**Natural Hazards
and Earth
System Sciences**

The mean frequency of recurrence of in-time-multidimensional events for drought analyses

J. González[1] and J. B. Valdés[2]

[1]Department of Civil Engineering, The University of Castilla - La Mancha, Avda. Camilo José Cela s/n, 13071 Ciudad Real, Spain
[2]Department of Civil Engineering and Engineering Mechanics, and Center for Sustainability of Semi-Arid Hydrology and Riparian Areas (SAHRA), The University of Arizona, Tucson, AZ 85721, USA

Abstract. Droughts are related with prolonged periods when moisture is significantly below normal values. Drought indices attempt to scale the main drought features to facilitate comparisons. Numerous indices are found in the literature based on different drought features. Many of them were created for particular places and specific objectives, and therefore not suitable to generalize the results. However, there have been attempts to develop a general index, which would provide full characterization of drought events. Two of the most well known are the Palmer Drought Severity Index (PDSI) and the Standard Precipitation Index (SPI). Each one has particular advantages and disadvantages. Still neither of them or any other includes a full representation of droughts in a single value index, being useful for all general application. The fact that droughts have a random nature prescribes the statistical theory for the foundation of a complete and generic index, which would meet this goal.

In this work, a procedure that allows a complete statistical characterization of drought events is presented. Droughts are characterized, from a statistical point of view, based both on the deviation from a normal regime and persistence. The events are represented as multivariate ones, whose dimensionality depends on the duration. Equal duration events are discriminated through their deviations from normality. The mean frequency of recurrence (MFR) is theoretically derived for such multivariate events, and it is used to scale such deviations. Therefore, events with different dimensionalities can be compared on a common dimension of interest, the MFR. This may be used as a drought index for drought characterization, both for analyzing historical events and monitoring current conditions. It may also be applied to analyze precipitation, streamflows and other hydroclimatic records. Its statistical nature and its general conception support its universality. Results may be applied not only to drought analysis, but also to analyze other random natural hazards. Applications of the procedure for drought analysis in Texas (USA) and in Gibraltar (Iberian Peninsula) are made and compared

with PDSI and SPI results. The MFR applied over drought analysis allows the representation of the main drought characteristics in a single value, based on the statistical feature of the phenomenon, and scaled on the mean frequency of recurrence.

1 Introduction

Droughts are related with extremely long periods of abnormally low water availability. Such low levels are coming from reduced precipitation periods, and they can be observed over different hydroclimatic variables related with water availability. Examples are precipitation (meteorological drought), soil moisture (agricultural drought), or streamflows (hydrological drought).

Whichever variable is analyzed to study droughts, the common feature looked for on the records is the downwards deviation from normality. If persistence is not considered, droughts may be represented as univariate extreme events, like other extreme hydrological events such as floods. However, for droughts is not only an issue the maximum deviation below a fixed threshold, but also the amount of time that remains below that threshold value is of high relevance. Runs theory has been applied to analyze droughts, defined as periods with water availability below a threshold. Therefore, run analysis may be used to assign probability to drought events based on one characteristic (e.g. cumulative deficit, duration, or mean intensity; Sen, 1976, 1977, 1980; Dracup et al., 1980a, b; Fernandez and Salas, 1999; Chung and Salas, 2000; Shiau and Shen, 2001), or two of them simultaneously (e.g. duration and severity; Gonzalez and Valdes, 2003). In this kind of statistical characterization, usually more that one drought variable is of relevance (i.e. duration, severity, intensity) to assign a probability, but only extraordinary long records provide data for a multivariate analyses. Another aspect of this kind of approach is how low the threshold should be. This influences the results, and it depends on the expected behavior of the variable under analysis.

Correspondence to: J. González (javier.gonzalez@uclm.es)
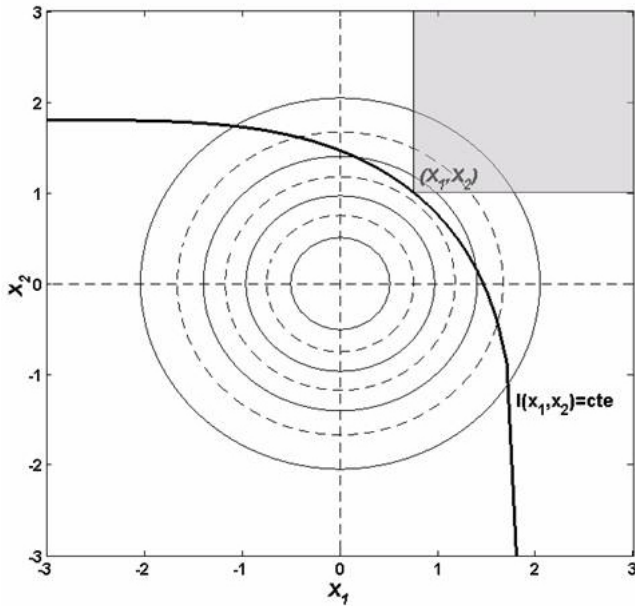
Fig. 1. Contour lines of standardized binormal density function for independent variables (x_1, x_2) and ℓ function for a point (X_1, X_2) defined as the probability of events (x_1, x_2) with $(x_1 \geq X_1 \cap x_2 \geq X_2)$

Since droughts are events that not only are characterized because they occur during prolonged periods, but also they affect large areas, spatial analyses are of high relevance. Dalezios et al. (2000) and Loukas et al. (2002) used several statistical techniques to analyze the temporal and space behavior of regional droughts along Greece. Kim et al. (2002) presented a methodology for estimating the return period of droughts as a function of intensity, duration and areal extent. The role of natural ecosystems in the soil-moisture dynamics are been studied in recent years to understand the behavior of plants in water-controlled ecosystems (Rodriguez-Iturbe et al., 2001; Porporato et al., 2001; Laio et al., 2001).

Water Resources Management (WRM) allocates water resources in time and space in order to meet demands. WRM requires not only the statistical characterization of historical drought events, but also the monitoring of current drought conditions based on real time observations or on derived indices. Numerous indices have been developed to be applied under particular objectives (for a complete summary see Hein, 2000). Keyantash and Dracup (2002) evaluate several of the most commonly used indices. The majority of them may be applied only in the particular location where they were created. One of the early generic drought indexes is the Palmer Drought Severity Index (PDSI; Palmer, 1965). This is an empirical dimensionless meteorological index, based on precipitation and temperature, and a conceptual representation of soil-water dynamics. Another well know index is the Standard Precipitation Index (SPI; McKee et al., 1993). This uses the statistical distribution of the precipitation record for a given time-aggregation length to compute the index. The value of the index is related with the probability of precipitation record below the observed.

