Impact of Streamflow Persistence on Hydrologic Design[1]

Ellen M. Douglas, S.M.ASCE[2]; Richard M. Vogel, M.ASCE[3]; and Charles N. Kroll, M.ASCE[4]

Abstract: Conventional methods for estimating the average return period, $E(T)$, and failure risk, R, generally ignore the impact of persistence in annual streamflows on associated probabilistic statements and streamflow statistics. Recent evaluations of streamflow observations indicate statistically significant serial correlations (persistence) associated with annual low flows in the United States. We define the average occurrence interval $E(T)$ as the expected time to the first event, and we present a method for estimating $E(T)$ and R in the presence of persistence. We show that for observed ranges of persistence, $E(T)$ can be nearly 100% greater and R more than 20% lower than conventional estimates. This implies that the expected design life of a system is longer when persistence is taken into account. Also, by ignoring persistence, low flow quantiles may be underestimated by 50% or more. An evaluation of the effect of persistence on drought risk estimation across the U.S. is presented.

Introduction

It is generally assumed that certain streamflow statistics evaluated at an annual time step, for example, floods or low flows, are serially independent. Consequently, frequency analyses for floods and low flows usually ignore serial correlation (a measure of persistence). The seven-day low flow, defined as the minimum annual flow obtained from a seven-day moving average of daily streamflow, is commonly used in engineering applications such as waste load allocation, instream flow regulation, drought risk, design of cooling plants, siting of treatment plants and sanitary landfills, and decision making regarding interbasin transfers and allowable basin withdrawals. Accurate estimation of design low flow quantiles is paramount to the success of many engineering designs. Contrary to the assumption of serial independence, Douglas et al. (2000) found statistically significant serial correlation in annual minimum seven-day low flows at approximately 25% of the stations analyzed across the U.S. The effect of persistence on engineering design quantiles must, therefore, be evaluated in low flow investigations.

Definition of Average Recurrence and Occurrence Intervals

A quantile, designated Q_p, is the streamflow value associated with a specified exceedance or nonexceedance probability, p. In practice, instead of referring to a quantile in terms of probability, it is often referred to in terms of its average return period. For example, $Q_{0.01}$ represents the streamflow with an annual exceedance probability equal to 0.01; this quantile is commonly referred to as the 100-year flood, and is denoted Q_{100}. In low flow analyses, a common design quantile is the seven-day low flow with an average return period of 10 years, which has a nonexceedance probability of 0.1 and is typically denoted $Q_{7,10}$.

Gumbel (1941) and Thomas (1948) defined the return period, T, as the interval between flood events, where a flood event was defined as a streamflow greater than a threshold value (design flood). They noted that, assuming the flood events are independent of one another, T follows a geometric probability distribution. Therefore

$$P(T=t)=p(1-p)^{t-1} \quad \text{for } t=1,2,3,\ldots,\infty \qquad (1)$$

where p = probability that the design flood will be exceeded in any one year (the annual exceedance probability). The mean, or expected value of T, denoted $E(T)$, is

$$E(T)=\sum_{t=1}^{\infty} t \cdot P(T=t)=\sum_{t=1}^{\infty} t \cdot p(1-p)^{t-1}=1/p \qquad (2)$$

The average return period, as originally derived by Gumbel and Thomas, is the time it takes, on average, for a design flood of a given probability, p, to be exceeded. In practice, when the term "return period" is used, it is usually meant to be interpreted as the average return period.

Interestingly, the average return period can be defined in different ways for different applications. Lloyd (1970) defined the average return period as the expectation of the interval between two flood events. Lloyd showed that the average return period defined in this way is insensitive to streamflow persistence. Vogel (1987) defined the average return period of a reservoir system failure as the expected number of years before the first occurrence of a system failure (i.e., flood or drought). Vogel showed that the

average return period defined in this way is indeed affected by streamflow persistence. Fernandez and Salas (1999a) and Sen (1999) developed more general formulations for estimating the average return period of design events in the presence of persistence. Tasker (1983) recognized the effect of persistence in streamflow records on the reliability of quantile estimates and developed a method for quantile estimation using the effective record length, n_e, which, in most cases, is less than the actual historical record length, n, when persistence is present in the streamflow record.

The concept of average return period can be applied to both exceedance events (i.e., floods) and nonexceedance events (i.e., low flows or droughts). In this paper, we accommodate these differing applications by redefining p as the probability of design failure regardless of whether that failure represents an exceedance event or a nonexceedance event. We will use the phrase "recurrence interval" to describe the time between a series of events, W, and coin a new phrase, "occurrence interval," to describe the time to the first event, T. We define the expected values of these two intervals as
- Average occurrence interval, $E(T)$=the expected value of the occurrence interval; and
- Average recurrence interval, $E(W)$=the expected value of the recurrence interval.

We believe that in practice, the average occurrence interval is more useful than the concept of the average recurrence interval, because it does not require an assumption regarding initial conditions; hence, it more closely corresponds to design conditions. In other words, one does not usually assume that a flood or drought has just occurred, when planning for the next flood or drought. We first review previous literature on this subject using a common notation and terminology, and derive equations for estimating the average occurrence interval and failure risk that incorporate streamflow persistence. We then explore the effect of streamflow persistence on average occurrence intervals, quantiles, and the risk of design failure.

