Nonparametric Approach for Estimating Return Periods of Droughts in Arid Regions

Tae-Woong Kim[1]; Juan B. Valdés[2]; and Chulsang Yoo[3]

Abstract: Droughts cause severe damage in terms of both natural environments and human lives, and hydrologists and water resources managers are concerned with estimating the relative frequencies of these events. Univariate parametric methods for frequency analysis may not reveal significant relationships among drought characteristics. Alternatively, nonparametric methods provide local estimates of the univariate and multivariate density function by using weighted moving averages of the data in a small neighborhood around the point of estimation and opposed to parametric methods. A methodology for estimating the return period of droughts using a nonparametric kernel estimator is presented in order to examine the univariate as well as the bivariate behavior of droughts. After evaluating and validating a nonparametric kernel estimator, a drought frequency analysis is conducted to estimate the return periods of droughts for the Conchos River Basin in Mexico. The results show that, for the univariate analysis, the return periods of the severe drought occurring in the 1990s are 100 years or higher. For the bivariate analysis, the return periods are approximately 50 years for joint distributions and more than 120 years for the conditional distributions of severity and duration.

DOI: 10.1061/(ASCE)1084-0699(2003)8:5(1)

CE Database subject headings: Droughts; Arid lands; Frequency; Estimation.

Introduction

Hydrologic systems are impacted by extreme events such as floods and droughts, which cause severe damage to natural environments and human lives. For example, the 1987–1989 drought, which lasted 3 years and covered 36% of the United States, caused approximately $39 billion in losses of energy, water, the ecosystem, and agriculture (National 2000). These extreme events are usually expressed by return periods in hydrology and water resources engineering. When the concept of the return period is applied to drought-related variables, the return period will be the average time between events with a certain magnitude or less (Haan 1977). In general, a drought is defined as a sustained period of significantly lower soil moisture levels and water supply relative to normal levels. During droughts, water supplies are inadequate to meet the water demand of management systems, and lack of rainfall adversely affects the environment and human society (Dracup et al. 1980).

A significant number of studies on droughts deal with the definition of droughts, low-flow frequency analysis, and climatic impacts of droughts (Smakhtin 2001). Lee et al. (1986) developed a practical approach for the frequency analysis of multiyear drought durations of annual streamflow series. A technique that smooths the frequency-curve irregularity of drought durations was used to reduce the statistical uncertainties associated with sample-size limitations. Nathan and McMahon (1990) evaluated the application of the Weibull distribution to low-flow frequency analysis. They investigated the differences between low-flow frequency estimates based on calendar and hydrologic years, and they made recommendations for the selection of an appropriate subset of the data.

In recent years, several papers were published addressing the methodology for estimating drought return periods (Fernández and Salas 1999a,b; Chung and Salas 2000). Fernández and Salas (1999a,b) summarized the definitions of the return period and estimated the risk of failure of hydraulic structures, especially applied to drought events. They estimated the return periods and the associated risks of failure of hydrologic events related to meteorological droughts, low flows, annual maximum floods, and hydrological droughts, which are either dependent or independent. Chung and Salas (2000) dealt with drought occurrence probabilities, return periods, and risks of drought events for dependent hydrologic processes. Rather than using Markovian models, which traditionally have been used for modeling hydrologic processes, low-order discrete autoregressive moving average (ARMA) models were used for modeling wet and dry years, since they are adequate for processes exhibiting longer time dependence. They concluded that low order ARMA models are useful for representing the occurrence of wet and dry periods and, consequently, are capable of modeling and simulating drought conditions that are observed historically.

A general approach used in drought-related frequency analyses is to derive the distributions of drought durations and severities separately (Sen 1976; Mathier et al. 1992; Stedinger et al. 1992). A drought event, however, is a multivariate event characterized by its duration, magnitude, and intensity (Salas 1992), which are

[1]Graduate Research Assistant, Dept. of Civil Engineering and Engineering Mechanics, and Center for Sustainability of Semi-Arid Hydrology and Riparian Areas (SAHRA), The Univ. of Arizona, Tucson, AZ 85721-0072. E-mail: taek@email.arizona.edu

[2]Professor and Head, Dept. of Civil Engineering and Engineering Mechanics, and Center for Sustainability of Semi-Arid Hydrology and Riparian Areas (SAHRA), The Univ. of Arizona, Tucson, AZ 85721-0072. E-mail: jvaldes@u.arizona.edu

[3]Associate Professor, Dept. of Civil and Environmental Engineering, Korea Univ., Seoul, Korea 136-701. E-mail: cyoo@korea.ac.kr

Note. Discussion open until February 1, 2004. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on October 18, 2001; approved on January 21, 2003. This paper is part of the ***Journal of Hydrologic Engineering***, Vol. 8, No. 5, September 1, 2003. ©ASCE, ISSN 1084-0699/2003/5-1–10/$18.00.

mutually correlated. Separate analysis of drought characteristics cannot reveal the significant correlation relationships among drought characteristics (Shiau and Shen 2001). Recent studies have applied bivariate distributions to hydrologic frequency analysis (Yue 1999; Yue et al. 1999; Shiau and Shen 2001). Shiau and Shen (2001) suggested that a better approach for describing drought characteristics is to derive the joint distribution of drought duration and drought severity. Bivariate distributions have been also applied to flood frequency analysis, providing examples showing that flood characteristics, such as flood peak, volume, and duration, can be successfully represented by bivariate distributions (Yue 1999; Yue et al. 1999).

The frequency analysis is primarily based on the estimation of the probability density function (PDF). The parametric approaches for estimating the probability density function must assume that data are drawn from a known parametric family of distributions. However, there is no universally accepted distribution for hydrologic variables (Silverman 1986; Moon and Lall 1994; Smakhtin 2001). During the last decade, nonparametric methods for frequency analysis were introduced and examined as an alternative to parametric methods (Adamowski 1985, 1996; Moon and Lall 1994; Lall 1995). Nonparametric methods provide local estimates of the density function by using weighted moving averages of the data in a small neighborhood around the point of estimation (Lall and Bosworth 1994).

In this study, a nonparametric density function estimator is used to estimate the return periods of droughts in the Conchos River Basin in Mexico. We examine the univariate behavior, as well as the bivariate behavior of droughts conditioned with other properties of droughts. After brief descriptions of the basin and data of interest, univariate and multivariate nonparametric methods for estimating the probability density function are presented. Then, they are evaluated and validated for estimating the cumulative density function (CDF) using data in Texas Climatologic Region 5, which has a longer record than the Conchos River Basin. Next, the return periods of droughts are estimated by using nonparametric methods proposed in this study. They include both the univariate and the bivariate behavior of droughts. Finally, evaluations of historical droughts in the basin and discussions of the methodology for drought frequency analysis are presented.