A good example of the need and level of effort to obtain a quantitative evaluation of drought situations is the Drought Monitoring Program (NOAA/National Climatic Data Center), with the National Drought Monitoring Center in the US. In Europe, studies like ARIDE and the requirement of drought index development for WRM in the European Water Framework Directive (200/60/CE) are additional examples. None of the developed indices is fully satisfactory for general applications. All of them are focused in one or a few drought aspects, but disregard many others. This is a consequence of the multiple nature of the droughts (i.e. meteorological, hydrological, agronomical, socio-economic). Even when all indices need to be adapted to the particular application, it does not imply that a general methodology cannot be developed.

In this paper a new general methodology to compute the Mean Frequency of Recurrence (MFR) of extreme events taken as multivariate in time is presented. An event is given by the succession of w consecutive extreme values, and thus the event-dimension is equal to its duration. An expression of the MFR, under the assumption of temporal independence, are analytically developed and approximately estimated. The MFR, or its inverse the Return Period, provides a useful tool to analyze and monitory droughts. The drought index derived from these results may be used to characterize either historical events or current drought condition as observed by a hydroclimatic variable (e.g. precipitation, soil moisture, flow rate, ground water level) according to the associated probability or MFR. The methodology uses the succession of values of the variable, and therefore takes implicitly into consideration all possible aspects that should be considered (i.e. duration, severity, intensity, minimum). Its application as drought index takes into account all characteristics of an extreme event, summarizing them in a single value. A demonstration of its performance is shown with cases in Texas Climatic Division 5, USA, and Gibraltar. Results were compared with the Palmer Index and SPI. The statistical nature of the approach validates its use on comparing drought severity in different locations. It may also be applied to study other extreme natural events.

2   Methodology

The main goal in this work is the quantification of the deviations from the normal behavior of a hydroclimatic variable related with water availability. In this way, deviation from mean values may be quantified. For example, for the case of analyzing extreme periods of two time intervals duration (x_t, x_{t+1}) of a continuous variable normally distributed (say standard distribution, i.e. $x \in N\mu=0, \sigma=1$)), the parameter $R^2 = X_t^2 + X_{t+1}^2$ may be applied. Each R value may be associated to a probability derived from the bivariate distribution of (x_t, x_{t+1}), that for the case of temporal independence is represented in Fig. 1. However, the parameter R is no the best way to statistically quantify deviations from normality

in drought analysis. The main reason is that this does not take into account the deviation direction, and equally treat positive, negative or combined deviations.

In the drought characterization problem, the interest is in the deviation towards a certain improper point of the $\mathbb{R}^w$ space, where w is the dimension of the considered event (i.e. for the case of two variables, positive $(+\infty, +\infty)$ or negative $(-\infty, -\infty)$ pole, depending on the established reference). Taking as working pole the improper point $(+\infty, +\infty)$, to quantify the deviations toward this pole is defined the function $\ell(x_1, x_2, \dots, x_w)$ as the probability of occurring events with variable values larger or equal than the observed (i.e. $\ell(X_1, X_2, \dots, X_w) = P(x_1 \geq X_1 \cap x_2 \geq X_2 \cap \dots, \cap x_w \geq X_w)$). In this way, an event is considered larger than a second is if the ℓ value of the first is lower than the second. Figure 1 illustrates this definition for the case of $w = 2$ and represents the shape of equal ℓ functions, when joint events are of equal magnitude.

To complete the introduction of the procedure, an additional consideration needs to be taken into account for tractability: The ℓ function can be used to compare events of equal duration. However, how can events with different duration periods be compared? The solution is not straightforward since ℓ values coming from different length periods can not be directly compared. This is because such probabilities are coming from probability spaces of different dimensions, so larger dimensions imply additional restrictions and thus lower probability. The approach proposed to compare events coming from different durations is by analyzing the associated characteristic of greater interest in the extreme analysis: the mean frequency of recurrence (MFR) or its inverse, the Return Period. When computing the MFR of events with different duration periods, all the probabilities resulting from spaces of different dimensions are expressed in the same scale. Either MFR or Return Period may be used to compare events of different durations. A method to compute MFR of events for any duration, assuming independence of the variables in each time interval, was analytically developed.

2.1   Computation of the MFR of events with duration w

In a set $X_1, X_2, \dots, X_w$ of w realizations of a random variable x, with probability density function $f(x)$, and cumulative density function $F(x)$, the function $\ell(X_1, X_2, \dots, X_w)$ is defined as the probability of events with $x_i \geq X_i$, for $i = 1, 2, \dots, w$.

$$\ell(X_1, X_2, \dots, X_w) = P[(x_1, x_2, \dots, x_w)/x_i \geq X_i,$$
$$i = 1, 2, \dots, w] \quad (1)$$

Under the assumption of independence between realizations, ℓ can be computed as:

$$\ell(X_1, X_2, \dots, X_w) = \prod_{i=1}^{w} (1 - F(X_i)) \quad (2)$$

and $\ell(x_1, x_2, \dots, x_w) = \alpha$, where α is assumed constant, represents a surface in the $\mathbb{R}^w$ space. Defining $X_c = F^{-1}(1 - \alpha)$, from Eq. (2) it follows that, over that surface, as $x_j \rightarrow -\infty$ for all $j \neq i$, then $x_i \rightarrow X_c$, since $\ell(x_1, x_2, \dots, x_w) \rightarrow 1 - F(x_i) = \alpha$. Therefore, over this surface, the range of values of each component x_i is $(-\infty, X_c)$. The surface $\ell(x_1, x_2, \dots, x_w) = \alpha$ will be referred as $\ell_w(X_c)$.

The function $\ell_w(X_c)$ delimits a volume for which $\ell(x_1, x_2, \dots, x_w) < \alpha = 1 - F(X_c)$. This volume in a $\mathbb{R}^w$ space is referred as L_w and is defined by X_c, or equivalently by α, as shown in Eq. (3).

$$L_w(X_c) = \{(x_1, x_2, \dots, x_w)/\ell(x_1, x_2, \dots, x_w)$$
$$< (1 - F(X_c))\} \quad (3)$$

In the volume $L_w(X_c)$ each component ranges between $(-\infty, X_c)$. When $k < w$, $(X_1, X_2, \dots, X_k)$ with $\ell(X_1, X_2, \dots, X_w) < (1 - F(X_c))$, are fixed, then the remainder components of the points into $L_w(X_c)$ in the $\mathbb{R}^w$ space lay in a narrower interval $(-\infty, H(X_c; X_1, X_2, \dots, X_k))$, where $H(X_c; X_1, X_2, \dots, X_k)$ is defined as follows:

$$H(X_c; X_1, X_2, \dots, X_k)$$
$$= F^{-1}\left(1 - \frac{1 - F(X_c)}{\prod_{i=1}^{k}(1 - F(X_i))}\right) \quad (4)$$