Two-State Markov Model

In any given year, an annual statistic, X, may be in one of two states relative to a design threshold, X_0; it either exceeds X_0 (constituting a failure state) or it does not (constituting a nonfailure or safe state). In the case of a flood control structure, exceeding the design threshold would constitute a failure state. Alternatively, in the case of a drought, exceeding a critical deficit would constitute a failure. In any given year, the probability of failure is equal to $P(X>X_0)=p$. Conversely, the probability that the design will operate in the safe state is $P(X\leq X_0)=1-p=q$. If the assumption of serial independence of X is valid, a simple two-state Bernoulli model is sufficient to model the transition from one state to the next. If U denotes the state of X in any given year t, then in the two-state Bernoulli case

$$P(U_t|U_{t-1},U_{t-2},...,U_0)=P(U_t) \qquad (3)$$

In other words, the state of X in any given year is independent of the state of X in any previous year. However, in the case where X is serially correlated, the situation is more complex. The simplest model that accommodates persistence is a two-state Markov chain model

$$P(U_t|U_{t-1},U_{t-2},\ldots,U_0)=P(U_t|U_{t-1}) \qquad (4)$$

which assumes that the state of X in any given year is only influenced by the state of X in the previous year. This model has been
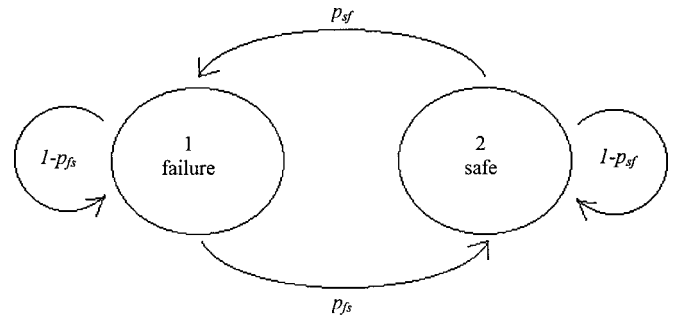

Fig. 1. Graphical representation of two-state Markov model

used by many authors (Lloyd 1970; Jackson 1975; Sen 1976; Hirsch 1979; Stedinger et al. 1983; Vogel 1987; Vogel and Bolognese 1995; Fernandez and Salas 1999a; Sen 1999) to model the effects of year-to-year streamflow persistence. Chung and Salas (2000) used more complex (discrete autoregressive and discrete autoregressive moving average) models to derive the probability distribution of drought occurrence and expressions for the average return period and risk.

Following Vogel (1987), let the row vector $\mathbf{P}_t=(p,q)$ specify the probability that a design is in either the failure state (with probability p) or the safe state (with probability q) in year t and that $\mathbf{P}_t$, $t=1,\ldots,N$ forms a Markov chain with the probability transition matrix

$$\mathbf{A}=\begin{bmatrix} 1-p_{fs} & p_{fs} \\ p_{sf} & 1-p_{sf} \end{bmatrix} \qquad (5)$$

where p_{sf}=probability that a failure year follows a safe year; and p_{fs}=probability that a safe year follows a failure year. The Markov chain model is given by $\mathbf{P}_t=P_{t-1}A$ and is depicted graphically in Fig. 1. Fig. 2 illustrates the joint probability regions with respect to the threshold, X_0. The marginal probabilities are defined as

$$P(X_t\leq X_0\cap X_{t-1}>X_0)+P(X_t>X_0\cap X_{t-1}>X_0)$$
$$=P(X_{t-1}>X_0)=p \qquad (6)$$
$$P(X_t\leq X_0\cap X_{t-1}\leq X_0)+P(X_t>X_0\cap X_{t-1}\leq X_0)$$
$$=P(X_{t-1}\leq X_0)=q \qquad (7)$$

Borrowing from Sen (1976), we define the autorun coefficient, r, as the conditional probability of two consecutive failure states, which is expressed as
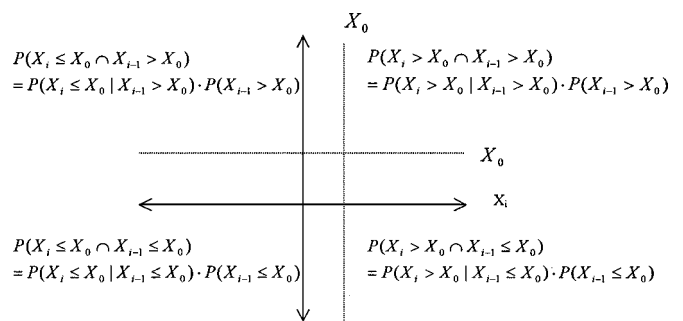

Fig. 2. Joint probability regions for four Markov transition states for specified design threshold, X_0

$$r = 1 - p_{fs} = P(X_i > X_0 | X_{i-1} > X_0) = \frac{P(X_i > X_0 \cap X_{i-1} > X_0)}{p}$$

$$(8)$$

With the relationships defined in Eq. (8) and Fig. 2, we can express the conditional probabilities for the other three transitions in terms of r, p, and q, as did Sen (1976)

$$p_{fs} = P(X_i \leq X_0 | X_{i-1} > X_0) = 1 - r \qquad (9)$$

$$p_{sf} = P(X_i > X_0 | X_{i-1} \leq X_0) = \frac{p}{q} \times (1 - r) \qquad (10)$$

$$1 - p_{sf} = P(X_i \leq X_0 | X_{i-1} \leq X_0) = 1 - \frac{p}{q} \times (1 - r) \qquad (11)$$

Average Occurrence Interval and Risk of Failure in Presence of Streamflow Persistence

For single events in Markov-dependent trials, Vogel (1987) and Fernandez and Salas (1999a) derived an expression for the average occurrence interval of a system failure defined by

$$E(T) = 1 + \frac{q^2}{p(1-r)} \qquad (12)$$

Similarly, Sen (1999) used a two-state Markov chain model to derive an expression for the average occurrence interval of a failure (in this case, the exceedance of a design flood) as

$$E(T) = \frac{q^2}{\left[1 - \frac{p}{q}(1-r)\right] p(1-r)} \qquad (13)$$

Eqs. (12) and (13) were derived with differing initial conditions. Eq. (13) is conditioned on the first year being a failure year, while for Eq. (12), the first year can be either a safe year or a failure year. We will use the unconditional equation (12) for the remainder of this paper to estimate the average occurrence interval of a system failure in the presence of persistence, since that situation most closely resembles the typical assumptions associated with a hydrologic design.

