Alternative model for precipitation probability distribution: application to Spain

Roberto Moncho[1,2,*], Vicente Caselles[2], Guillem Chust[3]

[1]Fundación para la Investigación del Clima, C/ Gran Vía 22, 6º izda., 28013 Madrid, Spain
[2]Departament de Física de la Terra i Termodinàmica, Universitat de València, C/ Dr. Moliner 50, 46100 Burjassot, Spain
[3]AZTI-Tecnalia, Marine Research Division, Txatxarramendi ugartea z/g, 48395 Sukarrieta, Spain

ABSTRACT: In climatology, there is difficulty in describing the probability distribution of rainfall because there are many days without precipitation, which usually causes the most likely daily rainfall to be zero. None of the widely used models is able to describe the overall variation of daily precipitation. This article proposes an alternative model for the probability of precipitation. The model, based on 4 parameters, has been applied to daily rainfall throughout all months of the year and for 108 stations in Spain. This alternative model provides better results than the commonly used probability models (Generalized Extreme Value, Pareto and Generalized Pareto Distribution, Gamma, Gumbel, Weibull, Exponential and Log-normal). Our model had a mean absolute error of <10% for most of the stations analyzed. Thus, this alternative model could be used to correct the probability distributions of daily precipitation obtained from weather forecasting and climate models.

KEY WORDS: Probability distribution · Return period · Precipitation · Spain

1. INTRODUCTION

The effect of projected climate change on precipitation is an issue of current interest (Kharin & Zwiers 2000), especially the prediction of changes in the occurrence of extreme rainfall and drought for agricultural and environmental applications. However, current climate models do not adequately reproduce the probability distribution of precipitation (Ebert 2001, Wilson & Toumi 2005, Perkins et al. 2007). The problem is intrinsic to the irregular nature of rainfall, and reflects the limitations of climate models and the constraints imposed by computation time required.

The rainfall probability distributions are very asymmetrical: for most weather stations, the maximum probability of daily rainfall is zero; thus, the empirical density function decreases as rainfall increases. Even if we complete the domain of the function of probability with evaporation as a mode of negative precipitation, distribution would still be asymmetric. This is because the potential evaporation rarely reaches

20 mm in 1 day at most stations worldwide (McGuinness & Leslie 1972, Granger 1989, Jeevananda Reddy 1995), although values >50 mm in 1 day are recorded quite frequently in some countries (Cerveny et al. 2007). A particular problem of the Mediterranean climate is that extreme precipitation is frequent. These factors hinder fitting the probability distribution to daily precipitation.

Thus, the commonly used models (Generalized Extreme Value [GEV], Generalized Pareto Distribution [GPD], Gamma, Gumbel, Weibull) do not fit the overall variation of daily precipitation (very low and very high precipitation in the same fit) (Bridges & Haan 1972, Öztürk 1981, Etoh et al. 1986, Begueria 2005). All of these models use 2 or 3 parameters; hence, it may be necessary to use probability models with more parameters (Park & Jung 2002, Hanson & Vogel 2008).

The objective of this study was to present an alternative model for the probability distribution of daily precipitation (including days without rain and with

rainfall <0.1 mm). The new model aims to characterize and correct the probability distribution of a generic set of daily precipitation data. This is an important requirement for improving the systematic correction of dynamic and statistical downscaling (Kysely 2002). This model should also improve the calibration of operational weather models and other tools for prediction and analysis of precipitation, such as weather radars.

2. DATA AND METHODS

2.1. Study area

The study area consisted of Spain, including the Canary and Balearic Islands (Fig. 1). The prevailing climate in Spain is Mediterranean, although other climatic conditions can be found regionally, including oceanic (in the north of the Iberian Peninsula), mountain (in the Sierra Nevada and Pyrenees), and dry subtropical climates (in the Canary Islands). Spain has a varied climate due to its geographical location and characteristics of the terrain. The Iberian Peninsula is located within the influence of the polar jet waves, or Rossby waves, and it is also occasionally affected by the subtropical jet stream (Alves & Verdière 1999, Peliz et al. 2002).

Therefore, climatic conditions are heterogeneous; e.g. annual rainfall ranges from 150 mm in the south-

east of Andalusia to 2500 mm in the northeast of the Basque Country, which has an oceanic climate (Capel-Molina 2000). The latter is characterized by regular precipitation throughout the year (in some places, with >50% of days experiencing rainfall). In contrast, the Mediterranean climate is characterized by dry summers and heavy rainfall concentrated in a few days (Lana et al. 2004, Martín-Vide 2004), especially in autumn and spring. In Spain, the number of days with precipitation usually ranges between 10 and 30% over the year, although it can attain values <10% in July, and in some cases, >50% in April (Capel-Molina 2000). For these reasons, Spain is a suitable pilot area for testing the methodology developed in the present study.

2.2. Precipitation data

The daily precipitation data used in this study were obtained from 108 stations with daily resolution from the State Meteorological Agency of Spain (AEMET) (Fig. 1). The selected time series have a length of at least 20 yr and 90% of days with records. Among these stations, 62 had a time series of >50 yr and 98% of days with records. The data showed a high frequency of negligible precipitation, between 0.6 and 16% of days with <0.1 mm rainfall. These values were changed to 0, and hence, the study took into account that the zeros actually represent the class of (0, 0.1) mm. In addition, some erroneous daily values (accumulated for several days) were found in the series analyzed; they constituted <0.1% of the series. Outliers were not found among these values, and hence did not affect the statistics. In these cases, the accumulated totals were divided by the corresponding number of days.

