

Estimating extreme monthly rainfall for Spain using non-stationary techniques

Diego Urrea Méndez & Manuel del Jesus

To cite this article: Diego Urrea Méndez & Manuel del Jesus (2023) Estimating extreme monthly rainfall for Spain using non-stationary techniques, Hydrological Sciences Journal, 68:7, 903-919, DOI: [10.1080/02626667.2023.2193294](https://doi.org/10.1080/02626667.2023.2193294)

To link to this article: <https://doi.org/10.1080/02626667.2023.2193294>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 04 May 2023.



Submit your article to this journal [↗](#)



Article views: 2135



View related articles [↗](#)



View Crossmark data [↗](#)

Estimating extreme monthly rainfall for Spain using non-stationary techniques

Diego Urrea Méndez  and Manuel del Jesus 

IHCantabria - Instituto de Hidráulica Ambiental de la Universidad de Cantabria, Santander, Spain

ABSTRACT

In hydrology, extreme value analysis is normally applied at stationary yearly maxima. However, climate variability can bias the estimation of extremes by partially invalidating the stationary assumption. Extreme value analysis for sub-yearly data may depart from stationarity (since maxima from one month may not be exchangeable with maxima from another) in terms of requiring to include it in the analysis. Here, we analyse the non-stationary structure of extreme monthly rainfall in Spain using two approaches: a parametric approach and an approach based on autoregressive time series models. Our analysis considers seasonality, climate variability and long-term trends for both approaches, and it compares both including their goodness of fit and complexity. The approach uses maximum likelihood estimation and Bayesian techniques. Our results show that autoregressive models outperform parametric models, providing a more accurate representation of extreme events when extrapolating outside of the period of fit.

ARTICLE HISTORY

Received 6 April 2022
Accepted 2 February 2023

EDITOR

A. Castellarin

GUEST EDITOR

A. Sharma

KEYWORDS

non-stationary; GEV distribution; maximum likelihood; extreme values; rainfall; maximum monthly rainfall; parametric models; autoregressive models; Monte Carlo; Bayesian techniques

1 Introduction

The analysis of extreme hydrometeorological events allows one to make inferences based on the historical records of hydroclimatic variables. Commonly, these inferences serve to predict the probability of occurrence of threshold levels of the variable for future extreme events. The selection of the maximum values of a variable in successive periods of time (maxima in a block of time, block maxima or BM) forms the fundamental basis for modelling climatic extremes. One of the simplest techniques for extreme event selection is the method of annual maxima (AMM) (Gumbel 1958), which consists of selecting the maximum value per year, its main limitation being a sub-optimal use of the available information. In addition, due to climate variability on time scales below the year and the low availability of hydrological series with sufficiently long records, other alternatives have been proposed. (Smith 1989) proposed the peaks-over-threshold (POT) method that focuses on defining the extreme events of interest as those values exceeding a given threshold. This approach presents some limitations as it rules out local extreme events, for instance those occurring in summer, so it cannot model all the inter-annual variability (Menéndez *et al.* 2009). Another type of study that overcomes the limitations of the AMM uses monthly maxima – instead of annual – for the study of environmental variables (Méndez *et al.* 2007, 2009, Mínguez *et al.* 2010). The use of monthly maxima has multiple applications that include flood risk management at specific times of the year (for instance, ice melting season) and reservoir management (to determine the

maximum flood that can be controlled during a given month depending on the stored volume). An additional advantage of using monthly maxima corresponds to the possible increase in data used, which improves the fit of the upper tail of the distribution by the maximum likelihood estimation (MLE) method (Méndez *et al.* 2007).

The generalized extreme value distribution (GEV) (Jenkinson 1955) is one of the most common distributions for fitting extreme events associated to environmental variables (Stedinger *et al.* 1993). This distribution has been widely used due to its ability to capture a wide range of tail behaviours (Coles 2001). Fitting the distribution implies finding the values of the three parameters of the GEV that best capture the distribution of the observed maxima. Although there are several methods to find the parameters, the MLE is the most used one (Smith 1985).

Classical frequency analysis assumes that the parameters of the distribution remain constant over time (stationarity condition). This assumption implies that the samples used for fitting are interchangeable. Any two values could be interchanged, and the original series could not be distinguished from the modified one. This is only the case when stationary periods are used for fitting, and that is why yearly maxima are commonly used. The stationarity assumption also implies that current and future extremes follow the same distribution, so we should not expect any difference in the behaviour of extremes in the future.

However, the stationarity assumption is not fully satisfied under real conditions since climate variability introduces some degree of non-stationarity – for example, the maximum rainfall in

a dry year and in a wet year may differ enough so as to be distinguishable. Climate change may also introduce trends in the time series of environmental variables that lead to some degree of time dependence in the distribution's parameters. Proof of this claim is the large amount of research in which the non-stationarity of extreme events has been analysed (Adlouni *et al.* 2007, Mínguez *et al.* 2010, Solari and Losada 2011, Salas and Obeysekera 2014). For some applications, then, the stationary GEV model – with constant parameters – is no longer valid under our current climate conditions (Leadbetter *et al.* 1983).

There is a non-null probability that extreme weather events will change in the future (Kharin and Zwiers 2000, 2005, Wang *et al.* 2004). In addition, there is evidence of the influence of climate change on the distribution of extreme hydroclimatic events, which suggests that they can hardly be treated as stationary (Wang *et al.* 2004, IPCC *et al.* 2013), reinforcing the importance of non-stationary analyses. This evidence has promoted advances in extreme value analysis in recent decades that have allowed practitioners to better characterize the natural climate variability and the non-stationarity of extreme events. For instance, Scarf (1992) considered an extension of the GEV distribution incorporating a trend in the location parameter; Coles (2001) provided an approach describing the use of covariates in parameters that incorporated teleconnection indices such as the Southern Oscillation Index (SOI); and Katz *et al.* (2002) provided examples that included the analysis of extremes associated with time series of precipitation and currents, including trends and the dependence on synoptic patterns in the atmosphere–ocean circulation.

Different authors have incorporated covariates in their non-stationary analysis of hydroclimatological extremes in recent decades. For instance Méndez *et al.* (2007) introduced a time-dependent GEV model incorporating harmonic functions (with annual periods), secular trends, and regional climate patterns from teleconnection indices in the location, scale, and shape parameters. Brown *et al.* (2008) studied global changes in daily temperature extremes since 1950, incorporating regional trends and atmospheric synoptic patterns such as the North Atlantic Oscillation (NAO). Menéndez *et al.* (2009) proposed a time-dependent GEV model considering the inter-annual variability and the seasonality of extreme wave heights. López and Francés (2013) implemented generalized additive models of location, scale, and shape (GAMLSS) for the analysis of non-stationary flood frequency in Spanish continental rivers, using climatic and reservoir indices as external covariates. Other approaches that identify plausible atmospheric indicators of changes in future precipitation extremes can be found in Roderick *et al.* (2020). Most of these analyses fit their models using the MLE approach.

With large sample sizes, the MLE method generates good fits. However, when the sample size is relatively small, the numerical solution applied to find the distribution parameters tends to diverge. To prevent this problem, some techniques have been proposed based on the use of prior distributions that contain information on the most probable values that the GEV parameters can take. These approaches have methodological similarities to the quasi-Bayesian method of maximum likelihood (QBML) proposed by Hamilton (1991), which removes the singularities associated with MLE and sets up Monte Carlo simulations showing its constant potential to improve mean squared

errors. Similar studies exist, such as that of Zhang *et al.* (2004), who use Monte Carlo simulations, comparing multiple methods to detect a trend in the magnitude of the extreme values; or Nakajima *et al.* (2012), who implemented a Bayesian approach using the Markov chain Monte Carlo (MCMC) method to find an accurate approximation of the GEV model that used an autoregressive (AR) or moving average (MA) process.

Integrating the methodologies and approaches described above to develop an analysis of extreme events at a regional scale poses an additional challenge, since all gauge analyses must show compatibility with one another while capturing the specific characteristics of each site. There are methods such as regional frequency analysis (RFA), based on the selection of data from several measurement sites, that assume all data follow the same probability distribution (Hosking and Wallis 1993, Bradley 1998); that is, they assume that the extremes are homogeneous over the region. Such analysis involves the division of a geographical area to test whether they are homogeneous and can be represented by a unique distribution model. By including multiple observations at different spatial points, RFA provides two main benefits over single-site models: (1) frequency estimates at uncalibrated locations, and (2) improved estimates of distribution parameters across locations (Bracken *et al.* 2018). Various approximations of the RFA can be found in the literature, including a regionalization for the study of floods and precipitations in non-stationary conditions (Renard 2011, Chebana *et al.* 2014, Bracken *et al.* 2016).

Although there are studies in the literature that integrate the non-stationary approach with regional analysis, we have not found any study that compares and evaluates the application of parametric and AR methods. Therefore, the objective of this study is to analyse the models that best characterize and estimate the temporal variation of extreme precipitation events, using data from Spain to make the comparisons. To complement the proposed objective, we propose to divide the territory into climatically homogeneous zones (following the Köppen-Geiger climate classification) and then verify whether extreme events are comparable between zones and can be captured by the same model. Two approaches will be used to formulate the models: (1) a parametric approach based on a non-stationary GEV model (Coles 2001), incorporating covariates and an explicit dependence on time; and (2) an AR approach in which the parameters of the GEV distribution will follow a time series model. The first approach will be fitted with the MLE method, while the second one will use Bayesian techniques.

2 Study area and information sources

We developed our study in Spain mainly for two reasons: first, because Spain has a dense network of pluviometers (Fig. 1) with available information; second, and more important, due to the wide range of climates that can be found there. Indeed, many different climates can be found in Spain, with strong changes occurring over very short distances. Four of the five climatic groups defined by the Köppen-Geiger climate classification (AEMET and IPMA 2011) can be found in Spain. Only group A, tropical climate, is absent. The most frequent climates are of type C (temperate) and of type B (dry), which can be found in continental Spain, as well as in the Canary and the