The fact that the average recurrence interval (time between successive events) is not affected by streamflow persistence, but the average occurrence interval (time to the first event) is, leads us to conclude that the effect of streamflow persistence on hydrologic design is most pronounced over the short term. For certain engineering applications, such as reservoir design, the expected time to the first failure of the system is of more interest than is the average time between successive failures over a long period of record. This information is used to evaluate the risk of failure over the design life of the system. If we define failure risk, R, as the probability of at least one failure over the design life, n, of a system (Chow et al. 1988) and assume that we are starting in a safe year, then failure risk can be expressed as

$$R = 1 - P(X_1 \leq X_0) \cdot \prod_{t=2}^{n} P(X_t \leq X_0 | X_{t-1} \leq X_0)$$

$$= 1 - q \cdot \left[1 - \frac{p}{q}(1-r)\right]^{n-1} \qquad (14)$$

which agrees with Sen (1999) and Fernandez and Salas (1999a).

Simplified Method for Relating Average Occurrence Interval and Failure Risk to Persistence

If we make the common assumption that streamflow is lognormally distributed and that X represents the annual statistic (maximum or minimum) calculated from the log-transformed streamflow data, then the joint density $P(X_t > X_0 \cap X_{t-1} > X_0)$ can be calculated using the double integral introduced by Sen (1976)

$$P(X_t > X_0 \cap X_{t-1} > X_0)$$

$$= \int_{X_0}^{\infty} \int_{X_0}^{\infty} \frac{1}{2\pi(1-\rho^2)^{1/2}}$$

$$\times \exp\left[-\frac{1}{2(1-\rho^2)}(x_i^2 - 2\rho x_i x_{i-1} + x_{i-1}^2)\right] dx_i dx_{i-1}$$

$$(15)$$

where x_i and x_{i-1} = normal variates with a mean of zero and unit variance; and ρ = lag-one autocorrelation coefficient of the log-transformed data. The autorun coefficient r [Eq. (8), which depends on the double integral in Eq. (15)] is a function of both p and ρ; therefore, Eq. (15) must be solved for each value of p and ρ. For Eqs. (12) and (14) to be of practical use, we developed an approximation for r

$$r = \frac{ab^p p^c}{p} \qquad (16)$$

The coefficients a, b, and c are functions of ρ. Using a polynomial model, the following equations were developed for calculating the coefficients in Eq. (16):

$$a = 1 - 1.514\rho + 2.601\rho^2 - 1.0016\rho^3 - 1.620\rho^4 + 1.475\rho^5 \qquad (17)$$

$$b = 1 + 1.553\rho - 0.7789\rho^2 - 3.5810\rho^3 + 5.5638\rho^4 - 2.701\rho^5 \qquad (18)$$

$$c = 2 - 2.0855\rho + 2.3419\rho^2 - 2.2995\rho^3 + 1.6817\rho^4 - 0.6232\rho^5 \qquad (19)$$

Eqs. (16)–(19) replace Eqs. (8) and (15), thereby simplifying the calculation of the average occurrence interval and failure risk in the presence of persistence. Relative errors for the approximation of r and the associated errors in the estimation of $E(T)$ and R are summarized in Douglas (2002). In general, relative errors for the approximation of r are less than 4% for $p \geq 0.1$, but grow larger as p decreases, ranging from 2.3 to 7% at $p = 0.05$ and from 4.3 to 15% at $p = 0.01$ (Fig. 3). The average normalized root-mean-square error for the approximation of r is 1.5%; for the approximation of $E(T)$ it is 15%, and for the approximation of R it is 0.6%.

Effect of Persistence on Average Occurrence Interval, Low Flow Quantiles, and Risk

Fig. 4 illustrates the effect of streamflow persistence on the average occurrence interval, $E(T)$, calculated from Eq. (12) [using Eqs. (16)–(19) to approximate the autorun coefficient, r]. Fernandez and Salas (1999a) present a similar figure, except with $E(T)$ as a function of p for selected values of ρ. For all levels of p, $E(T)$ increases as persistence increases. This is expected due to differential persistence, which is the phenomenon whereby higher values of persistence make it more likely to remain in a drought if one is already in a drought. For $p = 0.5$, $E(T)$ ranges from two years at $\rho = 0.0$ to 6.1 years at $\rho = 0.95$. For $p = 0.01$, $E(T)$ ranges from 100 years at $\rho = 0.0$ to 327 years at $\rho = 0.95$. Within the
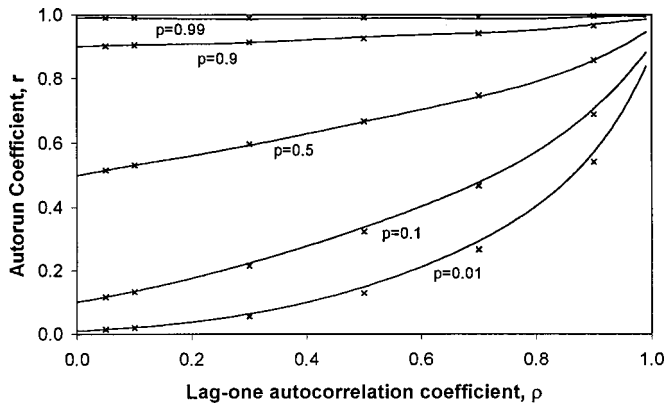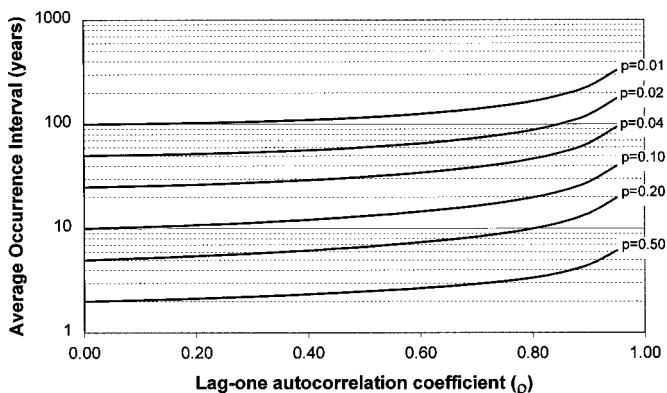
Fig. 3. Comparison of exact (curves) and approximate values (x) of autorun coefficient r as function of lag-one serial correlation ρ, for selected values of p. Approximate values of r are calculated with Eqs. (16)–(19).

