High return level estimates of daily ERA-5 precipitation in Europe estimated using regionalised extreme value distributions

Pauline Rivoire, Philomène Le Gall, Anne-Catherine Favre, Philippe Naveau, Olivia Martius

High return level estimates of daily ERA-5 precipitation in Europe estimated using regionalised extreme value distributions

Pauline Rivoire*

Oeschger Centre for Climate Change Research and Institute of Geography, University of Bern, Bern, Switzerland

Philomène Le Gall*

Univ. Grenoble Alpes, CNRS, IRD, Grenoble INP, IGE, F-38000 Grenoble, France

Anne-Catherine Favre

Univ. Grenoble Alpes, CNRS, IRD, Grenoble INP, IGE, F-38000 Grenoble, France

Philippe Naveau

Laboratoire des Sciences du Climat et de l'Environnement, ESTIMR, CNRS-CEA-UVSQ, Gif-sur-Yvette, France

Olivia Martius

Oeschger Centre for Climate Change Research and Institute of Geography, University of Bern, Bern, Switzerland

Corresponding author address:

E-mail: pauline.rivoire@giub.unibe.ch; philomene.le-gall@univ-grenoble-alpes.fr

* These authors contributed equally to this work

ABSTRACT

Accurate estimation of daily rainfall return levels associated with large return periods is needed for a number of hydrological planning purposes, including protective infrastructure, dams, and retention basins. This is especially relevant at small spatial scales. The ERA-5 reanalysis product provides seasonal daily precipitation over Europe on a $0.25^\circ \times 0.25^\circ$ grid (about 27×27 km). This translates more than 20,000 land grid points and leads to models with a large number of parameters when estimating return levels. To bypass this abundance of parameters, we build on the regional frequency analysis (RFA), a well-known strategy in statistical hydrology. This approach consists in identifying *homogeneous* regions, by gathering locations with similar distributions of extremes up to a normalizing factor and developing sparse regional models. In particular, we propose a step-by-step blueprint that leverages a recently developed and fast clustering algorithm to infer return level estimates over large spatial domains. This enables us to produce maps of return level estimates of ERA-5 reanalysis daily precipitation over continental Europe for various return periods and seasons. We discuss limitations and practical challenges and also provide a git hub repository. We show that a relatively parsimonious model with only a spatially varying scale parameter can compete well against statistical models of higher complexity.

1. Introduction

Heavy rainfall can cause natural hazards such as landslides, avalanches and floods (e.g., see Stocker et al. 2013; EEA 2018). Such hazards can cause casualties and damages, with direct and indirect economic impacts (MunichRE 2018; Prahl et al. 2018). To design protective infrastructure, for instance, a dam, one needs to know the frequency of a given intensity of precipitation (Madsen et al. 2014). The return-period of an event is the duration during which the event occurs once, *on average* (see e.g. Cooley 2013). Symmetrically, for a given duration, say 100 years, the 100-year return level is defined as the level that is exceeded once every 100 years, *on average*. Given a dataset (e.g. observation or reanalysis), time series are finite and observing an event exactly once in 100 years does not make an event the 100-year return level. One therefore needs a statistical model to predict the intensity of such events, and even unobserved events.

The aim of this paper is to provide return levels for large return periods over Europe. The station coverage being quite heterogeneous over Europe (Cornes et al. 2018) therefore, the use of gridded datasets is appropriate. Various types of gridded precipitation data sets are available (e.g., see Sun et al. 2018, for an overview). Precipitation gridded data can be derived from ground observations, satellite observations, combinations of ground observations and satellite observations and short-term numerical weather forecasts in reanalysis datasets. In reanalyses, past observations are assimilated in numerical weather forecast models to reconstruct past weather. The main advantage of this type of data set is its regular spatial and temporal coverage. Reanalyses also ensure consistency of the precipitation data with the atmospheric conditions, which is a valuable characteristic for weather and climate process studies. Precipitation in this study is extracted from the ERA-5 reanalysis data set (C3S 2017; Hersbach et al. 2020). We study daily precipitation over continental Europe. The region of interest covers more than 20,000 grid points over European land.

Extreme value theory (EVT) provides an asymptotic framework to model the distribution of extremes such as heavy precipitation. Two classical approaches for extreme modelling are the generalized extreme value (GEV) and the generalized Pareto (GPD). The GEV (Jenkinson 1955) aims at modelling maxima over large blocks (for instance, a year in Poschlod et al. 2021). The GPD (see e.g. Pickands III et al. 1975, and Section c) enables the modeling of exceedances over a given threshold (for instance, the 98-th quantile in Carreau et al. 2017). However, these two approaches only model extremes and our goal is to provide return levels in the full rainfall intensity range. We therefore need a class of distribution that can model the whole spectrum of precipitation intensities. Recently, Carreau and Bengio (2009), Papastathopoulos and Tawn (2013), Naveau et al. (2016) and Stein (2020) introduced distributions that model the whole spectrum of rainfall intensities. The methods model the upper tail with a Pareto distribution. Various types of transfer functions then fit the bulk and lower tail distribution. Tencaliec et al. (2020) defined a flexible version of

the extended generalized Pareto distribution (EGPD) and Rivoire et al. (2021) used it to fit the positive daily precipitation of ERA-5. The transfer function is estimated using Bernstein polynomials which bring flexibility to the transfer function estimation but require a large number of parameters (for example 30 for each grid point in Rivoire et al. 2021). In this paper, we simply use a monomial transfer function with a single flexibility parameter, see Section c for more details.

Poschlod (2021) fitted GEV distributions at each grid point of ERA-5 in Bavaria (Germany) to estimate 10-year precipitation return levels. However, extending this pointwise analysis of precipitation across Europe is quite onerous. Fitting a GEV and computing return levels for each grid point requires the estimation of more than 3×20'000 parameters (location, scale and shape parameters). In addition, estimates of the shape parameter at a specific location are quite sensitive to the length of the time series (e.g., see Zhang et al. 2012; Malekinezhad and Zare-Garizi 2014; Jalbert et al. 2017). Therefore reducing the dimensionality of the fitted parameters is of great practical importance. In contrast to this local approach, Sang and Gelfand (2009) and Naveau et al. (2014) assumed the shape parameter to be constant over the area of interest (Cape Floristic Region in South Africa and Switzerland, respectively). However, Europe is much larger than these areas, and the diverse climate and complex orography (Beck et al. 2018; Service 2020; ECMWF 2006) strongly influence the spatial distribution of precipitation (e.g., see Evin et al. 2016; Marra et al. 2021). The method used for dimensionality reduction should preserve the diverse spatial patterns of precipitation over Europe. In this paper, we therefore consider an intermede approach in which the shape parameter is common between grid points within homogeneous regions.

The regional frequency analysis (RFA), a concept from hydrology, attempts to build these homogeneous regions which consist of grid points with similar precipitation distributions (Dalrymple 1960; Hosking and Wallis 2005). In a homogeneous region, distributions are all equal to a common regional distribution up to a normalizing factor. In particular, their extreme behaviour should be analogous. Clustering grid points in homogeneous regions reduces the dimensionality of large precipitation data sets while preserving the spatial patterns. We use the definition of homogeneous distributions proposed by St-Hilaire et al. (2003) and Hosking and Wallis (2005): given a region of interest, say $\mathcal{R}$ (here Europe), a homogeneous cluster ($\mathcal{C}$) is defined as a sub-region where all spatial points s, have the same marginal distribution up to normalization:

$$\mathcal{C} = \{ s \in \mathcal{R} : \;\; Q_s = \lambda(s) \times q \}, \tag{1}$$

where Q_s is the quantile function at site s, the positive scalar $\lambda(s)$ varies in space, and q represents a positively-valued and dimensionless quantile function (common to every site in the cluster). As a consequence rescaled quantiles within a homogeneous cluster do not depend on localization s.

Several methods allow regions to be delineated as in Eq. (1). They often require climate and/or geograph-

ical covariates (see e.g. Fawad et al. 2018; Forestieri et al. 2018, for recent work) and work in three steps: i) selecting explanatory covariates, ii) grouping sites with similar covariates, and iii) testing the homogeneity of the groups obtained. Covariates are selected for their ability to explain the precipitation distribution (Evin et al. 2016; Ouarda et al. 2008). For instance, Darwish et al. (2021) selected explanatory covariates by applying a principal component analysis to available geographical and climate data. They found that longitude, latitude, elevation and seasonality of events explained most hourly precipitation in the UK. With these methods, choosing covariates is an essential step that requires expert knowledge. Moreover, covariate data must be available and may be complicated to transfer across regions with different climate characteristics. For example, the covariates that best describe precipitation may be different between the UK and Italy. To check the homogeneity of covariate-based groups, Hosking and Wallis (2005) proposed tests to examine the validity of the model corresponding to Eq. (1). The tests rely on two components: the moments that characterize the precipitation distribution, and the distributional assumption (Kappa-distributed, see e.g. Hosking 1994). The tests consist of measuring the dispersion of some estimated L-moments (for all sites in the region) around a theoretical regional value of L-moments. To compute the theoretical value, Hosking and Wallis (2005) assume that the precipitation follows a Kappa distribution. This distributional assumption is not necessarily satisfied in practice. To bypass the selection of covariates, Saf (2009) and Le Gall et al. (2021) proposed methodologies using precipitation data only. They started from the hypothesis that the distributions are partially characterized by their probability weighted moments (PWM, Greenwood et al. 1979).