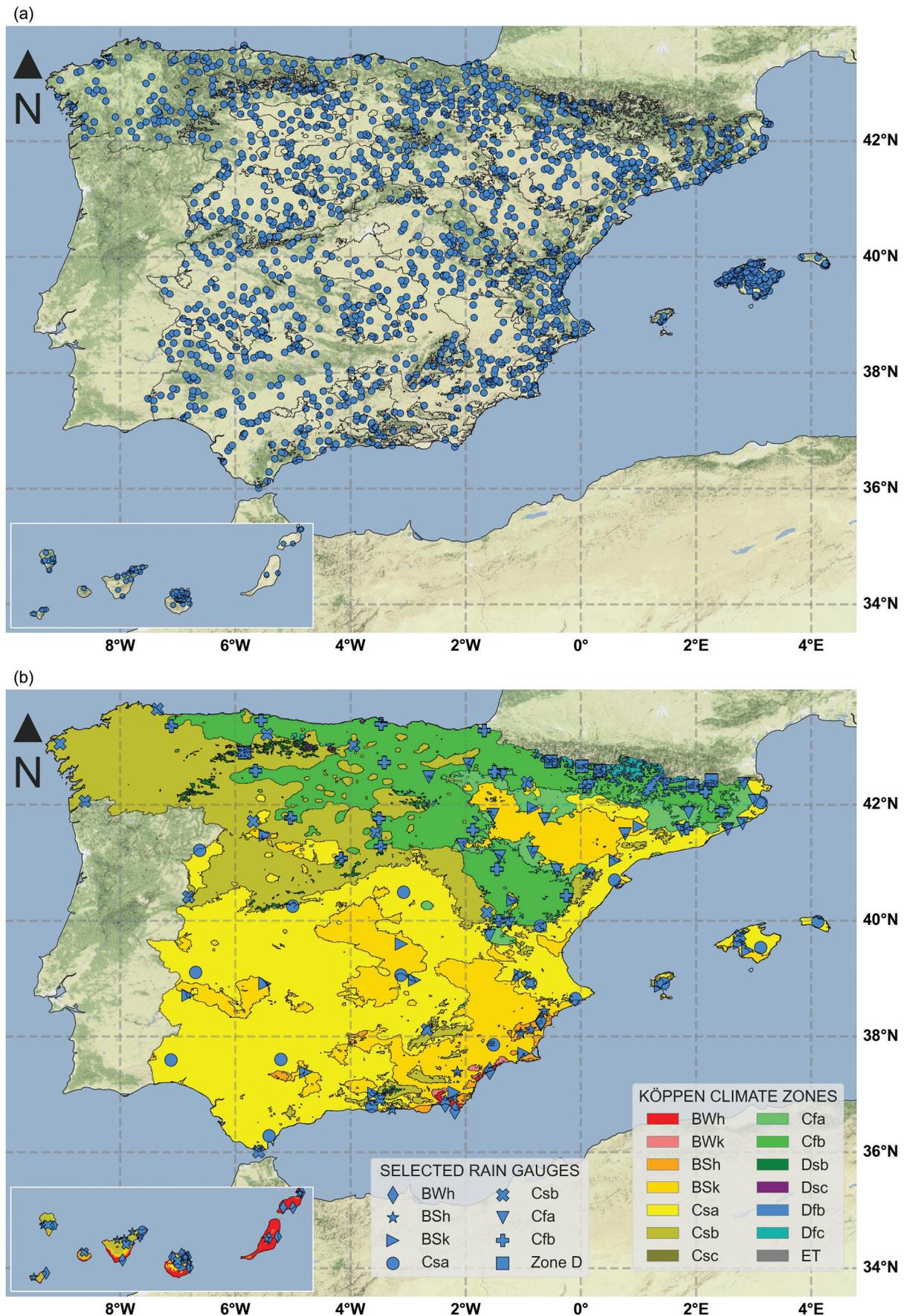


Figure 1. (a) Location of the Agencia Estatal de Meteorología de España (AEMET) daily raingauges (8310 gauges, blue points); (b) spatial distribution of the Köppen-Geiger climatic groups, created by AEMET, and selected raingauges with precipitation records in the period 1979–2019. The abbreviations that appear in the legend KÖPPEN CLIMATE ZONES refer to climatic sub-groups: Bsk, cold steppe; BSh, hot steppe; BWh, hot desert; BWk, cold desert; Csa, temperate with dry or hot summer; Csb, temperate with dry or temperate summer; Csc, temperate cold-summer; Cfa, temperate with a dry season and hot summer; Cfb, temperate with a dry season and temperate summer; Dsb, hot summer humid continental Mediterranean influence; Dsc and Dfc, sub-arctic climates; Dfb, hot summer humid continental climate; ET, tundra climate. The legend SELECTED RAIN GAUGES shows the shape of the selected station for each climatic zone.

Balearic Islands. Types D and E (continental and alpine, respectively) are present in mountain areas, although the latter only in high areas of the Pyrenees.

This study used daily precipitation data recorded in the Agencia Estatal de Meteorología de España (AEMET) network of pluviometers (Fig. 1). The network covers most of the climatic zones, except for the sub-groups Csc (climate type C) and ET (climate type E). These sub-groups occur in high areas of the Pyrenees and do not contain pluviometers with daily precipitation data, so they were not included in the analysis. Similarly, no information was available for the sub-group BWk (climate type B). Furthermore, mountain climates Dsb, Dsc, Dfb and Dfc only contain six raingauges with available information; for this reason, considering their geographical characteristics, they were integrated into a climatic group called D (Diez-Sierra and Del Jesus 2019).

In total, 8310 raingauges were processed, and we performed quality control on their time series. This process included the identification of atypical data, ensuring that they were physically possible (Gonzalez and Bech 2017); a control of repeated values for two or more consecutive days; and the implementation of quality indices that discarded time series with many missing data, false zeros, null data, etc. (Lez-Rouco 2001, Llabrés-Brustenga *et al.* 2019).

After selecting the time series with the best quality indicators, a second criterion for selection corresponded to the percentual coverage and the record period. Stations with at least 80% complete years in the period between 1979 and 2019 were considered. A full year was defined as one that had at least 80% complete data for all months. A total of 1576 stations passed the quality control. Due to the high computational cost involved in processing this information in subsequent analyses, the best 20 stations for each climate were selected, ensuring that they homogeneously covered the spatial distribution of each climate (Fig. 1). In total, 146 raingauges were used, since group D only contained six stations.

Teleconnection indices were used to capture inter-annual fluctuations in the occurrence of extreme precipitation events. These indices were used as covariates for the variation of the parameters of the GEV distribution. Specifically, we used the NAO and the Mediterranean Oscillation Index (MOI). These indices were selected considering the influence they had on the variability of rainfall in recent decades in Spain (Martin-Vide and Lopez-Bustins 2006). The MOI and NAO indices were obtained from the University of East Anglia Climate Research Unit (University of East Anglia 2021, data available at <http://www.cru.uea.ac.uk/cru>). All indices have daily data and cover the analysis period (1979–2019).

3 Methods

The main objective of our study is to determine an optimal non-stationary model to characterize extreme events of precipitation. We also aim to determine whether this optimal model changes depending on the analysed climate type.

To attain these objectives, we follow three steps (shown in Fig. 2). First, the historical precipitation and atmospheric data are reviewed and processed to build a database of quality-controlled observations. Second, non-stationary models of extremes (presented in section 3.1) are fit to the observations and compared to one another to determine the optimal

model (see section 3.2.1 and section 3.2.2). Two different techniques are used in this step: parametric models fitted by MLE, and AR time series models fitted using Bayesian techniques. Finally, the optimal model in each category (MLE and Bayesian) are compared to determine which approach provides the best approximation. The comparison is carried out using two measures: one based on a performance metric (see section 3.3), and another based on how well the approach estimates future precipitation (see section 3.3.2).

The optimal model in both approaches is not determined only based on the goodness of fit but accounts for model complexity also, to favour parsimonious models (assuming an equal goodness of fit, the model with fewer parameters should be preferred; see section 3.3). This trade-off is attained using the Akaike information criterion (AIC) (Akaike 1998) as the performance metric.

The proposed methodology aims to find the models that best characterize and estimate the temporal variation of extreme precipitation events in Spain. It also aims to verify whether the optimal model is climate-dependent, or if the temporal structure of extremes can be captured with a single model in the whole territory. The methodology is structured in four main modules, shown in Fig. 2, that are briefly described below:

- (1) Panel (a): review and pre-processing of historical precipitation and atmospheric data.
- (2) Panel (b): configuration and fitting of parametric models, using AIC as the optimization metric, separating extreme events by climate following the Köppen-Geiger classification.
- (3) Panel (c): configuration and fitting of AR time series models, using AIC as the optimization metric, separating extreme events by climate following the Köppen-Geiger classification.
- (4) Panel (d): comparison of optimal parametric and AR time series models by climatic zone, to determine the best-performing model family.

3.1 Non-stationary GEV model

Extreme value distributions were introduced by Fisher and Tippett (1928), and it was Jenkinson (1955) who combined the three families of extreme distributions – Gumbel, Fréchet and Weibull – into a unique cumulative distribution function (CDF), known as the GEV distribution, given by Equation (1):

$$F(z) = \begin{cases} \exp\left\{-\left[1 + \xi_{(t)}\left(\frac{z-\mu_{(t)}}{\sigma_{(t)}}\right)\right]^{\frac{1}{\xi_{(t)}}}\right\}; \xi_{(t)} \neq 0 \\ \exp\left\{-\exp\left[-\left(\frac{z-\mu_{(t)}}{\sigma_{(t)}}\right)\right]\right\}; \xi_{(t)} = 0 \end{cases} \quad (1)$$

where μ is the location parameter ($-\infty < \mu < \infty$), σ is the scale parameter ($\sigma > 0$) and ξ is the shape parameter ($-\infty < \xi < \infty$). The Gumbel, Weibull and Fréchet distributions correspond to $\xi = 0$, $\xi < 0$ and $\xi > 0$, respectively. In addition, the domain of the distribution verifies the equation $1 + \xi(z - \mu)/\sigma > 0$. For the evaluation of non-stationary extremes, it is possible to use the *BM* approach with monthly maxima per year $\{Z_{ym} = \max(X_{ym1}, X_{ym2}, X_{ym3}, \dots, X_{ymn})\}$, where Z_{ym} represents the maximum value of the

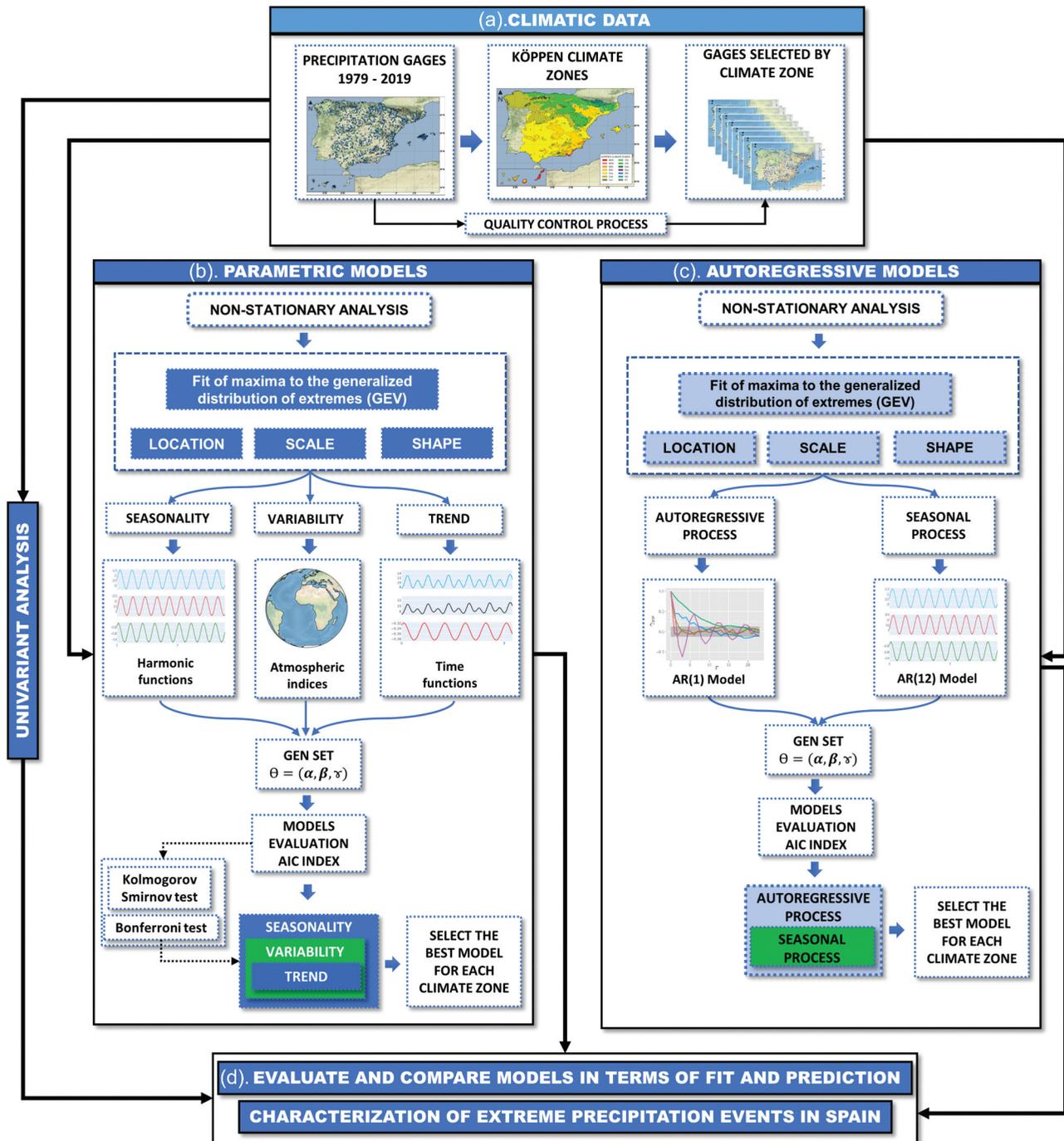


Figure 2. Proposed methodological approach, where panel (a) shows the framework for the reviewing and pre-processing of the historical precipitation and atmospheric data in Spain. Panels (b) and (c) present the methodology for selecting the optimal parametric and autoregressive models, respectively, in terms of goodness of fit and complexity. Panel (d) shows the comparison and selection of the optimal parametric or autoregressive model by climatic zone, in terms of goodness of fit and complexity, for the probabilistic characterization and prediction of extreme precipitation events in Spain.

selected variable for month m of year y ; X_{ymi} represents the maximum value of day i , in month m of year y ; and n refers to the total number of days that month m has in the year y .