The probability of occurring events $(x_1, x_2, \dots, x_w)$ belonging to $L_w(X_c)$ is defined as $P[L_w(X_c)] = P[(x_1, x_2, \dots, x_w) \in L_w(X_c)]$, and its computation results from integrating the PDF into $L_w(X_c)$, i.e.:

$$P[L_w(X_c)] = \int_{L_w(Xc)} \prod_{i=1}^{w} f(s_i) \cdot dV \quad (5)$$

where dV is the volume differential in the $\mathbb{R}^w$ space. As independence is assumed, the PDF in a point $(x_1, x_2, \dots, x_w)$ is equal to the cumulative product of all single PDFs, $f(x_i)$. This integral can be written using H functions as:

$$P[L_w(X_c)] = \int_{-\infty}^{X_c} f(s_w) \cdot \int_{-\infty}^{H(X_c; s_w)} f(s_{w-1}) \cdot \dots \cdot \int_{-\infty}^{H(X_c; s_w, s_{w-1}, \dots, s_2)} f(s_1) \cdot ds_1 \cdot \dots \cdot ds_{w-1} \cdot ds_w \quad (6)$$

However, the main interest is the associated MFR of events outside $L_w(X_c)$, for a fixed X_c (i.e. larger that this set of events), rather than the probability of $L_w(X_c)$. To compute it, let us assume $x_1, x_2, ..., x_{n-1}$ to be a set of consecutive realizations of the random variable x. After a new realization x_n is observed, define p_n as the conditional probability that the set of last w realizations (i.e. $x_{n-w+1}, x_{n-w+2}, ..., x_n$) belongs to $L_w(X_c)$, under the condition that all earlier sets of w consecutive realizations (i.e. $x_{i-w+1}, x_{i-w+2}, ..., x_i$ $i=w, w+1, ..., n-1$) belongs also to $L_w(X_c)$.

$$p_n = P[(x_{n-w+1}, x_{n-w+2}, ..., x_n)$$
$$\in L_w(X_c)/(x_{i-w+1}, x_{i-w+2}, ..., x_i)$$
$$\in L_w(X_c), \quad i = w, w+1, ..., n-1] \tag{7}$$

As n increases, p_n tends to a constant value that represents the probability that a $w-$event occurs (i.e. set of w consec-utive realizations) belonging to $L_w(X_c)$, given that earlier $w-$events also belong to $L_w(X_c)$. The backwards extension in time of these earlier events must be taken, from a practical point of view, as when the influences of earliest realizations have a negligible influence in p_n and it tends to a constant value. Under such practical point of view, as n increases the complementary $q_n=1-p_n$ tends to the mean frequency of recurrence (MFR) of $w-$events larger or equal than $\ell_w(X_c)$ assuming that previous $w-$events were smaller. Additionally, the return period of $w-$events larger or equal than $\ell_w(X_c)$, after sufficient realizations of $w-$events smaller than $\ell_w(X_c)$, is $1/q_n$, for large n.

Therefore, to evaluate MFR or the return period of $w-$events larger or equal than $\ell_w(X_c)$ requires the computation of conditional probabilities. Details of the analytical derivation of the expressions that provide the algorithms to compute the MFR are shown in Appendix A.

For $n \leq w$ it follows that:

$$\prod_{i=1}^{n} p_i = \int_{-\infty}^{X_c} f(s_n) \cdot \int_{-\infty}^{H(X_c;s_n)} f(s_{n-1}) \cdot ... \cdot \int_{-\infty}^{H(X_c;s_n,s_{n-1},...,s_2)} f(s_1) \cdot ds_1 \cdot ... \cdot ds_{n-1} \cdot ds_n \tag{8}$$

For $n > w + 1$, using a recurrence expression for computing the auxiliary function G_i:

$$G_1(s_w, s_{w-1}, ..., s_1) = \int_{-\infty}^{H(X_c;s_w,s_{w-1},...,s_2)} f(s_1) \cdot ds_1 \tag{9}$$

$$G_i(s_{i+w-1}, s_{i+w-2}, ..., s_{i+1}) = \int_{-\infty}^{H(X_c;s_{i+w-1},s_{i+w-2},...,s_{i+1})} f(s_i) \cdot G_{i-1}(s_{i+w-2}, s_{i+w-3}, ..., s_i) \cdot ds_i \quad \text{for } i > 1 \tag{10}$$

Thus:

$$\prod_{i=1}^{n} p_i = \int_{-\infty}^{X_c} f(s_n) \cdot \int_{-\infty}^{H(X_c;s_n)} f(s_{n-1}) \cdot ... \cdot \int_{-\infty}^{H(X_c;s_n,s_{n-1},...,s_{n-w+2})} f(s_{n-w+1}) \cdot G_{n-w}(s_{n-1}, s_{n-2}, ..., s_{n-w+1})$$
$$\cdot ds_{n-w+1} \cdot ... \cdot ds_{n-1} \cdot ds_n \tag{11}$$

Following the recurrence expressions (9) to (11) and designating P_n as

$$P_n = \prod_{i=1}^{n} p_i \tag{12}$$

p_n can be computed dividing P_n by P_{n-1}. The value of the MFR coincides with $q_n=1-p_n$ as n tends to infinity.

2.2   Transformation for the MFR computations

The computation of p_n involves a significant amount of integrals, especially for large w. For many PDFs, $f(x)$, those integrals do not have an analytical solution (e.g. the normal distribution) and numerical methods need to be applied to approximate the solution. Since it is a sequential process with multiplicative structure, care must be taken to control the propagation of errors. Furthermore, even with small errors, these may limit the application, because the probability characterization of extreme events is very sensitive to round off errors. A transformation is now presented that simplifies the integrals in the operations sequence, provides analytical solutions for the approximations that avoids numerical errors and allows better control of round of errors.