Le Gall et al. (2021) recently proposed a PWM-based algorithm to identify homogeneous spatial clusters of extreme precipitation and applied the algorithm to Swiss daily precipitation observations. The algorithm provided spatially coherent regions without using any geographical covariate. In this paper, we apply the clustering algorithm from Le Gall et al. (2021) to ERA-5 daily precipitation from all European land areas to group grid points with similar upper tails.

When clusters are delineated, information from all homogeneous grid points can be pooled to accurately estimate the EVT distribution parameters. For the regional distribution, we use an EGPD with three parameters see Section c. Only the scale parameter can vary within a homogeneous cluster. The flexibility and shape parameters are constant over the cluster. In a nutshell, the regional approach allows us to go from a model with $3 \times 20,000$ parameters to a model with $2 \times n_{\text{clusters}} + 20,000$ parameters, n_{clusters} being the number of clusters. We also compare the performance of this regional approach to the performance of a more flexible distribution where the flexibility parameter can vary between sites of the same homogeneous cluster.

This study is the first to provide ERA-5 return levels, which, to our knowledge, have never been provided

for the whole of Europe. Second, RFA is traditionally applied to smaller areas such as countries (see e.g. Fowler and Kilsby 2003; Evin et al. 2016, for RFA on the UK and Switzerland).

Section 2 introduces the precipitation data set and Section 3 describes the methods for the non-parametric clustering algorithm, the regional fitting and its assessment. The homogeneous regions, the assessment of the regional fitting and the corresponding 10, 50 and 100-year return levels are presented in the results section, Section 4. We discuss our results and compare our clusters to the regions obtained by national-scale studies in Section 5. We draw conclusions in Section 6.

2. Data

We use ERA-5 daily precipitation with 0.25° spatial resolution. ERA-5 is the latest global reanalysis data set provided by the European Center for Medium-Range Weather Forecasts (C3S 2017; Hersbach et al. 2020). Precipitation is provided with hourly resolution forecasts that we aggregate to daily precipitation. We study ERA-5 precipitation for the period 1979–2018 in Europe over land, which is a region in which the data set performs well (Rivoire et al. 2021). Because practical applications are mainly restricted to the continent, we do not include precipitation data over the oceans. We conduct a seasonal analysis to ensure the stationarity of the time series. We consider the daily positive precipitation for each season. Days are considered as wet when precipitation exceeds 1 mm (Maraun 2013).

3. Method

Here, we introduce the two stages of RFA: i) identify homogeneous regions (Sections a and b) and ii) use data from all grid points in the same region to model rainfall intensities (Section c). We also introduce the evaluation tools we used to assess the fitted distributions (Section d).

a. A scale-invariant ratio of PWM

Following the notations of Le Gall et al. (2021), we denote $\alpha_i(Z)$ the i-th PWM of the positive F-distributed random variable Z

$$\alpha_i(Z) = \mathbf{E}\left[ZF(Z)^i\right].$$

When self-evident, it is denoted simply as α_i. The first three moments are used to compute the scale-invariant ratio

$$\omega = \frac{3\alpha_2 - 2\alpha_1}{2\alpha_1 - \alpha_0}, \tag{2}$$

Le Gall et al. (2021) showed that ω can be seen as a ratio of two distances derived from norms.

Let i and j be two grid point locations, and Y_j and Y_j their two associated time series of seasonal positive precipitation. To spatially cluster daily rainfall, we need to compute a dissimilarity measure between two positive time series. Here, we use the ω-based distance defined by

$$\hat{d}_{ij} = \left| \widehat{\omega\,(Y_i)} - \widehat{\omega\,(Y_j)} \right|, \tag{3}$$

where $\widehat{\omega\,(Y_i)}$ is the estimate of $\omega(Y_i)$. We use this distance for two reasons. First, because the distance is based on PWM, it enables comparison of empirical distribution shapes, including heavy-tailed ones, without fitting a parametric distribution. Second, the key property of ω is its scale-invariance. For any precipitation variables Y_1, Y_2 in a homogeneous region, see Eq. (1),

$$\omega(Y_1) = \omega(Y_2).$$

The ratio ω can be interpreted as the heaviness of the tail within the mathematical framework of EVT. In the block maxima or peak-over-threshold approaches, ω only depends on the shape parameter. In the EGPD approach, ω depends on the shape and the flexibility parameter, see Section c. The distance between two grid points with homogeneous distributions should be close to zero. The clustering algorithm gather sites with similar ω estimates.

b. Clustering algorithm: partitioning around medoids (PAM)

Grouping close ω estimates is an unsupervised learning problem: we gather unclassified points that have common characteristics (here, their ω value). The grouping of estimates into clusters is based on geometric considerations: estimates are grouped if they are close to each other in the space of variables (here the axis of reals).

Several clustering methods are available (Kaufman and Rousseeuw 1990; Murty et al. 1999; Schubert and Rousseeuw 2021), most classic ones fall in two categories: partitioning or hierarchical methods. The k-medoids, or partitioning around medoids (PAM), and k-means are iterative algorithms that belong to the first group. They both require the final number of clusters k as input. The PAM algorithm is preferred to the k-means because of its ease of interpretation. Indeed, centres of the k-means clusters are barycentres and therefore virtual points whereas the centres of the PAM clusters are actual points of the dataset (see e.g. Jain et al. 1999; Bernard et al. 2013). For each of these methods, the choice of the dissimilarity measure is paramount. We work with the absolute difference as a distance, also called the Manhattan distance, as recommended in Bernard et al. (2013); Bador et al. (2015), see Eq. (3).

The centre of each cluster is the grid point with the smallest dissimilarity to all other grid points in the

cluster and is called the medoid. Each non-medoid point of the data-set is associated with its closest medoid. Generally speaking, PAM converges to an ensemble of medoids and clusters that is a local minimum of the total cost, see Eq. (9) in Le Gall et al. (2021).

To solve this optimization problem, PAM starts by selecting k initial medoids, here in a deterministic way. The first medoid is the medoid of the partition for one cluster: the most centrally located point. The set of k medoids is then completed by adding the medoids of partitions with an increasing number of clusters one by one until k is reached.

The second step consists of testing every swap possible between a medoid and any point non-medoid in the whole data set. If the total cost function decreases, then the point is kept as medoid. Clusters are then updated with respect to their new medoids. When no swap decreases the total cost (see e.g. Le Gall et al. 2021), then the algorithm stops.

The computational cost of these two steps increases with the size of the data set and the number of clusters. Because ERA-5 provides data for about 20‘000 grid points in European lands, we use a faster version (Reynolds et al. 2006; Schubert and Rousseeuw 2021) of the original algorithm. This variation removes some redundant computations in the swap step.

To measure the strength of the link between a point and its cluster, Rousseeuw (1987) introduced the silhouette score. The silhouette score for grid point i that belongs to the cluster k is defined as

$$1 - \left(\frac{d_{ik}}{\delta_{i,-k}} \right) \tag{4}$$

where d_{ik} is the average intra-cluster dissimilarity between grid point i and all other grid points in cluster k, and $\delta_{i,-k}$ the smallest of the $k-1$ average distance between site i and all other sites associated with a cluster different from k. When a grid point i is well classified, the intra-cluster average distance is significantly smaller than the distance between clusters. Its silhouette score is then close to 1. By contrast, a silhouette score close to -1 indicates a poorly classified grid point that should be in another cluster. Eventually, a grid point that is not significantly closer to points in the cluster than to other points has a silhouette score close to 0. In other words, it is not strongly linked to any cluster.

Finding the optimal number of clusters in a data set is a tricky task (Sugar and James 2003; Pansera et al. 2013). Numerous criteria that aim at identifying tight and well-separated clusters exist (Halkidi et al. 2002; Desgraupes 2013). We compute five of them (silhouette, Dunn, Davies Bouldin, Xie Beni, S_Dbw, see e.g. Halkidi et al. 2002; Desgraupes 2013) to determine the optimal number of clusters, between two and ten. These criteria are based on different distances and provide a different optimal number of clusters. We therefore choose the number of clusters subjectively. We visually compare the maps of the partitions for numbers of clusters. We compromise between a large number of clusters and a partition that is not

fragmented.

c. Regional fitting

To model the entire precipitation distribution, Naveau et al. (2016) and Tencaliec et al. (2020) proposed a simple scheme to build a flexible distribution by writing

$$F(z) = G(H_{\sigma,\xi}(z))$$

where the flexibility function G can be any cumulative distribution function such that there exists $\kappa > 0$ such that $\frac{G(u)}{u^\kappa}$ and $\frac{1-G(1-u)}{u}$ have finite limits when u goes to zero. These constraints ensure that F follows EVT for very low and high precipitation accumulations. Here, we use $G_\kappa(u) = u^\kappa, \quad \kappa > 0$, as flexibility function. Although simple, G is sufficiently flexible to model daily rainfall distributions while maintaining parsimony in the model (Evin et al. 2016).