For non-stationary models, the parameters of the GEV distribution can be expressed as a function of covariates that change with time (Coles 2001). The temporal dependence in models of extremes can be parameterized assuming that the monthly maxima of the successive months are independent random variables. The monthly maximum of the observed variable Z_t in month t is

modelled by a GEV distribution with time-dependent parameters (μ_t , σ_t , ξ_t) and with a probability density function (PDF) as expressed in Equation (2):

$$f(z) = \begin{cases} \frac{1}{\sigma_t} \left[1 + \xi_t \left(\frac{z - \mu_t}{\sigma_t} \right) \right]^{-\left(1 + \frac{1}{\xi_t}\right)} \exp \left\{ - \left[1 + \xi_t \left(\frac{z - \mu_t}{\sigma_t} \right) \right]^{\frac{1}{\xi_t}} \right\}; \xi \neq 0 \\ \frac{1}{\sigma_t} \exp \left(- \frac{z - \mu_t}{\sigma_t} \right) \exp \left[- \exp \left(- \frac{z - \mu_t}{\sigma_t} \right) \right]; \xi = 0 \end{cases} \quad (2)$$

3.1.1 GEV parametric models

Parametric models represent the time variation of the GEV parameters, expressing them as linear combinations of functions that change over time. We will call these functions that change over time covariates. Covariates may represent the effects of seasonality, variability, and trend. Seasonality can be introduced with sinusoidal functions (Menéndez *et al.* 2009) that explicitly consider annual and semi-annual cycles in the parameters of the GEV distribution. The inter-annual and decadal variability can be captured from regional climate patterns, which are generally represented by teleconnection indices (Hatzaki *et al.* 2010, Gregersen *et al.* 2013). The climatic variability in Spain is influenced by a great variety of indices, but the most relevant ones for precipitation are the MOI and NAO indices (Martin-Vide 2004, Lopez-Bustins *et al.* 2008). We assume that the effect of the teleconnection indices over the time variation of the parameters is linear. Two forms for the long-term trend are considered: a linear term (t) and a quadratic one (t^2). The model that integrates all the aforementioned effects is shown in Equations (3), (4) and (5).

$$\begin{aligned} \mu_{(t)} = & \alpha_{11} + \alpha_{21}\cos(2\pi t) + \alpha_{22}\sin(2\pi t) + \alpha_{23}\cos(4\pi t) \\ & + \alpha_{24}\sin(4\pi t) + \dots + \alpha_{31}MOI + \alpha_{32}NAO + \alpha_{41}t \\ & + \alpha_{42}t^2 \end{aligned} \quad (3)$$

$$\begin{aligned} \sigma_{(t)} = & \beta_{11} + \beta_{21}\cos(2\pi t) + \beta_{22}\sin(2\pi t) + \beta_{23}\cos(4\pi t) \\ & + \beta_{24}\sin(4\pi t) + \dots + \beta_{31}MOI + \beta_{32}NAO + \beta_{41}t + \beta_{42}t^2 \end{aligned} \quad (4)$$

$$\begin{aligned} \xi_{(t)} = & \gamma_{11} + \gamma_{21}\cos(2\pi t) + \gamma_{22}\sin(2\pi t) + \gamma_{31}MOI \\ & + \gamma_{32}NAO + \gamma_{41}t + \gamma_{42}t^2 \end{aligned} \quad (5)$$

where the coefficients α_{11} , β_{11} and γ_{11} represent the average values of the GEV parameters; the coefficients α_{2n} , β_{2n} and γ_{2n} correspond to the amplitudes of the sinusoidal functions that capture seasonality; the coefficients α_{3n} , β_{3n} and γ_{3n} indicate the sensitivity of extreme rainfall to the teleconnection indices; and α_{4n} , β_{4n} and γ_{4n} account for the trend.

The coefficients that multiply the covariates of the model can be packed into a vector $\Delta = (\alpha_{1n}, \alpha_{2n}, \alpha_{3n}, \alpha_{4n}, \beta_{1n}, \beta_{2n}, \beta_{3n}, \beta_{4n}, \gamma_{1n}, \gamma_{2n}, \gamma_{3n}, \gamma_{4n})$, which, in the case where all the variables are considered, will contain 25 coefficients, that would multiply 25 non-null regression covariates ($p = 25$). The vector of parameters of the candidate models, Δ , will be estimated using the maximum likelihood method. To ensure that the scale parameter does not have negative values, an exponential transformation is used such that $\sigma_{(t)}^* = \exp(\sigma_{(t)})$. Similarly, to ensure that the shape parameter lies in an acceptable range, a hyperbolic tangent transformation is applied $\xi_{(t)}^* = 0.5 \cdot \tanh(\xi_{(t)}/100)$. This transformation guarantees that the parameter values obtained lie between -0.5 and 0.5 , which are typical values that can be obtained when analysing extreme precipitation data in Spain.

3.1.2 GEV AR models

AR time series models are a representation of a random process where the behaviour of the variable of interest has a linear dependence on its past values. The notation of an autoregressive model of order p , $AR(p)$, is defined in Equation (6).

$$x_{(t)} = c + \sum_{i=1}^p \phi_i \cdot x_{(t-i)} + \varepsilon_{(t)} \quad (6)$$

where ϕ_1, \dots, ϕ_n are the model's AR parameters, c is a constant and $\varepsilon_{(t)}$ is a white noise term. Equation (6) can be written equivalently using a backtracking operator A , as presented in Equation (7).

$$X_t = c + \sum_{i=1}^p \phi_i A^i X_t + \varepsilon_{(t)} \quad (7)$$

Moving the summation term to the left-hand side and using polynomial notation, we obtain a more compact representation (Equation 8).

$$\varphi(A)X_t = c + \varepsilon_{(t)} \quad (8)$$

An AR model can be expressed in such a way that the infinite response input or impulse is the white noise. For the process of a generalized $AR(p)$ model to be stationary, the roots of the polynomial $z^p - \sum_{i=1}^p \phi_i z^{p-i}$ must be contained in the unit circle.

To accomplish this, each root Z_i must satisfy $|Z_i| < 1$. For example, for a model $AR(1)$ to be stationary in all directions, the inequality $|\phi| < 1$ must hold. If $|\phi| = 1$, then $x_{(t)}$ has infinite variance and is non-stationary. Assuming that for a given autoregressive model $AR(1)$, $|\phi| < 1$, the mean value $E_{(t)}$ is the same for all values of t .

$$E(X_t) = E(c) + \phi E(X_{t-1}) + E(\varepsilon_{(t)}) \quad (9)$$

If the mean, $E_{(t)}$, of the process is λ , then,

$$\lambda = c + \phi\lambda + 0 \quad (10)$$

And therefore:

$$\lambda = \frac{c}{1 - \phi} \quad (11)$$

Note that if $c = 0$ the mean of the process is 0. In turn, the variance is described as:

$$\text{var}(X_t) = E(X_t^2) - \lambda^2 = \frac{\sigma_\varepsilon^2}{1 - \phi^2} \quad (12)$$

where σ_ε is the standard deviation of ε_t . This equality can be shown by noting that:

$$\text{var}(X_t) = \phi^2 \text{var}(X_{t-1}) + \sigma_\varepsilon^2 \quad (13)$$

The GEV-AR model incorporates the AR model described above to account for the time variation of the GEV distribution parameters; the time dependence of the location, scale, and shape parameters is modelled by an AR process. The proposed model incorporates the time dependence using a state space

representation where the parameters of the GEV are modelled by the combination of an and an AR(12) model, depending on whether seasonality (the latter term) is included or not.

The autoregressive process AR(1) decomposes each parameter as shown in Equations (14), (15) and (16).

$$\mu_{(t)} = \phi_{\mu} \cdot \mu_{(t-1)} + (1 - \phi_{\mu}) \cdot \mu_m \quad (14)$$

$$\sigma_{(t)} = \phi_{\sigma} \cdot \sigma_{(t-1)} + (1 - \phi_{\sigma}) \cdot \sigma_m \quad (15)$$

$$\xi_{(t)} = \phi_{\xi} \cdot \xi_{(t-1)} + (1 - \phi_{\xi}) \cdot \xi_m \quad (16)$$

If the seasonal process, AR(12), is added to the AR representation, then the model parameters follow Equations (17), (18) and (19):

$$\begin{aligned} \mu_{(t)} = & \phi_{\mu} \cdot \mu_{(t-1)} + \lambda_{\mu} \cdot \mu_{(t-12)} + \phi_{\mu} \cdot \lambda_{\mu} \cdot \mu_{(t-13)} \\ & + (1 - \phi_{\mu} - \lambda_{\mu} + \lambda_{\mu} \cdot \phi_{\mu}) \cdot \mu_m \end{aligned} \quad (17)$$

$$\begin{aligned} \sigma_{(t)} = & \phi_{\sigma} \cdot \sigma_{(t-1)} + \lambda_{\sigma} \cdot \sigma_{(t-12)} + \phi_{\sigma} \cdot \lambda_{\sigma} \cdot \sigma_{(t-13)} \\ & + (1 - \phi_{\sigma} - \lambda_{\sigma} + \lambda_{\sigma} \cdot \phi_{\sigma}) \cdot \sigma_m \end{aligned} \quad (18)$$

$$\begin{aligned} \xi_{(t)} = & \phi_{\xi} \cdot \xi_{(t-1)} + \lambda_{\xi} \cdot \xi_{(t-12)} + \phi_{\xi} \cdot \lambda_{\xi} \cdot \xi_{(t-13)} \\ & + (1 - \phi_{\xi} - \lambda_{\xi} + \lambda_{\xi} \cdot \phi_{\xi}) \cdot \xi_m \end{aligned} \quad (19)$$

where $\mu_{(t)}$, $\sigma_{(t)}$ and $\xi_{(t)}$ correspond to the location, scale, and shape parameters of the GEV as a function of time; $\mu_{(t-n)}$, $\sigma_{(t-n)}$ and $\xi_{(t-n)}$ represent the value of each series of parameters with a lag time of n months; μ_m , σ_m and ξ_m correspond to the mean values of the parameters; ϕ_{μ} , ϕ_{σ} , ϕ_{ξ} are the coefficients of the AR(1) component; and λ_{μ} , λ_{σ} , λ_{ξ} are the coefficients of the AR(12) component.

3.2 Selection of the optimal models

The search for the optimal model explores combinations of the terms that capture seasonality, variability, and long-term trend, assuming that each time scale is well separated from the other two. Within each time scale, e.g. the seasonal one, different terms are also tested (e.g. a constant term, a yearly-period sinusoidal, etc.). To systematize this exploration, we assimilate each term to a gene, that may be expressed or not, and will thus change the behaviour of the model. Activating a gene activates and fits the corresponding parameter. A combination of genes, a gene set, will define a specific model. The performance metric of a specific model will thus be assigned to the gene set that generates that model. After testing all possible gene sets, a statistical test is carried out to verify which gene sets provide a statistically significant optimal performance, and the most parsimonious model from the optimal ones is selected.

