UNCERTAINTY OF FLOOD FREQUENCY ESTIMATES

Prepared by Nicholas C. Matalas
709 Glyndon St., S.E., Vienna, Va 22180

Under Subcontract to
Planning and Management Consultants, Ltd.
Carbondale, Il 62903

In Fulment of PMCL Contract
No. DACW 72-99-D-0005
Climate Variability and Climate Change
and the
Uncertainty of Flood Frequency Estimates

Under Contract to
U.S. Army Corps of Engineers
Institute of Water Resources
Ft. Belvoir, VA 22315

UPPER MISSISSIPPI RIVER SYSTEM FLOW FREQUENCY STUDY

APPENDIX I

Report II
Risk and Uncertainty Analysis

Contents

SUMMARY AND CONCLUSIONS

As prelude to an assessment of uncertainties in flood frequency analysis, the definitions of risk and uncertainty given in the Economic and Environmental Principles and Guidelines for Water and Land Related Resource Implementation Studies (U.S. Water Resources Council: 1983) are briefly reviewed. It is noted that while the Principles and Guidelines are clear in linking risk to probability, the Principles and Guidelines do not make it clear whether or not uncertainty is linked to probability. To render the definition of uncertainty a little clearer, Davidson's (1991) notion of "true uncertainty" is introduced, where true uncertainty cannot be substituted for by probability, leaving the notion of uncertainty to be used in a probability context, just as risk is used. In doing so, the uncertainties in flood frequency analysis, at least the uncertainties conditioned on the presumption of a probability distribution, objective or otherwise, can be dealt with in probabilistic terms. Thus the $T - year$ flood may be said to be unknown but knowable. It would be a contradiction in terms to say that the $T - year$ flood is unknowable. Probability is in reference to imperfect knowledge, not incomplete knowledge.

Traditionally flood frequency analysis has be based on the assumption that an observed sequence of floods is a realization of a sequence of independent and identically distributed (iid) random variables. Under this assumption, uncertainty in flood frequency estimates as mainly focused on 1) the length of the sequence, 2) the choice of the distribution function, 3) the method used to estimate the parameters of the chosen distribution function and 4) the definition of the plotting position. Because the length of a sequence can only increase with the passage of time, methods of regionalization are often used to transfer information to the site of interest from sites having longer sequences to effectively increase the length of the sequence of interest.

The issue of climate change has prompted hydrologist to question the iid assumption relative to hydrologic processes in general and to flood processes in particular. For the most part, attention has been given to detection of trends in hydrologic processes. Because hydrologic time series are short on a geologic scale, one can never be certain that a trend is not part of a slow oscillation unless the series ends. A trend contradicts the iid assumption, more specifically the assumption that the sequence of random variables are identically distributed. Moreover, one can never be certain that a slow oscillation is not a reflection of persistence which contradicts the iid assumption, more specifically the assumption that the sequence of random variables are independently distributed.

For this study, annual sequences of 1-Day high flows and annual sequences of peak flows at sites in the Upper Mississippi basin and in the Lower Missouri basin are considered. It was assumed that a flood sequence may be characterized by trend or persistence, where trend is linear and persistence is Markovian. Trend was a strong feature of the sequences while persistence was a less prominent feature. However, de-Markoving the sequences reduced the level of significance of trends in most of the sequences, and thereby, reduced the number of sequences having significant trends at the 1% or 5% levels. De-Markoving fully accounted for persistence, i.e. no sequence showed significant residual persistence at the 5% or higher level. De-trending the sequences reduced the level of significance of persistence in most of the sequences, and thereby, reduced the number of sequences having significant trends at the 1% or 5% levels. De-trending fully accounted for trend, i.e. no sequence showed significant residual trend at the %5 or higher level. These results suggests that their is an interaction between trend and persistence in that one partially accounts for the other.

The matter of trend and persistence is further assessed in relation to the frequency of flooding relative to a threshold elevation. For an arbitrary threshold elevation, flood water at time t, will either exceed the threshold elevation with probability p_t or not with probability $q_t = 1 - p_t$. Under the iid assumption, $p_t = p$, whereby $q_t = q = 1 - p \ \ \forall t$ – the flood process is said to be Bernoullian. Two non-Bernollian processes are considered, one marked by trend where the iid assumption is relaxed such that p_t, varies with t, and the other marked by persistence where the iid assumption is relaxed, such that the process of flooding is Markovian. In the case where the expected value of p_t is constant for the Bernoullian and the two Bernoullian processes, the non-Bernoullian process marked by trend yields a smaller standard deviation of the number of floods in a sequence of length n than the Bernoullian process. The non-Bernoullian process marked by persistence yields a larger standard deviation of the number of floods in a sequence of length n than the Bernoullian process if the first order autocorrelation coefficient is positive, and a smaller standard deviation if the first order autocorrelation coefficient is negative.

If the expected value of the probability defines the level of risk for a given level of negativity, then the expected risk is the same for the Bernollian and the two non-Bernoullian processes. However, the levels of uncertainty in the risk varies depending upon whether the iid assumption holds or not. In a more general setting of flood frequency analysis, it does not follow that the greater uncertainty, measured by the standard deviation, derives from trend or from persistence.

Under the iid assumption, a large share of the uncertainty in flood frequency analysis relating to the estimates of flood quantiles derives from the estimates of the parameters of the assumed distribution of floods. Of this share of uncertainty, most is attributed the skewness of the distribution. Because the estimate of skewness derived from a sequence of observed floods is subject to considerable sampling error, Bulletin 17-B places attention on improving at-site estimates of skewness . Bulletin 17-B also recognizes that the smaller floods effect the overall fit of the Log Pearson Type III distribution to an ordered sequence of observed floods. To reduce the uncertainty in the estimates of quantiles introduced through estimates of skewness, Bulletin 17-B gives attention to developing regionalized estimates of skewness and guidelines for censoring the smaller floods in a sequence.

Empirical evidence shows that for sequences of the logs of floods, in the Upper Mississippi and Lower Missouri basins at least, the right tails of observed distributions, on an at-site or on a regionalized basis, may be approximated by the right tails of Normal distributions. A Right-Tail Normal distribution is introduced, i.e. a Normal distribution that is fitted to the logs floods of magnitude greater than a specified threshold value. To assess the merit of a Right-Tail Normal distribution relative to a Pearson Type III distribution, the threshold is taken to be the median of the logs of the flows.

Three methods are considered for estimating the parameters of the Right-Tail Normal distribution, and are referred to as the inflection point method, the η – point method and the mirrored spread method, where $1 \le \eta \le n/2$ and n .denotes the sequence length. The three methods are neither inclusive or exclusive. The Right-Tail Normal distribution is seen to yield better estimates of the $T = 50 - Year$ flood than the Pearson Type III distribution in most of the 70 year sequences of annual 1-day high flows on an at-site basis and for the corresponding regionalized sequences. The at-site sequences in log space derived from the regionalized sequences in log space are said to be in quasi-log space. It is seen that all sequences in quasi log space are better fitted by Right-Tail Normal distributions than by Pearson Type III distributions in so far as estimating the $T = 50 - Year$ for 70 year sequences of annual 1-day high flows. The results for 70 year sequences of annual 1-day high flows carry over to 100 year sequences annual peak flows for both the estimates of the $T = 50 - Year$ floods and the $100 - year$ floods.

The Right-Tail Normal distribution is a Normal distribution whose parameters, the mean, μ, and standard deviation, σ, are estimated using only the observations of values greater than the threshold value, namely the median. None of the three methods are based on moments greater than the second and are therefore not dependent on the overall skewness of the observations. In effect, the Right-Tail Normal distributions side steps the need to either estimate the skewness of the observations or the censor some of the smaller observation followed by a redistribution of the total probability mass. In a sense, it can be said that using the Right-Tail Normal distribution instead of the Pearson Type III distribution reduces the uncertainty in the estimates of flood quantiles greater than the median.

If the chance negativities giving rise to flood risk are taken into account, it is quite likely that the negativities arise for floods above a threshold value. If that threshold is near the median flood, than a case can be made for choosing the Right-Tail Normal distribution over the Pearson Type III distribution. If the Right -Tail Normal distribution outperforms the Pearson Type III distribution in log space or in quasi-log space, then the Right-Tail Log Normal distribution will outperform the Log Pearson Type III distribution in real space, and conversely.

Figures

Figures (Continued)

Figures (Continued)

Figures (Continued)

Figures (Continued)

Figures (Continued)

Figures (Continued)

Figures (Continued)

Figures (Continued)

Figures (Continued)

Figures (Continued)

Figures (Continued)

Tables

Tables (continued)

Tables (Continued)

Tables (Continued)

UNCERTAINTY OF FLOOD FREQUENCY ESTIMATES

Introduction

The long standing practice of flood frequency analysis presumes sequences of observed annual floods to be realizations of independent and identically distributed (iid) random variables. In order to limit disagreements among federal agencies in estimating flood a quantile at a particular sites, the agencies have sought to identify a specific distribution to be used by all agencies in flood quantile estimation. Through the U.S. Water Resources Council, the federal agencies adopted the Log-Pearson Type III distribution with the understanding that unless another distribution could be shown to be better, the Log-Pearson Type III distribution would be used by all federal agencies in reference to a specific site. The procedures by which flood quantiles conditioned on the Log-Pearson Type III distribution are estimated are defined in Bulletin 17-B of the Interagency Committee on Water Data originally published by the U.S. Water Resources Council (1981).

Until recently, the iid assumption underlying flood frequency analysis has been unquestioned. As the issue of climate change moved higher on the public agenda, hydrologists have given greater attention to the study of the extent to which hydrologic change is a reflection of climate change induced by global warming or a reflection of change in land use. For the most part the studies have sought to show that flow sequences are marked by trends, whereby the iid assumption does not hold. Sequences of annual floods are seemingly less prone to reflect trends than sequences of annual mean or low flows. Because a trend may be part of a slow oscillator movement, an assessment of trend is best accompanied by an assessment of persistence. Without entering into study to establish the presence or absence of trends or of persistence in flood sequences, the uncertainty in flood frequency estimates induced by relaxation of the iid assumption is examined. Relaxation in the iid assumption leads to greater flood risk.

Uncertainty in flood frequency estimates is not limited to the questioning of the iid assumption. Even if the iid assumption holds, there is uncertainty related to statistical errors in the estimates of flood quantiles. Assessment of statistical errors has been and continues to be standard hydrologic practice.

Through flood frequency analysis, an estimate of a particular flood quantile, such as the magnitude of the 100-year flood, is derived. It is recognized that the estimate is subject to error. The error introduces uncertainty in using the estimated quantile in an operational situation. A decision for effecting mitigation of flood loses that in part, at least, incorporates the estimated quantile, is made under uncertainty. The uncertainty derives from several sources, of which the error of estimation in the flood quantile is one. It is that source which is a subject of investigation in this report.

The error in the estimation of a flood quantile is a function of 1) the length of the flood sequence, 2) the choice of probability distribution, 3) the method used to estimate the parameters of the chosen distribution and 4) the definition of the plotting position. Because the length of a sequence can only increase with the passage of time, efforts are made to effect an increase in record length through correlations with sequences of greater length, i.e. to effect a transfer of information to the site of interest from sites having longer sequences.

Of the various contributors to the error in the estimate of a flood quantile, only one is considered herein, namely, the contribution derived via the estimation of the skewness. The statistical sampling error in the estimate of skewness is recognized as a major source of uncertainty in the an estimate of a flood quantile. In the following discussion, the estimation of skewness per se is not addressed. Rather, the issue of estimating skewness is side stepped in log space by considering the use of a Right-Tail Normal distribution in place of the Pearson Type III distribution. The Right-Tail Normal distribution is a Normal distribution whose right tail is determined from the logs of the observation without fit of the left tail. If the Right-Tail Normal distribution provides better estimates of quantiles in log space than the Pearson Type III distribution, then the Right-Tail Log-Normal distribution provides better estimates of quantiles in real space than the Log-Pearson Type III distribution, and conversely. By considering a Right-Tail Normal distribution, there is no need to consider censoring some of the smaller floods to effect a better estimates of flood quantiles via a truncated Log Pearson Type III distribution as outlined by Bulletin 17-B.

The relative goodness of fit of the Right-Tail Normal distribution and the Pearson Type III distribution is examined on both an at-site and regional basis.

Assessing the effects of relaxation in the iid assumption provides better understanding of flood risk and uncertainty. Adopting a Right-Tail Log-Normal distribution as an alternative to the Log-Pearson Type III distribution effects, so to speak, a reduction in flood uncertainty without increasing costs or reducing benefits. Basically the Right-Tail Log-Normal distribution effects simplification of flood frequency analysis, and it puts the emphasis where it should be – on the fitting of the right tail to an ordered sequence of observed floods.

The study is based on observed sequences of annual 1-day high flows and annual peak flows at sites within the Upper Mississippi and Lower Missouri basins. With focus on flood risk and uncertainty, the definitions of risk and uncertainty given by the U.S. Water Resources Council (1983) in their report, Economic and Environmental Principals and Guidelines for Water and Related Resource Implementation Studies are briefly reviewed.

Prelude to the study, the spectrum of sequences of extreme flows of various durations, varying from annual 1-day low flow to annual mean flow to annual 1-day high flows are statistically described on both an at-site and a regional basis.

Background

Study Region and Data Base

The study is limited to sequences at sites within the Upper Mississippi and Missouri basins. The Missouri basin contributes about 528,000 mi^2 to the total drainage area of about 697,000 mi^2 for the two basins at St. Louis, MO. The two basins extend over 21° of longitude, approximately 1,000 mi and 8 ° of latitude, approximately 550 mi. Streamflow sequences at 44 sites, 26 in the Upper Mississippi basin and 18 in the Lower Missouri basin are used to assess the uncertainty of flood frequency estimates. The 44 sites are partitioned into two sets. The first set consists of 21 sites in the Upper Mississippi basin and 11 sites in the Lower Missouri basin. See Table 1.

Table 1: General Description of Flow Sites in the Upper Mississippi and Lower Missouri Basin

Stream	Locale	State	Latitude (deg.)	Longitude (deg.)	Drainage Area (sq. mi.)	Period of observed Sequence
			Upper Mississippi Basin			
St. Croix	St. Croix	WI	45.41	-92.65	6,240	1911-1998
Jump	Sheldon	WI	45.31	-90.98	576	1916-1998
Black	Neillsville	WI	44.58	-90.61	749	1914-1998
Maquaketa	Maquaketa	IA	42.08	-90.63	1,553	1914-1998
Mississippi	Clinton	IA	41.78	-90.25	85,600	1874-1998
Rock	Afton	WI	42.61	-89.07	3,340	1915-1998
Sugar	Broadhead	WI	42.61	-89.40	923	1915-1998
Pecatonica	Freeport	IL	42.30	-89.62	1,326	1915-1998
Cedar	Cedar Rapids	IA	41.97	-91.67	6,510	1903-1998
Skunk	Augusta	IA	40.75	-91.28	4,303	1915-1998
Mississippi	Keokuk	IA	40.39	-91.37	119,000	1879-1998
Des Moines	Stratford	IA	42.25	-94.00	5,452	1921-1998
Raccoon	Van Meter	IA	41.53	-93.95	3,441	1916-1998
Iroquois	Chebanse	IL	41.01	-87.82	2,091	1924-1998
Kankakee	Momence	IL	41.16	-87.67	2,294	1916-1998
Spoon	Seville	IL	40.49	-90.34	1,636	1915-1998
La Moines	Ripely	IL	40.03	-90.63	1,293	1922-1998
Meramec	Steelville	MO	38.00	-91.36	781	1923-1998
Bourbeuse	Union	MO	38.45	-90.99	808	1922-1998
Big	Byrnesville	MO	38.36	-90.65	917	1923-1998
Meramec	Eureka	MO	38.51	-90.59	3,788	1922-1998
			Lower Missouri Basin			
Yellowstone	Corwin Springs	MT	45.11	-109.21	2,623	1911-1998
Clarks Fork	Belfry	MT	45.01	-108.94	1,154	1922-1998
Yellowstone	Billings	MT	45.80	-107.53	11,795	1929-1998
Big Sioux	Akron	IA	42.84	-95.44	8,424	1929-1998
North Platte	Northgate	CO	40.94	-105.66	1,431	1916-1998
Bear	Morrison	CO	39.65	-104.80	164	1920-1998
Elkhorn	Waterloo	NE	41.29	-95.72	6,900	1929-1998
Nishnabottna	Hamburg	IA	40.63	-94.37	2,806	1929-1998
Grand	Gallatin	MO	39.93	-92.06	2,250	1921-1998
Thompson	Trenton	M)	40.08	-92.36	1,670	1929-1998
Gasconade	Jerome	MO	37.93	-90.02	2,840	1924-1998

Flow sequences of the first set of sites were structured from the daily flows taken from the Hydro-Climatic Data Network (HCDN) developed by Slack, Lumb and Landwehr (1993). The structured flows were the set of annual (October 1 through September 30) extreme flows for various durations, varying from the annual $1-day$ low flow to the annual mean flow to the annual *1 – day* high flow. The set of structure flows is referred to as the extreme flow spectrum. See Appendix A and Appendix B.

For purpose of study, the flow sequences all spanned the 70 year period 1927 through 1998. The period is the longest common period at the sites in the two basins.

The second set formed by the partition of the 44 sites in the two basin consists of 7 sites in the Upper Mississippi basin and 7 sites in the Lower Missouri basin, where 2 sites in the Upper Mississippi basin are also in the first set. See Table 2.

Table 2: General Description of Flow Sites on the Upper Mississippi and Lower Missouri Rivers

Locale	State	Latitude (deg.)	Longitude (deg.)	Drainage Area (sq. mi.)	Period of observed Sequence
Upper Mississippi River					
St. Paul	MN	45.00	-93.17	36,800	1867-1995
Winona	MN	44.03	-91.62	59,200	1885-1995
Dubuque	IO	42.52	-90.68	82,000	1879-1996
Clinton	IO	41.25	-90.20	85,600	1875-1996
Keokuk	IO	40.38	-91.42	119,000	1875-1996
Hannibal	MO	39.68	-91.33	137,000	1879-1996
St. Louis	MO	38.67	-90.25	697,013	1861-1995
Lower Missouri River					
Sioux City	IO	42.50	-96.47	314,600	1898-1997
Omaha	NE	41.25	-96.00	322,820	1889-1997
Nebraska City	NE	40.68	-95.83	414,439	1889-1997
St. Joseph	MO	39.77	-94.87	429,340	1889-1997
Kansas City	MO	39.03	-94.55	489,162	1889-1997
Booneville	MO	38.97	-92.71	505,710	1889-1997
Hermann	MO	38.71	-91.43	528,200	1889-1997

The flow sequences of the second set of sites are annual peak flows. For purpose of study, all the sequences at sites along the Upper Mississippi river spanned the 100-year period 1896-1995, and all the sequences along the Lower Missouri river spanned the 100-year period 1898-1997.

Geographic Distribution of Sites

The geographic distributions of the sites belonging to the two sets are shown schematically in Figures 1 and 2.

Statistical Descriptors of Sequences

The sequences are assessed in real space and in log space, and on an at-site basis and on a regionalized basis. An arbitrary sequence of length n, $\{x_t: t = 1, \ldots, n\}$ in real space, on an at-site or regionalized basis, is described in terms of the standard statistical descriptors, namely, the mean , $\bar{x}$, standard deviation, s_x, the coefficient of variation, Cv, coefficient of skewness, Sk, and the coefficient of kurtosis, Ku, where

$$\bar{x} = \sum_{t=1}^{n} x_t \Big/ n \qquad (1)$$

$$s_x = \left\{ \sum_{t=1}^{n} (x_t - \bar{x})^2 \Big/ (n-1) \right\}^{1/2} \qquad (2)$$

$$Cv = s / \bar{x} \qquad (3)$$

$$Sk = \{ n / (n-1)(n-2) \} \sum_{t=1}^{n} \left[(x_t - \bar{x}) / s_x \right]^3 \qquad (4)$$

$$Ku = \{ n(n+1) / (n-1)(n-2)(n-3) \} \sum_{t=1}^{n} \left[(x_t - \bar{x}) / s_x \right]^4 - \left\{ 3(n-1)^2 / (n-2)(n-3) \right\} \qquad (5)$$

Description in log space, on an at-site or regionalized basis, is given by replacing x_t by $y_t = ln(x_t)$.

See Appendix C.

Risk and Uncertainty

The notions of risk and uncertainty are central to water management. Federal water agencies deal with risk and uncertainty following the definitions of risk and uncertainty given by the U.S. Water Resources Council (1983) in the Economic and Environmental Principles and Guidelines for Water and Related Land Resource Implementation Studies. The Council's report is referred to herein as the Principles and Guidelines. The definition of risk is fairly clear, however, the definition of uncertainty is murky. Often precise definitions of terms are not necessary. But in cases where imprecision in the definition of terms limits the range of view of the issues under study, clear definitions of terms are warranted. Such is the case with the definition of uncertainty in water management.

Principles and Guidelines Perspective

In the Principles and Guidelines, risk and uncertainty are defined in the context of situations that relate to the ability to describe potential outcomes in probabilistic terms. Outcomes may be interpreted as possible consequences sequel to an act, i.e. a decision. Within this context, the Council states that

> situations of risk are detained [defined] as those in which the potential outcome can be described in reasonably well-known probability distributions such as the probability of particular flood events

and that

> situations of uncertainty are defined as those in which potential outcomes cannot be described in objectively known probability distribution.

In the case of risk, the potential outcomes can be described in "reasonably well known probability distributions." It does not matter if the probability distribution is reasonably well known or not. What matter is that risk entails probability. The extend to which the probability distribution is known affects how well risk can be assessed. It should be noted that linking risk to probability leaves open the question whether probability is best interpreted as being objective or personalistic. Addressing the question is outside the scope of this study.

