Bivariate Drought Recurrence Analysis Using Tree Ring Reconstructions

Javier González[1] and Juan B. Valdés, F.ASCE[2]

Abstract: Droughts may be represented by two main characteristics—duration and severity. In this paper, a general methodology to evaluate the frequency and risk of the occurrence of droughts is presented using a bivariate drought characterization. The theory of runs is applied to model drought recurrence as an alternating renewal process, describing droughts simultaneously in terms of their durations and severities. Short historical records usually do not allow reliable bivariate analyses. However, tree ring reconstructions of droughts provide information about past events, allowing the analysis. An approach to adapt and include dendrochronology reconstructions combined with historical records to characterize droughts is presented. The proposed approach uses the stochastic structure of the residuals of paleo reconstructions to generate equally likely representations of past drought events. The procedure was applied to paleo and historical records in Texas Climatic Division 5 and compared with univariate analyses. The application shows the bivariate analysis advantages in drought characterization.

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

CE Database subject headings: Droughts; Frequency analysis; Dendrochronology; Probability distribution functions.

Introduction

A drought is ‘‘a period of abnormally dry weather sufficiently prolonged for the lack of water to cause serious hydrologic imbalance in the affected area’’ (Huschke 1959). This departure from the norm can produce serious agricultural, environmental, and socioeconomical damage. Historically, a measure to describe how extreme a drought is has been sought.

The Palmer drought severity index (PDSI, now usually named PDI) was developed for this purpose (Palmer 1965). Several studies have examined the PDSI in terms of its assumptions and limitations (Karl 1983, 1986; Alley 1984; Guttman 1991; Heddinghaus and Sabol 1991; Guttman et al. 1992). The oversimplistic characterization of drought physical processes, the arbitrariness involved in its definition, and its space dependence are the most common drawbacks associated with the PDSI. However, it is computed for and used in all 48 contiguous states in the United States. Based on the extensive use of the PDSI, Guttman et al. (1992) recommended to use parameters derived from the statistical analysis of the index, such as the return period (already widely used in water resources), to characterize the rareness of a drought event in space independent form.

Return periods and, in general, recurrence and risk analysis are common issues in hydrologic engineering. The return period is an important design parameter in water resources systems. The risk of failure during the life of the project must be limited. However, the estimation of these statistical properties depends on the availability of historical data, which is, usually, the most important limitation to directly inferring the probability of extreme episodes of droughts. To overcome this lack of information, stochastic representation of hydrologic events has been employed to establish probabilistic models that describe reality and make it possible to theoretically infer the associated likelihood of drought events. Yevjevich (1967) successfully proposed the theory of runs as a major tool to use in objectively defining droughts and studying their statistical properties. Runs theory has been applied in several drought models and analyses in the literature (Sen 1976, 1980; Dracup et al. 1980a,b; Loaiciga and Leipnik 1996; Fernández and Salas 1999; Chung and Salas 2000; Shiau and Shen 2001; Douglas et al. 2002). Traditionally, drought duration and severity are the two most common characteristics of a drought (Dracup et al. 1980a,b). In these applications of runs theory, drought durations and severities have been analyzed separately. This may result in assigning two different probabilities for a single event (Frick et al. 1990). Nevertheless, the high correlation between duration and severity allows the estimation of severity based on duration or vice versa.

Kim et al. (2003) presented a methodology for estimating the return period of droughts using a nonparametric kernel estimator for univariate and bivariate behaviors. Bivariate analyses have also been applied to other hydrologic events, such as storms and streamflows (Yue 1999; Yue et al. 2001). However, bivariate analyses require record lengths that are usually not available, and such analyses remain in many cases infeasible. Paleoclimatology studies, and especially dendroclimatology, form a valuable source of information to analyze drought recurrence. Many efforts have been made in reconstructing drought records from tree rings in the United States (Stockton and Meko 1975; Blasing and Duvick 1984; Cook et al. 1992). An attempt to expand these drought reconstructions in a homogeneous way to the entire continental United States was initiated by Meko et al. (1993), and

[1]Graduate Student, Dept. of Civil Engineering and Engineering Mechanics, The Univ. of Arizona, Tucson, AZ 85721-0072.

[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.

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 August 16, 2002; approved on December 4, 2002. This paper is part of the ***Journal of Hydrologic Engineering***, Vol. 8, No. 5, September 1, 2003. ©ASCE, ISSN 1084-0699/2003/5-1–12/$18.00.

was developed further by Cook et al. (1996). A 2° latitude×3° longitude grid of summer drought reconstruction for the continental United States was developed using annual tree ring chronologies (Cook et al. 1999). The PDSI was used as a drought metric. However, the incorporation of tree ring reconstructions of PDSI series in statistical analyses and its combination with PDSI series evaluated from instrumental data are not straightforward. Reconstructions do not increase the number of degrees of freedom of the time series in the number of reconstructed years (Brockway and Bradley 1995). Such reconstructions explain only part of the variance of instrumental PDSI series, and an important proportion of the PDSI variance remains unexplained. As such, PDSI reconstructions can be used to predict tendencies, but cannot be compared directly with instrumental series.

A methodology is presented in this paper that, taking into account the uncertainty associated with the tree ring reconstructions, generates likely realizations of past drought events useful for drought recurrence analysis. The conjunction of tree ring reconstructions and historical data may provide, in many cases, enough information for a complete bivariate analysis of drought occurrence. For those cases, a theoretical derivation of the return period of droughts being jointly characterized by durations and severities is developed based on stochastic properties and runs theory. Finally, for risk analysis, a general algorithm is presented to compute the distribution of the interarrival time of droughts. The methodology is applied to the reconstruction of summer PDSI in Texas Climatic Division 5. Univariate and bivariate analyses are compared for this case.

Dendrochronology Reconstruction of Droughts

Dendroclimatology is the dating of climatic past events through the study of tree ring growth. The precise dating and annual resolution of tree rings are superior to most sources of paleoclimatic information. Tree rings probably offer the best means of reconstructing large-scale and highly resolved patterns of climate (Fritts 1976, 1991; Cook and Kairiukstis 1990; Cook et al. 1994).

Drought reconstructions from tree rings explain a fraction of the variance of the index applied to study droughts (PDSI, streamflow, rainfall). Thus, a reconstructed index has less variability than the original index. This has influence in drought analyses, because the estimated anomalies are generally lower than the expected values. That is, the mean of the reconstructed drought deficit is underestimated. These reconstructions can be used for identifying dominant frequencies of climatic variability [for example, the bidecadal drought rhythm analyzed in Cook et al. (1997)]. However, combining them in a statistical analysis of droughts with instrumental data is not simple. As an example, in the drought reconstruction for the continental United States (Cook et al. 1999; www.ncdc.noaa.gov), the PDSI Grid Reconstruction Number 63, in Texas, has a correlation coefficient of $\rho=0.67$. However, in dry years ($\text{PDSI}\leqslant -1$) the original PDSI and the reconstruction simultaneously indicate dry years only 47% of the time. In this paper, a methodology to convert such valuable information in a form that can be compared with instrumental data was developed. Following, the fundamentals of such a methodology are described through a tree ring drought reconstruction for Texas Climatic Division 5.

The PDSI series from Texas Climatic Division 5 (National Climate Data Center, www.ncdc.noaa.gov) was selected as a case study. Monthly PDSI data from 1895–2001 were used in this study. This division is close to the Bravo/Grande River, where drought recurrence analysis has important relevance in a U.S.–Mexico treaty that regulates the use of the water resources in the lower Grande Basin (Texas 2000). A 6°×6° search area of tree ring chronologies was established at 29N-105W, close to the centroid of Texas Climatic Division 5. Seven chronologies of different length were obtained from contributors of the international tree ring data bank (IGBP 2002).

Correlation analysis of the chronologies with the monthly PDSI series showed a higher correlation of tree rings with the PDSI of summer months (June, July, and August), as already mentioned in Cook et al. (1992). Consequently, the reconstruction is only of the average PDSI during the summer months [as in Cook et al. (1999)]. However, the summer season coincides with the rainiest season in this region; therefore, the summer PDSI is also a good estimator of the annual PDSI ($\rho_{\text{Summer-Annual}}=0.94$). Thus, the temporal scale in this work is annual. The tree ring variables are called the predictors, and the summer's PDSI is the predictand (Fritts 1976, 1991). Given the differences in length of the chronologies, the final reconstruction is composed of five different sub-reconstructions, coming from five different regression models. Those models are mathematically identical, but each one uses a different number of predictors [similar to Meko et al. (2001)]. The models start in 1827 (Model I, with seven chronologies), 1752 (II, with six), 1675 (III, with five), 1634 (IV, with four), and 1569 (V, with two), with all of them ending in 1965. Thus, starting with the set of longest chronologies in 1569, Model V produces Sub-reconstruction V, which ends in 1633. In 1634, Sub-reconstruction IV starts, provided by Model IV, and it ends in 1674. In that way, successively, the whole reconstruction is made.