2.3. Models

For this study, the probability for each station was analyzed separately. If P is any precipitation and p is precipitation greater than P, then the cumulative probability $\pi(p \geq P)$ is defined as the probability that a station or group of stations registers a daily precipitation greater than the rainfall threshold P on a day of the year or of a particular month. The occurrence frequency for each rainfall threshold, P, can be esti-
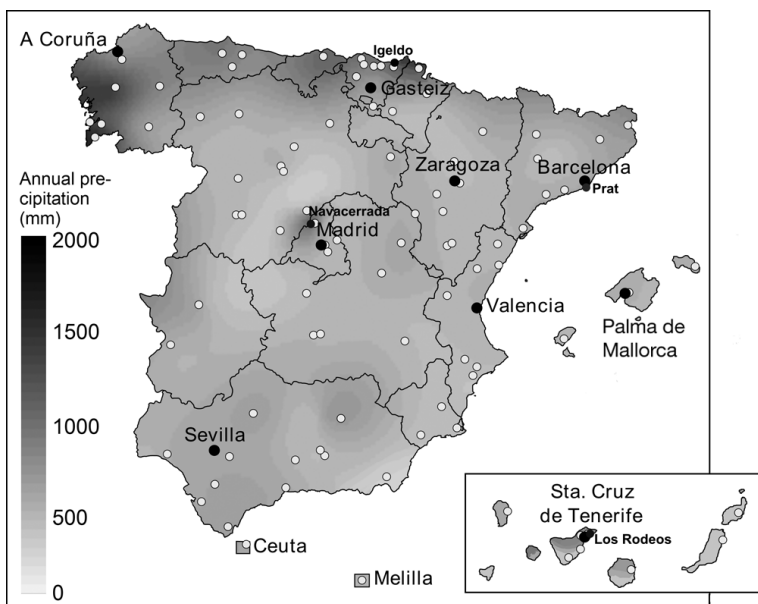

Fig. 1. Location of the meteorological stations used in this study and the annual precipitation for the period 1950–2000, interpolated with Thin-Plate spline method

mated by counting the records of daily rainfall of the station or a set of stations, and the cumulative probability $\pi(p \geq P)$ is estimated with the normalized sum of those frequencies. The return period T or expected time (in days) between 2 precipitation events with $p \geq P$ is given by the inverse of the cumulative probability:

$$T(P) \equiv \frac{1}{\pi(p \geq P)} \quad (1)$$

To fit the probability of daily precipitation (of a full year or a particular month), first we tested the commonly used models: GEV, Pareto and GPD, Gamma, Gumbel, Weibull, Exponential and Log-normal II and III (Bridges & Haan 1972, Öztürk 1981, Etoh et al. 1986, Begueria 2005). Second, other models were also tested, such as modified versions of GEV (Eq. 2) and Gumbel (Eq. 4).

$$\pi(\lambda_0 \geq \lambda) = e^{[-(1+\varepsilon\lambda)^{1/\varepsilon}]} \quad (2)$$

$$\pi(\lambda_0 \geq \lambda) = e^{-e^{(\lambda^w + k)}} \quad (3)$$

where ε, w, and k are parameters to be estimated and λ is the relative precipitation, which depends on a parameter P_0, 'the most probable precipitation', as well as a normalization parameter, P_1:

$$\lambda \equiv \frac{P - P_o}{P_1} \quad (4)$$

Therefore, the cumulative probability of the occurrence of precipitation $p \geq P$ satisfies the same probability distribution (Eq. 2), i.e. $\pi(p \geq P) = \pi(\lambda_0 \geq \lambda)$ where $\lambda_0 = (p - P_1)/P_1$. As an alternative, we propose another function of accumulated probability $\pi(\lambda)$:

$$\pi(\lambda_0 \geq \lambda) \approx \frac{1}{1 + \lambda^{w + \lambda^w e^{-k}}} \quad (5)$$

where w is a 'shape' parameter, k is the 'smoothness', and w, $k > 0$ (see Appendix 1), and λ is the relative precipitation ($\lambda > 0$, i.e. $P_0 \leq 0$ and $P_1 > 0$). It must be noted that we can make a change of variables as $r \equiv e^{-k}$, although we prefer to use k, because it is a parameter that varies more smoothly than r. Hence, 4 parameters for the return period (with Eqs. 1, 4 & 5) were used:

$$T \approx 1 + \left(\frac{P - P_o}{P_1}\right)^{w + e^{-k}\left(\frac{P - P_o}{P_1}\right)^w} \quad (6)$$

where P_1 is a 'scale' parameter, and P_0 is a 'location' parameter. P_0 represents the minimum value (return period = 1 d). P_1 is defined as an amplitude that indicates the difference between the expected precipitation in 2 d, P_2, and the minimum value P_0, i.e. $P_1 \equiv P_2 - P_0$. The inverse of Eq. (6) can be approxi-

mated to the Pareto distribution for short intervals (see Appendix 2).

In addition, from Eq. (6), it is possible to construct a series of the return period of each value of daily precipitation. This resulting series is henceforth called return-period series (RPS). This process is called total standardization, as opposed to partial standardization, which is defined as the process of generating an RPS with parameters extracted from a portion of the total length of the time series.

To estimate the parameters for all the functions, we used maximum likelihood inference, in particular, the Profile Log-Likelihood (PLL) approach (Raue et al. 2009). However, the quantity minimized into PLL was not exactly the square of the difference between the empirical probability (π) and the probability predicted by the model fit (π'). These were normalized by the empirical probability (both π and $1 - \pi$, in particular with $[1 - \pi]\pi$) to minimize the error at the tails of the probability distribution.

To compare the goodness-of-fit of the models, we used the measure of the Normalized Mean Absolute Error (NMAE) and the Akaike Information Criterion (AIC). The NMAE is a better index than the others for precipitation because it compares the relative errors for the lowest and highest precipitation. The model with the lowest AIC is that which best fits to the dataset (Akaike 1974, Burnham & Anderson 2002). In addition, to compare the probability distribution of daily precipitation predicted by the models and the observed precipitation distribution, 2 tests were employed: the Bootstrap Kolmogorov-Smirnov test (Marsaglia et al. 2003, Sekhon 2010) and the Anderson-Darling test (Scholz & Stephens 1987). As the precipitation data showed many ties, we used an estimation of p-value adjusted for ties in the Anderson-Darling test and bootstrap resampling in the Kolmogorov-Smirnov test (Sekhon 2010). Statistical analysis of the data was done by using the R-language (Muenchen & Hilbe 2010, R Development Core Team 2010).

