Risk Analysis, Vol. 18, No. 4, 1998

# Risk of Extreme Events Under Nonstationary Conditions

J. Rolf Olsen,[1] James H. Lambert,[1] and Yacov Y. Haimes[1,2]

The concept of the return period is widely used in the analysis of the risk of extreme events and in engineering design. For example, a levee can be designed to protect against the 100-year flood, the flood which on average occurs once in 100 years. Use of the return period typically assumes that the probability of occurrence of an extreme event in the current or any future year is the same. However, there is evidence that potential climate change may affect the probabilities of some extreme events such as floods and droughts. In turn, this would affect the level of protection provided by the current infrastructure. For an engineering project, the risk of an extreme event in a future year could greatly exceed the average annual risk over the design life of the project. An equivalent definition of the return period under stationary conditions is the expected waiting time before failure. This paper examines how this definition can be adapted to nonstationary conditions. Designers of flood control projects should be aware that alternative definitions of the return period imply different risk under nonstationary conditions. The statistics of extremes and extreme value distributions are useful to examine extreme event risk. This paper uses a Gumbel Type I distribution to model the probability of failure under nonstationary conditions. The probability of an extreme event under nonstationary conditions depends on the rate of change of the parameters of the underlying distribution.

**KEY WORDS:** Extreme events; nonstationary conditions; climate change; return period; risk-based engineering.

## 1. INTRODUCTION

The art and science of systems modeling have their virtues and vices. On the one hand management decisions of complex systems are made on sound and rational bases through modeling and analyses. At the same time, to apply modeling tools and methodologies, the analyst must adhere to fundamental simplifications of the real system under study. One can hardly refute the fact that most, if not all, technologically-based systems follow a pattern which is nonlinear, dynamic, probabilistic, and distributed in nature. For many reasons, however, such complexities rarely are fully accommodated in systems modeling. Furthermore, even when probabilistic models are developed, stationary behavior of the random variables is commonly assumed. Given the dynamic nature of our constantly changing world due to manmade or non-manmade effects, it seems unreasonable to assume that such random variables are stationary. The major impact of such nonstationary behavior is not manifested through its central tendency; rather, the major impact is realized through the extremes. Extreme high and low temperatures can affect agriculture, energy use, and human health. Floods can cause large amounts of damage. Droughts can harm agriculture and reduce water supplies. Nonstationarity may affect both the severity and frequency of these extreme events.

One example of current concern regarding nonstationarity and extreme events is the effect of climate change on water resources. The management of water

[1] Center for Risk Management of Engineering Systems, University of Virginia, Charlottesville, Virginia 22903.

[2] To whom all correspondence should be addressed at Center for Risk Management of Engineering Systems, Thornton Hall, University of Virginia, Charlottesville, Virginia 22903.



0272-4332/98/0800-0497$15.00/1 

resources and the design of water projects could be affected by a rise in global temperatures. For example, higher temperature leads to faster evaporation, since the capacity of air for evaporated water rises about 6% per degree Celsius.[1] Increased evaporation could lead to increased precipitation in some areas of the world. On the other hand, it could also lead to reduced soil moisture and decreased runoff. Climate change could therefore have a significant effect on water resources, especially on the occurrence of extreme events such as floods and droughts. The possibility of global climate change and its effect on extreme events could become a major concern for decision makers. New methods need to be developed to show the implications of nonstationarity to the management of extreme events such as floods and drought.

Climate variables can be considered to be random variables with a probability distribution $F(x) = Pr\{X < x\}$. A standardized random variable is defined as $Z = \frac{X - \mu}{\sigma}$, where $\mu$ is the location parameter and $\sigma$ is the scale parameter. Katz[2] suggests that climate change potentially involves both a shifting of the distribution (a change in the location parameter) and a rescaling of the distribution function (a change in the scale parameter) over time. Either of these changes would affect the frequency of extreme events as represented in the tails of the distribution. Mearns *et al.*[3] and Katz and Brown[4,5] have examined the possible effects of climate change on extreme climate events. Wigley[6] used the definition of the return period as the expected waiting time to look at the effect of nonstationary conditions on the risk of extreme events. He assumed that the underlying distribution of the variables was a normal distribution, the variables were independent, and there was a linear increasing trend in the mean with time. He then used a stochastic simulation to find the expected number of years before a failure occurred.

If climate change is occurring, precipitation patterns may be changing, which in turn may change the probability distributions for runoff. If $X_i$ is the observation of the largest flood in year $i$, the probability distribution of $X_i$ may be changing as the frequency and amount of rainfall changes. This paper considers how trends in the annual peak flow will affect the probability of extreme events occurring. We will first consider how the return period is affected by a trend in a parameter of the underlying distribution. Second, we will look at how limiting distributions can be used for sequences where there is a trend in the mean of the underlying distribution. For both analyses, the trend is assumed to be increasing. However, since the effect of a possible climate change on hydrologic trends is unknown, this paper intends to show how measures of risk are affected if such a trend were present.

## 2. THE RETURN PERIOD UNDER NONSTATIONARY CONDITIONS

### 2.1. Definitions of Return Period

The concept of the return period is often used in the analysis of extreme events and in engineering design. For example, a flood control structure such as a levee can be designed to protect against the 100-year flood, the flood which on average occurs once in 100 years. The designer wants to know the probability of occurrence of the flood which would overtop the structure. The minimum flood which causes damage would be the flood exceeding a threshold $c$. Let $X_i$ be the maximum flood in year $i$. Under stationary conditions, each $X_i$ is assumed to be independently and identically distributed with a cumulative distribution function given by $F_X(x)$. For an $n$ year period, the maximum flood is $Y_n$, where $Y_n = \max\{X_1, X_2, \ldots, X_n\}$. The probability that $Y_n$ exceeds $c$ could be used to estimate this probability. In this case, $X_i$ would be the annual maximum peak flow. In any year, it is possible that several floods could exceed $c$. However, if the probability of exceeding $c$ is very low, then the probability of exceeding $c$ multiple times in a given year is low.

In the statistics of extremes, the magnitude of an extreme event with a return period of $n$ years is identical to the characteristic largest value $u_n$. The characteristic largest value $u_n$ is the value of a random variable which has an annual exceedance probability of $1/n$.[7,8] The characteristic largest value $u_n$ is the value of $X$ such that in a sample of size $n$, the expected number of sample values larger than $u_n$ is one[7–13]:

$$n[1 - F_X(u_n)] = 1 \tag{1a}$$

or

$$1 - F_X(u_n) = \frac{1}{n} \tag{1b}$$

An equivalent definition of the return period is the expected waiting time to an extreme event.[6,8] Consider that if conditions are stationary, the probability of failure $p$ is the probability that a threshold will be exceeded in any year. The probability of failure is the inverse of the return period, $p = 1/n = 1 - F_X(u_n)$, where $n$ is the return period in years. If $k$ is the actual number of years before the extreme event occurs, the probability distribution for $k$ is

$$\begin{aligned} p_k &= Prob[\text{first extreme event occurs in year } k] \\ &= (1-p)^{k-1}p, \; k = 1, 2, \ldots \end{aligned} \tag{2}$$

The return period can be interpreted as the expected value of the number of years before a failure occurs:

$$E[k] = \sum_{k=1}^{\infty} k p_k \tag{3}$$

Under stationary conditions, $E[k]$ equals $1/p$[6]:

$$E[k] = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p = \frac{1}{p} \tag{4}$$

The number of exceedances of $c$ in $n$ years is a binomial random variable with probability of exceedance $[1 - F_X(c)]$ and number of trials $n$. Let $N(x_n)$ be the number of exceedances of $x_n$ where $\{x_n\}$ is a sequence of real numbers. If $\{x_n\}$ satisfies the condition

$$\lim_{n\to\infty} \{n[1 - F_X(x_n)]\} = \tau \text{ and } 0 < \tau < \infty \tag{5}$$

then the number of exceedances $N(x_n)$ can be modeled as a Poisson random variable for a large sample size $n$. The probability distribution of $N(x_n)$ for large sample size $n$ is

$$\lim_{n\to\infty} P[N(x_n) = r] = \frac{\exp(-\tau)\tau^r}{r!} \tag{6}$$

This result is based on the Poisson distribution being the limit for a binomial distribution as $n$ gets large.[8] Leadbetter *et al.*[14] and Falk *et al.*[15] have looked at the application of Poisson processes to extreme value theory. The expected arrival time of the first event of the Poisson process is another way to consider the waiting time before failure of the system.