Basin Information

The Conchos River Basin, shown in Fig. 1, lies within 26°N–30°N and 104°W–108°W in the arid/semiarid area of the Mexican state of Chihuahua, and has an area of 71,964 km^2. The waters of the Conchos River are used primarily for irrigation of nearly 80,000 ha, as well as for use in hydroelectric power plants located at La Boquilla and other minor dams. The Conchos River is the most important tributary of the Bravo/Grande River, since it supplies approximately 70–80% of the mainstream flow of the Lower Bravo/Grande River above the binational reservoir of Amistad. Important Mexican cities, such as Chihuahua, Hidalgo del Parral, and Delicias, are in the basin and are growing rapidly due to the increasing industrialization. Sustainability of the expanding water usage for agriculture, in addition to urban purposes and water rights of this region are of concern to the United States and Mexico.

As an indicator of drought severity, the Palmer drought severity index (PDSI) was used in this study. Palmer (1965) based this index on a balance between moisture supply and demand. It measures the departure of the moisture supply by taking into account

Fig. 1. Conchos River Basin, Mexico (Schmandt 2002)

the precipitation deficit at a certain location. Because it is a standardized value, the PDSI allows a comparison and assessment of a regional drought. In addition, it provides a measurement of the abnormality of weather for a region, as well as an opportunity to place current conditions in a historical perspective. Regional values of the PDSI were calculated using the areal mean of precipitation and temperature reconstructed across the basin.

In this study, a drought is defined, using the theory of runs, as an event during which the PDSI is continuously below a certain level, as shown in Fig. 2. Palmer (1965) categorized a drought condition using the PDSI. In dry spells (PDSI$<$0), a PDSI of -0.5 to 0 is considered near normal, -1.0 to -0.5 is an incipient drought, -2.0 to -1.0 is a mild drought, -3.0 to -2.0 is a moderate drought, -4.0 to -3.0 is a severe drought, and -4.0 or less is an extreme drought. For a selected threshold level (e.g., PDSI$=$0), a drought event is the set of negative runs (represented as shaded sets in Fig. 2), a drought duration (D_i) is the length of a drought event, a drought magnitude (M_i) is the area below the threshold level, and a drought intensity (I_i) is the ratio of magnitude over duration ($I_i = M_i / D_i$). In addition, a drought peak (P_i) is defined as the maximum deviation (minimum PDSI) during a drought event. Fig. 3 shows a time series of drought characteristics for the Conchos River Basin. Drought characteristics are individually uncorrelated random variables, since each time series of drought characteristics has the autocorrelation function, which is not significant at the 95% confidence level.

Kim et al. (2002) analyzed the drought characteristics in the Conchos River Basin using the PDSI and standardized precipitation index as indicators of drought severity. A drought intensity–

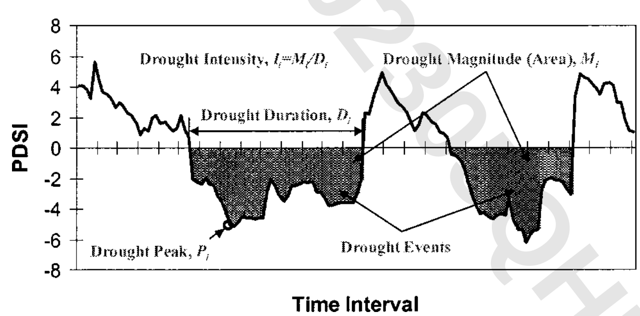

Fig. 2. Definition of drought characteristics using PDSI and concept of runs

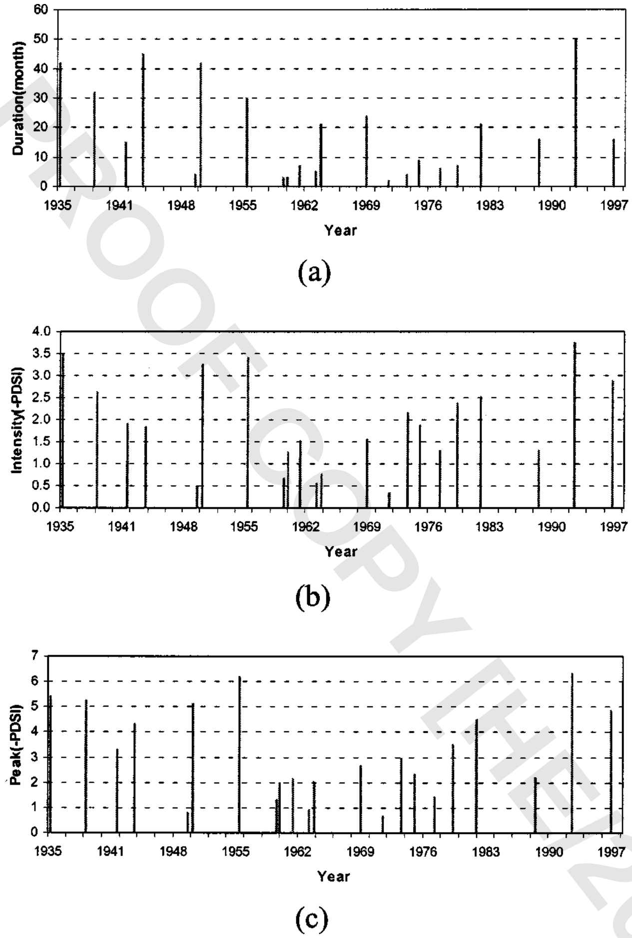

Fig. 3. Time series of drought characteristics in Conchos River Basin with PDSI threshold level=0, 1935–1998: (a) drought duration; (b) drought intensity; (c) drought peak

areal extent–frequency curve to characterize the spatial patterns of drought was also developed. The PDSI was shown to be useful for both defining and assessing the spatial extents of droughts in that basin.

Return Period of Droughts

The objective of a frequency analysis is to relate the magnitude of extreme events to their frequencies of occurrence through the use of probability distributions. It is usually assumed that the variables being analyzed are independent and identically distributed (Chow et al. 1988). The return period of an event in any observation is the inverse of its exceedance probability, $p=P(X>x_T)$; i.e.

$$T=\frac{1}{P(X>x_T)}=\frac{1}{1-P(X\leqslant x_T)} \tag{1}$$

where x_T=magnitude of the event having a return period of T.

In this study, the random variable X in Eq. (1) is drought duration. Because droughts sometimes last more than 1 year, the drought characteristics, shown in Fig. 3, can be analyzed as a partial duration series of independent events. Eagleson (1972) and Willems (2000) suggested that the distribution for partial duration series of independent events should be converted to an equivalent distribution for an annual exceedance series using the peak-over-threshold or partial-duration-series methods. If the marginal cumulative distribution of drought duration, d, for a given threshold level, is denoted by $F_D(d)$, the return period of drought duration, T_d, is defined in Eq. (2)

$$T_d(\text{years})=\frac{N}{n[1-F_D(d)]}=\frac{1}{\theta[1-F_D(d)]} \tag{2}$$

where $\theta=n/N$; N=total length of the observed PDSI (years); and n=total number of drought events, d, during N.