The period of 1920–1965 was selected as the calibration period for all of the models. The remaining 20 years of overlapping data, 1895–1919, were used for model validation (verification period). The regression models for the reconstructions are based on the principal component analysis method. They are a particular case (one predictand) of orthogonal spatial regression (OSR); OSR and canonical regression are traditionally used in dendroclimatology (Cook et al. 1994). The mathematical details of these regression models are briefly described next.

Let $\overline{\mathbf{Y}}$ be the vector of the standardized (i.e., zero mean, unit standard deviation) predictand to be reconstructed and let $\overline{\overline{\mathbf{TR}}}$ be the tree ring matrix, where each column represents a standardized chronology. Let $\overline{\overline{\mathbf{C}}}$ be the correlation matrix of $\overline{\overline{\mathbf{TR}}}$. The eigenstructure analysis of $\overline{\overline{\mathbf{C}}}$ provides the matrix of principal component vectors of the predictor set, $\overline{\overline{\mathbf{E}}}$ (normalized eigenvectors). The components of $\overline{\overline{\mathbf{TR}}}$ in the orthonormal base formed by $\overline{\overline{\mathbf{E}}}$ are the amplitudes in each principal mode. The application of the OSR method leads to the following expression:

$$\overline{\mathbf{Y}}=\overline{\overline{\mathbf{U}}}\cdot\overline{\boldsymbol{\beta}}+\overline{\boldsymbol{\varepsilon}} \tag{1}$$

where $\overline{\boldsymbol{\varepsilon}}$=vector of errors; $\overline{\boldsymbol{\beta}}$=vector of regression coefficients; and $\overline{\overline{\mathbf{U}}}$=matrix of selected amplitudes. A subset of amplitudes is chosen to remove principal components not related to the global climatic signal, and to reduce the level of *artificial predictability* that the inclusion of too many predictors in the model may produce (Cook et al. 1994). The amplitudes are selected to explain most of the variance in the data set. The method applied to choose this subset of amplitudes, or potential predictors, starts with an initial selection by the PVP criterion [i.e., in decreasing order of eigenvalues, select those whose cumulative products of eigenvalues are larger than or equal to 1; Guiot (1985)]. Then, the amplitudes from smaller eigenvalues are sequentially disqualified for

Table 1. Reconstruction Statistics

Model	Period	Calibration		Verification			
		R_c^2	$\sigma_{a,c}^2$	R_v^2	RE	CE	$\sigma_{a,v}^2$
I	1827–1965	0.63	0.31	0.50	0.43	0.41	0.30
II	1752–1826	0.68	0.30	0.48	0.42	0.40	0.30
III	1675–1751	0.67	0.30	0.47	0.40	0.38	0.31
IV	1634–1674	0.54	0.44	0.22	0.10	0.10	0.62
V	1569–1633	0.26	0.57	0.20	0.22	0.20	0.62

the regression until the F-test criterion, with 5% significance, rejects the hypothesis of identical residual variance of the models.

Cook et al. (1999) found an improvement of this regression by prewhitening the tree ring chronologies and the PDSI data before doing the regression. This is due to the sometimes large differences in short-lag autocorrelation between climate and tree rings. The short-lag autocorrelation—red noise—of the PDSI is added a posteriori to the regression's result. In our particular case, no improvement was found by prewhitening the tree ring data; probably the filtering done in the model by the components selection was sufficient. On the other hand, significant correlation between the regression residuals and the PDSI in short-lags was found.

Hence, a low-order autoregressive model, AR(p), was added to the principal components regression, leading to the next global model

$$y_t=\sum_{i=1}^{q} u_{i,t}\cdot\beta_i+\sum_{i=1}^{p} y_{t-i}\cdot\phi_i+a_i \quad (2)$$

where q=number of selected components; ϕ_i=autoregressive parameters; and a_i=model residual, assumed to be independent and normally distributed, with zero mean, and variance of σ_a^2. For consistency, principal components and autoregressive parameters are finally jointly recalibrated for the calibration period, 1920–1965, using the sum of squares of the residuals as the objective function. For each set of chronologies, several model orders are proposed—with and without including the tree ring information of $t+1$ as predictors, combined with autoregressive models of Orders 1, 2, or 3. Starting from higher-order models, the F-test (5% significance) criterion is again implemented to find the reduction of order that is statistically significant, where lower order implies loss in predictability. For all models, for the different set of chronologies, the only significant predictors were tree rings with lag=0 and second order ($p=2$) for the autoregressive component. Results fit the assumptions of normality for the residual model.

To measure the performance of each model over the verification period (1895–1919), a number of statistics were computed (Table 1 shows numerical results and Appendix I contains the expressions). The average explained variance over the calibration period (R_c^2) is a direct measure of the least-squares goodness of fit of the model. The average squared Pearson correlation (R_v^2) is a measure of the covariance between the actual and estimated PDSI in the verification period. The average reduction of error (RE) has a theoretical range of $-\infty$ to $+1$, where RE>0 indicates that the reconstruction is better than the calibration period mean. The average coefficient of efficiency (CE) has the same theoretical range as RE, and a positive value indicates reconstruction skill in excess of the verification period mean. For the AR part, the variances of the residual noise for the calibration ($\sigma_{a,c}^2$) and verification ($\sigma_{a,v}^2$) periods were computed. They belong to the same population with a 10% significance, based on the F-test. Usually, this part of the reconstruction is rejected in dendrochronology reconstructions. The reason why it is included in this work is explained in the next section.

Extraction of Drought Characteristics

As noted before, droughts are distinguished mainly by their durations and severities. However, from the usual lengths of historical records, it is only possible to separately estimate the distribution of both characteristics. For example, in a historical record of 100 years, about 20 drought events may be found (depending on the definition of droughts). This number of events is insufficient to properly characterize the bivariate distribution of droughts in terms of durations and severities. To get longer records, synthetic series may be generated based on the stochastic properties of the historical data (Shiau and Shen 2001). In our work, semisynthetic series are generated from the information of the tree ring reconstructions. Each subreconstruction model [Eq. (2)] has a deterministic part and a stochastic part. The deterministic part comes from the principal components regression (y_t^{TR}), assuming zero error. The stochastic part comes from the autoregressive model (x_t) and provides equally likely realizations

$$y_t=y_t^{TR}+x_t \quad (3)$$

with

$$y_t^{TR}=\sum_{i=1}^{q} u_{i,t}\cdot\beta_i+\sum_{i=1}^{p} y_{t-i}^{TR}\cdot\phi_i \quad (4)$$

$$x_t=\sum_{i=1}^{p} x_{t-i}\cdot\phi_i+a_t \quad (5)$$

The sum of the deterministic and stochastic terms is assumed to provide possible realizations of the droughts that could happen, taking into account what was recorded on the tree rings and the drought recurrence observed in the historical data. It is assumed that the stochastic part transforms the uncertainty of the regression into what would be expected based on the red noise structure of historical data. To verify this, additional validation tests of the stochastic model were performed—for the common period, 1895–1965, in which summer PDSI is known as well as tree ring chronologies; the drought stochastic properties in the historical record are compared with those of the realizations coming from the reconstruction, for each of the five models. Statistically significant differences between historical and reconstructed drought stochastic properties are checked (see the details in Appendix II).

If the stochastic nature of the droughts were not fully present in the historical data, the realizations, with their deterministic part, would manifest a more complete signal of the droughts' stochastic nature (mainly for low frequencies). Moreover, the sum of both terms is assumed to have a closer approximation to the unknown variability of the real process, in magnitude and structure. Therefore, measures such as mean drought severity will be larger than those obtained using ordinary reconstructions (i.e., the deterministic part), and closer to the real ones. Thus, adding the stochastic part allows for comparisons between different subreconstructions, and between reconstruction and historical data, since their variabilities are similar. The difference with generating synthetic series is that, based on the stochastic properties of known data, instead of rewriting the drought history on a straight line, it is rewritten on the remaining traces recorded in the tree rings (deterministic part—continuous line in Fig. 1). From this

