

Effects of disregarding seasonality on the distribution of hydrological extremes

P. Allamano, F. Laio, and P. Claps

Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

Received: 10 May 2011 – Published in Hydrol. Earth Syst. Sci. Discuss.: 13 May 2011

Revised: 17 October 2011 – Accepted: 18 October 2011 – Published: 20 October 2011

Abstract. This paper deals with the seasonality of hydroclimatic extremes and with the problem of accounting for their non-homogeneous character in determining the design value. To this aim we devise a simple stochastic experiment in which extremes are produced by a non-homogeneous extreme value generation process. The design values are estimated in closed analytical form both in a peak over threshold framework and by using the standard annual maxima approach. In this completely controlled world of generated hydrological extremes, a statistical measure of the error associated to the adoption of a homogeneous model is introduced. The sensitivity of this measure, named return period ratio, to the typology and strength of seasonality is investigated. We find that neglecting seasonality induces a downward bias in design value estimators. The magnitude of the bias may be large when the peak over threshold approach is adopted, while the return period distortion is limited when the annual maxima are considered. An application to rainfall data from a 30 000 km² region located in North-Western Italy is presented to better clarify the effects of disregarding seasonality in a real case.

1 Introduction

Hydrological variables, like precipitation depth and river discharge, often exhibit marked periodic variability at the annual time scale (Black and Werritty, 1996; Sivapalan et al., 2005; Viglione et al., 2010; Villarini et al., 2011). As a consequence, also the corresponding extreme values, i.e. the precipitation or discharge values associated to relevant and unfrequent events, will likely be dependent on the

non-homogeneity of their respective date of occurrence (e.g. Maraun et al., 2009). The seasonal variability of extreme events represents an issue for the standard applications of the frequency analysis, because the presence of seasonality can make some commonly used assumptions inappropriate. In particular, one should relax either the assumption of a standard probability distribution to represent the data (e.g. Gumbel or GEV) or the assumption that all available data are sampled from the same distribution (see Allamano et al., 2011b, p.3684-5). The manner how these date-dependent sample values should be treated represents an open problem in statistical hydrology, tackled in different ways.

The standard procedures for design value estimation are based on the analysis of series of annual maxima (AM). Estimation of the distribution of annual maxima from data with seasonal variability has received attention in the literature. Attempts to obtain the annual maximum distribution by fitting different distributions to the maxima in separate seasons – and to derive the expressions for the bias and variance of the resulting annual quantile estimators – were done by Lamberti and Pilati (1985) and Buishand and Demaré (1990). The basic idea is to have, in each season, random variables that satisfy the “identically distributed” hypothesis, which is essential in standard statistical inference. The use of monthly maxima has also been advocated (e.g. Carter and Challenor, 1981; Ettrick et al., 1987; Rust et al., 2009) because the hypothesis of identically distributed random variables in classical extreme-value theory is better satisfied for monthly than for annual maxima.

The problem with these approaches is that the number of parameters to be estimated increases, while the amount of information useful to describe the upper tail of the distribution increases at a lower rate, thus inflating the estimation uncertainty of design values. In addition, the application of these methods requires non-standard data (maxima in separate seasons) which may not be available for long time periods.



Correspondence to: P. Allamano
(paola.allamano@polito.it)

Maxima in separate seasons are, in fact, usually extracted from the continuous time measurements, which are systematically collected since the beginning of the digital era.

If a continuous time series is available, a valid alternative is offered by the peak over threshold (POT) or partial duration series approach. This approach considers several over-threshold values instead of a unique extreme value per year (Madsen et al., 1997; Lang et al., 1999; Claps and Laio, 2003). The strongest motivation for adopting the POT approach is that over-threshold values constitute precious additional information about the upper tail of the distribution, which is very important in hydrological applications (Revfeim, 1991; Katz et al., 2002). The POT approach naturally lends itself to being used with a rate of occurrence of the exceedances (λ) and an exceedances distribution having an annual cycle (Todorovic and Zelenhasic, 1970; Rasmussen and Rosbjerg, 1991). However, there are numerous applications of the POT approach which do not consider time-dependent parameters (e.g. Madsen et al., 1997; Claps and Laio, 2003; Bacova-Mitkova and Onderka, 2010; Fruh et al., 2010).

The aim of this paper is to provide indications about the effects of neglecting seasonality within standard POT or AM approaches. A very simple and controllable framework (which we call “toy-model” in Sect. 2) is adopted, suitable for sharpening the questions regarding the possible effects of seasonality on design event estimation. The scope of our contribution, hence, is not to propose a new model for use in POT or AM analyses, but to analyze the magnitude of under- (or over-) estimation of design events in the presence of seasonality. In other words, our aim is to address the question “what is the impact of neglecting seasonality on design event estimation?”.

2 A “toy-model” for understanding seasonality effects on the distributions of extremes

A simplified analytical model of the distribution of hydroclimatic extremes is devised, in which hydroclimatic variables are assumed to follow a non-homogeneous Poisson process of event arrivals in time with rate $\lambda(t)$. The intensity x of each event is distributed according to an exponential function

$$F(x|t) = 1 - \exp[-x/\alpha(t)] \quad (1)$$

where the mean $\alpha(t)$ of the distribution is dependent on time. Despite its simplicity, the model is widely adopted in hydrology, as demonstrated by the numerous applications that can be found in the literature (e.g. Eagleson, 1978; Rodriguez Iturbe et al., 1987; Rasmussen and Rosbjerg, 1989; Laio et al., 2001). Observe that the exponential is merely used as an example (simple and commonly used) distribution. To mimic the effects of seasonality, a sinusoidal (time-dependent) variation is imposed on the parameters of the