The bivariate return period of droughts can be estimated by substituting the univariate CDF into the bivariate CDF. For example, the bivariate return periods of drought duration (d) and drought intensity (i) are given by

$$T_{d,i}=\frac{1}{\theta[1-F_{D,I}(d,i)]} \tag{3}$$

$$T_{d|i}=\frac{1}{\theta[1-F_{D|I}(d|i)]} \tag{4}$$

where $T_{d,i}$=joint return period of drought duration and intensity; $T_{d|i}$=conditional return period of drought duration given intensity; $F_{D,I}(d,i)\,[=P(D\leqslant d,I\leqslant i)]$=joint cumulative distribution of drought duration and intensity; and $F_{D|I}(d|i)\,[=P(D\leqslant d|I\leqslant i)]$=conditional cumulative distribution of drought duration given intensity.

Parametric and Nonparametric Methods for Estimating PDF/CDF of Droughts

Basically, parametric methods for estimating the density function assume that a sample comes from a population with a given PDF. Nonparametric methods, however, are distribution free. Many studies on droughts and low-flow frequency analysis indicate that there is not a universally accepted distribution for drought-related variables (Smakhtin 2001). Further, the global assumption of parametric methods sometimes results in strongly biased estimates of the high or low quantiles when a variable of interest has a bimodal PDF (Sharma 2000).

Using the weighted moving averages of the records from a small neighborhood of the point of estimation, nonparametric function estimations have advantages that always reproduce the attributes represented by the sample (Lall 1995; Sharma 2000). Nonparametric density function estimations have become an active research topic in hydrology for frequency estimation and time series analysis (Lall 1995). Most nonparametric density estimation methods can be expressed by a kernel density estimator, which entails a weighted moving average of the empirical frequency distribution of the sample (Scott 1992; Sharma 2000).

Univariate Kernel Estimator

Given a set of observations $x_1,\ldots,x_n$, a mathematical expression of a univariate kernel probability density estimator is

$$\hat{f}_X(x)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h}\right) \tag{5}$$

where K=kernel function; and h=bandwidth that controls the variance of the kernel function. Table 1 shows examples of kernel functions typically used in hydrology and water resources engineering (Silverman 1986; Lall et al. 1996).

The choice of the bandwidth, h, is an important issue in estimating the probability density function, since the kernel estimator

Table 1. Examples of Univariate Kernel Functions

Kernel	$K(t)$
Epanechnikov	$K(t)=0.75(1-t^2),\ \lvert t\rvert\leqslant 1$
	$K(t)=0$, otherwise
Triangular	$K(t)=1-\lvert t\rvert,\ \lvert t\rvert\leqslant 1$
	$K(t)=0$, otherwise
Gaussian (normal)	$K(t)=(2\pi)^{-1/2}e^{-(t^2)/2}$
Rectangular	$K(t)=0.5,\ \lvert t\rvert\leqslant 1$
	$K(t)=0$, otherwise

is very sensitive to bandwidth (Moon and Lall 1994; Adamowski 1996). When h decreases, the estimated function displays the variance associated with individual observations. In other words, the bias diminishes, while the variance increases. As h increases, however, the opposite occurs. A large h results in a smoother function over the entire data space. For an asymptotically optimal choice for h, an overall measure of the effectiveness of $\hat{f}$ is provided by the mean integrated squared error (MISE), described in Eq. (6) (Bowman and Azzalini 1997)

$$\text{MISE}=E\left[\int\{\hat{f}(x)-f(x)\}^2dx\right]\approx\frac{1}{4}h^4\sigma_K^4\int f''(x)^2dx+\frac{1}{nh}\alpha \qquad (6)$$

where $\alpha=\int K^2(x)dx$; and σ_K^2 denotes the variance of the kernel function, $\int x^2K(x)dx$. The bandwidth is estimated such that it minimizes the MISE given by Eq. (6).

Bivariate Kernel Estimator

Bivariate kernel density estimators can be constructed in a similar manner. To expand the kernel density estimators to the bivariate distribution, bivariate kernel estimators are applied. This technique uses a product of univariate kernel functions with a set of bandwidths, (h_x,h_y). The joint PDF is estimated as follows:

$$\hat{f}_{XY}(x,y)=\frac{1}{nh_xh_y}\sum_{i=1}^{n}\left\{K\left(\frac{x-x_i}{h_x}\right)K\left(\frac{y-y_i}{h_y}\right)\right\} \qquad (7)$$

where n=number of observations (x_i,y_i); and h_x and h_y=bandwidths for the x- and y-direction, respectively. K is the kernel function, which is a radially symmetric unimodal probability density function.

A number of comparative studies of bandwidth selection for kernel estimation have been conducted concerning the choice of bandwidth that can be extended to the bivariate case. For example, Wand and Jones (1993) compared the performances of alternative smoothing parameterizations and indicated that the choice of appropriate bandwidths for each of the coordinate directions is important. They assumed that the kernel K is a radially symmetric probability density function, and the unknown density f has bounded and continuous second derivatives. Using the multidimensional form of Taylor's theorem, the approximate MISE for the bivariate case is (Silverman 1986)

$$\text{MISE}=\frac{1}{4}h^4\alpha^2\int\int\{\nabla^2f(x,y)\}^2dxdy+\frac{1}{nh^2\beta} \qquad (8)$$

where $\alpha=\int xyK(x,y)dxdy$; and $\beta=\int K(x,y)^2dxdy$. The optimal bandwidth for a bivariate kernel function will be obtained by minimizing the MISE given in Eq. (8).

Comparison of Nonparametric and Parametric Methods

Several studies have argued the advantage of using nonparametric methods for the frequency analysis of extreme events as an alternative to parametric methods (Adamowski 1985, 1996; Moon and Lall 1994; Guo et al. 1996). Adamowski (1985, 1996) showed that the nonparametric method is competitive with parametric counterparts for estimating floods and low-flow quantiles. In Monte Carlo simulation experiments, which compared the nonparametric method with the parametric log-Pearson Type III distribution, the nonparametric method gave more accurate results for simulated quantiles. Guo et al. (1996) developed a nonparametric kernel estimation model for estimating low-flow quantiles. Their results indicated that the nonparametric method had smaller bias and root-mean-square error in low-flow quantile estimates, and the nonparametric approach was a viable alternative to the Weibull models in applications to real data. Moon and Lall (1994) compared the kernel quantile estimator with several tail estimators for flood frequency. They also concluded that the nonparametric kernel estimator was shown to be competitive with other estimators.

Three parametric distributions, the lognormal, Gumbel (Extreme Value I), and Pearson Type III (gamma) distributions, were used to compare with the nonparametric kernel estimator. The probability-weighted moments method was used for estimating the parameters of the parametric distributions. Fig. 4 and Table 2 give comparisons of the PDF and the CDF for drought durations by the nonparametric kernel estimator with those of parametric methods. The nonparametric kernel estimator represents the multimodal characteristics of the observations better than parametric methods, since it is not bounded by a fixed functional form. Similarly, in the case of the cumulative density function, the nonparametric kernel estimator fits well the empirical points plotted using the Gringorten formula. Table 2 compares the sum of absolute bias (SAB) between empirical points and estimated points in the PDF and the CDF for the entire period of record. The nonparametric kernel estimator has the lowest SAB for both PDF and CDF in comparison to the three alternative parametric distributions when the entire period of record is used in the calibration. The nonparametric kernel estimator will improve the accuracy of the frequency analysis if the observed points in the tail of the PDF are significant, which are not outliers or observational errors, because our main concern lies in estimating the CDF in the tail of the PDF.

