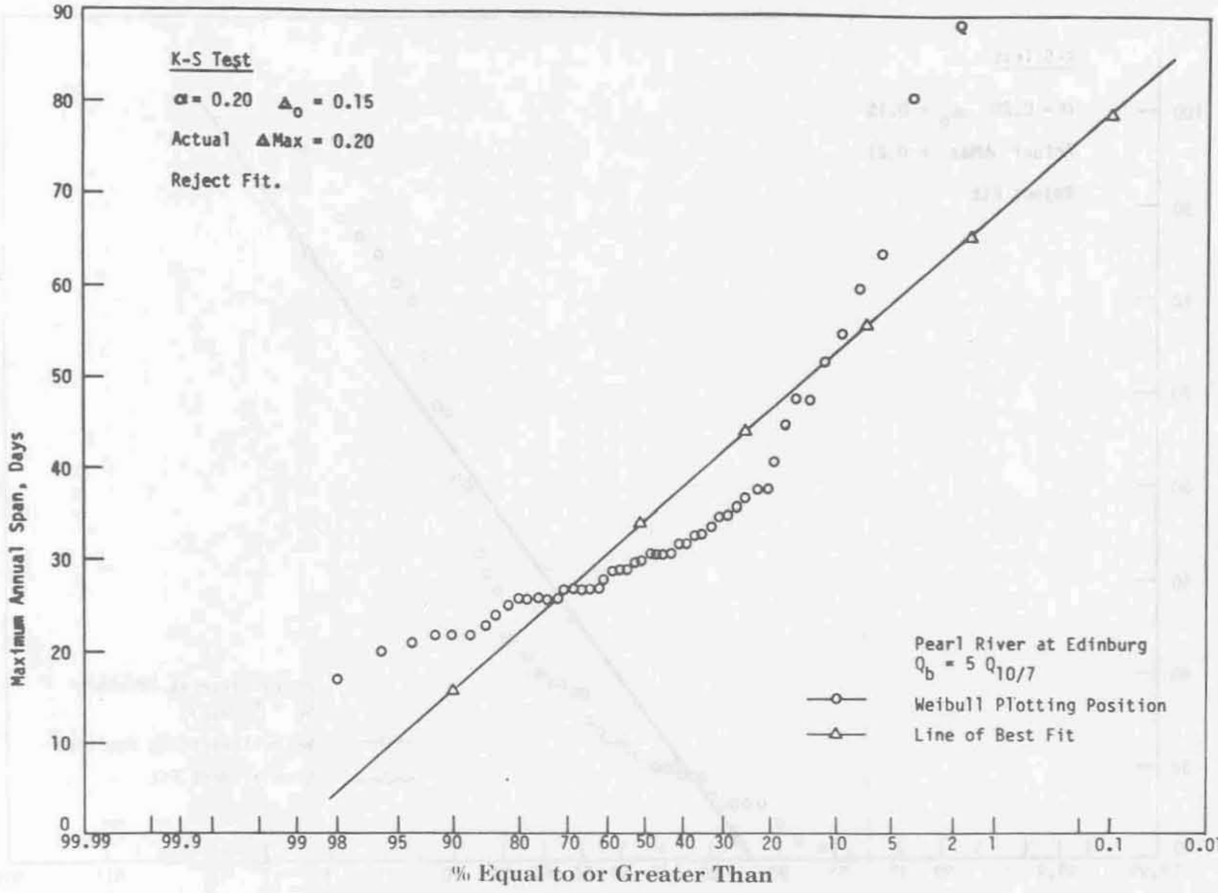

Figure 3-7. Normal Distribution of Best Fit.