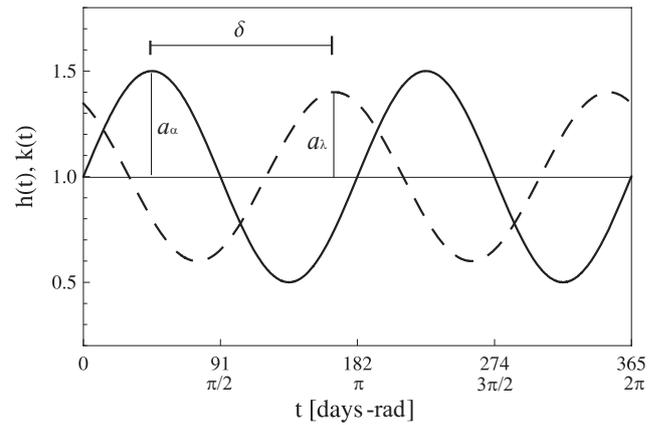


Fig. 1. Example of the $h(t)$ (solid) and $k(t)$ (dashed) curves, see Eq. (2). The temporal shift δ between the seasonal peaks is indicated, as well as their amplitudes, a_α and a_λ . In this case $n=2$.

Poisson process. The sine assumption is taken as a compromise between the need to provide a realistic representation of a periodic time series and the quest for simplicity that drives our research. We write $\alpha(t)$ and $\lambda(t)$ as

$$\alpha(t) = \alpha_0 \left(1 + a_\alpha \sin \left(\frac{2\pi}{365} nt \right) \right) = \alpha_0 \cdot h(t), \quad (2)$$

$$\lambda(t) = \frac{\lambda_0}{365} \left(1 + a_\lambda \sin \left(\frac{2\pi}{365} nt + \delta \right) \right) = \frac{\lambda_0}{365} \cdot k(t),$$

where α_0 and λ_0 are the average intensity and rate values; t is time in days ($t \in [0; 365]$); n is the number of cycles per year; $h(t)$ is the non-dimensional α regime and $k(t)$ the non-dimensional λ regime; a_α and a_λ are, respectively, the amplitudes of the sinusoids $h(t)$ and $k(t)$; and δ is the temporal shift between the two (i.e. if the sinusoids are in phase events are more intense when they are more frequent). Note that, to constrain $h(t)$ and $k(t)$ above zero, a_α and a_λ can vary in the range (0, 1). Observe also that the initial time $t=0$ is taken at any point where $\alpha(t)=\alpha_0$ and $d\alpha(t)/dt > 0$. An example for the $h(t)$ and $k(t)$ curves is reported in Fig. 1.

The distribution in Eq. (1) is the conditioned distribution $F(x|t)$ of the event intensities x on the date of occurrence t . The marginal distribution of the exceedances $F(x)$ can then be obtained by applying the Law of Total Probability (or Bayes theorem):

$$F^{(\text{POT})}(x) = \frac{1}{365} \int_0^{365} F(x|t) k(t) dt. \quad (3)$$

An analytical solution to the integral in Eq. (3) is not available. To overcome this problem we substitute $F(x|t)$ with $F'(x|t) = 1 - \exp[-x \cdot \alpha'(t)]$ (see the Appendix for details). This guarantees the analytical tractability of Eq. (3). Of course $F(x|t)$ and $F'(x|t)$ behave similarly, however it is not essential to have a precise correspondence between the two functions. In reality, in fact, the average intensities are

not strictly sinusoidal and (in a toy-model) could be represented also by an inverse sinusoidal function, $\alpha'(t)$. Our initial guess to start with a standard sinusoidal form for the event intensities serves the purpose of making our results more easy to interpret.

Setting $F'(x|t)$ for $F(x|t)$ in Eq. (3) leads to an expression of the distribution of exceedances in closed analytical form:

$$F^{(POT)}(x) = 1 - e^{-\frac{x}{\alpha_0}} \left\{ \Phi \left[0, \frac{a_\alpha}{\alpha_0 (1 + a_\alpha)} x \right] + a_\lambda \Phi \left[1, \frac{a_\alpha}{\alpha_0 (1 + a_\alpha)} x \right] \cos(\delta) \right\} \quad (4)$$

where $\Phi[\cdot]$ is the modified Bessel function of the first kind (Abramowitz and Stegun, 1965, ch. 9) and $k(t)$ is as defined in Eq. (2). An important property of $F(x)$ is that its expression does not change by changing the oscillation frequency (parameter n in Eq. (2)). Observe also that the mean of $F(x)$ can be obtained analytically as the first moment of the distribution in Eq. (4)

$$\mu_x = \frac{\alpha_0 (a_\alpha - (1 + a_\alpha) \left(\sqrt{\frac{1 + 2a_\alpha}{(1 + a_\alpha)^2}} - 1 \right) a_\lambda \cos(\delta))}{a_\alpha \sqrt{\frac{1 + 2a_\alpha}{(1 + a_\alpha)^2}}} \quad (5)$$

Moreover, the distribution in Eq. (4) tends to an exponential distribution (also called “base POT distribution”) in the homogeneous case, i.e. when $a_\alpha = a_\lambda = 0$.

The distribution of exceedances Eq. (4) can then be transposed into the correspondent annual maximum distribution. In fact, despite the non-homogeneity of the process, the standard relation $F^{(AM)}(x) = e^{-\lambda_0(1-F(x))}$ (relating the distribution of exceedances to the distribution of annual maxima, $F^{(AM)}(x)$) still holds (see Allamano et al., 2011a, p.C2850). On this premise, the following expression for the distribution of annual maxima is obtained

$$F^{(AM)}(x) = \exp \left[-\exp \left[\frac{x}{\alpha_0} \right] \lambda_0 \left(\Phi \left[0, \frac{a_\alpha x}{\alpha_0 (1 + a_\alpha)} \right] + a_\lambda \Phi \left[1, \frac{a_\alpha x}{\alpha_0 (1 + a_\alpha)} \right] \cos(\delta) \right) \right] \quad (6)$$

in which α_0 plays the role of the scale parameter. Observe that the distribution of maxima becomes a Gumbel distribution (also called “base distribution”) when $a_\alpha = a_\lambda = 0$ in Eq. (2). Observe also that for Eq. (6) the mean and standard deviation of the distribution are not analytical.