We fit the parameters to different levels of regionalization, from σ, ξ and κ computed individually at every grid point to σ computed individually and κ and ξ being common between grid points in a homogeneous region (see Table 1).

PWM can quickly be estimated non-parametrically and used for estimation of EGPD parameters (see Appendix of Naveau et al. 2016). Estimates of local parameters are taken as initial values for the iterative estimation of regional or semi-regional parameters, see Algorithm (1).

The quantile with probability p can be computed using the explicit formula

$$y_p = F^{-1}(p) = \frac{\sigma}{\xi}\left[\left\{1 - G^{-1}(p)\right\} - 1\right], \quad \text{if } \xi > 0, \tag{5}$$

$0 < p < 1$. The return level associated with return period of T years is y_p for $p = \frac{1}{T \times n_{\text{wds}}}$, where n_{wds} is the number of wet days per season. We use the mean of the number of wet days per season during the period under study as an approximation for n_{wds}.

For every grid point, we assume that the random variable modelling daily positive precipitation is independent and identically distributed. However, precipitation events can last for several consecutive days (Buriticá et al. 2021). To ensure independence in a time series of wet days, we extract one wet day out of three to fit the EGPD models. Despite the climate change, Donat et al. (2014) did not detect any clear trend in the whole precipitation distribution over the period of interest. The absence of a trend and the seasonal analysis are necessary to ensure identical distribution.

Algorithm 1 Regional fit of the EGPD in cluster C, see last row of Table 1

$cond = TRUE, eps = .001,$ and $u = 1$(mm)

procedure INPUT(Rainfall Matrix for cluster C)

  Remove dry days by only taking $\{Y(s)|Y(s) > u\}$

  Fit local Model at each location $s \in C$

  Denote κ_0 and ξ_0 the cluster means of κ and ξ from Step 4

  Compute $m(s)$ the sample mean at each $s \in C$

  while $cond = TRUE$ **do**

    Compute

$$\sigma_{new}(s) = \frac{\xi_0 \; m(s)}{\frac{\kappa_0}{F(u)} IB\left(H_{\xi_0}\left(\frac{u}{\sigma_0}\right), 1, \kappa_0, 1 - \xi_0\right) - 1}$$

    where $IB(.,.,.)$ is the incomplete Beta function

    if $|\sigma_{new} - \sigma_0| < eps$ **then**

    $cond = FALSE$

    end if

      $\sigma_0 \leftarrow \sigma_{new}$

  end while

  Return $(\kappa_0, \sigma_0, \xi_0)^T$

end procedure

d. Assessment of the fitting

We evaluate the goodness-of-fit with standard statistical tools focusing on accuracy, flexibility of estimation, and rewarding of the parsimony (smaller number of parameters).

First, quantile-quantile plots (QQ-plots) provide visual information on the proximity between two distributions. For selected grid points, we present QQ-plots, contrasting the empirical quantiles with the quantiles parametrically estimated with the local, semi-regional and regional fits, and EGPD with Bernstein flexibility function (see Table 1).

We assess the agreement between the fitting and the empirical distribution with the Anderson-Darling test (see e.g. Anderson and Darling 1952; Scholz and Stephens 1987). To ensure independence between the empirical and fitted distribution at a given grid point, we use a third of the positive precipitation time series that was not used in the fitting process as empirical data. Table 2 summarizes the results of the Anderson–Darling test over Europe for the regional, the semiregional, and the local fittings. To ensure spatial independence, we perform the tests for 1/8th of the grid points, randomly chosen. This way we avoid repetition of information between neighbouring grid points.

To evaluate the goodness-of-fit, we compute the Akaike information criterion (AIC) (Akaike 1987) for the local, the semiregional, and the regional models. This criterion combines a measure of the goodness of fit (log-likelihood) with the parsimony andsparsity of the model. The AIC has to be minimized. A smaller number of fitted parameters is a bonus for the model because this reduces the risk of overfitting. For example, the local model requires the estimation of about 3×20‘000 parameters, whereas the regional model only needs the estimation of about 20‘$000 + (\text{number of clusters}) \times 2$ parameters.

4. Results

a. Partition of ERA-5 over Europe

We apply the clustering algorithm introduced in Section 3b to ERA-5 positive daily precipitation for each season independently.

The optimal number of clusters is three for September-October-November (SON), December-January-February (DJF), and March-April-May (MAM), and five for June-July-August (JJA, see Section 5 for a discussion about this number). Figure 1 shows these partitions. The shade of colour indicates the silhouette coefficient of the grid points; light colours indicate low silhouette coefficients and therefore a weak association with the cluster. There are very few isolated grid points. For all the seasons, the borders between clusters follow the orography, for example in the Alps, the Carpathians, and the UK. This orographic link is present

in all seasons. Hence, the ratio ω captures spatial structures associated with physical features such as orography without requiring additional covariates such as longitude, latitude, or elevation. Silhouette scores are lowest at the borders between clusters, and downward-pointing triangles, which indicate grid points with low and minimum silhouette coefficients, are often located in transition zones between clusters (Fig. 1).

b. Assessment of the fitting

The fitting models are assessed with the Anderson–Darling test, the AIC criterion and QQ-plots.

The Anderson–Darling test indicates similar performance for the fitting of the regional and local EGPD models; see Table 1. The null hypothesis is that the fitted and the empirical distribution are the same. Table 2 displays the nonrejection rates of the null hypothesis for the Anderson–Darling test for each season and model across the entire domain. The null hypothesis is not rejected for 87% of the grid points in JJA and 91% in SON for the local fit. For the regional fit, it is not rejected for 84% of grid points in JJA and 88% in SON, DJF, and MAM. The nonrejection rate for the semiregional fitting is very similar to that the regional fitting. The percentage is lower for the local fit than for the regional fit in all seasons. Nonetheless, the difference between local and regional is smaller than 3% in all seasons. For all seasons and all fittings, the nonrejection rate indicates good performance of the model, the perfect nonrejection rate being 95% on a test with a confidence level of 5%.

The variability of meteorological processes tends to increase with the altitudinal gradient. Around complex topography, local-scale variations in precipitation may occur. Precipitation distributions might differ substantially between grid points, even within a homogeneous region, and the quality of the regional fit might decrease. We therefore distinguished the rejection rate of Anderson–Darling test between grid points below and above 1000 meters above sea level. We did not find any significant difference in the rejection rate of the Anderson–Darling test between grid points at low and high altitudes (not shown). Moreover, the goodness of the classification in the clustering procedure might impact the accuracy of the fit. At a grid point with a poor connection to its cluster, the regional value of ξ (and κ) might not be accurate and the distribution fitted regionally might be significantly different from the empirical distribution. We distinguished the Anderson–Darling test between grid points with a silhouette greater or lower than 0.2. Here too, we observe no significant difference between grid points with low and high silhouettes, for either the local or regional fits (not shown). Even if the local model is the most adaptable, the regional model seems to be sufficiently flexible to (i) take into account the local-scale variations caused by complex topography and (ii) compensate for the regionalization of two parameters out of three.

The AIC criterion summarizes the information contained in the likelihood and penalizes the number of

parameters. It should be as low as possible. The AIC is much lower for the regional model than for the semiregional and local models independent of the season (see table 3). AIC values across all grid points vary between $-115,106$ in JJA and $-107,250$ in DJF for the regional fitting, between $-79,400$ in JJA and $-67,614$ in DJF for the semiregional fitting and between $-43,704$ in JJA and $-27,984$ in SON for the local fitting. These AIC values highlight the trade-off between parsimony and goodness of fit of the regional fitting. Having only one EGP parameter varying within one cluster in the regional model substantially reduces the AIC.

Figure 2 displays the QQ plots for cluster medoid, cluster minimum, and cluster maximum silhouette coefficient in each cluster in SON (see partition in Fig.1(a)). For the most centrally located grid point, the medoid, all the fittings perform similarly well. One exception is the upper tail in the northern and southern clusters, which is slightly overestimated with the regional fitting compared to the local one. For the grid point with a minimum silhouette, the regional and semiregional fit have a similar performance to the local one or even outperform them in the southern cluster. In the intermediate and southern clusters, for the semiregional and regional fittings only the most extreme precipitation is overestimated compared to the local fit. For the grid point with the maximum silhouette, the regional and semiregional fits outperform the local fit for the whole distribution. Extremes are well captured with the regional method, except in the northern cluster, where the highest precipitation is overestimated for all the fittings. The semiregional and regional fittings seem to significantly improve the quality of estimation for the best-classified points. The semiregional and regional fittings have similar performances. We also compared with the local Bernstein fitting; see Table 1. Its performance is similar to the semiregional and regional fittings except in the southern cluster.

c. Return levels

The estimate of the return levels is spatially smooth despite the regionalization of two out of three parameters in the EGPD. Figures 3, 4 and 5 show the 10-, 50-, and 100-year return levels for all seasons. Even though the shape and flexibility parameters ξ and κ are constant across each cluster, the variability of scale parameter σ (estimated locally) accounts for the high level of spatial detail of the fit. Regions with high return levels are shown in deep blue and purple colours on the map. Specific regions known to experience heavy precipitation are highlighted, such as the Cévennes, South of France (with Cévenols episodes, see e.g. Ducrocq et al. 2008; Vautard et al. 2015) in SON and the Canton of Ticino in southern Switzerland (Isotta et al. 2014; Barton et al. 2016; Panziera et al. 2018) in SON, MAM, and JJA.