Define the transformation of x to a new variable v by Eq. (13).

$$v = -L_n(1 - F(x)) \tag{13}$$

Under this transformation, when $x \to -\infty$, $v \to 0$; and when $x \to X_c$, $v \to V_c=-L_n(1 - F(X_c))$, it follows that

$$V_c = V_1 + V_2 + ... + V_w \tag{14}$$

The function $\ell(x_1, x_2, \ldots, x_w)$ written in terms of v_i has a simpler form

$$\ell(v_1, v_2, \ldots, v_w) = e^{-\sum_{i=1}^{w} v_i} \tag{15}$$

Therefore, the surface represented by $\ell_w(V_c)$ forms, in term of v's, the region of the plane $v_1 + v_2 + \ldots + v_w = v_c$, with $v_i > 0$ for $i = 1, 2, \ldots, w$. The volume defined by $L_w(V_c)$ now takes the form of a polyhedron delimited by $\ell_w(V_c)$ and the planes $v_i = 0$ for $i = 1, 2, \ldots, w$. The function H, used to set the integration ranges of components of points belonging to $L_w(V_c)$. After fixing several components Eq. (4), may now be expressed as

$$H(V_c, v_1, v_2, \ldots, v_k) = V_c - \sum_{i=1}^{k} v_i \tag{16}$$

To perform the integrals, the Jacobian of the transformation is required. From Eq. (13)

$$\frac{dv}{dx} = \frac{f(x)}{1 - F(x)} = \frac{f(v)}{e^{-v}} \tag{17}$$

and the Jacobian function is

$$J(v) = \frac{dx}{dv} = \frac{e^{-v}}{f(v)} \tag{18}$$

This simplification allows the representation of the integrals in a simpler form. For example, the probability of $L_w(V_c)$, Eq. (6), using Eqs. (17) and (18) now takes the form

$$P[L_w(V_c)] = \int_0^{V_c} e^{-v_w} \cdot \int_0^{V_c - v_w} e^{-v_{w-1}} \cdot \ldots \cdot \int_0^{V_c - \sum_{i=2}^{w} v_i} e^{-v_1} \cdot dv_1 \cdot \ldots \cdot dv_{w-1} \cdot dv_w \tag{19}$$

$$P[L_w(V_c)] = 1 - \left(\sum_{i=1}^{w} \frac{V_c^i}{i!} \right) \cdot e^{-v_c} \tag{20}$$

This result is independent of the distribution function used for x. The only role of such distribution is in computing V_c.

The recurrence expressions (9) to (10) after this transformation yield to:

$$p_1 = 1 - e^{-v_c} \tag{21}$$

$$P_n = \prod_{i=1}^{n} p_i = \int_0^{V_c} e^{-v_n} \cdot \int_0^{V_c - \sum_{i=n}^{n} v_i} e^{-v_{n-1}} \cdot \ldots \cdot \int_0^{V_c - \sum_{i=2}^{n} v_i} e^{-v_1} \cdot dv_1 \cdot \ldots \cdot dv_{n-1} \cdot dv_n \tag{22}$$

for $n = 2, \ldots, w$

$$G_1(v_w, v_{w-1}, \ldots, v_2) = \int_0^{V_c - \sum_{i=2}^{w} v_i} e^{-v_1} \cdot dv_1 \tag{23}$$

$$G_i(v_{i+w-1}, v_{i+w-2}, \ldots, v_{i+1}) = \int_0^{V_c - \sum_{j=i+1}^{i+w-1} v_j} e^{-v_1} \cdot G_{i-1}(v_{i+w-2}, v_{i+w-3}, \ldots, v_i) \cdot dv_i \tag{24}$$

for $i > 1$

$$P_n = \int_0^{V_c} e^{-v_n} \cdot \int_0^{V_c - \sum_{i=n}^{n} v_i} e^{-v_{n-1}} \cdot \ldots \cdot \int_0^{V_c - \sum_{i=n-w+2}^{n} v_i} e^{-v_{n-w+1}} \cdot G_{n-w}(v_{n-1}, v_{n-2}, \ldots, v_{n-w+1}) \cdot dv_{n-w+1} \cdot \ldots \cdot dv_{n-1} \cdot dv_n \tag{25}$$

for $n > w$

Based on this transformation, the analytical expression of P_n may be calculated, and it is possible to evaluate p_n for different n's, studying its convergence. However, the number of terms in this analytical solution grows fast as n or w increases. P_n has the general form:

$$P_n = 1 + \sum_{j=1}^{n} q_j(V_c) \cdot e^{-j \cdot V_c} \qquad (26)$$

where $q_j(V_c)$ with $j = 1, 2, \ldots, n$ are polynomials and a function of V_c. An aspect of the computation of P_n is that as n increases the coefficients in the polynomials $q_j(V_c)$ become larger. For V_c smaller than a given value ε the round-off errors associated with the computation of Eq. (26) get larger or equal order of magnitude that the exact solution. The value of ε, delimiting the range for which round-off errors are relevant, increases with n. Therefore, care must be taken with the maximum value of n adopted and the precision of the calculation to avoid divergence.

Additionally, for large values of V_c, the terms in the summation, that is larger than a fixed ε, the terms in the summation associated with larger indexes (i.e. j) can be neglected

compared with first indexes. So, a good approximation to the exact solution can be achieved by truncating, in the sequence of integrals, the results to the first m terms.

$$P_n \approx 1 + \sum_{j=1}^{m} q_j(V_c) \cdot e^{-j \cdot V_c} \qquad \text{for} \quad m < n \qquad (27)$$

2.3 Approximate solution of MFR of events with duration w

Using truncation techniques an approximation to the function that relates the conditional probability p with V_c, for a given w, may be calculated. This function will have a fractional form, where both the numerator and denominator are summations of terms composed by a polynomial in V_c multiplied by an integer power of the negative exponential of V_c.

$$p_w = A_w(V_c)/B_w(V_c) \qquad (28)$$

$A_w(V_c)$ and $B_w(V_c)$ may be expressed in matrix form. For example, with $w = 2$ and truncation level $m=5$, $p_{w=2}$ is approximated by

$$A_{w=2}(V_c) = (1 \ V_c \ V_c^2 \ V_c^3 \ V_c^4 \ V_c^5) \cdot \begin{pmatrix} 1 & 7 & 19 & 21 & -6 & -42 \\ 0 & -9 & -42 & -65 & -24 & 14 \\ 0 & 0 & \frac{49}{2} & \frac{125}{2} & 36 & 0 \\ 0 & 0 & 0 & -\frac{125}{6} & -18 & -\frac{2}{3} \\ 0 & 0 & 0 & 0 & \frac{27}{8} & \frac{1}{8} \\ 0 & 0 & 0 & 0 & 0 & -\frac{1}{120} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ e^{-V_c} \\ e^{-2V_c} \\ e^{-3V_c} \\ e^{-4V_c} \\ e^{-5V_c} \end{pmatrix} \qquad (29)$$

$$B_{w=2}(V_c) = (1 \ V_c \ V_c^2 \ V_c^3 \ V_c^4 \ V_c^5) \cdot \begin{pmatrix} 1 & 6 & 13 & 8 & -14 & -14 \\ 0 & -8 & -30 & -32 & 0 & 0 \\ 0 & 0 & 18 & 32 & 8 & 0 \\ 0 & 0 & 0 & -\frac{32}{3} & -4 & 0 \\ 0 & 0 & 0 & 0 & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ e^{-V_c} \\ e^{-2V_c} \\ e^{-3V_c} \\ e^{-4V_c} \\ e^{-5V_c} \end{pmatrix} \qquad (30)$$

The approximate expressions for larger w have a similar structure and may be computed with Eqs. (21) to (25) using or not truncation techniques. In Fig. 2 the p_w function is plotted for values of w from 2 to 8. For a given p, as w increases the V_c value that provides the same p increases too. This tendency is shown in Fig. 3, where, for a set of p values, it is plotted the values of V_c versus w. The curves are smooth concave curves. This allows interpolations in w, so no all p_w's functions require to be calculated.