3 Model application

Ignoring the possible effects of seasonality when adapting a distribution to a set of data is a rather common simplification, especially when few observations are available. In

the following we explore the consequences of such a simplification, as they could affect the return period estimation both for the distribution of exceedances and the distribution of maxima.

To quantify the error associated to the selection of the base (non-seasonal) distribution, we consider the T -year event with the non-seasonal distribution and compute the corresponding return period, T^* , with the (correct) seasonal distribution. The return period ratio

$$R_T = \frac{T}{T^*} \quad (7)$$

provides an indication of the magnitude of the underestimation (overestimation) in design values related to neglecting seasonality. In particular, if $R_T > 1$ the adoption of the base distribution corresponds to an underestimation of the real design value.

3.1 Peak over threshold

Consider process exceedances to follow the distribution in Eq. (4), having mean μ_x as defined by Eq. (5). As anticipated, in the homogeneous case $a_\alpha = a_\lambda = 0$, and the distribution of exceedances becomes an exponential distribution, also called “base distribution”. The design event $x_T^{(POT)}$ with the base distribution is determined as

$$x_T^{(POT)} = -\mu_x \log \left(\frac{1}{\lambda_0 T} \right) \quad (8)$$

Using the seasonal distribution Eq. (4) one finds that the same event $x_T^{(POT)}$ has a return period T^* , obtained by setting Eq. (8) for x in Eq. (4), and considering that $\lambda_0 T^* = 1/(1 - F(x_T^{(POT)}))$ (see e.g. Claps and Laio, 2003).

Given the simplicity of the model the return period ratio is obtained in analytical form, and reads

$$R_T^{(POT)} = \lambda_0 T \left(\frac{1}{\lambda_0 T} \right)^{\frac{z(1+a_\alpha)}{a_\alpha}} \left(\Phi \left[0, z \log \left(\frac{1}{\lambda_0 T} \right) \right] - a_\lambda \Phi \left[1, z \log \left(\frac{1}{\lambda_0 T} \right) \right] \cos(\delta) \right) \quad (9)$$

with $z = [a_\alpha - (1 + a_\alpha) (\sqrt{(1 + 2a_\alpha)/(1 + a_\alpha)^2} - 1) a_\lambda \cos(\delta)] / \sqrt{1 + 2a_\alpha}$. Note that Eq. (9) is independent of the mean intensity α_0 because α_0 is a scale parameter for both the base and the seasonal distribution.

The sensitivity of the return period ratio to the parameters of Eq. (4) is explored in Fig. 2 (gray-shaded contour areas). The results are obtained by varying two parameters at a time and by fixing the other parameters. In particular, in Fig. 2, λ_0 and a_α are assumed to vary. In each point of the plane the intensity of the grey-shade is proportional to the value of the return period ratio, with large values of the return period ratio $R_T^{(POT)}$ corresponding to events $R_T^{(POT)}$ -times more frequent in the “seasonal reality” than if the exponential distribution was adopted. We find a strong sensitivity of the return period

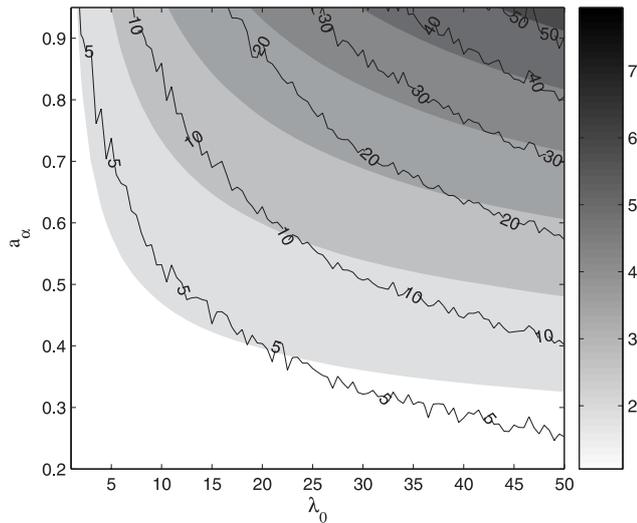


Fig. 2. Sensitivity of the return period $R_T^{(POT)}$ to variations of λ_0 and a_α , with $T = 100$ years, $\delta = \pi$, $n = 1$ and $a_\lambda = 0.5$. In each point of the plane the intensity of the grey-shade is proportional to the value of the return period ratio $R_T^{(POT)}$. The percentage of cases recognized as non-exponential by the Anderson-Darling test (at a 5 % level) is indicated by the black (labelled) contour lines. $N = 20$ years is considered.

ratio to large values of a_α and λ_0 . The design event dependence on this latter parameter (λ_0) is formalized by Eq. (8). From this equation it can be inferred that considering larger λ_0 values correspond to move further towards the right tail of the distribution, thus exacerbating the differences between the base and seasonal distribution.

The sensitivity of $R_T^{(POT)}$ to the variation of the two other parameters T and δ is shown, instead, in Fig. 3: it emerges that $R_T^{(POT)}$ significantly increases with both the return period T and the phase shift δ . The reason behind the dependence of $R_T^{(POT)}$ on δ is to be found in the event generation mechanism. In fact, when the two regimes are in phase we have a very high probability of picking all the exceedances in the same season, with the consequence that they will be almost identically and exponentially distributed. Conversely, when the two regimes are out of phase, we will have a well-mixed distribution of the events deriving by the coexistence of one season characterized by few very intense events with another season with a lot of small events. In this case, the non-exponentiality of the distribution will be very marked, with consequent high R_T values. From Fig. 3 it can be also deduced that $R_T^{(POT)}$ assumes slightly lower values when the regimes are in phase.