Validation of Nonparametric Kernel Estimator

To examine the stability of estimating the CDF of the nonparametric kernel estimator, the absolute relative bias (ARB) was chosen as another criterion for comparison

$$\text{ARB}_i=\left|\frac{EP_{i,T}-\text{CDF}_{i,t}}{EP_{i,T}}\right| \qquad (9)$$

where $EP_{i,T}$=Gringorten plotting position points for observations during entire record period, T; and $\text{CDF}_{i,t}$=estimated CDF points for the subrecord period, t. To further analyze the performance of the methods, the PDSI for Texas Climatologic Region 5 (National Climatic Data Center, http://www.ncdc.noaa.gov/) was used for validation. Texas Climatologic Region 5 is in the immediate vicinity of the Conchos River Basin and has a much longer record period. We used two subperiods, 1895–1950 and 1935–1998, and compared the CDF for the subperiods with the CDF for the entire

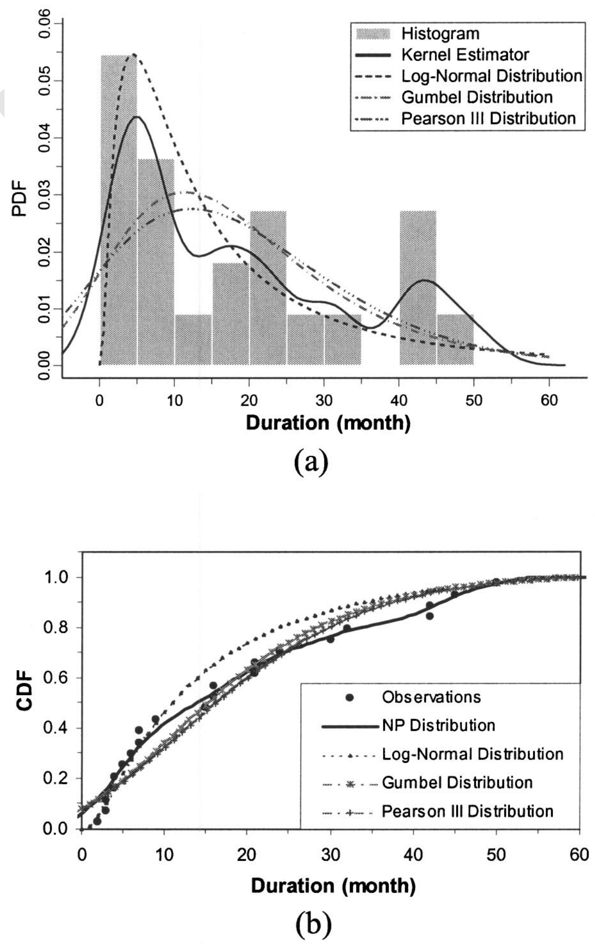

Fig. 4. PDFs and CDFs for drought duration estimated by parametric and nonparametric method (NP), Conchos River Basin, Mexico, 1935–1998: (a) histogram and PDFs; (b) CDFs

record period, 1895–1998. Both SAB and ARB measure systematic errors in the estimation of the CDF, which means the consistency of estimation of the CDF above or below the actual value (EP) over the entire period and the subperiod of observations, respectively. Table 3 compares the sum of ARB (SARB) and the average of ARB (AARB). The results in Table 3 show that the CDF estimated by the nonparametric kernel estimator has less bias and gives more stable results for estimating the CDF than the parametric methods under both criteria for both validation periods.

Applications

The nonparametric kernel estimator for estimating the PDF and CDF of droughts has proven to be more attractive than parametric

Table 2. Sum of Absolute Bias of PDF and CDF Estimated by Parametric and Nonparametric Method for Conchos River Basin, 1935–1998

Function	NP	Lognormal	Gumbel	Pearson III
PDF	0.0661	0.0950	0.1220	0.1237
CDF	0.3902	0.9504	0.8677	0.9062

Note: NP=nonparametric.

Table 3. Absolute Relative Bias of CDF Estimated by Parametric and Nonparametric Method for Texas Climatologic Region 5

Criterion	Period	NP	Lognormal	Gumbel	Pearson III
SARB	1895–1950	0.9155	6.1038	2.5230	2.0005
	1935–1998	3.7728	11.1731	6.6045	4.3118
AARB	1895–1950	0.0381	0.2543	0.1051	0.0834
	1935–1998	0.1509	0.4469	0.2642	0.1725

Note: SARB=sum of the absolute relative bias; AARB=average of the absolute relative bias; NP=nonparametric.

methods, using the criteria for this case study under the assumption that the PDF and CDF are alternative smooth curves in order to represent the histogram and the cumulative distribution, respectively. Bowman and Azzalini (1997) provided an introduction to kernel-based methods with an easy-to-use S-plus library, SM, which we applied to drought frequency analysis. A Gaussian density function was used for a smooth kernel function with a mean of zero and standard deviation of h. When a Gaussian kernel function is used, $\hat{f}$ will be a smooth curve having derivatives of all orders to satisfy the condition that $\hat{f}$ is inherently continuous and differentiable by definition. When K is a normal density, the evaluation of the optimal formula for h yields

$$h_{d,\mathrm{opt}} = \left[\frac{4}{n(p+2)}\right]^{1/(p+4)} \sigma_d \tag{10}$$

where $h_{d,\mathrm{opt}}$=optimal bandwidth; σ_d denotes the standard deviation of the distribution in dimension d; and p=number of dimensions, e.g., $p=1$ for a univariate kernel estimator and $p=2$ for a bivariate kernel estimator.

Univariate Analysis

To conduct the univariate analysis for droughts using the nonparametric technique, time series of drought durations were constructed. The time series have PDSI thresholds of 0.0, −0.5, −1.0, −2.0, −3.0, and −4.0 for the Conchos River Basin. For example, Fig. 3(a) shows the drought duration time series for a threshold of PDSI=0.0. The sample statistics for the sets of drought durations are shown in Table 4, where θ is the number of drought events divided by the data period (years) (n/N), which was used in Eqs. (2)–(4). The sample size for estimating the PDF is identical to the number of drought events. As the sample size increases, both the parametric and the nonparametric methods have better stability in estimating the PDF. Lall and Moon (1993) have shown that the nonparametric estimator is more useful for small samples than parametric methods, since the relative efficiency calculated from the mean square error

Table 4. Sample Statistics of Drought Duration (Month) Series in Conchos River Basin, 1935–1998

Parameter	0.0[a]	−0.5[a]	−1.0[a]	−2.0[a]	−3.0[a]	−4.0[a]
Number of events	22.000	23.000	26.000	30.000	18.000	14.000
Mean	18.360	16.910	13.380	7.700	7.060	5.430
Standard deviation	15.500	15.530	13.470	10.980	7.460	7.130
Skewness	0.780	0.960	1.420	2.270	2.190	2.740
Kurtosis	−0.680	−0.400	1.220	4.240	5.750	8.550
θ	0.339	0.354	0.400	0.461	0.277	0.215

[a]Threshold.