observed range of at-site estimates of ρ (0.0–0.8), $E(T)$ increases by 66% for $p=0.01$, 75% at $p=0.02$, and 98% for $p=0.1$. This means that for the $Q_{7,10}$, the average occurrence interval could be underestimated by nearly 100% using the conventional method of $E(T)=1/p$.
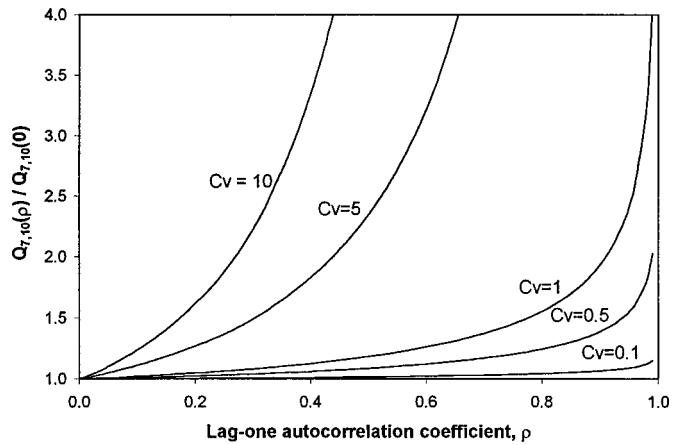
The effect of persistence on the seven-day, 10-year low flow ($Q_{7,10}$) is illustrated in Fig. 5. Annual minimum seven-day low flows were assumed to follow a three-parameter lognormal (LN3) probability distribution, which is obtained by letting $\ln(X-\tau)$ be normally distributed. The LN3 model for the $Q_{7,10}$ quantile is

$$Q_{7,10}=\tau+\exp(\mu+\sigma Z_p) \qquad (20)$$

where $\tau=$ lower bound; μ and $\sigma=$ mean and standard deviation of $\ln(X-\tau)$; and $Z_p=\Phi^{-1}(p)$, the inverse cumulative distribution function of a standard normal variate. For the illustrative purposes of this section, τ was set equal to zero, reducing the distribution to two-parameter lognormal. The LN3 distribution is used in a later section. For independent low flows, $p=1/E(T)$ or 0.1 in the case of the $Q_{7,10}$. However, when $E(T)$ remains constant (i.e., 10 years), the value of p increases with persistence. Z_p for ρ ranging from 0.01 to 0.99 and for coefficients of variation (C_v) from 0.1 to 10 was calculated by specifying C_v, setting $E(T)=10$ years, and solving Eqs. (16)–(19) simultaneously for p. Fig. 5 illustrates that as C_v increases, so too does the ratio of the low flow statistic based on serially correlated flows, $Q_{7,10}(\rho)$, to the low flow sta-


Fig. 5. Impact of persistence on $Q_{7,10}$ estimates as function of C_v

tistic based on independent flows, $Q_{7,10}(0)$. For watersheds with low C_v (less than 1), the effect of persistence on quantile estimation is somewhat subdued. For instance, within the observed range of at-site estimates of ρ (0.0–0.8), the $Q_{7,10}$ quantile increases only 4.5% at $C_v=0.1$. For streamflows with higher C_v, the effect of persistence becomes much more dramatic. For $C_v=1$, the $Q_{7,10}$ estimate increases by 56% (ratio of 1.56) at $\rho=0.8$, and for $C_v=5$, the ratio is 9.1. At very large values, the ratios increase even faster. For $C_v=10$, the ratio at $\rho=0.1$ is 1.24, the ratio at $\rho=0.2$ is 1.62, at $\rho=0.5$ the ratio is 5.5, and at $\rho=0.8$, the ratio increases to 83. The range of the $L-C_v$ values for annual minimum flows across the U.S. is illustrated in Figs. 7 and 8 in Vogel and Wilson (1996). Those values can be converted to a range in C_v values using Fig. 8 in Limbrunner et al. (2000).