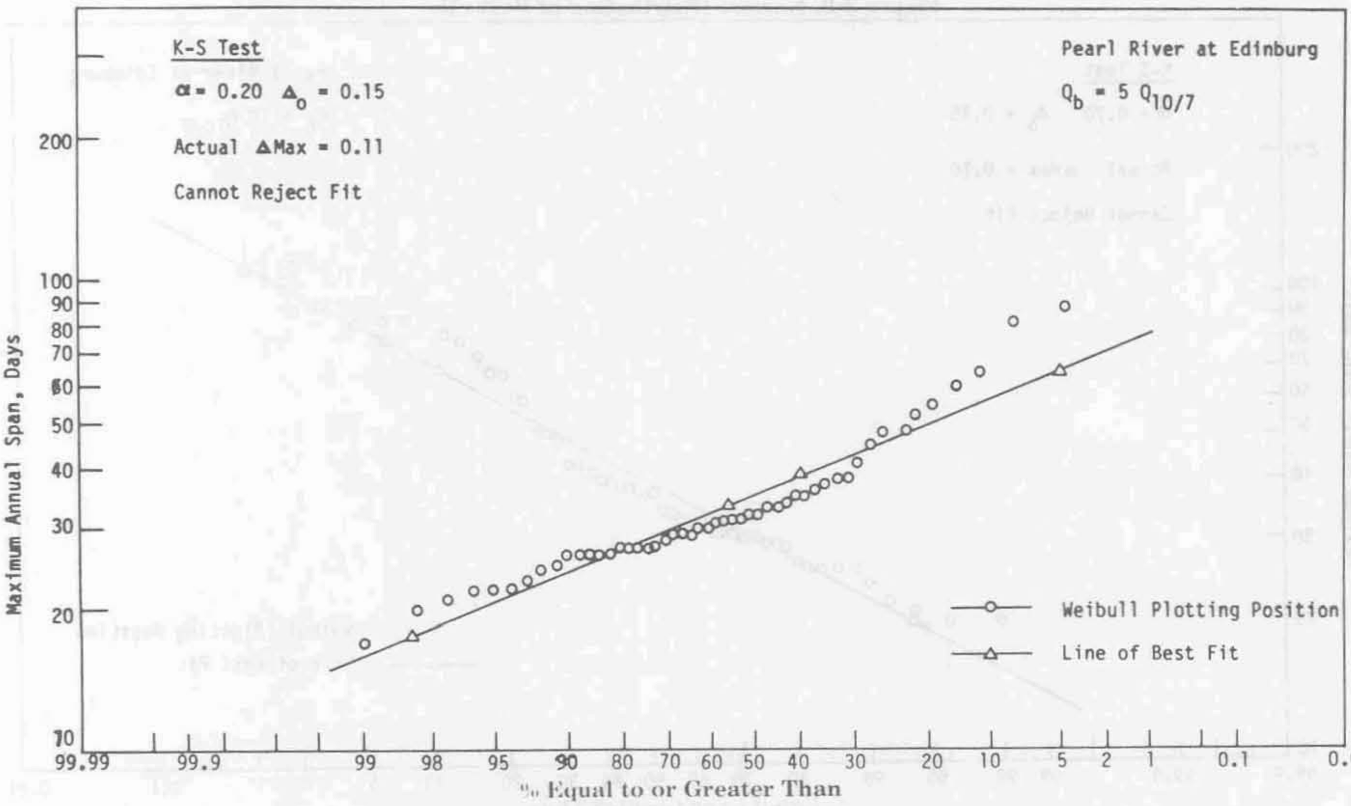

Figure 3-8. Log Normal Distribution of Best Fit.