From the conditional probabilities p_w, its complementary $q_w = 1 - p_w$ is the Mean Frequency of Recurrence (MFR) and the inverse $1/q_w$ is the Return Period of such events.

An important outcome of this work is that the results are distribution-free. The statistical distribution of the variable x only plays a role in the X_c transformation to V_c.

3   Application: the MFR as a drought frequency index

The technique developed above can be applied to characterize droughts in term of their Mean Frequency of Recurrence (MFR). Both historical events and present events may be characterized. The index expresses the degree of rareness of the event or situation based on the statistical normal behavior.

In order to illustrate how to characterize a given situation, Fig. 4 shows the analysis of the situation at time step n of a hypothetic signal that follows a standard normal distribution. The signal at that time is finishing a period of persistent low values. Different durations may be analyzed at that time to characterize the current situation. The p values estimated using similar structure expressions that Eq. (28) are shown
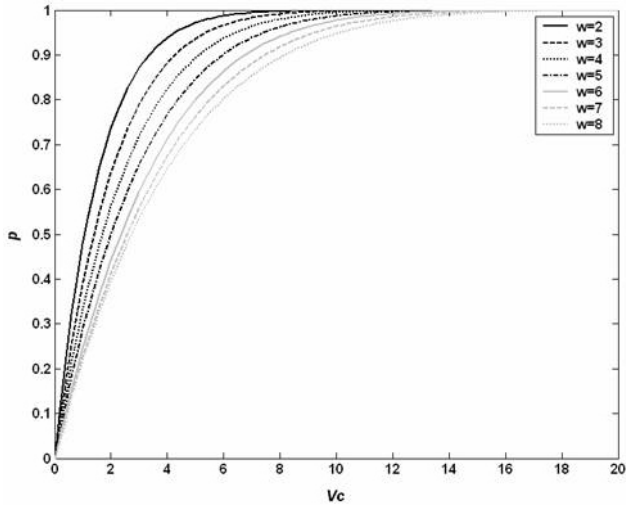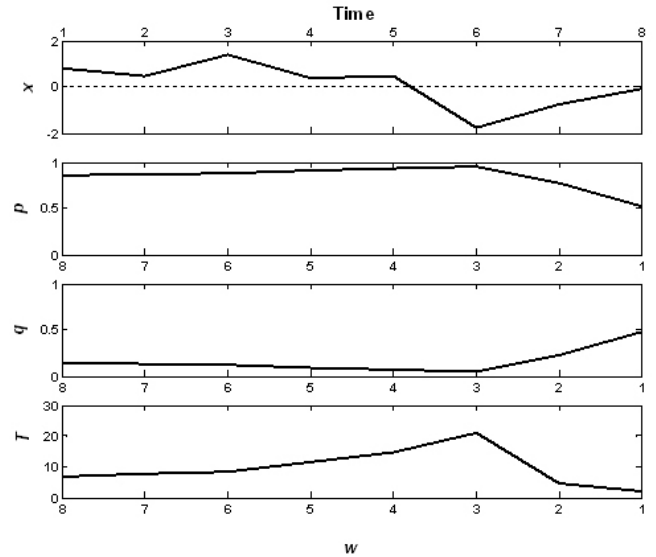
Fig. 2. Marginal probability of events of duration w, belonging to $L(X_c = F^{-1}(1 - e^{-V_c}))$ after indefinite succession of events belonging to $L(X_c)$
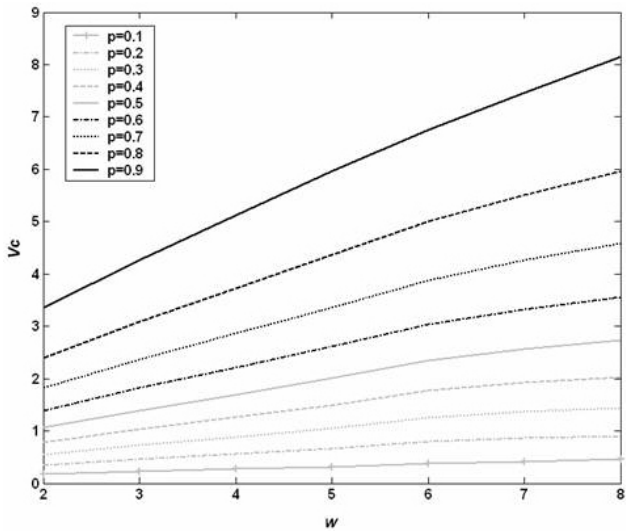

Fig. 3. V_c versus w for a fixed marginal probability p, which smooth variation allows interpolation in w

Fig. 4. Illustration of the computation of the drought index in time step n for time series x. At this time step the marginal probability p, the MFR or q, and the Return Period T are computed for events of different duration w, all ended at time step 8. The index value is defined as the maximum of all Return Periods of events that ended at this time step. In this case, the drought index value is equal to 21 for (X_{n-2}, X_{n-1}, X_n)
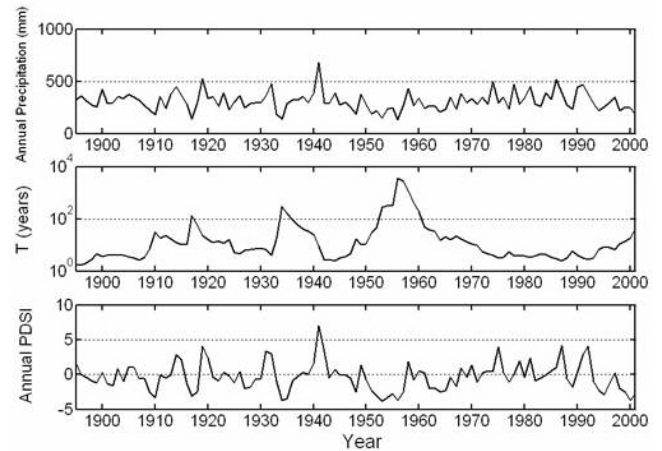
Fig. 5. Records of annual precipitation and drought study with recurrence analysis (T) and PDSI, for Texas Climatic Division 5 (USA)