In the next section, we will build on Wigley's work by examining the definition of the return period as the expected waiting time before failure and its implications under nonstationary conditions. The model assumes that the random variables are independent from year to year but that there is a trend in either the mean or the variance of the distribution. This analysis also assumes that the first and second moments of the distributions exist. This assumption may not hold for the Frechet distribution (Type II).

### 2.2. Model of an Independent Sequence of Random Variables

Again let $X_i$ be the maximum flood in year $i$, but each $X_i$ has a cumulative distribution function given by $F_{X,i}(x)$. Assuming that the $X_i$ are independent each year, the cumulative distribution function of $Y_n$ is given by

$$F_{Y,n}(y) = [F_{X,1}(y)]\,[F_{X,2}(y)] \ldots [F_{X,n}(y)] \tag{7}$$

The probability of a failure exceeding a threshold of $c$ occurring in the $n$ year period is $1 - F_{Y,n}(c)$:

$$R = 1 - ([F_{X,1}(c)]\,[F_{X,2}(c)] \ldots [F_{X,n}(c)]) \tag{8}$$

### 2.3. Return Period Under Nonstationary Conditions

Consider the definition of the return period as the expected number of years before the occurrence of a failure. Using a binomial distribution to model the probability of failure in each year, the probability $p$ of an event occurring in a year is no longer constant if climate change is occurring. The probability that a failure first occurs in $k$ years starting from year $t$ is

$$\begin{aligned} p_k(t) = {} & [F_{X,k+t-2}(c)]\,[F_{X,k+t-3}(c)] \ldots \\ & [F_{X,t+1}(c)]\,[F_{X,t}(c)]\,[1 - F_{X,t+k-1}(c)] \\ & k = 1, 2, \ldots; \qquad t = 1, 2, \ldots \end{aligned} \tag{9a}$$

$$\begin{aligned} p_k(t) = {} & \prod_{i=t}^{k+t-2} [F_{X,i}(c)]\,[1 - F_{X,k+t-1}(c)] \\ & k = 1, 2, \ldots \qquad t = 1, 2, \ldots \end{aligned} \tag{9b}$$

Substituting $p_k(t)$ into Eq. (3), the expected number of years from year $t$ before a failure occurs becomes

$$\begin{aligned} E_t[k] = {} & \sum_{k=1}^{\infty} \Big\{k\Big(\prod_{i=t}^{k+t-2} [F_{X,i}(c)]\Big) \\ & [1 - F_{X,k+1}(c)]\Big\} \qquad t = 1, 2, \ldots \end{aligned} \tag{10}$$

Another possible model of the non-stationarily is to consider the number of failures to follow a nonstationary or nonhomogeneous Poisson process and determine the expected arrival time of the first failure. Kulkarni[16] discusses these processes and calculation of the event times. This paper will only consider the binomial distribution in the calculation of the expected waiting time.

An alternative way to look at the probability of extreme events under nonstationary conditions is to note that the flood associated with an $n$-year return period is also changing each year. Let $u_n(t)$ be the magnitude of the flood with an expected waiting time of $n$ years start-

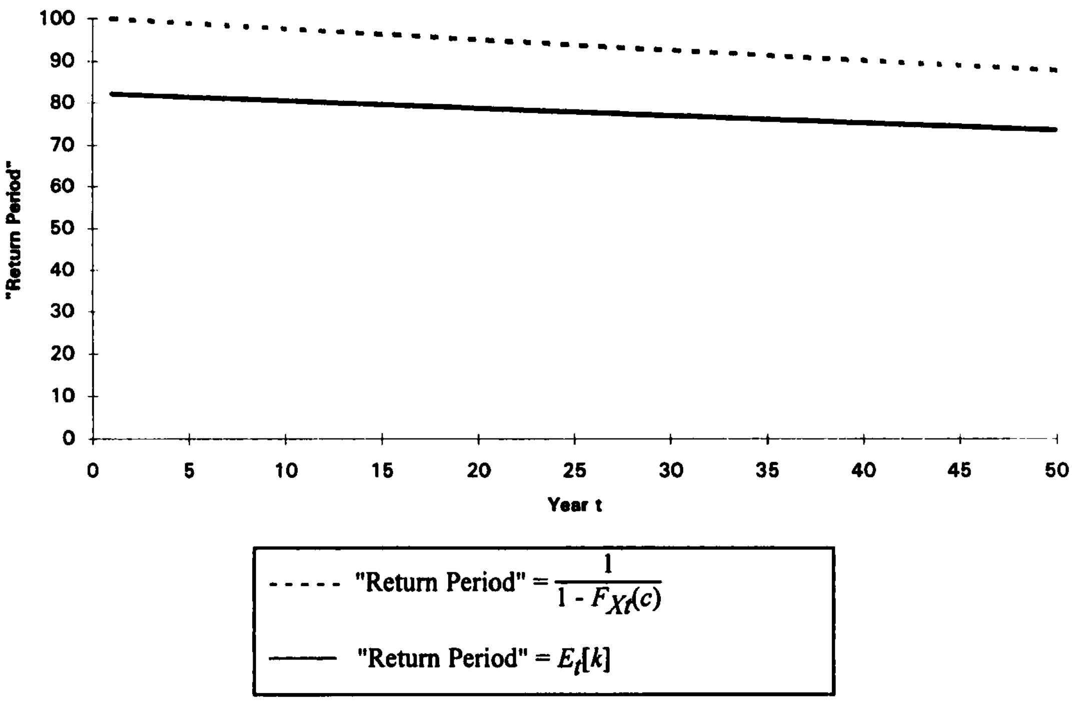


**Fig. 1.** Two definitions of return periods for a linear change in the mean each year of 0.001 standard deviation units.

ing in year $t$. Thus, $u_n(t)$ is the magnitude of the flood corresponding to the equation

$$n = \sum_{k=1}^{\infty} \left\{ \left(k \prod_{i=t}^{k+t-2} [F_{X,i}(u_n(t))]\right) [1 - F_{X,k+t-1}(u_n(t))] \right\} \tag{11}$$

If the expected waiting time before a failure is decreasing each year, then $u_n(t)$ would be increasing. In some future year $i$ years from year $t$, the probability of exceedance of a flood of magnitude $u_n(t)$ could greatly exceed $1/n$:

$$1 - F_{X,t+i}\,(u_n(t)) > (1/n) \tag{12}$$

If the probability of exceeding a threshold is increasing, the expected waiting time until a failure occurs will be less than the inverse of the probability of failure in a given year. In order to demonstrate the differences in the two definitions under nonstationary conditions, the expected waiting time was calculated numerically by summing over a large number of years $k$ and assuming that the trend would continue for every year. Figure 1 depicts the two alternative definitions of the return period for nonstationary conditions with an underlying normal distribution with unit variance and zero mean initially, but where the mean is changing each year by just 0.001 standard deviation units. Because it was assumed that the trend continued forever, the expected waiting time before failure in the first year is less than the inverse of the probability throughout the 50-year life of the project. Initially, in year $t = 1$, the inverse of the probability of exceeding the threshold is 100 years, while the expected waiting time is about 82 years. After 50 years, the $1/p$ method gives a return period of about 88 years, while the expected waiting time is about 74 years. Figure 2 shows the return periods for a change in the standard deviation of 0.001 standard deviation units per year. Initially in year $t = 1$, $1/p = 100$ years, while the expected waiting time is about 72 years. In year $t = 50$, the return period given by the inverse probability method is 75 years, while the expected waiting time is 58 years.