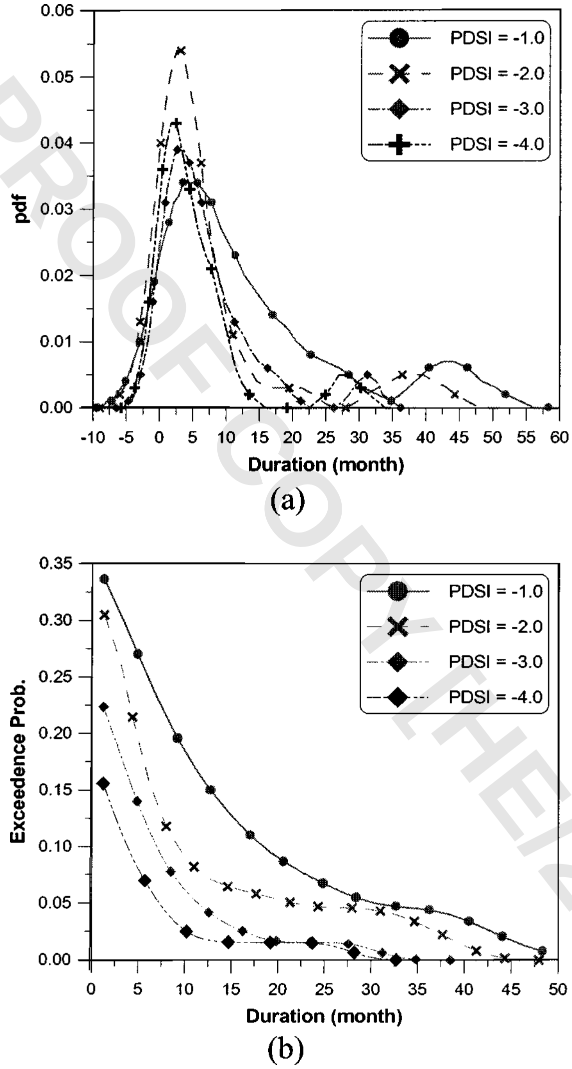

Fig. 5. Comparison of marginal PDF and exceedance probability for drought durations estimated by nonparametric method for threshold levels (TH), Conchos River Basin, Mexico, 1935–1998: (a) marginal PDFs; (b) exceedance probabilities

$[\mathrm{MSE}(T_P)/\mathrm{MSE}(Q_P)]$, where Q_P is a nonparametric quantile estimator and T_P is the parametric estimator, is greater than 1 (Harrell and Davis 1982).

Fig. 5 shows the marginal PDF and exceedance probability for drought durations in the Conchos River Basin. The marginal PDF was estimated by the univariate kernel density estimator with optimized bandwidth for the sample size. The exceedance probability was calculated from the CDF, which was estimated by integrating the PDF, multiplied by θ {e.g., $\theta[1-F_D(d)]$}. Fig. 5(a) displays the PDFs constructed from four data sets corresponding to the threshold levels, and Fig. 5(b) compares the exceedance probabilities estimated from the PDF, in which the underlying shapes contrast more effectively. The PDF for the Conchos River Basin has a more apparent bimodal shape, which causes the exceedance probabilities to be flat in the tail, as shown in Fig. 5(b). Classical parametric estimation procedures are weighted toward fitting the main body of the probability density, and the weights are negligible in the tail of the distribution. Considerable uncertainties for the magnitude of extreme events exist even if the

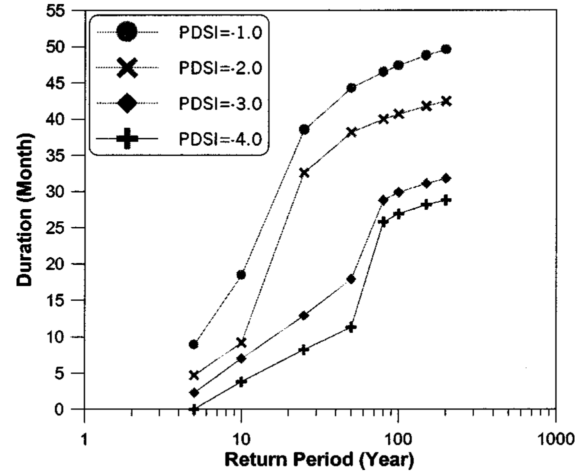

Fig. 6. Return periods of drought durations as function of threshold levels (TH), Conchos River Basin, Mexico, 1935–1998

parametric PDF fits well (Moon and Lall 1994). The return periods of drought durations for given threshold levels are shown in Fig. 6.

Bivariate Analysis

Shiau and Shen (2001) presented a joint PDF of drought duration and severity, expressed as the product of the conditional distribution of drought severity given duration and the marginal distribution of duration. The derivation of the joint PDF of drought characteristics is not mathematically tractable, because no known analytical solutions are available.

In this study, a bivariate kernel estimator is also employed to represent the bivariate behavior of droughts (duration-intensity, duration-peak) in the Conchos River Basin. For the observations of 64 years (1935–1998), 22 drought events are below a PDSI of 0.0, which represents a drought condition. The sample statistics of drought severities of interest for the bivariate study are presented in Table 5, where drought duration is highly correlated with intensity and peak. Fig. 7 shows the joint PDF (JPDF) for drought duration and intensity, and for drought duration and peak estimated directly from the data using a bivariate kernel estimator. The JPDF for duration and peak has an apparent principal axis in the diagonal direction due to their high correlation. Parametric models are difficult to apply because of the high correlation between the two variables of interest. For example, in a Gumbel mixed model with standard Gumbel marginal distributions, the cross-correlation coefficient should have a value between 0.0 and 0.67 (Oliveria 1975; Yue et al. 1999). For both duration-intensity and duration-peak, the correlations are above the upper limit.

Table 5. Sample Statistics of Drought Severities for Bivariate Analysis in Conchos River Basin, 1935–1998

Parameter	Intensity (PDSI)	Peak (PDSI)
Maximum	−0.340	−0.680
Minimum	−3.740	−6.310
Mean	−1.910	−3.190
Standard deviation	1.030	1.790
Skew	−0.210	−0.300
Kurtosis	−1.000	−1.180
Correlation with duration	0.724	0.823