Finally, we compare the return levels obtained with the local fit and the regional fit. Figure 6 displays

the relative difference between the 50-year return levels for regional and local fittings. The return levels differ by less than 10% for about 60% of the grid points in SON and for up to 80% of the grid points in DJF. The mean value of the absolute difference lies between 7% (DJF) and 10% (SON). Areas with the highest relative differences are generally located in the cluster with the highest shape parameters: those with more frequent extremes. The same maps for the 10- and 100-year return levels can be found in appendix (Fig. 10 and 11).

5. Discussion

We conduct a PAM clustering procedure based on the PWM ratio ω. We find that the optimal number of clusters is three in SON, DJF and MAM, and five in JJA. The higher number of clusters in JJA might be explained by a larger spatial variability of precipitation extremes in Europe in summer (see e.g Cortesi et al. 2014, in Spain). The same analysis was conducted for hierarchical partitioning, leading to the same results for some parametrization (results available upon request).

The choice of the optimal number of clusters is challenging. The various criteria for the choice proposed in the literature did not agree on the optimal number of clusters (see e.g. Pansera et al. 2013, for the use of two criteria). This can be explained by the large number of grid points in the analysis, resulting in more noise in the criteria than actual information about the goodness of the partitioning. In general, if many grid points are involved, we recommend using more than one criterion and checking the maps visually for the plausibility of the partition obtained.

We analyse the impact of the number of clusters on the regional fit. For this purpose, we compare the difference between the 50-year return levels based on the PAM partition with three clusters and the one with four clusters, in SON (not shown). For a large majority of the grid points, the difference in return levels is lower than 5%. The difference is a bit larger for a few outliers but remains lower than 25%. The outliers are generally located in regions with a very low silhouette score for the partition with 3 clusters.

We compare our partition in Central Europe to that obtained by Gvoždíková et al. (2019) over Germany, Poland, Austria and the Czech Republic. They considered extreme events between 1961 and 2013. Their approach was based on the weather extremity index. They computed Ward's linkage in the hierarchical clustering algorithm. The clusters they found exhibit a West-East pattern. The partition we obtain over these four countries also tends to separate eastern and western regions. Darwish et al. (2021) also found this West-East pattern in the UK. They delineated the regions using the most explanatory covariates (among those that were available) and then assessed their homogeneity by computing tests of Hosking and Wallis (2005) on hourly precipitation. Our results generally agree with the partition they obtained.

The regional model is more parsimonious than the local model; see Table 1. It is also more precise on well-classified points (see Figure 2). The semi-regional and regional models have similar performance; hence, the regional model should be preferred because it is more parsimonious. An alternative to our fitting method would be to select only grid points with a satisfactory silhouette score (e.g. greater than 0.2) to estimate the regional parameters. The quantiles of points with very low silhouette scores would then be estimated locally. This could increase the likelihood of the fitted distribution in some cases but would also increase the number of parameters to fit. However, the performance of the regional fit was not substantially lower than the local fit for the border areas between the clusters, and the rate of rejection in the Anderson–Darling test was not substantially higher at grid points with low silhouettes. For the sake of simplicity and parsimony, we choose to keep the regional approach for all points.

The local Bernstein distributions do not seem to be substantially closer to the empirical distribution than the regional ones; see Figure 2. Hence the flexibility brought by the scale parameter σ in the regional model is sufficient to fit the data well and therefore the most parsimonious model is as precise as the others.

We did not conduct any false discovery rate control for the Anderson–Darling test despite the large number of p-values. With a false discovery rate control, the null hypothesis would be more difficult to reject. The percentage of grid points rejecting the null hypothesis would be smaller and we would draw the same conclusion about the good performance of the EGPD fit. The rejection rate of the null hypothesis is slightly larger for the regional fitting than for the local fitting. However, we did not conduct a test of significance of the difference between these two rejection rates because of the computational cost.

The spatial pattern of our seasonal 10-year return levels (Fig. 3) is similar to that of the yearly 10-year return levels obtained by Poschlod et al. (2021) with an observational data set and the Canadian Regional Climate Model.

We also compare the return levels over Switzerland with those provided by MeteoSwiss (2019) for all the seasons (see Fig. 7, 8 and 9 in appendix). This small country provides a good test case for our study because the complex orography leads to a wide variety of precipitation patterns (Schmidli et al. 2002; Umbricht et al. 2013; Isotta et al. 2014; Evin et al. 2018). Return levels obtained by MeteoSwiss (2019) were computed by fitting a GEV to observed seasonal maxima, with a much higher spatial grid resolution than ERA-5 (up to 1km, see MeteoSwiss 2020). The maps of return levels are close in terms of magnitude and exhibit very similar spatial patterns. Only the small-scale structures are not captured by ERA-5 which is due to the coarser grid resolution of ERA-5. The magnitudes of extremes are slightly underestimated in ERA-5, especially in MAM. This agrees with the study of Hu and Franzke (2020) over Germany. They state that ERA-5 generally underestimates extremes of daily precipitation compared to observation-based gridded datasets (and weather station observations). In our analysis, despite the regionalization (three or

five clusters in Switzerland depending on the season) of two parameters out of three, the scale parameter σ presents sufficient variability to have correct return levels. The variability of σ alone is sufficient to provide accurate fitting, even in a country with a complex topography and high local spatial variability of extreme precipitation.

6. Conclusion

We derive return levels of extreme daily precipitation ($>$ 1mm) over Europe using regionalized parameters for the EGPD fits. The regionalization requires two steps. First, all land grid points are partitioned into a few homogeneous regions with a clustering algorithm. As distance measure, we estimate a scale-invariant ratio of PWM for each grid point, focusing on the tail of the distribution, and then use the PAM clustering algorithm to group these estimates into regions. The second step is the choice and fitting of a model to estimate return levels. We choose to fit an EGP distribution that models the full range of precipitation intensity. Only the scale parameter is allowed to vary within a homogeneous cluster, and the tail and flexibility parameters are common to all grid points in that cluster.

We assessed our regional analysis with classical statistical tools and compared it to previous analyses and return level estimates. Although parsimonious, the regional model is sufficiently flexible to capture the strong spatial variability of rainfall intensities.

This paper provides two main contributions. We provide maps of 10-, 50- and 100-year return levels for European precipitation of ERA-5, and we have made the algorithms for clustering and regional model available in a GitHub repository[1].

Acknowledgment

Within the CDP-Trajectories framework, this work is supported by the French National Research Agency in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02). P.L. and A-C.F. gratefully acknowledge financial support for this study provided by the Swiss Federal Office for Environment (FOEN), the Swiss Federal Nuclear Safety Inspectorate (ENSI), the Federal Office for Civil Protection (FOCP), and the Federal Office of Meteorology and Climatology, MeteoSwiss, through the project EXAR ("Evaluation of extreme Flooding Events within the Aare-Rhine hydrological system in Switzerland").

P.R. and O.M. acknowledge funding from the the Swiss National Science Foundation (grant number 178751).

[1] `https://github.com/PhilomeneLeGall/RFA_regional_EGPDk.git`

Part of this work was supported by the DAMOCLES-COST-ACTION on compound events, the French national program (FRAISE-LEFE/INSU and 80 PRIME CNRS-INSU), and the European H2020 XAIDA (Grant agreement ID: 101003469). P.N. also acknowledges the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-20-CE40-0025-01 (T-REX project), and the ANR-Melody.

The authors declare that they have no conflict of interest.

Data availability

We use ERA-5 global total daily precipitation data at resolution 0.25°. Hourly values were downloaded from the ECMWF MARS server (a valid ECMWF account required): `https://apps.ecmwf.int/data-catalogues/era5/?type=fc&class=ea&stream=oper&expver=1` Forecast steps 6 to 17, variable: Total precipitation (228.128). The MARS / EMOSLIB interpolation library has been used.

REFERENCES

Akaike, H., 1987: Factor analysis and AIC. *Selected papers of Hirotugu Akaike*, Springer, 371–386.

Anderson, T. W. and D. A. Darling, 1952: Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. *The annals of mathematical statistics*, 193–212.