### 2.4. Variance of the Waiting Time Under Nonstationary Conditions

The variance of the return period is one measure of the spread in the uncertainty of the return period. Under stationary conditions, the variance is

$$\mathrm{Var}(k) = \frac{1 - p}{p^2} \tag{13}$$

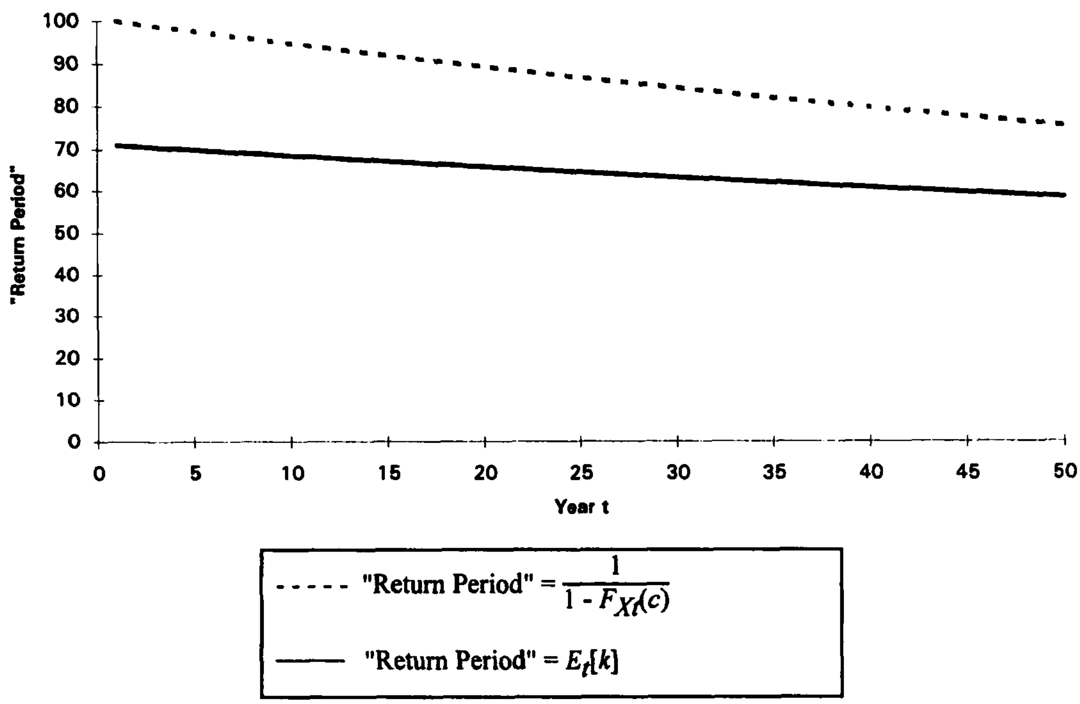


**Fig. 2.** Two definitions of the "return period" for a linear change in the standard deviation each year of 0.001 standard deviation units.

If conditions are not stationary, the variance of the waiting time before failure can be calculated by

$$\mathrm{Var}_t(k) = E_t\left[(k - E_t(k))^2\right] \qquad t = 1, 2, \ldots \tag{14a}$$

or

$$\mathrm{Var}_t(k) = \sum_{k=1}^{\infty} \{[k - E_t(k)]^2 \left(\prod_{i=t}^{k+t-2} [F_{X,i}(c)]\right) [1 - F_{X,k+t-1}(c)]\} \quad t = 1, 2, \ldots \tag{14b}$$

As the mean or standard deviation of the underlying distribution increases, the probability that a random variable will exceed a threshold also increases. The higher probability of failure occurring earlier will reduce the variance of the waiting time before failure, since more of the failures will occur earlier and closer to the expected time before failure and fewer failures will occur in later years where there is a greater distance from the mean.

Figure 3 is a graph of two different measures of the standard deviation of the expected time before failure of a random variable with an increasing trend in the mean of the underlying distribution plotted as a function of the year $t$. Figure 4 is is a similar graph where there is an increasing trend in the standard deviation of the underlying distribution. The standard deviation $sd(k)$ is defined as

$$sd(k) = \sqrt{\mathrm{Var}(k)} \tag{15}$$

When the return period is defined as the inverse of the probability of failure in any year, Eq. (13) is used to calculate the variance. Equation (14) is used to calculate the variance of the expected waiting time before failure. For an increasing trend, the standard deviation is lower when Eq. (14) is used, due to the higher probability of failure in earlier years.

Another view of the uncertainty can be seen from examining the coefficient of variation, defined as the ratio of the standard deviation of a random variable to its mean. Figures 5 and 6 show the coefficient of variation for the two methods for trends in the mean and standard deviation, respectively. For the inverse probability method, the coefficient of variation is almost constant for each year. However, for the expected waiting time before failure, the coefficient of variation is increasing, showing that the standard deviation is not decreasing by as much as the mean. This result indicates that the decrease in the variance is due to the decrease in the expected waiting time. The uncertainty in the expected waiting time, however, is increasing relative to the mean over time.

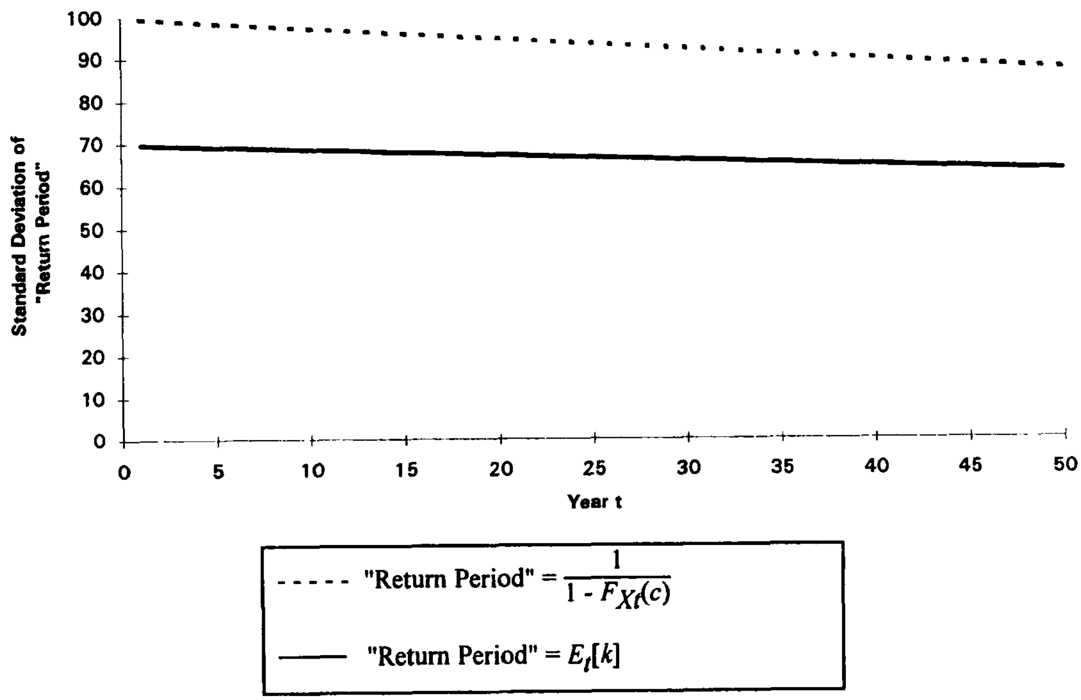


Fig. 3. Standard deviation of the "return period" using two alternative definitions for a linear change in the mean each year of 0.001 standard deviation units.