PROOF COPY [HE/2001/022449] 002305QHE

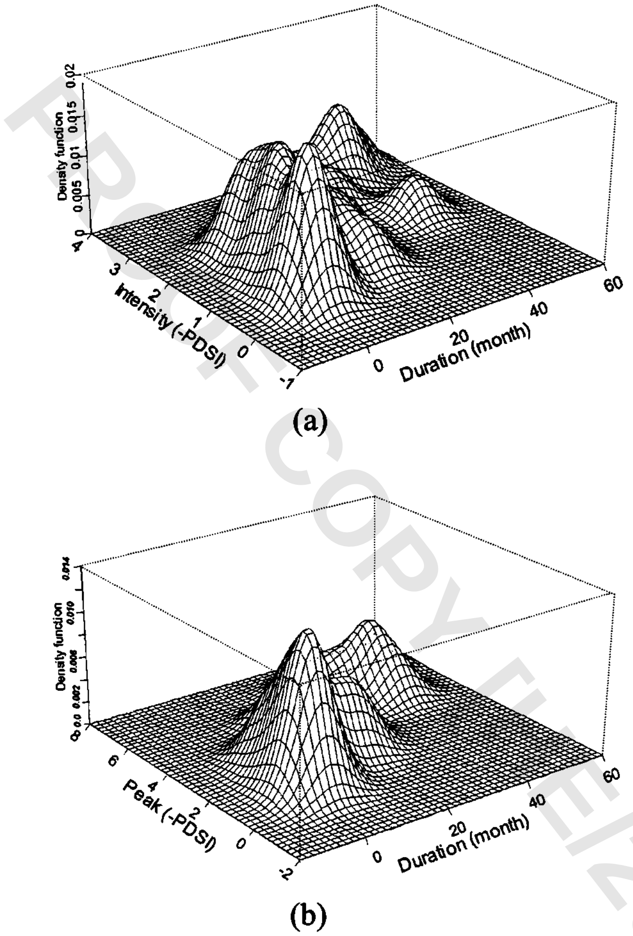

Fig. 7. Joint probability density function (JPDF) for: (a) duration-intensity; (b) duration-peak, Conchos River Basin, Mexico, 1935–1998

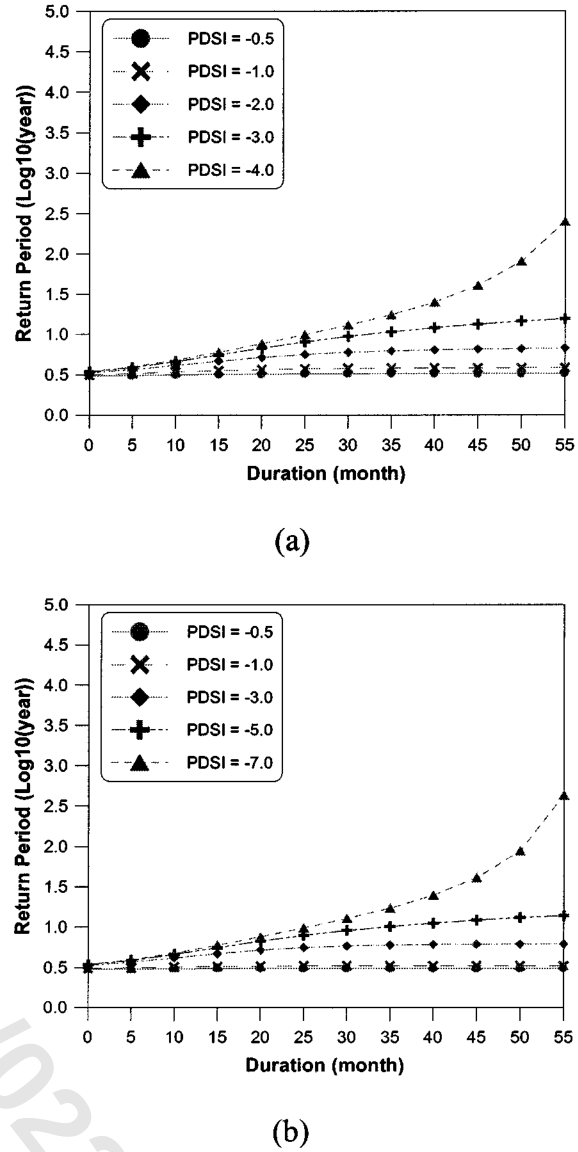

Fig. 8. Joint return period for duration corresponding to: (a) intensity; (b) peak

Joint Distributions of Drought Duration and Intensity/Peak

The estimation of the joint return periods of T_{di} (duration and intensity) and T_{dp} (duration and peak) is of interest in this study. First, the joint probability density function for durations corresponding to a certain intensity value, $P(D,I\leqslant i_0)$, was estimated from the JPDF by slicing the JPDF between 0 and i_0 and summing up for each duration. Then, the joint cumulative density function for duration corresponding to a certain intensity value, $P(D\leqslant d,I\leqslant i_0)$, was constructed by integrating the joint probability density function, $P(D,I\leqslant i_0)$. Similarly, the joint probability density function for durations corresponding to a certain peak value, $P(D,P\leqslant p_0)$, and the joint cumulative density function for durations corresponding to a certain peak value, $P(D\leqslant d,P\leqslant p_0)$, were constructed. Finally, using Eq. (3), the joint return periods of T_{di} and T_{dp} were evaluated, and are shown in Fig. 8.

Conditional Distributions of Drought Duration and Intensity/Peak

The estimation of the conditional return periods of $T_{d|i}$ (duration given intensity) and $T_{d|p}$ (duration given peak) is also of interest in this study. The conditional probability density function for duration, given a certain intensity, $P(D|I\leqslant i_0)$, was estimated by dividing the joint probability density function, $P(D,I\leqslant i_0)$, by the marginal distribution for intensity, $P(I\leqslant i_0)$. Then, the conditional cumulative density function for duration given a certain intensity value, $P(D\leqslant d|I\leqslant i_0)$, was constructed by integrating the conditional probability density function, $P(D|I\leqslant i_0)$. Similarly, the conditional probability density function for duration given a certain peak value, $P(D|P\leqslant p_0)$, and the conditional cumulative density function for duration given a certain peak value, $P(D\leqslant d|P\leqslant p_0)$, were constructed. Finally, using Eq. (4), the conditional return periods of duration given intensity, $T_{d|i}$, and given peak, $T_{d|p}$, were calculated, and are presented in Fig. 9.

Historical Droughts in Conchos River Basin

Historical droughts occurring in the study area were evaluated and their return periods were calculated. The climate of the Conchos River Basin is characterized by low precipitation and high variability, because the Conchos River Basin is isolated from moisture sources and affected by the subtropical high-pressure belt. However, the 1990s rainfall in the Conchos River Basin was significantly below normal. Using the drought intensity–areal