One could argue that the effects of seasonality would be recognized if the data were closely scrutinized. This is verified by applying the Anderson-Darling goodness of fit test for exponentiality to the samples generated by the seasonal model. The Anderson-Darling test is especially well suited

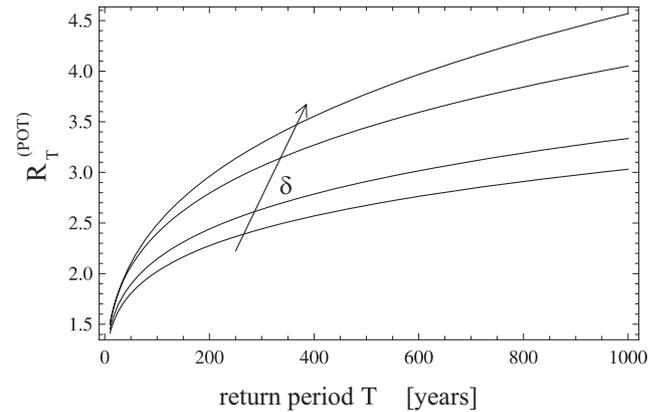


Fig. 3. Relation between the return period ratio $R_T^{(POT)}$ and T for increasing values of δ (0, 365/8, 365/4, 365/2), with $a_\alpha = a_\lambda = 0.5$, $n = 1$ and $\lambda_0 = 20$.

to recognize deviations in the tails of the hypothetical and operational distributions, and its failure to recognize the difference between the two distributions is precisely a demonstration that there might be a serious problem of model selection in the presence of seasonality (see Laio, 2004). In details, 1000 samples of length N' , generated from the distribution Eq. (4) under different parameterizations, are tested against the hypothesis of exponential distribution. In Fig. 2, black labelled contour lines show the percentage of cases recognized as non-exponential by the Anderson-Darling test (at a 5 % level) for the case of $N = 20$ years (which entails a sample length of $N' = \lambda_0 N$). It is observed that, for small a_α values, the percentages in the graph are very close to the significance level of the test, meaning that the distribution in this region is substantially undistinguishable from an exponential distribution. The percentage of cases recognized as non-exponential increases up to 50 % for larger values of a_α (i.e. for very peaky seasonal regimes) and λ_0 .

By comparing the two sets of curves (gray-shaded areas and contours) we observe that the pattern of the $R_T^{(POT)}$ values and of the test performances (as functions of λ_0 and a_α) are rather similar: low efficiencies of the Anderson-Darling test emerge in correspondence of low $R_T^{(POT)}$ values, whereas better performances are found where $R_T^{(POT)}$ is high. However, it is clear that the power of the test remains rather low even when the $R_T^{(POT)}$ values are large, with a 70 % probability of not being able to recognize a 5-fold increase in return period, with possible serious detrimental consequences in design-value estimation.

3.2 Annual maxima

Suppose that a sample of annual maxima follows the distribution function in Eq. (6), which is described by the five parameters α_0 , λ_0 , a_α , a_λ and δ . The Gumbel distribution (which is characterized by two parameters, called $\alpha_0^{(AM)}$ and $\lambda_0^{(AM)}$

in the following) is the form that Eq. (6) would assume for $a_\alpha = a_\lambda = 0$, i.e. in the absence of seasonality. An analysis of the error that would derive from the adoption of a Gumbel distribution in the presence of seasonal maxima is provided in the following.

As in the previous case, the two parameters $\alpha_0^{(AM)}$ and $\lambda_0^{(AM)}$ of a Gumbel distribution are estimated with the method of moments as

$$\alpha_0^{(AM)} = \frac{\sqrt{6} \sigma^{(AM)}}{\pi}, \tag{10}$$

$$\lambda_0^{(AM)} = \exp \left[\frac{\mu^{(AM)}}{\alpha_0} - \gamma_E \right],$$

where $\mu^{(AM)}$ is the mean and $\sigma^{(AM)}$ the standard deviation of the samples taken from the distribution in Eq. (6) and γ_E is the Euler constant. We then define $x_T^{(AM)} = \alpha_0 \log[-1/\lambda_0 \log[1 - 1/T]]$ as the design T -year event with the Gumbel distribution. The same event would be characterized by a return period T^* with the seasonal distribution Eq. (6). The return period ratio $R_T^{(AM)}$ is defined by Eq. (7) as the ratio of the reference return period T to the return period T^* of the $x_T^{(AM)}$ event. In this case the relation between the return period ratio and the distribution parameters cannot be expressed in closed analytical form. The ratio is computed several times by varying the parameters of Eq. (6). The values assumed by $R_T^{(AM)}$, after considering the variation of λ_0 and a_α and by keeping the other parameters constant, are shown in Fig. 4 (shaded areas). It can be observed that $R_T^{(AM)}$ assumes values up to 1.45, for high values of a_α and low λ_0 , whereas $R_T^{(AM)}$ is unaffected by the variation of the scale parameter α_0 (not shown). This implies that the assumption of a Gumbel model leads to less relevant overestimation of the return period of the design event than in the POT case.

To check if these errors would be detectable we apply the same testing procedure as for the POT case. The application demonstrates that the test is unable, for sample dimensions $N \leq 50$ years, to discern between a sample generated from Eq. (6) and a Gumbel-distributed sample. In Fig. 4 the percentage of cases recognized as non-Gumbel by the Anderson-Darling test (at a 5% level) is shown by the black contour lines, for variable values of a_α and λ_0 . It can be observed that the percentages in the graph are very close to the test significance level, i.e. the two distributions are statistically undistinguishable.