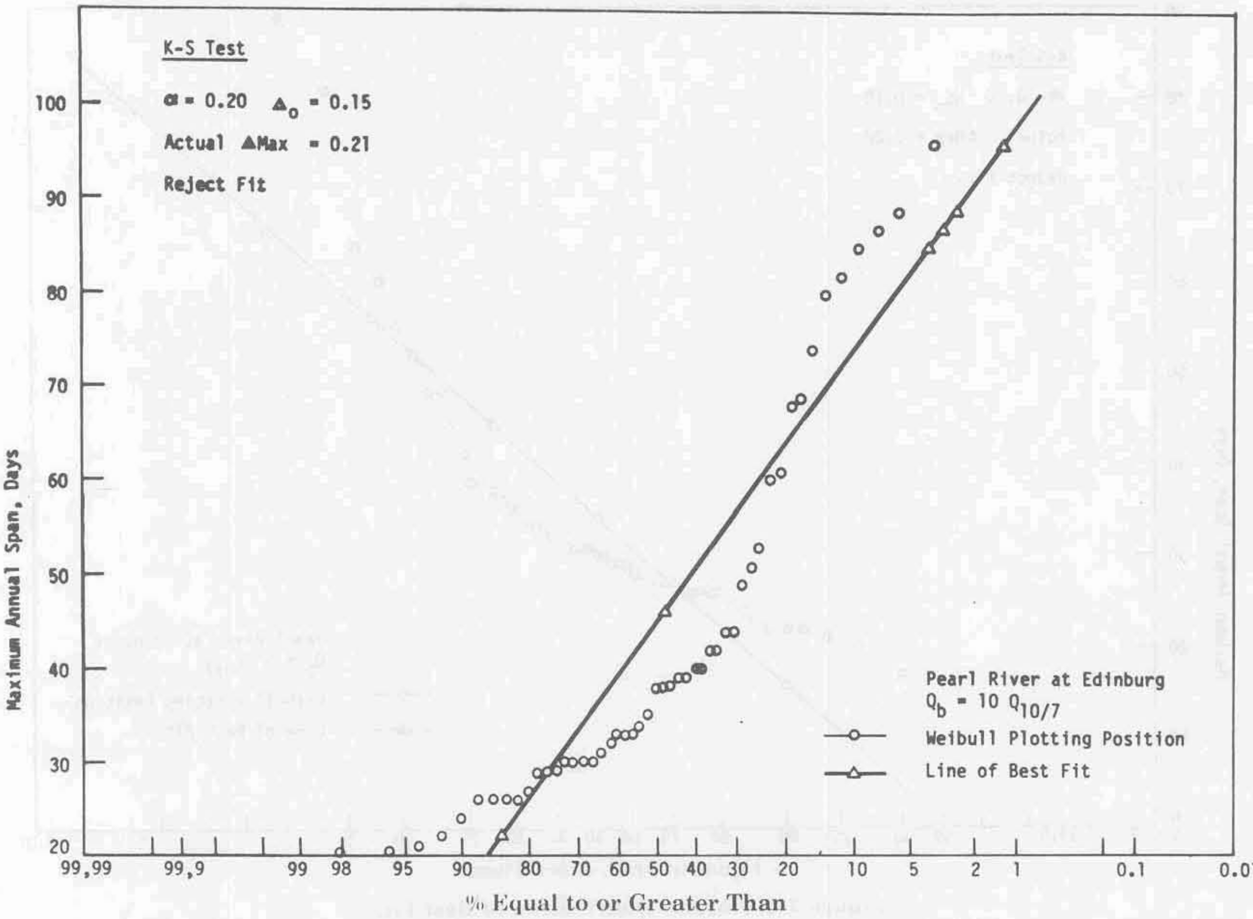

Figure 3-9. Normal Distribution of Best Fit.