3. RESULTS

First, the commonly used models (GEV, Pareto and GPD, Gamma, Weibull, Gumbel, Exponential Lognormal II and III) and the 3 proposed models (Modified GEV, Modified Gumbel, and alternative model) were fitted to the empirical distributions of the daily precipitation data for every month of the year and station. Fig. 2 shows the empirical and theoretical
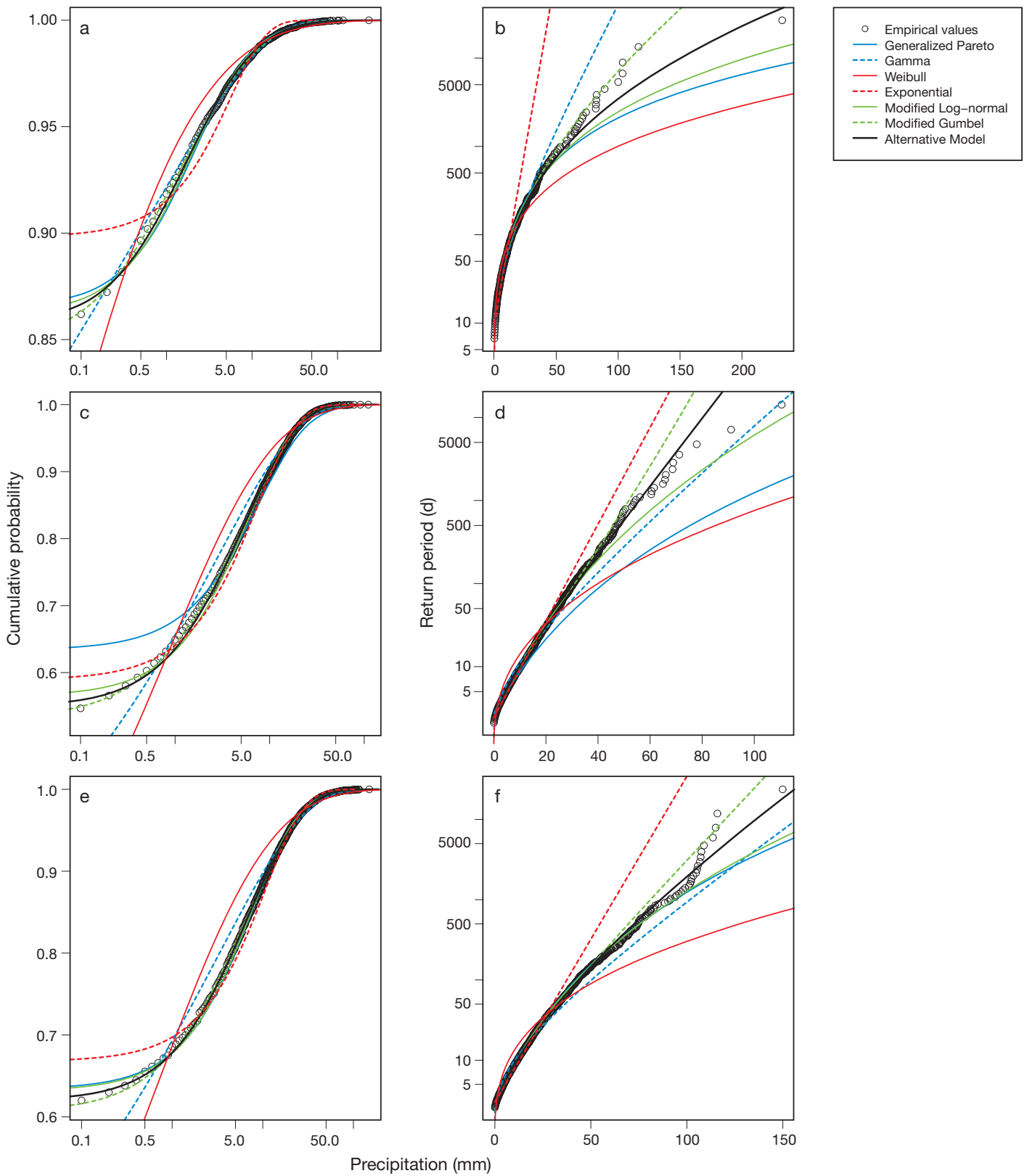
Fig. 2. Examples of cumulative probability and return period, according to the empirical values and several models fitted, for 3 stations with different climates: (a,b) Tenerife-Los Rodeos (subtropical climate); (c,d) A Coruña (oceanic climate); (e,f) Navacerrada (mountain climate)

distributions for stations corresponding to the 3 Spanish climate types: A Coruña (Galicia) has an oceanic climate, Tenerife-Los Rodeos (Canary Islands) a subtropical climate, and Navacerrada (Madrid) a mountain climate. In most cases, the modified Gumbel and alternative models fit the high and low precipitation better than the other models. On the other hand, the Log-normal III and Generalized Pareto distributions fit low precipitation better than high precipitation.

The NMAE was obtained by comparing observed and modelled precipitation for all stations and months. Application of the Gumbel, GEV, Lognormal II, and Pareto distributions produced a total NMAE of >1 for most of the stations, especially for the highest and lowest rainfall; therefore, these models were discarded. However, all the other models showed NMAE of <1 for all values of precipitation and for all stations, except for the Weibull model, which produced high NMAE for some stations (Fig. 3a). The modified Gumbel and alternative model produced the best NMAE (<0.10), while the other models, such as Log-normal III, GPD, and Modified GEV had an NMAE generally between 0.10 and 0.20 (Fig. 3a). The AIC indicated that the alternative model is often better (Fig. 3b). Fig. 4 shows the AIC for 3 stations (Igeldo, Valencia and Prat), which indicates that in most of the months, the alternative model fits better than the Gumbel model.

Finally, a comparison of the goodness-of-fit of the 2 best models (the modified Gumbel and alternative models) was undertaken using Kolmogorov-Smirnov and Anderson-Daring tests (Fig. 5). The comparison shows that in both models observed and predicted precipitation do not differ significantly; in some cases, the alternative model fit better than the modified Gumbel Model. The fit was worse for the summer months, partly due to the relatively few days that experience precipitation.