Figs. 6 and 7 illustrate the impact of serial correlation on the failure risk, R. The presence of persistence causes R to grow more slowly with time. For example, with a 25-year design life and $p=0.1$, the risk of failure is 23% lower at $\rho=0.8$ ($R=0.72$) than at $\rho=0.0$ ($R=0.93$). Another way to look at this is, for a specified risk, the design life increases as streamflow persistence increases. For instance, the horizontal dashed line in Fig. 6 represents $R=0.65$. By defining the design occurrence interval as the average occurrence interval and equating this to the expected life of an

Fig. 4. Variation in average occurrence interval, $E(T)$, with ρ for selected values of p
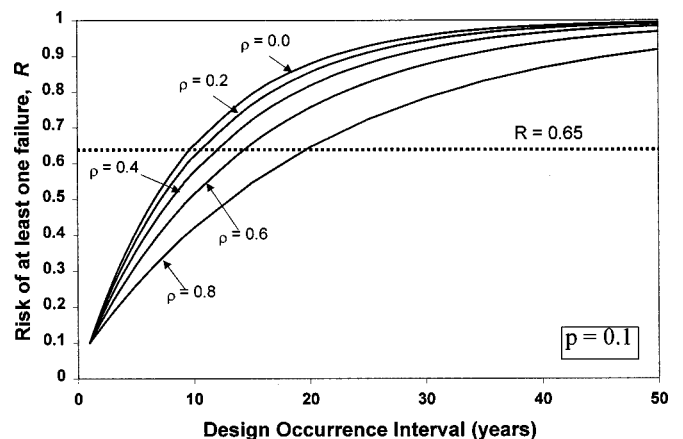
Fig. 6. Effect of persistence on risk of failure (R) and design occurrence interval for $p=0.1$. Dashed line represents a specified $R=0.65$. Design occurrence interval is set equal to the average occurrence interval.
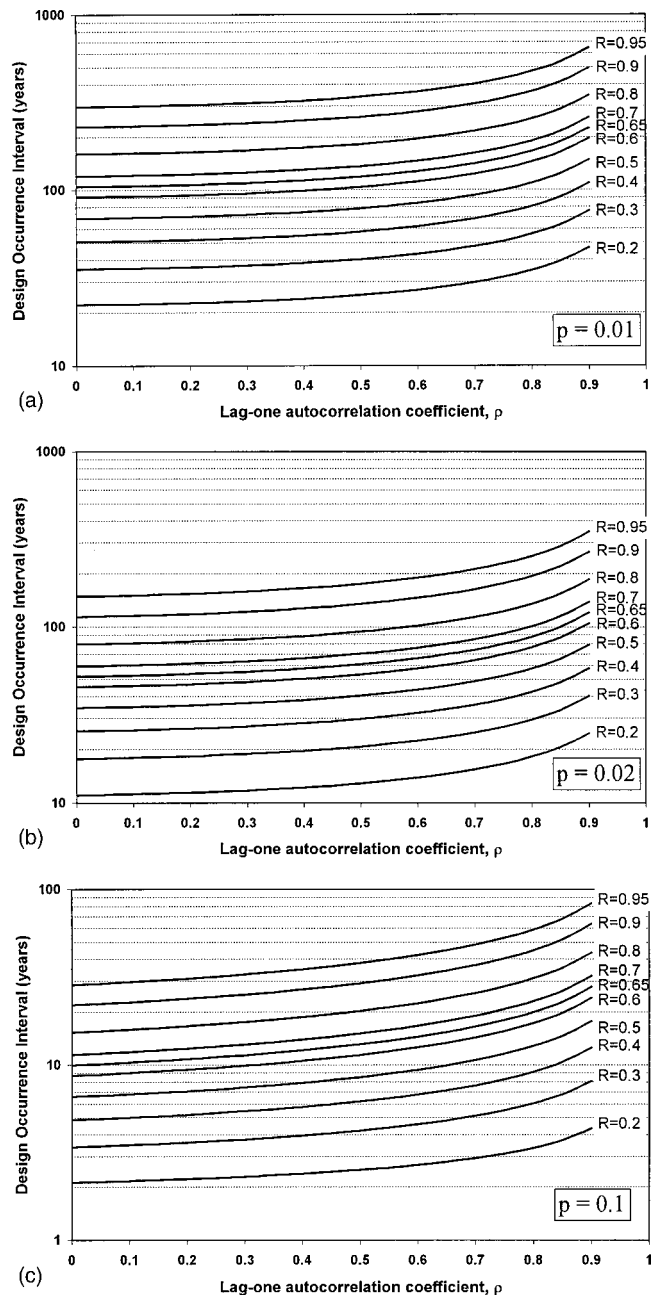
Fig. 7. (a) Design occurrence interval at specified values of risk of failure (R) for $p = 0.01$; (b) Design occurrence interval at specified values of risk of failure (R) for $p = 0.02$; (c) Design occurrence interval at specified values of risk of failure (R) for $p = 0.1$ (assumes that design occurrence interval is equal to average occurrence interval)

engineered system, the expected life increases from 10 years at $\rho = 0.0$ to 12 years (a 20% increase) at $\rho = 0.4$ and 20 years (a 100% increase) at $\rho = 0.8$. Clearly, incorporating the effects of persistence into hydrologic design can be beneficial because the expected life of a system will be longer when persistence is accounted for. Figs. 7(a–c) are plots of design occurrence intervals at specified values of R for $p = 0.01$, 0.02, and 0.1, which correspond to conventional design return periods of 100, 50, and 10 years, respectively.