In the case of uncertainty, the Principles and Guidelines state that the outcomes "can not be described [by] objectively known probability distributions" because as stated later, "there are no known probability distributions to describe uncertain outcomes." It is unclear what the Principles and Guidelines mean by there being "no known probability distributions." Are uncertain outcomes governed by a probability law with which one has no familiarity? Is the governing probability law unknown because all the potential outcomes are unknown? Is the probability law unknown because the potential outcomes are not governed by a probability law? Is the probability law unknown because the law is not an objective law but a stated degree of belief that accords with the calculus of probability?

Risk and uncertainty are said to derive from measuring error and from the variability that is inherent in natural, social and economic processes. No less important are sampling errors in parameter estimates that derived from "short" finite lengths of observations on natural, social and economic processes in cases where the processes may be considered to be stochastic processes.

The Principles and Guidelines state that some risks and uncertainties can be eliminated through increased project costs or through reduction in program benefits, while others can not. Elimination of risk to a degree, i.e. reduction in risk, may follow from various actions, e.g., collection of additional information, using techniques that offer higher orders of approximation, incorporating more stringent factors of safety into the system design, hedging on committing to large capital investments early in the development stage, and conducting sensitivity analyses of the estimated costs and benefits of alternative system designs. Adding robustness, redundancy and resilience to project designs effects a reduction in risk with the occurrence of greater project costs. See Matalas and Fiering (1977).

Alternative Perspective

Risk relates to the chance occurrence of something unwanted. Rescher (1983) expressed risk as the chancing of negativity and noted that there is a difference in taking a risk and in facing a risk. To take a risk is to make a conscious decision that enhances the occurrence of some unfortunate event., e.g. the loss of property or life as a result of an extreme flood sequel to deciding to dwell on the flood plain. To face a risk is to be so positioned that harm might be experienced, e.g. the situation of the residents of Johnstown, Pennsylvania on the morning of May 31, 1889.

Risk, R, is often measured as

$$R = CP \tag{6}$$

where C denotes the adverse consequence whose occurrence with probability P is sequel to a specific act, i.e. a (flood management) decision. Adverse consequence refers to damage or loss of property or to injury or loss of life. Eq. (6) defines risk as an expected value. In choosing among risks, one may choose that risk having the smaller expected value. However, the expected value does not preclude the use other measures of risk.

In taking or facing risk, one must contend with uncertainty, where uncertainty may derive from imperfect knowledge or from incomplete knowledge. Imperfect knowledge presumes that all the possible consequences (outcomes) of an act, i.e. decision, are known or could be known with the expenditure of resources prior to the act. In effect, the sample space is known and therefore probabilities, objective or otherwise, can be assigned to the possible outcomes. Knowledge is said to be imperfect because, though we know that one of a set of possible outcomes will follow from a decision, we do not know with certainty which outcome it will be. Because our knowledge is imperfect, our distribution of the total probability mass over all the outcomes is very likely to be questionable.

Incomplete knowledge, i.e., ignorance, implies that not only are some if not all the outcomes unknown at the time a decision is to be made, they are unknowable. Thus the sample space is incomplete and expenditure of resources can not reveal the unknown outcomes. Because we do not know just what it is that is unknown, resources can not be directed to make that which is unknown known. As put by Shackle (1949), knowledge can not be gained before its time. Because the sample space is incomplete, probabilities, personalistic or otherwise, can not be assigned to those outcomes that are not known.

A situation of incomplete knowledge is a situation of a special case of uncertainty that Davidson (1991) termed "true" uncertainty. By defining incomplete knowledge as a situation of true uncertainty, the term uncertainty can be used in cases where risk and uncertainty are understood within the same linguistic context, namely, that of probability. For example, it seems natural to speak of the uncertainty of a gain or a loss and to speak of the risk of a loss rather than the uncertainty of a gain. It seems more natural to speak of the uncertainty in statistical estimates rather than the risk in statistical estimates. Though a risky situation may not be a strictly uncertain situation, risk and uncertainty may each be mapped into probability. The sampling errors associated with estimates of statistical parameters are described by a probability distribution.

By differentiating between uncertainty and true uncertainty, one may identify situations where a metric other than probability must be used to measure true uncertainty. If such situations do not exist within the domain of water management, then a risky situation may be viewed similarly as an uncertain situation, situations calling for a probability measure. It follows that all possible outcomes are known and thereby, the total probability mass may be distributed over the outcomes. If there are situations of true uncertainty, then the general framework under which water management is conducted will need to be reviewed by the federal water agencies and other responsible parties in the business and academic community. Clearly, the definition of uncertainty in the Principles and Guidelines needs clarification. Discussion of a metric of true uncertainty is outside the scope of this study.

Flood Frequency Analysis

Rafter (1895) was among the first to suggest adopting a probabilistic perspective of floods. Within 20 years, flood frequency analysis evolved into its present form through the work of Hazen (1914), Fuller (1914) and others. Basically, flood frequency analysis presumes, implicitly at least, a sequence of observed annual floods to be a realization of a sequence of independent and identically distributed random variables. The observations are ranked from smallest to largest and to each, a probability is assigned. The assigned probability, defined in terms of the total number of observations in the sequence, n, and the rank value, i, of the ordered observations, where $i = 1, \ldots, n$, is referred to as the plotting position, an estimate of the probability, $P(i)$, of exceeding the magnitude of the $i-th$ ranked flood. A probability distribution of specific mathematical form is fitted to the ranked observations, where the parameters of the distribution are estimated from the observations by a specific statistical method. From the fitted distribution, a flood magnitude, $x(P')$, corresponding to a specified exceedence probability, P', can be obtained graphically, analytically or by numerical integration.

The inverse of the exceedence probability is referred to as the return period,

$$T = 1/P \tag{7}$$

The quantity $x(T') = x(P')$ is referred to as the $T' - year$ flood, the flood having an exceedence probability of P'.

In the United States, flood frequency analysis is currently conducted by federal agencies in accordance with Bulletin 17-B. Basically, the bulletin spells out the procedure for estimating the $T - year$ flood via fitting the Log-Pearson Type III distribution to an ordered set of observed floods. The estimates of the exceedence probabilities of the observations are defined as Weibull plotting positions,

$$\begin{aligned} \hat{P}\left[x > x^{(i)}\right] &= P(i) \\ &= i/(n+1) \end{aligned} \tag{8}$$

where, $\hat{P}\left[x > x^{(i)}\right] = P(i)$ denotes the estimated exceedence probability of the $i-th$ ranked flood, i.e. the probability of a flood x exceeding the observed flood of rank i, $x^{(i)}$, in a sequence of n observed floods.

The Log-Pearson Type III distribution is fitted using the method of moments to estimate the location and scale parameters of the distribution. The measure of the distribution's skewness (degree of asymmetry) is estimated through a procedure that takes into account estimates of measures of skewness of observed flood sequences at other sites in the "vicinity" of the primary site of interest, the site where an estimate of the $T-year$ flood is specifically sought.

Bulletin 17-B allows for estimating the $T-year$ flood via fitting an alternative distribution if it can be shown that use of an alternative distribution is warranted, i.e. better in some meaningful way. However, the bulletin does not set forth explicit terms on which judgment can be made as to whether or not the use of an alternate distribution is warranted.

The $T-year$ flood is unknown, but it is, of course, knowable. It would be a contradiction in terms to say that the $T-year$ flood is unknowable. An estimate of the $T-year$ flood is a conditional statement – the estimate is conditioned on an assumed distribution, in particular the Log-Pearson Type III distribution. As a conditional statement, reference may be made to the uncertainty in the value of the $T-year$ flood. Traditionally, concern over the uncertainty in the estimate of the $T-year$ flood has focused on the following factors: 1) the length of the sequence, 2) the choice of the distribution function, 3) the method used to estimate the parameters of the chosen distribution function and 4) the definition of the plotting position. Because the length of a sequence can only increase with the passage of time, efforts are made to effect an increase in record length through correlation with sequences of greater length, i.e. to effect a transfer of information to the site of interest from sites having longer sequences. These efforts are referred to as regionalization.

The issue of global warming has broaden the concern over uncertainty in the $T-year$ flood with focus on trend in the observations of annual floods. This focus is on questioning the presumption that a sequence of observed floods to be a realization of a sequence of identically distributed random variables. Hydrologic time series are short on a geophysical scale so we can never be certain that a trend is not part of a slow oscillation unless the series ends. See e.g. Kendall and Stuart (1966). Moreover, we can never be certain that oscillatory movement is not a reflection of persistence implying that a sequence of observed floods are not a realization of a sequence of independently distributed random variables. Thus, uncertainty in the $T - year$ flood as a consequence of global warming calls into question the assumption that a sequence of observed floods is a realization of a sequence of independent and identically distributed random variables.

An Assessment of Trend and Persistence

The issue of global warming has prompted an examination of hydrologic time series for trends. Trend is viewed as a sustained change in flow with time, where change is in terms of the mean or some other statistical characteristic of an observed flow sequence. An observed hydrologic flow sequence is regarded as a realization of a stochastic process. In the presence of trend, the stochastic process is in some way non-stationary. For example, the process may be nonstationary in the mean, but stationary in all other respects. If the process is nonstationary in the $k-th$ order moment, presuming that the moment exists, then the process is non-stationary in all lower order moments.

If a trend is indeed a trend, positive or negative, it has a beginning and an end. A trend can not persist indefinitely. If a trend were to persist indefinitely, flow would either exceed the carrying capacity of the basin or the basin would become dry. See Olsen et al (1999). A trend may in fact be a "trend" reflecting an oscillatory wave of low frequency. More generally, a time series may be a composition of a large number of oscillatory waves of varying frequency and such that the time series is stationary though perhaps persistent. Stationary persistence is measured by the dependence between flows at different times, where the degree of dependence is determined by the interval between the times and not by the times themselves. If the frequency of a wave translates into a period greater than the observation period of the time series, then the series may appear to be trending. See e.g. Planning & Management Consultants, Ltd. (1999).

The distinction between trend and persistence is operationally important. In particular, the distinction is important to flood frequency analysis. Heretofore, flood frequency analyses have been conducted under the assumption, implicit or otherwise, that flood sequences are realizations of stationary stochastic processes. More specifically, a flood sequence has been viewed as a sequence of independent and identically distributed (iid) random variables, a realization of a purely random process.

Under the iid assumption, the future mirrors the past, whereby an estimate of a specific flow quantile based on an observed flow sequence differs from the quantile by amount within the bounds of sampling error. The distribution of sampling errors is the same for all future sequences of length equal to that of the observed sequence spanning the past. In the presence of trend, flows may be independently distributed, but not identically distributed. In the presence of persistence, flows may be identically distributed, but not independently distributed. In the presence of both trend and persistence, the flows are neither identical or independently distributed. If the iid assumption does not hold, then the uncertainty in estimates of flow quantiles can not be accounted for in terms of classical sampling errors, i.e. by sampling errors presuming the iid assumption holds implying that an observed flow sequence is a realization of a purely random process.

In the following discussions, trend is limited to trend in the mean and is measured by the regression of flow on time. Persistence is limited to Markovian persistence and is measured by the first order autocorrelation coefficient. Limiting trend to linear trend over the period of observation and limiting persistence to Markovian persistence is meant to provide a first-order account of trend and persistence. In actuality, trend and persistence, if they exist, may be of more complicated forms. Linear trend is measured by the coefficient of linear regression of flow on time. Markovian persistence is measured as linear temporal dependence by the first order autocorrelation coefficient.

Account is taken of residual trend, i.e. trend following de-Markoving flow sequences, and residual persistence, i.e. persistence following de-trending flow sequences. The analytical structure of de-Markoving and de-trending are given in Appendix D.

Selected Elements of Flow Spectrum: Annual 1-Day High, Mean and 1-Day Low

At the selected sites in the Upper Mississippi and Missouri basins, an assessment is made of trends and persistence in the sequences of annual 1-day high flows, annual mean flows and annual 1-day low flows. The assessment is summarized in Tables 3 through 5. For other elements of the flow spectrum, the assessment of trend and persistence is given in Appendix E

Table 3: Trend and Persistence of Annual 1-Day High Flows

	Trend			Persistence		
	Obs.	DT	DM	Obs.	DT	DM
			Upper Mississippi Basin			
St. Croix	0.003	0.000	0.004	0.186	0.182	0.013
Jump	-0.002	0.000	-0.002	0.083	0.081	-0.003
Black	-0.006	0.000	-0.006	-0.112	-0.127	-0.005
Maquaketa	-0.007	0.000	-0.007	0.082	0.070	-0.010
Mississippi	0.013*	0.000	0.011	0.139	0.072	-0.020
Rock	0.003	0.000	0.003	0.015	0.005	0.005
Sugar	-0.014	0.000	-0.014*	0.081	0.017	0.017
Pectonica	-0.010	0.000	-0.010	-0.007	-0.032	0.008
Cedar	0.000	0.000	0.000	0.081	0.080	0.014
Skunk	0.008	0.000	0.009	-0.048	-0.079	-0.002
Mississippi	0.016**	0.000	0.013*	0.143	0.034	-0.006
Des Moines	0.006	0.000	0.007	-0.037	-0.056	-0.005
Raccoon	0.013*	0.000	0.012*	-0.051	-0.118	0.017
Iroquois	0.017**	0.000	0.013*	0.215	0.115	0.031
Kankakee	0.025**	0.000	0.019**	0.223	-0.083	-0.052
Spoon	0.014**	0.000	0.014*	0.039	-0.054	-0.002
La Moines	0.017**	0.000	0.014*	0.241*	0.126	0.007
Meramec	0.007	0.000	0.003	-0.063	-0.074	0.034
Bourbeuse	0.015**	0.000	0.014*	0.053	-0.038	-0.007
Big	0.015**	0.000	0.013*	0.118	0.031	-0.006
Meramec	0.012*	0.000	0.012	0.028	-0.031	-0.003
Average	0.007	0.000	0.006	0.067	0.006	0.001
Stdev	0.010	0.000	0.009	0.101	0.085	0.018
			Missouri Basin			
Yellowstone	0.014*	0.000	0.013*	0.153	0.076	0.009
Clarks Fork	0.009	0.000	0.009	0.211	0.183	0.066
Yellowstone	0.011*	0.000	0.011	0.214	0.169	0.059
Big Sioux	0.008	0.000	0.009	-0.022	-0.052	0.008
North Platte	0.002	0.000	0.002	-0.033	-0.035	0.001
Bear	-0.006	0.000	-0.007	-0.035	-0.051	-0.008
Elkhorn	0.010*	0.000	0.011	-0.090	-0.133	0.001
Nisnabottna	0.014**	0.000	0.010	0.039	-0.036	0.006
Grand	0.009	0.000	0.010	0.007	-0.043	-0.001
Thompson	0.010*	0.000	0.007	0.263*	0.232	-0.019
Gasconade	0.007	0.000	0.008	-0.036	-0.058	0.006
Average	0.008	0.000	0.007	0.061	0.023	0.012
Stdev	0.006	0.000	0.005	0.125	0.121	0.026

Table 4: Trend and Persistence in Annual Mean Flows

	Trend			Persistence		
	Obs.	DT	DM	Obs.	DT	DM
			Upper Mississippi Basin			
St. Croix	0.017**	0.000	0.011	0.535**	0.473**	0.033
Jump	0.004	0.000	0.004	0.261*	0.252*	0.068
Black	0.010*	0.000	0.009	0.179	0.142	0.035
Maquaketa	0.014**	0.000	0.008	0.366**	0.309*	0.020
Mississippi	0.022**	0.000	0.012*	0.471**	0.333*	0.022
Rock	0.020**	0.000	0.013*	0.376**	0.235	0.050
Sugar	0.016**	0.000	0.010	0.347*	0.261*	0.030
Pectonica	0.015**	0.000	0.010	0.312*	0.230	0.044
Cedar	0.021**	0.000	0.014*	0.343*	0.197	-0.013
Skunk	0.015**	0.000	0.011	0.173	0.087	0.021
Mississippi	0.024**	0.000	0.014*	0.441**	0.260*	0.004
Des Moines	0.021**	0.000	0.015*	0.351*	0.204	-0.034
Raccoon	0.022**	0.000	0.015*	0.296*	0.128	0.002
Iroquois	0.021**	0.000	0.014*	0.304*	0.149	-0.014
Kankakee	0.027**	0.000	0.017**	0.394**	0.122	-0.067
Spoon	0.015**	0.000	0.013*	0.071	-0.040	0.006
La Moines	0.013**	0.000	0.010	0.169	0.091	0.018
Meramec	0.011*	0.000	0.007	0.276*	0.238	0.004
Bourbeuse	0.012*	0.000	0.009	0.185	0.137	-0.021
Big	0.011*	0.000	0.008	0.267*	0.228	0.030
Meramec	0.014**	0.000	0.010	0.299*	0.236	0.002
Average	0.016	0.000	0.011	0.306	0.203	0.011
Stdev	0.005	0.000	0.003	0.111	0.106	0.030
			Missouri Basin			
Yellowstone	0.017**	0.000	0.013*	0.266*	0.178	-0.077
Clarks Fork	0.008	0.000	0.007	0.157	0.135	-0.008
Yellowstone	0.016**	0.000	0.011	0.347*	0.271*	-0.038
Big Sioux	0.023**	0.000	0.014*	0.487**	0.348*	-0.045
North Platte	0.008	0.000	0.006	0.192	0.162	0.007
Bear	0.003	0.000	0.002	0.020	0.018	-0.002
Elkhorn	0.024**	0.000	0.013*	0.490**	0.335*	-0.136
Nisnabottna	0.024**	0.000	0.016**	0.280*	0.080	-0.041
Grand	0.013**	0.000	0.011	0.088	0.008	-0.002
Thompson	0.013**	0.000	0.011	0.132	0.057	-0.001
Gasconade	0.012*	0.000	0.008	0.302*	0.260*	-0.007
Average	0.015	0.000	0.010	0.251	0.168	-0.032
Stdev	0.007	0.000	0.004	0.152	0.122	0.043

Table 5: Trend and Persistence in 1-Day Low Flows

	Trend			Persistence		
	Obs.	DT	DM	Obs.	DT	DM
			Upper Mississippi Basin			
St. Croix	0.032**	0.000	0.015*	0.719**	0.514**	-0.008
Jump	0.029**	0.000	0.019**	0.533**	0.268*	0.011
Black	0.031**	0.000	0.020**	0.458**	0.109	-0.081
Maquaketa	0.017**	0.000	0.011	0.427**	0.340*	0.058
Mississippi	0.019**	0.000	0.010	0.495**	0.395**	0.002
Rock	0.027**	0.000	0.018**	0.347*	0.053	-0.052
Sugar	0.036**	0.000	0.013*	0.748**	0.446**	0.000
Pectonica	0.030**	0.000	0.012*	0.633**	0.411**	0.007
Cedar	0.020**	0.000	0.010	0.489**	0.387**	0.054
Skunk	0.017**	0.000	0.012	0.365**	0.271*	0.062
Mississippi	0.024**	0.000	0.014*	0.452**	0.273*	0.032
Des Moines	0.026**	0.000	0.013*	0.586**	0.426**	-0.012
Raccoon	0.022**	0.000	0.018**	0.172	-0.023	-0.009
Iroquois	0.023**	0.000	0.018**	0.255*	0.057	-0.003
Kankakee	0.020**	0.000	0.013*	0.333*	0.191	-0.057
Spoon	0.015**	0.000	0.012*	0.121	0.032	0.010
La Moines	0.007	0.000	0.006	0.193	0.169	0.038
Meramec	0.021**	0.000	0.008	0.634**	0.552**	-0.029
Bourbeuse	0.014**	0.000	0.006	0.476**	0.435**	-0.054
Big	0.020**	0.000	0.011	0.459**	0.357*	0.010
Meramec	0.017**	0.000	0.008	0.476**	0.402*	-0.044
Average	0.022	0.000	0.013	0.446	0.289	-0.003
Stdev	0.007	0.000	0.004	0.170	0.169	0.040
			Missouri Basin			
Yellowstone	0.008	0.000	0.004	0.151	0.133	-0.050
Clarks Fork	-0.010	0.000	-0.008	0.404**	0.375**	-0.008
Yellowstone	0.002	0.000	0.001	0.160	0.157	-0.005
Big Sioux	0.029**	0.000	0.011	0.698**	0.539**	-0.086
North Platte	0.017**	0.000	0.004	0.571**	0.517**	-0.030
Bear	0.022**	0.000	0.014*	0.414**	0.271*	-0.041
Elkhorn	0.028**	0.000	0.008	0.729**	0.603**	-0.099
Nisnabottna	0.026**	0.000	0.017**	0.354*	0.100	-0.034
Grand	0.015**	0.000	0.011	0.306*	0.233	0.050
Thompson	0.016**	0.000	0.013*	0.203	0.106	0.000
Gasconade	0.013**	0.000	0.004	0.569**	0.530**	-0.036
Average	0.015	0.000	0.007	0.414	0.324	-0.031
Stdev	0.012	0.000	0.007	0.205	0.195	0.041

Annual 1-Day High Flows

The flow sequences do not suggest a strong propensity toward either trend or persistence, particular in term of residual trend and residual persistence. Of the 32 sequences, half have significant trends, but only a third have significant residual trends after de-Markoving. Eight sequences have significant trends at the 1% level but for only one sequence is the trend significant following de-Markoving. De-trending fully accounts for trend

Only two sequences have significant persistence, but none of the sequences have significant residual persistence after de-trending. De-Markoving fully accounts for persistence.

See Figures 3 and 4, below.

Annual Mean Flows

The flow sequences suggest a strong propensity toward trend and persistence. Of the 32 sequences, 27 have significant trends, of which 22 have significant trends at the 1 % level. Following de-Markoving, only 13 sequences have significant residual trends, of which only 2 have significant trends at the 1% level.