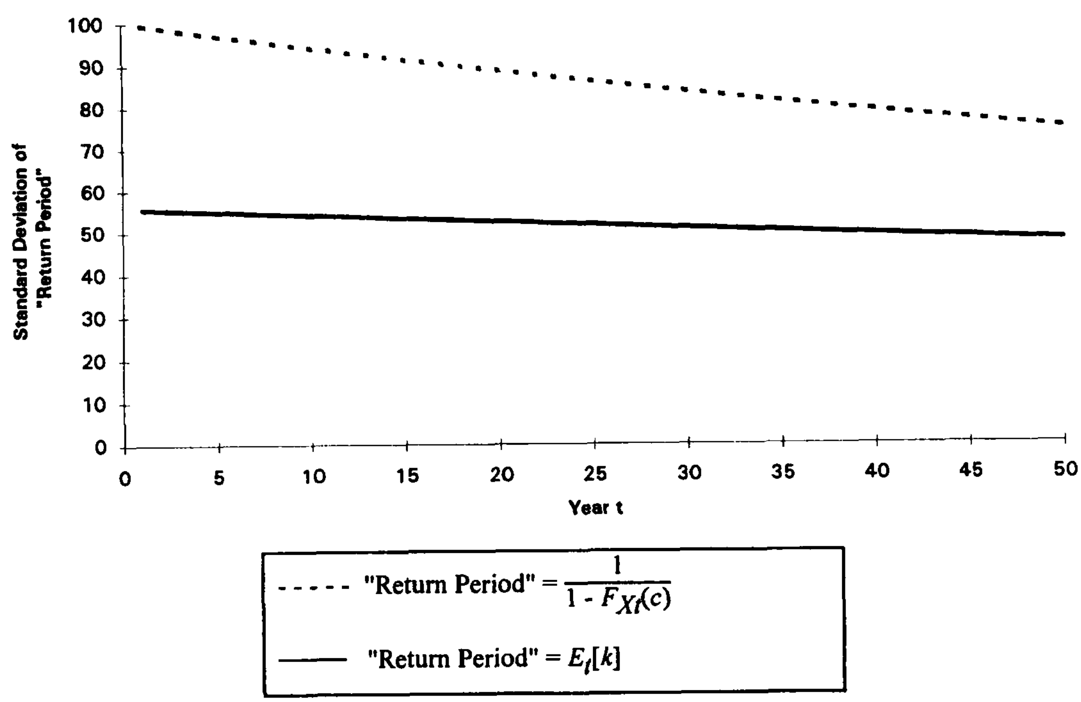


Fig. 4. Standard deviation of the "return period" using two alternative definitions for a linear change in the standard deviation each year of 0.001 standard deviation units.

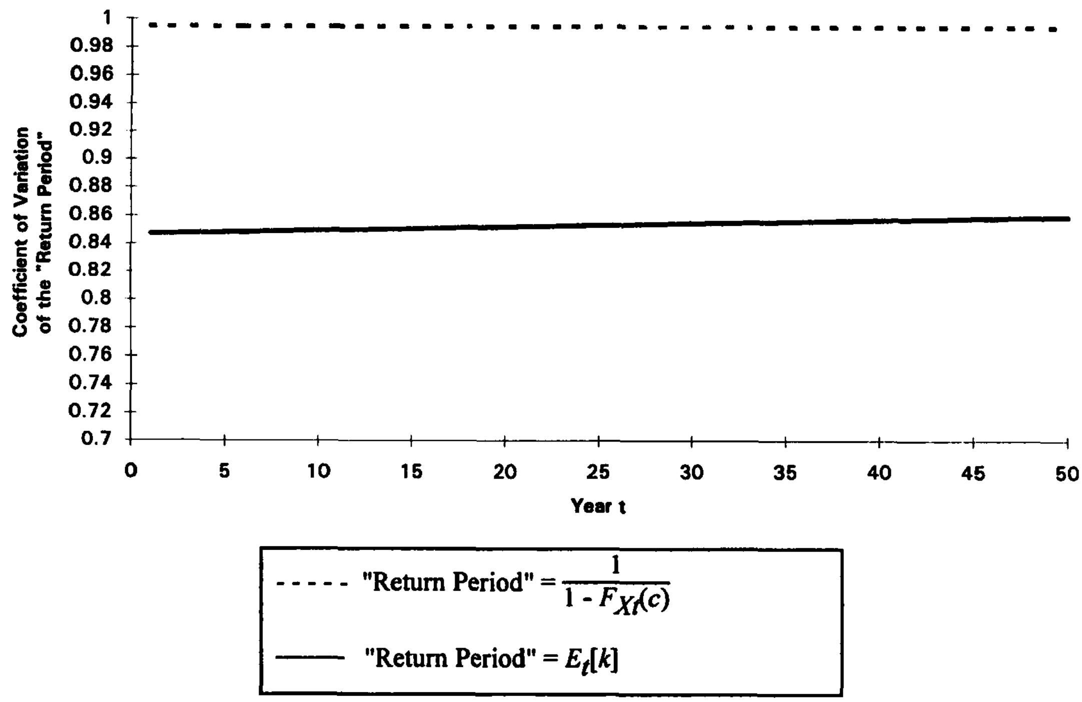


**Fig. 5.** Coefficient of variation of the "return period" using two alternative definitions for a linear change in the mean year of 0.001 standard deviation units.

