



The Markovian arrival process: A statistical model for daily precipitation amounts



Pepa Ramírez-Cobo^{a,*}, Xavier Marzo^b, Alba V. Olivares-Nadal^c, José Álvarez Francoso^b, Emilio Carrizosa^c, M. Fernanda Pita^b

^a Department of Statistics and Operations Research, University of Cádiz, Spain

^b Department of Physical Geography, University of Sevilla, Spain

^c IMUS, Institute of Mathematics, University of Sevilla, Spain

ARTICLE INFO

Article history:

Received 22 June 2013

Received in revised form 16 December 2013

Accepted 21 December 2013

Available online 7 January 2014

This manuscript was handled by Andras

Bardossy, Editor-in-Chief, with the

assistance of Peter F. Rasmussen, Associate

Editor

Keywords:

Daily precipitation data

Markovian arrival process

Hidden Markov models

Correlation

Moment matching method

ABSTRACT

The Markovian arrival process (*MAP*) is a stochastic process that allows for modeling dependent and non-exponentially distributed observations. Due to its versatility, it has been widely applied in different contexts, from reliability to teletraffic. In this work we show the suitability of the *MAP* for modeling daily precipitation data, which are often characterized by a non-negligible correlation structure. Specifically, a set of daily precipitation amounts series from the region of Andalusia (Spain) is shown to be correctly fitted with a two-state *MAP*.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The fit of climatological series by theoretical models is an essential task in order to synthetically describe the behavior of the data, obtain estimates and predictions given the observed values or simulate series under the same statistical pattern. In rainfall series, usually characterized by a large variability (especially in Mediterranean climates) and a significative empirical autocorrelation function, there is a need of finding suitable models that correctly capture the data behavior. Most works dealing with such a problem have considered the modeling of annual or monthly totals, where the normal distribution in the first case, and several asymmetrical distributions (as the lognormal, or incomplete gamma distributions) in the second case, have been suggested (Lana and Burgueño, 2000; Husak et al., 2007; Martínez et al., 2010).

The modeling of daily precipitation amounts, or daily intensity, is still an open problem, even if recently there has been an increasing interest in the problem of fitting daily rainfall data, mainly due

to their social, environmental and economic impacts (Mekis and Hogg, 1999; Romero et al., 1998, 1999; Wilks, 1999). These impacts are especially intense in the Mediterranean climates since annual and monthly precipitations in such climates are generated by a small number of rainy days. This results in a high daily rainfall irregularity, with alternating high and low values, and a huge number of days without precipitation. These particular features are not so noteworthy in other climatic areas (especially oceanic areas), in which monthly and annual rainfall are the product of an abundant number of rainy days with moderate and regular rainfall totals. Such an extreme behavior in Mediterranean climates is responsible for two phenomena with high socio-economic and environmental impacts: floods and strong water erosion, as a consequence of heavy rains, and droughts, caused by long dry periods, which considerably affect agriculture and water supply. Long dry periods cause a non-negligible persistence in the series, which needs to be considered in the modeling problems.

Due to the strong asymmetry of the empirical histograms of daily rainfall amounts, some generalization of the exponential distribution as the Gamma or Mixed Exponential distributions have been considered in the literature to capture the skewness and possible extreme behavior of the data, see for example Wilks (1998, 1999), Gronewold et al. (2012). For the modeling of daily rainfall

* Corresponding author. Address: Departamento de Estadística e I.O., Universidad de Cádiz, Campus Universitario de la Asunción. Avenida de la Universidad s/n 11405 Jerez de la Frontera, Cádiz, Spain. Tel.: +34 956 037 890.

E-mail address: pepa.ramirez@uca.es (P. Ramírez-Cobo).

extremes, probability distributions with tails longer than the exponential model have been applied. For example, the Weibull and Gumbel distributions have been suggested to capture the extreme behavior of daily rainfall series in Catalonia and Valencia, the Spanish regions with a most extreme behavior, see Peñarrocha et al. (2002) and Casas et al. (2007). It is worth pointing out that if the probability distributions previously referred (Gamma, Mixed Exponential, Weibull or Gumbel) are not combined with specific patterns of temporal or spatial autocorrelation (as in Casas et al. (2007)), then the consecutive daily amounts will be considered as independent observations (see Wilks (1999)), an assumption that may be unrealistic in some cases.

Apart from the modeling of daily precipitation intensity, it may also exist an interest in capturing the precipitation occurrence behavior (rainfall frequency, distribution of wet and dry spells, and transition probabilities). To this end, different statistical models that take into account the persistence factor have been considered in the literature. The Besson persistence coefficient (Besson, 1924) can be considered as a pioneering index and it has been one of the most widely used throughout the twentieth century (Brooks and Carruthers, 1953; Holawe and Dutter, 1999). Markov chains of different orders have been traditionally used to describe sequences of dry and wet days (Feyerherm et al., 1967; Gates and Tong, 1976; Lana and Burgueño, 1998; Martín-Vide and Gómez, 1999), to deal with periodicity (Kottogoda et al., 2004, 2008), or to compute the probability of precipitation and transition probabilities (Hosseini et al., 2011). The duration of dry sequences has also been estimated from a Weibull model (Lana et al., 2008), or by applying the generalized extreme value (GEV) and generalized Pareto (GP) distributions (Lana et al., 2006).

Some works consider the combination of two models, one for precipitation intensity and the other for modeling the precipitation occurrence. As examples, Lennartsson et al. (2008) combines Markov and Gaussian processes, and Wilks (1998) suggests a second order Markov chain and a mixture of exponentials for the occurrence and intensity precipitation models, respectively. On the other hand, Stern and Coe (1984) consider non-stationary Markov chains to model the occurrence of daily rain and the Gamma distribution to fit the rainfall amounts. Wang and Nathan (2007) merges a Markov chain with a Gamma distribution taking into account both daily and monthly statistical properties to generate daily rainfall data. Some other authors as Mehrotra and Sharma (2007a,b), Srikanthan and Pegram (2009), not only have combined Markov chain to model occurrence with various procedures to deal with intensity, but have also developed some techniques to simulate spatially correlated data simultaneously.

Hidden Markov processes have recently become very popular in the modeling of real data of different contexts, from medicine to engineering or signal processing. Also, in climatology and particularly in rainfall data, they have played an important role (Hughes and Guttorp, 1999; Bellone et al., 2000; Ailliot et al., 2009). In this paper we consider the fit of daily precipitation amounts by a Markovian arrival process (MAP), a class of hidden Markov process which, unlike previous approaches, not only takes into account the non-exponentiality and high variability of the observed series, but also their dependence structure. MAPs were introduced by Neuts (1979) as a wide class of versatile stochastic processes that generalizes the Poisson process and the exponential distribution. The MAP combines a versatile probability distribution, the phase-type (PH) model, with a specific autocorrelation function in such a way that it is able to capture a wide range of statistical marginal patterns under the assumption of a correlation structure. Phase-type distributions (O’Cinneide, 1990; Asmussen et al., 1996; Asmussen, 2000) constitute a large class of probability distributions that include well known models as the Exponential, Erlang, Coxian, Hyper- and Hypo-exponential distributions. MAPs have

proven long ago as suitable models in various contexts where dependent data are observed, such as teletraffic, reliability or queueing (Kang et al., 2002; Altman et al., 2005; Neuts et al., 2000; Montoro-Cazorla et al., 2009), but up to our knowledge they have never been considered in climatology.

The purpose of this paper is to show the suitability of the simplest version of MAP, the second order MAP, noted MAP_2 , to model daily precipitation intensity. An estimation approach based on a moments matching method is derived and applied to fit the MAP_2 to a set of real daily precipitation series in Andalusia, Spain. The fitting process, though technically sophisticated, is fully automated thanks to the numerical routines developed, which are available for practitioners. An extensive numerical study is developed to compare the MAP_2 with alternative models for daily precipitation intensity.

The paper is organized as follows. In Section 2 the MAP_2 is introduced and its main properties are reviewed. In Section 3, we derive a scheme for inference for the MAP_2 , based on the empirical moments of the observed data. Illustrations to the modeling of real rainfall data are given in Section 4, where comparisons with benchmark models are also provided. Finally, in Section 5, some conclusions and possible extensions to this work are discussed.

2. The second order Markovian arrival process

The MAP is a stochastic process that generalizes the well-known Poisson model. It is characterized by an unobservable, irreducible and continuous Markov process $\{J(t), t > 0\}$, with state space S and initial probability α , and a sequence of observable continuous and non-negative random variables $(A_1, A_2, \dots, A_n, \dots)$, which are conditionally independent given the realizations of the underlying Markov process $\{J(t), t > 0\}$. The Markov process $\{J(t), t > 0\}$ behaves jumping from one state to another according to a rate matrix D . The transitions remain unobserved but determine the value of the observed random variables $\{A_n\}_{n \geq 1}$. In this sense, the MAP can be regarded as a hidden Markov model, Ephraim and Merhav (2002). In the context of daily rainfall modeling, we will interpret the variable A_n as the total precipitation at the n th day.