below for durations $w = 1 \ldots 8$ time steps, all periods ending at time step n (i.e. (X_n), (X_{n-1}, X_n), $\ldots$, $(X_{n-7}, X_{n-6}, \ldots, X_n)$). The MFR, $q = 1 - p$, and the Return Period, $T = 1/q$ are also presented. If only the last value X_n is of interest, this looks normal, and T is low. However, when the last two or three values are analyzed, T increases, especially for $w = 3$. This includes the extreme period of low values that still prevails at this time. As w increases, T decreases because the mean behavior of the signal gets closer to normal. Therefore, to characterize the situation at time step 8, from the point of view of period of low extreme values, the natural way is associating the observed largest Return Period with the events that ended at this time step. The value in this case is $T = 21$ time step units and this will be the value

of the drought index that would characterize the situation at this time step. This corresponds with the idea that droughts are not only extreme low values of a water-availability signal, but also persistency must be taken into account. With this procedure, persistence is included by going back in time looking for the most extreme period.

The above procedure characterizes the situation at a fixed time step, getting the drought index value, equal to the largest T of the periods that ended at that moment. Characterizing in this way each time, in a time series, produce a representation of the variation of the degree of dryness. As an
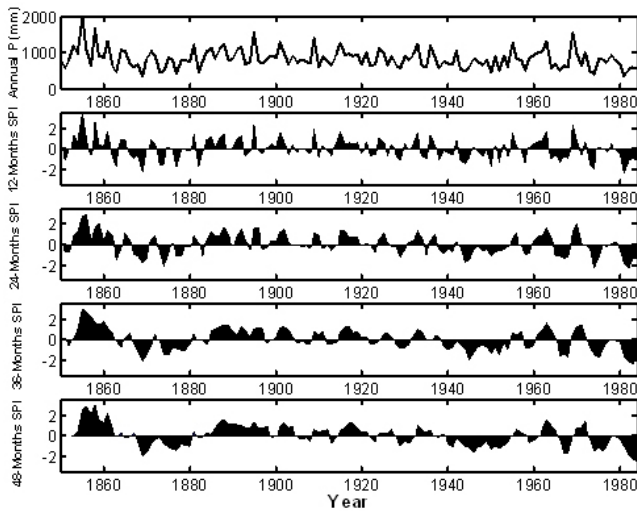
Fig. 6. Precipitation records and SPI in Gibraltar.

example, Fig. 5 presents the application of the index to the drought analysis of annual precipitation in Texas Climatic Division 5. The index characterizes the degree of extreme dryness of current periods after each year. Main droughts occurred in the 1930's, 1950's, and in current years. The results of applying the Palmer drought index, PDSI, are also presented. PDSI expresses the current precipitation regime, in addition to temperature, as a dimensionless index. Values below -1 represent drought periods with different grades of intensity depending on the index value (from -1 to -2 indicate mild drought, from -2 to -3 moderate drought, from -3 to -4 severe drought, and lower than -4 extreme drought). Since expressed in a dimensionless scale, PDSI values may be compared among drought events from different geographical locations (always keeping in mind that PDSI was originally developed for arid and semiarid regions and its application in different climate regions are outside the original scope).

During the Texas 1950's drought the Palmer index was low, with values below -4 for several consecutive years. However, while drought was occurring and its persistence increased, the index kept similar low values without providing a clear idea of the drought magnitude which was increasing in time. Another examples to support the point of no persistence representation of the Palmer index are found by analyzing the 1930's and 1950's droughts. In both cases the index achieves values lower than -4, so both are extreme droughts. However, the index does not distinguish which event is more exceptional. Drought durations must be included in addition to the index value to account for persistence, but no integral quantification of the magnitude of each drought is supplied.

On the other hand the use of the proposed approach, computing the return period of extremely low events in the precipitation records analysis overcomes these shortcomings. When the 1950's drought starts, the return period of the event becomes large and increases as the situation continues, given

by the higher persistence, which implies a more exceptional event. As normal conditions returns, the return period decreases going back to normal values. Analyzing the 1930's and 1950's events, even when the minimum annual precipitation was lower for the 1930's, the 1950's drought achieved a return period of the order of 1000 years, while the 1930's was 100 years. This is obviously due to persistence. For the case of the 1950's drought the return period is of similar magnitude to that calculated by Gonzalez and Valdes (2003), in that case coming from paleoclimatic data and the bivariate statistical analysis of droughts using PDSI ($T=700$ years). For the case of the 1930's drought the return period coming from the recurrence analysis increases significantly ($T=44$ years). Nevertheless, relative increments are expected because the presented approach computes return periods with larger degrees of freedom.

The MFR or the return period, T, provides a direct idea of the event magnitude that is represented by an unique value. In both real time analyses and retrospective analyses, the approach is useful for quantification proposes. Another comparable aspect between PDSI and MFR is the spatial behavior of the index. As analyzed by Guttman et al. (1992) the Palmer index does not account with the a priori desired property of spatial comparability of the index. As these authors suggest, magnitudes of probabilistic nature are prescribed for this aim. In this way, the MFR or T values are fully suitable. Its probability groundwork allows comparability between different locations, even with various climates. According with this idea, the SPI is also a drought index with a probabilistic nature, which Guttman et al. (1998) compared its spatial behavior with PDSI. Their work showed the better spatial comparability properties of the SPI index with respect to the PDSI.

SPI represents the distribution of cumulative precipitation for different periods transformed as normalized standard population. Figure 6 presents the SPI analysis applied over Gibraltar precipitation records for periods of 12, 24, 36, 48 months. The index allows for the quantification of both dry and wet periods. The different aggregation time provides information about the patter occurrence of different kind of droughts (e.g. short droughts, useful for agricultural purposes, or long droughts, of interest for WRM). This fact and its spatial comparability are the main advantages of the SPI drought index.