Specific details differ for parametric and AR models, since the former are fitted via MLE, while the latter are fitted via Bayesian techniques. However, after selecting the optimal model for each category, the two are compared, and the best-performing model is selected. Moreover, as the analysis period is long enough, we

decomposed it into two sub-periods, fitting (training) the models on the first one and making estimations on the second one, to evaluate the estimation capability of each of the approaches.

3.2.1 Parametric model selection

To find the optimal model, it is necessary to explore all possible combinations of covariates. Iterative methodologies have been proposed in the literature to determine the optimal combination of covariates (Menéndez *et al.* 2009, Mínguez *et al.* 2010). However, a viable alternative is to implement a set of gene sets, that turns covariates on or off to explore all possible combinations. We propose a method equivalent to a genetic algorithm (a type of stochastic global optimization algorithm) but instead of testing only some gene sets, we test them all to find the optimum one. This data structure is inspired by biological theory, based on natural selection with binary representation (Sheppard 2018). The structure of this gene set is shown in Fig. 3. The figure shows that seasonality, variability, and trend are analysed separately, each with a different combination of covariates. The gene set is represented mathematically as a binary vector (θ) of 17 components, one for each covariate that may be activated. Each possible combination of the components of the gene set corresponds to a unique model. It is important to note that the coefficients α_{11} , β_{11} and γ_{11} (Equations 3, 4, 5) represent the mean value of each parameter and are always active to guarantee the stability of the model.

Due to the large number of possible gene sets (2^{17}) and the time it takes to fit each one of them to all the selected raingauges, we divide the fitting procedure into three steps: first we define the optimal seasonal components (panel (a) of Fig. 3, totalling 32 (2^5) models), then the optimal variability component (panel (b) of Fig. 3, totalling 64 (2^6) models) and, finally, the optimal trend component (panel (c) of Fig. 3, totalling 64 (2^6) models).

To avoid over-parameterization, and thus reduce overfitting, for model selection we used a procedure analogous to that in other studies (Menéndez *et al.* 2009) that incorporate the minimization of the AIC. In this way, the selected model will be one that generates a good fit and at the same time is the simplest possible model (preserving the principle of parsimony and reducing overfitting to a minimum). The criterion is described by Equation (20):

$$AIC = 2(n_p) - 2\ell(x, t; \theta) \quad (20)$$

where n_p is the number of covariates (sinusoidal functions, trends, and teleconnection indices) active in the gene set, and $\ell(x, t; \theta)$ is the value of the likelihood function for the optimal parameters of the given gene set. The lower the number calculated with this criterion, the better the model represented by the given gene set is.

The specific procedure followed to carry out this analysis (panel (b) of Fig. 2) proceeds as follows:

- (1) A series of monthly maxima is created for all gauges. All models will be fitted to all the time series.
- (2) The 32 models that introduce seasonality (represented by the corresponding 32 components of the vector θ) are fitted to the monthly maxima time series using MLE. AIC is computed for each one of them. Gene

(a) SEASONALITY	
Location	$\mu_s(t) = \alpha_{11} + \alpha_{21} \cos(2\pi \cdot t) + \alpha_{22} \sin(2\pi \cdot t) + \alpha_{23} \cos(4\pi \cdot t) + \alpha_{24} \sin(4\pi \cdot t)$
Scale	$\sigma_s(t) = \beta_{11} + \beta_{21} \cos(2\pi \cdot t) + \beta_{22} \sin(2\pi \cdot t) + \beta_{23} \cos(4\pi \cdot t) + \beta_{24} \sin(4\pi \cdot t)$
Shape	$\xi_s(t) = \gamma_{11} + \gamma_{21} \cos(2\pi \cdot t) + \gamma_{22} \sin(2\pi \cdot t)$
(b) VARIABILITY	
Location	$\mu_v(t) = \mu_s(t) + \alpha_{31} \cdot MOI + \alpha_{32} \cdot NAO$
Scale	$\sigma_v(t) = \sigma_s(t) + \beta_{31} \cdot MOI + \beta_{32} \cdot NAO$
Shape	$\xi_v(t) = \xi_s(t) + \gamma_{31} \cdot MOI + \gamma_{32} \cdot NAO$
(c) TREND	
Location	$\mu_T(t) = \mu_v(t) + \alpha_{41}t + \alpha_{42}t^2$
Scale	$\sigma_T(t) = \sigma_v(t) + \beta_{41}t + \beta_{42}t^2$
Shape	$\xi_T(t) = \xi_v(t) + \gamma_{41}t + \gamma_{42}t^2$
d. GENE SET AND COVARIATES	
Gene set	$\theta = [\overset{1}{\color{red}}, \overset{1}{\color{green}}, \overset{1}{\color{blue}}, \overset{1}{\color{red}}, \overset{1}{\color{blue}}]$
Covariates	$\Delta = [\alpha_{11}, \alpha_{21}, \alpha_{22}, \alpha_{23}, \alpha_{24}, \alpha_{31}, \alpha_{32}, \alpha_{41}, \alpha_{42}, \beta_{11}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{24}, \beta_{31}, \beta_{32}, \beta_{41}, \beta_{42}, \gamma_{11}, \gamma_{21}, \gamma_{22}, \gamma_{31}, \gamma_{32}, \gamma_{41}, \gamma_{42}]$

Figure 3. Summary of the parametric model terms, grouped by non-stationarity type. Panel (a) presents the proposed seasonal component, where the parameters of the generalized extreme value distribution (GEV) are expressed as a function of harmonic functions. Panel (b) describes the climate variability component, introduced by atmospheric indices. Panel (c) shows the trend component. Panel (d) describes the vector θ , which represents the gene set that activates or deactivates the covariates and their coefficients, represented in the vector Δ .

set 0, where all components are deactivated, is the stationary model; all other gene sets incorporate non-stationarity (seasonality in this case).

- (3) A ratio is computed among the AIC values of all non-stationary models and the stationary one ($AIC_{ratio} = AIC_{Non-Stationary}/AIC_{Stationary}$). Only those gene sets that show a better performance than the stationary one ($AIC_{ratio} < 1$) in at least 95% of the stations are considered for further analysis.
- (4) The distribution of the AIC_{ratio} for the selected gene sets are compared pairwise using a Kolmogorov-Smirnov test (Stephens 1992) to determine whether the samples may come from the same distribution or not. Bonferroni corrections are applied to account for multiple comparisons (Bonferroni 1936). From all the possible optimal gene sets (those with the lowest median AIC_{ratio}) that are indistinguishable in performance (those whose AIC_{ratio} may come from the same distribution), we select the one with the fewest parameters. Note here that another principle of parsimony is used: a model that was optimal over a larger number of climate zones was preferred over another with fewer parameters but optimal for fewer climate zones. We should not forget that all these parameters are indistinguishably optimal, and thus the global parsimony principle is preferred over the local one.
- (5) The same procedure is followed for gene sets representing climate variability and trend. For each new category, the previously determined optimal gene sets are kept, adding only the new components to the vector. Moreover, the AIC_{ratio} is not computed over the stationary model anymore, but using the optimal model of the previous category as a reference. Following this procedure, we determined the overall optimal parametric model.

3.2.2 AR model selection

The method used to select the optimal gene set from the AR approach is similar to the parametric model selection. However, in this case, only two elements are considered: an AR component and a seasonal one. Since the AR term may model long-range correlations, we hypothesize that it may naturally incorporate climate variability, making it unnecessary to explicitly incorporate a climate variability component. The AR term will only consider a lag-1 component, while the seasonal term will be included using a lag-12 component. Figure 4 shows a scheme of the different terms (genes) that are explored. In this case, the gene set has six components: three for the AR part and three for the seasonal one: two for each GEV parameter.

It is important to note that the seasonal model components are only active when the AR components are active. Therefore, for one parameter there are three possibilities: neither the AR nor the seasonal terms are active, only the AR term is active, or both the AR and the seasonal terms are active. Each parameter admits three options, and as we have three parameters, the total number of models in this case is 27.

Moreover, Bayesian techniques are used for fitting the AR models. These techniques require setting a prior distribution for the parameters to be fitted. In this study, we have used the following prior distributions:

$$\mu \sim N(100, 20)$$

$$\sigma \sim N(40, 8)$$

$$\xi \sim \text{beta}(4, 4)$$

(a) AUTOREGRESSIVE MODEL	
Location	$\mu_A(t) = \phi_\mu \cdot \mu_{t-1} + (1 - \phi_\mu) \cdot \mu_m$
Scale	$\sigma_A(t) = \phi_\sigma \cdot \sigma_{t-1} + (1 - \phi_\sigma) \cdot \sigma_m$
Shape	$\xi_A(t) = \phi_\xi \cdot \xi_{t-1} + (1 - \phi_\xi) \cdot \xi_m$
(b) SEASONALITY MODEL	
Location	$\mu_S(t) = \mu_A(t) + \lambda_\mu \cdot \mu_{t-12} - \phi_\mu \cdot \lambda_\mu \cdot \mu_{t-13} + (\lambda_\mu \cdot \phi_\mu - \lambda_\mu) \cdot \mu_m$
Scale	$\sigma_S(t) = \sigma_A(t) + \lambda_\sigma \cdot \sigma_{t-12} - \phi_\sigma \cdot \lambda_\sigma \cdot \sigma_{t-13} + (\lambda_\sigma \cdot \phi_\sigma - \lambda_\sigma) \cdot \sigma_m$
Shape	$\xi_S(t) = \xi_A(t) + \lambda_\xi \cdot \xi_{t-12} - \phi_\xi \cdot \lambda_\xi \cdot \xi_{t-13} + (\lambda_\xi \cdot \phi_\xi - \lambda_\xi) \cdot \xi_m$
(c) GENE SET AND COVARIATES	
Gene set	$\theta = [1, 1, 1, 1, 1, 1]$
Coefficients	$\Delta = [\phi_\mu, \lambda_\mu, \phi_\sigma, \lambda_\sigma, \phi_\xi, \lambda_\xi]$

Figure 4. Summary of the autoregressive model terms, grouped by non-stationarity type. Panel (a) shows the autoregressive component, where the time dependence of the generalized extreme value distribution (GEV) parameters is expressed by a time series model. Panel (b) shows the seasonal component, which is included by means of a 12-month lagged autoregressive term. Panel (c) describes the vector θ , which represents the gene set that activates or deactivates the covariates and their coefficients, represented in the vector Δ . Note that the coefficients μ_m , σ_m and ξ_m are always kept active to guarantee the stability of the models and represent the mean value of each parameter.

$$(\phi, \lambda) \sim \text{beta}(2, 4)$$

The selected distributions allow a wide range of possibilities that do not limit the model to specific values. In addition, each distribution includes information about reasonable expected values. Imposing a prior distribution is equivalent to constraining the value of the parameters in an optimization problem.

The specific procedure followed to carry out this analysis (panel (c) of Fig. 2) proceeds as follows:

- (1) A series of monthly maxima is created for all gauges. All models will be fitted to all the time series.
- (2) The time series of monthly maxima are fitted for each gauge, obtaining the posterior distribution of the parameters $((\phi_\mu, \phi_\sigma, \phi_\xi), (\lambda_\mu, \lambda_\sigma, \lambda_\xi))$, instead of an optimal set of parameters. For this reason, the median AIC is computed for each gene set, and the procedure described in the previous section (testing the hypothesis that the AIC ratio follows the same distribution) is followed to determine the optimal models.

3.3 Comparing parametric and AR models

Once the optimal parametric and AR models have been selected, the interest turns towards determining whether one category of models performs better than the other one. To achieve this goal, two different procedures are proposed: one based on AIC ratios and another one on estimating future precipitation.