In this paper we will restrict ourselves to the simplest version of MAP, namely, the second-order (or two-state) MAP, which is governed by $\{J(t), t > 0\}$ with state space $S = \{1, 2\}$. As will be shown in our empirical study this version of the process, noted MAP_2 , is already flexible enough to accurately fit rainfall data. A description of the basic properties of the MAP_2 follows. For a more detailed description, see Neuts (1979), Lucantoni (1991), Asmussen (2000) or Ramírez-Cobo et al. (2010).

Bodrog et al. (2008) provide a compact representation for the MAP_2 . In particular, for most MAP_2 s with positive lag-one autocorrelation (correlation between consecutive observations), as is the case of daily rainfall data, such MAP_2 s are characterized in terms of two matrices D_0, D_1 , which depend on 4 parameters, x, u, y, v :

$$D_0 = \begin{pmatrix} x & y \\ 0 & u \end{pmatrix}, \quad D_1 = \begin{pmatrix} -x - y & 0 \\ v & -u - v \end{pmatrix}. \quad (1)$$

These 4 parameters satisfy the following constraints:

$$\begin{aligned} x, u &\leq 0 \\ y, v &\geq 0 \\ x + y &\leq 0 \\ u + v &\leq 0. \end{aligned} \quad (2)$$

Matrices D_0 and D_1 in (1) are rate matrices that satisfy $D = D_0 + D_1$, where D is the infinitesimal generator of the underlying Markov process $\{J(t), t > 0\}$ whose stationary probability vector is denoted by $\pi = (\pi, 1 - \pi)$ and computed as $\pi D = \mathbf{0}$.

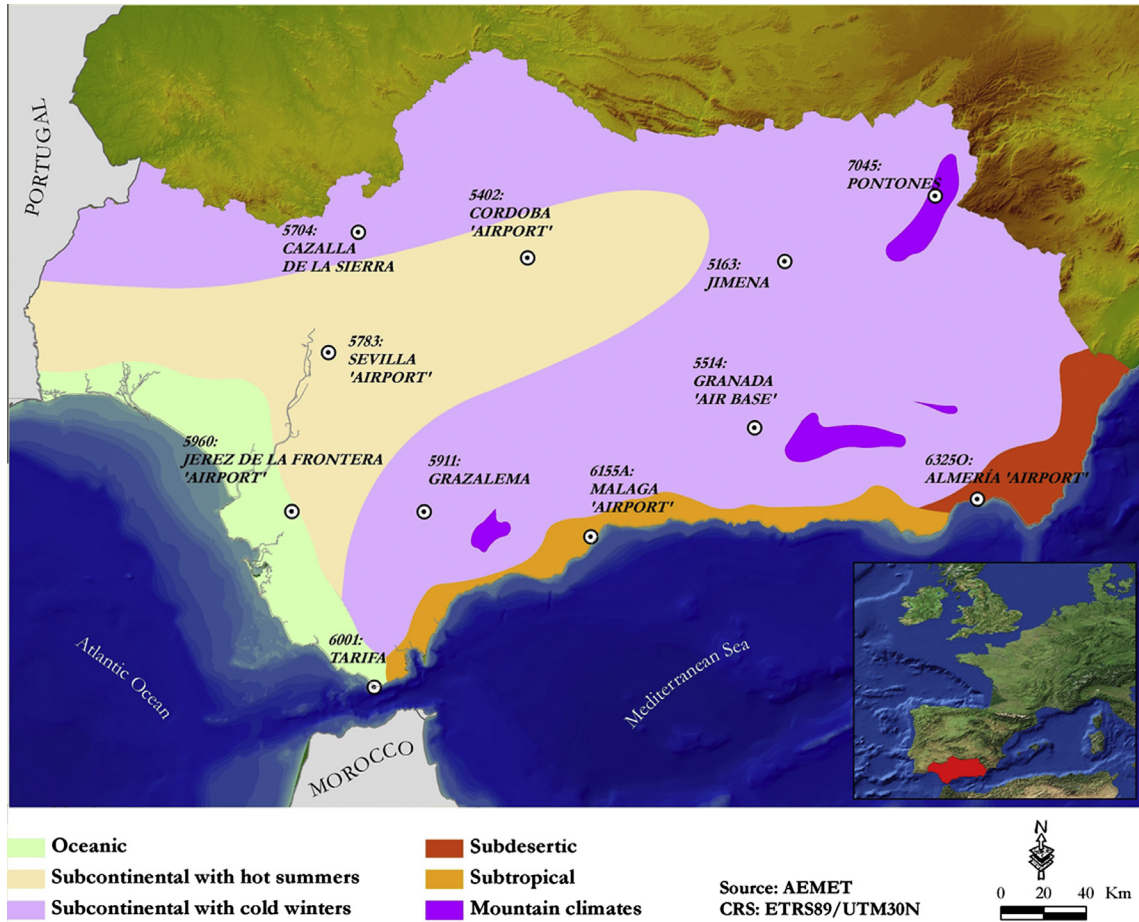


Fig. 1. Types of Mediterranean climates in Andalusia and location of meteorological stations.

An important property of MAPs concerns Markov renewal theory (Çinlar, 1975). If Δ_n denotes the total precipitation at the n th day, and X_n is the (unobserved) state of the underlying Markov chain at that moment, then $\{X_{n-1}, \Delta_n\}_{n=1}^\infty$ is a Markov renewal process, and in particular, $\{X_n\}_{n=1}^\infty$ is a Markov chain whose transition matrix P^* is given by

$$P^* = (-D_0)^{-1}D_1, \tag{3}$$

with stationary probability vector denoted by $\phi = (\phi, 1 - \phi)$ such that $\phi P^* = \phi$.

We next describe properties of the random variables of interest, $\{\Delta_n\}_{n \geq 1}$. For inference purposes we will assume that the MAP_2 is in its stationary version, that is, the initial probability vector equals the stationary distribution, $\alpha = \phi$. In the stationary version of the process, $\{\Delta_n\}_{n \geq 1}$ are identically distributed and follow a phase-type (PH) distribution with representation $\{\phi, D_0\}$. This implies

that $\Delta_n \stackrel{d}{=} \Delta$, for all $n \geq 1$, and Δ possesses the cumulative distribution function given by

$$F_\Delta(t) = P(\Delta < t) = \phi(I - e^{D_0 t})P^* \mathbf{e}, \tag{4}$$

$$= 1 - \phi e^{D_0 t} \mathbf{e}, \text{ for } t \geq 0,$$

where \mathbf{e} is a unit vector. Expression (4) can be understood as the probability of a total rainfall less than or equal to t at a given day. The density function of Δ is obtained as

$$f_\Delta(t) = \phi e^{D_0 t} (-D_0) \mathbf{e},$$

and the moments of Δ can be computed as

$$\mu_n = E(\Delta^n) = n! \phi (-D_0)^{-n} \mathbf{e}. \tag{5}$$

The variance σ^2 and the coefficient of asymmetry γ of Δ follow from (5).

Table 1
Observatories and percentage of missing data at each season.

Observatory code	Name	Winter	Spring	Summer	Autumn
63250	Almería, Airport	0	0	0	0
5704	Cazalla de la Sierra	4.6	4.2	1.9	4.6
5402	Córdoba, Airport	10.9	9.5	3.3	9.5
5514	Granada, Air Base	0	0	0	0
5911	Grazalema	0	0.7	0.3	0.1
5960	Jerez de la Frontera, Airport	0.3	0	0.3	0.2
5163	Jimena	2.5	2.7	1.7	4.6
6155A	Málaga, Airport	0	0	0	0
7045	Pontones	10.1	7.5	2.5	9.5
5783	Sevilla Airport	0	0	0	0
6001	Tarifa	4.4	0.5	0.6	4.2

A crucial property of the MAP_2 for statistical modeling is that it allows for correlated precipitation amounts, the autocorrelation coefficient of lag k , ρ_k , being given by

$$\rho_k = \rho(A_n, A_{n+k}) = \gamma^k \frac{\mu_2 - \mu_1^2}{\mu_2 - \mu_1^2}, \quad \text{for } k > 0, \quad (6)$$

where $0 \leq \gamma < 1$ is one of the two eigenvalues of the transition matrix P^* . As P^* is stochastic, then necessarily the other eigenvalue is equal to 1 (Bodrog et al., 2008).

3. Parameter estimation

Bodrog et al. (2008) prove that the first autocorrelation coefficient and the first three moments $\{\rho(1), \mu_1, \mu_2, \mu_3\}$ completely characterize the process MAP_2 . We have already given formulae for such values in Section 2, which depend on the parameters x, u, y, v in (1) and they must satisfy (2).