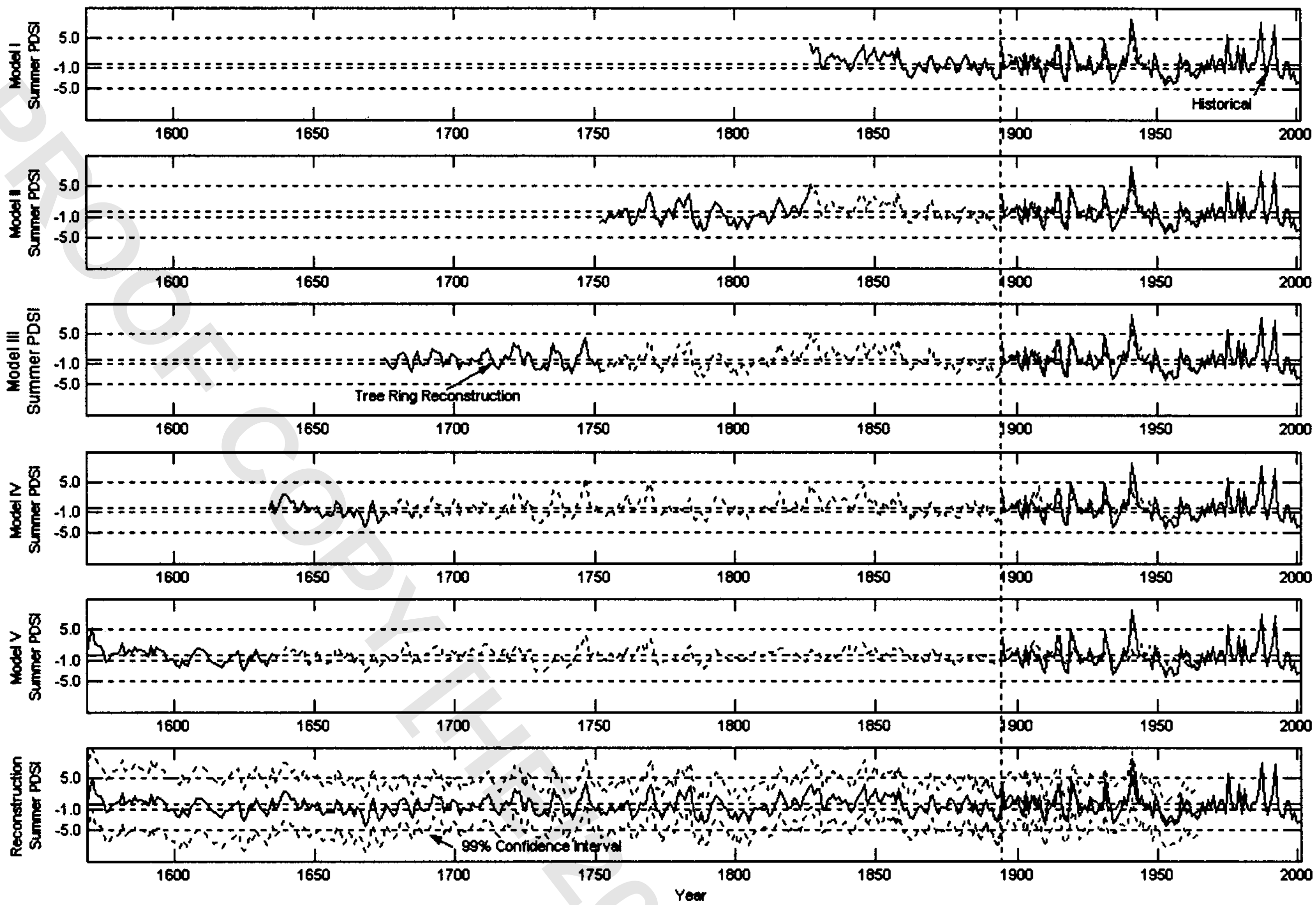

Fig. 1. Tree ring reconstructions of droughts for each model and final reconstruction with 99% confidence interval

point of view, even though Model V has a low R_c^2, its reconstruction results reveal the possible occurrence of droughts in the early 1600s that have the same amplitudes, in that subreconstruction, as the 1950s drought. There is evidence of a megadrought in the 1600s in other paleo records in the United States. Therefore, the model will reflect a higher probability of drought in those years, due to the trend observed in the tree rings regression. As the R_c^2 increases, the range of likely reconstructed droughts during a fixed period decreases (width of confidence interval in Fig. 1). If, instead, white noise were added, the reconstruction would be losing such an important characteristic in droughts as persistence.

A Monte Carlo technique was applied to evaluate the characteristics of the droughts that could happen in the reconstruction. A set of 1,000 realizations was used to represent this information. An important difference in dealing with semisynthetic series is that even when an infinite number of possible realizations are generated, the uncertainty associated with the inferred statistics must be calculated as coming from a sample size equal to the mean number of events that occur in the reconstructed period.

Each reconstruction is linked with historical records to extract its drought characteristics using runs analysis. To properly connect reconstructions and historical records, the known future (beginning of historical record) value should be taken into account in the last years of the reconstruction. In a stochastic process, if a future value is known, the distribution of the actual value of the variable is conditioned by this expectation. A procedure was used that considers the dependence of a stochastic variable on future values (Appendix III).

Bivariate Characterization of Drought Recurrence

Based on formulations of the drought occurrences as an alternating renewal process, expressions for the recurrence interval and the risk of occurrence have been developed for droughts with different durations (Loaiciga and Leipnik 1996; Fernández and Salas 1999), and different severities (Shiau and Shen 2001). In the present work, the alternating renewal process is also applied—in this case, to develop a theoretical expression of the recurrence interval for the general case of droughts described in terms of durations and severities. In addition, an algorithm to obtain the distribution of the recurrence interval to be used in risk analysis is presented. Even though in the formulation time is a continuous process, discrete time can be represented by substituting integrals for sums, as is suitably indicated. The summer PDSI is the variable used as an indicator of a moisture deficit situation. The PDSI may represent how significant a drought is, but its spatial comparability has been criticized (Guttman et al. 1992) and it does not have the associated probability of occurrence. Here, the theory of runs (Yevjevich 1967) is used to describe droughts based on the Palmer index.

As Fig. 2 illustrates, a drought is defined as the event during which the PDSI is below or equal to a fixed truncation level. A drought is characterized by two principal magnitudes—drought duration and drought severity. Drought duration (D) is defined as the period of time during which the PDSI is below or equal to the truncation level, and drought severity is the cumulative difference between the truncation level and the PDSI values during the

PROOF COPY [HE/2002/022550] 005305QHE

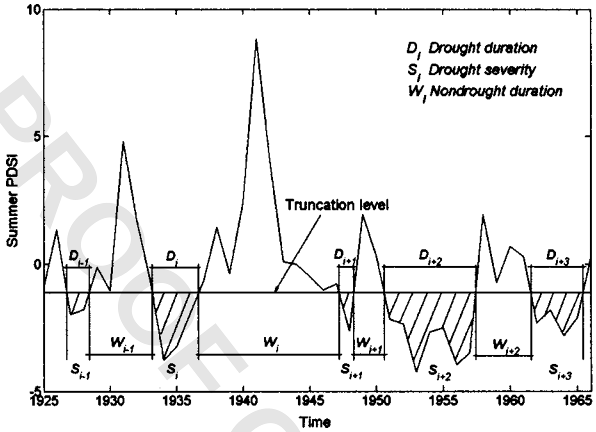

Fig. 2. Illustration of runs analysis—Definition of drought duration, severity, and nondrought duration for fixed truncation level

drought. Both magnitudes are random, and are usually significantly correlated (Yevjevich 1967). Together they describe the principal characteristics of a drought. Therefore, it is appropriate to describe their joint probability-density function (PDF), which, for convenience, will be expressed as the product of the marginal distribution of drought durations and the conditional distribution of drought severities for a given drought duration

$$f_{D,S}(d,s)=f_D(d)\cdot f_{S|D=d}(s) \tag{6}$$

To finally describe the sequence of droughts, the interarrival time is defined as the elapsed time (L) between the beginnings of two consecutive droughts. The nondrought period is the period of time between the end and beginning of two consecutive droughts. Here, the magnitude of interest is its duration (W), and its PDF $f_W(w)$. Hence

$$L=D+W \tag{7}$$

The joint probability cumulative function for the bivariate distribution of droughts is defined as $F_{D,S}(d,s)=P(D<d\cup S<s)$. In addition, the joint probability failure function is $\overline{F}_{D,S}(d,s)=P(D\geqslant d\cup S\geqslant s)$, which in other fields is called the survivor function.