3.3.1 AIC ratios method

The first procedure uses AIC ratios to compare the parametric and AR approaches. The performance of the stationary model

must be equal whether computed through MLE or Bayesian techniques (since it is the exact same model in both cases), but the former results in just a number, and the latter in a complete distribution. Therefore, to compare the AIC of the two approaches, the AIC of the AR approach must be summarized into a number (we selected the median) and then multiplied by a normalization factor. This factor explains the change between the performance of the stationary model fitted by MLE and the median of the performance of the stationary model obtained by Bayesian techniques.

The normalization factor is the ratio between the AIC of the stationary model for the parametric approach (MLE) and the median of the AIC distribution for the stationary model for the AR approach (Bayesian technique), as shown in Equation (21).

$$NR = \frac{AIC_{Sta}^{PAR}}{AIC_{Sta}^{AR}} \quad (21)$$

The renormalized AIC values for the optimal AR model and the AIC values of the parametric model are compared using a Kolmogorov-Smirnov test. This test determines whether the distributions associated with the AIC values are different, and defines which of the approaches is more appropriate to model the non-stationarity in extreme monthly rainfall.

3.3.2 Estimating future precipitation

Since one of the objectives of this work is to determine an optimal estimation model, we propose to compare the optimal model in each category using their estimation accuracy for future rainfall. To make this comparison, the global period of analysis is divided into two: a first period that covers the years 1979–1998 and a second one that covers the

years 1999–2019. The first period is used to fit the optimal models. These models are then used to estimate rainfall in the second period. Another fitting will be carried out in the second period to validate the estimations obtained from the first fitting.

Based on this decomposition of the analysis period, two different statistical characterizations will be compared:

- (1) **Historical fit:** fit of the optimal models to data of the second period (1999–2019), summarized by the estimations that it provides for the mid-year of the range (corresponding to January 2009).
- (2) **Estimation fit:** fit of the optimal models to data of the first period (1979–1998), summarized by the estimations that it provides for the mid-year of the second period (corresponding to January 2009).

The comparison will be focused on the values of the precipitation corresponding to a return period of 10 years. In the parametric case, the historical fit and the estimation fit provide a value for the 10-year return period rainfall for each gauge. Thus, the accuracy is characterized by the difference between the two values. In the AR case, the rainfall for the 10 year return period is a distribution. The comparison is thus carried out analysing the differences between 5000 samples obtained by sampling the historical and the estimation fit distributions.

3.4 Software tools used

Parametric models were fitted using the global black box optimization module `BlackBoxOptim.jl` available in the Julia programming language packages (Bezanson *et al.* 2017). Bayesian methods were fit using `PyStan` (Stan Dev Team 2021) and Python.

`PyStan` uses the *MCMC* method, which is an alternative to conventional numerical methods. We used `PyStan` with four chains of size $N = 1000$, applying a burning period of 2000 samples. The construction of the chain is done using the Metropolis-Hastings algorithm.

4 Results

4.1 Optimal parametric model

The results of the selection of the optimal parametric model are shown in Fig. 5. Panel (a) shows that the non-stationarity of extremes in all of Spain can be optimally captured by a specific model form (area shaded with small circles; green in the colour figure). Only in climates of type BWh, in the most arid regions, does the stationary model do a better job.

Panels (b) to (i) show box plots for the AIC of the best-performing model for each climate type. In those panels, *ST* refers to the stationary model; *SE* includes seasonality; *SE + V* includes both seasonality and variability; and *SE + V + T* includes the seasonality, variability, and trend. The panels also include a dashed horizontal line showing the minimum median AIC (the median AIC for the best model) of all the comparisons for a climate.

The best-performing model accounts for seasonality through two sinusoidal components, one of yearly period

and another with a period of six months, in the location and scale parameters, with no seasonality in the shape parameter. It also requires including the effect of MOI and NAO in both the location and shape parameters to account for climate variability. A linear and a quadratic trend are required for the location parameter, and only a linear one is required for the scale and shape parameters. The optimal model is described in Equations (22), (23) and (24).

$$\begin{aligned} \mu_{(t)} = & \alpha_{11} + \alpha_{21}\cos(2\pi t) + \alpha_{22}\sin(2\pi t) + \alpha_{23}\cos(4\pi t) \\ & + \alpha_{24}\sin(4\pi t) + \dots + \alpha_{31}MOI + \alpha_{32}NAO + \alpha_{41}t + \alpha_{42}t^2 \end{aligned} \quad (22)$$

$$\begin{aligned} \sigma_{(t)} = & \beta_{11} + \beta_{21}\cos(2\pi t) + \beta_{22}\sin(2\pi t) + \beta_{23}\cos(4\pi t) \\ & + \beta_{24}\sin(4\pi t) + \dots + \beta_{31}MOI + \beta_{32}NAO + \beta_{41}t \end{aligned} \quad (23)$$

$$\xi_{(t)} = \gamma_{11} + \gamma_{41}t \quad (24)$$

Figure 5 shows that including every additional component (seasonality, variability, and trend) improves the results for all climates each time, except for climates BWh and BSh. Climate BSh only improves once the trend term is included, but the improvement is noticeable. Climate BWh does not improve with any addition, showing that the stationary model is the best-performing one for that climate.

The results of the optimal parametric model indicate that the non-stationarity analysis can be optimally included, almost exclusively, in the location and scale parameters. The influence of seasonality, variability and trend is relevant in all the climatic groups except for the BWh group, which represents dry desert zones located in small areas of the southeast of the Iberian Peninsula and in the Canary Islands (Fig. 5(a)), where the stationary model does a better job. These regions coincide with the minimum pluviometric values of the peninsula (AEMET and IPMA 2011), which may indicate that not enough extreme events exist to optimally fit a non-stationary model.

4.2 Optimal AR model

Of the 27 models tested, only five provided reasonable fits, which were:

- Stationary model.
- μ -AR(1): location parameter modelled with the autoregressive term.
- μ -AR(1/12): location parameter modelled with the autoregressive and the seasonal terms.
- σ -AR(1): scale parameter modelled with the autoregressive term.
- σ -AR(1/12): scale parameter modelled with the autoregressive and the seasonal terms.

Figure 6 shows the AIC results for each climate (panels (b) to (i)), as well as the spatial distribution of the best-fitting gene sets (panel (a)). The figure shows that model μ -AR(1/12) is the best performing in climates BSk, Csa, BSh and BWh; σ -AR(1)

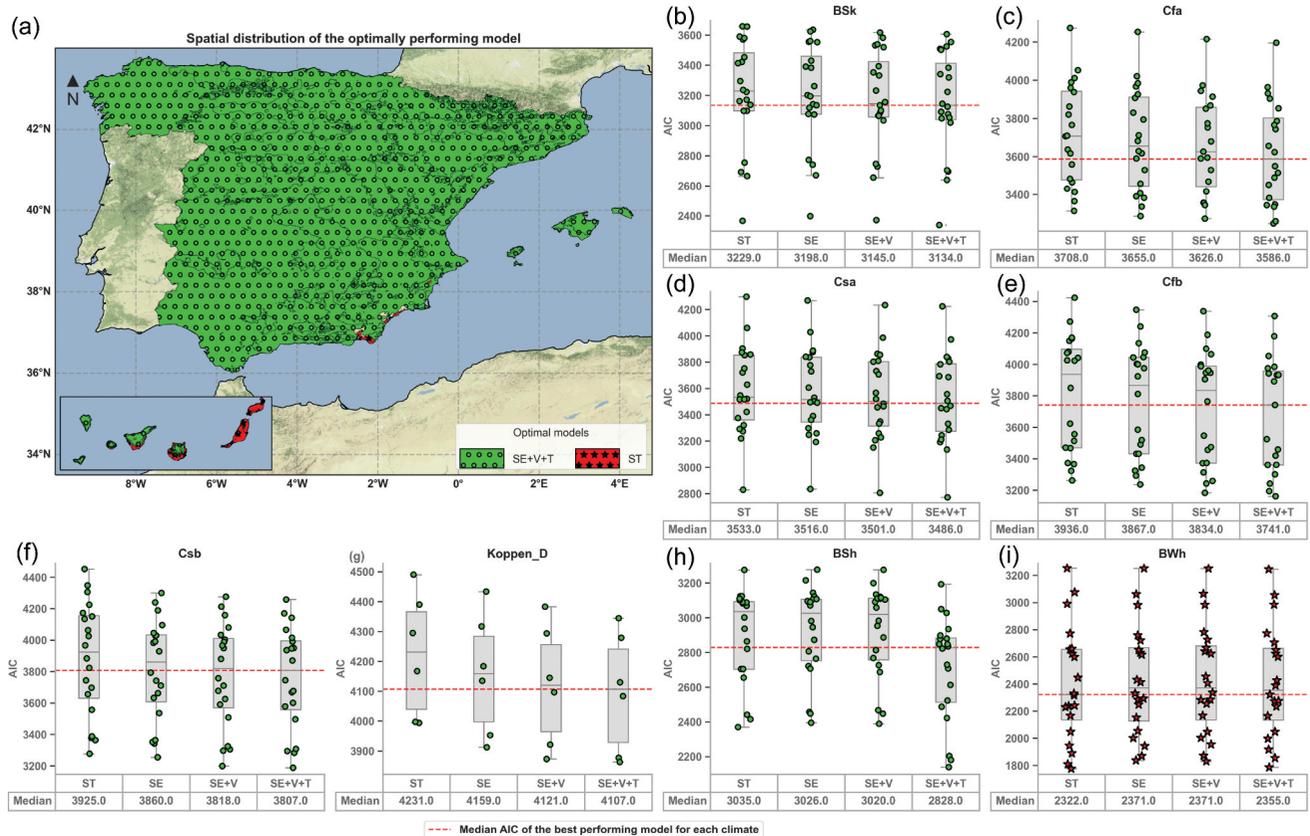


Figure 5. Spatial distribution of the optimal parametric model and performance evaluation of different parametric models for each climate. Panel (a) shows the spatial distribution of the optimal model. The shaded area with little circles (covering the BSh (hot steppe), Bsk (cold steppe), Cfa (temperate with a dry season and hot summer), Csa (temperate with dry or hot summer), Cfb (temperate with a dry season and temperate summer), Csb (temperate with dry or temperate summer) and Köppen_D climates) represents the region over which the variability of extremes is best captured by including seasonality, variability, and trend ((SE + V + T)). This area is optimally represented by a specific model. The shaded area with small stars (it only includes climate BWh (hot desert)) is best represented by a stationary model. Panels (b) to (i) present the box plots that show the performance in terms of Akaike Information Criterion (AIC) of the optimal parametric models obtained increasing the complexity of the model. ST refers to the stationary model, SE to the seasonal model, SE + V to the model that introduces seasonality and variability, and SE + V + T to the complete model. The horizontal dashed line shows the minimum median AIC of the box plots (the AIC of the best performing model). A table containing the values of the median AIC per gene set and per climatic sub-group is added at the bottom of each sub-figure.

in climate Csb; and σ -AR(1/12) in climates Cfa, Cfb and D. μ -AR(1) is never the best-performing model. The results for the stationary model are not shown because it takes AIC values higher (indicating worse performance) than the other gene sets tested.

The results indicate that in the AR framework the non-stationarity is captured in the location and scale parameters and does not require incorporating any modification in the shape parameter.

AR and seasonal terms are needed to properly capture the behaviour of extremes over the whole region analysed, except for areas of climate Csb, with markedly drier periods in summer (AEMET and IPMA 2011), where the seasonality term may be less important. Note, however, that the differences between σ -AR(1) and σ -AR(1/12) for the climate Csb are small, so additional experiments may be required to properly account for the need of a seasonal component.