Effect of Streamflow Persistence on Hydroeconomic Analysis

Chow et al. (1988) showed that the optimum design occurrence interval can be determined by knowing the probability of failure in any one year and the damage cost associated with this failure. As the design occurrence interval increases, the capital cost of a structure increases (because the magnitude of the event increases) but the expected cost of damages (or risk cost) decreases because the probability of occurrence decreases. The optimal design occurrence interval is the average occurrence interval that minimizes the sum of the capital cost and expected risk cost. The risk cost, D_0, is the product of the annual probability, $f(x)dx$, of an event of magnitude x, and the expected damage, $D(x)$, resulting from the event, integrated for $x > x_0$

$$D_0 = \int_{x_0}^{\infty} D(x) \cdot f(x) dx \tag{21}$$

Neither the exceedance probability of an event $[p = \int_{x_0}^{\infty} f(x) dx]$ nor the risk cost, D_0, which is based on both the magnitude and the probability of the event, is influenced by persistence. However, as we have already shown, the average occurrence interval associated with each event is influenced by persistence. Fig. 8 is an extension of Example 13.2.3 in Chow et al. (1988). By introducing a modest value of streamflow persistence ($\rho = 0.3$) into the analysis, the total cost curve (sum of capital and risk costs) is shifted to the right, resulting in the optimal design occurrence interval increasing from 25 years (as in the original analysis) to 28 years. This means that the expected design life of this system (if assumed equal to the optimal occurrence interval) is three years (12%) greater than it would have been if persistence had been ignored.

Impact of Persistence on Low Flow Risk Estimation in the United States

Fernandez and Salas (1999b) demonstrated the effects of persistence on low flow quantiles, risk, and occurrence interval estimation using natural streamflow records in Colorado and Argentina. Tasker (1983) used Monte Carlo simulations to compute effective record lengths for low flow quantiles estimated from specified values of record length and streamflow persistence. To investigate the general effect of incorporating streamflow persistence on hydrologic design parameters, we computed the average occurrence interval and the seven-day, 10-year low flow quantile ($Q_{7,10}$) accounting for persistence at over 1,000 sites across the U.S. These estimates were then compared to the average occurrence intervals and quantiles computed by conventional methods that ignore the impact of persistence.

Analyses in this study were performed using historical average daily streamflow records contained in the Hydro-Climatic Data Network, a dataset compiled by Slack et al. (1993) for 1,571 gauging stations across the continental U.S. Results were summarized for the 18 water resources regions illustrated in Fig. 9. Low flows were assumed to follow an LN3 distribution and $Q_{7,10}$ quantiles were estimated using Eq. (20). The following estimators of the LN3 distribution parameters were used (Loucks et al. 1981):

$$\hat{\tau} = \frac{x_{(1)}x_{(n)} - \hat{x}_{0.50}^2}{x_{(1)} + x_{(n)} - 2\hat{x}_{0.50}} \tag{22}$$
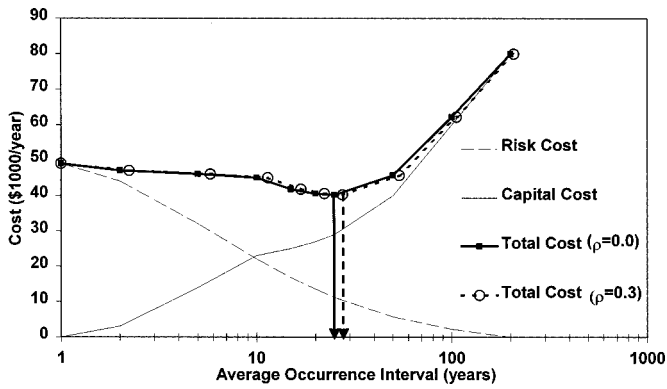
Fig. 8. Optimal design occurrence interval with persistence (28 years) and without persistence (25 years); modification of example 13.2.3 in Chow et al. (1988)

Fig. 10. Box plots of $E(T)$ estimates of $Q_{7,10}$ for stations across the U.S. Each box plot represents the distribution of $E(T)$ estimates for individual stations within each region.
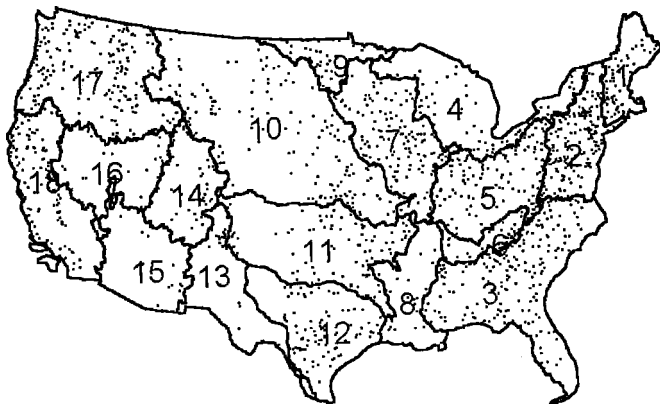
$$\hat{\mu} = \ln\left[\frac{\bar{x} - \hat{\tau}}{\sqrt{1 + s_x^2/(\bar{x} - \hat{\tau})^2}}\right] \qquad (23)$$

$$\hat{\sigma}^2 = \ln\left[1 + \frac{s_x^2}{(\bar{x} - \hat{\tau})^2}\right] \qquad (24)$$
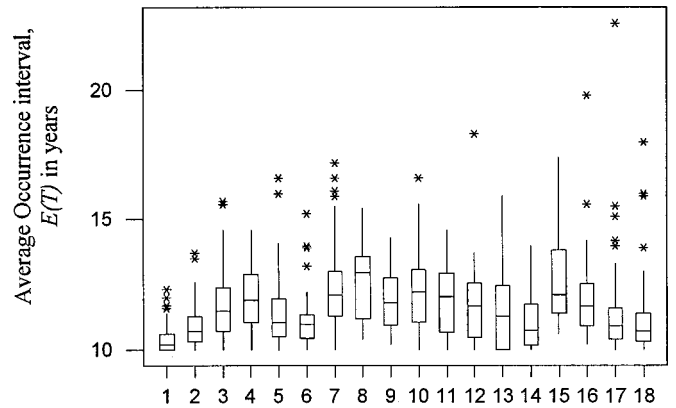
Records that contained zero flow values were excluded from this analysis because Kroll and Vogel (2002) found that streamflow records that contain zeros have very different distributions than those without zeros, and records with zero flows are poorly approximated by an LN3 distribution function.