Therefore, a natural estimation procedure is based on the method of moments, in which values of the parameters x, u, y, v are sought so that the resulting values of $\mu_1, \mu_2, \mu_3, \rho(1)$ exactly fit

$\bar{\mu}_1, \bar{\mu}_2, \bar{\mu}_3$ and $\bar{\rho}(1)$, the empirical moments and lag-one autocorrelation coefficient. In other words, a first estimation attempt is to find values of x, u, y, v satisfying (2) and the following equalities:

$$\mu_1 = \bar{\mu}_1, \quad \mu_2 = \bar{\mu}_2, \quad \mu_3 = \bar{\mu}_3, \quad \rho(1) = \bar{\rho}_1, \quad (7)$$

where μ_i , for $i = 1, 2, 3$ and $\rho(1)$ are expressed in terms of the model parameters as

$$\begin{aligned} \mu_1 &= \frac{-v-y}{uy+yv+vx}, \\ \mu_2 &= 2 \frac{vu-yv+yx}{xu(uy+yv+vx)}, \\ \mu_3 &= -6 \frac{vu^2-yvu-yvx+yx^2}{x^2u^2(uy+yv+vx)}, \\ \rho(1) &= \frac{yv(u^3x+2vu^2x+v^2ux-3xu^2y-5xuyv-3x^2uv+u^3y+x^3v-v^2x^2+2yx^2v-2yv^2x)}{xu(2vu^2y+2v^2uy+v^2ux-2y^2vu-2y^2v^2-2yv^2x+y^2xu+2y^2xv+2yx^2v-2xuyv)} \\ &\quad + \frac{yv(-2y^2vu-y^2u^2+v^2uy+2vu^2y+y^2xu+2yx^2u+y^2xv+x^3u-2x^2u^2-y^2v^2)}{xu(2vu^2y+2v^2uy+v^2ux-2y^2vu-2y^2v^2-2yv^2x+y^2xu+2y^2xv+2yx^2v-2xuyv)}. \end{aligned}$$

Solving directly the nonlinear system of Eq. (7) is not straightforward. Instead, we propose to address (7) via solving numerically the following optimization problem:

Table 2
Columns 3–8: Empirical and estimated first moment, variance, coefficient of asymmetry and first three autocorrelation parameters by the MAP_2 , the Gamma and Mixed Exponential, for daily precipitation (winter) Andalusian series. Last column: obtained p -Values from the χ^2 -goodness-of-fit tests.

Station		μ_1	σ^2	γ	$\rho(1)$	$\rho(2)$	$\rho(3)$	p -Value
Almería	Empirical	0.76	13.75	10.54	0.09	0.03	0.01	
	MAP_2	0.76	13.75	10.54	0.09	0.01	0.00	0.07
	Gamma	0.76	13.75	9.75				0.48
	ME	0.57	6.68	5.32				$p < 0.01$
Cazalla	Empirical	3.25	92.02	4.40	0.38	0.18	0.15	
	MAP_2	3.25	91.72	4.43	0.38	0.32	0.28	0.07
	Gamma	3.25	92.02	5.89				$p < 0.01$
	ME	10.58	71.45	4.17				$p < 0.01$
Córdoba	Empirical	2.17	38.17	4.83	0.28	0.22	0.12	
	MAP_2	2.15	38.90	4.64	0.28	0.18	0.12	0.67
	Gamma	2.17	38.17	5.68				0.05
	ME	4.71	28.84	4.00				$p < 0.01$
Granada	Empirical	1.38	13.63	4.23	0.32	0.15	0.10	
	MAP_2	1.38	13.63	4.23	0.32	0.23	0.17	0.22
	Gamma	1.38	13.63	5.33				$p < 0.01$
	ME	1.91	11.65	3.99				$p < 0.01$
Grazalema	Empirical	8.88	556.15	4.48	0.44	0.25	0.16	
	MAP_2	8.80	566.85	4.30	0.42	0.42	0.42	0.69
	Gamma	8.88	556.15	5.30				0.02
	ME	78.94	399.06	3.69				$p < 0.01$
Jerez	Empirical	2.45	48.53	4.32	0.26	0.19	0.13	
	MAP_2	2.45	48.51	4.32	0.26	0.16	0.09	0.55
	Gamma	2.45	48.53	5.68				0.28
	ME	6.01	36.29	3.98				$p < 0.01$
Jimena	Empirical	1.66	22.02	5.11	0.26	0.13	0.11	
	MAP_2	1.62	23.15	4.59	0.26	0.15	0.09	0.06
	Gamma	1.66	22.03	5.64				$p < 0.01$
	ME	2.76	16.50	3.96				$p < 0.01$
Málaga	Empirical	2.47	85.25	6.83	0.23	0.12	0.14	
	MAP_2	2.47	85.23	6.83	0.23	0.12	0.06	$p < 0.01$
	Gamma	2.47	85.25	7.44				0.13
	ME	6.14	46.71	4.40				$p < 0.01$
Pontones	Empirical	3.44	77.71	4.76	0.36	0.21	0.15	
	MAP_2	3.47	76.46	4.91	0.37	0.36	0.34	0.02
	Gamma	3.44	77.71	5.12				0.20
	ME	11.83	44.80	3.81				0.02
Sevilla	Empirical	2.17	42.71	4.49	0.29	0.18	0.14	
	MAP_2	2.17	42.56	4.53	0.29	0.19	0.13	0.06
	Gamma	2.17	42.71	6.01				$p < 0.01$
	ME	4.71	31.77	4.17				$p < 0.01$
Tarifa	Empirical	3.05	70.76	5.27	0.29	0.17	0.17	
	MAP_2	2.97	75.01	4.65	0.29	0.19	0.13	0.05
	Gamma	3.05	70.76	5.50				0.01
	ME	9.35	48.75	3.73				$p < 0.01$

$$\begin{cases} \min \{ \rho(1) - \bar{\rho}(1) \}^2 + \tau \left\{ \left(\frac{\mu_1 - \bar{\mu}_1}{\mu_1} \right)^2 + \left(\frac{\mu_2 - \bar{\mu}_2}{\mu_2} \right)^2 + \left(\frac{\mu_3 - \bar{\mu}_3}{\mu_3} \right)^2 \right\} \\ \text{s.t. } x, u \leq 0, \\ y, v \geq 0, \\ x + y \leq 0, \\ u + v \leq 0. \end{cases} \quad (8)$$

Here τ is a penalty parameter, which in our experiments is set to $\tau = 1$. This choice follows Carrizosa and Ramírez-Cobo (2012), which in addition provides a deeper discussion of the numerical solution of problem (8) and its performance. We stress that solving (8) is doable with standard numerical/statistical packages such as MATLAB or R. In <http://personal.us.es/jrcobo/www/Software.html> a MATLAB© code for solving the optimization problem is available.

4. Application to the analysis of Andalusian rainfall data

4.1. Data description

The collection of analyzed daily precipitation amounts was obtained from the Spanish Meteorological Agency (AEMET). The data

base includes daily records in a total of 11 observatories in Andalusia (south of Spain), from 1968 to 2009 (with the exception of Córdoba, where only the information up to 2006 was available). The observatories are shown in Fig. 1.

To remove the strong seasonal component of the original series associated with the annual cycle of precipitation, the records have been grouped into four sets corresponding to the seasons of winter, spring, summer and autumn. Instead of using the conventional approach of climatological seasons, the astronomical approach has been considered since in this way the seasonal cycle of the Andalusian precipitation is more accurately recorded. To clarify this point consider the Andalusian summer season, which is characterized by extremes rainfall shortages from the end of June until mid September. If climatological seasons were used, the real pattern of the summer rainfall shortages would be distorted since on one hand, part of the spring precipitation amounts (those corresponding to the beginning of June) would be taken into account, and on the other hand, the shortage period related to the beginning of September would be displaced to the fall season. The astronomical approach allows for each seasonal series to be internally rather homogeneous, and additionally, the seasonal component within

Table 3

Columns 3–8: Empirical and estimated first moment, variance, coefficient of asymmetry and first three autocorrelation parameters by the MAP₂, the Gamma and Mixed Exponential, for daily precipitation (spring) Andalusian series. Last column: obtained p-Values from the χ^2 -goodness-of-fit tests.