## 3. LIMITING DISTRIBUTIONS UNDER NONSTATIONARY CONDITIONS

### 3.1. Background

It is often difficult to know the exact probability distribution to represent, for example, the potential for floods. However, according to the statistics of extremes, there are only three possible families of distributions for the maximum of a sequence of random variables. It is useful to know the family or domain of attraction to which the maximum value would belong. Although these three distributions are possible distributions to estimate the probability of extreme events, an implicit assumption in their use is that the number of observations approaches infinity. However, in considering hydrologic extremes for example, the number of observations is finite. Therefore, other distributions such as the Log-Pearson III are often used to estimate flood probabilities.[17]

If $Y_n$ is the maximum of n observations: $Y_n = \max \{X_1, X_2, \ldots, X_n\}$, and the $X_i$ are independent and identically distributed random variables with a known initial distribution function, $F_X(x)$, the exact distribution of $Y_n$ can be determined for a given sample of $n$ observations:

$$F_{Y_n}(y) = [F_X(y)]^n \tag{16}$$

As $n$ becomes large, the probability distribution of $Y_n$ depend only on the tail of the initial distribution. The asymptotic forms of the extreme value distribution are useful in approximating the risk of an extreme event if the parent distribution is unknown or if the sample size is large but unknown. A cumulative distribution function $F(x)$ belongs to a domain of attraction (asymptotic form) for maxima if

$$\lim_{n\to\infty} H_n(a_n x + b_n) = \lim_{n\to\infty} F^n(a_n x + b_n) = H(x) \tag{17}$$

is satisfied for sequences of normalizing constants $\{a_n > 0\}$ and $\{b_n\}$, and $H(x)$ is an extreme value distribution. There are only three types of nondegenerate asymptotic extreme value distributions for the maximum[7–13]:

Type I (Gumbel): (18a)

$$H(x) = \exp(-e^{-x}) \qquad -\infty < x < \infty$$

Type II (Frechet): (18b)

$$H(x) = 0 \qquad x \le 0$$

$$= \exp(-x^{-\alpha}) \quad \text{for some } \alpha > 0 \qquad x > 0$$

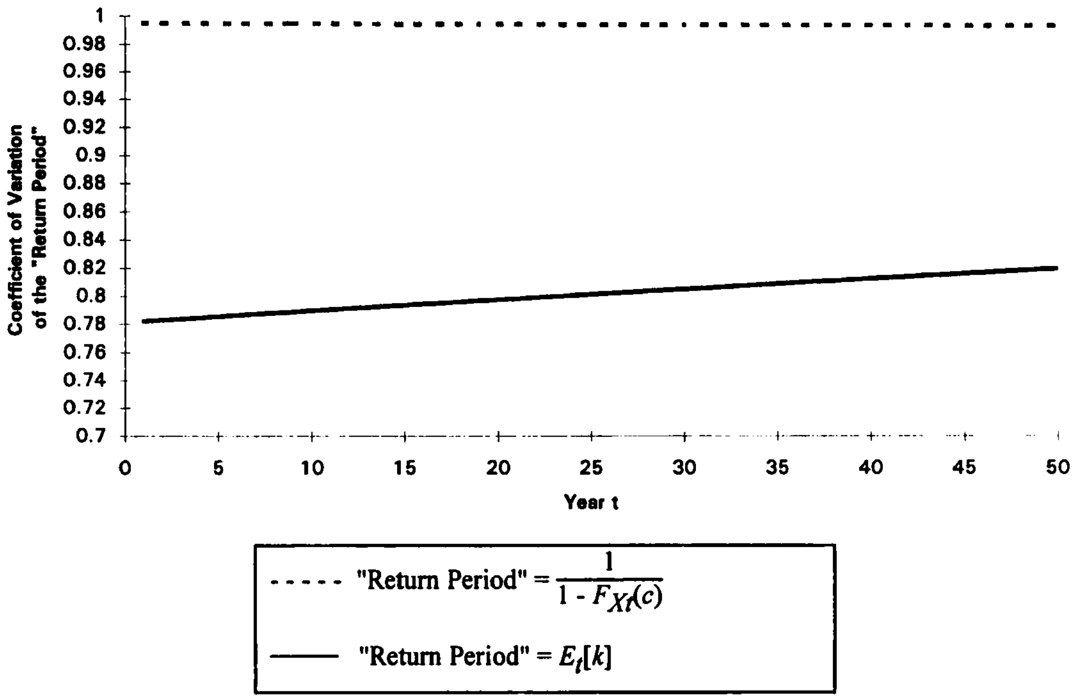


**Fig. 6.** Coefficient of variation of the "return period" using two alternative definitions for a linear change in the standard deviation each year of 0.001 standard deviation units.

Type III (Weibull): (18c)

$$H(x) = \begin{cases} \exp\left(-(-x)^{\alpha}\right) & x \le 0 \\ 1, & x > 0 \end{cases}$$

for some $\alpha > 0$. The limiting distributions for the minima of $X_i$ can be obtained by making the transformation $\overline{X}_i = -X_i$

When determining the limiting distributions for the maximum or minimum of random variables $X_i$, it is often assumed that the observations of the $X_i$s are independent and identically distributed. However, the limiting distribution for the $X_i$s may be one of the three Gumbel types even if these two assumptions do not hold. The limiting extreme value distribution is first determined for an independent sequence of random variables with a trend in the distribution. The assumption of independence is then relaxed to allow for some dependence among the random variables in a sequence.

### 3.2. Domain of Attraction for an Independent Sequence of Random Variables

The domain of attraction for $F_{Y,n}(y)$ can be determined by noting that $F_{Y,n}(y)$ is the product of $n$ cumulative distribution functions (Eq. 7). If the extreme value types of $F_{X,i}(y)$, $i = 1, \ldots, n$, are known, then the extreme-value type of $F_{Y,n}(y)$ can often be determined based on Resnick.[18] The product of Type II extreme-value functions is of a Type II function, and the product of Type III extreme-value functions is of a Type III. Furthermore, the product of Type I and Type II functions is a Type II, and the product of Type I and Type III functions is a Type III. The case of the products of Type I functions is more difficult. The product of Type I functions need not belong to any of the Types I, II, or III. $F_{Y,n}(y)$ would be Type I if the original distribution functions are tail equivalent.[18] Two distribution functions $F(x)$ and $G(x)$ are said to be right tail equivalent, if and only if,

$$\omega(F) = \omega(G) \tag{19}$$

and

$$\lim_{x \to \omega(F)} \frac{1 - F(x)}{1 - G(x)} = 1$$

where $\omega(F)$ is defined as the upper end point of the distribution.[8]

### 3.3. Extremes of Nonstationary Sequences

Another approach is to consider the underlying random variables as a sequence from a time series with some autocorrelation, rather than as independent observations. Climate and hydrologic variables are often autocorrelated where the variables show some interdependence with the preceding variables in the sequence. According to Matalas,[17] flood sequences show little serial correlation, except for the discharges from large lakes. The serial correlation ($\rho$) of low flows ($L$) is greater than the serial correlation of mean flows ($M$) which is greater than that of flood flows ($F$): $\rho(L) > \rho(M) > \rho(F)$.

Let $\{\xi_n\}$ be a sequence of stationary normal variables that are correlated, so that $E(\xi_i) = 0$, $\mathrm{Var}(\xi_i) = 1$ and $\mathrm{Cov}(\xi_i, \xi_j) = r_{ij}$. Define $M_n$ as the maximum of this sequence

$$M_n = \max(\xi_1, \ldots, \xi_n) \tag{20}$$

If some restriction on the dependence between $\xi_i$ and $\xi_j$ is assumed, Leadbetter *et al.*[14] have shown that the domain of attraction of $M_n$ is still a Type I. Leadbetter *et al.* assume that $|r_{ij}| < \rho_{|i-j|}$ for $i \neq j$ and $\rho_n \log n \to 0$ and $\rho_n < 1$.[14] This assumption implies that the dependence between the variables becomes small as the number of observations increases. The normalizing constants $a_n$ and $b_n$ are the same as in the independent case for stationary random variables:

$$P\{a_n(M_n - b_n) \leq x\} \to \exp(-e^{-x}) \tag{21a}$$

where

$$a_n = (2 \log n)^{1/2} \tag{21b}$$

and

$$b_n = (2 \log n)^{1/2} - (1/2)(2 \log n)^{-1/2} (\log \log n + \log 4\pi) \tag{21c}$$