Define $\hat{D}(d,s)$ as the set of droughts ($\hat{d}$) with durations larger than or equal to the critical duration d and severities larger than or equal to the critical severity s. Between two successive droughts $\hat{d}\in\hat{D}(d,s)$ will occur $N_{d,s}$ droughts ($N_{d,s}\geqslant 1$). The first of those droughts belongs to $\hat{D}(d,s)$, and $N_{d,s}-1$ are out of this set, $\hat{d}\notin\hat{D}(d,s)$. The time ($T_{d,s}$) between the beginning of those two successive droughts, $\hat{d}\in\hat{D}(d,s)$, can be expressed as the sum of the interarrival times of the droughts that happen in that period

$$T_{d,s}=L_{|D\geqslant d\cup S\geqslant s}+\sum_{i=1}^{N_{d,s}-1}L_{i|D<d\cap S<s} \tag{8}$$

where

$$L_{|D\geqslant d\cup S\geqslant s}=D_{|D\geqslant d\cup S\geqslant s}+W \tag{9}$$

$$L_{|D<d\cap S<s}=D_{|D<d\cap S<s}+W \tag{10}$$

Assuming that droughts are independent events, identically distributed, and that drought and nondrought durations are independent, the expected value of $T_{d,s}$ is

$$E(T_{d,s})=E(L_{|D\geqslant d\cup S\geqslant s})+[E(N_{d,s})-1]\cdot E(L_{|D<d\cap S<s}) \tag{11}$$

Substituting Eqs. (6) and (7) in Eq. (8)

$$E(T_{d,s})=E(N_{d,s})\cdot E(W)+E(D_{|D\geqslant d\cup S\geqslant s})+[E(N_{d,s})-1]\cdot E(D_{|D<d\cap S<s}) \tag{12}$$

From the independence assumption, the number of trials (droughts) at which the next drought $\hat{d}\in\hat{D}(d,s)$ occurs follows a geometric distribution, and the expected value is $1/P[\hat{d}\in\hat{D}(d,s)]$

$$E(N_{d,s})=1/P(D\geqslant d\cup S\geqslant s) \tag{13}$$

Given that $P(D<d\cap S<s)+P(D\geqslant d\cup S\geqslant s)=1$

$$E(N_{d,s})-1=P(D<d\cap S<s)/P(D\geqslant d\cap S\geqslant s) \tag{14}$$

$$E(D_{|D\geqslant d\cup S\geqslant s})=\frac{\int_{\ell=0[1]^*}^{\infty}\ell\cdot P(D=\ell\cup S\geqslant s)\cdot d\ell}{P(D\geqslant d\cup S\geqslant s)} \tag{15}$$

$$E(D_{|D<d\cap S<s})=\frac{\int_{\ell=0[1]^*}^{d[d-1]^*}\ell\cdot P(D=\ell)\cdot dt+\int_{\ell=d}^{\infty}\ell\cdot P(D=\ell\cup S<s)\cdot d\ell}{P(D<d\cap S<s)} \tag{16}$$

with [ ]* for the discrete case.

Substituting Eqs. (14), (15), and (16) into Eq. (12) results in

$$E(T_{d,s})=E(N_{d,s})\cdot E(W)+\frac{\int_{\ell=0[1]^*}^{\infty}\ell\cdot P(D=\ell)\cdot d\ell}{P(D\geqslant d\cup S\geqslant s)}=E(N_{d,s})\cdot[E(D)+E(W)] \tag{17}$$

$T_{d,s}$ will be in units of time. The expected recurrence interval of droughts $\hat{d}\in\hat{D}(d,s)$ is equal to the expected number of events between these droughts times the expected duration of the interarrival time between droughts. Thus, the expected interarrival time is not a function of the condition over the drought duration. For the particular case of the univariate drought description based on the severities, Shiau and Shen (2001) demonstrated that the final expression in Eq. (17) provided the return period. In our work, it is generalized for the bivariate analysis.

Other statistics may be calculated based on this model. For example, the expected number of years, $Y_{d,s}$, from a nondrought year until the beginning of the next drought $\hat{d}\in\hat{D}(d,s)$ is

$$E(Y_{d,s})=E(N_{d,s})\cdot E(W)+[E(N_{d,s})-1]\cdot E(D_{|D<d\cap S<s}) \tag{18}$$

For risk analysis, it is necessary to know the distribution and not only the expected value of the recurrence interval of droughts $\hat{d}\in\hat{D}(d,s)$. This distribution can be computed using the following algorithm.

Define the conditional probability density distribution (PDD) of the interarrival time as

$$f_{L|D\geqslant d\cup S\geqslant s}(L)=\int_{\ell=0[1]^*}^{L[L-1]^*}P(D=\ell|S\geqslant s\cup D\geqslant d)\cdot P(W=L-\ell)\cdot d\ell \tag{19}$$

$$f_{L|D<d\cap S<s}(L) = \int_{\ell=0[1]^*}^{L[L-1]^*} P(D=\ell|S<s\cap D<d)\cdot P(W=L-\ell)\cdot d\ell \tag{20}$$

For $N_{d,s}=1$, two consecutive droughts $đ \in Đ(d,s)$ occur, and the distribution of $T_{d,s}$ results in

$$f_{T_{d,s}|N_{d,s}=1}(T=L)=f_{L|D\geqslant d\cup S\leqslant s}(L) \tag{21}$$

$$f_{N_{d,s}}(1)=P(D\geqslant d\cup S\geqslant s) \tag{22}$$

When time is considered as a discrete variable, in order to limit the number of possible combinations, two computational variables are defined, as follows: the minimum T, equal to $T_{\min}(N_{d,s}=1)=d+1$ time units (at least one drought of $D=d$ and one nondrought occur); and the maximum probable T, equal to $T_{\max}(N_{d,s}=1)=D_{\max}+W_{\max}$, where $D_{\max}$ is the maximum probable drought duration and $W_{\max}$ is the maximum probable duration of the nondrought period. $D_{\max}$ and $W_{\max}$ can be estimated for a given tolerance level as

- $D_{\max}=d$ such that $[1-F_D(d)]<$tolerance and
- $W_{\max}=w$ such that $[1-F_W(w)]<$tolerance.

The computation of the distribution is completed successively considering the case of one additional drought event occurring, smaller than the critical, between the two droughts belonging to $\hat{D}(d,s):N_{d,s}=n$; with $n=2,3,\ldots$. For those cases

$$f_{T_{d,s}|N_{d,s}=n}(T) = \int_{\ell=0\{\max[T_{\min}(n-1),T-T_{\max}(1)]\}^*}^{T\{\min[T_{\max}(n-1),T-T_{\min}(1)]\}^*} f_{T_{d,s}|N_{d,s}=n-1}(\ell)\cdot f_{L|D<d\cap S<s} \times(T-\ell)\cdot d\ell \tag{23}$$

$$f_{N_{d,s}}(n)=f_{N_{d,s}}(n-1)\cdot P(D<d\cup S<s) \tag{24}$$

$$F_{N_{d,s}}(n)=F_{N_{d,s}}(n-1)+f_{N_{d,s}}(n) \tag{25}$$

$$f_{T_{d,s}|N_{d,s}\leqslant n}(T) = \frac{f_{T_{d,s}|N_{d,s}\leqslant n-1}(T)\cdot F_{N_{d,s}}(n-1)+f_{T_{d,s}|N_{d,s}=n}(T)\cdot f_{N_{d,s}}(n)}{F_{N_{d,s}}(n)} \tag{26}$$

where $T_{\min}(n)=(d+1)n$ and $T_{\max}(n)=(D_{\max}+W_{\max})n$; $n=2,3,\ldots,N_{\max}$, where $N_{\max}$ is the maximum probable N; and $N_{\max}=n$ such that $[1-F_{Nd,s}(n)]<$tolerance.

Using the same methodology, other additional probability distributions can be computed.

Application of Proposed Methodology

In the case study, the tree ring drought reconstruction allows us to expand the PDSI series from 107 years (1895–2001) to 433 years (1569–2001). This increase in the series length makes it possible to jointly characterize the droughts in terms of durations and severities. The runs analysis of the semisynthetic series from the Monte Carlo realizations provides the information necessary to analyze the stochastic properties of droughts, verify important assumptions, and fit the appropriate distribution functions. The truncation level chosen for the runs analysis is PDSI$=-1$, defined by Palmer (1965) as the beginning of a drought.

From the pool of drought durations (D), severities (S), and nondrought durations (W), the correlation between drought durations and severities is significant, with a value of $\rho_{D,S}=0.87$. Simultaneously, very low correlation values between D_t and D_{t+1} ($\rho=0.06$), S_t and S_{t+1} ($\rho=0.07$), W_t and W_{t+1} ($\rho=0.06$), W_t and D_t ($\rho=0.00$), and W_t and D_{t+1} ($\rho=0.03$) justify the hypotheses of the independence of drought durations, severities, and nondrought durations, and between drought durations and nondrought durations (i.e., they are not significantly larger than zero).