Station		μ_1	σ^2	γ	$\rho(1)$	$\rho(2)$	$\rho(3)$	p-Value
Almería	Empirical	0.50	8.84	14.30	0.11	0.07	0.00	
	MAP ₂	0.50	8.83	14.30	0.11	0.02	0.00	0.01
	Gamma	0.50	8.84	11.66				0.79
	ME	0.26	4.36	6.33				$p < 0.01$
Cazalla	Empirical	1.92	42.15	6.97	0.33	0.14	0.06	
	MAP ₂	1.92	42.14	6.97	0.33	0.24	0.17	$p < 0.01$
	Gamma	1.92	42.15	6.75				0.08
	ME	3.61	31.61	4.63				0.01
Córdoba	Empirical	1.32	18.47	5.16	0.29	0.09	0.07	
	MAP ₂	1.32	18.50	5.14	0.29	0.18	0.12	0.92
	Gamma	1.32	18.47	6.49				0.35
	ME	1.75	13.85	4.47				$p < 0.01$
Granada	Empirical	1.04	12.09	6.42	0.18	0.07	0.02	
	MAP ₂	1.04	12.09	6.42	0.18	0.07	0.03	$p < 0.01$
	Gamma	1.04	12.09	6.63				$p < 0.01$
	ME	1.09	8.08	4.34				$p < 0.01$
Grazalema	Empirical	3.86	148.22	5.82	0.34	0.16	0.10	
	MAP ₂	3.86	148.20	5.81	0.34	0.26	0.20	0.36
	Gamma	3.86	148.22	6.30				0.27
	ME	14.93	99.79	4.15				$p < 0.01$
Jerez	Empirical	1.23	20.52	5.74	0.28	0.06	0.06	
	MAP ₂	1.23	20.51	5.74	0.28	0.17	0.10	0.31
	Gamma	1.23	20.52	7.32				0.06
	ME	1.52	14.78	4.90				$p < 0.01$
Jimena	Empirical	1.60	28.28	7.97	0.14	0.07	0.01	
	MAP ₂	1.60	28.28	7.97	0.14	0.04	0.01	$p < 0.01$
	Gamma	1.60	28.28	6.62				0.16
	ME	2.57	17.19	4.15				$p < 0.01$
Málaga	Empirical	1.03	22.70	9.17	0.32	0.10	0.03	
	MAP ₂	1.03	22.70	9.17	0.32	0.21	0.14	0.03
	Gamma	1.03	22.70	9.17				0.82
	ME	1.07	11.54	5.13				$p < 0.01$
Pontones	Empirical	2.87	51.38	3.97	0.24	0.09	0.05	
	MAP ₂	2.87	51.20	4.00	0.24	0.14	0.08	0.11
	Gamma	2.87	51.38	4.99				$p < 0.01$
	ME	8.83	49.40	3.96				0.1
Sevilla	Empirical	1.23	18.94	5.47	0.28	0.08	0.06	
	MAP ₂	1.23	18.93	5.47	0.28	0.17	0.11	0.93
	Gamma	1.23	18.94	7.03				0.47
	ME	1.53	14.05	4.78				$p < 0.01$
Tarifa	Empirical	1.30	22.47	6.37	0.32	0.13	0.10	
	MAP ₂	1.30	22.47	6.37	0.32	0.23	0.16	0.2
	Gamma	1.30	22.47	7.28				0.35
	ME	1.69	15.27	4.74				$p < 0.01$

the series is removed without having to use 12 monthly series, which may be too disaggregated.

Crude data presented missing values. Table 1 shows, for the different observatories and seasons, the percentage of missing values, which ranges between 0% and 10.9%. Therefore, an interpolation procedure was used to complete the series. The interpolated values were obtained by linear regression from neighboring series better correlated with the series under study. To avoid the appearance of negative values and keep homogeneity, the regression line was forced to pass through the origin. This approach has been widely applied in the literature to reconstruct daily precipitation series, see for example Eischeid et al. (2000) or Serrano et al. (2010). The obtained correlation coefficients (Pearson's R) were larger than 0.7 in winter, spring and autumn. In the cases of Córdoba and Pontones (those with the largest proportion of missing values), it exceeded the value of 0.9. In summer season the correlations were slightly lower (around 0.6) but this was compensated by the small number of missing values in that season. Finally, following Peterson et al. (1998) and González-Rouco et al. (2000), a quality

control procedure was applied to detect and correct possible errors in the database. A total of six outliers were detected and replaced according to the interpolation procedure.

4.2. Fitting results

In this section we show the suitability of the MAP_2 for the modeling of daily precipitation data. According to Wilks (1999) daily precipitation stochastic models can be considered as either precipitation amounts models, or precipitation occurrence models. In this work we focus on the MAP_2 as a daily precipitation amounts model and therefore, we will fit the probability model defined by (4) and (6) to the real sequences of daily precipitation amounts described in Section 4.1.

Given a sequence of daily precipitations, the four characterizing parameters of a MAP_2 as in (1) are estimated by the solution to the problem (8). Given the estimates of (x, y, u, v) , noted $(\hat{x}, \hat{y}, \hat{u}, \hat{v})$, then the estimated moments and autocorrelation function are obtained by applying formulae (5) and (6) to $(\hat{x}, \hat{y}, \hat{u}, \hat{v})$. From these,

Table 4
Columns 3–8: Empirical and estimated first moment, variance, coefficient of asymmetry and first three autocorrelation parameters by the MAP_2 , the Gamma and Mixed Exponential, for daily precipitation (summer) Andalusian series. Last column: obtained p-Values from the χ^2 -goodness-of-fit tests.

Station		μ_1	σ^2	γ	$\rho(1)$	$\rho(2)$	$\rho(3)$	p-Value
Almería	Empirical	0.11	3.47	37.12	0.01	0.00	0.01	
	MAP_2	0.11	3.47	37.12	0.01	0.00	0.00	0.42
	Gamma	0.11	3.47	32.05				0.59
	ME	0.02	1.03	9.48				0.95
Cazalla	Empirical	0.33	8.15	14.85	0.15	0.03	0.02	
	MAP_2	0.33	8.15	14.85	0.15	0.04	0.01	0.03
	Gamma	0.33	8.15	16.80				0.49
	ME	0.16	6.44	9.49				0.01
Córdoba	Empirical	0.27	5.86	14.76	0.24	0.06	0.00	
	MAP_2	0.27	5.86	14.76	0.24	0.11	0.05	0.7
	Gamma	0.27	5.86	17.76				0.55
	ME	0.09	3.73	9.48				$p < 0.01$
Granada	Empirical	0.22	4.32	17.75	0.09	0.04	0.03	
	MAP_2	0.22	4.32	17.75	0.09	0.01	0.00	0.96
	Gamma	0.22	4.32	18.61				0.76
	ME	0.05	1.87	9.32				$p < 0.01$
Grazalema	Empirical	0.43	15.92	18.56	0.09	0.05	0.02	
	MAP_2	0.43	15.92	18.56	0.09	0.01	0.00	$p < 0.01$
	Gamma	0.43	15.92	18.16				0.6
	ME	0.19	7.39	9.41				0.06
Jerez	Empirical	0.25	9.25	20.21	0.06	0.01	0.00	
	MAP_2	0.25	9.25	20.21	0.06	0.00	0.00	0.76
	Gamma	0.25	9.25	23.96				1
	ME	0.14	5.63	9.48				$p < 0.01$
Jimena	Empirical	0.37	10.41	20.14	0.05	0.02	0.01	
	MAP_2	0.37	10.40	20.14	0.05	0.00	0.00	0.06
	Gamma	0.37	10.41	17.36				0.76
	ME	0.13	4.07	8.29				0.8
Málaga	Empirical	0.18	3.99	19.15	0.16	0.09	0.05	
	MAP_2	0.18	3.99	19.15	0.16	0.05	0.01	0.38
	Gamma	0.18	3.99	22.13				0.75
	ME	0.08	3.39	9.48				$p < 0.01$
Pontones	Empirical	0.63	12.42	9.60	0.12	0.06	0.05	
	MAP_2	0.63	12.42	9.60	0.12	0.03	0.00	0.37
	Gamma	0.63	12.42	11.15				0.23
	ME	0.39	9.97	7.65				0.21
Sevilla	Empirical	0.22	5.71	17.68	0.17	0.01	0.00	
	MAP_2	0.22	5.71	17.68	0.17	0.06	0.02	0.77
	Gamma	0.22	5.71	20.96				0.81
	ME	0.09	3.73	9.48				$p < 0.01$
Tarifa	Empirical	0.16	3.11	23.31	0.13	0.03	0.01	
	MAP_2	0.16	3.11	23.31	0.13	0.03	0.00	0.92
	Gamma	0.16	3.11	21.70				0.62
	ME	0.04	1.60	9.48				$p < 0.01$

the estimated variance and coefficient of asymmetry, σ^2 and γ , are also derived.

In Wilks (1999), the Gamma distribution and Mixed Exponential distribution (from now on, ME) are suggested for modeling daily precipitation amounts. The Gamma distribution is a two-parameters model that generalizes the exponential distribution. The ME model, expressed in terms of three parameters $ME(p, \lambda_1, \lambda_2)$, is indeed a mixture of two exponentials of parameters λ_1 and λ_2 (and p is the probability of the first component), and is also called in the literature hyperexponential distribution. Both the Gamma and ME distribution are particular cases of the phase-type (PH) model defined by the cumulative distribution function (4), and therefore it is expected that these models performs equal to or poorer than the MAP_2 when estimating the marginal distribution of the sample data. In order to test the fit to the empirical distribution function, a χ^2 -goodness-of-fit test was run at a 5% significance level for each of the three considered models.

Columns 3-8 from Tables 2–5 show the empirical values of μ_1, σ^2, γ and $\rho(k)$, for $k = 1, 2, 3$, and their estimated counterparts

by fitting a MAP_2 (via the estimation procedure in Section 3), and the Gamma and ME models (by a simple matching moments method and maximum likelihood, respectively), for winter, spring, summer and fall series, respectively. The fitted values which are closest to the empirical ones are highlighted in bold style. The last column of Tables 2–5 show the p -values obtained from the χ^2 -goodness-of-fit tests. To shorten notation, we only write 2 decimals.