Of the 32 sequences, 19 have significant persistence, of which only 7 have significant trends at the 1 % level. Following de-trending, 9 sequences have significant trends, but only 1 sequence has significant trend at the 1% level.

See Figures 5 and 6, below.

The averages of the measures of trend and persistence of the sequences of annual mean flows are about twice as large as the averages of the measures for the sequences of annual 1-day high flows. The standard deviations of the measures are nearly the same for the sequences of annual 1-day high flows and the sequences of annual mean flows. Refer to Tables 1 and 2.

Annual 1-Day Low Flows

The flow sequences show a very strong propensity toward trend and persistence. Of the 32 sequences, 28 have significant trends, all which are significant at the 1% level. Following de-Markoving, 15 sequences have significant trends, but only 6 sequences have significant residual trends at the 1% level.

Of the 32 sequences, 26 have significant persistence, and of these, 21 have significant persistence at the 1 % level. Following de-trending, 20 sequences have significant persistence, of which 13 have significant residual persistence.

See Figures 7 and 8, below.

The averages of the measures of trend in the sequences of annual 1-day low flows are about the same as the averages of the measures of trend in the sequences of annual mean flows, however, the standard deviations of the measures of trend in the annual 1-day low flow sequences is greater than that in the annual mean flow sequences. The annual 1-day low flow sequences yield somewhat larger averages and standard deviations of the measures of persistence. Refer to Tables 4 and 5.

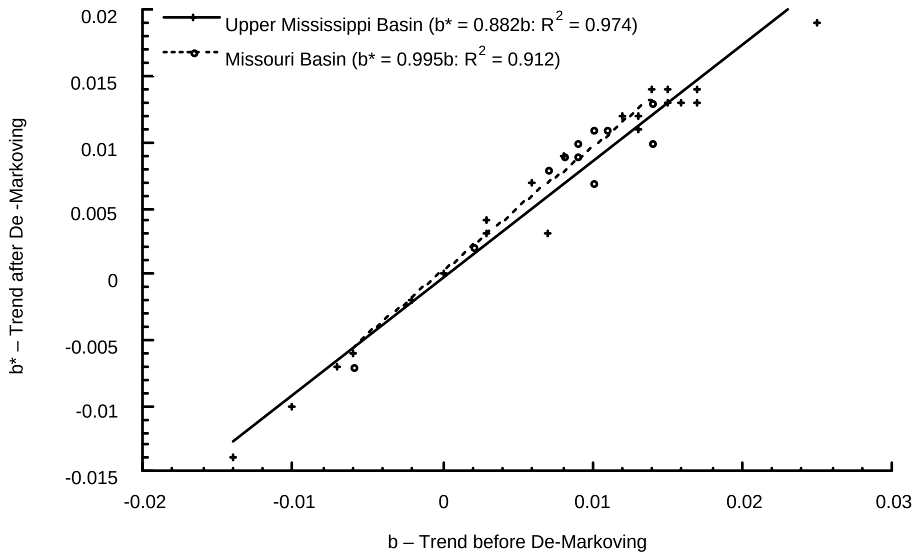

Figure 3: Effect of De-Markoving on Trend in 1-Day High Flows Sequences

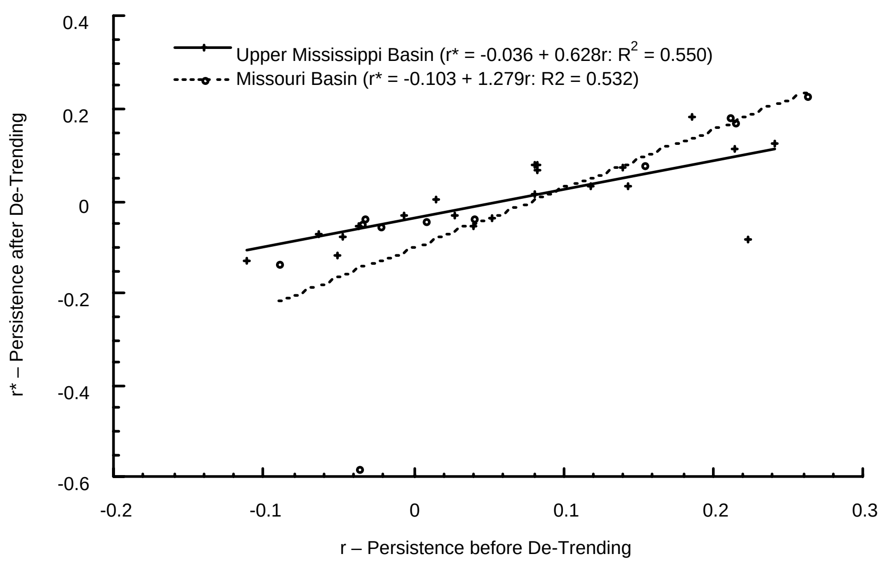

Figure 4: Effect of De-Trending on Persistence in 1-Day High Flow Sequences

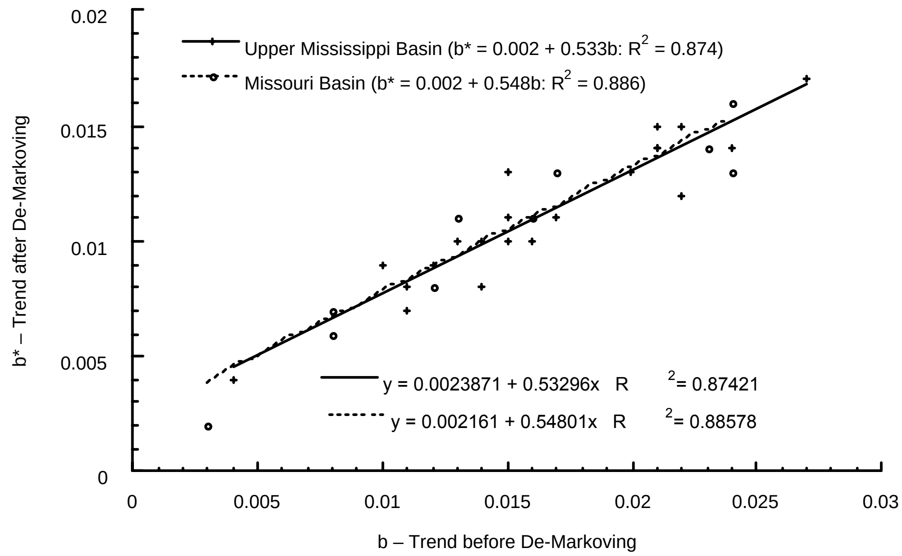

Figure 5: Effect of De-Markoving on Trend in Annual Mean Flow Sequences

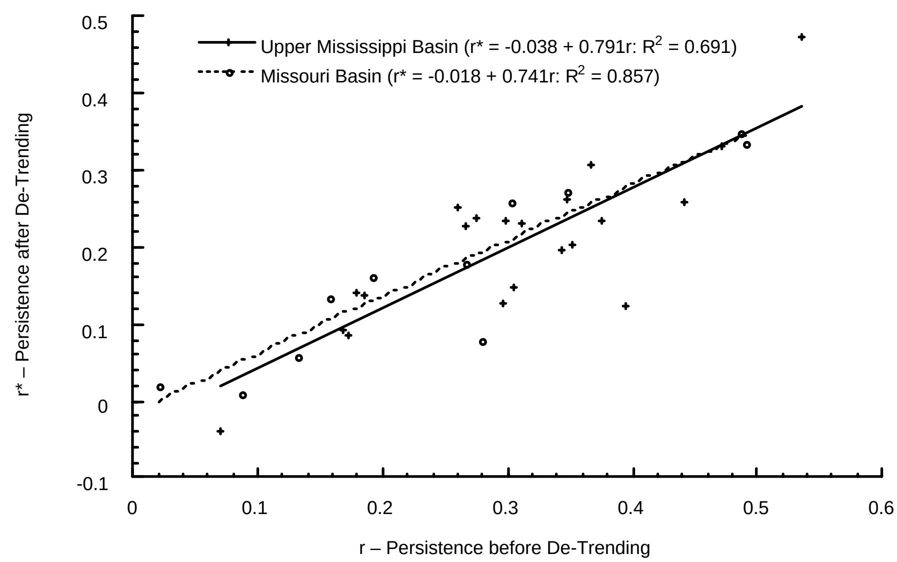

Figure 6: Effect of De-Trending on Persistence in Annual Mean Flow Sequences

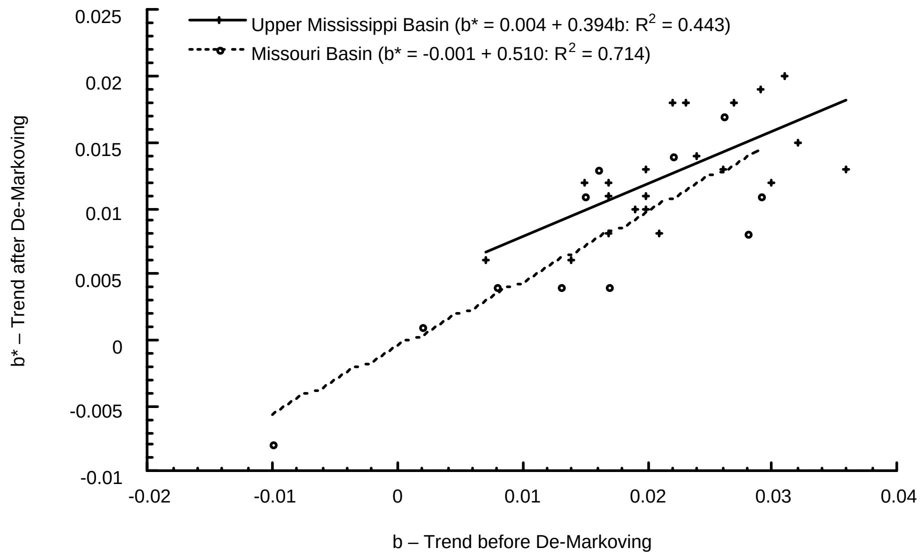

Figure 7: Effect of De-Markoving on Trend in 1-Day Low Flow Sequences

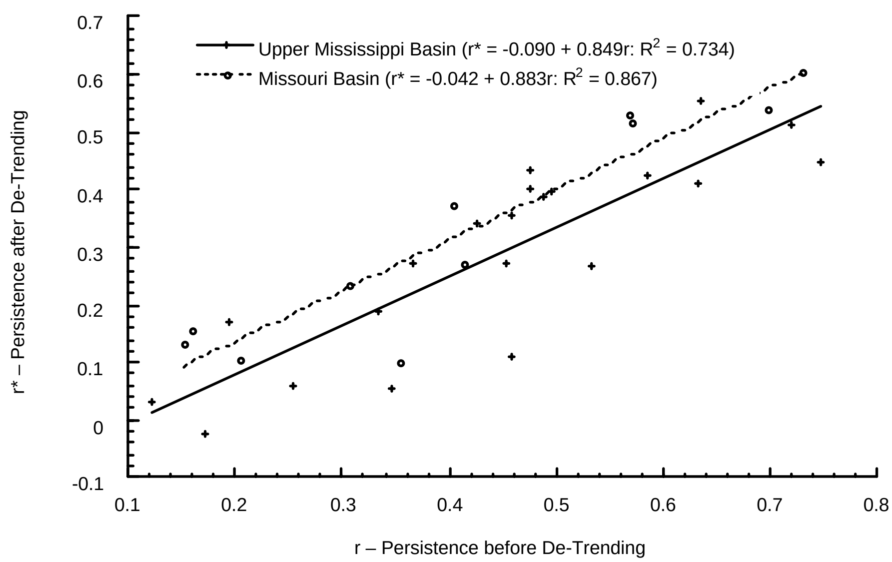

Figure 8: Effect of De-Trending on Persistence in 1-Day Low Flow Sequences

Other Elements of Flow Spectrum

Assessment of trend and persistence is summarily given in Appendix E. A numerical account of the shift in level of significance of trend and persistence following de-Markoving and de-trending is also presented in Appendix E, where account includes the shifts with respect to the sequences of annual 1-day high flows, annual mean flows and annual 1-day low flows.

The geographic distributions of the levels of significant of trends and persistence and the levels of significant residual trends and residual persistence are shown in Appendix E.

A Comparison of Effects of Trend and Persistence

Let h denote the elevation of water in a basin corresponding to a flood flow of magnitude Q, and let h^* denote a specific value of elevation. Define

$$x = \begin{cases} 1; \text{if } h \geq h^* \\ 0; \text{otherwise} \end{cases} \tag{9}$$

where $x = 1$ denotes a "flood" and $x = 0$ denotes "no flood". In addition to binomial outcomes, assume that floods are identically and independently distributed over time, whereby the flow process is Bernoullian. The probabilities of the outcomes are

$$Prob[x = 1] = p \tag{10}$$

$$\begin{aligned} Prob[x = 0] &= 1 - p \\ &= q \end{aligned} \tag{11}$$

The number of floods, y, that occur in a period of n years may vary from 0 to n and is distributed with expectation

$$\begin{aligned} E[y] &= \sum_{t=1}^{n} p \\ &= np \end{aligned} \tag{12}$$

and variance

$$\begin{aligned} V[y] &= \sum_{t=1}^{n} pq \\ &= npq \end{aligned} \tag{13}$$

See e.g. Johnson and Kotz (1969).

Trend

Assume that the probability of a flood changes over time due to various factors, notably due to changes in climate or changes in land use. Assume further that floods are temporally independent, i.e the flow process in non-Bernoullian – floods are non-identically and independently distributed over time. In year t, where $t = 1, 2, \ldots, n$, the probability of a flood is p_t, whereby $q_t = 1 - p_t$. In general, $p_1 \neq p_2 \neq \cdots \neq p_n$.

The number of floods in a period of n years is distributed with expectation

$$\begin{aligned} E'[y] &= E\left[\sum_{t=1}^{n} x_t\right] \\ &= \sum_{t=1}^{n} p_t \\ &= n\overline{p} \end{aligned} \tag{14}$$

and variance

$$\begin{aligned} V'[y] &= \sum_{t=1}^{n} p_t q_t \\ &= \sum_{t=1}^{n} p_t - \sum_{t=1}^{n} p_t^2 \\ &= n\overline{p} - \sum_{t=1}^{n} p_t^2 \end{aligned} \tag{15}$$

If $p_t = p \; \forall t$, then Eqs. (14) and (15) reduce to Eqs. (12) and 13).

The variance of the p_i, $V[p]$, is given by

$$\begin{aligned} V[p] &= \sum_{t=1}^{n} \left(p_t - \overline{p}\right)^2 \Big/ n \\ &= \sum_{t=1}^{n} p_t^2 / n - \left(\sum_{t=1}^{n} p_t\right)^2 \Big/ n^2 \\ &= \sum_{t=1}^{n} p_t^2 / n - \overline{p}^{\,2} \end{aligned} \tag{16}$$

whereby Eq. (15) may be expressed as

$$
\begin{aligned}
V'[y] &= n\bar{p} - n\left(V[p] + \bar{p}^2\right) \\
&= n\bar{p} - n\bar{p}^2 - nV[p] \\
&= n\bar{p}\bar{q} - nV[p] \\
&= V[y] - nV[p]
\end{aligned}
\quad (17)
$$

where $\bar{q} = 1 - \bar{p}$. If $p_t = p \ \forall t$, Eq.(17) reduces to Eq. (13). See e.g. Kenny and Keeping (1956) and Uspensky (1937).

The expected number of floods within a period of n years is the same whether the p_t vary with t or not. From Eq. (17), it is noted that $V'[y] < V[y]$ regardless of how the p_t vary with t. The variability of flooding within an n year period is less if the p_t vary with t than if p_t is a constant $\forall t$. In general, the manner in which the p_t vary with t effects the degree to which $V'[y] < V[y]$. If the slope of an upward linear trend in the p_t is equal to the absolute value of the slope of a downward linear trend, then the degrees to which $V'[y] < V[y]$ are the same

Persistence

Consider the non-Bernoullian flood process where floods are identically and non-independently distributed over time. It is assumed that temporal dependence, persistence, has the following Markovian structure formulated by Thomas (personal communication: 1957).

Eqs. (10) and (11) hold. The conditional, i.e. transition probabilities, the probabilities that particular outcomes will occur in one year given that a certain outcome occurred in the previous year are noted as

$$
\left.
\begin{aligned}
Prob\left[x_t = 1 \middle| x_{t-1} = 1\right] &= p(1|1) = (a+b) \\
Prob\left[x_t = 0 \middle| x_{t-1} = 1\right] &= p(0|1) = 1-(a+b) \\
Prob\left[x_t = 1 \middle| x_{t-1} = 0\right] &= p(1|0) = b \\
Prob\left[x_t = 0 \middle| x_{t-1} = 0\right] &= p(0|0) = 1-b
\end{aligned}
\right\}
\quad (18)
$$

The joint probabilities, i.e. the probabilities that particular outcomes will be manifest in consecutive years are

$$\left.\begin{aligned} p(x_{t-1}=1, x_t=1) &= p(1,1) = p(1|1)p(1) = (a+b)p \\ p(x_{t-1}=1, x_t=0) &= p(0,1) = p(0|1)p(1) = [1-(a+b)]p \\ p(x_{t-1}=0, x_t=1) &= p(1,0) = p(1|0)p(1) = bq \\ p(x_{t-1}=0, x_t=0) &= p(0,0) = p(0|0)p(1) = (1-b)q \end{aligned}\right\} \tag{19}$$

From the calculus of probability

$$p(0|1)p(1) = p(1|0)p(0) \tag{20}$$

whereby

$$b = (1-a)p \tag{21}$$

It follows that $\forall t$

$$E[x_t] = p \tag{22}$$

$$E[x_t^2] = p \tag{23}$$

and therefore

$$V[x_t] = pq \tag{24}$$

Furthermore, $\forall t$

$$E[x_t x_{t-1}] = (a+b)p \tag{25}$$

and therefore

$$\begin{aligned} Cov[x_t, x_{t-1}] &= E[x_t x_\tau] - E[x_t]E[x_\tau] \\ &= (a+b)p - p^2 \end{aligned} \tag{26}$$

whereby, the first order autocorrelation, ρ_1, is

$$\begin{aligned} \rho_1 &= \frac{Cov[x_t, x_{t-1}]}{\{V[x_t]\}^{1/2}\{V[x_{t-1}]\}^{1/2}} \\ &= \frac{apq}{pq} \\ &= a \end{aligned} \tag{27}$$

It can be shown that

$$\rho_k = a^k \tag{28}$$

further illustrating the Markovian structure of the non-Bernoullian process having the transition probabilities defined by Eq. (18).

The number of floods within a period of n years is distributed with expectation

$$\begin{aligned} E''[y] &= E\left[\sum_{t=1}^{n} x_t\right] \\ &= \sum_{t=1}^{n} E[x_t] \\ &= np \end{aligned} \tag{29}$$

and variance

$$\begin{aligned} V''[y] &= E\left[\sum_{t=1}^{n} x_t - np\right]^2 \\ &= E\left[\left(\sum_{t=1}^{n} x_t^2\right) - 2np\sum_{t=1}^{n} x_t + n^2p^2\right] \\ &= npq + 2\sum_{t=1}^{n-1}\sum_{\tau=t+1}^{n} E[x_t x_\tau] - n(n-1)p^2 \end{aligned} \tag{30}$$

The term $E[x_t x_\tau]$ may be expressed as

$$E[x_t x_\tau] = a^{\tau - t} pq + p^2 \tag{31}$$

whereby

$$\begin{aligned} 2\sum_{t=1}^{n-1}\sum_{\tau=t+1}^{n} E[x_t x_\tau] &= 2\sum_{t=1}^{n-1}\sum_{\tau=t+1}^{n}\left[a^{\tau-t}pq + p^2\right] \\ &= 2\left\{npq\sum_{t=1}^{n-1}a^t - pqa\sum_{t=1}^{n-1}ta^{t-1} + \sum_{t=1}^{n-1}\sum_{\tau=t+1}^{n}p^2\right\} \\ &= 2apq\left\{\frac{n(1-a)-(1-a^n)}{n(1-a)^2}\right\} + n(n-1)p^2 \end{aligned} \tag{32}$$

and therefore, Eq. (30) may be expressed as

$$\begin{aligned} V''[y] &= npq + 2apq\left\{\frac{n(1-a)-(1-a^n)}{(1-a)^2}\right\} \\ &= V[y]\left\{1 + 2a\left[\frac{n(1-a)-(1-a^n)}{n(1-a)^2}\right]\right\} \end{aligned} \tag{33}$$

The expected number of floods in the non-Bernoullian case marked by temporal dependence, is the same as in the strictly Bernoullian case or in the non-Bernoullian case marked by trend. It is readily noted that

$$V''[y]\begin{cases} > V[y];\ if\ a > 0 \\ = V[y];\ if\ a = 0 \\ < V[y];\ if\ a < 0 \end{cases} \tag{34}$$

Trend vs. Persistence

Trend and persistence have no effect on the expected number of floods above a threshold level of elevation. The effects of Markovian persistence with $a > 0$ and of trend on the variability of the number of floods above a threshold level of elevation are counter to one another – one tending to mitigate the effect of the other. The net effect of trend and Markovian persistence with $a > 0$ on the variability of the number of floods above a threshold level of elevation may be negligible if the levels of trend and persistence are low. Thus, variability in the number of floods above a threshold level of elevation in keeping with a Bernoullian structure of floods does not necessarily imply the absence of trend and persistence.

The effect of Markovian persistence with $a < 0$ compounds the effect of trend – both reduce the variability of the number of floods above a threshold level of elevation without either effecting the expected number of floods. Thus, variability in the number of floods above a threshold level of elevation smaller than expected via a Bernoullian process is not necessarily the consequence of trend alone or persistence alone.