Nevertheless, the SPI index does not allow considering different time aggregations, when comparing events that, even when they correspond to the same temporal scale, have different durations. To illustrate this point, the drought analysis performed with SPI in Gibraltar (Fig. 6) reveals several droughts in the record. Two of them occurred in 1960's and 1970's. These episodes have different characteristics: the first one has a duration of 5 years, while the second extends over a shorter period of 2 years. The 1960's drought has a smaller intensity than the 1970's drought, so the first corresponds with an event of small intensity but prolonged over a long period, while the second coincides with a short high intensity event. To evaluate which event is more exceptional,
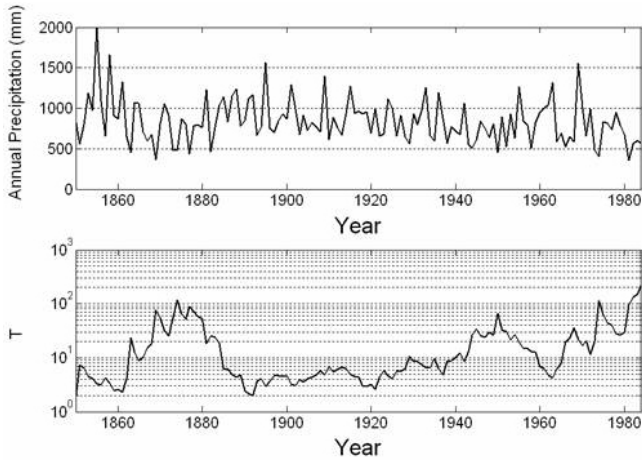
Fig. 7. Precipitation records and drought recurrence analysis in Gibraltar.

if cumulative periods of 12 and 24 months are used, the SPI index indicated that the 1970's has a higher magnitude. However, if a cumulative period of 48 months is considered, the index set the 1960's drought as more severe.

The question of ranking droughts is not completely resolved using SPI, however, this is an important question. For example, in WRM the design and management of reservoir systems is different depending on the expected duration of shortage periods. Figure 7 presents the drought analysis in Gibraltar using the proposed approach. Two more extreme events occurred in 1879's and 1980's, corresponding to 60 and 100 years return period respectively. The procedure presented in this paper determined that the 1960's drought has a return period of 20 years and the 1970's a return period of approximately 50 years. Thus the 1970's drought is more exceptional, even when shorter.

The main advantages of this analysis over the SPI is that it contains in one unique curve all this information, taking into account all the characteristics such as duration, severity, or distribution of the deficit. In addition, recurrence analysis has a physical meaning: the return period, which provides a direct idea of the grade of exceptionality of the event. As a statistical property, the MFR may be used to compare drought events from different climates and regions. In the analyzed case studies the precipitation regimes are completely different with significant differences in their variability. In the case of Texas Climatic Division 5, it belongs to a semiarid region, whose coefficient of variation of the annual precipitation reaches a value of 0.650. On the other hand, Gibraltar is a humid region, with an annual precipitation coefficient of variation of 0.324. Even with so different precipitation regimes, the MFR efficiently characterizes drought episodes and makes possible the comparison of the drought conditions between both regions. This is a desired property in WRM dealing with extensive systems affected by different precipitation regimes. The MFR could provide information about which region is under a more severe event, so resource compensation could be performed. This may allow for the management of diversions between different watersheds.

The work provides the recurrence frequency of $w-$events. This frequency is used to distinguish between events of different durations. Low frequencies represent a higher exceptionality of the event. Future work is required to compute the return period of events for any duration. The frequency of droughts of a given magnitude is higher when the possibility of any event duration is considered.

4   Conclusions

The proposed approach to compute the MFR of in-time-multidimensional events allows the practical use of this concept in the characterization of droughts. The approach requires the condition of temporal independence of the involved variable, but it does not depend on the probability distribution of the variable. The application of the MFR of multidimensional events in drought characterization, and the formulation of the associated index, has the advantage of dealing with droughts in term of probabilities, and it characterizes droughts integrating all the aspects of an event: duration, severity, maximum intensity, etc. The probability nature of the index coming from the recurrence analysis justifies its correct application on comparing drought events of different places and climates. This is an advantage from other indices not based on statistic properties.

The main advantage with respect the SPI, which is also based on statistical concepts, is that it embeds the duration on the analysis, and does not require a parallel analysis like the SPI does. Furthermore, the presented approach may be applied to any hydroclimatic variable of interest, not only precipitation. The resulting index allows the representation of the main drought characteristics in a single value, based on the stochastic feature of the phenomenon, and scaled on the mean frequency of recurrence. The index value has complete meaning, in the sense way that provides the Return Period of the actual situation, considered as the more extreme from the recent observations.

The approach is useful not only for drought characterization, but also in any other analysis of natural hazards from extreme deviations from the normality of a variable during a certain extension in time or space (e.g. floods), as well as, in system state monitoring for water resources management, analyzing precipitation, stream flows, aquifer recharges, soil moisture, etc.

Appendix A   Analytical derivation of the MFR

To evaluate the MFR or the return period of $w-$events larger or equal than $\ell_w(X_c)$ requires to compute a succession of conditional probabilities. The evaluation of p_n, as defined in Eq. (7), is as follows:

1. After the first realization, x_1, no $w-$event may be studied, but as x_1 must belong to a set $(x_1, x_2, \dots, x_w) \in L_w(X_c)$, then $x_1 \in (-\infty, X_c)$. Therefore $p_1 = F(X_c)$ and the conditional PDF of x_1 is given by Eq. (A.1).

$$f_1(x_1) = f(x_1)/p_1 \qquad\qquad x_1 \in (-\infty, X_c) \tag{A1}$$

2. When the second realization occurs, x_2, the set (x_1, x_2) must be from a $w-$event belonging to $L_w(X_c)$. If $w=2$, then $(x_1, x_2) \in L_w(X_c)$. When $w>2$, the remaining components, $(x_3, x_4, \dots, x_w)$ are still undetermined, and it is only known that they are in the interval $(-\infty, X_c)$. From this fact, when $x_i \to -\infty$, $i=3, 4, \dots w$, at least (x_1, x_2) must belong to $L_2(X_c)$. Therefore, even for $w>2$, $(x_1, x_2) \in L_2(X_c)$ the conditional probability p_2 follows Eq. (A.2).

$$p_2 = \int_{-\infty}^{X_c} f(s_2) \cdot \int_{-\infty}^{H(X_c;s_1)} f_1(s_1) \cdot ds_1 \cdot ds_2 \tag{A2}$$

From (A.1) it follows that:

$$p_1 \cdot p_2 = \int_{-\infty}^{X_c} f(s_2) \cdot \int_{-\infty}^{H(X_c;s_2)} f(s_1) \cdot ds_1 \cdot ds_2 \tag{A3}$$

3. Similarly, for $n \le w$ it follows that

$$p_n = \int_{-\infty}^{X_c} f(s_n) \cdot \int_{-\infty}^{H(X_c;s_n)} f_{n-1}(s_{n-1}) \cdot \dots \cdot \int_{-\infty}^{H(X_c;s_n,s_{n-1},\dots,s_2)} f_1(s_1) \cdot ds_1 \cdot \dots \cdot ds_{n-1} \cdot ds_n \tag{A4}$$

with $f_i(x_i) = f(x_i)/p_i$ $i=1, 2, \dots, n-1$, and therefore

$$\prod_{i=1}^{n} p_i = \int_{-\infty}^{X_c} f(s_n) \cdot \int_{-\infty}^{H(X_c;s_n)} f(s_{n-1}) \cdot \dots \cdot \int_{-\infty}^{H(X_c;s_n,s_{n-1},\dots,s_2)} f(s_1) \cdot ds_1 \cdot \dots \cdot ds_{n-1} \cdot ds_n \tag{A5}$$