4. DISCUSSION

4.1. Advantages and limitations

This study has focused on the examination of the complete distribution of daily precipitation variation, instead of investigating the probability of a specific statistic (e.g. maximum) of daily precipitation in a month or a year (e.g. Bridges & Haan 1972, Etoh et al. 1986, Durman et al. 2001, Eslamian & Feizi 2007). Some authors have investigated wider ranges of probability of daily precipitation above a specified threshold (Wilson & Toumi 2005). However, the standard probability functions cannot accurately fit the entire range of precipitation, from the frequency of days without precipitation to the frequency of highest extremes. This is because usually the probability
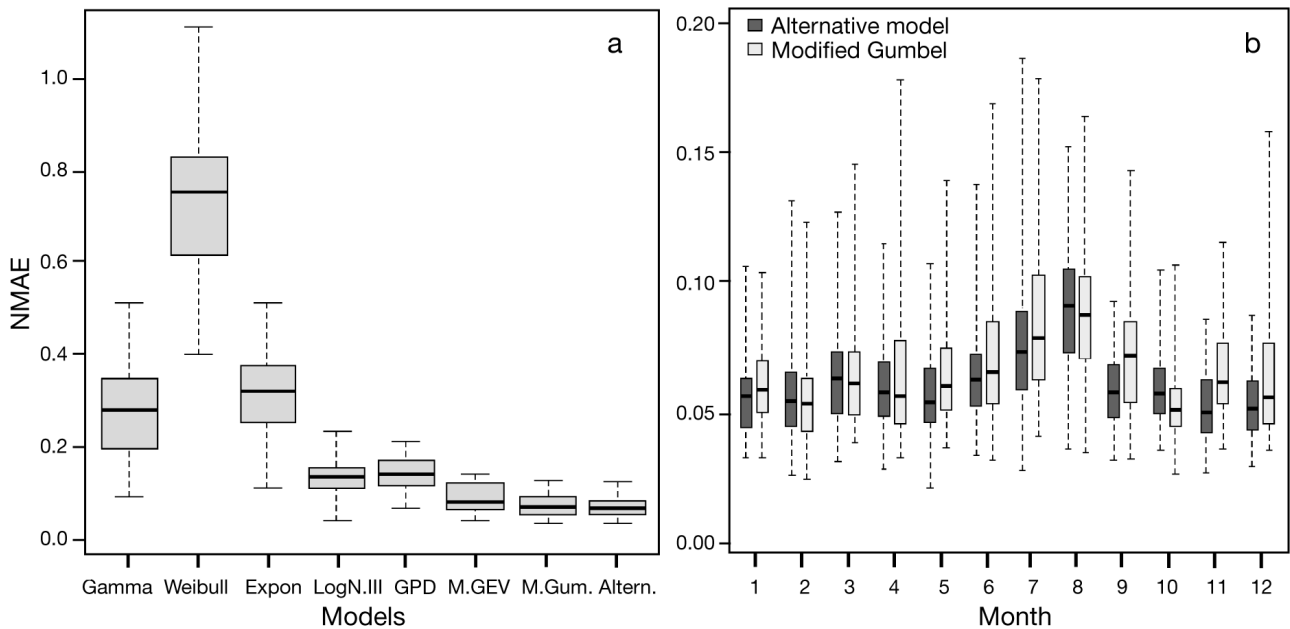

Fig. 3. Comparison of the Normalized Mean Absolute Error (NMAE) of the predicted precipitation versus the observed empirical distribution function, for the set of 108 stations, applied to (a) annual precipitation: Gamma, Weibull, Exponential, Lognormal III, Generalized Pareto Distribution (GPD), Modified Generalized Extreme Value (M. GEV), Modified Gumbel and alternative models; (b) monthly precipitation: modified Gumbel and alternative models. Horizontal line: median; box: 25th and 75th percentiles; whiskers: 5th and 95th percentiles
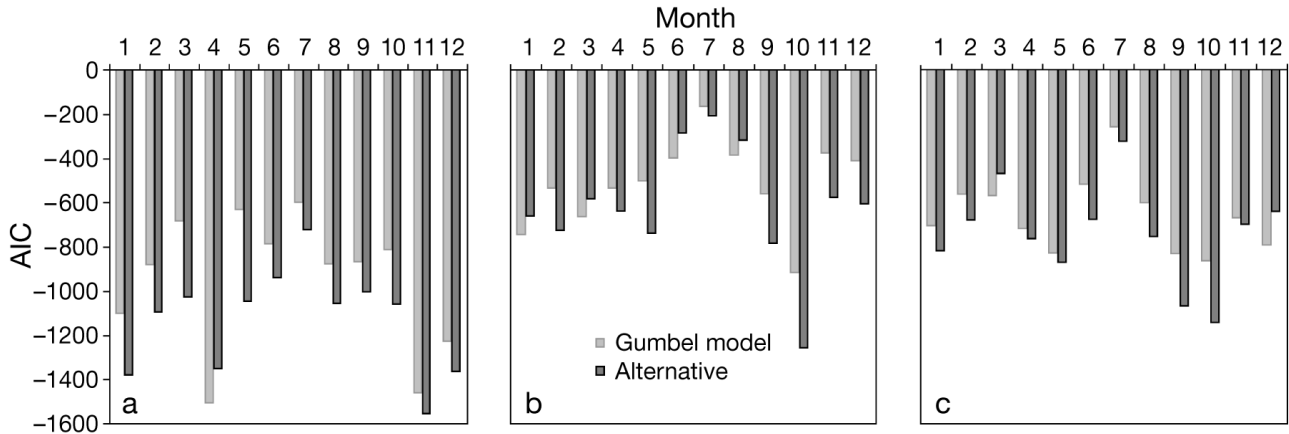
Fig. 4. Comparison of the monthly AIC of the fits to Modified Gumbel and alternative models of accumulated probability for 3 stations: (a) Igeldo (Gipuzkoa); (b) Valencia (Valencia); (c) Prat (Barcelona)
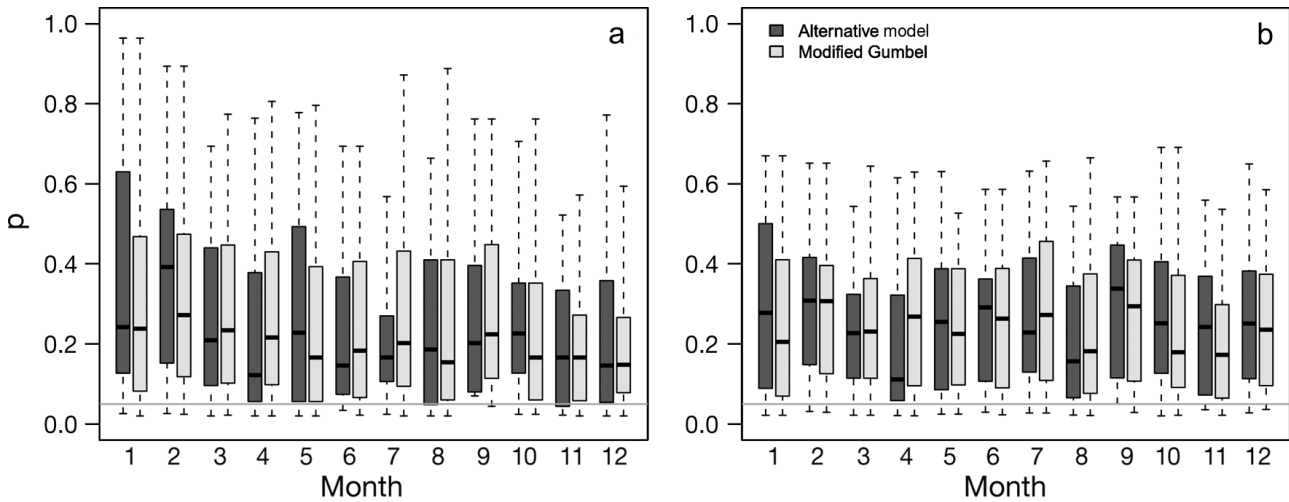