Several remarks need to be made at this point. First, note that, as commented in the introduction, one of the major strengths of the MAP_2 with respect to other models for daily precipitation amounts which consider the observations as independent, is its ability to preserve not only the marginal, but also the joint structure of the data, by fitting the correlation coefficients, which, according to Tables 2–5 are non-negligible. In this case, unlike the MAP_2 , which combines a probability distribution (PH) with the specific autocorrelation function (6), neither the Gamma nor the ME distributions are able to capture the correlation coefficients. Concerning the empirical values it can be seen that the average precipitation amounts (μ_1) are similar between the

Table 5

Columns 3–8: Empirical and estimated first moment, variance, coefficient of asymmetry and first three autocorrelation parameters by the MAP_2 , the Gamma and Mixed Exponential, for daily precipitation (autumn) Andalusian series. Last column: obtained p -Values from the χ^2 -goodness-of-fit tests.

Station		μ_1	σ^2	γ	$\rho(1)$	$\rho(2)$	$\rho(3)$	p -Value
Almería	Empirical	0.81	13.16	7.34	0.17	0.03	0.01	
	MAP_2	0.81	13.16	7.34	0.17	0.06	0.02	0.83
	Gamma	0.81	13.16	8.87				0.45
	ME	0.66	7.89	5.37				$p < 0.01$
Cazalla	Empirical	3.61	125.47	4.92	0.34	0.16	0.09	
	MAP_2	3.61	125.47	4.92	0.34	0.26	0.20	0.34
	Gamma	3.61	125.47	6.20				0.07
	ME	13.04	92.96	4.27				$p < 0.01$
Córdoba	Empirical	2.34	61.15	6.68	0.23	0.13	0.08	
	MAP_2	2.34	61.14	6.68	0.23	0.12	0.06	$p < 0.01$
	Gamma	2.34	61.15	6.66				0.18
	ME	5.50	36.86	4.16				$p < 0.01$
Granada	Empirical	1.34	15.64	4.25	0.22	0.08	0.06	
	MAP_2	1.35	15.43	4.37	0.22	0.11	0.05	0.1
	Gamma	1.34	15.64	5.85				0.01
	ME	1.82	12.31	4.18				$p < 0.01$
Grazalema	Empirical	6.96	417.77	5.05	0.29	0.17	0.09	
	MAP_2	6.83	435.14	4.63	0.29	0.19	0.12	0.67
	Gamma	6.96	417.77	5.87				0.29
	ME	48.45	277.60	3.89				$p < 0.01$
Jerez	Empirical	2.60	66.28	4.86	0.31	0.13	0.09	
	MAP_2	2.60	66.19	4.88	0.31	0.21	0.15	0.06
	Gamma	2.60	66.28	6.24				0.01
	ME	6.80	46.29	4.19				$p < 0.01$
Jimena	Empirical	1.57	23.76	5.94	0.20	0.06	0.04	
	MAP_2	1.57	23.76	5.94	0.20	0.09	0.04	0.06
	Gamma	1.57	23.76	6.19				0.19
	ME	2.47	16.47	4.15				$p < 0.01$
Málaga	Empirical	2.45	88.64	6.49	0.24	0.12	0.07	
	MAP_2	2.42	91.27	6.11	0.24	0.12	0.06	0.9
	Gamma	2.42	88.64	7.65				0.56
	ME	6.04	48.64	4.51				$p < 0.01$
Pontones	Empirical	2.97	63.59	4.34	0.28	0.14	0.08	
	MAP_2	2.97	63.70	4.32	0.28	0.18	0.12	0.85
	Gamma	2.97	63.59	5.35				0.32
	ME	8.87	56.65	4.07				0.01
Sevilla	Empirical	2.41	61.85	5.69	0.25	0.12	0.07	
	MAP_2	2.37	63.69	5.35	0.25	0.13	0.07	0.14
	Gamma	2.37	61.85	6.52				0.05
	ME	5.80	43.49	4.37				$p < 0.01$
Tarifa	Empirical	2.78	68.84	4.83	0.25	0.13	0.09	
	MAP_2	2.78	68.82	4.83	0.25	0.14	0.08	0.8
	Gamma	2.78	68.84	5.95				0.75
	ME	7.77	46.54	3.96				$p < 0.01$

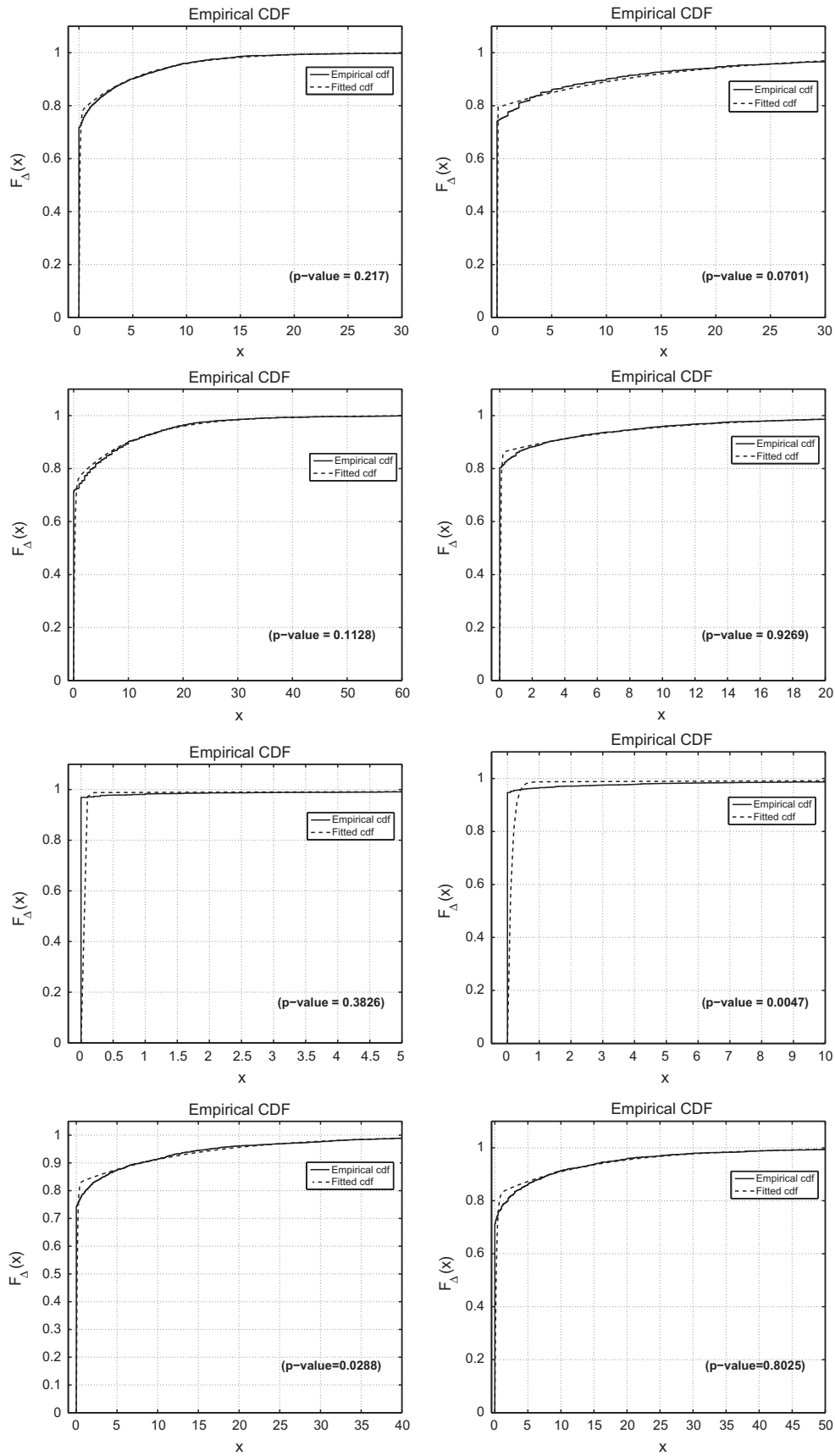


Fig. 2. Empirical (solid line), and estimated (dashed line) cumulative distribution functions $F_{\Delta}(x)$ as in (4) with a MAP_2 model. Each row corresponds to two random stations per season (Granada and Cazalla in winter, Pontones and Sevilla in spring, Málaga and Grazalema in summer, and Jerez and Tarifa in fall).

stations (with the exception of Grazalema). More variability can be found in the distribution of the variance (see for example the values of σ^2 in Cazalla and Grazalema). The positive coefficients of asymmetry imply that the distributions of the data are skewed to the right, as a consequence of possible extreme values. Most series present a significant positive correlation, in some cases around 0.35–0.40 (see Cazalla, Pontones, Grazalema in winter, Cazalla and Grazalema in spring, Córdoba in summer and Cazalla in autumn). With regards to the performance of the MAP_2 , it can be concluded from the tables that the MAP_2 properly fits the marginal moments (mean, variance and skewness coefficient) of the daily precipitation data. In addition, according to the p -values, the MAP_2 can be considered the generator model of the data in 34 (out of 44) stations. From Fig. 2, which depicts the empirical (solid line) and estimated (dotted line) cumulative distribution functions (4) for eight different stations, it can be seen that both the body and the tail of the empirical distributions are correctly captured by the estimated model.