This analysis can also be applied to a nonstationary sequence consisting of a deterministic trend and a stationary normal sequence. Following Leadbetter *et al.*,[14] let $(\eta_1, \eta_2, \ldots, \eta_n)$ be a normal sequence where $\eta_i = m_i + \xi_i$, where $m_i$ are added deterministic components and $\xi_i$ is a stationary normal sequence with $E(\xi_i) = 0$, $\mathrm{Var}(\xi_i) = 1$ and $\mathrm{Cov}(\xi_i, \xi_j) = r_{ij}$. Let

$$M_n = \max\{\eta_l, \eta_2, \ldots, \eta_n\} \tag{22}$$

Some assumptions must also be made concerning the deterministic trend. Leadbetter *et al.* assume that the $m_i$ are such that

$$\beta_n = \max |m_i| = o((\log n)^{1/2}) \text{ as } n \to \infty^{(14)} \tag{23}$$

This assumption includes the cases where the $m_i$s are bounded. Leadbetter *et al.* prove that the asymptotic distribution of the maximum is a Type I distribution:

$$P\{a_n(M_n - b_n - m_n^*) \leq x\} \to \exp(-e^{-x}) \tag{24a}$$

where $a_n$ and $b_n$ are the same as the stationary case, and $m_n{}^*$ is chosen such that

$$|m_n^*| \leq \beta_n \tag{24b}$$

and

$$\frac{1}{n} \sum_{i=1}^{n} \left(a_n^* - (m_i - m_n^*) - \frac{1}{2}(m_i - m_n^*)^2\right) \to 1 \text{ as } n \to \infty \tag{24c}$$

where

$$a_n^* = a_n - \log \log (n/2a_n) \tag{24d}$$

The term $m_n{}^*$ represents the amount that the normalizing constant $b_n$ is shifted to account for the nonstationary trend. Assuming only that the $m_i$ are bounded, then $m_n{}^*$ is

$$m_n^* = m_n^{*(1)} + a_n^{-1} \log k_n^{(1)} \tag{25a}$$

where

$$k_n^{(1)} = \sum_{i=1}^{n} \left(a_n - (m_i - m_n^{*(1)}) - \frac{1}{2}(m_i - m_n^{*(1)})^2\right) \tag{25b}$$

and

$$m_n^{*(1)} = a_n^{-1} \log \left(n^{-1} \sum_{i=1}^{n} e^{m_i a_n}\right) \tag{25c}$$

A stronger assumption concerning the deterministic trend could be made. In this case, the $m_i$ are bounded and the max $m_i = \beta$. Furthermore, it is assumed that v of $m_1, \ldots, m_n$ are equal to $\beta$, and that v $\sim n$. The value of $m_n{}^*$ is now given by[14]:

$$m_n^* = a_n^{-1} \log \left(n^{-1} \sum_{i=1}^{n} e^{m_i a_n}\right) = m_n^{*(1)} = \beta + o(a_n^{-1}) \tag{26}$$

In this case, $\beta$ is the maximum shift of the mean and it is assumed that most of the observations come from a distribution where $\beta$ is the mean.

Although these models were developed for an underlying normal distribution, they can be extended to include the lognormal distribution, which is often used to model the probability of an annual flood. If $Y$ is distributed lognormally, then the transformation

$$x = \log(y) \tag{27}$$

will give a normally distributed random variable $X$. Let

$$W_n = \max (Y_1, \ldots, Y_n) \tag{28}$$

The asymptotic distribution of $W_n$ is again a Type I extreme value distribution[19]:

$$P\{d_n(W_n - c_n) \leq x\} \rightarrow \exp(-e^{-x}) \tag{29a}$$

where

$$c_n = \exp [b_n + m_n^*] \tag{29b}$$

and

$$d_n = a_n/c_n \tag{29c}$$

Theoretically, any univariate distribution $F(x)$ can be transformed into a normal distribution. However, it may be necessary to approximate the $F(x)$ and the normal distribution as a Taylor series, since a closed form expression does not exist for the normal cumulative distribution function (or $F(x)$ also). Lambert and Li studied the impact of monotone transformations to transform or preserve the domain of attraction.[20] Assuming that the underlying distribution is normal or lognormal, the Type I extreme value distribution can then be used to estimate the probability that the maximum over an $n$ year period is below a threshold.

If the yearly change in the mean can be determined, then $m_n^*$ can be determined using Eq. (25). Alternatively, if the yearly change in the mean is unknown, but the maximum shift in the mean can be estimated, then the Type I distribution can be used with $m_n^* = \beta$, the maximum shift in the mean. This probability distribution provides an upper bound to the probability of failure under nonstationary conditions. In the next section, these models will be applied to the analysis of flood risk under nonstationary conditions.

## 4. APPLICATION OF NONSTATIONARY MODELS TO FLOOD PROTECTION

### 4.1. Modeling Approach

A flood protection structure such as a levee is often designed to protect for a level of flood that on average occurs once in every 100 years (assuming stationary conditions). If conditions are not stationary, then the levee is actually protecting for a flood with a shorter return period. The models of the previous section will be used to estimate the probability that the maximum of the series does not exceed a threshold $c$.

A *future yearly trend* in the mean or the variance could either be hypothesized or estimated from time series data. Assuming the trend continues in the future and the yearly data are independent, an exact distribution for the probability of failure can be determined. However, the underlying distribution and parameters are often not known exactly. In these cases, an extreme value distribution could be used to model the annual maximum flood. A Type I extreme value distribution is often used for this purpose. If a trend in the parameters of the Type I distribution can be determined, the model that builds on Eq. (25) can be used to estimate the probability of failure. If the underlying trend is not known but the total shift in the parameters can be estimated, then an upper bound on the probability of failure can be estimated using the Type I distribution.

The probability of failure calculated using these models depends greatly on future trends. There is, of course, uncertainty concerning whether a trend exists in precipitation or runoff. There is additional uncertainty in quantifying such a trend. The next section will look at how the probability of failure changes as the mean changes, using the three different models.

### 4.2. Application of Nonstationary Models for Different Changes in the Mean

For this example, it is assumed that the underlying flood random variable has been transformed to be of a normal distribution. In the first year, the distribution was a standard normal distribution with the mean equal to 0 and the standard deviation equal to 1.0. In each year, the mean was assumed to increase by a small amount. This change was assumed to occur for 50 years. The mean was increased linearly (a constant change each year for 50 years). Different annual changes in the mean were used. The threshold was 2.326, which corresponds to $P(Z < 2.326) = 0.99$ for a standard normal distribution. In the stationary case (no increase in the mean), this threshold corresponds to protection against a flood with a return period of 100 years.