Bador, M., P. Naveau, E. Gilleland, M. Castellà, and T. Arivelo, 2015: Spatial clustering of summer temperature maxima from the cnrm-cm5 climate model ensembles & e-obs over europe. *Weather and climate extremes*, **9**, 17–24.

Barton, Y., P. Giannakaki, H. von Waldow, C. Chevalier, S. Pfahl, and O. Martius, 2016: Clustering of Regional-Scale Extreme Precipitation Events in Southern Switzerland. *Monthly Weather Review*, **144 (1)**, 347–369, doi:10.1175/MWR-D-15-0205.1, NIHMS150003.

Beck, H. E., N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood, 2018: Present and future köppen-geiger climate classification maps at 1-km resolution. *Scientific data*, **5 (1)**, 1–12.

Bernard, E., P. Naveau, M. Vrac, and O. Mestre, 2013: Clustering of maxima: Spatial dependencies among heavy rainfall in France. *Journal of Climate*, **26 (20)**, 7929–7937.

Buriticá, G., M. Nicolas, T. Mikosch, and O. Wintenberger, 2021: Some variations on the extremal index. *arXiv preprint arXiv:2106.05117*.

C3S, C. C. C. S., 2017: Era5: Fifth generation of ecmwf atmospheric reanalyses of the global climate. Accessed: 2021-03-02, https://cds.climate.copernicus.eu/cdsapp#!/home, accessed: 2021-03-02.

Carreau, J. and Y. Bengio, 2009: A hybrid Pareto model for asymmetric fat-tailed data: the univariate case. *Extremes*, **12 (1)**, 53–76.

Carreau, J., P. Naveau, and L. Neppel, 2017: Partitioning into hazard subregions for regional peaks-over-threshold modeling of heavy precipitation. *Water Resources Research*, **53 (5)**, 4407–4426.

Cooley, D., 2013: Return periods and return levels under climate change. *Extremes in a changing climate*, Springer, 97–114.

Cornes, R. C., G. van der Schrier, E. J. van den Besselaar, and P. D. Jones, 2018: An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. *Journal of Geophysical Research: Atmospheres*, **123 (17)**, 9391–9409, doi:10.1029/2017JD028200.

Cortesi, N., J. C. Gonzalez-Hidalgo, M. Brunetti, and M. de Luis, 2014: Spatial variability of precipitation in Spain. *Regional Environmental Change*, **14 (5)**, 1743–1749, doi:10.1007/s10113-012-0402-6.

Dalrymple, T., 1960: Flood-frequency analyses, manual of hydrology: Part 3. Tech. rep., USGPO,.

Darwish, M. M., M. R. Tye, A. F. Prein, H. J. Fowler, S. Blenkinsop, M. Dale, and D. Faulkner, 2021: New hourly extreme precipitation regions and regional annual probability estimates for the uk. *International Journal of Climatology*, **41 (1)**, 582–600.

Desgraupes, B., 2013: Clustering indices. *University of Paris Ouest-Lab Modal'X*, **1**, 34.

Donat, M. G., J. Sillmann, S. Wild, L. V. Alexander, T. Lippmann, and F. W. Zwiers, 2014: Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. *Journal of Climate*, **27 (13)**, 5019–5035, doi:10.1175/JCLI-D-13-00405.1.

Ducrocq, V., O. Nuissier, D. Ricard, C. Lebeaupin, and T. Thouvenin, 2008: A numerical study of three catastrophic precipitating events over southern France. II: Mesoscale triggering and stationarity factors. *Quarterly Journal of the Royal Meteorological Society*, **134 (630)**, 131–145, doi:10.1002/qj.199, URL website:http://www.kemkes.go.idhttp://doi.wiley.com/10.1002/qj.199.

ECMWF, 2006: URL https://sites.ecmwf.int/era/40-atlas/docs/section_B/parameter_tp.html#, URL https://sites.ecmwf.int/era/40-atlas/docs/section_B/parameter_tp.html#.

EEA, E. E. A., 2018: URL https://www.eea.europa.eu/publications/national-climate-change-vulnerability-2018, URL https://www.eea.europa.eu/publications/national-climate-change-vulnerability-2018.

Evin, G., J. Blanchet, E. Paquet, F. Garavaglia, and D. Penot, 2016: A regional model for extreme rainfall based on weather patterns subsampling. *Journal of Hydrology*, **541**, 1185–1198.

Evin, G., A.-C. Favre, and B. Hingray, 2018: Stochastic generation of multi-site daily precipitation focusing on extreme events. *Hydrology and Earth System Sciences*, **22 (1)**, 655–672, doi:10.5194/hess-22-655-2018, URL https://hess.copernicus.org/articles/22/655/2018/.

Fawad, M., I. Ahmad, F. A. Nadeem, T. Yan, and A. Abbas, 2018: Estimation of wind speed using regional frequency analysis based on linear-moments. *International Journal of Climatology*, **38 (12)**, 4431–4444.

Forestieri, A., F. Lo Conti, S. Blenkinsop, M. Cannarozzo, H. J. Fowler, and L. V. Noto, 2018: Regional frequency analysis of extreme rainfall in sicily (italy). *International Journal of Climatology*, **38**, e698–e716.

Fowler, H. and C. Kilsby, 2003: A regional frequency analysis of united kingdom extreme rainfall from 1961 to 2000. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **23 (11)**, 1313–1334.

Greenwood, J. A., J. M. Landwehr, N. C. Matalas, and J. R. Wallis, 1979: Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. *Water resources research*, **15 (5)**, 1049–1054.

Gvoždíková, B., M. Müller, and M. Kašpar, 2019: Spatial patterns and time distribution of central european extreme precipitation events between 1961 and 2013. *International Journal of Climatology*, **39 (7)**, 3282–3297.

Halkidi, M., Y. Batistakis, and M. Vazirgiannis, 2002: Clustering validity checking methods: Part ii. *ACM Sigmod Record*, **31 (3)**, 19–27.

Hersbach, H., et al., 2020: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **(September 2019)**, 1–51, doi:10.1002/qj.3803.

Hosking, J. R., 1994: The four-parameter kappa distribution. *IBM Journal of Research and Development*, **38 (3)**, 251–258.

Hosking, J. R. M. and J. R. Wallis, 2005: *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press.

Hu, G. and C. L. Franzke, 2020: Evaluation of daily precipitation extremes in reanalysis and gridded observation-based data sets over germany. *Geophysical Research Letters*, **47 (18)**, e2020GL089 624.

Isotta, F. A., et al., 2014: The climate of daily precipitation in the alps: development and analysis of a high-resolution grid dataset from pan-alpine rain-gauge data. *International Journal of Climatology*, **34 (5)**, 1657–1675, doi:Doi10.1002/Joc.3794.

Jain, A. K., M. N. Murty, and P. J. Flynn, 1999: Data Clustering: A Review. *ACM Comput. Surv.*, **31 (3)**, 264–323, doi:10.1145/331499.331504, URL `https://doi.org/10.1145/331499.331504`.

Jalbert, J., A.-C. Favre, C. Bélisle, and J.-F. Angers, 2017: A spatiotemporal model for extreme precipitation simulated by a climate model, with an application to assessing changes in return levels over North America. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **66 (5)**, 941–962, doi:10.1111/rssc.12212, URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssc.12212`, `https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/rssc.12212`.

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.

Kaufman, L. and P. J. Rousseeuw, 1990: *Finding groups in data: an introduction to cluster analysis*, Vol. 344. John Wiley & Sons.

Le Gall, P., A.-C. Favre, P. Naveau, and C. Prieur, 2021: Improved regional frequency analysis of rainfall data.

Madsen, H., D. Lawrence, M. Lang, M. Martinkova, and T. R. Kjeldsen, 2014: Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. *Journal of Hydrology*, **519 (PD)**, 3634–3650, doi:10.1016/j.jhydrol.2014.11.003, URL `http://dx.doi.org/10.1016/j.jhydrol.2014.11.003`.

Malekinezhad, H. and A. Zare-Garizi, 2014: Regional frequency analysis of daily rainfall extremes using L-moments approach. *Atmósfera*, **27 (4)**, 411 – 427, doi:https://doi.org/10.1016/S0187-6236(14)70039-6, URL `http://www.sciencedirect.com/science/article/pii/S0187623614700396`.

Maraun, D., 2013: Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *Journal of Climate*, **26 (6)**, 2137–2143, doi:10.1175/JCLI-D-12-00821.1.

Marra, F., M. Armon, M. Borga, and E. Morin, 2021: Orographic effect on extreme precipitation statistics peaks at hourly time scales. *Geophysical Research Letters*, **48 (5)**, e2020GL091 498.

MeteoSwiss, 2019: Maps of extreme precipitation. Accessed: 2021-07-25, `https://www.meteoswiss.admin.ch/home/climate/swiss-climate-in-detail/extreme-value-analyses/maps-of-extreme-precipitation.html?filters=1-day-sum_JJA_retlev_X100`, accessed: 2021-07-25.