Fig. 5. The p-values corresponding to (a) Kolmogorov-Smirnov test with bootstrap resampling and (b) Anderson-Darling test with adjustment for ties, comparing the observed and predicted precipitation using modified Gumbel and alternative models. Grey horizontal line: threshold value of p = 0.05. Black horizontal line: median; box: 25th and 75th percentiles; whiskers: 5th and 95th percentiles
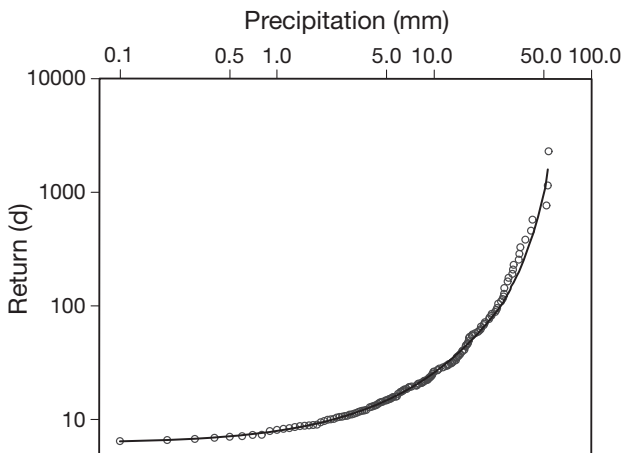

Fig. 6. Return period versus daily precipitation for the months of the year for the station of Prat (Barcelona): observed values (grey circles) and predicted values (black line)

function uses 2 or 3 parameters, which appears to be insufficient to fit the observed data (Hanson & Vogel 2008). The alternative model of probability developed in this work can accurately represent the entire range of probability of daily precipitation (e.g. Figs. 2 and 6) because it uses 4 parameters.

Fig. 6 is an example of the cumulative probability of monthly precipitation at the Prat station (Barcelona). The fitted line corresponds to the alternative model. The fit was made for all values of precipitation, except for the 2 highest and the 2 lowest values; thus, extreme values fit well even if they are excluded from the determination of the parameters (NMAE = 0.15). If we represent all standardized absolute errors throughout the year (Fig. 7), the normalized absolute error is similar for all scales of precipitation, except for the minimum measurable value (0.1 mm).
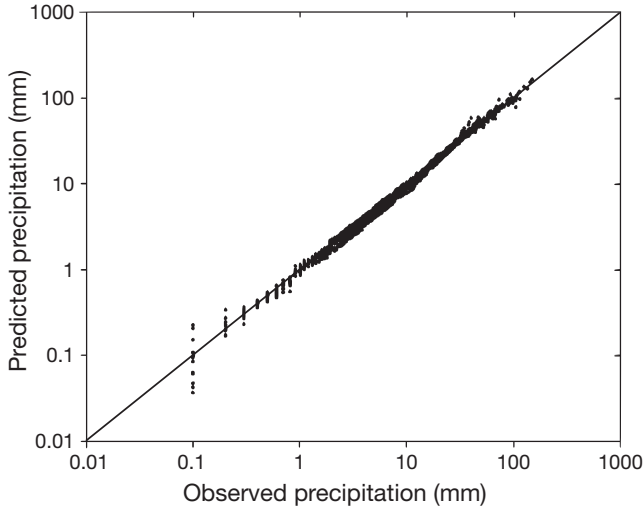
Fig. 7. Validation of the daily precipitation for the months of the year for the station of Prat (Barcelona)

However, the use of 4 parameters can lead to difficulties in estimation by computer algorithms, because it requires the initial values to be very close to the best estimates of the parameters. If not adequately initiated, the algorithms can produce poor fits. However, the great flexibility of the alternative model can cause significant differences in the fits to similar datasets. Therefore, there may be some interdependence between the parameters, which subtracts robustness from the interpretation or comparison between different settings. For example, in this work, a correlation between P_0 and P_1 is found, with $R^2 = 0.49$, and between w and $\log(P_1)$ with $R^2 = 0.60$, which indicate interdependence. In some cases, other models are superior to the proposed alternative model. This is true especially for dry conditions, for which the modified Gumbel sometimes provided better results (e.g. see Fig. 3b in August and October).