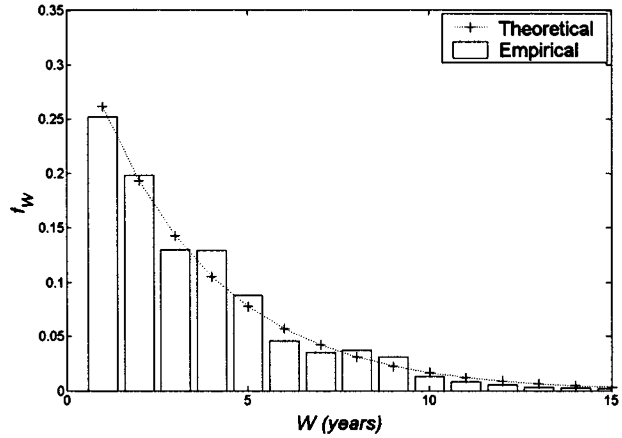

Fig. 3. Theoretical geometric distribution of nondrought durations and empirical distribution in Texas Climatic Division 5

The distributions of drought and nondrought durations can be derived theoretically by using the theory of runs (Yevjevich 1967; Sen 1976). The PDSI value in a year can be considered in two possible states—dry (PDSI$\leqslant-1$) or wet (PDSI>-1). If the PDSI state in year t depends on the previous-year state, a Markov-dependent process with two states and a homogeneous transition probability matrix (i.e., a simple Markov chain) can be applied to describe this fact (Fernández and Salas 1999). The transition probability matrix results in $P_{DD}=P(\text{PDSI}_{t+1}\leqslant-1|\text{PDSI}_t\leqslant-1)$, $P_{DW}=P(\text{PDSI}_{t+1}>-1|\text{PDSI}_t\leqslant-1)$, $P_{WD}=P(\text{PDSI}_{t+1}\leqslant-1|\text{PDSI}_t>-1)$, and $P_{WW}=P(\text{PDSI}_{t+1}>-1|\text{PDSI}_t>-1)$. It follows that $P_{DD}+P_{DW}=1$ and $P_{WW}+P_{WD}=1$. Therefore, the drought and nondrought duration distributions are

$$f_D(d)=P_{DW}\cdot P_{DD}^{d-1};\quad d=1,2,3,\ldots \tag{27}$$

$$f_W(w)=P_{WD}\cdot P_{WW}^{w-1};\quad w=1,2,3,\ldots \tag{28}$$

The expected values of D and W are $E(D)=1/P_{DW}$ and $E(W)=1/P_{WD}$. Thus, the expected value of the interarrival time in Eq. (15) is $E(L)=1/P_{DW}+1/P_{WD}$. Theoretical and empirical distributions of W are shown in Fig. 3. The application of the χ^2 test validated the selected model.

The drought duration distribution will be used in Eq. (20) to describe the joint PDD of drought durations and severities. No analytical solution of the conditional distribution of severities is known. Theoretical distributions are usually fitted from observed data. Gamma and exponential distributions are commonly employed (Zelenhastic and Salvai 1987; Mathier et al. 1992; Shiau and Shen 2001). In this case, the writers also found it appropriate to use the gamma distribution for the conditional distribution of severities

$$f_{S|D=d}(s)=\frac{s^{\alpha_d-1}\cdot e^{-s/\beta_d}}{\beta_d^{\alpha_d}\cdot\Gamma(\alpha_d)};\quad s>0 \tag{29}$$

PROOF COPY [HE/2002/022550] 005305QHE

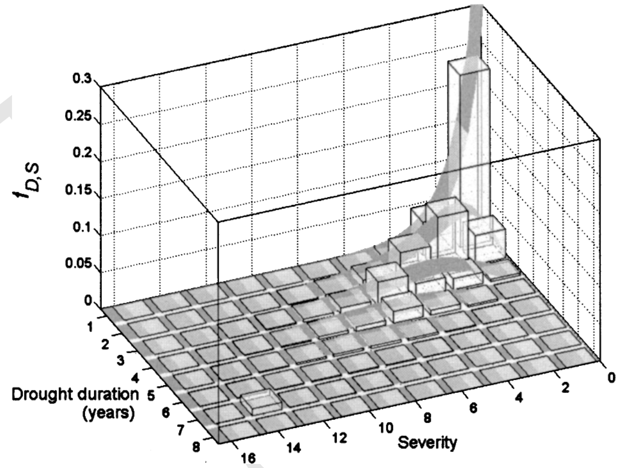

Fig. 4. Theoretical [Eq. (30)] and empirical bivariate distribution of drought durations and severities in Texas Climatic Division 5

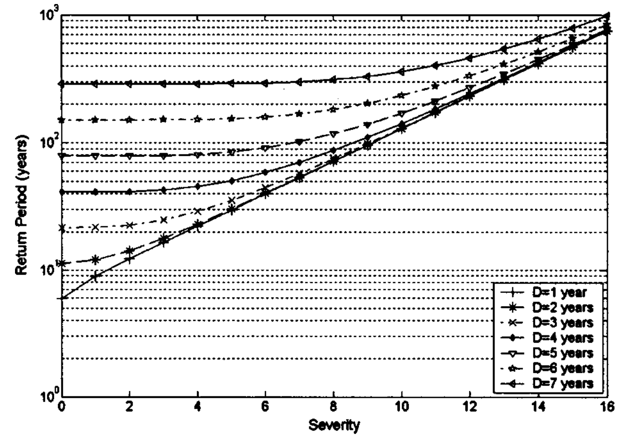

Fig. 5. Return period of droughts with bivariate analysis in Texas Climatic Division 5

From data exploration, the parameters of the gamma distributions (α_d, β_d) appear linearly dependent on drought durations for α_d, and constant for β_d [as also found by Shiau and Shen (2001)]. This is related to the high correlation between drought durations and severities. If it is written that $\alpha_d = \rho_\alpha \cdot d + \kappa_\alpha$ and assumed that $\beta_d = \beta$ (constant), then the mean severity for a given duration, based on the gamma distribution, is $E(S|D=d) = \alpha_d \cdot \beta_d = (\rho_\alpha \cdot d + \kappa_\alpha) \cdot \beta$, which is linearly dependent on *d*. Therefore, taking the proposed parameterization of gamma conditionals distributions, the joint PDD is

$$f_{D,S}(d,s) = \frac{s^{\alpha_d - 1} \cdot e^{-s/\beta}}{\beta^{\alpha_d} \cdot \Gamma(\alpha_d)} \cdot P_{DW} \cdot P_{DD}^{d-1}; \quad s>0 \quad \text{and} \quad d=1,2,... \tag{30}$$

with $\alpha_d = \rho_\alpha \cdot d + \kappa_\alpha$. This bivariate distribution has four parameters (ρ_α, κ_α, β, and P_{DW}), which are jointly calibrated using the maximum likelihood method. Fig. 4 compares empirical and theoretical distributions. A χ^2 goodness-of-fit test was performed, accepting the fitted distribution (Read and Cressie 1988).

The drought duration follows a geometric distribution, which is the discrete case of the exponential distribution. At the same time, the exponential distribution is a special case of the gamma distribution (fixing shape parameter equal to one). This suggests that a bivariate gamma distribution might be adequate to describe the bivariate distribution of drought durations and severities, treating duration as a continuous variable. In other hydrological events such as storms and floods, a bivariate gamma distribution is used to describe the joint probability distribution of two correlated random variables. A detailed discussion is found in Yue et al. (2001). However, in the case study the number of drought events resulting from the reconstruction and instrumental data is 71, and the number of parameters of a general bivariate gamma distribution with correlated variables is six. Such a distribution cannot be fitted with the available data without infringing upon the Pareto parsimony principle. For longer records, it is recommended to use this general function, mathematically more correct.

Hence, the probabilities used in the theoretical formulation of the recurrence model are

$$P(D \geq d \cup S \geq s) = \overline{F}_{D,S}(d,s) = \sum_{i=d}^{\infty} f_D(i) \cdot [1 - F_{S|D=i}(s)] \tag{31}$$

$$P(D<d \cap S<s) = 1 - \overline{F}_{D,S}(d,s) \tag{32}$$

$$P(W=w) = P_{wd} \cdot P_{ww}^{w-1} \tag{33}$$

$$P(D=\ell | D \geq d \cup S \geq s) = \frac{f_D(\ell) \cdot [1 - F_{S|D=\ell}(s)]}{F_{D,S}(d,s)} \quad \text{for} \quad \ell \geq d \tag{34}$$

$$P(D=\ell | D<d \cap S<s) = \frac{f_D(\ell)}{1 - F_{D,S}(d,s)} \quad \text{for} \quad \ell < d \tag{35}$$

$$P(D=\ell | D<d \cap S<s) = \frac{f_D(\ell) \cdot F_{S|D=\ell}(s)}{1 - F_{D,S}(d,s)} \quad \text{for} \quad \ell \geq d \tag{36}$$

Therefore, the return period of droughts of durations equal to or greater than *d* and severities equal to or larger than *s*, Eq. (18), results in

$$E(T_{d,s}) = \frac{1/P_{DW} + 1/P_{WD}}{F_{D,S}(d,s)}; \quad s>0 \quad \text{and} \quad d=1,2,... \tag{37}$$

Fig. 5 shows the return periods for the bivariate characterization of droughts. The return period of droughts with a duration of *d* or larger converges to the return period with longer duration conditions as severity increases. This is due to the high correlation between duration and severity (i.e., the probability of drought events with low durations and high severities is small). The correlation also explains the horizontal shape of the curves for long durations, when small severities are unlikely. The 1-year-duration curve corresponds to the return period of droughts characterized in terms of severities, since duration is not conditioned. Similarly, the values of each duration curve for zero severity represent the return period of droughts described in terms of duration. For perfect correlation, the transition of every duration-curve from horizontal to the severity marginal curve would be one point. As the correlation decreases, this transition is larger.