We focus now on the performance of the MAP_2 , in comparison with that of the Gamma and ME models. Note first that the Gamma distribution exactly fits the mean and variance of the data. This is an expected result since the estimation of (α, β) is based on the equality between the theoretical and empirical mean and variance. On the contrary, in some cases the ME performs poorer than the MAP_2 and Gamma distribution when estimating the mean and variance (see for example the series of Grazalema in winter, spring and fall, Tarifa in summer or Cazalla in fall). Concerning the coefficient of asymmetry, it can be seen how the MAP_2 perfectly matches it, while the Gamma distribution tends to generate longer tails (since the coefficient of asymmetry is usually larger). On the contrary, the ME model tends to underestimate the skewness coefficient, which implies that less extreme values will be generated according to the estimated ME distribution. The p -values obtained when running the χ^2 -goodness-of-fit tests indicate that the Gamma and ME are not rejected in 32 and 5 times (out of 44), respectively. The MAP_2 and Gamma models are accepted at the same time for 23 series, in 15 of which the MAP_2 's p -values are higher than the corresponding Gamma's p -values.

In order to examine the MAP_2 with respect to its capacity to reproduce the largest observed values, an analogous analysis to that conducted by Wilks (1999) was performed. Similarly as in the cited reference, the comparison between the observed maxima and their theoretical counterparts is obtained by estimating the cumulative probability associated with each observed maximum $x_{(n)}$ as

$$P(X \leq x_{(n)}) = \frac{n - 1/3}{n + 1/3}, \tag{9}$$

where n represents the sample size. Fig. 3 depicts the relationships between the largest observed daily precipitation amounts for each station and season, in relation to the corresponding (through Eq. (9)) theoretical amounts from the MAP_2 model (top panel), the Gamma model (central panel) and the ME model (bottom panel). Also, the mean squared error was computed for each model.

As a conclusion from Fig. 3 it can be said that both the MAP_2 and Gamma models perform similarly in terms of the obtained mean squared errors. The MAP_2 seems to match better the observed extremes for those series where the maximum daily amounts are smaller than 100 mm, but the Gamma model outperforms the MAP_2 for maximum daily amounts larger than 180 mm, approximately. Also, it is observed that the MAP_2 underestimates the observed extremes in more series than the Gamma model does. On the contrary, the Gamma model sometimes generates largest values which are larger than the real ones. The ME model presents the poorest results: in all series the observed extremes are under-

represented (very significantly in some cases). These facts are supported by the previous discussion about the estimated coefficients of asymmetry shown in Tables 2–5.

Two additional experiments to examine the extreme-value characteristics were considered. In both of them, two stations per

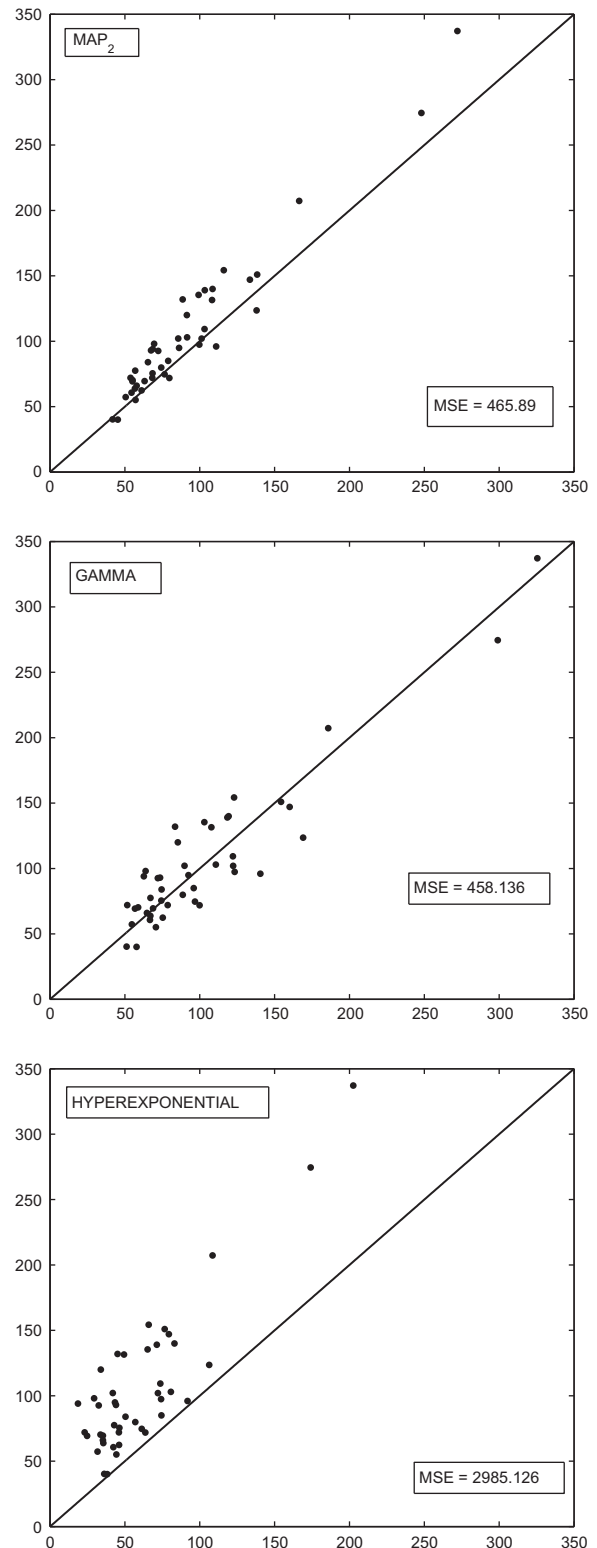


Fig. 3. Relationships between the largest observed daily precipitation amounts (vertical) and the corresponding (through Eq. (9)) model-derived extremes (horizontal) for the MAP_2 model (top panel), Gamma distribution (central panel), and Mixed Exponential distribution (bottom panel).

season were randomly selected, for which three samples of length equal to that of the observed data ($n = 3749$ for winter series, $n = 3833$ for spring series, $n = 3979$ for summer series, and $n = 3780$ for fall series) were generated according to the estimated MAP_2 , Gamma and ME models. Fig. 4 shows the box and whisker

plots of the four series: the real daily precipitation amounts series at each considered station, and the synthetic precipitation amounts series generated by the estimated Gamma, ME and MAP_2 model. The dots in the plots represent the series' extreme values.