On the other hand, an important advantage of this methodology is that it can completely transform a series of daily rainfall into a series of probability: RPS. An RPS is a 'naked' time series, i.e. information about the mean and variability of daily precipitation is contained in the monthly parameters extracted. However, there is a percentage of variability, which is different between 2 RPS of equal length, because some *a priori* factors are random. This means that if 2 RPS are very long, then they tend to be equal in terms of their probability distribution, especially for similar climates. For example, an RPS obtained with partial standardization (long enough) will be similar to that obtained with total standardization, except for small random events related to the short length of the series. Another percentage of RPS variability may be due to climatic features; e.g. 2 sets of precipitation can have the same probability distribution, although they may differ in the prevailing duration of dry and wet periods and alternation between them.
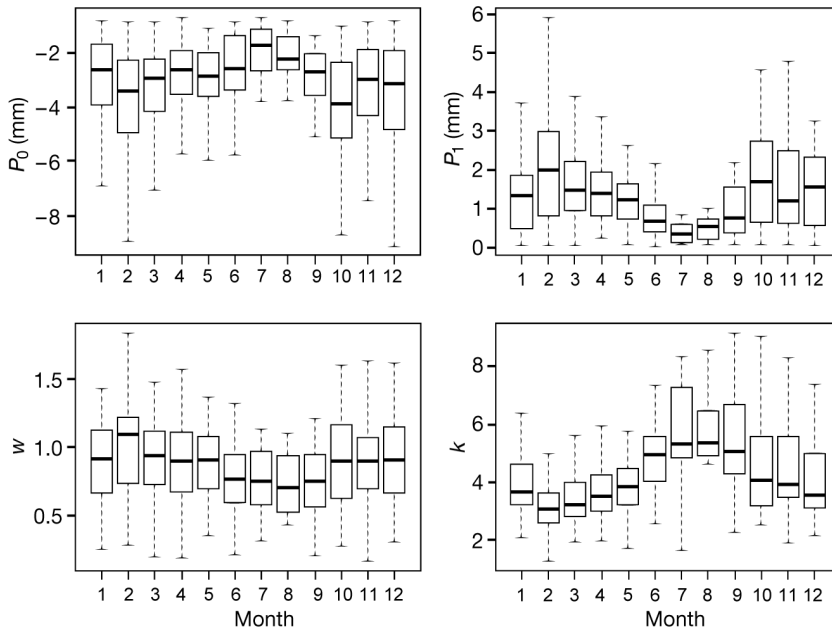
In addition, this model can be used to characterize and correct the probability distribution of a generic set of daily precipitation data. Its use could introduce improvements in the statistical downscaling of climate models (Widmann et al. 2003), and thus improvements in predicting changes in the precipitation regimes (Kidson & Thompson 1998, Perkins et al. 2007). The model could also lead to enhanced forecasting weather models and improved calibration of data obtained from weather radars (Marks et al. 2011). To capture the rainfall data containing zeros, some authors have proposed using a mixed distribution (Smith et al. 2011). However, the mixed distribution is not suitable, because it is not robust for climatic analysis in other applications, such as to correct precipitation distributions obtained from climate models.
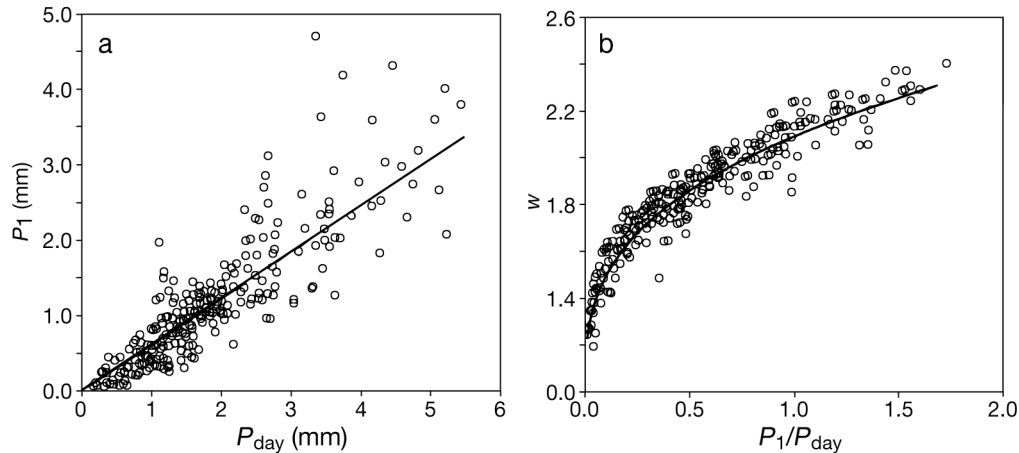

Fig. 8. Monthly values of 4 estimable parameters of the alternative model (P_0, P_1, w, and k) applied to 108 stations in Spain. Horizontal line: median; box: 25th and 75th percentiles; whiskers: 5th and 95th percentiles

Fig. 9. (a) Relationship between 'scale' parameter (P_1) of the annual fit and the daily average precipitation (P_{day}). (b) Relationship between 'form' parameter and the ratio P_1/P_{day}

4.2. Climatic relationships

Some of the 4 parameters estimated for the alternative model exhibit seasonal cycles. Fig. 8 shows this property, especially for the 'scale' (P_1) and 'smoothness' (k). Logically, the 'scale' of the probability distribution is lower and the 'smoothness' is higher for drier months. This means that the probability of low rainfall is highest for summer. However, the 'shape' parameter (w) is slightly lower in summer, which implies a higher probability of extreme rainfall. This is consistent with the characteristics of a Mediterranean climate.

Accordingly, a correlation between the monthly 'scale' (P_1) and the daily average precipitation (P_{day}) was found, with $R^2 = 0.63$ and slope of 0.6 ± 0.1 (Fig. 9a). When this ratio was extracted from the 'scale' parameter (P_1), we observed a higher correlation with the 'shape' (w). In particular, we found that w is approximately equal to $b \cdot (P_1/P_{day})^c$, where $b = 1.1 \pm 0.1$ and $c = 0.34 \pm 0.05$, with $R^2 = 0.90$ (Fig. 9b). This function relates the shape of the probability distribution and the ratio between the scale and average precipitation. In other words, the average precipitation can be simulated using the relation $P_{day} \sim P_1 \cdot (b/w)^{1/c}$, which is simpler than the theoretical expression of the average (see Appendix 1).

A new probability model should be applicable to various climates, not only to the Mediterranean. In this sense, the seasons in Spain may represent different climates; e.g. the summer of coastal Mediterranean climate is similar to the desert climate, spring is similar to oceanic climate, and the winter of the continental-Mediterranean climate could be similar to cold climates. Therefore, the fit of the alternative probability model for the different months of the year suggests a good performance for different climates. In fact, most of the stations studied in Spain, representing the ocean, mountain, and subtropical climates, were adequately modeled (Figs. 2 & 3). The alternative model works better for wet climates, because the empirical probability distribution shows a low random error, and thus, it appears to be a function with very regular derivatives.