Traditionally, droughts have been described in univariate form, in terms of duration or severity, in most cases due to an insufficient number of recorded events. The high correlations between durations and severities justified this approach. Fig. 6 shows what could be the discrepancy between univariate analyses, in a case of high correlation, $\rho_{D,S} = 0.87$. In this figure, the conditional distribution of severity given the duration is represented by several quantiles. Parallel to the axes are presented the results of the

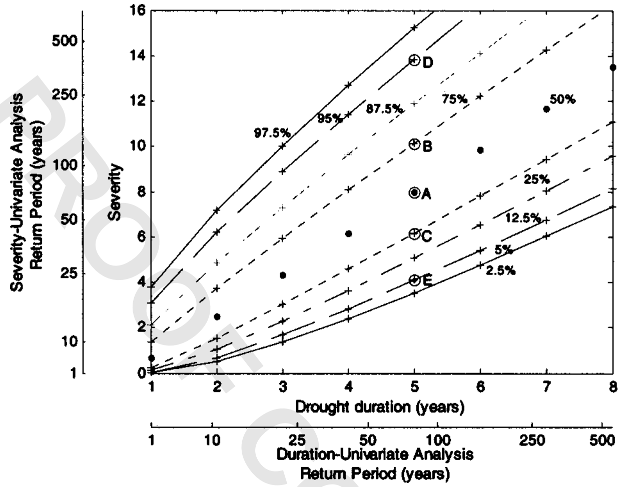

Fig. 6. Conditional distribution of severities given durations, represented by different quantiles, and univariate analyses of drought recurrence based on severity and duration—Texas Climatic Division 5

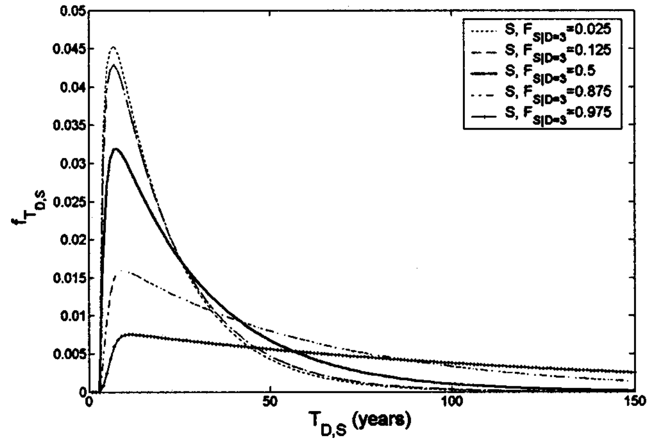

Fig. 8. Probability density function of interarrival time between droughts (bivariate analysis) for drought duration equal to three years and different quantiles of conditional distribution of severity—Texas Climatic Division 5

univariate analyses of droughts based on duration and severity, respectively. The amplitude of the conditional distribution functions produces significant discrepancies. For example, in the case of drought durations equal to 5 years, the return period based on duration yields $T_D = 80$ years. The median of the severities that occur with this duration (quantile 50%, Point A) has a return period based on a severity of $T_S = 80$ years. Therefore, for the drought represented by Point A, both analyses agree. However, if the quantiles of 25 and 75% are used (Points B and C), the return period based on severity may vary between 40 and 135 years. Higher discrepancies occur with the 5 and 95% quantiles (Points D and E), which correspond to 19 and 360 years, respectively. Both duration and severity are of interest for drought description. From this fact, the discrepancies between univariate analyses make the use of univariate characterization not approachable.

The discrepancies disappear when the bivariate analysis is employed. Fig. 7 shows the superposition of the contours of the return period, using the bivariate analysis, and the quantiles of the conditional distribution of severity given the duration. The

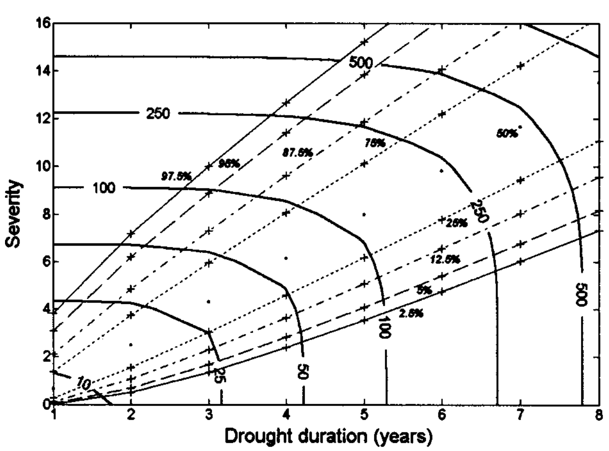

Fig. 7. Superposition of return period (years) of droughts based on bivariate analysis and conditional distribution of drought severities, given duration, represented by different quantiles—Texas Climatic Division 5

droughts are jointly characterized by the probabilities associated with duration and severity. The amplitude of the conditional distribution of severity shows the variability of the droughts that occur with a given duration. In this figure, the univariate analyses results are also presented; values of the return period contour on the axes. The same level of discrepancies may occur if instead of giving the duration, severity is selected. In this case, the amplitude of the marginal distribution is also too large to apply univariate analyses.

The bivariate description provides a more complete characterization of droughts—not only because it produces unique return periods, but also because duration and severity are the most important items for water resources management in drought analysis. A drought of severity S can cause the failure of the system, depending on its duration. For a 1-year-duration drought, probably no severity causes failure (this always depends on the system vulnerability). What makes a drought a hazard is its persistence. Instead, the severity S starts to be problematic when it takes place over a long period. On the other hand, when such severity is spread over a longer duration, the probability of failure may be again lower. Therefore, a bivariate analysis of droughts is also required for a more complete study of the risk of failure of a water resource system. Fig. 8 shows the distribution of the interarrival time between droughts of durations equal to 3 years or greater and severities equal to or greater than several quantiles of its conditional distribution. These distributions are of interest, to know the risk of failure of a system during the life of the project. As the quantile increases, the probability of occurrence in a given period decreases (area below the curve until the project life). In this case, univariate analysis would lead to a conservative result.

Results and Conclusions

In this paper, a methodology is proposed to evaluate the recurrence of droughts, described jointly in terms of durations and severities. The occurrence of droughts is described as an alternating renewal process, under the assumptions of equal distribution and independence of drought and nondrought events. The two most important characteristics of the recurrence of droughts are developed in this bivariate analysis—the return period and the risk of occurrence. Following the same methodology, other statis-

PROOF COPY [HE/2002/022550] 005305QHE

Table 2. Return Periods (Years) of Historical Droughts Based on Historical Data Besides Tree Ring Reconstruction (HTRR), Only Historical Data, and Raw Tree Ring Reconstruction (RTRR)

	HTRR 1569–2001			Historical 1895–2001		RTRR 1675–1965		
Drought	$T_{D,S}$	T_D	T_S	T_D	T_S	$T_{D,S}$	T_D	T_S
1951–1957	699	290	465	465	619	1,451	186	1,447
1962–1965	51	41	30	49	37	56	37	52
1934–1936	44	22	39	23	48	72	22	72
1916–1918	37	22	31	23	38	55	22	55
1909–1910	22	11	21	11	24	34	13	34
1994–1995	19	11	18	11	20	27	13	27

tical properties and distributions, which could be of interest for water resources management, may be calculated.

The possibility of developing a bivariate analysis of droughts is restricted to the availability of a sufficiently long record of drought events. Tree ring reconstructions of droughts provide a valuable source of information for the recurrence analysis. The reconstructions themselves indicate the likely trends of the drought index, but they cannot be compared directly with historical records. A procedure is presented to allow the inclusion of reconstruction information in the statistical analysis of droughts jointly with the historical records. The procedure takes into account the residuals of the reconstructions and their stochastic structure, based on the historical data. The residuals are modeled by an autoregressive model. This model is assumed to transform the deterministic part of the reconstruction in likely realizations of the drought indexes that could occur with those tree ring records. The deterministic and stochastic parts of the reconstructions were validated. A Monte Carlo technique was used to extract the stochastic component of the reconstructed droughts. The reconstructions are properly connected with the historical record, by the conditional expectation, and are combined for the statistical analysis.