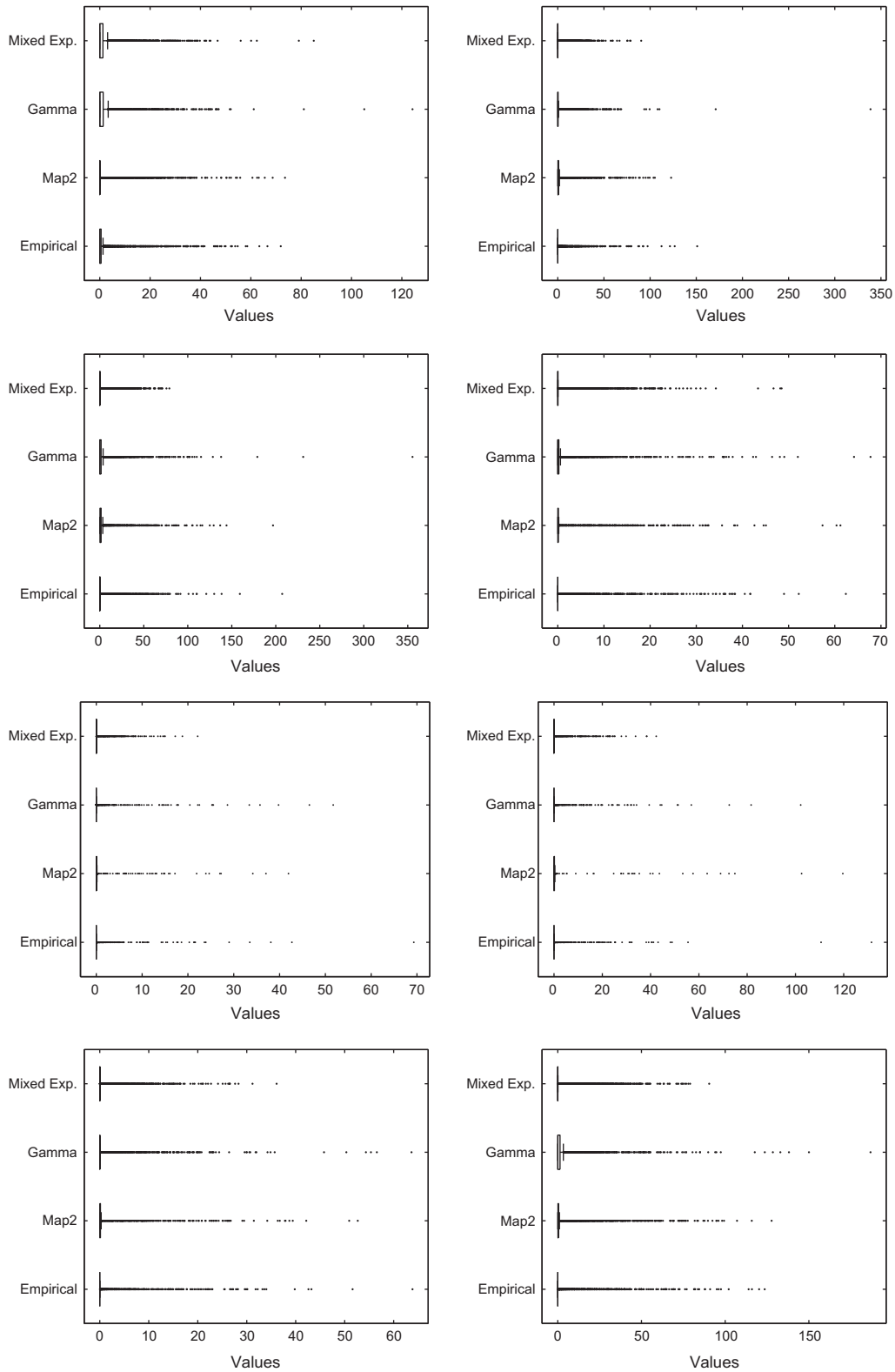


Fig. 4. Box and whisker plots of original daily precipitation amounts series and of generated series according to the estimated ME, Gamma and MAP_2 models. Each row corresponds to two random stations per season (Jerez and Málaga in winter, Grazalema and Jerez in spring, Granada and Grazalema in summer, and Almería and Cazalla in fall).

Note how the MAP_2 model seems to generate samples of the same order as the original data in the considered eight simulated series; in the case of the Gamma distribution, larger values than those in the original series are sometimes generated, while in general, the ME model generates data with lighter tails. Note too that for some series the Gamma and MAP_2 perform similarly (Jerez in spring, Granada in summer), and the ME and MAP_2 also do in some cases (Málaga in winter or Jerez in spring).

In the third experiment, the 95th, 99th and 99.5th percentiles of the real and synthetic precipitation amounts series were computed. They are shown in Fig. 5, where the square, star, circle and diamond symbols represent the percentiles of the real and generated series according to the MAP_2 , Gamma and ME models, respectively. As shown by the box and whisker plots in Fig. 4, the considered models seem to generate extremes values under different patterns, and therefore differences in the highest percentiles are expected.

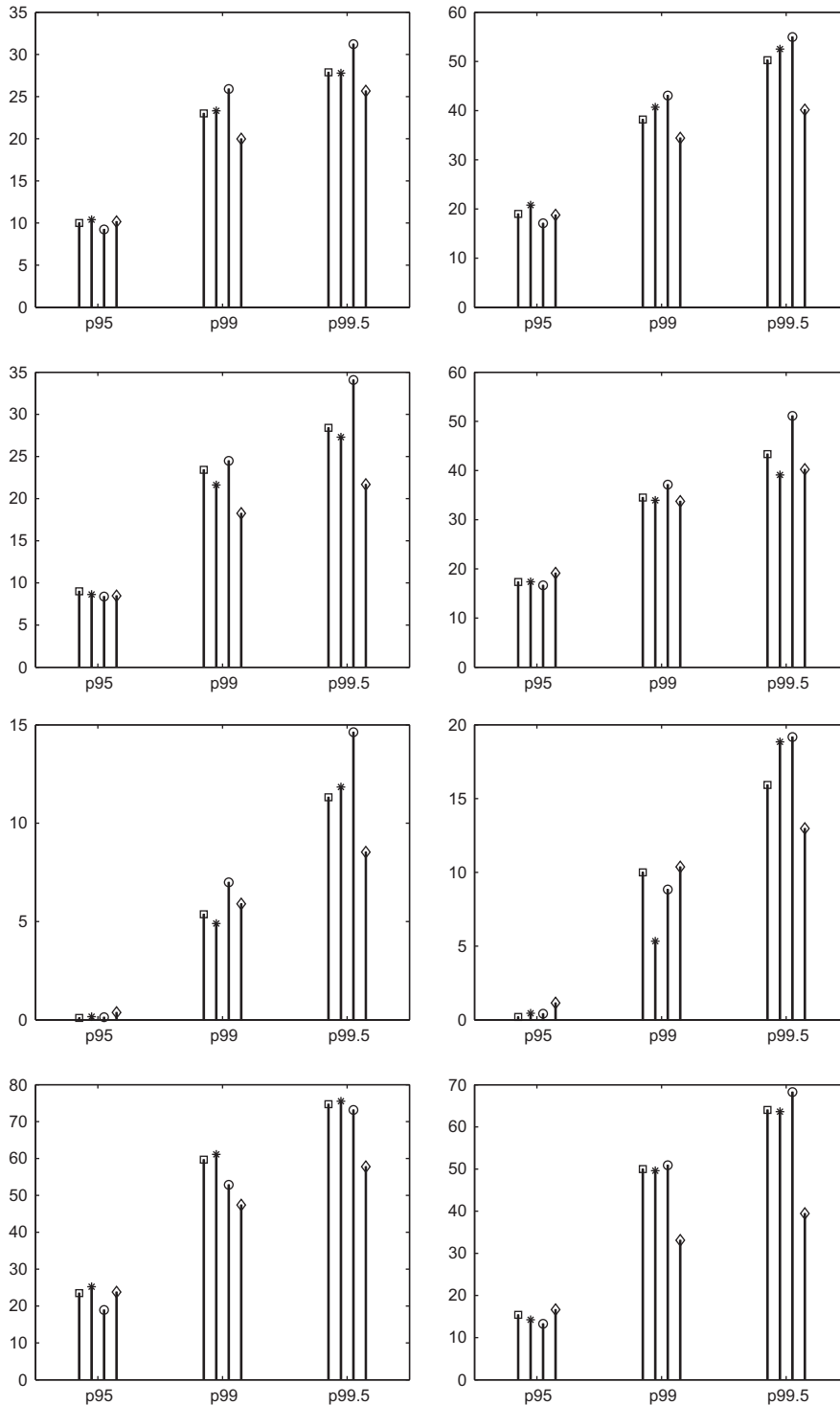


Fig. 5. 95th, 99th and 99.5th percentiles of the real daily precipitation amounts series (\square) and of the generated series from the estimated MAP_2 ($*$), Gamma (\circ) and ME (\diamond) models. Each row corresponds to two random stations per season (Jimena and Tarifa in winter, Córdoba and Pontones in spring, Granada and Jimena in summer, and Cazalla and Málaga in fall).

Indeed, the higher the values of q , the most differences are found between the models. As a general rule, the 99th and 99.5th percentiles obtained under the Gamma distribution seem to be larger than they are under the MAP_2 or ME models, while those from the ME model are the smallest ones (being the summer season an exception). The MAP_2 seems to correctly fit the percentiles in the eight considered series but shows a poorer performance in comparison with the other approaches in summer station.

5. Discussion

In this work we present the two-state MAP as a stochastic model for the daily precipitation intensity. An estimation approach based on a moments matching method has been described and applied to fit the MAP_2 to a set of 44 real daily precipitation amounts series. The MAP_2 is compared to benchmark models for daily precipitation intensity suggested in the literature, as the Gamma and Mixed Gamma distributions. The results show that it performs similarly or better when modeling the marginal distribution of the series and it has a similar performance to the Gamma distribution if the extreme characteristics are examined. However, unlike the competing approaches the MAP_2 not only captures the marginal statistical pattern of the data as the mean, variance and coefficient of asymmetry, but it also correctly fits the empirical persistence which, as shown by the numerical results, may be significative.

A number of extensions to this work are possible. First, we plan to explore the capability of the MAP_2 as a model for precipitation occurrence. Dry spells of Mediterranean climates are known to be difficult to model by popular probability distributions (as the Geometric, Mixture of two Geometric components or Negative Binomial distribution). It is natural to wonder if a better performance of the MAP_2 would be obtained under a different optimization criteria (for example, the maximization of the likelihood function of the spell lengths). We also find of interest to investigate the suitability of the MAP_2 in other climatology contexts where the assumption of independence is not realistic. It would be desirable as well to study the potential of the MAP_2 as a model for prediction and classification purposes. Given that the problem of missing data is not infrequent in precipitation series, a different and promising research line consists in generalizing our estimation methodology for data presenting missing values, by applying an inference technique along the lines of the well-known Expectation–Maximization (EM) algorithm. Finally, the increasing interest in climate change leads to the study of models able to detect trend changes, and the authors believe the MAP_2 could be used with that end. Work on these issues is underway.

In the spirit of a reproducible research the codes utilized in this paper to simulate and estimate the MAP_2 are available at <http://personal.us.es/jrcobo/www/Software.html> as a stand-alone MATLAB© toolbox.

Acknowledgements

The authors thank AEMET for providing real data sets. The research of the authors is supported by Grants MTM2012–36163, Spain, P11-FQM-7603, FQM-329 and HUM7922, Andalucía, all financed in part with EU ERD Funds, and SWAN (Sustainable Water Action – The Seventh Framework programme). The first author is supported by Consolider “Ingenio Mathematica” through her post-doc contract.