Fig. 10 shows the variability of $E(T)$ estimates [using Eq. (12)] for 1,048 stations across the United States. Median values range from 10.2 years in region 1 (New England) to 12.9 years in region 8 (lower Mississippi River). Upper quartiles range from 10.6 years in region 1 to 13.5 years in region 8. The maximum value in each region ranges from 12.3 years in region 1 to 22.6 years in region 17 (Pacific Northwest). These results indicate that, while the effect of persistence on average occurrence interval estimates is at most 3.5% for three-quarters of the stations evaluated, $E(T)$ estimates in each region can be from 23 to greater than 100% higher than $E(T) = 10$ years estimated by the conventional method, $E(T) = 1/p$.

Fig. 11 illustrates the quantile ratios $[Q_{7,10}(\rho)/Q_{7,10}(0)]$ summarized for the same stations. Ratios higher than one indicate that the $Q_{7,10}$ quantile estimated by our method is higher than would be estimated by conventional methods (with constant $p = 0.1$).

Values of C_v for the $Q_{7,10}$ flows ranged from 0.1 to 1.7; therefore, the impact of persistence on quantile ratios for these stations was expected to be small to moderate (Fig. 5). Median quantile ratios range from 1.004 in region 1 to 1.068 in region 7 (upper Mississippi River). Upper quartiles ranged from 1.017 in region 1 to 1.155 in region 12 (Texas Gulf). The maximum ratio in each region ranged from 1.090 in region 1 to 1.855 in region 10 (Missouri River). In all regions, there were stations where the conventional method of quantile estimation would underestimate $Q_{7,10}$ flows by at least 9%. In nine of the 18 regions (regions 2, 3, 4, 9–13, 17, and 18), there were stations where this underestimation would be 20% or more.

At-site estimates of the lag-one autocorrelation coefficient (r_1) for LN3-transformed flows $[\ln(X - \tau)]$ were used to estimate $E(T)$ at each site. A summary of at-site r_1 values for each region is included in Table 1. These at-site estimates ranged from less than 0 (in which case r_1 was set equal to 0) to 0.848. At-site estimates of r_1 for the untransformed flows ranged from less than 0 to 0.841. There is enormous sampling variability associated with at-site estimates of lag-one correlation coefficients, and regional estimates may reflect the serial correlation structure better than the at-site estimates (Vogel et al. 1998). However, not having performed the type of sampling experiments as did Vogel et al. (1998) for annual flow series, we were unsure of the size of the region needed to ensure homogeneity.

Fig. 9. Water resources regions used for this study
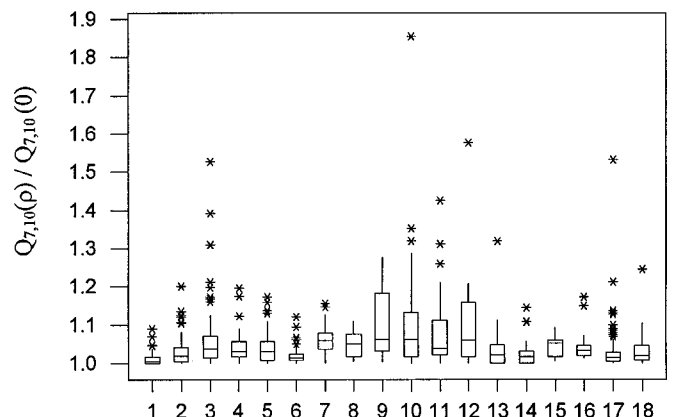
Fig. 11. Box plots of $Q_{7,10}$ quantile ratios for stations across the U.S.

Table 1. Lag-One Autocorrelation Coefficients for LN3-Transformed $Q_{7,10}$ Flows at Stations used in this Study

Region	Median r_1	Upper quartile	Lower quartile	Maximum	Minimum
1	0.05	−0.05	0.15	0.41	−0.33
2	0.17	0.07	0.28	0.54	−0.26
3	0.31	0.18	0.43	0.66	−0.27
4	0.36	0.24	0.46	0.60	−0.23
5	0.24	0.14	0.37	0.70	−0.27
6	0.22	0.12	0.28	0.64	−0.29
7	0.39	0.29	0.48	0.73	0.00
8	0.47	0.29	0.52	0.65	0.10
9	0.35	0.32	0.45	0.58	0.04
10	0.40	0.26	0.49	0.70	−0.17
11	0.38	0.17	0.47	0.60	−0.01
12	0.33	0.16	0.43	0.76	0.00
13	0.28	0.03	0.41	0.67	−0.17
14	0.18	0.07	0.34	0.56	−0.35
15	0.39	0.33	0.52	0.73	0.15
16	0.34	0.22	0.42	0.80	0.05
17	0.21	0.10	0.33	0.85	−0.28
18	0.18	0.07	0.30	0.76	−0.39

Conclusions

We defined the average occurrence interval, $E(T)$, as the expected time to the first event of interest. Here an event is defined as the exceedance (or nonexceedance) of some threshold event. The risk associated with an event, R, was defined as the probability of at least one event over the design life of the system. Conventional methods for estimating $E(T)$ and R generally ignore serial correlation (persistence) in streamflow. This is largely because traditionally, engineers have worked with the concept of a recurrence interval, which is the time interval *between* two events, as opposed to the occurrence interval, which is the time to the first event. The recurrence interval is not influenced by the persistence of the flow process, whereas the occurrence interval is. The occurrence interval is better suited to hydrologic design problems because it does not require the engineer to make assumptions about initial conditions, prior to the design life. However, if the occurrence interval is to find use in practice, one should include the impacts of persistence in the calculation.