Drought durations and severities both play an important role in drought management. Even when there is a significant positive correlation between durations and severities, a bivariate analysis is recommended because of the discrepancies that may arise from univariate analyses. Table 2 illustrates this fact in the case study, and the benefits of the presented approach. The drought analyses were implemented over three different data sets. The first data set was composed of the historical data and the tree ring reconstruction (HTRR), using the technique described in this paper, in the complete period 1569–2001 (Columns 2–4). Its results can be compared with univariate analyses over the period 1895–2001, using only the historical data (Columns 5 and 6). Comparing Columns 3 and 4 with 5 and 6, the inclusion of pass drought reconstruction (HTRR) increases the probability of rare events such as the 1950s drought. For drought events of lower magnitudes, the likelihood is slightly larger, with a tendency to be equal for frequent droughts, as expected. Again, there are significant differences in the return period computed from univariate analyses, where the droughts are not completely characterized. The bivariate analysis (Column 2) shows that the 1950s drought in Texas Climatic Division 5 is an extreme drought, with a return period of about 700 years. It must be noted that bivariate analyses always produce a larger return period than univariate analyses, since one additional condition is imposed. Finally, the return periods obtained from the raw tree ring reconstruction (i.e., the deterministic part) for the period with better correlation, 1675–1965, were computed using both bivariate and univariate analyses (Columns 7–9). The results illustrate the limitations if the stochastic part is not considered in the statistical analyses. The severity is always underestimated because the deterministic part of the reconstruction has lower variability than the original.

Future Research

The writers encourage the application of this methodology in other sites, and its comparison to and discussion with other studies. The most important step in the application of the bivariate analysis is the estimation of the bivariate distribution of durations and severities. The theoretical derivation of a function for this probability distribution is still not developed.

One of the main reasons why droughts are studied is because of their associated losses (socioeconomical, environmental, agricultural). Therefore, the inclusion of these losses in the probability analysis is recommended. In the actual bivariate analysis of droughts, the return period definition considers the probability of droughts with durations and severities larger than or equal to a critical duration d and severity s. But smaller durations d may produce similar damage with a higher severity than s, and those cases are not included. Thus, the return period definition should be based on equal-damage curves.

Finally, the singular spectrum analysis of the results shows an important periodicity with a period close to the sunspots period. Taking this into account and the advantages of longer records, the possibility of a drought index forecast must be studied.

Acknowledgments

Professors K. Lansey, B. Nijssen, and S. Sorooshian provided valuable discussions and comments that improved the content of this paper. The comments of three anonymous reviewers also improved the paper. This material is based upon work supported in part by the Center for Sustainability of Semi-Arid Hydrology and Riparian Areas under the STC Program of the National Science Foundation, Agreement Number EAR-9876800. All contributions are gratefully acknowledged.

Appendix I. Statistics Used for Reconstruction Validation

The average explained variance (R_c^2) over the calibration period is

$$R_c^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - \overline{y}_c)^2} \tag{38}$$

The average squared Pearson correlation (R_v^2) over the verification period is

$$R_v^2 = \frac{\lfloor \Sigma(y_i - \overline{y}_v)(\hat{y}_i - \overline{\hat{y}}_v)\rfloor}{\Sigma(y_i - \overline{y}_v)^2 \Sigma(\hat{y}_i - \overline{\hat{y}}_v)^2} \tag{39}$$

The average RE in the verification period is

$$\mathrm{RE} = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - \overline{y}_c)^2} \tag{40}$$

The average CE over the verification period is

$$\mathrm{CE}=1-\frac{\Sigma(y_i-\hat{y}_i)^2}{\Sigma(y_i-\overline{y_v})^2} \tag{41}$$

where y_i and $\hat{y}_i$=actual and estimated data in year i; $\overline{y_c}$=mean of the actual data over the calibration period; and $\overline{y_v}$ and $\hat{y}_v$=means of the actual and reconstructed data for the verification period (Cook et al. 1999).

Appendix II. Drought Stochastic Structure Test

The stochastic nature of the droughts in the historical record is represented by the parameters of the distribution functions of drought durations, drought severities, and nondrought durations. Such parameters are sample estimators of the population parameters, and as such they are assumed normally distributed. Their means and variances are estimated using the criteria given by Bury (1999). Similar studies may be followed for each possible realization resulting from the reconstruction with each model.

Every parameter distribution coming from a realization is compared with the distribution of the corresponding parameter in the historical record. An F-test is used for those comparisons, where the variance of the joint distribution is compared with the variance of the populations. The test is made for a significant level of α (5%) for successive realizations coming from Monte Carlo runs. Based on the assumption that in each run the probability of test acceptance is $p=1-\alpha$ (assuming equally distributed parameters for the historical and modeled population), the number of accepted runs after N runs follows a binomial distribution. For $N=100$, the critical number of accepted runs (minimum) with a significance of α is 91. The test accepts a model when all parameters have failed fewer than nine runs. In our reconstructions, all models were accepted. The most sensible parameter was the shape parameter of the gamma distribution. For this parameter, Model V produced 92 accepted runs and Models IV, III, II, and I produced 95, 99, 98, and 99, respectively.

Appendix III. Conditional Expectation

A stochastic system (x_t), following an autoregressive model, AR(p), can be expressed in terms of its Green's function terms (G_i) [also referred to as ψ weights (Box and Jenkins 1976)] in the form

$$x_t=a_t+\sum_{i=1}^{\infty} G_i\cdot a_{t-i} \tag{42}$$

When the model is stable, Eq. (42) can be approximated for a finite sum of $N+1$ terms. In year $t_0=1895$, y_{t_0} and $y_{t_0}^{TR}$ are known, and therefore x_{t_0}. Before t_0, x_t is no longer known, but the conditional expectation modifies the distribution of its values. The problem can be formulated as a condition for the values of a_t in the N-later years and a_{t_0}

$$x_{t_0}=a_{t_0}+\sum_{i=1}^{N} G_i\cdot a_{t_0-i} \tag{43}$$

This is an equation with $N+1$ degrees of freedom. The statistical distribution of the unknowns must be taken into account to give the probable solution. To solve it, let us formulate a general problem in the following form:

$$Z_k=B_k+\alpha_k\cdot b_k \tag{44}$$

where Z_k and α_k are known; $B_k\in N(0,\sigma_{B_k}^2)$; and $b_k =\in N(0,\sigma_{b_k}^2)$. B_k and b_k are independents. The joint distribution of (b_k,B_k) is a bivariate normal distribution. Since the variables are independents, the joint PDF is the product of their univariate distributions. The condition defined by Eq. (44) has a probability in this bivariate distribution, and the values of both variables that obey Eq. (44) are normally distributed. Using b_k as an independent variable to express the pairs (b_k,B_k) that satisfy Eq. (44), the conditional distribution of b_k has the following moments:

$$\text{Mean,}\quad \mu_{b_k|(44)}=\alpha_k\cdot Z_k\frac{\sigma_{b_k}^2}{\sigma_{B_k}^2+\alpha_k^2\cdot\sigma_{b_k}^2} \tag{45}$$

$$\text{Variance,}\quad \sigma_{b_k|(44)}^2=\frac{\sigma_{B_k}^2\cdot\sigma_{b_k}^2}{\sigma_{B_k}^2+\alpha_k^2\cdot\sigma_{b_k}^2} \tag{46}$$

Generating a random value of b_k for this conditional distribution, the difference in variance of both variables, (b_k,B_k), is considered.

Using this procedure and the properties of linear combination of independent normal distributions, Eq. (43) can be solved starting from $k=N$, $N-1$, $N-2$,...,1 and

$$Z_k=Z_{k+1}-b_{k+1},\quad \text{for } k=1,2,\ldots,N-1;$$

and

$$Z_k=x_{t_0},\quad \text{for } k=N \tag{47}$$

$$B_k=a_{t_0}+\sum_{i=1}^{k-1} G_i\cdot a_{t_0-i}\in N(0,\sigma_{B_k}^2),$$

with

$$\sigma_{B_k}^2=\left(1+\sum_{i=1}^{k-1} G_i^2\right)\cdot\sigma_a^2 \tag{48}$$

$$b_k=a_{t_0-k}\in N(0,\sigma_a^2) \tag{49}$$

$$\alpha_k=G_k \tag{50}$$

for $k=1,2,\ldots,N$.