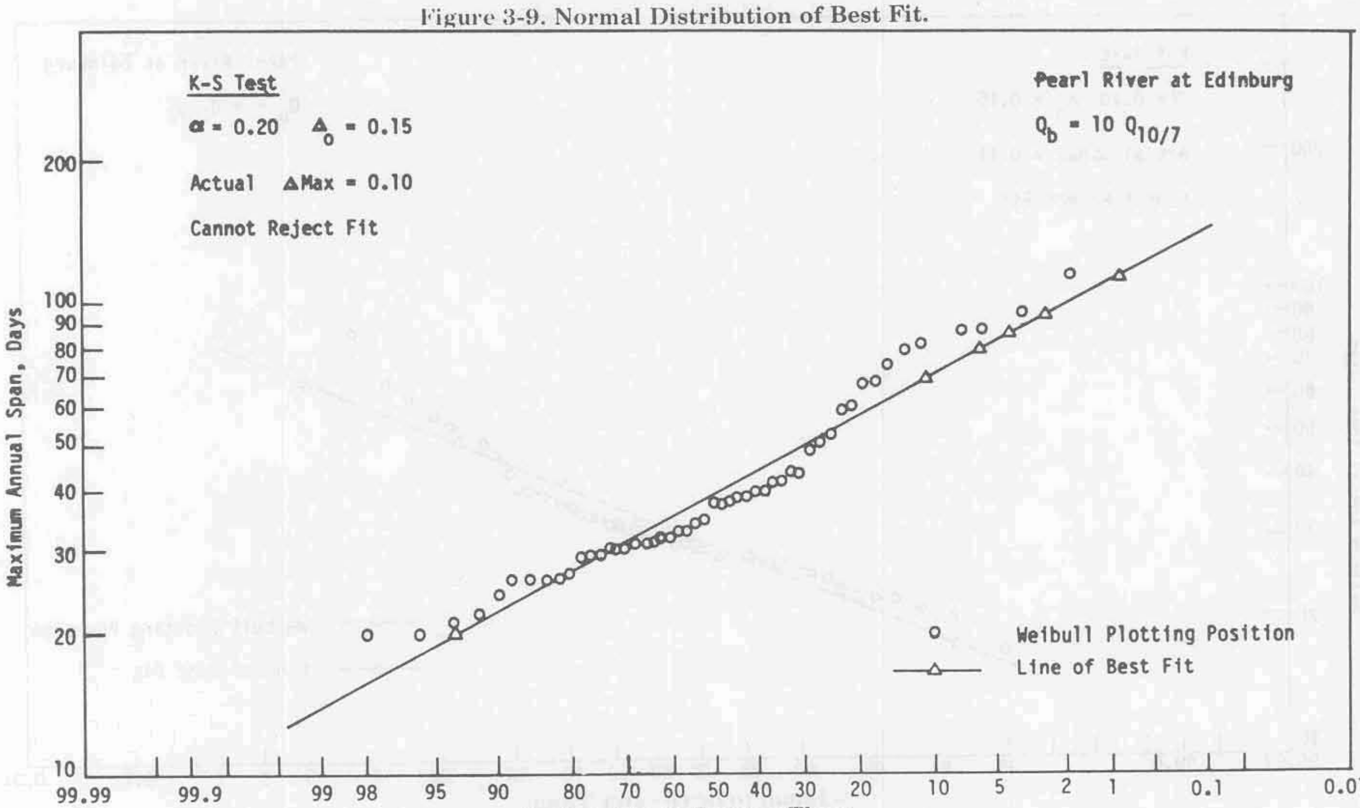

Figure 3-10. Log Normal Distribution of Best Fit.

seems to fit the data well. This is indicated by the fact that the K-S test cannot be used as a basis to reject the fit on any of the three plots at the 80% significance level.

The same distribution was applied to stream records within the state of Mississippi with acceptable fits. Thus it is the authors' opinion that the lognormal distribution is adequate to describe the distribution of maximum annual spans with levels of Q_0 from 0 to 10 $Q_{10,7}$. This does not rule out the fact that other distributions may also fit the data but the lognormal is adequate for application to Mississippi streams based on the Kolmogorov-Smirnov test.

SUMMARY AND CONCLUSIONS

Data from long term records of daily streamflows in Mississippi were analyzed for spans between hydrologic events. In this study an event was defined as a maximum in the time series irrespective of magnitude. Based on a statistical analysis of spans between events, the following conclusions can be drawn:

1. The mean monthly spans are cyclic having maximums in October and February or March and minimums in December and July or August.
2. Average spans between events tend to increase with increasing drainage area. More events occur annually within small basins than on large basins.
3. A two parameter lognormal distribution was deemed adequate by the Kolmogorov-Smirnov test to describe the distribution of maximum annual spans at the 0.80 significance level from $0 \leqslant Q_0 \leqslant 10\ Q_{10,7}$.

ACKNOWLEDGEMENTS

This work was supported in part by funds provided by the United States Department of the Interior, Office of Water Research and Technology, as authorized under the Water Research and Development Act of 1978 and by the State of Mississippi, Bureau of Pollution Control. The authors gratefully acknowledge the assistance of Bob Rice and Richard Bradley, students in the Department of Civil Engineering, who did much of the routine.

REFERENCES

1. Belanger, Stephen A., **Evaluation of Controlled Discharge of Stabilization Pond Effluent to Minimize Receiving Stream Impact,** M. S. Thesis, Mississippi State University, December 1978.
2. Haan, C. T., **Statistical Methods in Hydrology,** The Iowa State Press, 1977.
3. Shindala, A., V. L. Zitta, D. O. Hill and M. W. Corey, **Water Quality Studies on the Upper Tombigbee River,** Engineering and Industrial Research Station, Mississippi State University, CE-76-1, 1976.
4. Todorovic, P., "Stochastic Models of Floods," **Water Resources Research,** April 1978.
5. U. S. Geological Survey, **Water Resources Data for Mississippi.**
6. Yevjevich, Vujica, **Probability and Statistics in Hydrology,** Water Resources Publications, Fort Collins, 1972.
7. Yevjevich, Vujica, **Stochastic Processes in Hydrology,** Water Resources Publications, Fort Collins, 1972.