MeteoSwiss, 2020: Spatial climate analyses. Accessed: 2021-09-09, `https://www.meteoswiss.admin.ch/home/climate/swiss-climate-in-detail/raeumliche-klimaanalysen.html`, accessed: 2021-09-09.

MunichRE, 2018: A stormy year: Natural Catastrophes 2017. *Topics Geo*, 65, URL `https://www.munichre.com/topics-online/en/climate-change-and-natural-disasters/natural-disasters/topics-geo-2017.html`.

Murty, M. N., A. Jain, and P. Flynn, 1999: Data clustering: a review acm compt. surv. *ACM Computing Surveys*, **31 (3)**.

Naveau, P., R. Huser, P. Ribereau, and A. Hannart, 2016: Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, **52 (4)**, 2753–2769.

Naveau, P., A. Toreti, I. Smith, and E. Xoplaki, 2014: A fast nonparametric spatio-temporal regression scheme for generalized pareto distributed heavy precipitation. *Water Resources Research*, **50 (5)**, 4011–4017.

Ouarda, T., A. St-Hilaire, and B. Bobée, 2008: Synthèse des développements récents en analyse régionale des extrêmes hydrologiques. *Revue des sciences de l'eau/Journal of Water Science*, **21 (2)**, 219–232.

Pansera, W. A., B. M. Gomes, A. Boas, and E. L. Mello, 2013: Clustering rainfall stations aiming regional frequency analysis. *Journal of Food, Agriculture & Environment*, **11 (2)**, 877–885.

Panziera, L., M. Gabella, U. Germann, and O. Martins, 2018: A 12-year radar-based climatology of daily and sub-daily extreme precipitation over the swiss alps. *International Journal of Climatology*, **38 (10)**, 3749–3769, doi:10.1002/joc.5528.

Papastathopoulos, I. and J. A. Tawn, 2013: Extended generalised pareto models for tail estimation. *Journal of Statistical Planning and Inference*, **143 (1)**, 131–143.

Pickands III, J. et al., 1975: Statistical inference using extreme order statistics. *Annals of statistics*, **3 (1)**, 119–131.

Poschlod, B., 2021: Using high-resolution regional climate models to estimate return levels of daily extreme precipitation over bavaria. *Natural Hazards and Earth System Sciences Discussions*, 1–32.

Poschlod, B., R. Ludwig, and J. Sillmann, 2021: Ten-year return levels of sub-daily extreme precipitation over europe. *Earth System Science Data*, **13 (3)**, 983–1003.

Prahl, B. F., M. Boettle, L. Costa, J. P. Kropp, and D. Rybski, 2018: Data Descriptor: Damage and protection cost curves for coastal floods within the 600 largest European cities. *Scientific Data*, **5**, 1–18, doi:10.1038/sdata.2018.34.

Reynolds, A. P., G. Richards, B. de la Iglesia, and V. J. Rayward-Smith, 2006: Clustering rules: a comparison of partitioning and hierarchical clustering algorithms. *Journal of Mathematical Modelling and Algorithms*, **5 (4)**, 475–504.

Rivoire, P., O. Martius, and P. Naveau, 2021: A comparison of moderate and extreme ERA-5 daily precipitation with two observational data sets. *Earth and Space Science*, doi:10.1029/2020ea001633.

Rousseeuw, P. J., 1987: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, **20**, 53–65.

Saf, B., 2009: Regional flood frequency analysis using l-moments for the west mediterranean region of turkey. *Water Resources Management*, **23 (3)**, 531–551.

Sang, H. and A. E. Gelfand, 2009: Hierarchical modeling for extreme values observed over space and time. *Environmental and ecological statistics*, **16 (3)**, 407–426.

Schmidli, J., C. Schmutz, C. Frei, H. Wanner, and C. Schar, 2002: Mesoscale precipitation variability in the region of the european alps during the 20th century. *International J. Climatol.*, **22 (9)**, 1049–1074.

Scholz, F. W. and M. A. Stephens, 1987: K-sample anderson–darling tests. *Journal of the American Statistical Association*, **82 (399)**, 918–924.

Schubert, E. and P. J. Rousseeuw, 2021: Fast and eager k-medoids clustering: O (k) runtime improvement of the pam, clara, and clarans algorithms. *Information Systems*, 101804.

Service, C. C., 2020: .

St-Hilaire, A., T. Ouarda, M. Lachance, B. Bobée, M. Barbet, and P. Bruneau, 2003: La régionalisation des précipitations: une revue bibliographique des développements récents. *Revue des sciences de l'eau/Journal of Water Science*, **16 (1)**, 27–54.

Stein, M. L., 2020: Parametric models for distributions when interest is in extremes with an application to daily temperature. *Extremes*, doi:10.1007/s10687-020-00378-z, URL `https://doi.org/10.1007/s10687-020-00378-z`, https://doi.org/10.1007/s10687-020-00378-z.

Stocker, T., et al., 2013: *Technical Summary*, book section TS, 33–115. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, doi:10.1017/CBO9781107415324.005, URL `www.climatechange2013.org`.

Sugar, C. A. and G. M. James, 2003: Finding the number of clusters in a dataset: An information-theoretic approach. *Journal of the American Statistical Association*, **98 (463)**, 750–763.

Sun, Q., C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, and K.-L. Hsu, 2018: A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Reviews of Geophysics*, **56 (1)**, 79–107, doi:10.1002/2017RG000574.

Tencaliec, P., A.-C. Favre, P. Naveau, C. Prieur, and G. Nicolet, 2020: Flexible semiparametric generalized Pareto modeling of the entire range of rainfall amount. *Environmetrics*, **31 (2)**, e2582, doi:https://doi.org/10.1002/env.2582, URL `https://onlinelibrary.wiley.com/doi/abs/10.`

1002/env.2582, https://doi.org/10.1002/env.2582, https://onlinelibrary.wiley.com/doi/pdf/10.1002/env.2582.

Umbricht, A., S. Fukutome, C. Liniger, M. A.and Frei, and C. Appenzeller, 2013: Seasonal variation of daily extreme precipitation in switzerland. Report, MeteoSwiss.

Vautard, R., et al., 2015: Extreme Fall 2014 Precipitation in the Cévennes Mountains. *Bulletin of the American Meteorological Society*, **96 (12)**, S56–S60, doi:10.1175/BAMS-D-15-00088.1, URL https://journals.ametsoc.org/doi/10.1175/BAMS-D-15-00088.1.

Zhang, Q., M. Xiao, V. P. Singh, and J. Li, 2012: Regionalization and spatial changing properties of droughts across the Pearl River basin, China. *Journal of Hydrology*, **472**, 355–366.

Appendices

A. Return levels in Switzerland

List of Tables

Models	Flexibility function G	ξ	σ
Local Bernstein	G_i = Bernstein polynomials, site specific	site specific	site specific
Local	$G_i(u) = u^{\kappa_i}, \kappa_i > 0$ site-specific	site-specific	site-specific
Semi-regional	$G_i(u) = u^{\kappa_i}, \kappa_i > 0$ site-specific	constant on each cluster	site-specific
Regional	$G_i(u) = u^{\kappa}, \kappa > 0$ constant on each cluster	constant on each cluster	site-specific

Table 1: Description of the four EGPD models, with various complexity compared in Section 4b. The Bernstein EGPD is presented in Tencaliec et al. (2020). The local EGPD is introduced in Naveau et al. (2016) and its regional version in Le Gall et al. (2021). The comparison is mainly conducted between the local, the semi-regional and the regional fitting (in bold).

Model	SON	DJF	MAM	JJA
local	91%	89%	90%	87%
semi-regional	89%	87%	88%	83%
regional	88%	88%	88%	84%

Table 2: Anderson–Darling test at a risk level of 5%: Percentages of grid points for which the hypothesis of equality between the empirical distribution and the fitted distribution is not rejected. Distributions are fitted locally semiregionally and regionally; see second, third and last row of Table 1.