The following examples illustrates the effects of trend and persistence on the variability in the number of floods above a threshold level of elevation.

Bernoullian Case

In the Bernoullian case, floods are identically and independently distributed – the probability that a flood will exceed a threshold elevation does not vary from one year to the next, and the fact that the threshold had or had not been exceeded in one year has no bearing on whether the threshold will or will not be exceeded in any subsequent year. The expected values and variances of the distribution of the number of floods exceeding the threshold in a period of $n = 5, 10, 25, 50, 100$ given the exceedence probabilities $p = 0.04, 0.02, 0.01$ are shown in Table 6. The corresponding probabilities of non-exceedence are $q = 0.96, 0.98, 0.99$. The specific exceedence probabilities translate as the $T = 25, 50, 100$ year floods.

Table 6: Expected Value and Variance of the Distribution of the Number of Floods exceeding a Threshold Elevation – Bernoulli Process

n	p=0.04		p=0.02		p=0.01	
	$E[y]$	$V[y]$	$E[y]$	$V[y]$	$E[y]$	$V[y]$
5	0.20	0.1920	0.10	0.0980	0.05	0.0495
10	0.40	0.3840	0.20	0.1960	0.10	0.0990
25	1	0.9600	0.50	0.4900	0.25	0.2475
50	2	1.9200	1	0.9800	0.50	0.4950
100	4	3.8400	2	1.9600	1	0.9900

Non-Bernoullian Case Marked by Trend

In the non-Bernoullian case marked by trend, floods are non-identically and independently distributed – the probability that a flood will exceed a threshold elevation varies from one year to the next, but the fact that the threshold had or had not been exceeded in one year has no bearing on whether the threshold will or will not be exceeded in a subsequent year.

For $t = 1, \ldots, n$,

$$\begin{aligned} \bar{t} &= \sum_{t=1}^{n} t/n \\ &= (n+1)/2 \end{aligned} \tag{35}$$

Given p_1 and p_n,

$$\bar{p} = (p_n + p_1)/2 \tag{36}$$

$$b = (p_n - p_1)/n \tag{37}$$

whereby

$$p_t = \bar{p} + b(t - \bar{t}) \tag{38}$$

The parameter b is the slope of the linear trend in the p_t. The effect of a negative linear trend in the p_t on the expected value and variance of the number of times floods exceed the threshold elevation in a period of n years is the same as in the case of a positive trend where the absolute values of the positive and negative slopes are equal.

The values of $\bar{t}$, $\bar{p}$ and b for specific values of n, p_1 and p_n are given in Table 7.

Table 7: Parameters Describing the Variation of p_t with t

n	*5*	*10*	*25*	*50*	*100*
			$p_1 = 0.005, p_n = 0.075$		
$\bar{t}$	3	5.5	13	25.5	50.5
$\bar{p}$	0.04	0.04	0.04	0.04	0.04
b	0.0175	0.0078	0.0029	0.0014	0.007
			$p_1 = 0.005, p_n = 0.035$		
$\bar{t}$	3	5.5	13	25.5	50.5
$\bar{p}$	0.02	0.02	0.02	0.02	0.02
b	0.0075	0.0033	0.0012	0.0006	0.0003
			$p_1 = 0.005, p_n = 0.015$		
$\bar{t}$	3	5.5	13	25.5	50.5
$\bar{p}$	0.01	0.01	0.01	0.01	0.01
b	0.0025	0.0011	0.0004	0.0002	0.0001

The expected values and variances of the distribution of the number of floods exceeding the threshold in a period of $n = 5, 10, 25, 50, 100$ for the mean exceedence probabilities $\bar{p} = 0.04, 0.02, 0.01$ are shown in Table 8.

Table 8: Expected Value and Variance of the Distribution of the Number of Floods exceeding a Threshold Elevation as Affected by Trend

n	$\bar{p} = 0.04$		$\bar{p} = 0.02$		$\bar{p} = 0.01$	
	$E'[y]$	$V'[y]$	$E'[y]$	$V'[y]$	$E'[y]$	$V'[y]$
5	0.2	0.1889	0.1	0.0974	0.05	0.0494
10	0.4	0.3790	0.2	0.1951	0.1	0.0989
25	1	0.9489	0.5	0.4880	0.25	0.2473
50	2	1.8988	1	0.9761	0.5	0.4946
100	4	3.7976	2	1.9523	1	0.9891

From Eqs. (12) and (14), it is noted that the expected value of the distribution of the number of floods that exceed the threshold elevation within a period of n years in the non-Bernoullian case marked by trend is the same as that in the Bernoullian case – $E'[y] = E[y]$. For a given value of n, $E[y]$ and $E'[y]$ decrease as $p = \overline{p}$ decreases. For a given value of $p = \overline{p}$, $E[y]$ and $E'[y]$ increase as n increases. Refer to Tables 7 and 8.

From Eqs. (13) and (17), it is noted that the variance of the distribution of the number of floods that exceed the threshold elevation within a period of n years in the non-Bernoullian case marked by trend is less than that in the Bernoullian case – $V'[y] < V[y]$. For a given value of n, $V[y]$ and $V'[y]$ decrease as $p = \overline{p}$ decreases. For a given value of $p = \overline{p}$, $V[y]$ and $V'[y]$ increase as n increases. Moreover, for a given value of n, the difference between $V[y]$ and $V'[y]$ decreases at an increasing rate as $p = \overline{p}$ decreases. In the case where $p = \overline{p}$ is very small, the effect of trend on the variance of the distribution of the number of floods that exceed the threshold elevation within a period of n years is small. Refer to Tables 7 and 8.

Non-Bernoullian Case Marked by Markovian Persistence

In the non-Bernoullian case marked by persistence, floods are identically and non-independently distributed – the probability that a flood will exceed a threshold elevation does not vary from one year to the next, but the fact that the threshold had or had not been exceeded in one year has a bearing on whether the threshold will or will not be exceeded in any subsequent year.

For specific values of n, p and a, the expected value and variance of the distribution of the number of floods that exceed a threshold elevation are given in Tables 9a and 9b.

Table 9a: Expected Value and Variance of the Distribution of the Number of Floods exceeding a Threshold Elevation as Affected by Persistence – $a = 0.1$

n	p=0.04		p=0.02		p=0.01	
	$E''[y]$	$V''[y]$	$E''[y]$	$V''[y]$	$E''[y]$	$V''[y]$
5	0.2	0.2252	0.1	0.1149	0.05	0.0581
10	0.4	0.4599	0.2	0.2347	0.1	0.1186
25	1	1.1639	0.5	0.5940	0.25	0.3001
50	2	2.3372	1	1.1929	0.5	0.6026
100	4	4.6839	2	2.3907	1	1.2076

Table 9b: Expected Value and Variance of the Distribution of the Number of Floods exceeding a Threshold Elevation as Affected by Persistence – $a = -0.1$

n	p=0.04		p=0.02		p=0.01	
	$E''[y]$	$V''[y]$	$E''[y]$	$V''[y]$	$E''[y]$	$V''[y]$
5	0.2	0.1634	0.1	0.0834	0.05	0.0421
10	0.4	0.3205	0.2	0.1636	0.1	0.0826
25	1	0.7918	0.5	0.4041	0.25	0.2041
50	2	1.5773	1	0.8051	0.5	0.4066
100	4	3.1482	2	1.6069	1	0.8116

From Tables 9a and 9b, it readily seen that persistence effects an increase (decrease) in the variance of the distribution of the number of floods that exceed the threshold elevation within a period of n years if $a > 0$ $(a < 0)$.

Right and Left Tails of Flood Distributions

Overview

In conducting flood frequency analyses, Federal agencies are guided by Bulletin 17-B The Bulletin calls for fitting the Log-Pearson Type III distributions to ordered sequences of observed flood flows, unless there are strong grounds for fitting another distribution. The Log-Pearson Type III distribution presumes that the logs of the observed flows are distributed as Pearson Type III. The Pearson Type III distribution may be positively or negatively skewed. As skewness approaches zero, the Pearson Type III distribution approaches the Normal distribution. Thus, the Log-Normal distribution is a special case of the Log-Pearson Type III distribution.

The logs of observed flood flows tend to yield negative values of skewness, whereby, the Normal distribution provides a relatively poor fit to the logs of the flows, and consequently the Log-Normal distribution provides a relatively poor fit to the flows. The Pearson Type III distribution in log space and the Log-Pearson Type III distribution in real space provide better fits. It is noted that the Normal distribution is defined by two parameters, whereas the Pearson type III distribution is defined by two or at most three parameters, one of which measures the skewness of the distribution.

Evaluation of the goodness of fit of a particular distribution is generally in reference to the overall fit of the distribution, i.e. the fit over the full range of an ordered set of flows. An analytically defined probability distribution may be fitted to a set of observations by one of several statistical methods. In hydrology, the most common procedures are those of moments, maximum likelihood, and L-moments. These methods are based on relations between the parameters of the distribution to be fitted and specific statistical averages derived from integration over the range of the distribution. For example, in fitting a Normal distribution, the integration is over the interval $(-\infty, \infty)$. In fitting a Log-Normal, the interval of integration is (m, ∞), where the lower bound $m \geq 0$ is physically interpreted as base flow. Bulletin 17-B is based on the method of moments.

The relative poor fit of the Normal distribution in log space, judged in reference to the overall fit, does not preclude the right tail of the Normal distribution, or of some other distribution, providing a relatively better fit to the distribution of the larger of the observed flows than the right tail of the Pearson Type III distribution. It is the right tail of the distribution that matters in flood frequency studies. In drought studies, the left tail matters. Few studies consider fitting one tail independently of the other.

To improve the overall fit of the Log-Pearson Type III distribution, attention is given in Bulletin 17-B to estimating a regional coefficient of skewness applicable to the site at issue. The estimate of skewness is regarded as a principal source of uncertainty in evaluating flood risks at specific sites. The extent to which the left and right tails of the presumed distribution of floods contribute to the skewness of the distribution is open to investigation. Seeking to fit the right tail independently of the left tail affords a basis for assessing the contributions of the tails to the skewness of the distribution. If only the right tail is to be fitted, then the estimate of skewness of the complete distribution, i.e. the distribution extending over both tails may or may not be relevant. The degree relevance is an open question.

The following discussions of fitting the right tail of a distributions to an ordered set of flows are in reference to two cases. The first case considers the regional vector derived from the logs of the flows at specific sites within a region, and the second case considers vector formed by the logs of the elements of the regional vector derived from the flows at the specific sites within the region. For the first case, the distribution of the elements of the regional vector is said to be in log space, whereas, for the second case, the distribution of the logs of the elements of the regional vector is said to be in quasi-log space.

For the purpose of this study, the left tail extends over flows less than the median, and the right tail extends over flows equal to or greater than the median. The median flow is the $T = 2 - year$ flow, i.e. the flow having an exceedence probability of $1 - P = 0.5$, where $P \equiv Prob(X < x)$. Further study of fitting the right tail independently of the left tail perhaps should consider a three part partition of a distribution to include a central part. With such a partition, the right tail of a distribution would extend over all flows greater than the $T > 2 - year$. flood.

Left and Right Tails

From Appendix B, it is noted that in log space, the right tails of the distributions of the elements of the high end of the flow spectrum, i.e. the $k-day$ high flows, tend to vary linearly over the probability scale. The linear tendency suggests that the right tails accord reasonably well with the Normal distribution in log space, and consequently with the Log Normal distribution in real space. In log space, the left tails of the distributions reflect varying degrees of stretchiness relative to the left tail of the Normal distribution.

The following terminology is introduced. If the right tail of a regionalized distribution is well approximated by the right tail of a Normal distribution, the right tail of a regionalized distribution is said to be right tail Normal. If the left tail of the regionalized distribution lies below (above) the left tail of the Normal distribution, the left tail of the regionalized distribution is said to be left tail super- (sub-) Normal. A left tail that is super- (sub-) Normal is in effect a tail that is stretched (compressed) relative to a Normal left tail. If in fact, the left tail of a regionalized distribution is left tail Normal, then the right tail of the regionalized distribution is said to be sub- (super-) Normal if the right tail of the regionalized distribution lies below (above) the right tail of a Normal distribution.

For exploratory purposes, the Normal distribution was fitted to the right tail of the regionalized distributions of $k-day$ high flows as follows. The Normal distribution is defined by two parameters, namely the mean and standard deviation. Given that the mean and median are the same for the Normal distribution, the mean was set equal to the median, which for the regionalized distribution in log space is equal to 1 and in quasi-log space is equal to 0. The standard deviation of the Normal distribution is equal to the absolute difference in the values of the normalized logs of flow for exceedence probabilities of 0.5 and 0.84. The probability 0.5 marks the median of the distribution, and the probability 0.84 marks one of the two inflection points of the Normal distribution. The other inflection point is marked by the probability 0.16.

Let $\{x_i: i = 1, \ldots, n\}$ denote a regional vector in either log or quasi-log space. By the manner – the median-median method – in which the vector was obtained, the vector is an ordered set of values, from smallest, x_1, to largest, x_n. The probability distribution of the elements of the regional vector, $F(x) = Prob(x < X)$ is defined in terms of the Weibull plotting position —

$$F(x_i) = i/(n+1) \tag{39}$$

Let $F(y)$ denote the Normal cumulative distribution function: $y \sim N(\mu, \sigma)$, where μ, mean=median, and σ, standard deviation, are given by the procedure outlined above. The Normal variate y_i corresponding to x_i is that value for which

$$F(y_i) = F(x_i) \tag{40}$$

The right tail of the Normal distribution is given by $F(y_i)$ for $i = \nu, \ldots, n$, where

$$\nu = \begin{cases} \dfrac{n+1}{2}; \; n \;\; odd \\ \dfrac{n}{2}; \; n \;\; even \end{cases} \tag{41}$$

Normal distributions fitted to the right tails of regionalized distributions in log space of the $1-day$ high flows for the Upper Mississippi and Missouri basins are shown in Figures 9 and 10, respectively.

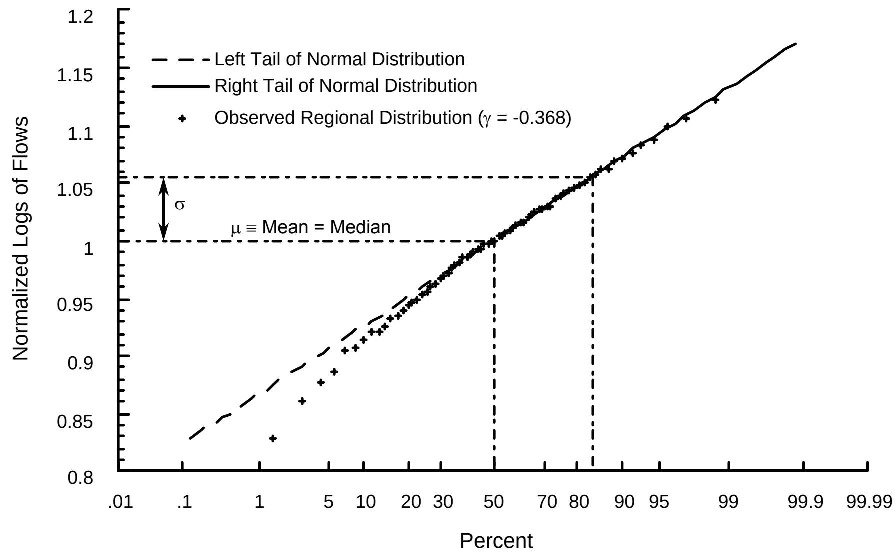

Figure 9: Upper Mississippi Basin - Regionalized Distribution in Log Space of Annual 1-Day High Flows with Right Tail Fitted with a Normal Distribution (μ = 1, σ = 0.057)

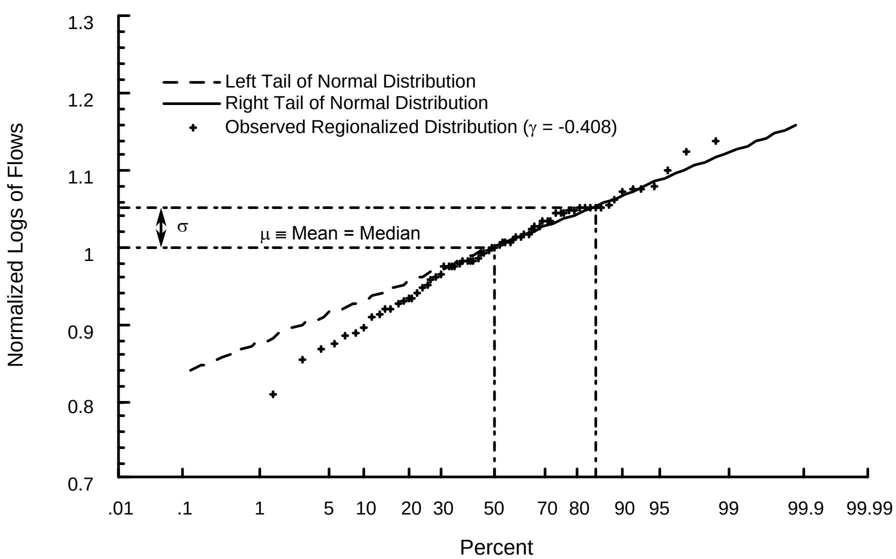

Figure 10: Missouri Basin - Regionalized Distribution in Log Space of Annual 1-Day High Flows with Right Tail Fitted with a Normal Distribution (μ = 1, σ = 0.053)

From Figures 9 and 10, it is seen that 1) the Normal distribution provides a reasonably good fit to the right tail of the observed regional distribution in log space of the $k = 1 - Day$ annual high flows. The left tail of the observed distribution is sub-Normal, and the observed distribution is negatively skewed. These features of the observed regional distribution in log space of the $k = 1 - Day$ annual high flows relative to the Normal distribution are displayed by the observed regional distributions in log space of the $k - Day$ annual high flows for $k > 1$, as well as by the observed regional distributions in log space of the annual mean flows. See Appendix F.

The regional distributions in quasi log space of the $k = 1 - Day$ annual high flows are shown in Figure 11 for the Upper Mississippi basin and in Figure 12 for the Missouri basin.

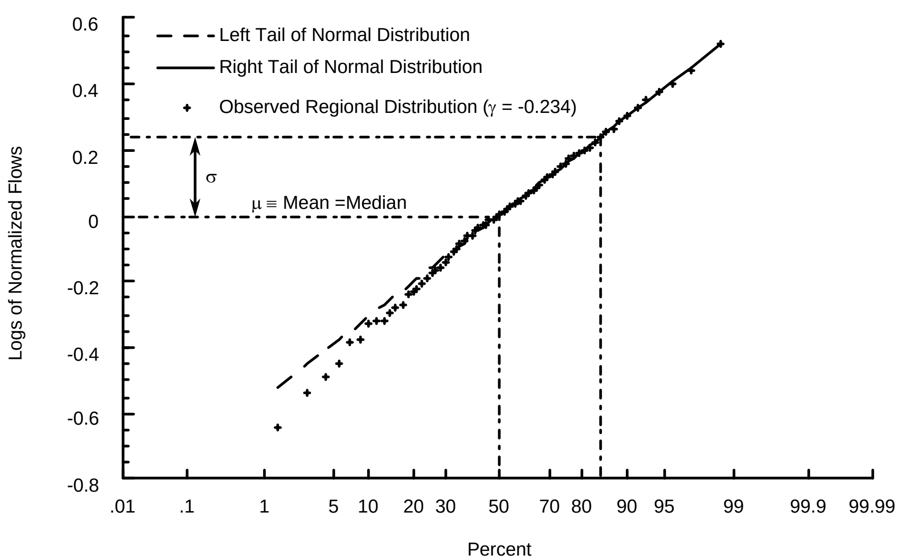

Figure 11: Upper Mississippi Basin – Regionalized Distrinution in Quasi-Log Space of Annual 1-Day High Flows with Right Tail Fitted with a Normal Distribution (μ = 0, σ = 0.236

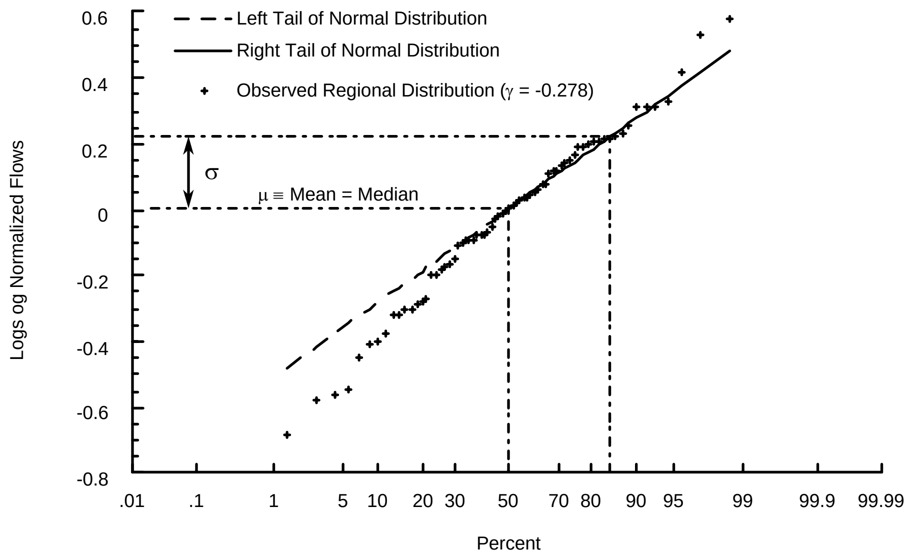

Figure 12: Missouri Basin – Regionalized Distribution in Quasi-Log Space of 1-Day Annual High Flow with Right Tail Fitted with a Normal Distribution (μ = 0, σ = 0.218)

The features of the regional distribution of the annual $k = 1-Day$ high flows displayed in log space are displayed in quasi-log space. These features are also displayed in quasi-log space by the annual $k > 1-Day$ high flows and by the annual mean flows. See Appendix G.