PROOF COPY [HE/2001/022449] 002305QHE

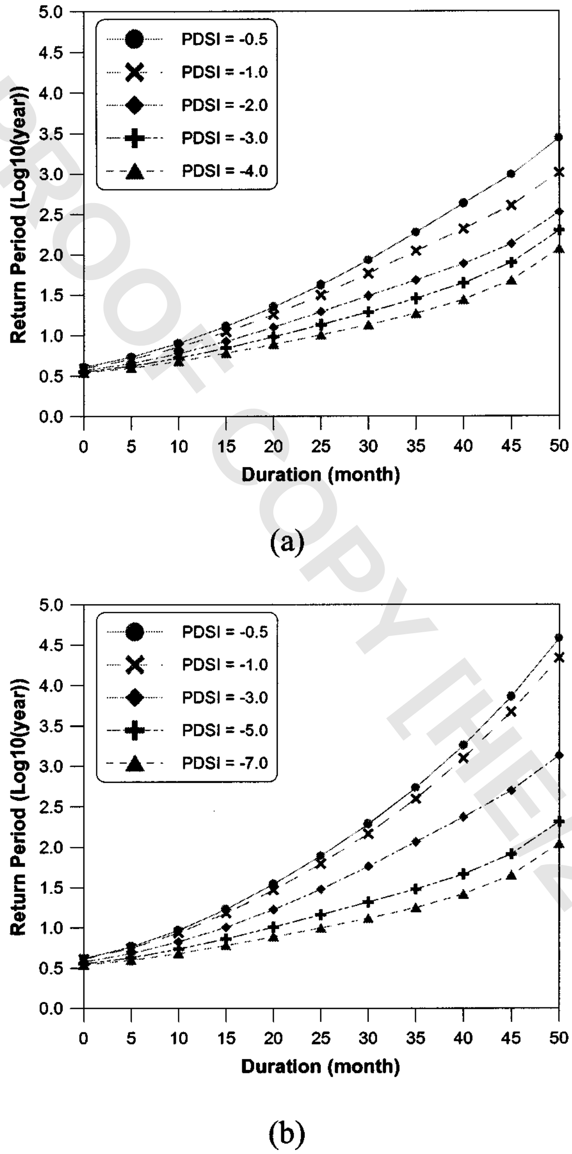

Fig. 9. Conditional return period for duration given: (a) intensity; (b) peak

extent–frequency curve, Kim et al. (2002) indicated that the droughts occurring in the 1990s in this basin were associated with return periods of 80–100 years with large areal extents. In the Conchos River Basin, the droughts in the 1990s drove many small farmers off the land and produced a 30% reduction in the cattle inventory (Comisión 1997). In addition, due to the severe reduction of the Conchos inflow to the Bravo/Grande River in the 1990s, Mexico owed the United States about 1,240 Mm^3 with respect to water delivery under the 1944 U.S./Mexico Treaty on the use of the Bravo/Grande River (Texas 2001).

Table 6 contains characteristics of the severest drought in the 1990s with respect to threshold levels. Fig. 10(a) shows that the marginal return periods of droughts in the Conchos River Basin in the 1990s, estimated from the univariate analysis, were higher than 100 years. The droughts with threshold levels of PDSI $=-3.0$ (severe drought) and -4.0 (extreme drought) have a return period of 140 years. For the bivariate analysis (threshold $=0.0$), the severest drought in the Conchos River Basin lasted 50 months (duration=50) from May 1993 with an intensity of -3.74, and peak of -6.31. Fig. 10(b) shows that the return periods of this drought are around 50 years for the joint distribution of duration-intensity $[T_{d(=50)i(=-3.74)}]$, and duration-peak $[T_{d(=50)p(=-6.31)}]$, and 130 and 125 years for the conditional distribution of duration given intensity $[T_{d(=50)|i(=-3.74)}]$, and given peak $[T_{d(=50)|p(=-6.31)}]$, respectively. The joint and conditional return periods (Figs. 8 and 9) are useful for practical design and planning purposes, since both drought duration and severity may be considered. For example, it is possible to obtain various combinations of drought duration and drought severity for several purposes in hydrological engineering design and management.

Table 6. Characteristics of Severest Drought in Conchos River Basin

Parameter	0.0[a]	−0.5[a]	−1.0[a]	−2.0[a]	−3.0[a]	−4.0[a]
Year	1993	1993	1993	1994	1994	1994
Month	May	May	May	June	July	August
Duration (Months)	50	50	48	34	31	28
Intensity (PDSI)	−3.74	−3.24	−2.87	−2.79	−2.02	−1.18
Peak (PDSI)	−6.31	−5.81	−5.31	−4.31	−3.31	−2.31

[a]Threshold.

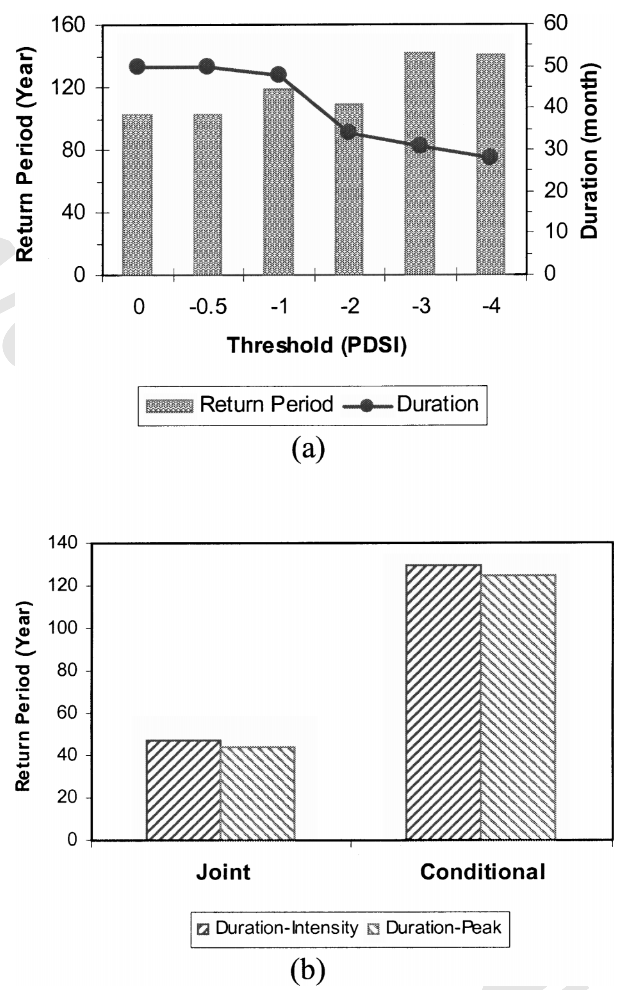

Fig. 10. Return period of historical severest droughts in Conchos River Basin, 1935–1998: (a) univariate analysis; (b) bivariate analysis $[T_{d(=50)i(=-3.74)}$, $T_{d(=50)|i(=-3.74)}$, $T_{d(=50)p(=-6.31)}$, and $T_{d(=50)|p(=-6.31)}$, from left]

Concluding Remarks

A nonparametric methodology for frequency analysis to determine the return period of drought is presented in this paper. The drought characteristics were defined using a drought index, and the univariate behavior and bivariate behavior of drought characteristics were analyzed using the proposed methodology. The nonparametric kernel estimator used in this study is capable of constructing the PDF and the CDF as an alternative tool to the parametric method, since it provides local estimates using weighted moving averages of points of estimation. The nonparametric method is also shown to be competitive for frequency analysis, because it does not require any assumption of distribution and can fit a multimodal density function.

In this study, using the PDSI, we were able to consider drought severity as well as its duration. The methodology for estimating the univariate and bivariate return periods of drought durations, which can be estimated from the nonparametric kernel estimator, has been presented and applied to an arid region, the Conchos River Basin in Mexico. Because the droughts are characterized by highly correlated random variables, such as duration, intensity, and peak, their joint and conditional probabilistic behaviors, as well as their marginal distributions, need to be analyzed in order to understand multivariate drought events. In the Conchos River Basin, the drought occurred in the 1990s; it was the severest drought in the region during the record period, 1935–1998. It had a marginal return period of more than 100 years and, considering the bivariate behavior, it had a return period of around 130 years for the conditional distribution of drought duration given intensity.