The models presented in Sec. 2.2 and 3.3 are used to estimate the probability of a failure occurring. The probability of failure is defined to be

$$Pr\{\text{failure}\} = 1 - Pr\{M_{50} \leq 2.326\} \tag{30}$$

or the probability that the maximum in the 50-year period exceeds the threshold. We will demonstrate three methods reflecting alternative sets of assumptions. The first method assumes that the sequence is independent and calculates the exceedance probability exactly from the underlying annual distribution according to Eq. (8). The second method uses Eq. (24):

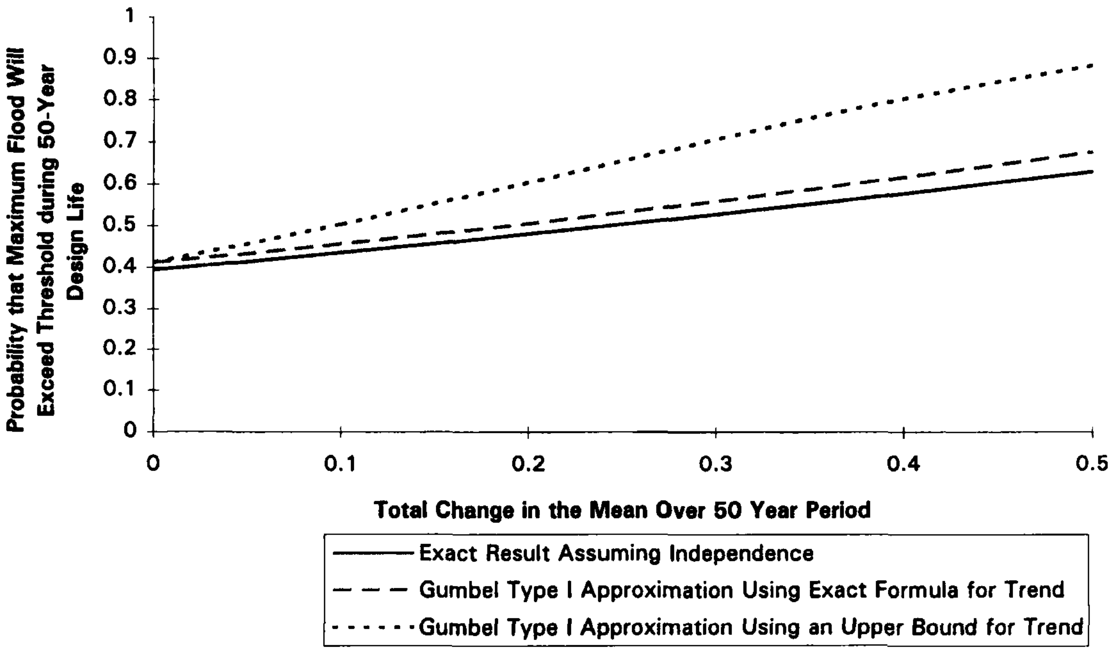


**Fig. 7.** Probability of failure over a 50-year design life for an underlying normal probability distribution with a linear change in the mean.

$$P\{M_n \leq x\} = \exp(-\exp[-a_n(x - b_n - m_n^*)]) \quad (31)$$

where $m_n{}^*$ is calculated using Eq. (25). The third method assumes a Type I distribution where the upper bound of the trend $\beta$ is used to approximate the nonstationary trend:

$$P\{M_n \leq x\} = \exp(-\exp[-a_n(x - b_n - \beta)]) \quad (32)$$

The results of using the three different methods to calculate probability are shown in Fig. 7. Notice that the estimate of the probability of failure is higher using the approximation of the Type I limiting distribution than it is for the exact calculation based on Eq. (8). The limiting distribution using the maximum shift $\beta$ for $m_n^*$ is in turn higher than the estimate using the limiting distribution and an exact formula for $m_n{}^*$. This result is to be expected, since it captures the trend by its upper bound only.

### 4.3. Sensitivity Analysis: Effect of Rate of Change

The first two methods to calculate the probability of failure depend on the amount of change each year. This section will look at how the probability of failure depends on the rate of change. If the rate of change were initially higher and gradually leveled off, it would be expected that the probability of failure over the design life of a project would be higher than for a case of constant change. We assumed that the mean in year k was given by

$$\mu_k = \rho(1 - \exp[\omega(k - 1)]) \quad (33)$$

with $\rho$ being the total change, and $\omega$ chosen to ensure a rapid initial rate of change with all the change occurring in a 50-year time period. Figure 8 is a graph comparing the failure probabilities using Eq. (24) with an exponential rate of change with the linear rate. The total change in the mean was the same for both the linear and exponential rates of change. Therefore, the probability of failure in year 50 is the same for both trends. Figure 8 shows that the risk of failure over the 50-year design life is higher when there is a higher initial increase. Using a Type I distribution with $m_n{}^* = \beta$ provides an upper bound on the probability of failure no matter how fast the change occurs.

### 4.4. Management of Extreme Events Under Nonstationary Conditions

Figures 1 and 2 show another aspect of risk management under nonstationary conditions if the probability of a failure is increasing in each year. A flood protection structure, for example, is often designed to provide protection for floods with a "return period" of 100 years under stationary conditions. The designer must

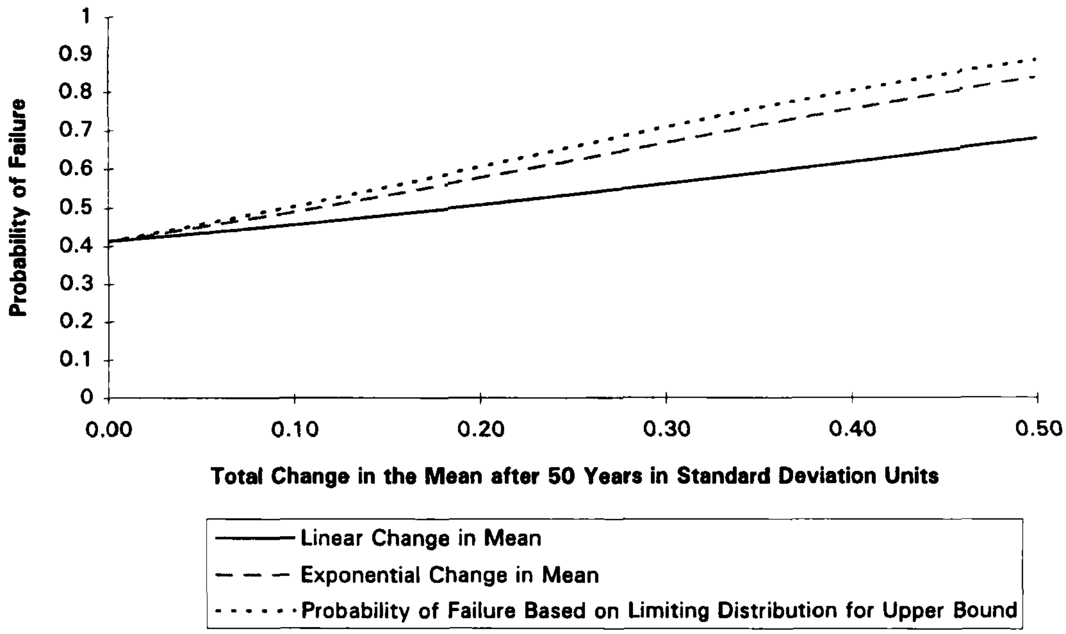


**Fig. 8.** Probability of failure for different rates of change in the mean for a 50-year period with a changing mean and a normal underlying distribution.

be aware that if conditions are nonstationary, then the expected number of years before a failure could be less than 100, even at the start of the project. The designer should consider alternative design criteria if conditions are nonstationary. Three possible definitions of the return period are given in Table I. The first possible criterion is to ensure that the structure provides protection in all years for at least a 100-year flood (1/probability = 0.01). A second possible criterion would be to ensure that the expected time before failure at the beginning of the project's life is 100 years. However, if this criterion is used, the expected waiting time before failure in a later year may be much less than 100 years. An approach even more potentially conservative would be to require the expected waiting time to be at least 100 years in all years of the project.