References

Ailliot, P., Thompson, C., Thomson, P., 2009. Space time modeling of precipitation using a hidden Markov model and censored gaussian distributions. *J. Roy. Stat. Soc. Ser. C (Appl. Stat.)* 58, 405–426.

- Altman, E., Avrachenkov, K., Barakat, C., 2005. A stochastic model of TCP/IP with stationary random losses. *IEEE/ACM Trans. Network.* 13, 356–369.
- Asmussen, S., 2000. Matrix-analytic models and their analysis. *Scand. J. Stat.* 27, 193–226.
- Asmussen, S., Nerman, O., Olsson, M., 1996. Fitting phase-type distributions via the EM algorithm. *Scand. J. Stat.* 23, 419–441.
- Bellone, E., Hughes, J., Guttorp, P., 2000. A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts. *Climate Res.* 15, 1–12.
- Besson, L., 1924. On the probability of rain. *Mon. Weather Rev.* 52, 308.
- Bodrog, L., Heindlb, A., Horváth, G., Teleka, M., 2008. A Markovian canonical form of second-order matrix-exponential processes. *Euro. J. Operat. Res.* 190, 459–477.
- Brooks, C., Carruthers, N., 1953. *Handbook of Statistical Methods in Meteorology*. Her Majesty Stationery Office, London.
- Carrizosa, E., Ramírez-Cobo, P., 2012. Maximum likelihood estimation of the two-state Markovian arrival process (submitted for publication) arXiv:1401.3105v1.
- Casas, M., Herrero, M., Ninyerola, M., Pons, X., Rodríguez, R., Rius, A., Redaño, A., 2007. Analysis and objective mapping of extreme daily rainfall in Catalonia. *Int. J. Climatol.* 27, 399–409.
- Çınlar, E., 1975. *Introduction to Stochastic Processes*. Prentice-Hall, Usa.
- Eischeid, J., Pasteris, P., Diaz, H., Plantico, M., Lott, N., 2000. Creating a serially complete, national daily time series of temperature and precipitation for the western United States. *J. Appl. Meteorol.* 39, 1580–1591.
- Ephraim, Y., Merhav, N., 2002. Hidden markov processes. *IEEE Trans. Inform. Theory* 48, 1518–1569.
- Feyerherm, A., Bark, L., Dean, J., 1967. Goodness of fit of a Markov chain model for sequences of wet and dry days. *J. Appl. Meteorol.* 6, 770–773.
- Gates, P., Tong, H., 1976. On Markov chain modeling to some weather data. *J. Appl. Meteorol.* 15, 1145–1151.
- González-Rouco, J., Quesada, V., Valero, F., 2000. Quality control and homogeneity of precipitation data in southwest of Europe. *Int. J. Climatol.* 14, 964–978.
- Gronewold, A., Stow, C., Crooks, J., Hunter, T., 2012. Quantifying parameter uncertainty and assessing the skill of exponential dispersion rainfall simulation models. *Int. J. Climatol.* <http://dx.doi.org/10.1002/joc.3469>.
- Holawe, F., Dutter, R., 1999. Geostatistical study of precipitation series in Austria: time and space. *J. Hydrol.* 219, 70–82.
- Hosseini, R., Le, N., Zidek, J., 2011. Selecting a binary Markov model for precipitation process. *Environ. Ecol. Stat.* 18, 795–820.
- Hughes, J., Guttorp, P., 1999. A non-homogeneous hidden Markov model for precipitation occurrence. *J. Roy. Stat. Soc. Ser. C (Appl. Stat.)* 48, 15–30.
- Husak, G., Michaelson, J., Funk, C., 2007. Use of the gamma distribution to represent monthly rainfall in africa for drought monitoring applications. *Int. J. Climatol.* 27, 935–944.
- Kang, S., Han Kim, Y., Sung, D., Choi, B., 2002. An application of Markovian arrival process to modeling superposed ATM cell streams. *IEEE Trans. Commun.* 50, 633–642.
- Kottegoda, N., Natale, L., Raiteri, E., 2004. Some considerations of periodicity and persistence in daily rainfalls. *J. Hydrology* 296, 23–37.
- Kottegoda, N., Natale, L., Raiteri, E., 2008. Stochastic modelling of periodicity and trend for multisite daily rainfall simulation. *J. Hydrol.* 361, 319–329.
- Lana, X., Burgueño, A., 1998. Daily dry-wet behavior in Catalonia (NE Spain) from the viewpoint of Markov chains. *Int. J. Climatol.* 18, 793–815.
- Lana, X., Burgueño, A., 2000. Some statistical characteristics of monthly and annual pluviometric irregularity of the Spanish Mediterranean coast. *Theor. Appl. Climatol.* 65, 79–97.
- Lana, X., Burgueño, A., Martín-Vide, J., Serra, C., 2006. Statistical distributions and sampling strategies for the analysis of extreme dry spells in Catalonia (NE Spain). *J. Hydrol.* 324, 94–114.
- Lana, X., Burgueño, A., Serra, C., Martín-Vide, J., Gómez, L., 2008. Spatial and temporal patterns of dry spell lengths in the Iberian peninsula for the second half of the twentieth century. *Theor. Appl. Climatol.* 91, 99–116.
- Lennartsson, J., Baxevani, A., Chen, D., 2008. Modeling precipitation in Sweden using multiple step markov chains and a composite model. *J. Hydrol.* 363, 42–59.
- Lucantoni, D., 1991. New results for the single server queue with a batch Markovian arrival process. *Stoch. Models* 7, 1–46.
- Martínez, M., Lana, X., Burgueño, A., 2010. Long-term rainfall monthly shortage in Spain: spatial patterns, statistical models and time trends. *Int. J. Climatol.* 30, 1668–1688.
- Martín-Vide, J., Gómez, L., 1999. Regionalization of peninsular Spain based on the length of dry spells. *Int. J. Climatol.* 19, 537–555.
- Mehrotra, R., Sharma, A., 2007a. Preserving low-frequency variability in generated daily rainfall sequences. *J. Hydrol.* 345, 102–120.
- Mehrotra, R., Sharma, A., 2007b. A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability. *J. Hydrol.* 335, 180–193.
- Mekis, E., Hogg, W., 1999. Rehabilitation and analysis of Canadian daily precipitation time series. *Atmos. Ocean* 37, 53–85.
- Montoro-Cazorla, D., Pérez-Ocón, R., Segovia, M., 2009. Replacement policy in a system under shocks following a Markovian arrival process. *Reliab. Eng. Syst. Safety* 94, 497–502.
- Neuts, M.F., 1979. A versatile Markovian point process. *J. Appl. Probab.* 16, 764–779.
- Neuts, M.F., Pérez-Ocón, R., Torres-Castro, I., 2000. Repairable models with operating and repair times governed by phase type distributions. *Adv. Appl. Probab.* 32, 468–479.
- O’Cinneide, C., 1990. Characterization of phase-type distributions. *Stoch. Models* 6, 1–57.

- Peñarrocha, D., Estrela, M., Millán, M., 2002. Classification of daily rainfall patterns in a Mediterranean area with extreme intensity levels: the Valencia region. *Int. J. Climatol.* 22, 677–695.
- Peterson, T.C., Vose, R., Schmoyer, R., Razuvaev, V., 1998. Global Historical Climatology Network (GHCN) quality control of monthly temperature data. *Int. J. Climatol.* 18, 1169–1179.
- Ramírez-Cobo, P., Lillo, R., Wiper, M., 2010. Non-identifiability of the two-state Markovian arrival process. *J. Appl. Probab.* 47, 630–649.
- Romero, R., Guijarro, J., Ramis, C., Alonso, S., 1998. A 30 year (1964–1993) daily rainfall data base for the Spanish Mediterranean regions: first exploratory study. *Int. J. Climatol.* 18, 541–560.
- Romero, R., Guijarro, J., Ramis, C., Sumner, G., 1999. Daily rainfall affinity areas in Mediterranean Spain. *Int. J. Climatol.* 19, 557–578.
- Serrano, S.V., Beguería, S., López-Moreno, J., Vera, M.G., Stepanek, P., 2010. A complete daily precipitation database for northeast Spain: reconstruction, quality control, and homogeneity. *Int. J. Climatol.* 30, 1146–1163.
- Srikanthan, R., Pegram, G., 2009. A nested multisite daily rainfall stochastic generation model. *J. Hydrol.* 371, 142–153.
- Stern, R., Coe, R., 1984. A model fitting analysis of daily rainfall data. *J. Roy. Stat. Soc. Ser. A* 147, 1–34.
- Wang, Q., Nathan, R., 2007. A method for coupling daily and monthly time scales in stochastic generation of rainfall series. *J. Hydrol.* 346, 122–130.
- Wilks, D., 1998. Multisite generalization of a daily stochastic precipitation generation model. *J. Hydrol.* 210, 178–191.
- Wilks, D., 1999. Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agric. Forest Meteorol.* 93, 153–169.