4 Case study

In this section 294 rainfall stations located in the North-West of Italy are examined, with the aim of demonstrating the importance of precipitation seasonality in the area of study.

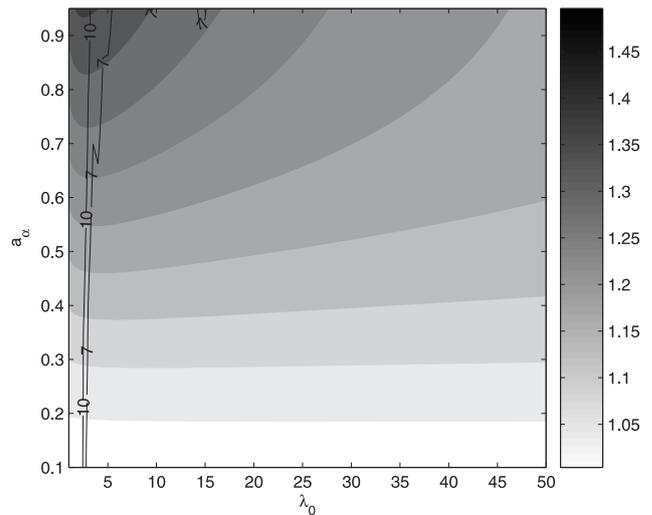


Fig. 4. Sensitivity of the return period ratio $R_T^{(AM)}$ to variations of λ_0 and a_α , with $T = 100$ years, $n = 1$, $\delta = \pi$ and $a_\lambda = 0.5$. In each point of the plane the intensity of the grey-shade is proportional to the value of the return period ratio. The percentage of cases recognized as non-Gumbel by the Anderson-Darling test (at a 5% level) is indicated by the black (labelled) contour lines. $N = 50$ years is considered.

Daily totals of precipitation depths are available between year 1913 and 1986. The peak over threshold series are extracted by imposing a threshold value of 20 mm on the precipitation depths. The threshold corresponds to about $\lambda = 20$ events per year over the whole data set, which seems a reasonable approximation for the average number of significant rain events in a year. The site-specific regime curves are obtained by aggregating the exceedance amounts and the occurrence rates at the monthly timescale and by subsequently averaging the results for each station. A representation of the at-site regimes $h(t)$ and $k(t)$ (respectively, the amount and occurrence regimes divided by their mean, see Eq. 2) is given in Fig. 5 for all the stations in the region (grey lines).

A marked seasonal behavior in both the amount and occurrence rate of the regimes emerges: in particular most of the curves are characterized by a double peak. The two curves also appear to have approximately the same oscillation period and to be in phase (i.e. the maximum rainfall occur when the storms are more frequent). The time shift between the two curves is determined by varying the value of δ and evaluating, for each station, a distance measure (D) between the site-specific regimes. D is defined as the average of the absolute differences between the 12 monthly values of $h(t)$ and $k(t)$. The minimum D is found in correspondence of $\delta = 0$ for 291 (out of 294) stations. The global regimes (black curves in Fig. 5) of the mean precipitation intensity and rainfall rate, derived by averaging the station-specific curves, are then assumed to be in phase for the whole region of study.

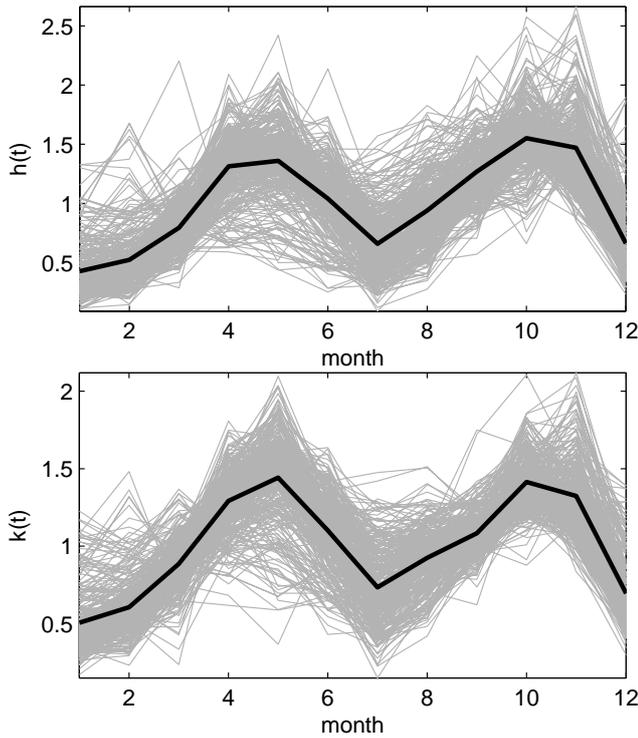


Fig. 5. Non-dimensional monthly average rainfall intensity $h(t)$ and rainfall rate $k(t)$ for 294 gauging stations in North-Western Italy (grey lines). The overall spatial mean values are reported in black.

The parameters α_0 and λ_0 are estimated in each station as the average (over-threshold) rainfall intensity and rainfall rate. a_α and a_λ are estimated as functions of the coefficients of variation (CV) of $\alpha(t)$ and $\lambda(t)$. The reasoning to find the relation between CV_λ and a_λ is described hereinafter. Similarly one can find the relation between CV_α and a_α .