A summary account of the Normal distributions fitted to the right tails of the distributions of the regional vectors in log space and in quasi-log space is given in Table 10.

Table 10: Standard Deviations of Normal Distributions Fitted to Right Tails of Regional Vectors

k	Upper Mississippi Basin		Missouri Basin	
	Log Space	Quasi-Log Space	Log Space	Quasi-Log Space
1-Day	0.057	0.236	0.053	0.218
3-Day	0.057	0.224	0.052	0.228
7-Day	0.057	0.210	0.053	0.221
14-Day	0.054	0.210	0.057	0.221
30-Day	0.054	0.207	0.063	0.227
60-Day	0.053	0.182	0.083	0.284
90-Day	0.056	0.174	0.081	0.274
180-Day	0.053	0.157	0.079	0.260
AM	0.052	0.172	0.083	0.249

Over the high end of the flow spectrum, the Normal distribution provides a better fit of the right tails of the observed regional distributions for the Upper Mississippi basin than for the Missouri basin. It is noted that 21 sites form the region for the Upper Mississippi basin, whereas 11 sites form the region for the Missouri basin. The Missouri region is more heterogeneous in terms of climate than the Upper Mississippi region. Whether these attributes of the two regions have a bearing on the goodness of fit of the right tails of the observed regional distributions in log space and in quasi-log space by Normal distributions is an open question.

For further discussion of the right and left tails of flood distributions see Appendix H.

Comparison of Pearson Type III and Right-Tail Normal Distributions – Annual 1-Day High Flows

In the above discussions, it was shown that in log space and in quasi-log space the Normal distribution provides a good fit to the right tails of regionalized distributions spanning the "high end" of the flow spectrum, that is for the regionalized sequences of $k-day$ high flows, where $1 \leq k \leq 365$. The two spaces are in reference to two cases. The first case refers the regional vector derived from the logs of the flows at specific sites within a region, and the second case refers vectors formed by the logs of the elements of the regional vector derived from the flows at the specific sites within the region. For the first case, the distribution of the elements of the regional vector is said to be in log space, whereas, for the second case, the distribution of the logs of the elements of the regional vector is said to be in quasi-log space.

In flood studies, it is the right tail of the probability distribution that matters. By an overall fit of a distribution, i.e. the fit over all the observations, the left tail wags the right tail, so to speak. The motivation for a right tail fit is to diminish, if not eliminate the impact of the left tail on the right tail through the overall fitting procedure. If the smaller floods are censored, than the distribution is a truncated distribution with appropriate adjustment in the exceedence probabilities of the uncensored floods. The methodology of right tail fitting of the Normal distribution outlined above does not require any censoring and thus no adjustment in the exceedence probabilities is needed.

How well the right tail fit of the Normal distribution compares with the overall fit of the Pearson Type III is examined. The comparison carries over to real space in terms of the right tail fit of the Log-Normal Distribution relative to the overall fit of the Log-Pearson Type III distribution.

The comparison of the Right-Tail Normal with the overall Pearson Type III is made in terms of the distributions of the elements of the regional vectors in log space and in quasi-log space, and in terms of the distributions of the logs of the annual 1-day high flows at each of the 32 sites, where 21 sites are in the Upper Mississippi basin and the other 11 sites are in the Missouri basin.

The sequences of annual 1-day high flows at each of the 32 sites in the two regions span the 70-year period, 1929-1998. This period is the longest concurrent period of the sequences. The comparison of the Right-Tail Normal distribution with the overall Pearson Type III distribution is in terms of the 50-year flood, i.e. the 50-year, annual 1-day high flow. The 50-year flow has an exceedence probability of $P = (1/50) = 0.02$.

Rationale for Right-Tail Distribution

Previously it was noted that in the case of an overall fit of a distribution, the fit of the right tail is affected by the fit of the left tail, and conversely. In flood frequency analysis it is the left tail that matters, and therefore a good fit of the right at the expense of the goodness of fit of the left tail merits attention.

Another reason for improving the fit of the right tail at the expense of the left tail is as follows. In a given locale, assume that the damages, $D(x)$, that would be incurred from a flood of magnitude x are defined as

$$D(x) \begin{cases} = 0 \text{ if } m \le x \le x^* \\ > 0 \text{ if } x > x^* \end{cases} \tag{42}$$

where x^* denotes the threshold flood, a value below which no damages are incurred, and m denotes the lower bound on flood magnitude. The expected value of damages is given by

$$\begin{aligned} E[D(x)] &= \int_m^{x^*} D(x) f(x) dx + \int_{x^*}^{\infty} D(x) f(x) dx \\ &= 0 + \int_{x^*}^{\infty} D(x) f(x) dx \\ &= \int_{x^*}^{\infty} D(x) f(x) dx \end{aligned} \tag{43}$$

where $f(x)$ denotes the probability density function of floods which is not known. A specific density function is accepted $g(x)$ as the density function that best represents $f(x)$. Using $g(x)$, the estimate of the expected value of the damages is

$$\hat{E}[D(x)] = \int_{x^*}^{\infty} D(x)g(x)dx \qquad (44)$$

No matter what density function is used to represent $f(x)$, floods of magnitude less than the threshold value contribute nothing to the estimated expected value of damages.

In the above discussions, empirical evidence was offered that in log space, the Right-Tail Normal distribution provides better estimates of the $T-year$ flood than does the Pearson Type III distribution for $T>2$, i.e. for floods greater than the median flood. If in log space the Right-Tail Normal distribution out-performs the Pearson Type II distribution, then in real space, the Log-Normal distribution will out-perform the Log-Pearson Type III distribution. It is noted that the left tail of the Right-Tail Normal distribution does not provide as good a fit to the observations as does the left tail of the Pearson Type III distribution.

Let $g(x)$ be the Pearson Type III density function, and let $h(x)$ be the Right-Tail Normal density function. Assume that $x^* > \tilde{x}$, i.e. the threshold flood exceeds the median flood. No matter how poorly the left tail of the Right-Tail Normal distribution fits the observations in comparison to the goodness of fit of the Pearson Type III distribution, the left tail of neither distribution has any effect on the estimated expected value of damages. The difference between estimates of the expected value of damages conditioned on $g(x)$ and $h(x)$ is small, thought the Right-Tail Normal distribution better fits the observed floods greater than the median flood than does the Pearson Type III distribution.

Flow Sequences

Of the 32 sequences of annual 1-day high flows spanning the 70-year period, 1929-1998, 21 are in the Upper Mississippi basin and 11 are in the Missouri basin. The geographic locations of the sequences along with statistical descriptors of the sequences in log space, namely, the mean, μ, the standard deviation, σ, and the coefficient of skewness, γ, are given in Table 11. Because the matter of differences between sampling and population values of statistical descriptors is not dealt with, the sample statistics are denoted in terms of the Greek letters that are generally reserved for the corresponding population statistics.

Table 11: Location and Statistical Description of Sequences of Annual 1-Day High Flows

Stream	Locale	State	Flow Descriptors μ	σ	γ
		Upper Mississippi Basin			
St. Croix	St. Croix	WI	4.374	0.185	-0.638
Jump	Sheldon	WI	3.870	0.233	-0.062
Black	Neillsville	WI	4.018	0.251	-0.644
Maquaketa	Maquaketa	IA	4.045	0.277	-0.147
Mississippi	Clinton	IA	5.126	0.165	-0.608
Rock	Afton	WI	3.755	0.192	-0.598
Sugar	Broadhead	WI	3.428	0.285	-0.093
Pecatonica	Freeport	IL	3.706	0.241	0.144
Cedar	Cedar Rapids	IA	4.335	0.320	-0.640
Skunk	Augusta	IA	4.271	0.268	-1.013
Mississippi	Keokuk	IA	5.256	0.164	-0.717
De Moines	Stratford	IA	4.132	0.315	-0.741
Raccoon	Van Meter	IA	4.092	0.304	-0.391
Iroquois	Chebanse	IL	4.091	0.208	-0.548
Kankakee	Momence	IL	3.806	0.154	-0.583
Spoon	Seville	IL	4.067	0.226	-0.272
La Moines	Ripely	IL	3.955	0.288	-0.871
Meramec	Steeville	MO	4.030	0.354	-0.940
Bourbeuse	Union	MO	4.108	0.245	0.024
Big	Byrnesville	MO	4.154	0.288	-0.552
Meramec	Eureka	MO	4.550	0.279	-0.378
Average			*4.151*	*0.250*	*-0.489*
Std. Deviation			*0.424*	*0.056*	*0.319*
		Missouri Basin			
Yellowstone	Corwin Springs	MT	4.203	0.118	-0.307
Clarks Fork	Belfry	MT	3.844	0.104	-0.046
Yellowstone	Billings	MT	4.583	0.135	-0.409
Big Sioux	Akron	IA	3.949	0.434	-0.318
North Platte	Northgate	CO	3.391	0.220	-0.628
Bear	Morrison	CO	2.367	0.341	0.437
Elkhorn	Waterloo	NE	4.063	0.351	-0.022
Nishabottna	Hamburg	IA	4.096	0.324	-1.273
Grand	Gallatin	MO	4.337	0.261	-0.479
Thompson	Trenton	MO	4.247	0.291	-0.285
Gasconade	Jerome	MO	4.446	0.295	-0.477
Average			*3.957*	*0.261*	*-0.346*
Std. Deviation			*0.617*	*0.106*	*0.423*

All the coefficients of skewness are negative except three, two in the Upper Mississippi basin (Pecatonica at Freeport, IL and Bourbeuse at Union, MO), and one in the Missouri basin (Bear at Morrison, CO). The negative skews range from -1.013 to 0.024 with an average of -0.434 in the Upper Mississippi basin, and from -1.273 to 0.434 with an average of -0.489 in the Missouri Basin.

On a regional basis, the statistical descriptors of the distributions of the annual 1-Day high flows in log space and in quasi-log space are given in Table 12.

Table 12: Ststistical Descriptors of Regional Sequences of Annual 1-Day High Flows

	Statistical Descriptors		
	μ	σ	γ
Upper Mississippi Basin			
Log Space	1	0.060	-0.368
Quasi-Log Space	0	0.244	-0.234
Missouri Basin			
Log Space	1	0.065	-0.408
Quasi-Log Space	0	0.262	-0.278

It is noted from Table 11, that the values of the coefficients of skewness in log space exceed the values in quasi-log space. The values of the standard deviation in log space are an order of magnitude smaller than the values in quasi-log space. In log space the mean is nearly equal to the median of unit value, whereas in quasi-log space, the mean is nearly equal to the median of zero value. It should be noted, that the regionalization process that was used follows the median-median procedure, where in log space, the median = 1 for the regional sequence, and in quasi-log space, the median = 0 for the regional sequence.

Pearson Type III Distribution

The fitting procedure outlined in Bulletin 17- B for the Pearson Type III distribution is basically the method of moments. In lieu of the at-site estimate of the skewness, a regional estimate of skewness is used in fitting the distribution to the observed distribution where the ordered floods are assigned exceedence probabilities defined by the Weibull plotting position. For a sequence of n observations ordered from smallest to largest, the $i-th$ observation, where $i = 1, \ldots, n$, is assigned the Weibull plotting position, $i/(n+1)$. The associated exceedence probability is given by

$$\begin{aligned} P &= 1 - i/(n+1) \\ &= (n+1-i)/(n+1) \end{aligned} \tag{45}$$

In the following discussions, the fitting of the Pearson Type III distribution is based on the at-site estimates or on the regionalized estimates of skewness. No account is taken of regionalizing the at-site estimates of skewness or of regionalizing the regional estimates of skewness in terms of Bulletin 17-B. The regionalized estimates of skewness are, in effect, regional estimates of skewness.

The Pearson Type III distribution is defined as

$$f(x) = \frac{1}{|b|\Gamma(a)} \left(\frac{x-m}{b} \right)^{a-1} exp\left[-\left(\frac{x-m}{b} \right) \right] \tag{46}$$

If b is positive, then x is positively skewed, where $m \le x \le \infty$. If b is negative, then x is negatively skewed, where $-\infty \le x \le m$.

The $\kappa - th$ moment about the origin of x is

$$E\left[x^{\kappa}\right] = \begin{cases} \int_{m}^{\infty} x^{\kappa} f(x) dx;\ if\ a > 0 \\ \int_{-\infty}^{m} x^{\kappa} f(x) dx;\ if\ a < 0 \end{cases} \tag{47}$$

It follows that

$$\left.\begin{aligned} \mu &= m + ab \\ \sigma &= |b| a^{1/2} \\ \gamma &= \frac{2b}{|b| a^{1/2}} \end{aligned}\right\} \tag{48}$$

whereby,

$$\left.\begin{aligned} m &= \mu - \sigma/\gamma \\ a &= 4/\gamma^2 \\ b &= \sigma\gamma/2 \end{aligned}\right\} \tag{49}$$

Let $\{x_i : i = 1, \ldots, n\}$ denote an arbitrary sequence. The values of μ, σ and γ, actually the estimates of the values μ, σ and γ, are given by

$$\left.\begin{aligned} \mu &= \sum_{i=1}^{n} x_i / n \\ \sigma &= \left\{ \sum_{i=1}^{n} (x - \mu)^2 / n \right\}^{1/2} \\ \gamma &= \left\{ \sum_{i=1}^{n} (x - \mu)^3 / n \right\} / \sigma^3 \end{aligned}\right\} \tag{50}$$

Annual 1-Day High Flows

The values of m, a and b are given in Table 13 for the at-site sequences of annual 1-day high flows in log-space.

Table 13: Pearson Type III Parameter for At-Site Sequences of Annual 1-Day High Flows in Log Space

	m	*a*	*b*
		Upper Mississippi Basin	
St. Croix	4.958	9.989	-0.058
Jump	11.131	1055.841	-0.007
Black	4.799	9.657	-0.081
Maquaketa	7.802	183.970	-0.020
Mississippi	5.667	10.806	-0.050
Rock	4.398	11.204	-0.057
Sugar	9.577	463.502	-0.013
Pecatonica	7.040	191.858	0.017
Cedar	5.334	9.751	-0.102
Skunk	4.800	3.897	-0.136
Mississippi	5.713	7.773	-0.059
De Moines	4.983	7.293	-0.117
Raccoon	5.645	26.086	-0.060
Iroquois	4.849	13.331	-0.057
Kankakee	4.336	11.754	-0.045
Spoon	5.727	53.874	-0.031
La Moines	4.615	5.277	-0.125
Meramec	4.783	4.532	-0.166
Bourbeuse	24.639	7,028.689	0.003
Big	5.198	13.149	-0.079
Meramec	6.029	28.042	-0.053
		Missouri Basin	
Yellowstone	4.974	42.981	-0.018
Clarks Fork	8.324	1,868.663	-0.002
Yellowstone	5.241	23.933	-0.027
Big Sioux	6.674	39.524	-0.069
North Platte	4.092	10.138	-0.069
Bear	3.991	20.981	0.075
Elkhorn	36.204	8,367.870	-0.004
Nishabottna	4.605	2.468	-0.206
Grand	5.427	17.404	-0.063
Thompson	6.287	49.173	-0.041
Gasconade	5.681	17.553	-0.070

On the basis of regionalized annual 1-day high flows, the parameter values are given in Table 14.

Table 14: Pearson Type III Parameter for Regional Sequences of Annual 1-Day High Flows in Log Space

	m	*a*	*b*
Upper Mississippi Basin			
Log Space	1.323	29.530	-0.011
Quasi-Log Space	2.071	72.959	-0.029
Missouri Basin			
Log Space	1.313	23.972	-0.013
Quasi-Log Space	1.856	51.656	-0.036

Pearson Type III → Normal Distribution

As $a \to \infty$ the Pearson Type III distribution tends to the Normal distribution, $N(\mu, \sigma)$,

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} exp\left[-\left(\frac{x-\mu}{\sigma} \right)^2 \right] \qquad (51)$$

where $-\infty \le x \le \infty$. For an arbitrary sequence, $\{x_i: i = 1, \ldots, n\}$, the values of the parameters, μ and σ , of the Normal distribution are given by

$$\left.\begin{aligned} \mu &= \sum_{i=1}^{n} x_i / n \\ \sigma &= \left\{ \sum_{i=1}^{n} (x_i - \mu)^2 / n \right\}^{1/2} \end{aligned}\right\} \qquad (52)$$

For $m = 0$ and $b = 1$, the Pearson Type III distribution becomes the standard Gamma distribution

$$f(x) = \frac{1}{\Gamma(a)} x^{a-1} exp[-x] \qquad (53)$$

If $U_1, \ldots, U_v$ are independent random variables, each distributed as $N(0,1)$, then

$\sum_{j=1}^{\nu} U_j = \chi_\nu^2$ is distributed as chi square with ν degrees of freedom. The random variable $\chi_\nu^2/2$ is distributed as standard Gamma with $a = \nu/2$

$$f(\chi^2) = \frac{1}{\left[2^{\frac{\nu}{2}}\Gamma(\nu/2)\right]}(\chi^2)^{\frac{\nu}{2}-1} exp\left(-\frac{\chi^2}{2}\right) \tag{54}$$

where $\chi^2 \geq 0$ and ν is an integer.

The standard Gamma distribution tends to the unit normal distribution as $a \to \infty$, i.e. as skewness $\gamma \to 0$:

$$\begin{aligned} \lim_{a\to\infty} Prob\left[(x-a)a^{-\frac{1}{2}} \leq u\right] &= \Phi(u) \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u} exp\left(-\frac{t^2}{2}\right) \end{aligned} \tag{55}$$

The Pearson Type III distribution tends to the unit normal distribution as $a \to \infty$:

$$\lim_{a\to\infty} Prob\left\{\left[\frac{(x-m)}{b} - a\right]a^{-\frac{1}{2}} \leq u\right\} = \Phi(u) \tag{56}$$

The chi square distribution tends to the unit normal distribution as $a \to \infty$:

$$\lim_{a\to\infty} Prob\left[\frac{(\chi_\nu^2 - \nu)}{(2\nu)^{1/2}} \leq u\right] = \Phi(u) \tag{57}$$

Refer to Johnson and Kotz (1970).

For a large, Fisher (1922) proposed approximating the Normal distribution by the following transformation of χ_ν^2

$$Prob\left[\chi_\nu^2 < x\right] \approx \Phi\left(\sqrt{2x} - \sqrt{2\nu - 1}\right) \tag{58}$$

More rapid convergence is given by the transformation of Wilson and Hilferty (1931)

$$Prob\left[\chi_\nu^2 < x\right] \approx \Phi\left\{\left[(x/\nu)^{1/3} - 1 + 2/(9\nu)\right]\sqrt{9\nu/2}\right\} \tag{59}$$

See Wadsworth and Bryan (1960).

Fitting Right-Tail Normal Distributions

For a Right-Tail Normal distribution, the values of μ and σ are partitioned by the values of x greater than the median. Assume that the elements of $\{x_i: i = 1, \ldots, n\}$ are in rank order from smallest to largest. To each element of $\{x_i: i = 1, \ldots, n\}$, a probability, $F(x_i)$, is assigned, where for the Weibull plotting position,

$$F(x_i) = i/(n+1) \tag{60}$$

It is assumed that the right and left tails of the Normal distribution are marked by the median, which for a Normal distribution is the mean, μ: the left tail extend from $-\infty$ to the median, and the right tail extends from the median to ∞. Thus the right tail of the Normal distribution is given by $F(x_i)$ for $i = \nu, \ldots, n$, where

$$\nu = \begin{cases} \dfrac{n+1}{2}; \; n \text{ odd} \\ \dfrac{n+2}{2}; \; n \text{ even} \end{cases} \tag{61}$$

A Right-Tail Normal distribution is defined by values of μ and σ that are themselves defined by the partial sequence $\{x_i: i = \nu, \ldots, n\}$

The parameter μ of a Right-Tail Normal distribution may be defined as

$$\mu = \begin{cases} \dfrac{x_{\nu'} + x_{\nu''}}{2}; \; n \text{ even} \\ x_\nu; \; n \text{ odd} \end{cases} \tag{62}$$

where

$$\left.\begin{aligned} \nu' &= \frac{n}{2} \\ \nu'' &= \frac{n+2}{2} \end{aligned}\right\} n\ even \tag{63}$$

$$\nu = \frac{n+1}{2};\ n\ odd \tag{64}$$

By definition, for $\mu = median$,

$$F(\mu) = 0.5 \tag{65}$$

The parameter σ may be defined in several ways. Herein, three different method of determining σ are considered, 1) the inflection point method, 2) the $\eta-point$ method, and 3) the mirrored spread method.

Inflection Point Method

Let x^* denote the value for which

$$F(x^*) = 0.8413\cdots \tag{66}$$

A general property of the Normal distribution is
that

$$x^* - \mu = \sigma \tag{67}$$

where x^* marks the right inflexion point of the Normal density function. The left inflection point is given by $x^{**} - \mu = \sigma$. However, since μ and σ are defined by the observations greater than the median, the left inflection point may not match the left inflection point of a Left-Tail Normal distribution. In any case, the left tail of the observed distribution does not enter into determining the fit of a Right-Tail Normal distribution.