By defining p as the probability of the event, regardless of whether that event is defined by the exceedance or nonexceedance of a threshold value, equations for estimating the expected occurrence interval $E(T)$ and risk R in the presence of persistence were derived. This study has shown that for observed ranges of streamflow persistence, $E(T)$ can be up to 98% greater and R can be 23% or more lower than conventional estimates. This means that the expected design life of a system is longer when persistence is taken into account. An evaluation of historical low flow data at over 1,000 sites across the United States indicates that $E(T)$ estimates for the $Q_{7,10}$ at some stations in each region were 20 to >100% higher than $E(T) = 10$ years estimated by the conventional method, and that quantile estimates at stations in half of the regions were underestimated by 20% or more.

Acknowledgements

Notation

The following symbols are used in this paper:

$E(T)$ = average occurrence interval;

$E(W)$ = average recurrence interval;

N,n = sample size, i.e., number of years of record, design life;

p = probability of failure;

p_{fs} = probability that failure year follows safe (nonfailure) year;

p_{sf} = probability that safe (nonfailure) year follows failure year;

Q_p = quantile;

q = probability of operating in safe (nonfailure) state;

R = failure risk, probability of at least one failure over life of system;

r = autorun coefficient, probability of two successive failure years;

r_1 = at-site estimate of lag-one autocorrelation coefficient;

T = occurrence interval;

U = state of X (failure or nonfailure) relative to threshold value;

X = annual statistics, i.e., annual maximum streamflow, annual low flow;

X_0 = design threshold, i.e., design flood, critical deficit;

Z_p = inverse cumulative distribution function of standard normal variate $= \Phi^{-1}(p)$;

μ, σ = mean and standard deviation of normal distribution;

ρ = lag-one autocorrelation coefficient; and

τ = lower-bound parameter of three-parameter lognormal distribution (LN3);

References

Chow, V. T., Maidment, D. R., and Mays, L. W. (1988). *Applied hydrology*, McGraw-Hill, New York.

Chung, C., and Salas, J. D. (2000). "Drought occurrence probabilities and risks of dependent hydrologic processes." *J. Hydrologic Eng.,* 5(3), 259–268.

Douglas, E. M. (2002). "The behavior of floods and low flows in the United States." PhD dissertation, Tufts Univ., Medford, Mass.

Douglas, E. M., Vogel, R. M., and Kroll, C. N. (2000). "Trends in flood and low flows across the U.S." *J. Hydrol.,* 240(1-2), 90–105.

Fernandez, B., and Salas, J. D. (1999a). "Return period and risk of hydrologic events. I: Mathematical formulation." *J. Hydrologic Eng.,* 4(4), 297–307.

Fernandez, B., and Salas, J. D. (1999b). "Return period and risk of hydrologic events. II: Applications." *J. Hydrologic Eng.,* 4(4), 308–316.

Gumbel, E. J. (1941). "The occurrence interval of flood flows." *Ann. Math. Stat.,* 12(2), 163–190.

Hirsch, R. M. (1979). "Synthetic hydrology and water supply reliability." *Water Resour. Res.,* 15(6), 1603–1615.

Jackson, B. B. (1975). "Markov mixture models for drought levels." *Water Resour. Res.,* 11(1), 64–74.

Kroll, C. N., and Vogel, R. M. (2002). "The probability distribution of low streamflow series in the United States." *J. Hydrologic Eng.*, in press.

Limbrunner, J. F., Vogel, R. M., and Brown, L. C. (2000). "Estimation of harmonic mean of lognormal variable." *J. Hydrologic Eng.,* 5(1), 59–66.

Lloyd, E. H. (1970). "Occurrence interval in the presence of persistence." *J. Hydrol.,* 10(3), 291–298.

Loucks, D. P., Stedinger, J. R., and Haith, D. A. (1981). *Water resource systems planning and analysis*, Prentice-Hall, Englewood Cliffs, N.J.

Sen, Z. (1976). "Wet and dry periods of annual flow series." *J. Hydraul. Div., Am. Soc. Civ. Eng.,* 102(10), 1503–1514.

Sen, Z. (1999). "Simple risk calculations in dependent hydrological series." *Hydrol. Sci. J.,* 44(6), 871–878.

Slack, J. R., Lumb, A. M., and Landwehr, J. M. (1993). "Hydro-climatic data network (HCDN)." *Water-Resource Investigations Rep. 93-4076*, U.S. Geological Survey, Washington, D.C.

Stedinger, J. R., Sule, B. F., and Pei, D. (1983). "Multiple reservoir system screening models." *Water Resour. Res.,* 19(6), 1383–1393.

Tasker, G. D. (1983). "Effective record length for the T-year event." *J. Hydrol.,* 64, 39–47.

Thomas, H. A. (1948). "Frequency of minor floods." *J. Boston Soc. Civ. Eng.,* 35(1), 425–442.

Vogel, R. M. (1987). "Reliability indices for water supply systems." *J. Water Resour. Plan. Manage.,* 113(4), 563–579.

Vogel, R. M., and Bolognese, R. A. (1995). "Storage-reliability-resilience-yield relations for over-year water supply systems." *Water Resour. Res.,* 31(3), 645–654.

Vogel, R. M., Tsai, Y., and Limbrunner, J. F. (1998). "The regional persistence and variability of annual streamflow in the United States." *Water Resour. Res.,* 34(12), 3445–3459.

Vogel, R. M., and Wilson, I. (1996). "Probability distribution of annual maximum, mean, and minimum streamflow in the United States." *J. Hydrologic Eng.,* 1(2), 69–76.