5. CONCLUSIONS

Most of the widely used distributions for daily precipitation data (GEV, GPD, Gamma, Gumbel, Weibull, Exponential and Log-normal) were unable to precisely fit the entire range of daily precipitation, including the days without rain. It was therefore necessary to use models with 3 or 4 parameters, and we tested 3 models: a modified Gumbel and the GEV, and an alternative model. The alternative model with 4 parameters produced the best results for most of the stations, with a mean absolute error of <10% in most of the stations analyzed, and its performance was similar for all scales of precipitation. The fit was poor for summer months, due to the few days of precipitation. Thus, the alternative model might work best for wet climates. The model can be used to characterize the seasonal cycles of precipitation; hence, it can also be used to correct the probability distribution of a generic set of daily precipitation.

LITERATURE CITED

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19:716–723

Alves M, Verdière AC (1999) Instability dynamics of a subtropical jet applications to the Azores Front current system: eddy driven mean flow. J Phys Oceanogr 29: 837–864

Ashkar F, Mahdi S (2006) Fitting the log-logistic distribution by generalized moments. J Hydrol (Amst) 328:694–703

Begueria S (2005) Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value. J Hydrol (Amst) 303:215–230

Ben-Zvi A (2009) Rainfall intensity-duration-frequency relationships derived from large partial duration series. J Hydrol (Amst) 367:104–114

Bridges TC, Haan CT (1972) Reliability of precipitation probabilities estimated from the gamma distribution. Mon Weather Rev 100:607–611

Burnham KP, Anderson D (2002) Model selection and multimodel inference. Springer-Verlag, New York

Capel Molina JJ (2000) El clima de la península ibérica. Editorial Ariel, Barcelona

Cerveny RS, Lawrimore J, Edwards R, Landsea C (2007) Extreme weather records: compilation, adjudication and publication. Bull Am Meteorol Soc 88:853–860

Durman CF, Gregory JM, Hassell DC, Jones RG, Murphy JM (2001) A comparison of extreme European daily precipitation simulated by a global and a regional climate model for present and future climates. Q J R Meteorol Soc 127:1005–1015

Ebert EE (2001) Ability of a poor man's ensemble to predict the probability and distribution of precipitation. Mon Weather Rev 129:2461–2480

Eslamian SS, Feizi H (2007) Maximum monthly rainfall analysis using L-moments for an arid region in Isfahan Province, Iran. J Appl Met Clim 46:494–503

Etoh T, Murota A, Nakanishi M (1986) SQRT-exponential type distribution of maximum. In: Singh VP (Ed) Hydrologic frequency modeling. Reidel, Dordrecht, p 253–264

Granger RJ (1989) An examination of the concept of potential evaporation. J Hydrol (Amst) 111:9–19

Hanson LS, Vogel R (2008) The probability distribution of daily rainfall in the United States. Proc World Environmental and Water Resources Congress 2008, Honolulu, HI: 1–10 doi 10.1061/40976(316)585

Hardy M (2010) Pareto's Law. Math Intel 32:38–43

Jeevananda Reddy S (1995) Sensitivity of some potential evapotranspiration estimation methods to climate change. Agric For Meteorol 77:121–125

Kharin VV, Zwiers FW (2000) Changes in the extremes in an ensemble of transient climate simulations with a coupled

atmosphere–ocean GCM. J Clim 13:3760–3788

Kidson JW, Thompson CS (1998) A comparison of statistical and model-based downscaling techniques for estimating local climate variations. J Clim 11:735–753

Kysely J (2002) Comparison of extremes in GCM-simulated, downscaled and observed central European temperature series. Clim Res 20:211–222

Lana X, Martínez MD, Serra C, Burgueño A (2004) Spatial and temporal variability of the daily rainfall regime in Catalonia (northeastern Spain), 1950–2000. Int J Climatol 24:613–641

Marks DA, Wolff DB, Carey LD, Tokay A (2011) Quality control and calibration of the dual-polarization radar at Kwajalein, RMI. J Atmos Oceanic Technol 28:181–196

Marsaglia G, Tsang WW and Wang J (2003) Evaluating Kolmogorov's distribution. J Stat Softw 8:1–4

Martín-Vide J (2004) Spatial distribution of a daily precipitation concentration index in peninsular Spain. Int J Clim 24:959–971

McGuinness JL, Leslie HP (1972) Maximum potential evapotranspiration frequency. J Irrig Drain Div 2:207–214

McLaughlin MP (2001) A compendium of common probability distributions. Available at: www.causascientia.org/math_stat/Dists/Compendium.pdf

Muenchen RA, Hilbe JM (2010) R for Stata users. Statistics and computing, Springer, New York

Öztürk A (1981) On the study of a probability distribution for precipitation totals. J Appl Met 20:1499–1505

Park JS, Jung HS (2002) Modelling Korean extreme rainfall using a Kappa distribution and maximum likelihood estimate. Theor Appl Climatol 72:55–64

Peliz A, Rosa TL, Santos AMP, Pissarra JL (2002) Fronts, jets, and counterflows in the western Iberian upwelling system. J Mar Syst 35:61–77

Perkins SE, Pitman AJ, Holbrook NJ, McAneney J (2007) Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. J Clim 20:4356–4376

R Development Core Team (2010) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25:1923–1929

Scholz FW, Stephens MA (1987) K-sample Anderson-Darling tests. J Am Stat Assoc 82:918–924

Sekhon JS (2010). Matching: multivariate and propensity score matching with balance optimization. R package version 4.7-11. Available at: cran.r-project.org/package= Matching.

Smith JA, Villarini G, Baeck ML (2011) Mixture distributions and the hydroclimatology of extreme rainfall and flooding in the eastern US. J. Hydrometeor 12:294–309

Widmann M, Bretherton CS, Salathé EP (2003) Statistical precipitation downscaling over the northwestern United States using numerically simulated precipitation as a predictor. J Clim 16:799–816

Wilson PS, Toumi R (2005) A fundamental probability distribution for heavy rainfall. Geophys Res Lett 32:L14812 doi:10.1029/2005GL022465

Appendix 1. Probability density function

From Eq. (4), we can deduce that the probability density function, $f(\lambda; w, k)$, is given by:

$$f(\lambda;w,k) \equiv \frac{d}{d\lambda}\pi(\lambda;w,k) = \frac{\lambda^{w+\lambda^w e^{-k}-1}}{\left(1+\lambda^{w+\lambda^w e^{-k}}\right)^2}\left[w+(1+w\ln\lambda)\,\lambda^w e^{-k}\right] \quad \text{(A1)}$$

where w is the shape parameter and k is a smoothness parameter, while λ is the relative precipitation defined in Eq. (3). In Figs. A1 & A2, the parameter k is only related to the probability of medium and high values of λ, while w is related to the shape of all curves of the probability density. Parameters w and k and the variable λ are defined as always positive.