In general, x^* is determined by interpolation. Let $x_{i'}$ and $x_{i'+1}$ denote the elements of $\{x_i: i = 1, \ldots, n\}$ for which $F(x_{i'}) = (i'/n+1) < 0.8413\cdots$ and $F(x_{i'+1}) = [(i'+1)/(n+1)] > 0.8413\cdots$, where by linear interpolation

$$x^* = x_{i'+1} - \frac{(x_{i'+1} - x_{i'})(F(x_{i'+1}) - 0.8413\cdots)}{(F(x_{i'+1}) - F(x_{i'}))} \tag{68}$$

whereby

$$\sigma = x^* - \mu \tag{69}$$

$\eta - Point$ Method

Let $\{x_{i_j}: j = 1, \ldots, \eta\}$ denote an ordered sub-set of elements belonging to $\{x_i: i = 1, \ldots, n\}$ where $x_{i_j} \geq \mu \ \forall j$ and

$$\eta \leq \begin{cases} \dfrac{n+1}{2}; \ n \ odd \\ \dfrac{n}{2} \ n \ even \end{cases} \tag{70}$$

Let $F(u)$ denote the unit Normal distribution: $u \sim N(0,1)$. For

$$F(u_{i_j}) = F(x_{i_j}) = i_j/(n+1) \tag{71}$$

$$u_{i_j}\sigma = x_{i_j} - d \tag{72}$$

Therefore

$$\sigma \sum_{j=1}^{\eta} u_{i_j} = \sum_{j=1}^{\eta} x_{i_j} - \eta d \tag{73}$$

whereby

$$\sigma = \left[\sum_{j=1}^{\eta} x_{i_j} - d\eta\right] \Big/ \sum_{j=1}^{\eta} u_{i_j} \tag{74}$$

where

$$d = \begin{cases} 1;\ Regional\ Log\ Space \\ 0;\ Regional\ Quasi-Log\ Space \\ \mu;\ At-Site\ Log\ Space \end{cases} \tag{75}$$

For a given value of η, there are θ $\eta-point$ values of σ, where

$$\theta = \begin{cases} \begin{pmatrix} \frac{n+1}{2} \\ \eta \end{pmatrix};\ n\ odd \\ \begin{pmatrix} \frac{n}{2} \\ \eta \end{pmatrix};\ n\ even \end{cases} \tag{76}$$

For $\eta = 1$, there are $(n+1)/2$ possible values of σ, if n is odd or $n/2$ if n is even. The inflection point method is a special case of the $\eta-point$ method where $\eta = 1$ and $x_{i_1} = x^*$, where x^* is defined by Eq. (68). For η equal to $(n+1)/2$ if n is odd or $n/2$ if n is even, $\theta = 1$.

Mirrored Spread Method

The mirrored spread method takes σ to be twice the mean sum of squares about the mean of the values greater than the mean. The method yields

$$\sigma = \left\{(2/n)\sum_{i=\nu}^{n}(x_i - d)^2\right\}^{1/2} \tag{77}$$

where ν and d are defined by Eqs. (61) and (75).

Parameter Values

For the various sequences, the values of the parameters, μ and σ, of the Right-Tail Normal distribution were determined. Each of the three methods for determining σ were used In the case of the $\eta-point$ method, $\eta = 35 -$ the flow sequences are of length $n = 70$.

The at-site values of μ and σ for sequences of annual 1-day high flows in log space are given in Table 15.

Table 15: Right-Tail Normal Parameters for Sequences of Annual 1-Day High Flows in Log Space

	μ	σ Inflection Point	σ 35-Point	σ Mirrored Spread
		Upper Mississippi Basin		
St. Croix	4.400	0.151	0.157	0.144
Jump	3.886	0.192	0.204	0.205
Black	4.051	0.179	0.202	0.198
Maquaketa	4.051	0.295	0.298	0.263
Mississippi	5.149	0.151	0.134	0.131
Rock	3.761	0.205	0.184	0.173
Sugar	3.423	0.276	0.308	0.282
Pecatonica	3.703	0.290	0.263	0.247
Cedar	4.386	0.281	0.261	0.242
Skunk	4.274	0.190	0.256	0.232
Mississippi	5.279	0.126	0.133	0.127
De Moines	4.179	0.274	0.244	0.239
Raccoon	4.090	0.330	0.312	0.287
Iroquois	4.109	0.215	0.195	0.175
Kankakee	3.815	0.159	0.142	0.135
Spoon	4.088	0.213	0.214	0.196
La Moines	3.984	0.241	0.248	0.230
Meramec	4.086	0.243	0.258	0.258
Bourbeuse	4.107	0.245	0.254	0.242
Big	4.167	0.238	0.275	0.252
Meramec	4.570	0.275	0.248	0.245
Average	*4.169*	*0.227*	*0.228*	*0.214*
Std. Deviation	*0.430*	*0.056*	*0.055*	*0.050*
		Missouri Basin		
Yellowstone	4.212	0.096	0.106	0.105
Clarks Fork	3.841	0.116	0.113	0.105
Yellowstone	4.584	0.128	0.135	0.126
Big Sioux	4.009	0.337	0.360	0.360
North Platte	3.421	0.175	0.188	0.170
Bear	2.315	0.418	0.438	0.405
Elkhorn	4.104	0.315	0.318	0.310
Nishabottna	4.149	0.218	0.229	0.227
Grand	4.394	0.180	0.194	0.193
Thompson	4.273	0.245	0.275	0.251
Gasconade	4.442	0.289	0.297	0.276
Average	*3.977*	*0.229*	*0.241*	*0.230*
Std. Deviation	*0.635*	*0.102*	*0.106*	*0.102*

The three methods yield comparable Right-Tail Normal values of σ. The average values of σ are nearly the same for each of the methods. The average values are somewhat larger for the Missouri basin than for the Upper Mississippi basin. The standard deviations of the values of σ within the Upper Mississippi basin are about half those within the Missouri basin.

The average values of the Right-Tail Normal values of σ are somewhat smaller than the average values overall values of σ. Refer to Tables 1 and 10. The variability of the Right-Tail Normal values σ over the sequences in either basin is nearly equal to the variability among the at-site values of σ. Refer to Tables 11 and 15.

On a regional basis, the values of μ and σ are given in Table 16.

Table 16: Right-Tail Normal Parameters for Regional Sequences of Annual 1-Day High Flows

	μ	σ Inflection Point	σ 35-Point	σ Mirrored Spread
		Upper Mississippi Basin		
Log Space	1	0.057	0.056	0.053
Quasi-Log Space	0	0.236	0.234	0.223
		Missouri Basin		
Log Space	1	0.053	0.057	0.054
Quasi-Log Space	0	0.218	0.234	0.227

From Table 16, it is noted that the values of μ and σ in a given space log space for one basin are almost equal to the values for that space for the other basin. In either log space or quasi-log, the three methods yield almost equal values of σ.

Goodness of Fit

The comparison of the Pearson Type III distribution with the Right-Tail Normal distribution is in terms of magnitudes of the estimates of the 50-year event for the 70-year sequences and of both the 50-year and the 100-year event for the 100-year sequences.

Let $\{x_i: i = 1, \ldots, n\}$ denote a sequence of flows ordered from smallest, x_1, to largest, x_n. Each flow is assigned a probability, defined as the Weibull plotting position. Refer to Eq. (16). The $T-year$ flow, $X(T)$, where T equals 50 or 100, is given by linear interpolation. Let $x_{i'}$ and $x_{i'+1}$ denote the elements of $\{x_i: i = 1, \ldots, n\}$, for which $F(x_{i'}) = [i'/(n+1)] < F^*$ and $F(x_{i'+1}) = [(i'+1)/(n+1)] > F^*$. By linear interpolation

$$X(T) = x_{i'+1} - \frac{(x_{i'+1})(F(x_{i'+1}) - F^*)}{(F(x_{i'+1}) - F(x_{i'}))} \tag{78}$$

where

$$F^* = \begin{cases} 0.98;\ if\ T = 50 \\ 0.99;\ if\ T = 100 \end{cases} \tag{79}$$

Let $X(T|PIII)$ denote the value of the $T-year$ flow obtained from a fitted Pearson Type III distribution:

$$F^* = \begin{cases} \int_m^{X(T|PIII)} f(x)dx;\ if\ a > 0 \\ \int_m^{X(T|PIII)} f(-x)dx;\ if\ a < 0 \end{cases} \tag{80}$$

where F^* is defined by Eq. (79), $f(x)$, by Eq. (46) and a, by Eq.(49)

Let $X(T|RTN)$ denote the value of the $T-year$ flow obtained from a fitted Right-Tail Normal distribution:

$$F^* = \int_{-\infty}^{X(T|RTN)} f(x)dx \tag{81}$$

where F^* is defined by Eq. (79), and $f(x)$, by Eq. (51).

Let

$$\Delta(PIII) = X(T) - X(T|PIII) \quad (82)$$

denote the difference between the estimates of the $T-year$ flow obtained from the observations and the fitted Pearson Type III distribution. Let

$$\Delta(RTN) = X(T) - X(T|RTN) \quad (83)$$

If $|\Delta(PIII)| < |\Delta(RTN)|$, then the Pearson Type III distribution provides a better estimate of the $T-year$ flow than the Right-Tail Normal distribution. If, however, $|\Delta(PIII)| > |\Delta(RTN)|$, then the Right-Tail Normal distribution provides a better estimate of the $T-year$ flow.

The comparison of the goodness of fit of the Pearson Type III distribution and the Right-Tail Normal distribution among the at-site sequences of annual 1-day high flows in log space is given in Table 17.

Table 17: Goodness of Fit per $T = 50 - year$ Flow Relative to At-Site Sequences of 1-Day high Flows in Log Space

River	Skewness	$\Delta(PIII)$	$\Delta(RTN)$ Inflection Point	$\Delta(RTN)$ 35-Point	$\Delta(RTN)$ Mirrored Spread
			Upper Mississippi Basin		
St. Croix	-0.633	0.008	-0.013	-0.026	0.000*
Jump	-0.062	0.131*	0.181	0.155	0.154
Black	-0.644	0.100	0.125	0.077*	0.084
Maquaketa	-0.147	-0.074	-0.141	-0.147	-0.074*
Mississippi	-0.608	0.030	-0.021	0.014*	0.020
Rock	-0.598	0.013*	-0.085	-0.043	-0.020
Sugar	-0.093	-0.034	-0.011*	-0.077	-0.023
Pecatonica	0.144	0.026	-0.092	-0.035	-0.003*
Cedar	-0.640	-0.029*	-0.117	-0.076	-0.036
Skunk	-1.013	0.086	0.088	-0.049	0.001*
Mississippi	-0.717	0.066	0.055	0.039*	0.053
Des Moines	-0.741	0.078	-0.019*	0.046	0.054
Raccoon	-0.392	0.021	-0.099	-0.060	-0.009*
Iroquois	-0.548	-0.034*	-0.132	-0.091	-0.050
Kankakee	-0.583	0.046	-0.023	0.011*	0.026
Spoon	-0.272	0.007*	-0.021	-0.024	0.013
La Moines	-0.871	0.009*	-0.068	-0.083	-0.045
Meramec	-0.940	0.077	0.058	0.028	0.028*
Bourbeuse	0.024	0.095	0.095	0.078*	0.102
Big	-0.552	0.063	0.062	-0.013*	0.035
Meramec	-0.378	0.043	-0.028	0.027*	0.034
Average	*-0.489*	*0.035*	*-0.010*	*-0.012*	*-0.016*
Std. Deviation	*0.319*	*0.052*	*0.089*	*0.070*	*0.054*
			Missouri Basin		
Yellowstone	-0.307	0.044	0.066	0.040*	0.041
Clark's Fork	-0.046	0.012	-0.010	-0.004*	0.012
Yellowstone	-0.409	0.042	0.023	0.008*	0.027
Big Sioux	-0.318	0.042*	0.104	0.057	0.056
North Platte	-0.628	0.024	0.008*	-0.019	0.018
Bear	0.437	0.126	-0.065	-0.107	-0.038*
Elkhorn	-0.022	0.044*	0.079	0.081	0.089
Nishabottna	-1.273	0.185	0.114	0.092*	0.096
Grand	-0.479	0.080*	0.121	0.090	0.093
Thompson	-0.285	0.036	0.059	-0.005*	0.045
Gasconade	-0.477	0.073	0.010	-0.007*	0.035
Average	*-0.346*	*0.064*	*0.046*	*0.021*	*0.043*
Std. Deviation	*0.423*	*0.051*	*0.058*	*0.059*	*0.040*

* Best Fit

From Table 17, it is noted that on average the Right-Tail Normal distribution provides a better estimate of the $T = 50 - year$ flow than the Pearson Type III distribution in both the Upper Mississippi basin and the Missouri basin. For 15 of the 21 sequences in the Upper Mississippi basin, the Right-Tail Normal distribution out performs the Pearson Type III distribution. Of those 15 sequences, method 2 – the $\eta - point$ method – is somewhat better than method 3 – the mirrored spread method – by a factor of 8-to-6. However, for 14 of the 15 sequences, the Right-Tail Normal distribution conditioned on method 3 is better than the Pearson Type III distribution, whereas, for 10 of the 15 sequences, the Right-Tail Normal distribution conditioned on method 2 is better than the Pearson Type III distribution.

For 8 of the 11 sequences in the Missouri basin, the Right-Tail Normal distribution out performs the Pearson Type III distribution. Of those 8 sequences, method 2 performs best, although in each case the Right-Tail Normal distribution conditioned on method 3 better fits the flows than the Pearson Type III distribution.

For the 32 sequences in the two basins, 23 are better fitted with the Right-Tail Normal distribution than with the Pearson Type III distribution.

The probability distributions fitted with the Pearson Type III distribution and with the Right-Tail Normal distribution are shown for each of the 32 sequences in Appendix I.

The comparison of the goodness of fit of the Pearson Type III distribution and the Right-Tail Normal distribution among the regional sequences of annual 1-day high flows in log space is given in Table 18.

Table 18: Goodness of Fit per $T = 50 - year$ Relative to Regional Sequences of 1-Day high Flows in Log Space

River	Skewness	$\Delta(PIII)$	$\Delta(RTN)$ Inflection Point	$\Delta(RTN)$ 35-Point	$\Delta(RTN)$ Mirrored Spread
			Upper Mississippi Basin		
Log Space	-0.368	0.007	-0.003	0.002*	0.005
Quasi-Log Space	-0.234	0.029	0.000*	0.004	0.027
			Missouri Basin		
Log Space	-0.408	0.019	0.024	0.013*	0.019
Quasi-Log Space	-0.278	0.081	0.106	0.072*	0.087

* Best Fit

For both the Upper Mississippi and the Lower Missouri basin, the regional sequences are best fitted with a Right-Tail Normal distribution. Although, the Right-Tail Normal distribution conditioned on the inflection point method is best in quasi-log space in the Upper Mississippi basin, Right-Tail Normal distribution conditioned on the 35-point method is better is better than the Pearson Type III distribution in all cases.

The probability distribution fitted with the Pearson Type III distribution and with the Right-Tail Normal distribution are shown for each of the 4 regional sequences in Appendix J.

Real Space

The distributions of the at-site sequences in log space fitted with a Pearson Type III distribution and with a Right-Tail Normal distribution are transformed into real space through exponentiation. The Pearson Type III distribution in log space exponentiates to the Log-Pearson Type III distribution in real space, and the Right-Tail Normal distribution in log space exponentiates to the Right-Tail Log-Normal distribution in real space. The transform is monotonic, and therefore it does not affect the relative magnitudes of estimates of the $T - year$ flow. Whichever distribution fits best in log space, its real space transform fits best in real space.

The distributions of the regional sequences in log space fitted with a Pearson Type III distribution and with a Right-Tail Normal distribution are transformed into real space as follows. Let $X'(T)$ denote one of the three estimates of the $T-year$ flow, namely, $X(T)$, $X(T|PIII)$ or $X(T|RTN)$. Let $\tilde{x}_j$ denote the median of the logs of the flows at the $j-th$ site in the region. The real space estimate of the $T-year$ flow at the $j-th$, site $Y_j'(T)$, is given by

$$Y_j'(T) = Exp\left[\tilde{x}_j X'(T)\right] \tag{84}$$

Multiplying each of the log space estimates of the $T-year$ flow by a constant, namely, $\tilde{x}_j$, does not effect the relative magnitudes of the estimates. Because exponentiation does not effect the relative magnitudes of the products of the log space estimates of the $T-year$ flows by a constant, whichever distribution fits best in log space, its real space transform fits best in real space.

The distributions of the regional sequences in quasi-log space fitted with a Pearson Type III distribution and with a Right-Tail Normal distribution are transformed into real space as follows. Let $X''(T)$ denote one of the three estimates of the $T-year$ flow, namely, $X(T)$, $X(T|PIII)$ or $X(T|RTN)$. Let w_j denote the median of the real space flows at the $j-th$ site. The real space estimate of the $T-year$ flow at the $j-th$ site, $Y_j'(T)$, is given by

$$Y_j'(T) = w_j Exp\left[X'(T)\right] \tag{85}$$

It follows that whichever distribution fits best in log space, its real space transform fits best in real space.

If the estimates of the $T-year$ flows at specific sites in a region are obtained through at-site analysis, then for a good majority of the sites the better estimates are those based on the Right-Tail Normal distribution in log space or the Right-Tail Log-Normal distribution in real space. On the other hand, if the estimates of the $T-year$ flows at the specific sites are obtained through regional analysis, then for all sites he better estimates are those based on the Right-Tail Normal distribution in log space or the Right-Tail Log-Normal distribution in real space.

With the Right-Tail Normal distribution in log space, the matter of whether one or more of the "low" flows should be censored need not be addressed. By using the Right-Tail Normal distribution, there is no need to determine the skewness at site or through regionalization.

Comparison of Pearson Type III and Right-Tail Normal Distributions – Annual Peak Flows

A comparison between the Pearson Type III distribution and the Right-Tail Normal distribution is made with respect to estimates of the $T-year$, for $T = 50, 100$. The estimates the $T-year$ floods are obtained from sequences of annual peaks, 7 along the Upper Mississippi river and 7 along the Missouri river. The Mississippi river sequences span the 100 year period 1896-1995, and the Missouri river sequences span the 100 year period 1898-1997. See Table 19.

Table 19: Location and Statistical Description of Sequences of Annual Peak Flows

Locale	State	Flow Descriptors		
		μ	σ	γ
		Upper Mississippi Basin		
St. Paul	MN	4.576	0.258	-0.295
Winona	MN	4.931	0.212	-0.803
Dubuque	IA	5.092	0.163	-0.623
Clinton	IA	5.112	0.154	-0.432
Keokuk	IA	5.241	0.158	-0.584
Hannibal	MO	5.296	0.188	-1.350
St. Louis	MO	5.691	0.169	-0.477
Average		*5.134*	*0.186*	*-0.650*
Std. Deviation		*0.342*	*0.038*	*0.347*
		Missouri Basin		
Sioux City	IA	5.158	0.196	-0.522
Omaha	NE	5.157	0.186	-0.471
Nebraska City	NE	5.216	0.187	-0.494
St. Joseph	MO	5.222	0.159	-0.185
Kansas City	MO	5.322	0.184	-0.017
Booneville	MO	5.409	0.186	-0.097
Hermann	MO	5.492	0.196	0.025
Average		*5.282*	*0.185*	*-0.252*
Std. Deviation		*0.128*	*0.012*	*0.226*

For further description of these sequences refer to Planning & Management Consultants, Ltd. (1999.)

The comparison between the Pearson Type III distribution and the Right-Tail Normal distribution is in reference to the relative goodness of fit of the distributions with respect to estimates of the $T-year$ flood, where $T = 50, 100$. The comparison is made in reference to both at-site sequences and regionalized sequences. The assessment of the relative goodness of fit of the distributions is limited to estimates of the $T-year$ flood, for $T \geq 2$. To consider values of $T < 2$, the definition of the right tail would need to be revised.

In contrast to the comparison in terms of the annual 1-day high flows discussed above, the comparison in terms of the annual peak flows takes into account the statistical significance in the differences in the estimates of the $T-year$ floods derived from the Pearson Type III distribution and the Right-Tail Normal distribution.

Goodness of Fit

The Pearson Type III and the Right-Tail Normal distributions were fitted to ordered sequences of flows using the Weibull plotting position. To fit the Pearson Type III distribution, the parameters of the distribution were determined by the method of moments. See Table 20.

Table 20: Pearson Type III Parameter for At-Site Sequences of Annual Peak Flows in Log Space

	m	*a*	*b*
	Upper Mississippi Basin		
St. Paul	6.325	45.942	-0.295
Winona	5.458	6.202	-0.085
Dubuque	5.616	10.287	-0.051
Clinton	5.824	21.451	-0.033
Keokuk	5.781	11.717	-0.046
Hannibal	5.574	2.195	-0.127
St. Louis	6.4	17.567	-0.040
	Missouri Basin		
Sioux City	5.908	14.654	-0.051
Omaha	5.949	18.067	-0.044
Nebraska City	5.972	16.369	-0.046
St. Joseph	6.944	117.139	-0.015
Kansas City	26.258	12,994.935	-0.002
Booneville	9.261	426.874	-0.009
Hermann	21.354	6,573.286	0.002

The Right Tail Normal distribution was fitted to an observed distribution using each of the three methods for estimating the parameters of the Right-Tail Normal distribution – inflection point, $\eta-$ point and mirrored spread– where $\eta = 50$. See Table 21.

Table 21: Right-Tail Normal Parameters for At-Site Sequences of Annual Peak Flows in Log Space

	μ	σ Inflection Point	σ 50-Point	σ Mirrored Spread
		Upper Mississippi Basin		
St. Paul	4.600	0.117	0.224	0.244
Winona	4.956	0.178	0.173	0.173
Dubuque	5.119	0.132	0.125	0.128
Clinton	5.125	0.120	0.137	0.133
Keokuk	5.252	0.141	0.140	0.136
Hannibal	5.324	0.153	0.140	0.140
St. Louis	5.694	0.181	0.170	0.156
Average	*5.153*	*0.146*	*0.158*	*0.159*
Std. Deviation	*0.366*	*0.026*	*0.034*	*0.041*
		Missouri Basin		
Sioux City	5.154	0.168	0.196	0.179
Omaha	5.161	0.165	0.176	0.165
Nebraska City	5.248	0.132	0.147	0.144
St. Joseph	5.239	0.122	0.136	0.139
Kansas City	5.329	0.153	0.166	0.175
Booneville	5.401	0.192	0.197	0.189
Hermann	5.484	0.211	0.214	0.201
Average	*5.288*	*0.163*	*0.176*	*0.170*
Std. Deviation	*0.123*	*0.031*	*0.028*	*0.023*

On a regional basis, in log space and quasi-log space, the parameters of the Fitted Pearson Type III and Right-Tail Normal distributions are given in Tables 22 and 23.