Using joint and conditional distributions, the limitation of the univariate frequency analysis could be overcome, and the multivariate behavior of droughts could be understood more thoroughly. Further, applying a nonparametric method to estimate the probability distribution of droughts could eliminate some of the limitations of parametric methods. The overall results in this study indicate that the proposed method is useful for determining the return period of a drought and evaluating the drought that occurred in the past. In particular, the bivariate return periods estimated in this study are useful for both design and water management, since both drought duration and severity may be determined more easily and simultaneously in the bivariate nonparametric method for frequency analysis.

Acknowledgments

This study is based on works supported by the Center for Sustainability of Semi-Arid Hydrology and Riparian Areas at the University of Arizona under the STC Program of the National Science Foundation, Agreement Number EAR-9876800. We are grateful to Dr. Javier Aparicio of the Mexican Institute of Water Technology and Dr. Tereza Cavazos of the Department of Physical Oceanography, CICESE, Mexico, for assisting with data preparation and providing valuable comments.

References

Adamowski, K. (1985). ''Nonparametric kernel estimation of flood frequencies.'' *Water Resour. Res.*, 21(11), 1885–1890.

Adamowski, K. (1996). ''Nonparametric estimation of low-flow frequencies.'' *J. Hydraul. Eng.*, 122(1), 46–49.

Bowman, A., and Azzalini, A. (1997). *Applied smoothing techniques for data analysis: The kernel approach with S-Plus illustrations*, Oxford University Press, New York.

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

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

Comisión Nacional del Agua (CNA). (1997). ''*Plan Hidráulico Gran Visión del Estado de Chihuahua 1996–2020*.'' *Rep.*, Chihuahua, Mexico (in Spanish).

Dracup, J. A., Lee, K. S., and Paulson, E. G. (1980). ''On the definition of droughts.'' *Water Resour. Res.*, 16(2), 297–302.

Eagleson, P. S. (1972). ''Dynamics of flood frequency.'' *Water Resour. Res.*, 8(4), 879–898.

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

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

Guo, S. L., Kachroo, R. K., and Mngodo, R. J. (1996). ''Nonparametric kernel estimation of low flow quantiles.'' *J. Hydrol.*, 185, 335–348.

Haan, C. T. (1977). *Statistical methods in hydrology*, Iowa State University Press, Ames, Iowa.

Harrell, F. E., and Davis, C. E. (1982). ''A new distribution free quantile estimator.'' *Biometrika*, 69(3), 635–640.

Kim, T., Valdés, J. B., and Aparicio, J. (2002). ''Frequency and spatial characteristics of droughts in the Conchos River Basin, Mexico.'' *Water Int.*, 27(3), 420–430.

Lall, U. (1995). ''Nonparametric function estimation: Recent hydrologic contributions.'' *Rev. Geophys.*, ■, 1093–1102.

Lall, U., and Bosworth, K. (1994). ''Multivariate kernel estimation of functions of space and time hydrologic data.'' *Stochastic and statistical methods in hydrology and environmental engineering*, Vol. 3, K. V. Hipel, A. I. Mcleod, U. S. Panu, and V. P. Singh, eds., Kluwer Academic, Boston, 301–315.

Lall, U., and Moon, Y. (1993). ''Kernel flood frequency estimators: Bandwidth selection and kernel choice.'' *Water Resour. Res.*, 29(4), 1003–1015.

Lall, U., Rajagopalan, B., and Tarboton, D. G. (1996). ''A nonparametric wet/dry spell model for resampling daily precipitation.'' *Water Resour. Res.*, 32(9), 2803–2823.

Lee, K. S., Sadeghipour, J., and Dracup, J. A. (1986). ''An approach for frequency analysis of multiyear drought durations.'' *Water Resour. Res.*, 22(5), 655–662.

Mathier, L., Perreault, L., Bobe, B., and Ashkar, F. (1992). ''The use of geometric and gamma-related distributions for frequency analysis of water deficit.'' *Stochastic Hydrol. Hydr.*, 6(4), 239–254.

Moon, Y., and Lall, U. (1994). ''Kernel quantile function estimator for flood frequency analysis.'' *Water Resour. Res.*, 30(11), 3095–3103.

Nathan, R. J., and McMahon, T. A. (1990). ''Practical aspect of low-flow frequency analysis.'' *Water Resour. Res.*, 26(9), 2135–2141.

National Oceanic and Atmospheric Administration Paleoclimatology Program. (2000). ''*North American drought: A paleo perspective*.'' ⟨http://ngdc.noaa.gov/paleo/drought/⟩.

Oliveria, J. T. D. (1975). ''Bivariate extremes: Extensions.'' *Bulletin of the International Statistical Institute*, 46(2), 310–319.

Palmer, W. C. (1965). ''Meteorological drought.'' *Research Paper No. 45*, U.S. Weather Bureau, Washington, D.C.

Salas, J. D. (1992). ''Analysis and modeling of hydrologic time series.'' *Handbook of hydrology*, D. R. Maidment, ed., McGraw-Hill, New York, 19.1–19.72.

Schmandt, J. (2002). ''Bi-national water issues in the Rio Grande/Río Bravo Basin.'' *Water Policy*, 4(2), 137–155.

Scott, D. W. (1992). *Multivariate density estimation: Theory, practice and visualization*, Wiley, New York.

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

Sharma, A. (2000). ''Seasonal to interseasonal rainfall probabilistic forecasts for improved water supply management: Part 3—A nonparametric probabilistic forecast model.'' *J. Hydrol.*, 239, 249–258.

Shiau, J.-T., and Shen, H. W. (2001). ''Recurrence analysis of hydrologic droughts of differing severity.'' *J. Water Resour. Plan. Manage.*, 127(1), 30–40.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, Chapman & Hall, London.

Smakhtin, V. U. (2001). ''Low flow hydrology: A review.'' *J. Hydrol.*, 240, 147–186.

Stedinger, J. R., Vogel, R. M., and Foufoula-Georgiou, E. (1992). ''Frequency analysis of extreme events.'' *Handbook of hydrology*, D. R. Maidment, ed., McGraw-Hill, New York, 18.1–18.66.

Texas Center for Policy Studies (TCPS). (2001). ''*The Rio Conchos: A preliminary overview*.'' *Rep.*, Austin, Tex.

Wand, M. P., and Jones, M. C. (1993). ''Comparison of smoothing parameterizations in bivariate kernel density estimation.'' *J. Am. Stat. Assoc.*, 88(422), 520–528.

Willems, P. (2000). ''Compound intensity/duration/frequency-relationships of extreme precipitation for two seasons and two storm types.'' *J. Hydrol.*, 233, 189–205.

Yue, S. (1999). ''Applying bivariate normal distribution to flood frequency analysis.'' *Water Int.*, 24(3), 248–254.

Yue, S., Ouarda, T. B. M. J., Legendre, P., and Bruneau, P. (1999). ''The Gumbel mixed model for flood frequency analysis.'' *J. Hydrol.*, 226, 88–100.

PROOF COPY [HE/2001/022449] 002305QHE