4.3 Comparing the optimal parametric and AR models

Figure 7 compares the AIC distributions for the best performing model in the two approaches (parametric and AR) for each climatic zone. Note that the AIC for the AR model has been

represented by its median, which has been normalized (see Equation 21) to perform a fair comparison between the two approaches. The figure clearly shows that AR models outperform parametric models for every climate analysed. The distributions are, indeed, so different that no overlap exists, clearly showing the superiority of the AR approach to account for the non-stationary behaviour of extreme rainfall events in Spain.

Regarding the variability of the results and considering the scale observed, a very low dispersion can be seen, with no presence of atypical data, which marks another difference between the two approaches analysed, demonstrating the strength of the evidence in favour of the AR models. Because the AR approach is the optimal one, the same homogeneous climatic zones (see panel (a) in Fig. 7) described in section 4.2 are identified.

To complement the analysis and present additional evidence for the performance of the AR approach, Fig. 8 compares estimations for the rainfall with a return period of 10 years. The comparison presents the estimates obtained from the optimal AR model (distributions) and from the optimal parametric models (point estimates in the lower part of each panel of the figure). Estimates are computed using the historical fit and the estimation fit defined in section 3.3.2.

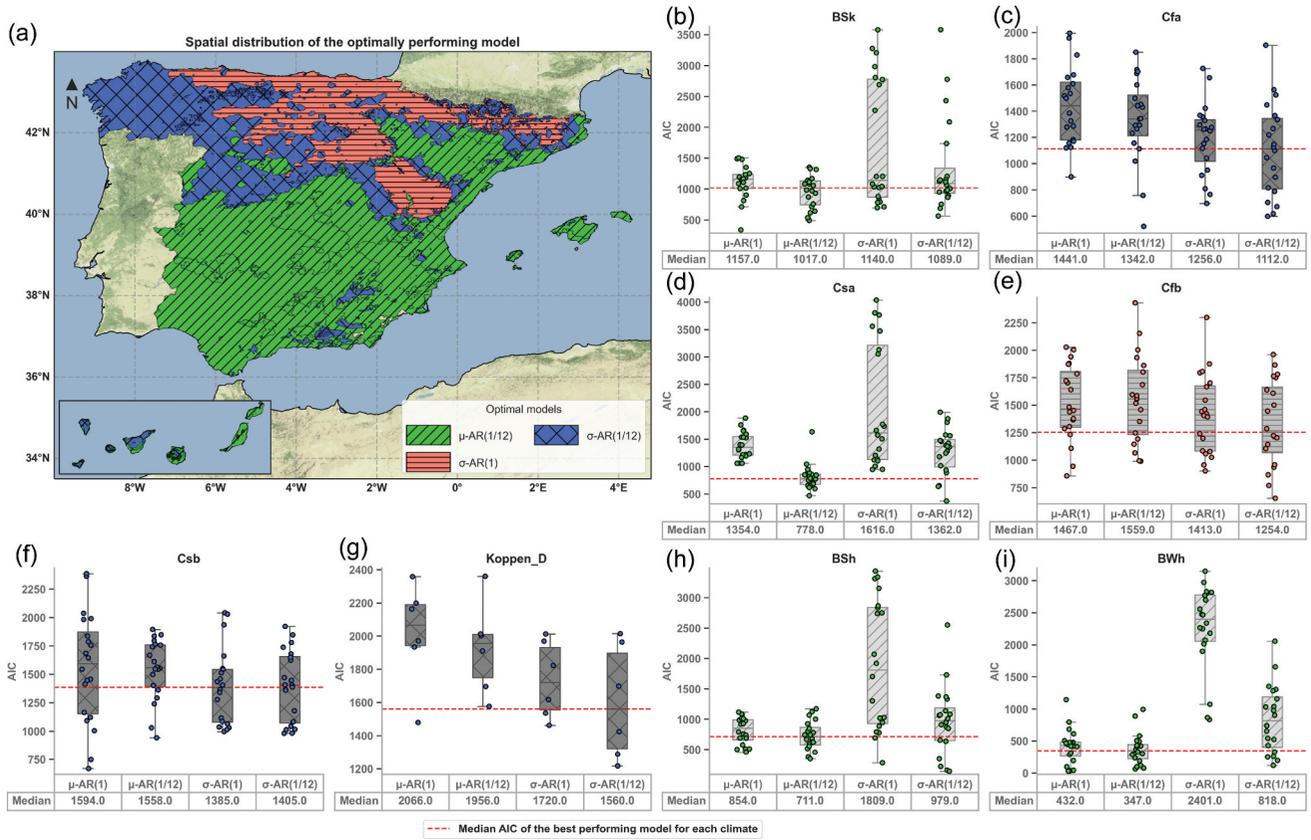


Figure 6. Spatial distribution of the optimal autoregressive model and performance evaluation of different autoregressive models for each climate. Panel (a) shows the spatial distribution of the optimal model. The area shaded with inclined lines (45°) is the area where μ -AR(1/12) is the optimal model (climates Bsk (cold steppe), Csa (temperate with dry or hot summer), BSh (hot steppe) and BWh (hot desert)); the area shaded with horizontal lines for σ -AR(1) (climate Cfb (temperate with a dry season and temperate summer)); and the one with crossed lines for σ -AR(1/12) (climates Cfa (temperate with a dry season and hot summer), Csb (temperate with dry or temperate summer) and D). Panels (b) to (i) show the box plots of median Akaike Information Criterion (AIC) for each gene set and climate type. The horizontal dashed line in box plots shows the minimum median AIC for the best-performing model for each climate. Note that the box plots are shaded and correspond to the described areas.

The figure only shows one selected rain gauge per climate to illustrate the differences between the two approaches. Moreover, the results apply to the value of the return period for January 2009, which is the midpoint of the analysed date range.

Figure 8 shows that both approaches provide, in general, compatible results, except for the most arid climate (BWh, centre and lower right panels), where the stationary parametric model predicts values that are 66% larger than the ones predicted by the optimal AR model; and for climate Cfa where the parametric estimations are outside of the range (historical fit: 22 mm and estimation fit: 23 mm). The AR models produce more coherent estimations, where the accord between the historical and the estimation fits seems better than for the parametric case.

In general terms, the distributions of the AR estimations for a given climate show a similar shape, with almost a perfect match between them (climate Cfb), with a small bias (climates Cfa and Csb) or with a larger one (climate Bsh), but still below 10%. The distributions for climates Csa and D show the largest differences, but still provide compatible predictions with biases below 10%. In absolute terms, the smallest differences are close to 0 (0.1 mm), while the maximum is slightly smaller than 5 mm (corresponding to the station of VILLAFRANCA BOSCANO, centre panel in Fig. 8).

However, the results of Fig. 8 refer only to one station (one with an average quality fit). To evaluate whether the two approaches perform differently, Fig. 9 shows a systematic comparison of estimated results for all the stations considered for each climate. The box plots shaded with inclined lines (45°) show the difference, at a specific station, between the 10-year return period values calculated using the historical fit and the estimation fit (defined in section 3.3.2) for the parametric approach. Each box plot summarizes the results for a different climate. Similarly, the shaded box plots with dots indicate the same difference, but using the AR approach. The VILLAFRANCA BOSCANO station of the Csa sub-group was excluded from the graph (it presented a difference of -4.7 mm) as it was the only station where the difference exceeded the lower limit of the graph (-1.5 mm).

Figure 9 shows that the parametric models generate predictions with a greater range of dispersion than the AR ones. Its bias is small, although larger than that of the AR models. Most of the parametric model results are within the safe zone (the area in Fig. 9 that is shaded with horizontal lines), the region where the values calculated from the estimation fit are larger than the values calculated from the historical fit. However, excursions into the unsafe zone (the area in Fig. 9 that is shaded with inclined lines (135°)) are more numerous and larger for the parametric approach than for the AR one.

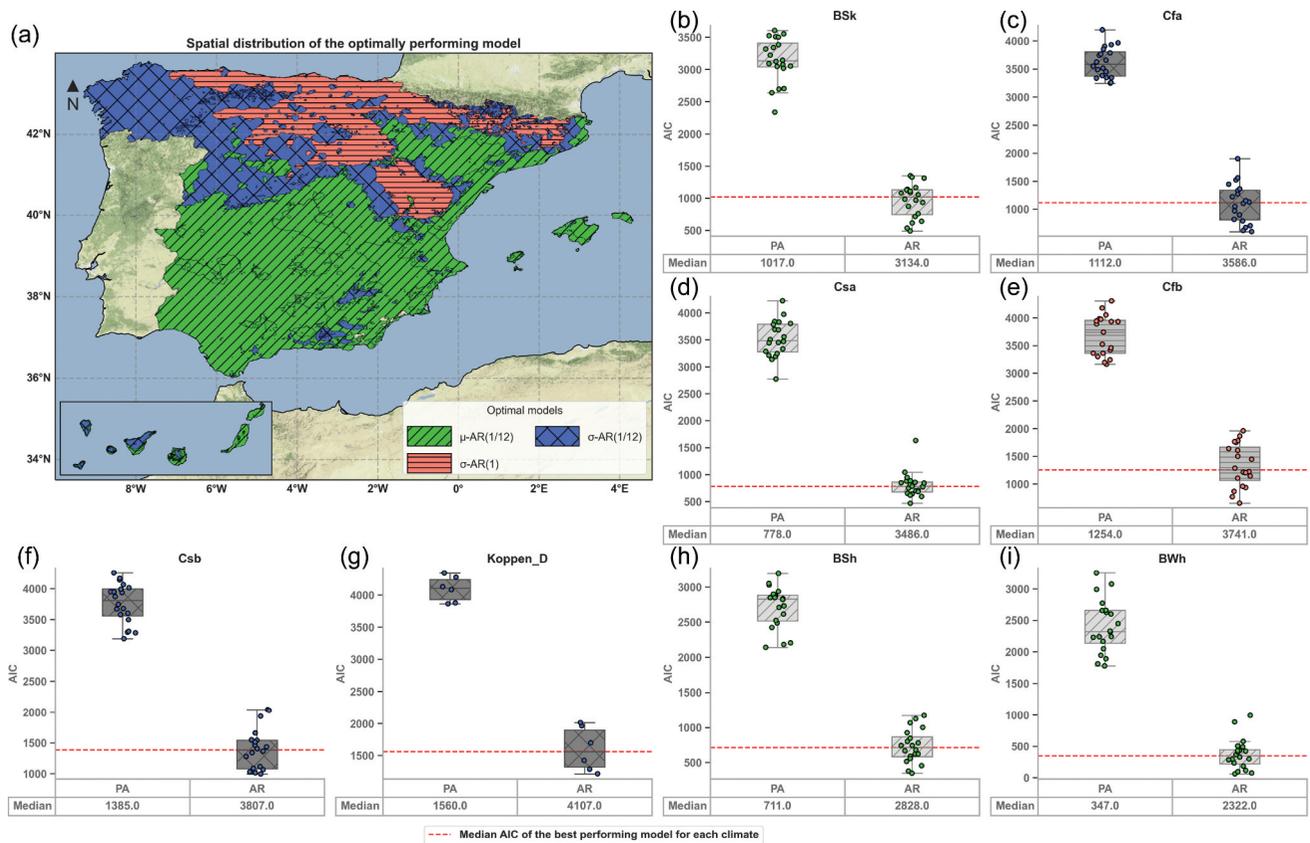


Figure 7. Spatial distribution of the optimal approach and comparison of performance between the parametric and the autoregressive approach. Panel (a) shows the spatial distribution of the optimally performing approach. Since the autoregressive approach outperforms the parametric model every time, the panel shows the same distribution as in Fig. 6. Panels (b) to (i) show the box plots of the Akaike Information Criterion (AIC) distribution for the optimal parametric and the optimal autoregressive models. The figure shows that the autoregressive models clearly outperform the parametric model for all climate types. Note that the AIC for the autoregressive model has been normalized (see Equation 21) to allow a fair comparison between the two approaches.