Table 22: Pearson Type III Parameter for Regional Sequences of Annual Peak Flows

	m	a	b
	Upper Mississippi Basin		
Log Space	1.104	10.820	-0.039
Quasi-Log Space	0.590	12.633	-0.048
	Missouri Basin		
Log Space	1.265	62.403	-0.004
Quasi-Log Space	1.295	53.874	-0.024

Table 23: Right-Tail Normal Parameters for Regional Sequences of Annual Peak Flows

	μ	σ Inflection Point	σ 50-Point	σ Mirrored Spread
		Upper Mississippi Basin		
Log Space	1	0.029	0.028	0.027
Quasi-Log Space	0	0.153	0.149	0.144
		Missouri Basin		
Log Space	1	0.029	0.031	0.030
Quasi-Log Space	0	0.149	0.161	0.159

Let $X(T)$ denote the estimate of the $T-year$ flood given by the observed distribution. Let $X(T|PIII)$ denoted the estimate of the $T-year$ flood given by the Pearson Type III distribution fitted to the observed distribution. Let $X(T|RTN_i)$ denote the estimate of the $T-year$ flood given by the Right-Tail Normal distribution fitted to the observed distribution using method i for estimating the parameter of the Right Tail Normal distribution, where $i = 1, 2, 3$ refers to the inflection point method, the $\eta = 50$-point method and the mirrored spread method.

Let

$$\Delta(PIII) = X(T) - X(T|PIII) \tag{78}$$

denote the difference between the estimates of the $T-year$ flood given by the observed and the fitted Pearson Type III distributions.
Let

$$\Delta(RTN_i) = X(T) - X(T|RTN_i) \tag{79}$$

denote the difference between the estimates of the $T-year$ flood given by the observed and the fitted Right-Tail Normal Distribution using parameter estimation method i.

If $|\Delta(PIII) < \Delta(RTN_i)|$ then the Pearson Type III distribution provides a better estimate of the $T-year$ flood, and thereby a better fit to the observed distribution, than the Right-Tail Normal distribution using parameter estimation method i. If $|\Delta(PIII) > \Delta(RTN_i)|$ then the Right-Tail Normal distribution using parameter estimation method i provides a better estimate of the $T-year$ flood, and thereby a better fit to the observed distribution, than the Pearson Type III distribution. If $|\Delta(PIII) = \Delta(RTN_i)|$, then there is no difference in the goodness of fits of the Pearson Type III distribution and the Right-Tail Normal distribution using parameter estimation method i.

Statistical Assessment of Goodness of Fit

At-Site Analysis

The statistical significance of the difference between $\Delta(PIII)$ and $\Delta(RTN_i)$ is assessed as follows. Let $\Delta_j(PIII)$ and $\Delta_j(RTN_i)$ denote the differences between the observed estimates of the $T-year$ flood given by the observed and the fitted Pearson Type III distribution and the Right-Tail Normal Distribution using parameter estimation method i, respectively at the $j-th$, where $j-th$. For either the Upper Mississippi basin or the Missouri basin, $M = 7$.

Let

$$\xi(PIII) = \sum_{j=1}^{M} \Delta_j(PIII) / M \tag{80}$$

$$\varsigma(PIII) = \left\{ \sum_{j=1}^{M} \left[\Delta_j(PIII) - \xi(PIII) \right]^2 / M \right\}^{1/2} \tag{81}$$

denote the mean and standard deviation of the values $\Delta_j(PIII)$. The mean and standard deviation of the values $\Delta_j(RTN_i)$ are

$$\xi(RTN_i) = \sum_{j=1}^{M} \Delta_j(RTN_i) / M \tag{82}$$

$$\varsigma(RTN_i) = \left\{ \sum_{j=1}^{M} \left[\Delta_j RTN_i - \xi RTN_i \right]^2 \right\}^{1/2} \tag{83}$$

Let $\tilde{\xi}(PIII)$ and $\tilde{\xi}(RTN_i)$, and $\tilde{\varsigma}(PIII)$ and $\tilde{\varsigma}(RTN_i)$ denote the "population" means and standard deviations, i.e., the values of the means and standard deviations of the $\Delta_j(PIII)$ and $\Delta_j(RTN_i)$ that would attain with $M = \infty$.

If the null hypothesis

$$H_0(\tilde{\varsigma}^2) : \tilde{\varsigma}^2(PIII) = \tilde{\varsigma}^2(RTN_i)$$

is not rejected at probability level α, then if the null hypothesis

$$H_0(\tilde{\xi}) : \tilde{\xi}(PIII) = \tilde{\xi}(RTN_i)$$

is not rejected at probability level β, then there is no statistical difference in the goodness of fits of the Pearson Type III distribution and the Right-Tail Normal distribution to the observed distribution. In general, $\alpha = \beta$.

The null hypothesis $H_0(\tilde{\varsigma}^2)$ is assess by means of the F-distribution. See e.g. Johnson and Kotz (1970). Let

$$\varsigma_1^2 = max\left\{\varsigma^2(PIII), \varsigma^2(RTN_i)\right\} \tag{84}$$

$$\varsigma_2^2 = min\left\{\varsigma^2(PIII), \varsigma^2(RTN_i)\right\} \tag{85}$$

Let

$$F = \frac{M_1\varsigma_1^2/(M_1-1)}{M_2\varsigma_2^2/(M_2-1)} \tag{86}$$

The probability that F will fall in the interval dF is

$$dP = \frac{\nu_1^{\nu_1/2}\nu_2^{\nu_2/2}F^{(\nu_1-2)/2}dF}{B(\nu_1/2, \nu_2/2)(\nu_1F+\nu_2)^{(\nu_1+\nu_2)/2}} \tag{87}$$

where $B(\nu_1/2, \nu_2/2)$ denotes the beta function with arguments $\nu_1/2$ and $\nu_2/2$. See e.g. Abramowitz and Stegun (1964).

Thus

$$Prob\left[F > F^*\right] = 1 - \int_0^F dP \tag{88}$$

If $Prob\left[F > F^*\right] \geq \alpha$, then the null hypothesis $H_0(\tilde{\varsigma}^2)$ is rejected.

Herein, $M_1 = M_2 = M$, where $M = 7$ for both the Upper Mississippi basin and the Lower Missouri basin, and $M = 14$ in the case where the basins are pooled.

Assume that the null hypothesis $H_0(\tilde{\varsigma}^2)$ is not rejected, i.e., $Prob\left[F \leq F^*\right] < \alpha$, in which case the null hypothesis $H_0(\tilde{\xi})$ may be assessed by means of the t-distribution. Let

$$t = \left[\xi(PIII) - \xi(RTN_i)\right]\left\{\frac{M_1\varsigma(PIII) + M_2\varsigma(RTN_i)}{M_1 + M_2 - 2}\right\}^{1/2}\left\{\frac{M_1M_2}{M_1 + M_2}\right\}^{1/2} \tag{89}$$

where t is distributed as

$$dt = \left(\frac{1}{\left(M^*\right)^{1/2} B\left(M^*/2, 1/2\right)} \right) \left(1 + \frac{t^2}{M^*} \right)^{-\left(M^*+1\right)/2} \tag{90}$$

where $B\left(M^*/2, 1/2\right)$ denotes the beta function with arguments $M^*/2$ and $1/2$. Refer to Abramowitz and Stegun (1964). If

$$Prob\left[t > t^*\right] = 1 - \int_{-\infty}^{t} dt < \beta \tag{91}$$

then the null hypothesis $H_0\left(\tilde{\xi}\right)$ is rejected.

Regional Analysis-Log Space

To statistically assess the goodness of fits of the Pearson Type III distribution and the Right-Tail Normal distribution to a regionalized distribution, the regional values of $\Delta(PIII)$ and $\Delta\left(RTN_i\right)$ are transposed into at-site values as follows. Let ω_j denote the median of the logs of the flows at site j. The at-site fit of the Pearson Type III distribution and the Right-Tail Normal distribution at the $j-th$ site derived via regionalization in log space are given as

$$\delta_j(PIII) = \omega_j \Delta(PIII) \tag{92}$$

$$\delta_j\left(RTN_i\right) = \omega_j \Delta\left(RTN_i\right) \tag{93}$$

It is noted that if the regional distribution is best fitted by the Pearson Type III (Right-Tail Normal distribution), then at-site distribution derived through regionalization is best fitted by the Pearson Type III (Right-Tail Normal distribution). The at-site goodness of fit at the $j-th$ site is measured by the regional goodness of fit times a constant, namely, the median of the logs of the flows at the $j-th$ site.

The means and the standard deviations of the $\delta_j(PIII)$ and the $\delta_j\left(RTN_i\right)$ are given by Eqs. (80) through (83) with $\Delta_j(PIII)$ and $\Delta_j\left(RTN_i\right)$ replaced by $\delta_j(PIII)$ and $\delta_j\left(RTN_i\right)$, respectively. Note,

$$\begin{aligned}\xi(PIII) &= \Delta(PIII)\sum_{j=1}^{M}\omega_j / M \\ &= \Delta(PIII)\overline{\omega}\end{aligned} \tag{94}$$

$$\varsigma(PIII) = \Delta(PIII)\left\{\sum_{j=1}^{M}\left[\omega_j - \overline{\omega}\right]^2 / M\right\}^{1/2} \tag{95}$$

$$\begin{aligned}\xi(RTN_i) &= \Delta(RTN_i)\sum_{j=1}^{M}\omega_j / M \\ &= \Delta(RTN_i)\overline{\omega}\end{aligned} \tag{96}$$

$$\varsigma(RTN_i) = \Delta(RTN_i)\left\{\sum_{j=1}^{M}\left[\omega_j - \overline{\omega}\right]^2 / M\right\}^{1/2} \tag{97}$$

Regional Analysis-Quasi-Log Space

Let

$$\Psi(PIII) = exp\left[\Delta(PIII)\right] \tag{98}$$

$$\Psi(RTN_i) = exp\left[\Delta(RTN_i)\right] \tag{99}$$

Let θ_j denote the median of the flows at the $j-th$ site. The at-site fit of the Pearson Type III distribution and the Right-Tail Normal distribution at the $j-th$ site derived via regionalization in quasi-log space are given as

$$\psi_j(RTN_i) = \theta_j\Psi(RTN_i) \tag{100}$$

$$\psi_j(RTN_i) = \theta_j\Psi(RTN_i) \tag{101}$$

whereby

$$\begin{aligned}\xi(PIII) &= \Psi(PIII)\sum_{j=1}^{M}\theta_j / M \\ &= \Psi(PIII)\overline{\theta}\end{aligned} \tag{102}$$

$$\varsigma(PIII) = \Psi(PIII)\left\{\sum_{j=1}^{M}\left[\theta_j - \overline{\theta}\right]^2 / M\right\}^{1/2} \tag{103}$$

$$\begin{aligned}\xi\left(RTN_i\right) &= \Psi\left(RTN_i\right)\sum_{j=1}^{M}\theta_j / M \\ &= \Psi\left(RTN_i\right)\overline{\theta}\end{aligned} \tag{104}$$

$$\varsigma\left(RTN_i\right) = \Psi\left(RTN_i\right)\left\{\sum_{j=1}^{M}\left[\theta_j - \overline{\theta}\right]^2 / M\right\}^{1/2} \tag{105}$$

It follows that if the regional distribution is best fitted by the Pearson Type III (Right-Tail Normal distribution), then at-site distribution derived through regionalization is best fitted by the Pearson Type III (Right-Tail Normal distribution). The at-site goodness of fit at the $j-th$ site is measured by the anti-log of the regional goodness of fit times a constant, namely, the median of the flows at the $j-th$ site.

At-Site Goodness of Fit

The at-site values of $\Delta(PIII)$ and $\Delta\left(RTN_i\right)$ conditioned on $T = 50, 100$ are given in Tables 24a and 24b.

Table 24a: Goodness of Fit per $T = 50 - year$ Flow Relative to At-Site Sequences of Annual Peak Flows in Log Space

River	Skewness	$\Delta(PIII)$	$\Delta(RTN)$ Inflection Point	$\Delta(RTN)$ 50-Point	$\Delta(RTN)$ Mirrored Spread
			Upper Mississippi Basin		
St. Paul	-0.295	0.127*	0.228	0.131	0.132
Winona	-0.803	0.067	0.016*	0.025	0.025
Dubuque	-0.623	0.012	-0.007	-0.007	-0.000*
Clinton	-0.432	-0.015	0.004*	-0.031	-0.022
Keokuk	-0.584	0.044	0.016*	0.019	0.027
Hannibal	-1.350	0.098	-0.001*	0.027	0.025
St. Louis	-0.477	-0.028*	-0.100	-0.078	-0.048
Average	*-0.650*	*0.044*	*0.022*	*0.012*	*0.020*
Std. Deviation	*0.347*	*0.058*	*0.099*	*0.064*	*0.057*
			Missouri Basin		
Sioux City	-0.522	-0.015	0.009*	-0.037	-0.019
Omaha	-0.471	0.031*	0.044	0.033	0.037
Nebraska City	-0.494	0.037	0.066	0.037*	0.043
St. Joseph	-0.185	0.071*	0.114	0.086	0.080
Kansas City	-0.017	0.059*	0.115	0.088	0.070
Booneville	-0.097	-0.051	-0.054	-0.065	-0.048*
Hermann	0.025	0.036	0.011	0.006*	0.032
Average	*-0.252*	*0.024*	*0.044*	*0.021*	*0.028*
Std. Deviation	*0.238*	*0.043*	*0.061*	*0.058*	*0.046*
			Pooled Basins		
Average	*-0.452*	*0.034*	*0.033*	*0.017*	*0.024*
Std. Deviation	*0.353*	*0.050*	*0.080*	*0.059*	*0.050*

* Minimum Absolute Difference in Fit

Table 24b: Goodness of Fit per $T = 100 - year$ Flow Relative to At-Site Sequences of Annual Peak Flows in Log Space

River	Skewness	$\Delta(PIII)$	$\Delta(RTN)$ Inflection Point	$\Delta(RTN)$ 50-Point	$\Delta(RTN)$ Mirrored Spread
			Upper Mississippi Basin		
St. Paul	-0.295	0.113	0.221	0.112*	0.113
Winona	-0.803	0.130	0.058*	0.068	0.068
Dubuque	-0.623	0.077	0.048*	0.064	0.056
Clinton	-0.432	0.066	0.081	0.041*	0.051
Keokuk	-0.584	0.104	0.064	0.066	0.076
Hannibal	-1.350	0.151	0.020*	0.051	0.049
St. Louis	-0.477	0.083	-0.008*	0.017	0.051
Average	*-0.650*	*0.103*	*0.069*	*0.082*	*0.066*
Std. Deviation	*0.347*	*0.030*	*0.073*	*0.045*	*0.023*
			Missouri Basin		
Sioux City	-0.522	0.177	0.195	0.142*	0.162
Omaha	-0.471	0.163	0.171	0.156*	0.162
Nebraska City	-0.494	0.114	0.141	0.107*	0.114
St. Joseph	-0.185	0.118*	0.165	0.133	0.126
Kansas City	-0.017	0.103*	0.166	0.136	0.116
Booneville	-0.097	0.117	0.112	0.099*	0.119
Hermann	0.025	0.039	0.010	0.005*	0.034
Average	*-0.252*	*0.119*	*0.137*	*0.111*	*0.119*
Std. Deviation	*0.238*	*0.045*	*0.062*	*0.051*	*0.043*
			Pooled Basins		
Average	*-0.452*	*0.111*	*0.137*	*0.111*	*0.119*
Std. Deviation	*0.353*	*0.038*	*0.074*	*0.048*	*0.043*

* Minimum Absolute Difference in Fit

In the case of the $50 - year$ flood, the Right-Tail Normal distribution provides a better fit to the observed distribution than the Pearson Type III distribution at a majority of the sites – 5 of the 7 in the Upper Mississippi basin and 4 of the 7 in the Missouri basin. In the case of the $100 - year$ flood, the Right-Tail Normal Distribution is more dominate than the Pearson Type III distribution. At all 7 site in the Upper Mississippi basin and at 5 of the 7 sites in the Missouri basin, the Right-Tail Normal distribution provides a better fit to the observed distribution than the Pearson Type III distribution.

The fits of the Pearson Type III distributions and the Right-Tail Normal distributions to the observed at-site distributions are shown in Appendix K. The Right-Tail Normal distributions are fitted using the mirrored spread method to estimate the distribution's parameters.

Assessment of Variances of Differences in Fits

The F values and their corresponding probabilities relating to the variances of the at-site differences in the fits of the Pearson Type III distribution and the Right-Tail Normal Distribution to the observed distribution are given in Table 25a and 25b.

Table 25a: F Values Based on At-Site Fits of the Pearson Type III and the Right-Tail Normal Distributions to Observed Distributions

Method	50-Year Peak Flow	100-Year Peak Flow
	Upper Mississippi Basin	
Inflection	2.951	5.767
50-Point	1.244	0.916
Mirrored Spread	0.970	0.565
	Missouri Basin	
Inflection	2.042	1.911
50-Point	1.837	1.295
Mirrored Spread	1.168	0.918
	Pooled Basins	
Inflection	2.570	3.877
50-Point	1.402	1.622
Mirrored Spread	1.004	1.299

Table 25b: Probability of F Values Based on At-Site Fits of the Pearson Type III and the Right-Tail Normal Distributions to Observed Distributions

Method	50-Year Peak Flow	100-Year Peak Flow
	Upper Mississippi Basin	
Inflection	0.107	0.026
50-Point	0.399	0.541
Mirrored Spread	0.514	0.747
	Missouri Basin	
Inflection	0.203	0.255
50-Point	0.239	0.382
Mirrored Spread	0.428	0.540
	Pooled Basins	
Inflection	0.050	0.010
50-Point	0.276	0.197
Mirrored Spread	0.497	0.322

In neither the Upper Mississippi basin or the Lower Missouri basin is the null hypothesis, $H_0(\tilde{\varsigma}^2)$, rejected at probability level $\alpha \le 0.05$ in estimating the $50-year$ flood with the Right-Tail Normal distribution. In the Upper Mississippi basin the null hypothesis, $H_0(\tilde{\varsigma}^2)$, is rejected at probability level $\alpha = 0.05$ in the case where the mirrored spread method is used to estimate the $100-year$ flood with the Right-Tail Normal distribution. In the case where the basins are pooled, the null hypothesis, $H_0(\tilde{\varsigma}^2)$, is rejected at probability level $\alpha = 0.05$ in estimating the $50-year$ flood and at probability level $\alpha = 0.01$ in estimating the $100-year$ flood by the inflection point method with the Right-Tail Normal distribution. For neither the estimate of the $50-year$ flood nor the $100-year$ flood is the null hypothesis, $H_0(\tilde{\varsigma}^2)$, rejected at probability level $\alpha \le 0.05$ in using either the 50-point method or the mirrored spread method with the Right-Tail Normal distribution.

Assessment of Means of Differences in Fits

The t values and their corresponding probabilities relating to the variances of the at-site differences in the fits of the Pearson Type III distribution and the Right-Tail Normal Distribution to the observed distribution are given in Table 26a and 26b.

Table 26a: t Values Based on At-Site Fits of the Pearson Type III and the Right-Tail Normal Distributions to Observed Distributions

Method	50-Year Peak Flow	100-Year Peak Flow
	Upper Mississippi Basin	
Inflection	0.490	1.144
50-Point	0.956	2.731
Mirrored Spread	0.774	2.576
	Missouri Basin	
Inflection	0.695	0.639
50-Point	0.105	0.296
Mirrored Spread	0.162	0.012
	Pooled Basins	
Inflection	0.380	0.344
50-Point	0.797	1.514
Mirrored Spread	0.507	1.165

Table 26b: Probability of t Values Based on At-Site Fits of the Pearson Type III and the Right Tail Normal Distributions to Observed Distributions

Method	50-Year Peak Flow	100-Year Peak Flow
	Upper Mississippi Basin	
Inflection	0.633	0.275
50-Point	0.358	0.018
Mirrored Spread	0.454	0.024
	Missouri Basin	
Inflection	0.500	0.535
50-Point	0.918	0.722
Mirrored Spread	0.874	0.990
	Pooled Basins	
Inflection	0.707	0.733
50-Point	0.433	0.142
Mirrored Spread	0.616	0.254

In no case is the null hypothesis, $H_0(\tilde{\xi})$, rejected at probability level $\alpha \leq 0.05$. Though on average, the differences in the estimates of the $T = 50, 100$-year floods yielded by the Pearson Type III and the Right-Tail Normal distribution are not statistically significant at probability level $\alpha \leq 0.05$, the minimum differences in the fits to the observed distributions is dominated by the Right-Tail Normal distribution.

Regional Goodness of Fit

The regional values of $\Delta(PIII)$ and $\Delta(RTN_i)$ conditioned on $T = 50, 100$ are given in Tables 27a and 27b.