Model	SON	DJF	MAM	JJA
local	-30,888	-27,984	-30,016	-43,704
semi-regional	-69,792	-67,614	-69,138	-79,400
regional	-108,702	-107,250	-108,266	-115,106

Table 3: Akaike information criterion over Europe for each model and season

List of Figures

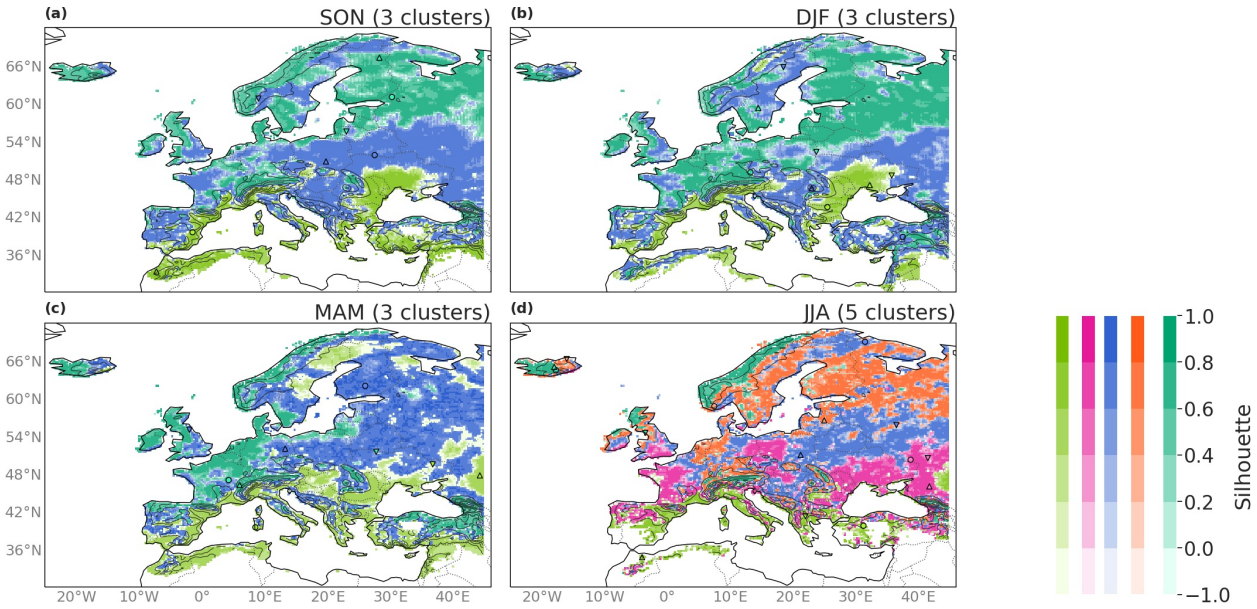

Figure 1: Partition with the PAM algorithm applied on ERA-5 daily positive precipitation over Europe for all seasons. Each cluster is identified by a colour. The shades of colour indicate the silhouette coefficient at every grid point. Intense hues indicate a strong association with the cluster. The black lines are 500m altitude isolines of the surface topography in ERA-5. Within a cluster, the circle indicates the location of the medoid, and the triangle pointing up (resp. down) indicates the grid point with the highest (resp. lowest) silhouette coefficient.

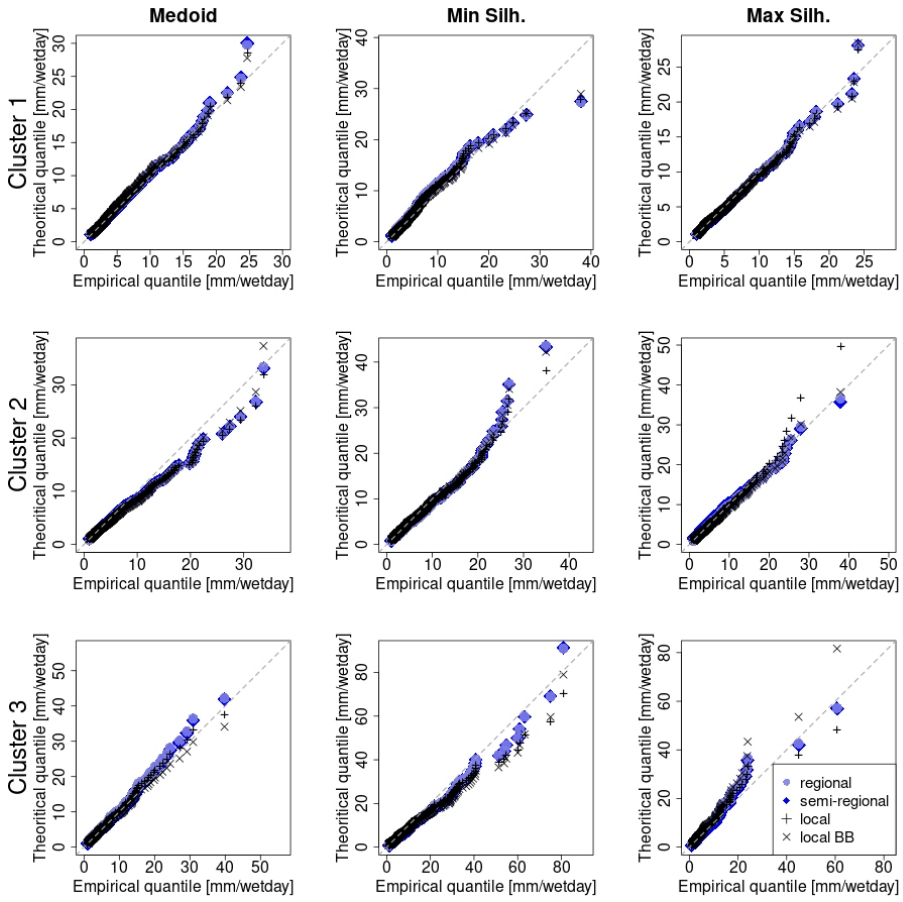

Figure 2: Example QQ plots of the regional, semiregional, local, and local Bernstein (local BB) fittings, for the medoid point (left) and the grid points with minimum (middle) and maximum (right) silhouettes in the northern (top row), intermediate (middle row), and southern (bottom row) clusters in SON (blue cluster in Figure 1).