In a risk analysis, it is also necessary to consider the level of damages associated with a failure and to compare these damages for different project alternatives. To show how nonstationarity may affect such an analysis, assume that there is a constant level of damages for a flood which exceeds the threshold $c$, denoted as $D$. In the stationary case, the expected value of damages in any year $t$ $E[D_t]$ would be the same in every year of the project's life:

$$E[D_t] = (1 - F_X(c))D \tag{34}$$

However, in the nonstationary case with an increasing trend, the expected value of damages would be increasing in every year due to the changing probability of failure:

$$E[D_t] = (1 - F_{X,t}(c))D \tag{35}$$

One measure of the future damages used to compare alternative projects is to calculate the present value of the sum of the damages using a discount rate $r$:

$$\begin{aligned} \text{PV(Damages)} &= \sum_{i=1}^{50} E[D_i] \frac{1}{(1+r)^{i-1}} \\ &= \sum_{i=1}^{50} (1 - F_{X,i}(c))D \frac{1}{(1+r)^{i-1}} \end{aligned} \tag{36}$$

A consequence of discounting is that the damages in future years are valued less than the value of damages in the earlier years, even though the expected damages could be much larger in the later years. This effect of discounting is not a problem in the stationary case, since the probability of failure is constant in each year. However, if discounting of future damages is used in the nonstationary case with an increasing trend, these damages are weighted less than those in earlier years, even though the probability of failure is greater. The risk to future generations will be larger than for the current generation.

For the dynamic case, it is not clear what is the best measure to characterize risk. The decision maker must decide how to value a failure in a distant future year vs.

**Table I.** Possible Definitions of the Return Period for Nonstationary Conditions

| Definition of return period | Mathematical formula for return period |
|---|---|
| Inverse of the probability of failure within the year $t$. (Expected observations of identical years before failure.) | $\tau_1 = \tau_1(t) = \frac{1}{p(t)} = \frac{1}{1 - F_{X,t}(c)}$ |
| Expected waiting time before failure at the beginning of the design life of a project. | $\tau_2 = \sum_{k=1}^{\infty} \left\{ k \left( \prod_{i=1}^{k-1} [F_{X,i}(c)] \right) [1 - F_{X,k}(c)] \right\}$ (not a function of $t$). |
| Expected waiting time before failure starting at any year $t$ during the project life. | $\tau_3 = \tau_3(t) = \sum_{k=1}^{\infty} \left\{ k \left( \prod_{i=t}^{k+t-2} [F_{X,i}(c)] \right) [1 - F_{X,k+t-1}(c)] \right\}$ $t = 1, 2, \ldots$ |

a failure in an earlier year. The decision maker also must consider the uncertainty of the trend which may be altering the probability of the extreme event occurring. The results here leave ample flexibility to the decision maker as to how to discount the future through various interpretations of the return period.

## 5. CONCLUSIONS

This paper looked at two aspects of characterizing extreme events and how they may be affected by nonstationary conditions. The first consideration was the definition of the return period. Engineered infrastructures that protect urban and rural areas from floods are designed and constructed on the basis of specified return periods. It is very common that levees along the Mississippi River, for example, protect for a level of flood that on average occurs once in every 100 years. When conditions are nonstationary, the probability of occurrence of a level of flood may be changing in each year. Defining the return period as the expected waiting time before failure may be a more accurate characterization. If there is an increasing trend in the mean, then using this definition for the return period may provide a more conservative design standard.

A second consideration in this paper was applying extreme value distributions to nonstationary conditions in order to estimate the probability of flooding. Given the difficulty in determining the exact probability distribution function that can adequately represent the hydrology of a region, the use of the statistics of extremes reduces these choices of probabilities to only one of three families (Gumbel Types I, II, and III). A Gumbel Type I can be used to estimate the probability of failure if there is a known trend in a parameter of the random variable. If the trend is unknown but the maximum shift in the parameter can be estimated, then an upper bound on the probability of failure can be determined.

The focus of this paper has been on the changes in the probability of flooding due to changes in hydrology. However, these models also can be used to estimate the increased probability of a threshold of damages due to economic development and population growth. They can also be applied to other systems that exhibit trends or nonstationarity. An example of such a nonstationary system would be computer workload performance. The theory and methodology developed here should find additional applications for the quantification of risk in other engineering and nonengineering systems.

## ACKNOWLEDGEMENTS

This research was partially supported by NSF Grant CMS-9322213, Risk Management of Extreme Events in Engineered Systems Vulnerable to Climate Change. We thank Nicholas Matalas for his review and helpful comments on this paper. We also thank Rick Katz for his insights into the problem of climate change and extreme events. This paper was first presented to the Annual Meeting of the Society for Risk Analysis; Honolulu, Hawaii; December 1995. The authors are grateful for the feedback from the audience at and subsequent to this presentation.

## REFERENCES

1. P. E. Waggoner (ed.), *Climate Change and U.S. Water Resources* (John Wiley & Sons, New York, 1990).

2. R. W. Katz, "Towards a Statistical Paradigm for Climate Change," Climate Res. **2**, 167–175 (1992).
3. L. O. Mearns, R. W. Katz, and S. H. Schneider, "Extreme High-Temperature Events: Changes in their Probabilities with Changes in Mean Temperature," *J. Clim. Appl. Meteorol.* **23**, 1601–1613 (1984).
4. R. W. Katz and B. G. Brown, "Extreme Events in a Changing Climate: Variability is More Important than Climate," *Climat. Change* **21**, 289–302 (1992).
5. R. W. Katz and B. G. Brown, "Sensitivity of Extreme Events to Climate Change: The Case of Autocorrelated Time Series," *Environmetrics* **5**, 451–462 (1994).
6. T. M. L. Wigley, "The Effect of Changing Climate on the Frequency of Absolute Extreme Events," *Clim. Monit.* **17**, 44–55 (1988).
7. Ang and Tang, *Probability Concepts in Engineering Planning and Design*, Vol. 2 (John Wiley, New York, 1984).
8. E. Castillo, *Extreme Value Theory in Engineering* (Academic Press, San Diego, 1988).
9. E. J. Gumbel, *Statistics of Extremes* (Columbia University Press, New York, 1958).
10. P.-O. Karlsson and Y. Y. Haimes, "Risk-Based Analysis of Extreme Events," *Water Res. Res.* **24**, 9–20 (1988).
11. P.-O. Karlsson and Y. Y. Haimes, "Risk Assessment of Extreme Events: Application," *J. Water Res. Plan. Mgmt* **115**, 299–320 (1989).
12. J. Mitsiopoulos and Y. Y. Haimes, "Generalized Quantification of Risk Associated with Extreme Events," *Risk Anal.* **9**, 243–254 (1989).
13. J. Mitsiopoulos, Y. Y. Haimes, and D. Li, "Approximating Catastrophic Risk Through Statistics of Extremes," *Water Res. Res.* **27**, 1223–1230 (1991).
14. M. R. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes* (Springer-Verlag, New York, 1983).
15. M. Falk, J. Husler, and R.-D. Reiss, *Laws of Small Numbers: Extremes and Rare Events* (Birkhauser Verlag, Basel, 1994).
16. V. G. Kulkarni, *Modeling and Analysis of Stochastic Systems* (Chapman and Hall, London, 1995).
17. N. Matalas, Personal communication (1995).
18. S. I. Resnick, "Products of Distribution Functions Attracted to Extreme Value Laws," *J. Appl. Probabil.* **8**, 781–795 (1972).
19. J. Horowitz, "Extreme Values from a Nonstationary Stochastic Process: An Application to Air Quality Analysis," *Technometrics* **22**, 469–478 (1980).
20. J. H. Lambert and D. Li, "Evaluating Risk of Extreme Events for Univariate Loss Functions," *J. Water Res. Plan. ASCE* **120**, 382–399 (1994).