Table 27a: Goodness of Fit per $T = 50 - year$ Flow Relative to Regional Sequences of Annual Peak Flows in Log Space

River	Skewness	$\Delta(PIII)$	$\Delta(RTN)$ Inflection Point	$\Delta(RTN)$ 50-Point	$\Delta(RTN)$ Mirrored Spread
		Upper Mississippi Basin			
Log Space	-0.608	0.005	-0.001	0.000*	0.002
Quasi-Log Space	-0.563	0.025	-0.009	-0.000*	0.009
		Missouri Basin			
Log Space	-0.253	0.003	0.007	0.002*	0.003
Quasi-Log Space	-0.272	0.014	0.036	0.010*	0.015

* Minimum Absolute Difference in Fit

Table 27b: Goodness of Fit per $T = 100 - year$ FlowRelative to Regional Sequences of Annual Peak Flows in Log Space

River	Skewness	$\Delta(PIII)$	$\Delta(RTN)$ Inflection Point	$\Delta(RTN)$ 50-Point	$\Delta(RTN)$ Mirrored Spread
		Upper Mississippi Basin			
Log Space	-0.608	0.015	0.007*	-0.010	0.009
Quasi-Log Space	-0.563	0.082	0.045*	-0.047	0.055
		Missouri Basin			
Log Space	-0.253	0.028	0.031	0.026*	0.027
Quasi-Log Space	-0.272	0.140	0.160	0.131*	0.137

* Minimum Absolute Difference in Fit

The fits of the Pearson Type III distributions and the Right-Tail Normal distributions to the regional distributions are shown in Appendix L. The Right-Tail Normal distributions are fitted using the mirrored spread method to estimate the distribution's parameters.

The at-site differences in fits derived through regionalization in log space are given in Tables 28a and 28b, and derived through regionalization in quasi-log space are given in Tables 29a and 29b.

Table 28a: Goodness of Fit per $T = 50-year$ Flow Relative to At-Site Sequences of Annual Peak Flows Derived through Regionalization in Log Space

River	Median	$\Delta(PIII)$	$\Delta(RTN)$ Inflection Point	$\Delta(RTN)$ 50-Point	$\Delta(RTN)$ Mirrored Spread
			Upper Mississippi Basin		
St. Paul	4.600	0.023	-0.005	0.005*	0.009
Winona	4.956	0.025	-0.005	0.005*	0.009
Dubuque	5.119	0.026	-0.005	0.005*	0.010
Clinton	5.125	0.026	-0.005	0.005*	0.010
Keokuk	5.252	0.026	-0.005	0.005*	0.011
Hannibal	5.324	0.027	-0.005	0.005*	0.011
St. Louis	5.694	0.028	-0.006	0.006*	0.011
Average	*5.153*	*0.026*	*-0.005*	*0.005*	*0.010*
Std. Deviation	*0.336*	*0.002*	*0.000*	*0.000*	*0.001*
			Missouri Basin		
Sioux City	5.173	0.016	0.036	0.010*	0.016
Omaha	5.182	0.016	0.036	0.010*	0.016
Nebraska City	5.248	0.016	0.037	0.010*	0.016
St. Joseph	5.239	0.016	0.037	0.010*	0.016
Kansas City	5.329	0.016	0.037	0.011*	0.016
Booneville	5.401	0.016	0.038	0.011*	0.016
Hermann	5.484	0.016	0.038	0.0118	0.016
Average	*5.294*	*0.016*	*0.037*	*0.011*	*0.016*
Std. Deviation	*0.116*	*0.000*	*0.001*	*0.000*	*0.000*
			Pooled Basins		
Average	*5.223*	*0.021*	*0.016*	*0.008*	*0.013*
Std. Deviation	*0.252*	*0.005*	*0.022*	*0.003*	*0.003*

* Minimum Absolute Difference in Fit

Table 28b: Goodness of Fit per $100-year$ Flow Relative to At-Site Sequences of Annual Peak Flows Derived through Regionalization in Log Space

River	Median	$\Delta(PIII)$	$\Delta(RTN)$ Inflection Point	$\Delta(RTN)$ 50-Point	$\Delta(RTN)$ Mirrored Spread
			Upper Mississippi Basin		
St. Paul	4.600	0.069	0.032*	-0.046	0.041
Winona	4.956	0.074	0.035*	-0.050	0.045
Dubuque	5.119	0.077	0.036*	-0.051	0.046
Clinton	5.125	0.077	0.036*	-0.051	0.046
Keokuk	5.252	0.079	0.037*	-0.052	0.047
Hannibal	5.324	0.080	0.037*	-0.053	0.048
St. Louis	5.694	0.085	0.040*	-0.057	0.051
Average	*5.15*	*0.077*	*0.036*	*-0.052*	*0.046*
Std. Deviation	*0.336*	*0.005*	*0.002*	*0.003*	*0.003*
			Missouri Basin		
Sioux City	5.173	0.145	0.160	0.134*	0.140
Omaha	5.182	0.145	0.161	0.135*	0.140
Nebraska City	5.248	0.147	0.163	0.136*	0.142
St. Joseph	5.239	0.147	0.162	0.136*	0.142
Kansas City	5.329	0.149	0.165	0.139*	0.144
Booneville	5.401	0.151	0.167	0.140*	0.146
Hermann	5.484	0.154	0.170	0.143*	0.148
Average	*5.294*	*0.148*	*0.164*	*0.138*	*0.143*
Std. Deviation	*0.116*	*0.003*	*0.004*	*0.003*	*0.003*
			Pooled Basins		
Average	*5.223*	*0.113*	*0.100*	*0.043*	*0.095*
Std. Deviation	*0.252*	*0.037*	*0.066*	*0.098*	*0.050*

* Minimum Absolute Difference in Fit

Table 29a: Goodness of Fit per $T = 50-year$ Flow Relative to At-Site Sequences of Annual Peak Flows Derived through Regionalization in Quasi-Log Space

River	Median	$\Delta(PIII)$	$\Delta(RTN)$ Inflection Point	$\Delta(RTN)$ 50-Point	$\Delta(RTN)$ Mirrored Spread
			Upper Mississippi Basin		
St. Paul	4.600	0.115	-0.041	-0.005*	0.041
Winona	4.956	0.124	-0.045	-0.005*	0.045
Dubuque	5.119	0.128	-0.046	-0.005*	0.046
Clinton	5.125	0.128	-0.046	-0.005*	0.046
Keokuk	5.252	0.131	-0.047	-0.005*	0.047
Hannibal	5.324	0.133	-0.048	-0.005*	0.048
St. Louis	5.694	0.142	-0.051	-0.006*	0.051
Average	*5.153*	*0.129*	*-0.046*	*-0.005*	*0.046*
Std. Deviation	*0.336*	*0.008*	*0.003*	*0.000*	*0.003*
			Missouri Basin		
Sioux City	5.173	0.072	0.186	0.052*	0.078
Omaha	5.182	0.073	0.187	0.052*	0.078
Nebraska City	5.248	0.073	0.189	0.052*	0.079
St. Joseph	5.239	0.073	0.189	0.052*	0.079
Kansas City	5.329	0.075	0.192	0.053*	0.080
Booneville	5.401	0.076	0.194	0.054*	0.081
Hermann	5.484	0.077	0.197	0.055*	0.082
Average	*5.294*	*0.074*	*0.191*	*0.053*	*0.079*
Std. Deviation	*0.116*	*0.002*	*0.004*	*0.001*	*0.002*
			Pooled Basins		
Average	*5.223*	*0.101*	*0.072*	*0.024*	*0.063*
Std. Deviation	*0.252*	*0.029*	*0.123*	*0.030*	*0.017*

* Minimum Absolute Difference in Fit

Table 29b: Goodness of Fit per $100-year$ Flow Relative to At-Site Sequences of Annual Peak Flows Derived through Regionalization in Quasi-Log Space

River	Median	$\Delta(PIII)$	$\Delta(RTN)$ Inflection Point	$\Delta(RTN)$ 50-Point	$\Delta(RTN)$ Mirrored Spread
			Upper Mississippi Basin		
St. Paul	4.600	0.377	0.207*	-0.216	0.253
Winona	4.956	0.406	0.223*	-0.233	0.273
Dubuque	5.119	0.420	0.230*	-0.241	0.282
Clinton	5.125	0.420	0.231*	-0.241	0.282
Keokuk	5.252	0.431	0.236*	-0.247	0.289
Hannibal	5.324	0.437	0.240*	-0.250	0.293
St. Louis	5.694	0.467	0.256*	-0.268	0.313
Average	*5.153*	*0.423*	*0.232*	*-0.242*	*0.283*
Std. Deviation	*0.336*	*0.028*	*0.015*	*0.016*	*0.018*
			Missouri Basin		
Sioux City	5.173	0.724	0.828	0.678*	0.709
Omaha	5.182	0.725	0.829	0.679*	0.710
Nebraska City	5.248	0.735	0.840	0.687*	0.719
St. Joseph	5.239	0.733	0.838	0.686*	0.718
Kansas City	5.329	0.746	0.853	0.698*	0.730
Booneville	5.401	0.756	0.864	0.708*	0.740
Hermann	5.484	0.768	0.877	0.718*	0.751
Average	*5.294*	*0.741*	*0.847*	*0.693*	*0.725*
Std. Deviation	*0.116*	*0.016*	*0.019*	*0.015*	*0.016*
			Pooled Basins		
Average	*5.223*	*0.582*	*0.539*	*0.226*	*0.504*
Std. Deviation	*0.252*	*0.167*	*0.320*	*0.486*	*0.230*

* Minimum Absolute Difference in Fit

On a regional basis, either in log space or quasi-log space, the Right-Tail Normal distribution provides a better fit to the regional distribution than the Pearson Type III distribution. Refer to Tables 27a and 27b above. Through regionalization in log space, the at-site fits are derived by multiplying the regional differences in fits by the at-site medians of the flows at the sites. Thus, the ranking of the regional differences in the fits is identical to the rankings of the derived at-site differences for any given site. Refer to Tables 27a, 28a and 28b above. Similarly, through regionalization in quasi-log space, the ranking of the regional differences in the fits is identical to the rankings of the derived at-site differences for any given site. Refer to Tables 27b, 29a and 29b above.

Assessment of Variances of Differences in Fits

The F values and their corresponding probabilities relating to the variances of the at-site differences in the fits of the Pearson Type III distribution and the Right-Tail Normal distribution derived through regionalization in log space and quasi-log space are given in 30a and 30b.

Table 30a: F Values Based on At-Site Fits of the Pearson Type III and the Right Tail Normal Distributions to Observed Distributions Derived through Regionalization in Log Space and Quasi-Log Space

Method	Log Space		Quasi-Log Space	
	50-Year Peak Flow	100-Year Peak Flow	50-Year Peak Flow	100-Year Peak Flow
		Upper Mississippi Basin		
Inflection	0.040	0.218	0.130	0.301
50-Point	0.040	0.444	0.002	0.329
Mirrored Spread	0.160	0.360	0.130	0.450
		Missouri Basin		
Inflection	5.444	0.001	6.612	1.306
50-Point	0.444	0.862	0.510	0.876
Mirrored Spread	1.000	0.930	1.148	0.958
		Pooled Basins		
Inflection	17.355	0.196	18.016	3.674
50-Point	0.291	7.033	1.083	8.487
Mirrored Spread	0.312	1.837	0.356	1.901

Table 30b: Probability of F Values Based on At-Site Fits of the Pearson Type III and the Right Tail Normal Distributions to Observed Distributions Derived through Regionalization in Log Space and Quasi-Log Space

Method	Log Space		Quasi-Log Space	
	50-Year Peak Flow	100-Year Peak Flow	50-Year Peak Flow	100-Year Peak Flow
		Upper Mississippi Basin		
Inflection	0.999	0.957	0.987	0.915
50-Point	0.999	0.827	1	0.899
Mirrored Spread	0.979	0.880	0.987	0.823
		Missouri Basin		
Inflection	0.029	0.999	0.018	0.377
50-Point	0.827	0.569	0.783	0.562
Mirrored Spread	0.500	0.534	0.436	0.520
		Pooled Basins		
Inflection	0.000	0.997	0.000	0.013
50-Point	0.983	0.001	0.444	0.000
Mirrored Spread	0.978	0.143	0.963	0.130

In neither log space or quasi log space is the null hypothesis, $H_0: \overline{\varsigma}^2$, for the Upper Mississippi basin and the Lower Missouri basin. However, if the basins are pooled, then the null hypothesis is rejected relative to certain methods of estimating the parameters of the Right-Tail Normal distribution. Rejecting the null hypothesis does not weaken the argument that in either log space or quasi-log space, the Right-Tail Normal distribution provides a better it to the regional distribution than the Pearson Type III distribution.

Assessment of Means of Differences in Fits

Table 31a: t Values Based on At-Site Fits of the Pearson Type III and the Right Tail Normal Distributions to Observed Distributions Derived through Regionalization in Log Space and Quasi-Log Space

Method	Log Space		Quasi-Log Space	
	50-Year Peak Flow	100-Year Peak Flow	50-Year Peak Flow	100-Year Peak Flow
		Upper Mississippi Basin		
Inflection	47.713	19.597	51.885	16.040
50-Point	31.808	56.230	42.136	55.343
Mirrored Spread	22.589	13.908	24.417	11.088
		Missouri Basin		
Inflection	63.306	116.739	68.650	11.334
50-Point	33.430	6.309	28.023	5.658
Mirrored Spread	0	3.099	5.874	1.846
		Pooled Basins		
Inflection	0.780	8.228	0.380	0.424
50-Point	7.818	2.395	6.888	2.501
Mirrored Spread	4.626	1.047	0.063	0.504

Table 31b: Probability of t Values Based on At-Site Fits of the Pearson Type III and the Right Tail Normal Distributions to Observed Distributions Derived through Regionalization in Log Space and Quasi-Log Space

Method	Log Space		Quasi-Log Space	
	50-Year Peak Flow	100-Year Peak Flow	50-Year Peak Flow	100-Year Peak Flow
		Upper Mississippi Basin		
Inflection	0.000	0.000	0.000	0.000
50-Point	0.000	0.000	0.000	0.000
Mirrored Spread	0.000	0.000	0.000	0.000
		Missouri Basin		
Inflection	0.000	0.000	0.000	0.000
50-Point	0.000	0.000	0.000	0.000
Mirrored Spread	1	0.009	0.000	0.009
		Pooled Basins		
Inflection	0.433	0.000	0.410	0.675
50-Point	0.000	0.024	0.000	0.019
Mirrored Spread	0.000	0.143	0.017	0.230

In both log space and quasi-log space, the null hypothesis is rejected for the Upper Mississippi basin and the Missouri basin. The differences in the means of the differences in the fits of the regional distribution by the Pearson Type III distribution and the Right-Tail Normal distribution cannot be ascribed to chance. Thus, the argument that the Right-Tail Normal distribution provides a better fit to the regional distribution, and consequently to the at-site distributions derived through regionalization, than the Pearson Type III distribution is strengthen. If the basins are pooled, then the null hypothesis is not rejected with certain methods used to estimate the parameters of the Right-Tail Normal distribution. In any case, an argument can be made in favor of the Right-Tail Normal distribution over the Pearson Type III distribution.

Postscript

Central to the above discussions of trend and persistence and of risk and uncertainty was the assumption iid assumption, the assumption that a sequence of n annual flood flows is a realization of a sequence of n independent and identically distributed random variables. An assessment of trend and persistence or of risk and uncertainty on a site by site basis, whether on at-site terms or on regionalized terms is not complete without some account of the covariance structure of the N flow sequences used in the assessment. The covariance structure defines the structure of linear correlation among the flow sequences. In a large region, if N is small and the sequence sites are widely dispersed, the covariance/correlation structure likely would not have a significant effect upon the assessment, except perhaps if all sites are on a single river course.

Recently, Douglas et al (2000) took direct account of spatial correlation – covariance/correlation structure – in assessment of trends in floods and low flows. An account of the covariance/correlation structure general presumes a specific multivariate distribution. In most cases, it is presumed that the multivariate distribution is Normal in either real space or log space. The assumption of normality provides greater depth and scope in hydrologic analyses than any other assumption of the multivariate distribution given the extensive development of the multivariate Normal distribution.

The Right-Tail Normal distribution introduced above as a univariate distribution of annual floods can be generalized as a multivariate Right-Tail Normal distribution to accommodate flood studies on a regional basis. Nonetheless, non-Normal multivariate distributions of annual floods and for other hydrologic phenomena, e.g. annual low floods, any arbitrary element of the spectrum of extreme flows, are important to hydrology. Various statistical techniques have been developed for generating variate values for bivariate distributions having specified marginal distributions. A technique developed by Johnson (1978) is presented in Appendix M. The technique is used to generate bivariate sequences of values having Pearson Type III, say in log space, and Log Pearson Type III, say in real space, marginal distributions.

The assumption of normality has its advantages not only in dealing with the spatial covariance/correlation structure, but also in dealing with the temporal covariance/correlation structure, i.e. in time series analyses.

The objectives of water resources development, or of any developmental enterprize, are future expectations, more specifically future economic expectations. Because the future cannot be totally comprehended, the decisions to effect future expectations are almost always made in a state of uncertainty. The more distant the future is from the present, the less the future will mirror the past. The uncertainty that arises from the future not mirroring the past is referred to by Davidson (1991) as true uncertainty. It is problematic as to whether or not true uncertainty may be substituted for by probability.

The matter of true uncertainty has not been explored in the field of water resources. It can be argued that the definition of uncertain in the Principles and Guidelines (U.S. Water Resources Council: 1983) is in effect a definition of true uncertainty. However, the wording does not make it clear that that is indeed the case. The Council states that uncertainty is an integral part of water resources development, implying that uncertain must be addressed in the course of water resources investigation. Even if the Council's definition of uncertainty is a definition of true uncertainty, then uncertainty (Council's definition) must be measured by a metric other than probability.

It remains to be seen if true uncertainty underlines the development of water resources, and if so how true uncertainty is to be measured. Is Shackle's (1949) index of surprise a meaningful measure of true uncertainty in water resources investigation? Is that that the uncertainty in water resources investigation is not true uncertainty, but uncertainty that may be substituted for by probability?

References

Abramowitz, Milton and Irene A. Stegun, 1964, Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables, U.S. Government Printing Office, Washington, D.C. , pp. 255-258.

Davidson, Paul, 1991, Is Probability Theory Relevant for Uncertainty? A Post Keynesian Perspective, Journal of Economic Perspectives, vol. 5, no. 1, pp. 129-143.

Douglas, E.M., R.E. Vogel and C.N. Kroll, 2000, Trends in Floods and Low Flows: Impact of Spatial Correlation, Journal of Hydrology, 240, pp. 90-105.

Fisher, R.A., 1922, On the Interpretation of χ^2 from Contingency Tables and Calculation of P, Journal of the Royal Society of London, Series A, vol. 85, pp.87-94.

Fuller, W.E., 1914, Flood Flows, Trans. Am. Soc. Civil Engr., vol. 77, pp. 564-617.

Galambos, Janos, 1978, The Asymptotic Theory of Extreme Order Statistics, John Wiley & Sons, New York, 352 pp.

Hazen, Allen, 1914, Storage to be Provided in Impounding Reservoirs for Municipal Water Supply, Trans. Am. Soc. Civil Engr., vol. 77, Paper no. 1308, pp. 1539-1640.

Interagency Committee on Water Data, 1981, (Bulletin 17-B of the Hydrology Committee originally published bt the U.S. Water Resources Council), Guidelines for Determining Flood Flow Frequency. U.S. Government Printing Office, Washington, D.C.

Johnson, Norman L. and Samuel Kotz, 1969, Discrete Distributions, Houghton Mifflin Company, Boston, Mass., pp. 50-86

Johnson, Norman L. and Samuel Kotz, 1970, Continuous Univariate Distributions-1, Houghton Mifflin Company, Boston, Mass. pp. 176-181.

Johnson, Mark E., 1979, Bivariate Distributions with given Marginals and Fixed Measures of Dependence,

Kendall, M.G. and Alan Stuart, 1966, The Advanced Theory of Statistics, pp. 349-350, Hafner Publishing Co., New York.

Kenny, J.F. and E.S. Keeping, 1956, Mathematics of Ststistics, Part 2, D. Van Nostrand, Co., Inc., Princeton, N.J., pp. 22-48.

Matalas, N.C. and M.B. Fiering, 1977, Water Resources Systems Planning, in Climate, Climate Change and Water Supply, Chap. 6, National Research Council, National Academy of Sciences, Washington D.C.

Olsen, J. Rolf, Jery R. Stedinger, Nicholas C. Matalas and Eugene Z. Stakhiv, 1999, Climate Variability and Flood Frequency Estimation for the Upper Mississippi and Lower Missouri Rivers, Journal American Water Rescources Association, vol. 35, no. 6, pp. 1509-1523.

Planning & Management Consultants, Ltd., 1999, Flood Frequency Analysis in the Upper Mississippi and Missouri Basins, prepared by Nicholas C. Matalas, subcontractor, 103 pp.

Rafter, George W., 1895, Hudson River Storage – Second Report on Survey of Upper Hudson Valley, Annual Report of the State Engineer and Surveyor, Appendix 8, pp. 803-858, Albany, New York

Rescher, Nicholas, 1983, Risk: A Philosophical Introduction to the Theory of Risk Evaluation and Management, University Press of America, 208 pp.

Shackle, G.L.S., 1949, Expectations in Economics, Cambridge University Press, London, 146 pp.

Slack, J.R., A.M. Lumb and J.M. Landwehr, 1993, Hydro-Climatic Data Network (HCDN), Water-Resource

Thomas, Jr., Harold A., 1957, Personal Communication.

U.S. Water Resources Council, 1983, Economic and Environmental Principals and Guidelines for Water and Related Resource Implementation Studies,

Uspensky, J.V., 1937, Introduction to Mathematical Probability, McGraw-Hill, New York.

Wadsworth, George P. and Joseph G. Bryan, 1960, Introduction to Probability and Random Variables, McGraw-Hill Book Co., Inc., New York, pp. 111-112.

Wilson, E.B. and M.M. Hilferty, 1931, The Distribution of Chi-Square, Proceedings of the National Academy of Sciences, Washington, D.C., pp. 684-688.