In fact, AR models show a lower bias than the parametric ones (the median of the differences is closer to zero) and a lower variance since most of the predictions cluster closely around the zero-difference error. The only climate where the two approaches are comparable is for the BWh climate because here the optimal model is the stationary one. For all the other climates, the results of the AR model are superior.

5 Discussion

The optimal parametric model was identified in terms of goodness of fit and complexity, favouring models that perform better over more types of climates. The results of this approach indicate models that introduced the non-stationary process in the GEV distribution parameters obtained a better representation, with respect to the stationary parametric models, in seven of the eight climatic zones studied. The introduction of seasonality, variability, and trend in the GEV parameters generated an improvement in the adjustment of said distribution to extreme precipitation events, except for the BWh sub-group, which represented dry desert zones located in small areas of the southeast of the Iberian Peninsula and the Canary Islands, where the stationary model obtained a better representation.

In the proposed models, we assume that the non-stationarity introduced by climate change is captured in the trend, since this is the only inter-annual variability included in the model and not explained by climate indices. It is important to consider this trend because it provides a long-term overview of the changes that could occur. In the case of the selected parametric model, it is framed in the linear term, t , included in the location, scale, and shape parameters of the selected model. In addition, the location parameter contains a quadratic term, t^2 , showing that this parameter represents a trend with greater magnitude than the others. Since our main interest is short-term estimation, where climate models predict little to no change, we assume that climate change introduces a drift in the model parameters to extrapolate the trends from the recent past, assuming that other climate change effects (like changes in seasonality, etc.) may be neglected.

In the AR models, the non-stationarity introduced by climate change could not be captured in the same way. In this case, the trend term did not represent an improvement in terms of fit with respect to the AR and seasonal components alone. Indeed, this result may indicate that long-term trends in extremes captured in the parametric models could be an artefact, but further analyses will be required to deepen our understanding of this specific result.

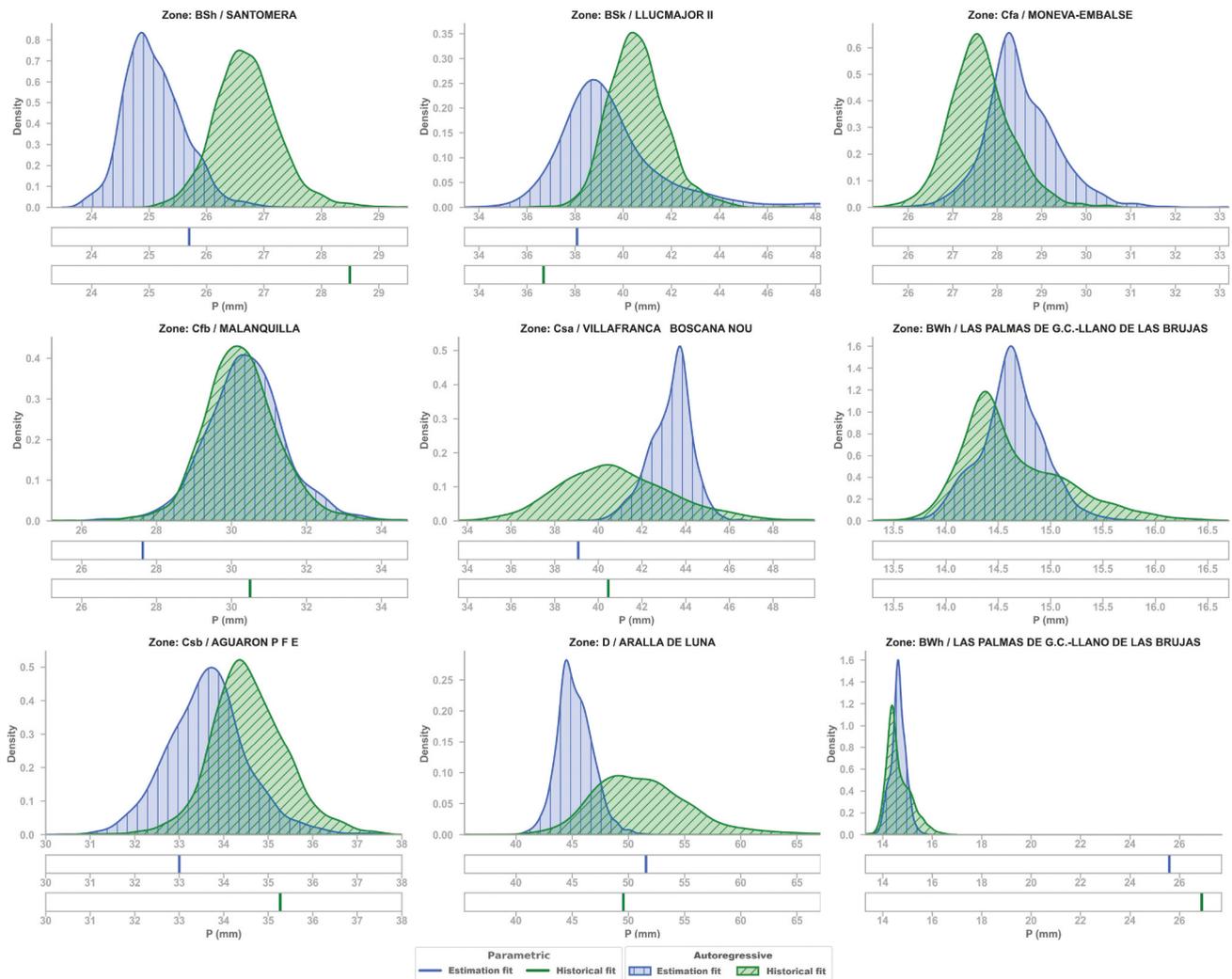


Figure 8. Comparison of the 10-year return period rainfall for January 2009 obtained from the historical and estimation fits for a representative gauge for each climate. The distributions result from sampling 2000 values of the corresponding optimal autoregressive model. The density function shaded with vertical lines represents estimations based on data from a previous period (estimation fit; see section 3.3.2). The density function shaded with sloped lines represents the rainfall estimated within the same period (historical fit; see section 3.3.2). Below the axis of the distribution plot, there are two additional axes that show the estimates of the corresponding optimal parametric model: the upper one showing the estimate from the estimation fit, and the lower one showing the estimates from the historical fit. Note that the BWh (hot desert) zone is represented in two sub-figures (bottom right and centre right), which contain the same distributions; however, in the bottom right it can be seen that this climatic zone has a greater difference between the estimates of the two models compared. The upper right sub-figure does not contain the estimates of the parametric model, because they are outside the range (historical fit: 22 mm and estimation fit: 23 mm).

The optimal AR models obtained a good representation of the extreme precipitation events in all climates. Indeed, these models obtained favourable results when the non-stationarity was included in the location parameter or in the scale parameter. The shape parameter, however, had to be kept constant to ensure the convergence of the models, probably because the increased expressivity of the model required many more data points to converge to a unique solution.

The results of the AR approach indicate that the monthly maximum precipitation events are influenced by extreme events in previous months, as well as by the seasonal dependence of climate. The effect is best incorporated into the scale parameter, although some regions are best captured by introducing non-stationarity in the location parameter. This behaviour was observed in all the climatic zones studied, being slightly less marked in the seasonal component of the Csb subgroup, which represented zones with markedly dry periods in summer (AEMET and IPMA 2011).

The two approaches considered in the study (parametric and AR) were used to estimate the maximum annual precipitation associated with a return period of 10 years in the analysed stations. According to the results obtained, the AR approach seems more accurate for probabilistic estimations of extreme precipitation events. Even when the two approaches constitute a tool to consider the dependencies between extreme values and the temporal evolution of the climate, the AR model provides more flexibility, resulting in more accurate estimations.

The application of non-stationary techniques to the characterization of extreme events is of great importance to managers and practitioners, who may generate more accurate descriptions of the forcing for their designs and systems. The results that we provide may guide and simplify the application of these methods, and thus assist in the management of irrigation and hydroelectric power plants in the Spanish territory. Moreover, our experimental

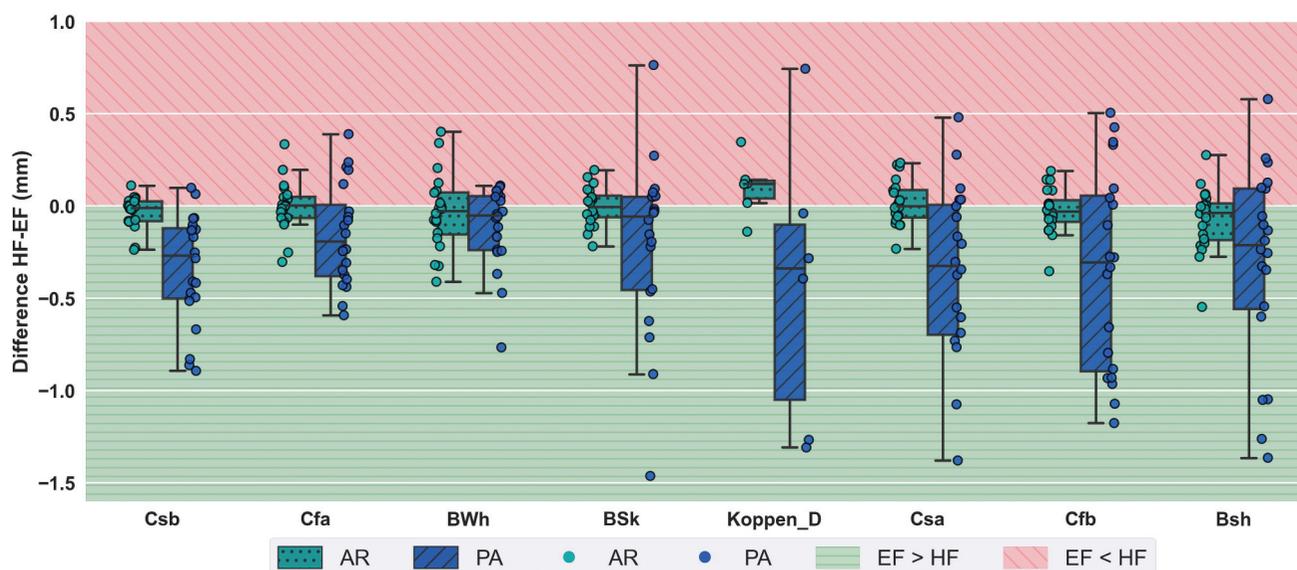


Figure 9. Differences in the estimated 10-year return period rainfall between the estimation fit and historical fit (defined in section 3.3.2) for every station of each climate. The box plots shaded with inclined lines (45°) show the differences for the parametric approach, and the box plots shaded with dots show those for the autoregressive approach. Each box plot summarizes the results for the respective climate. The area where the estimation fit (EF) estimates are larger than the historical fit (HF) ones is shaded with horizontal lines, as it is a safe region (estimates based on the past are larger than those based on current values). The area where historical fit estimates are larger is shaded with inclined lines (135°), as it is the unsafe zone.

design may be used to analyse other rainfall networks and find the optimal ways in which climate variability may be incorporated.

6 Conclusions

The AR models perform better in terms of quality of fit and complexity (AIC) than the parametric models for all the analysed climates. The results indicate that the AR models best capture and characterize the temporal variability of extreme precipitation events.

The improvement in terms of AIC of the AR approach can probably be attributed to two factors: first and most important, the structure and flexibility provided by the AR and seasonal terms, capable of modelling long-range correlations (lag up to 12 months); second, the fitting method, since the Bayesian method does not have overfitting problems.

The AIC differences between the AR and parametric models are driven by the likelihood function. However, the number of parameters – although not a determining factor – show large differences. Parametric models need about 19 parameters, while AR models need at most four parameters. AR models achieve a higher level of expressiveness with a much smaller number of parameters. This fact also contributes to the observed differences.