Under the hypothesis of poissonianity, the cumulative distribution of λ , $F_\Lambda(\lambda)$, can be obtained as a derived distribution of the times of occurrence, provided that the relation $\lambda(t)$ is monotonic. To meet this condition, we restrict the domain of $\lambda(t)$ to the interval $[\pi/2n; 3\pi/2n]$ (which should be multiplied by $365/2\pi$ to be expressed in days). The distribution of λ is unaffected by this assumption, due to the periodicity of the $\lambda(t)$ function. On this domain, the cumulative distribution of times $F_T(t)$ is uniform, Eq. (2) relates λ to t , and the cumulative distribution function of λ reads

$$F_\Lambda(\lambda) = \frac{1}{\pi} \arcsin \left(\frac{\lambda_0 - \lambda}{a_\lambda \lambda_0} \right) - \frac{1}{2}, \quad (11)$$

which is valid in the domain $[\lambda_0(1 - a_\lambda); \lambda_0(1 + a_\lambda)]$. The mean and standard deviation are respectively $\mu_\lambda = \lambda_0$ and $\sigma_\lambda = a_\lambda \lambda_0 / \sqrt{2}$ and the relation between CV_λ and a_λ is found accordingly as $a_\lambda = \sqrt{2} \cdot CV_\lambda$. Analogously, one obtains $a_\alpha = \sqrt{2} \cdot CV_\alpha$.

$R_T^{(\text{POT})}$ values are then obtained from Eq. (9) for each station, as shown in Fig. 6. Rather large values are found,

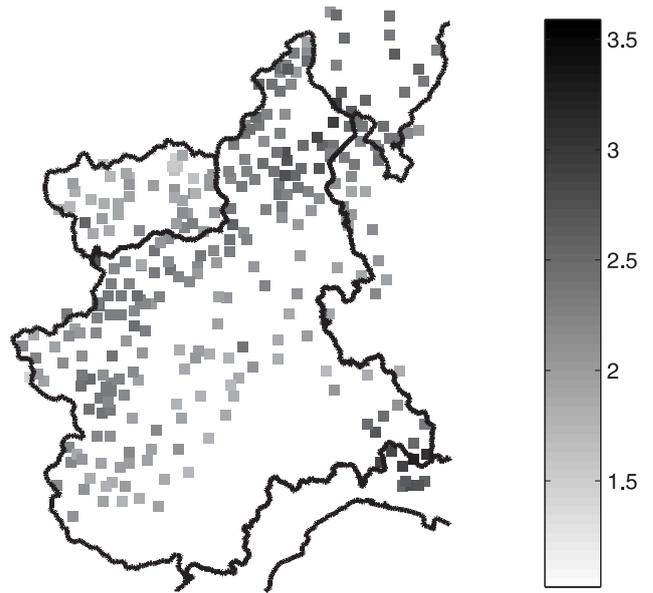


Fig. 6. Return period ratios ($R_T^{(\text{POT})}$) evaluated, for $T = 100$ years, in 294 rain gauging stations of the North-West of Italy. The color scale refers to the resulting range of variation of $R_T^{(\text{POT})}$.

pointing out the risk to significantly underestimate the design values if seasonality is neglected. When the annual maxima are considered, $R_T^{(\text{AM})}$ varies between 1 and 1.45 (not shown), demonstrating that in this case the errors induced by neglecting seasonality are not extremely relevant.

5 Discussion and conclusions

The effects of disregarding seasonality of extremes when evaluating the return period of a design event are investigated. Two very simple examples are illustrated: (1) the case of using an exponential distribution to describe POT values derived from a non-homogeneous Poisson process (Sect. 3.1); and (2) the case of using a Gumbel distribution for AM values extracted from a seasonal (i.e. with time dependent parameters) distribution (Sect. 3.2). The entity of the return period overestimation (i.e. apparently less frequent extreme events) when seasonality is not accounted for is quantified by a coefficient R_T . It is found that, when seasonality is ignored and threshold exceedances (POT) are analyzed, the return period of an event can be significantly overestimated. Neglecting seasonality therefore corresponds to significantly underestimating the design values; moreover, standard goodness-of-fit techniques are not very efficient at recognizing the unsuitability of the non-seasonal distribution.

Resorting to peak over threshold analysis is, nevertheless, an advantageous technique, especially in the presence of few years of observation, as acknowledged by numerous works (see e.g. Wilks, 1993; Lang et al., 1999). This allows, in fact, to capture more accurately the upper tail of the distribution,

increasing at the same time the sample size. In this sense, the difficulties pointed out by this study in analyzing seasonal POT values should not discourage the adoption of an over-threshold modelling, but rather reassert the need for accurate preliminary analyses of the data based on visual inspection of the regimes or, possibly, on unsupervised testing techniques (e.g. “analysis of variance” or ANOVA techniques, see e.g. Salas et al., 1980; Kottegoda and Rosso, 1998).

Less relevant errors are associated to the use of the annual maximum distribution, even if seasonality still induces a systematic downward bias in design value estimators. In particular, the Gumbel distribution is found to represent an acceptable choice to describe the annual maxima generated by a Poisson exponential process with seasonally varying parameters. The reason behind this better performance probably lies in the greater flexibility of the Gumbel distribution (compared to the exponential). Another possible motivation behind this behavior is that seasonality is often disguised by the AM approach: in fact, if seasonality is weak, the annual maxima will be evenly distributed across the year, and hence identically distributed. Conversely, if seasonality is strong, annual maxima would preeminently stem from a specific season, where the “identically distributed” hypothesis will be satisfied again.