Notation

The following symbols are used in this paper:

- a = independent and normally distributed variable;
- B, $\underline{b}$ = normally distributed variables;
- $\mathbf{C}$ = correlation matrix of tree rings chronologies;
- CE = average coefficient of efficiency;
- D, d = drought duration;
- $\hat{D}$ = droughts set;
- $D_{\max}$ = maximum probable drought duration;
- d = drought event;
- $E(\,)$ = expectation operator;
- $\mathbf{E}$ = amplitudes matrix;
- $\underline{F}(\,)$ = probability cumulative function;
- $F(\,)$ = probability failure function;
- $f(\,)$ = probability density function;
- G = Green's function;
- i, k = indexes;
- L = drought interarrival time;
- ℓ = integrand variable;

PROOF COPY [HE/2002/022550] 005305QHE

$N_{d,s}$ = number of drought events occurring between two drought events with duration equal to or greater than d and severity equal to or greater than s;
n = number of droughts;
$P(\ )$ = probability operator;
P_{DD} = $P(\mathrm{PDSI}_{t+1} \leqslant -1 | \mathrm{PDSI}_t \leqslant -1)$;
P_{DW} = $P(\mathrm{PDSI}_{t+1} > -1 | \mathrm{PDSI}_t \leqslant -1)$;
P_{WD} = $P(\mathrm{PDSI}_{t+1} \leqslant -1 | \mathrm{PDSI}_t > -1)$;
P_{WW} = $P(\mathrm{PDSI}_{t+1} > -1 | \mathrm{PDSI}_t > -1)$;
p = order of autoregressive model;
q = order of moving average model;
R_c^2 = average explained variance;
R_v^2 = average squared Pearson correlation;
RE = average reduction of error;
S, s = drought severity;
$T_{d,s}$ = interarrival time between two drought events with duration equal to or greater than d and severity equal to or greater than s;
$T_{\max}$ = maximum probable interarrival time;
$T_{\min}$ = minimum interarrival time;
$\overline{\overline{\overline{\mathbf{TR}}}}$ = tree rings chronologies matrix;
t = time index;
u = amplitude of tree rings chronologies in principal components;
W, w = nondrought duration;
$W_{\max}$ = maximum probable nondrought duration;
x = stochastic part of tree ring reconstruction;
$Y_{d,s}$ = time from nondrought year until beginning of first drought with duration equal to or greater than d and severity equal to or greater than s;
$\overline{\mathbf{Y}}$ = vector of standardized predictand;
y = tree ring reconstruction;
y^{TR} = deterministic part of tree ring reconstruction;
Z = constant;
α = constant;
α_d = shape parameter of gamma distribution;
β, β_d = scale parameter of gamma distribution;
$\underline{\boldsymbol{\beta}}$ = vector of regression coefficients;
$\underline{\boldsymbol{\varepsilon}}$ = vector of errors;
κ_α = constant of α_d linear regression;
μ = mean population;
ρ_α = regression coefficient of α_d linear regression;
σ^2 = variance; and
ϕ = autoregressive parameter.

References

Alley, W. M. (1984). ''The Palmer drought severity index: Limitations and assumptions.'' *J. Clim. Appl. Meteorol.*, 23, 1100–1109.

Blasing, T. J., and Duvick, D. N. (1984). ''Reconstruction of precipitation history in North American corn belt using tree rings.'' *Nature (London)*, 307, 143–145.

Box, G. E. P., and Jenkins, G. M. (1976). *Time series analysis forecasting and control*, 2nd Ed., Holden-Day, Oakland, Calif.

Brockway, C. G., and Bradley, A. A. (1995). ''Errors in streamflow drought statistics reconstructed from tree-ring data.'' *Water Resour. Res.*, 31(9), 2279–2293.

Bury, K. (1999). *Statistical distributions in engineering*, Cambridge University Press, London.

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

Cook, E. R., Briffa, K. R., and Jones, P. D. (1994). ''Spatial regression methods in dendroclimatology: A review and comparison of two techniques.'' *International Journal of Climatology*, 14, 379–402.

Cook, E. R., and Kairiukstis, L., eds. (1990). *Methods of dendrochronology—Applications in the environmental sciences*, Kluwer Academic, Boston.

Cook, E. R., Meko, D. M., Stahle, D. W., and Cleaveland, M. K. (1996). ''Tree-rings reconstruction of past drought across the coterminous United States: Tests of regression method and calibration/verification results.'' *Tree rings, environments, and humanity*, J. S. Dean, D. M. Meko, and T. W. Swetnam, eds., Radiocarbon, 155–169.

Cook, E. R., Meko, D. M., Stahle, D. W., and Cleaveland, M. K. (1999). ''Drought reconstructions for the continental United States.'' *J. Clim.*, 12(7), 1145–1162.

Cook, E. R., Meko, D. M., and Stockton, C. W. (1997). ''A new assessment of possible solar and lunar forcing of the Bidecadal drought rhythm in the western United States.'' *J. Clim.*, 10, 1343–1356.

Cook, E. R., Stahle, D. W., and Cleaveland, M. K. (1992). ''Dendroclimatic evidence from eastern North America.'' *Climate since A.D. 1500*, R. S. Bradley and P. D. Jones, eds., Routledge, Chapman & Hall, New York, 331–348.

Douglas, E. M., Vogel, R. M., and Kroll, C. N. (2002). ''Impact of streamflow persistence on hydrologic design.'' *J. Hydrologic Eng.*, 7(3), 220–227.

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

Dracup, J. A., Lee, K. S., and Paulson, E. G. (1980b). ''On the statistical characteristics of drought events.'' *Water Resour. Bull.*, 26(5), 823–829.

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

Frick, D. M., Bode, D., and Salas, J. D. (1990). ''Effect of drought on urban water supplies. I: Drought analysis.'' *J. Hydraul. Eng.*, 116(6), 733–753.

Fritts, H. C. (1976). *Tree rings and climate*, Academic, San Diego.

Fritts, H. C. (1991). *Reconstructing large-scale climate patterns from tree-ring data*, The University of Arizona Press, Tucson, Ariz.

Guiot, J. (1985). ''The extrapolation of recent climatological series with spectral canonical regression.'' *Journal of Climatology*, 5, 325–335.

Guttman, N. B. (1991). ''A sensitivity analysis of the Palmer hydrologic drought index.'' *Journal of the American Water Resources Association*, 27(5), 797–807.

Guttman, N. B., Wallis, J. R., and Hosking, J. R. M. (1992). ''Spatial comparability of the Palmer drought severity index.'' *Journal of the American Water Resources Association*, 28(6), 1111–1119.

Heddinghaus, T. R., and Sabol, P. (1991). ''A review of the Palmer drought severity index and where do we go from here?'' *Proc., 7th Conf. on Applied Climatology*, American Meteorological Society, Boston, 242–246.

Huschke, R. E. (1959). *Glossary of meteorology*, American Meteorological Society, Boston.

IGBP PAGES/World Data Center for Paleoclimatology. (2002). ''International tree ring data bank.'' ⟨http://www.ngdc.noaa.gov/paleo/treering.html⟩, National Oceanic and Atmospheric Administration/NGDC Paleoclimatology Program, Boulder, Colo.

Karl, T. R. (1983). ''Some spatial characteristics of drought duration in the United States.'' *J. Clim. Appl. Meteorol.*, 22, 1356–1366.

Karl, T. R. (1986). ''The sensitivity of the Palmer drought severity index and Palmer's A-index to their calibration coefficients including potential evapotranspiration.'' *J. Clim. Appl. Meteorol.*, 25, 77–86.

Kim, T., Valdés, J. B., and Yoo, C. (2003). ''A nonparametric approach for estimating return periods of droughts in arid regions.'' *J. Hydrologic Eng.*, in press.

Loaiciga, H. A., and Leipnik, R. B. (1996). ''Stochastic renewal model of low-flow streamflow sequences.'' *Stochastic Hydrol. Hydr.*, 10(1), 65–85.

Mathier, L., Perreault, L., Bobée, 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.

Meko, D. M., Cook, E. R., Stahle, D. W., Stockton, C. W., and Hughes,

M. K. (1993). ''Spatial patterns of tree-growth anomalies in the United States and southeastern Canada.'' *J. Clim.*, 6, 1773–1796.

Meko, D. M., Therrell, M. D., Baisan, C. H., and Hughes, M. K. (2001). ''Sacramento River flow reconstructed to A.D. 869 from tree rings.'' *Journal of the American Water Resources Association*, 37(4), 1029–1039.

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

Read, T. R. C., and Cressie, N. A. C. (1988). *Goodness-of-fit statistics for discrete multivariate data*, Springer, New York.

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

Sen, Z. (1980). ''Statistical analysis of hydrologic critical droughts.'' *J. Hydraul. Div., Am. Soc. Civ. Eng.*, 106(1), 99–115.

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.

Stockton, C. W., and Meko, D. M. (1975). ''A long-term history of drought occurrence in western United States inferred from tree rings.'' *Weatherwise*, 28, 244–249.

Texas Center for Policy Studies. (2000). ''Growth at any cost? Reconciling economic development policy and water conservation in the Rio Grande Basin.'' *Rep. of Panel Proc., 75*, ⟨http://www.texascenter.org/⟩.

Yevjevich, V. M. (1967). ''An objective approach to definitions and investigations of continental hydrologic droughts.'' *Hydrologic Paper 23*, Colorado State Univ., Fort Collins, Colo.

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

Yue, S., Ouarda, T. B. M. J., and Bobée, B. (2001). ''A review of bivariate gamma distributions for hydrological application.'' *J. Hydrol.*, 246, 1–18.

Zelenhastic, E., and Salvai, A. (1987). ''A method of stream drought analysis.'' *Water Resour. Res.*, 23(1), 156–168.

PROOF COPY [HE/2002/022550] 005305QHE