In terms of prediction, the optimal AR model is more accurate than the optimal parametric one (see Fig. 9), even though the differences in predictive performance, which are important, are smaller than in terms of AIC, which are remarkable (see Fig. 8). Associating AIC results with prediction performance might be intuitive, but it is not a correct comparison, since AIC is not a suitable metric to measure prediction goodness of fit (Gelman *et al.* 2014). In fact, Gelman *et al.* (2014) recognize that it is useful to compare very dissimilar models, and for this, predictive comparisons can make sense. In the proposed models, we

assume that the non-stationarity introduced by climate change is embedded in the trend, since this is the only term that considers inter-annual variability beyond that explained by climate indices. In the case of the selected parametric model, it is framed in the linear term, t , present in the location, scale, and shape parameters of the selected model. In addition, the location parameter contains a quadratic term t^2 , showing that this parameter represents a trend with greater magnitude than the others.

In the AR models, the non-stationarity introduced by climate change could not be captured in the same way. In this case, the trend term did not represent an improvement in terms of fit with respect to the AR and the seasonal component introduced in the location and scale parameters of the selected model.

Long-term trends, which represent climate change effects, were only important for the parametric models, and not for the AR ones. Indeed, a trend component was necessary in the three parameters for the parametric approach, while AR models performed better without this term. This result may indicate that trends in monthly maxima may be a mathematical artefact born from the interaction between a somewhat rigid seasonality term and the climate indices. Further analyses will be required to clarify the real causes of this result.

The proposed methodological framework constitutes a tool to analyse the temporal dependence of extreme precipitation events, including estimation under non-stationary conditions. The application of the proposed methodology provides a solid basis to address problems associated with the design of hydraulic structures and the management of water resources. Further lines of research may focus on the application of other statistical distributions that integrate different covariates to represent non-stationarity, including the development of a new framework for risk assessment.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work received financial support from the Government of Cantabria through the Fénix Programme and from Grant RTI2018-096449-B-I00, funded by MCIN/AEI/10.13039/501100011033 and by “ERDF A way of making Europe.”

ORCID

Diego Urrea Méndez  <http://orcid.org/0000-0001-6284-3087>

Manuel del Jesus  <http://orcid.org/0000-0003-0703-8960>

References

- Adlouni, S.E., *et al.*, 2007. Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research*, 43 (3), 3410. doi:10.1029/2005WR004545
- AEMET and IPMA, 2011. Atlas Climático Ibérico. In: *Iberian climate atlas*. Spain: Madrid, 15–18.
- Akaike, H., 1998. Information Theory and an Extension of the Maximum Likelihood Principle. In: *Springer Series in Statistics (Perspectives in Statistics)*, New York, NY.
- Bezanson, J., *et al.*, 2017. Julia: a fresh approach to numerical computing. *SIAM Review*, 59 (1), 65–98. doi:10.1137/141000671
- Bonferroni, C.E., 1936. *Teoria statistica delle classi e calcolo delle probabilità*. Vol. 8. Firenze, Italy: Pubblicazioni del R. Istituto superiore di scienze economiche e commerciali di Firenze.
- Bracken, C., *et al.*, 2016. Spatial Bayesian hierarchical modeling of precipitation extremes over a large domain. *Water Resources Research*, 52 (8), 6643–6655. doi:10.1002/2016WR018768
- Bracken, C., *et al.*, 2018. A Bayesian hierarchical approach to multivariate nonstationary hydrologic frequency analysis. *Water Resources Research*, 54 (1), 243–255. doi:10.1002/2017WR020403
- Bradley, A.A., 1998. Regional frequency analysis methods for evaluating changes in hydrologic extremes. *Water Resources Research*, 34 (4), 741–750. doi:10.1029/98WR00096
- Brown, S.J., Caesar, J., and Ferro, C.A.T., 2008. Global changes in extreme daily temperature since 1950. *Journal of Geophysical Research: Atmospheres*, 113 (D5), D05115. doi:10.1029/2006JD008091
- Chebana, F., *et al.*, 2014. Regional frequency analysis at ungauged sites with the generalized additive model. *Journal of Hydrometeorology*, 15 (6), 2418–2428. doi:10.1175/JHM-D-14-0060.1
- Coles, S., 2001. *An introduction to statistical modeling of extreme values*. London: Berlin.
- Diez-Sierra, J. and Del Jesus, M., 2019. Subdaily rainfall estimation through daily rainfall downscaling using random forests in Spain. *Water*, 11 (1), 125. doi:10.3390/w11010125
- Fisher, R.A. and Tippett, L.H.C., 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24 (2), 180–190.
- Gelman, A., Hwang, J., and Vehtari, A., 2014. Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24 (6), 997–1016. doi:10.1007/s11222-013-9416-2
- Gonzalez, S. & Bech, J., 2017. Extreme point rainfall temporal scaling: a long term (1805–2014) regional and seasonal analysis in Spain: EExtreme point rainfall temporal scaling in Spain. *International Journal of Climatology*, 37 (15), 5068–5079.
- Gregersen, I.B. *et al.*, 2013. A spatial and nonstationary model for the frequency of extreme rainfall events. *Water Resources Research*, 49 (1), 127–136.
- Gumbel, E.J., 1958. *Statistics of extremes*. Columbia University Press, EE. UU.: New York.
- Hamilton, J.D., 1991. A quasi-bayesian approach to estimating parameters for mixtures of normal distributions. *Journal of Business and Economic Statistics*, 9 (1), 27–39.
- Hatzaki, M., *et al.*, 2010. Future changes in the relationship of precipitation intensity in Eastern Mediterranean with large scale circulation. *Advances in Geosciences*, 23, 31–36. doi:10.5194/adgeo-23-31-2010
- Hosking, J.R.M. and Wallis, J.R., 1993. Some statistics useful in regional frequency analysis. *Water Resources Research*, 29 (2), 271–281. doi:10.1029/92WR01980
- IPCC, *et al.*, 2013. *The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.
- Jenkinson, A.F., 1955. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81 (348), 158–171. doi:10.1002/qj.49708134804
- Katz, R.W., Parlange, M.B., and Naveau, P., 2002. Statistics of extremes in hydrology. *Advances in Water Resources*, 25 (8–12), 1287–1304. doi:10.1016/S0309-1708(02)00056-8
- Kharin, V.V. and Zwiers, F.W., 2000. Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere–ocean GCM. *Journal of Climate*, 13 (21), 3760–3788. doi:10.1175/1520-0442(2000)013<3760:CITEIA>2.0.CO;2
- Kharin, V.V. and Zwiers, F.W., 2005. Estimating extremes in transient climate change simulations. *Journal of Climate*, 18 (8), 1156–1173. doi:10.1175/JCLI3320.1
- Leadbetter, M.R., Lindgren, G., and Rootzen, H., 1983. *Extremes and related properties of random sequences and processes*. Springer, EE. UU.: New York.
- Lez-Rouco, J.F.G., 2001. Quality control and homogeneity of precipitation data in the southwest of Europe. *Journal of Climate*, 14, 15.
- Llabrés-Brustenga, A., *et al.*, 2019. Quality control process of the daily rainfall series available in Catalonia from 1855 to the present. *Theoretical and Applied Climatology*, 137 (3–4), 2715–2729. doi:10.1007/s00704-019-02772-5
- Lopez-Bustins, J.-A., Martin-Vide, J., and Sanchez-Lorenzo, A., 2008. Iberia winter rainfall trends based upon changes in teleconnection and circulation patterns. *Global and Planetary Change*, 63 (2–3), 171–176. doi:10.1016/j.gloplacha.2007.09.002
- López, J. and Francés, F., 2013. Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates. *Hydrology and Earth System Sciences*, 17 (8), 3189–3203. doi:10.5194/hess-17-3189-2013
- Martin-Vide, J., 2004. Spatial distribution of a daily precipitation concentration index in peninsular Spain. *International Journal of Climatology*, 24 (8), 959–971. doi:10.1002/joc.1030
- Martin-Vide, J. and Lopez-Bustins, J.-A., 2006. The Western Mediterranean Oscillation and rainfall in the Iberian Peninsula. *International Journal of Climatology*, 26 (11), 1455–1475. doi:10.1002/joc.1388
- Méndez, F.J. *et al.*, 2007. Analyzing Monthly Extreme Sea Levels with a Time-Dependent GEV Model. *Journal of Atmospheric and Oceanic Technology*, 24 (5), 894–911.
- Menéndez, M., *et al.*, 2009. The influence of seasonality on estimating return values of significant wave height. *Coastal Engineering*, 56 (3), 211–219. doi:10.1016/j.coastaleng.2008.07.004
- Mínguez, R., *et al.*, 2010. Pseudo-optimal parameter selection of non-stationary generalized extreme value models for environmental variables. *Environmental Modelling and Software*, 25 (12), 1592–1607. doi:10.1016/j.envsoft.2010.05.008
- Nakajima, J., *et al.*, 2012. Generalized extreme value distribution with time-dependence using the AR and MA models in state space form. *Computational Statistics & Data Analysis*, 56 (11), 3241–3259. doi:10.1016/j.csda.2011.04.017
- Renard, B., 2011. A Bayesian hierarchical approach to regional frequency analysis. *Water Resources Research*, 47 (11), W11513. doi:10.1029/2010WR010089
- Roderick, T.P., Wasko, C., and Sharma, A., 2020. An improved covariate for projecting future rainfall extremes? *Water Resources Research*, 56 (8), 8. doi:10.1029/2019WR026924

- Salas, J.D. and Obeysekera, J., 2014. Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. *Journal of Hydrologic Engineering*, 19 (3), 554–568. doi:10.1061/(ASCE)HE.1943-5584.0000820
- Scarf, P.A., 1992. Estimation for a four parameter generalized extreme value distribution. *Communications in Statistics - Theory and Methods*, 21 (8), 2185–2201. doi:10.1080/03610929208830906
- Sheppard, C., 2018. *Genetic algorithms with python*. CreateSpace Independent Publishing Platform, United Kingdom.
- Smith, R.L., 1985. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72 (1), 67–90. doi:10.1093/biomet/72.1.67
- Smith, R.L., 1989. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science*, 4 (4), 367–377.
- Solari, S. and Losada, M.A., 2011. Non-stationary wave height climate modeling and simulation. *Journal of Geophysical Research: Oceans*, 116 (C9), C9. doi:10.1029/2011JC007101
- Stan Dev Team, 2021. *CmdStanPy*. Python 3.8. Available from: <https://pypi.org/project/pystan> [Accessed 10 October 2021].
- Stedinger, J.R., Vogel, R.M., and Foufoula-Georgiou, E., 1993. *Frequency analysis of extreme events, in handbook of hydrology*. EE. UU.: New York.
- Stephens, M.A., 1992. Introduction to Kolmogorov (1933) On the Empirical Determination of a Distribution. In: S. Kotz and N.L. Johnson, eds. *Breakthroughs in Statistics: Methodology and Distribution Springer Series in Statistics*. New York, NY, 93–105.
- University of East Anglia, 2021. *Climatic research unit - Groups and Centres - UEA*. United Kingdom. Available from: <https://www.uea.ac.uk/groups-and-centres/climatic-research-unit> [Accessed 8 February 2021].
- Wang, X.L., Zwiers, F.W., and Swail, V.R., 2004. North Atlantic Ocean wave climate change scenarios for the 21st century. *Journal of Climate*, 17 (12), 2368–2383. doi:10.1175/1520-0442(2004)017<2368:NAOWCC>2.0.CO;2
- Zhang, X., Zwiers, F.W., and Li, G., 2004. Monte carlo experiments on the detection of trends in extreme values. *Journal of Climate*, 17 (10), 1945–1952. doi:10.1175/1520-0442(2004)017<1945:MCEOTD>2.0.CO;2