However, while the Gumbel distribution generally performs well, it is of interest to note that, in the presence of seasonality, the estimated $\alpha_0^{(AM)}$ and $\lambda_0^{(AM)}$ values (see Eq. 10) lose (some of) their physical significance. In fact, $\alpha_0^{(AM)}$ and $\lambda_0^{(AM)}$ do not correspond to the average intensity and average rate, α_0 and λ_0 . In Fig. 7a the relative variations of the $\lambda_0^{(AM)}$ to λ_0 ratio are shown for varying a_α and a_λ . It is observed that the $\lambda_0^{(AM)}$ to λ_0 ratio takes values between 1 (in the homogeneous case) and 0.5, meaning that the estimated $\lambda_0^{(AM)}$ could be one half of the real λ_0 . In panel b the behavior of the non-dimensional product $\lambda_0^{(AM)} \alpha_0^{(AM)} / \lambda_0 \alpha_0$ is shown in the same parameter range. One can observe that the product values are fan-shaped around the unit value, with greater differences corresponding to higher a_α values. This entails that in the parameters estimation process the total “over-threshold” volume $\lambda_0^{(AM)} \cdot \alpha_0^{(AM)}$ tends to remain constant, with lower values assumed by λ being compensated by an increase in the estimated α . This systematic underestimation of λ and overestimation of α should be accounted for when the parameters $\lambda_0^{(AM)}$ and $\alpha_0^{(AM)}$ are used to infer the statistical properties of the extreme-value process in the area.

In conclusion, the message we want to convey with this work is about the need of accounting for seasonality when dealing with hydro-climatic extreme values. We show that disregarding their non-homogeneity may have detrimental consequences in design value estimation. Moreover, the errors may be difficult to detect by means of standard goodness of fit testing techniques. Future work will focus on the formulation of ad-hoc metrics to identify and quantify the seasonality in hydrological extremes.

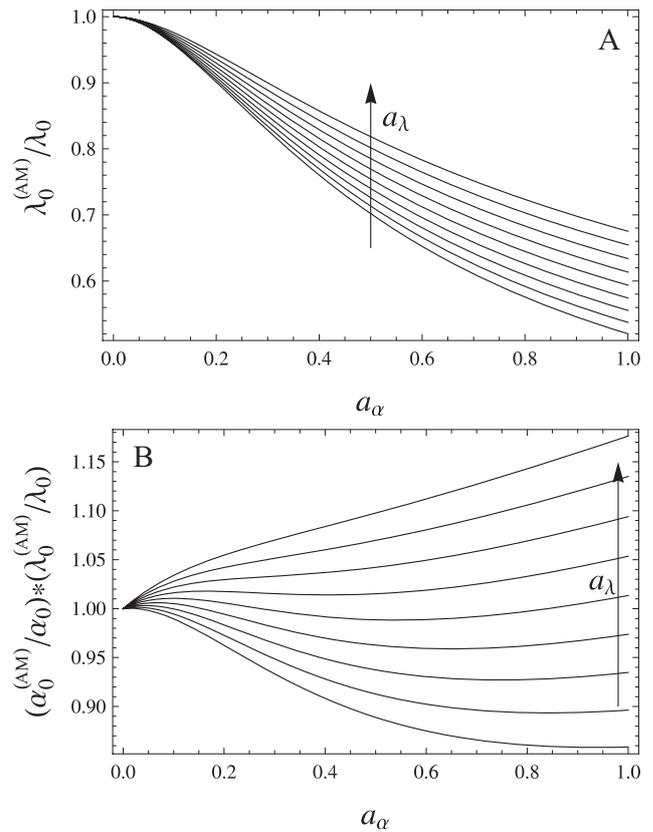


Fig. 7. Panel (A) Sensitivity of the $\lambda_0^{(AM)}$ to λ_0 ratio to varying a_α and a_λ values (with $T=100$ years, $\delta=0$, $n=1$ and $\lambda_0=20$ events/year). Panel (B) Sensitivity of the non-dimensional product $\lambda_0^{(AM)} \alpha_0^{(AM)} / \lambda_0 \alpha_0$ to the same parameter variations.

Appendix A

To solve the integral in Eq. (3) some further manipulation of Eq. (1) is needed. To this end we approximate $F(x|t)$ in Eq. (1) as

$$F'(x|t) = 1 - \exp[-x \cdot \alpha'(t)] \tag{A1}$$

with $\alpha'(t) = \alpha'_0 \cdot h'(t) = \alpha'_0 (1 - a'_\alpha \sin(2\pi nt/365))$. The minus sign before the sine assures that the maxima and minima of $\alpha(t)$ and $\alpha'(t)$ are in phase.

The values of the parameters α'_0 and a'_α should be found which minimize the distance between $F'(x|t)$ and $F(x|t)$. A possible strategy could be to calculate the total squared distance

$$\begin{aligned} \text{TSD}(\alpha'_0, a'_\alpha, x) &= \int_0^{365} [F(x|t) - F'(x|t)]^2 dt \tag{A2} \\ &= \int_0^{365} e^{-2x/\alpha(t)} [1 - e^{-x(\alpha'(t)-1/\alpha(t))}]^2 dt \end{aligned}$$

and to find the values of α'_0 and a'_α which minimize $\text{TSD}(\alpha'_0, a'_\alpha)$. The solution to this equation does not

provide an analytical result, but one can observe that the integral on the right hand side of Eq. (A2) can be seen as a weighted average of the distance between the curves, where the weighting factor is $w(t) = e^{-2x/\alpha(t)}$. It is clear that this weighting factor is maximum where $\alpha(t)$ is largest, i.e. in $t' = 365(\pi/2n + 2\pi/n)/2\pi$ and minimum in $t'' = 365(3\pi/2n + 2\pi/n)/2\pi$. The ratio between the minimum and maximum weighting factors reads

$$r = \frac{e^{-\frac{2x}{\alpha_0(1-a_\alpha)}}}{e^{-\frac{2x}{\alpha_0(1+a_\alpha)}}} = e^{-\frac{4x a_\alpha}{\alpha_0(1-a_\alpha^2)}} \quad (\text{A3})$$

and rapidly converges to zero for large values of x/α_0 , which are those of interest in this paper (see Allamano et al., 2011b, p.3684-5 for discussion). The differences between $F'(x|t)$ and $F(x|t)$ are therefore negligible when $\alpha(t)$ is small, and it is sensible to determine α'_0 and a'_0 when $\alpha(t)$ attains its maximum, i.e. in t' . By setting t' for t in $F(x|t)$ and $F'(x|t)$ and equating the two expressions one easily obtains $\alpha'_0 = 1/\alpha_0$ and $a'_0 = a_\alpha/(1+a_\alpha)$.