Some statistical characteristics are described as follows. For high values of w, the form of probability density function is similar to the Gaussian curve (see Fig. A1), with mean $\mu \approx 1$. This is because, if $k \gg 1$, Eq. (A1) corresponds approximately to the log-logistic distribution (Ashkar & Mahdi 2006). Therefore, the mean will be similar; in particular:

$$\text{mean}(\lambda) \approx \left(\frac{\frac{\pi}{w}}{\sin\left(\frac{\pi}{w}\right)}\right)^{g(w,k)} \quad \text{(A2)}$$

where $g(w, k)$ is a smooth function that tends to 1 for high values of w and k. The mean of λ is approximately $(\pi/w)^g$ for low values of w. On the other hand, the derivative of Eq. (A1) shows that the mode is a nonlinear function of w and k; however, it is always between 0 and 1, and the influence of k is very low. For this reason, the mode can be approximated by the log-logistic mode (McLaughlin 2001):

$$\text{mod}(\lambda) \approx \begin{cases} 0 & si \quad w < 1 \\ \left(\dfrac{w-1}{w+1}\right)^{\frac{1}{w}} & si \quad w \geq 1 \end{cases} \quad \text{(A3)}$$

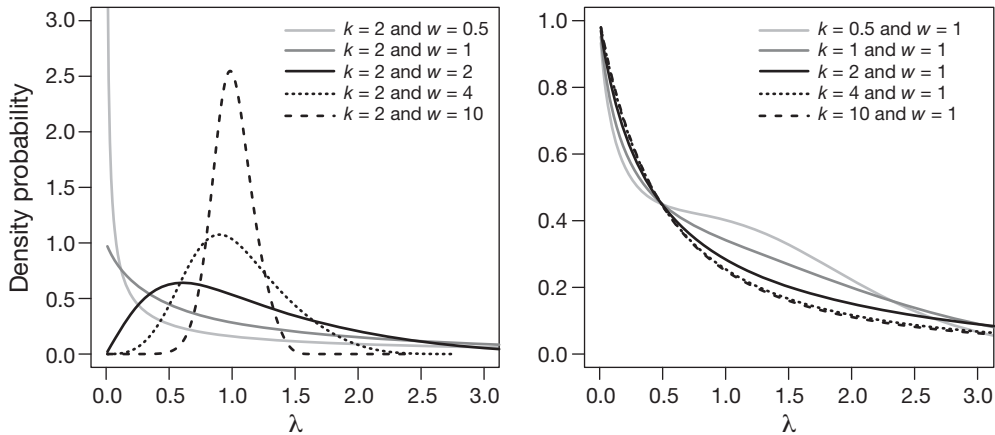where $\text{mod}(\lambda)$ is the mode of λ.

Fig. A1. Examples of the probability density curve, with $k = 2$ and $w = 0.5$ to 10, and with $w = 1$ and $k = 0.5$ to 10
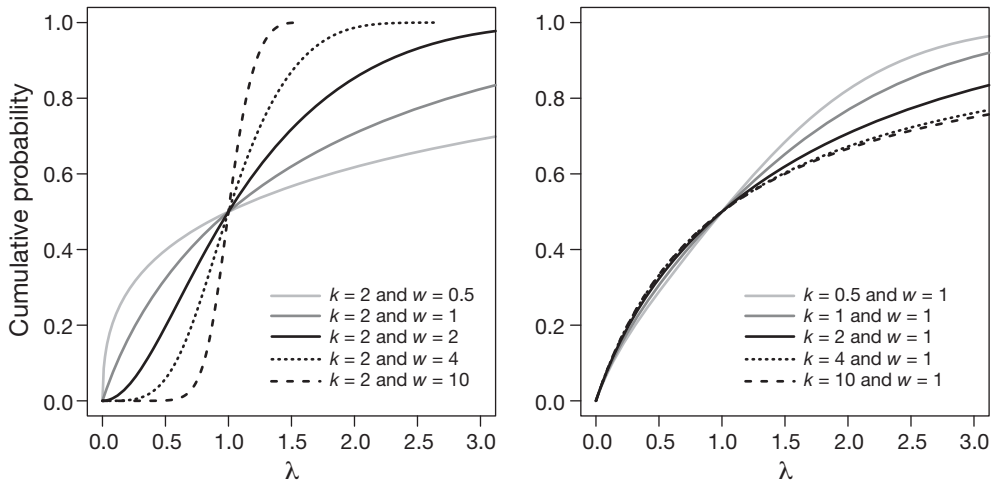

Fig. A2. Examples of the cumulative probability curve, with $k = 2$ and $w = 0.5$ to 10, and with $w = 1$ and $k = 0.5$ to 10

Appendix 2. Inverse function

Eq. (4) has no explicit inverse, and therefore, it is necessary to use programmed algorithms or find an approximate inverse function:

$$\frac{P - P_o}{P_1} = (T-1)^{m(T-1)} \tag{A4}$$

where $m(T-1)$ is an adjustable function of $(T-1)$, where T is the return period. For short intervals of periods, Eq. (A4) can also be approximated by the Pareto distribution (Hardy, 2010), according to:

$$\frac{P}{P_1} \approx T^{m_o} \tag{A5}$$

where P_0 and m_0 are the estimable parameters. For a particular rainfall event, we can consider the duration as an expected time for a given rainfall accumulation in the context of the event. Thus, it is possible to approximate the temporal distribution of maximum accumulation to another power distribution:

$$\frac{P}{P_1} \approx t^{h_o} \tag{A6}$$

where t is the duration, and P_1 and h_0 are the estimable parameters. By adding the expressions (A5) and (A6), an Intensity-Duration-Frequency (IDF) curve is obtained (Ben-Zvi 2009):

$$P(T, t) \approx P_1(T_1, t_1)\left(\frac{T}{T_1}\right)^{m_o}\left(\frac{t}{t_1}\right)^{h_o} \tag{A7}$$

where $P(T, t)$ is the expected rainfall accumulation for a T-return and t-duration, $P_0(T_0, t_0)$ is the expected rainfall accumulation for a T_0-return and t_0-duration, and m_0 and h_0 are the estimable parameters.