4. When $n=w+1$, the conditional probability p_w+1 is given by Eq. (A.6).

$$p_{w+1} = \int_{-\infty}^{X_c} f(s_{w+1}) \cdot \int_{-\infty}^{H(X_c;s_{w+1})} f_w(s_w) \cdot \dots \cdot \int_{-\infty}^{H(X_c;s_{w+1},s_w,\dots,s_3)} f_2(s_2) \cdot \int_{-\infty}^{H(X_c;s_w,s_{w-1},\dots,s_2)} f_1(s_1) \cdot ds_1 \cdot ds_2 \cdot \dots \cdot ds_w \cdot ds_{w+1} \tag{A6}$$

Defining the function

$$G_1(s_w, s_{w-1}, \dots, s_2) = \int_{-\infty}^{H(X_c;s_w,s_{w-1},\dots,s_2)} f(s_1) \cdot ds_1 \tag{A7}$$

it follows that

$$\prod_{i=1}^{n} p_i = \int_{-\infty}^{X_c} f(s_{w+1}) \cdot \int_{-\infty}^{H(X_c;s_{w+1})} f(s_w) \cdot \dots \cdot \int_{-\infty}^{H(X_c;s_{w+1},s_w,\dots,s_3)} f(s_2) \cdot G_1(s_w, s_{w-1}, \dots, s_2) \cdot ds_2 \cdot \dots \cdot ds_w \cdot ds_{w+1} \tag{A8}$$

5. For a general $n > w+1$ and taking into consideration the recurrence expression

$$G_i(s_{i+w-1}, s_{i+w-2}, \ldots, s_{i+1}) = \int_{-\infty}^{H(X_c; s_{i+w-1}, s_{i+w-2}, \ldots, s_{i+1})} f(s_i) \cdot G_{i-1}(s_{i+w-2}, s_{i+w-3}, \ldots, s_i) \cdot ds_i \tag{A9}$$

yields

$$\prod_{i=1}^{n} p_i = \int_{-\infty}^{X_c} f(s_n) \cdot \int_{-\infty}^{H(X_c; s_n)} f(s_{n-1}) \cdot \ldots \cdot \int_{-\infty}^{H(X_c; s_n, s_{n-1}, \ldots, s_{n-w+1})} f(s_{n-w+1}) \cdot G_{n-w}(s_{n-1}, s_{n-2}, \ldots, s_{n-w+1})$$

$$\cdot ds_{n-w+1} \cdot \ldots \cdot ds_{n-1} \cdot ds_n \tag{A10}$$

Following the recurrence expressions (A.9) and (A.10) and designating P_n as

$$P_n = \prod_{i=1}^{n} p_i \tag{A11}$$

p_n can be computed dividing P_n by P_{n-1}. Therefore, $q_n = 1 - p_n$ is the marginal probability of occurring a w−event larger than the defined by X_c. As n tends to infinity, q_n tends to the MFR of such events, and its inverse is the return period.

References

Chung, C. and Salas, J. D.: Drought occurrence probability and risks of dependent hydrologic processes, J. Hydrologic Engrg., 5(3), 259–268, 2000.

Dalezios, N. R., Loukas, A., Vasiliades, L., and Liakopoulos, H.: Severity-Duration-Frequency Analysis of Droughts and Wet Periods in Greece, Hydrol. Sci. J. Vol., 45(5), 751–770, 2000.

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

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

Fernandez, B. and Salas, J. D.: Return period and risk of hydrologic events, II, J. Hydrologic Engrg., 4(4), 297–307, 1999.

Gonzalez, J. and Valdes, J. B.: Bivariate drought recurrence analysis using tree rings reconstructions, J. Hydrologic Engrg., 8(5), 247–258, 2003.

Guttman, N. B., Wallis, J. R., and Hosking, J. R. M.: Spatial comparability of the Palmer Drought Severity Index, J. Am. Water Resour. Assoc., 28(6), 1111–1119, 1992.

Guttman, N. B.: Comparing The Palmer Drought Index and the Standardized Precipitation Index, J. Am. Water Resour. Assoc., 34(1), 1029–1039, 1998.

Heim Jr., R. R.: Drought indices: A review, in: Drought: A Global Assessment, edited by Wilhite, D. A., Routledge, 159–167, 2000.

Keyantash, J. and Dracup, J.A.: The quantification of drought: an evaluation of drought indices, Bull. Am. Meteorol. Soc., August, 1167–1180, 2002.

Kim, T. W., Valdes, J. B., and Aparicio, J.: Frequency and spatial characteristics of droughts in the Conchos River Basin, Mexico, Water Intern., 27(3), 420–430, 2003.

Laio, F., Porporato, A., Ridolfi, L., and Rodriguez-Iturbe, I.: Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress II, Probabilistic soil moisture dynamics, Adv. Water Resources, 24, 707–723, 2001.

Loukas, A., Vasiliades, L., and Dalezios, N. R.: Hidroclimatic variability of regional droughts in Greece using the Palmer Moisture Anomaly Index, Nord. Hydrol. J., 33(5), 425–442, 2002.

McKee, T. B., Doesken, N. J., and Kleist, J.: Drought monitoring with multiple timescales, Proceedings Eighth Conf. on Applied Climatology, Anaheim, CA, Amer. Meteor. Soc., 179–184, 1993.

Palmer, W. C.: Meteorological Drought. Research Paper No. 45, US Department of Commerce, Weather Bureau, Washington, D.C., 1965.

Porporato, A., Laio, F., Ridolfi, L., and Rodriguez-Iturbe, I.: Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress III, Vegetation water stress, Adv. Water Resources, 24, 725–744, 2001.

Rodriguez-Iturbe, I., Porporato, A., Laio, F., and Ridolfi, L.: Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress I, Scope and general outline, Adv. Water Resources, 24, 695–705, 2001.

Sen, Z.: Wet and dry periods of annual flows series, J. Hydr. Div., ASCE, 102(10), 1503–1514, 1976.

Sen, Z.: Run-sums of annual flow series, J. Hydro., Amsterdam, 35, 311–324, 1977.

Sen, Z.: Statistical analysis of hydrologic critical droughts, J. Hydr. Div., 106(1), 99–115, 1980.

Shiau, J. and Shen, H. W.: Recurrence analysis of hydrologic drought of differing severity, J. of Water Resources Plann. and Managm., 127(1), 30–40, 2001.