Acknowledgements. The work was funded by the Italian Ministry of Education (grants 2007HBTS85 and 2008KXN4K8). Comments and suggestions from P. F. Rasmussen, A. Evans, and two anonymous reviewers are gratefully acknowledged.

Edited by: N. Verhoest

References

- Abramowitz, M. and Stegun, I.: Handbook of Mathematical Functions, Applied Math. Series 55, Dover Publications, 1965.
- Allamano, P., Laio, F., and Claps, P.: Interactive comment on “Effects of seasonality on the distribution of hydrological extremes” by P. Allamano et al., Hydrol. Earth Syst. Sci. Discuss., 8, C2841–C2851, 2011a.
- Allamano, P., Laio, F., and Claps, P.: Interactive comment on “Effects of seasonality on the distribution of hydrological extremes” by P. Allamano et al., Hydrol. Earth Syst. Sci. Discuss., 8, C3680–C3690, 2011b.
- Bacova-Mitkova, V. and Onderka, M.: Analysis of extreme hydrological events on the Danube using the peak over threshold method, J. Hydrol. Hydromech., 58, 88–101, 2010.
- Black, A. and Werritty, A.: Seasonality of flooding: a case study of North Britain, J. Hydrol., 195, 1–25, 1996.
- Buishand, T. and Demaré, G.: Estimation of the annual maximum distribution from samples of maxima in separate seasons, Stoch. Hydrol. Hydraul., 4, 89–103, 1990.
- Carter, D. and Challenor, P.: Estimating return values of environmental parameters, Q. J. Roy. Meteorol. Soc., 107, 256–266, 1981.
- Claps, P. and Laio, F.: Can continuous streamflow data support flood frequency analysis? An alternative to the Partial Duration Series approach, Water Resour. Res., 39, 12–16, 2003.
- Eagleson, P.: Climate, soil, and vegetation 1. Introduction to water balance dynamics, Water Resour. Res., 14, 705–712, 1978.
- Ettrick, T., Mawdlsey, J., and Metcalfe, A.: The influence of antecedent catchment conditions on seasonal flood risk, Water Resour. Res., 23, 481–488, 1987.
- Fruh, B., Feldmann, H., and Panitz, H.: Determination of precipitation return values in complex terrain and their evaluation, J. Climate, 23, 2257–2274, 2010.
- Katz, R., Parlange, M., and Naveau, P.: Statistics of extremes in hydrology, Adv. Water Res., 25, 1287–1304, 2002.
- Kottegoda, N. and Rosso, R.: Statistics, probability, and reliability for civil and environmental engineers, International Edition, McGraw-Hill, 1998.
- Laio, F.: Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters, Water Resour. Res., 40, W09308, doi:10.1029/2004WR003204, 2004.
- 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 Res., 24, 707–723, 2001.
- Lamberti, P. and Pilati, S.: Probability distributions of annual maxima of seasonal hydrological variables, Hydrolog. Sci. J., 30, 111–135, 1985.
- Lang, M., Ouarda, T., and Bobée, B.: Towards operational guidelines for over-threshold modeling, J. Hydrol., 225, 103–117, 1999.
- Madsen, H., Rasmussen, P., and Rosbjerg, D.: Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events, 1. At-site modeling, Water Resour. Res., 33, 747–757, 1997.
- Maraun, D., Rust, H., and Osborn, T.: The annual cycle of heavy precipitation across the United Kingdom: a model based on extreme value statistics, Int. J. Climatol., 29, 1731–1744, 2009.
- Rasmussen, P. and Rosbjerg, D.: Risk estimation in partial duration series, Water Resour. Res., 25, 2319–2330, 1989.
- Rasmussen, P. and Rosbjerg, D.: Prediction uncertainty in seasonal partial duration series, Water Resour. Res., 27, 2875–2883, 1991.
- Revfeim, K.: Annual maxima and totals of seasonally varying processes, Stoch. Hydrol. Hydraul., 5, 147–153, 1991.
- Rodriguez Iturbe, I., Cox, D., and Isham, V.: Some models for rainfall based on stochastic point-processes, P. Roy. Soc. Lond. A, 410, 269–288, 1987.
- Rust, H., Maraun, D., and Osborn, T.: Modelling seasonality in extreme precipitation, Eur. Phys. J. Special Topics, 174, 2009.
- Salas, J., Delleur, J., Yevjevich, V., and Lane, W.: Applied Modeling of Hydrologic Time Series, Water Resources Publications, Littleton, Colorado, 1980.
- Sivapalan, M., Blöschl, G., Merz, R., and Gutknecht, D.: Linking flood frequency to long-term water balance: Incorporating effects of seasonality, Water Resour. Res., 41, W06012, doi:10.1029/2004WR003439, 2005.
- Todorovic, P. and Zelenhasic, E.: A stochastic model for flood analysis, Water Resour. Res., 6, 211–222, 1970.
- Viglione, A., Chirico, G., Komma, J., Woods, R., Borga, M., and Blöschl, G.: Quantifying space-time dynamics of flood event types, J. Hydrol., 394, 213–229, 2010.

Villarini, G., Smith, J., Serinaldi, F., and Ntelekos, A.: Analyses of seasonal and annual maximum daily discharge records for central Europe, *J. Hydrol.*, 399, 299–312, 2011.

Wilks, D.: Comparison of three-parameters probability distributions for representing annual extreme and partial duration precipitation series, *Water Resour. Res.*, 29, 3543–3549, 1993.