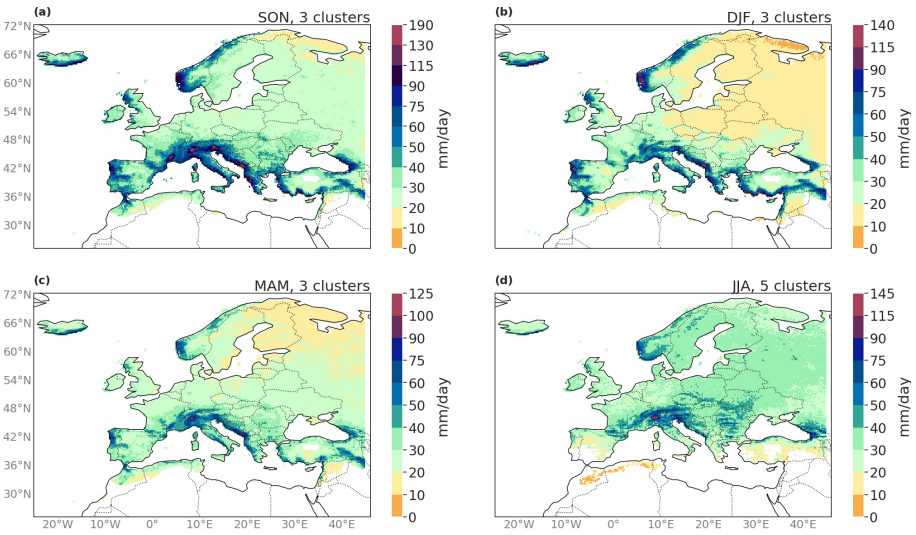

Figure 3: 10-year return levels computed with the regional fitting, see Table 1

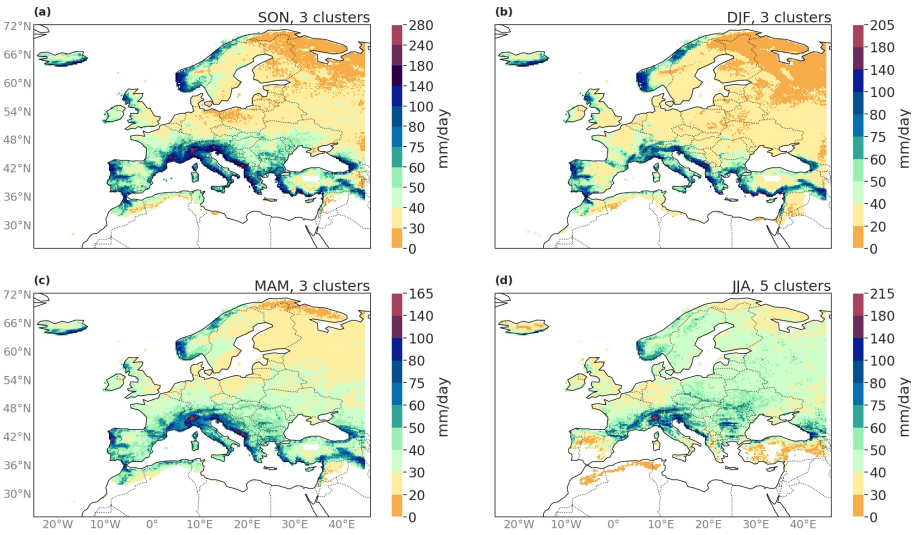

Figure 4: 50-year return levels computed with the regional fitting, see Table 1

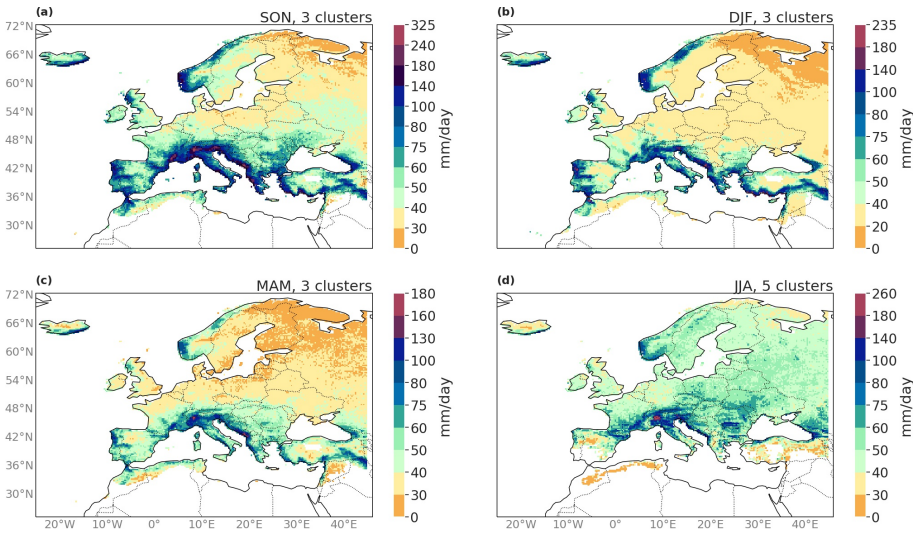

Figure 5: 100-year return levels computed with the regional fitting, see Table 1

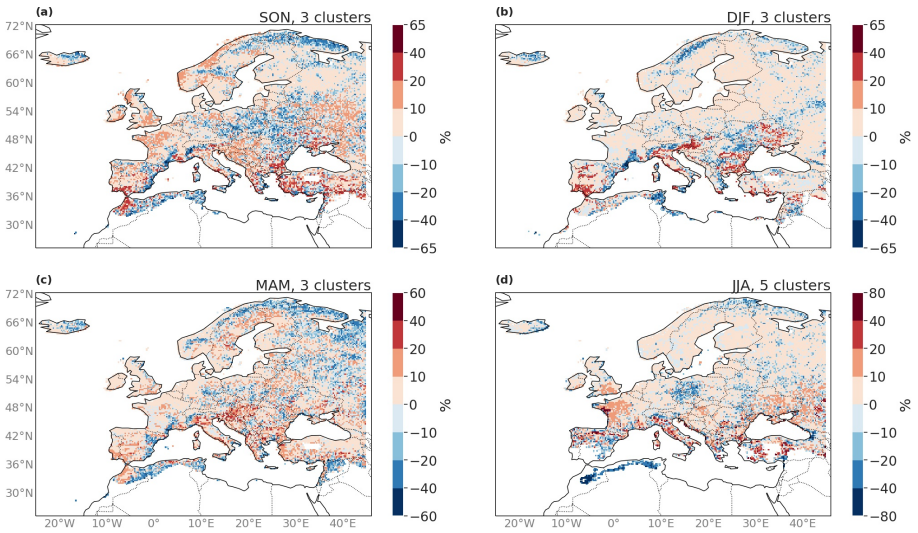

Figure 6: Relative difference between the 50-year return levels computed with the regional fitting and the local fitting

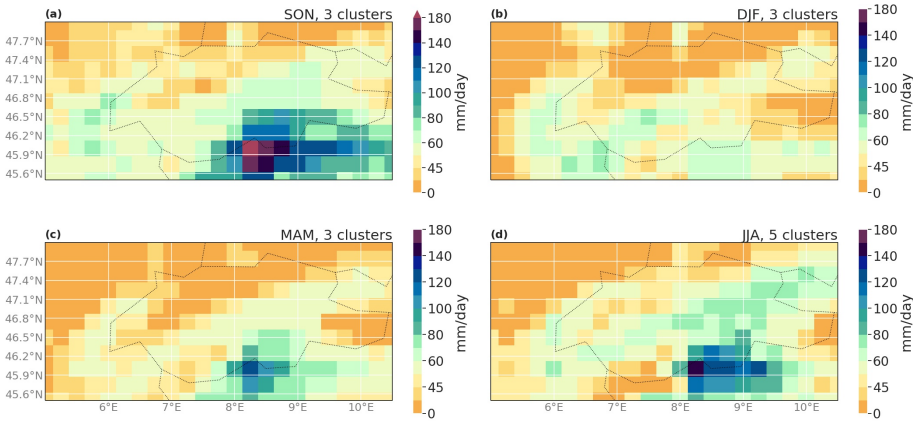

Figure 7: 10-year return levels in Switzerland computed with the regional fitting, for a comparison with the one provided by MeteoSwiss (2019)

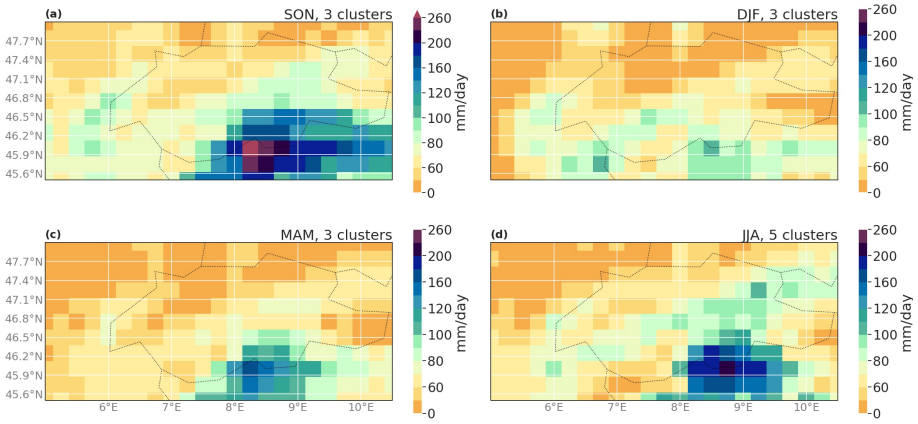

Figure 8: 50-year return levels in Switzerland computed with the regional fitting, for a comparison with the one provided by MeteoSwiss (2019)

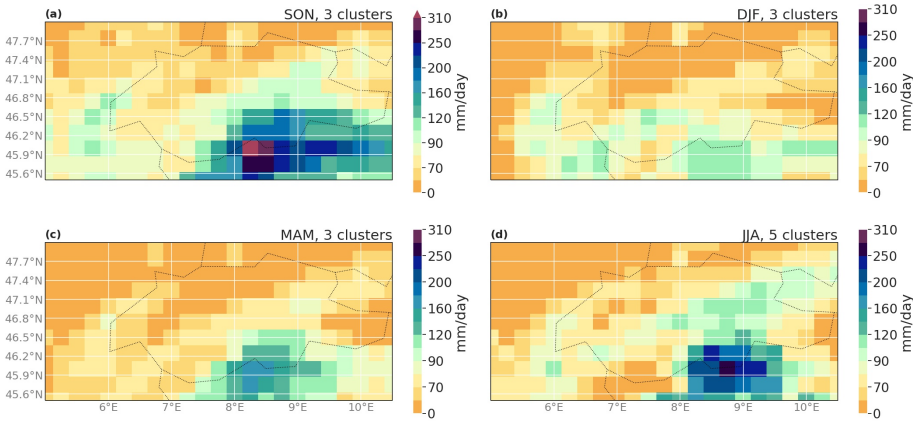

Figure 9: 100-year return levels in Switzerland computed with the regional fitting, for a comparison with the one provided by MeteoSwiss (2019)

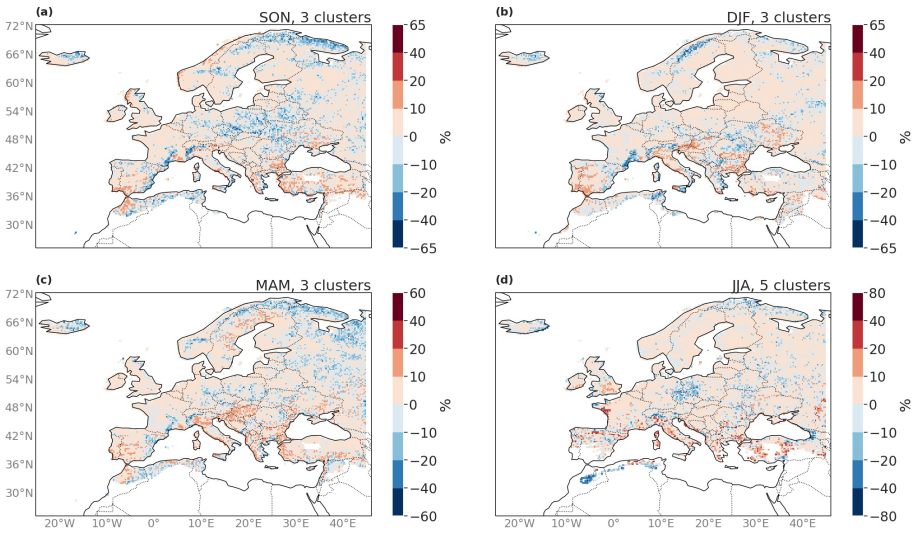

Figure 10: Relative difference between the 10-year return levels computed with the regional fitting and the local fitting

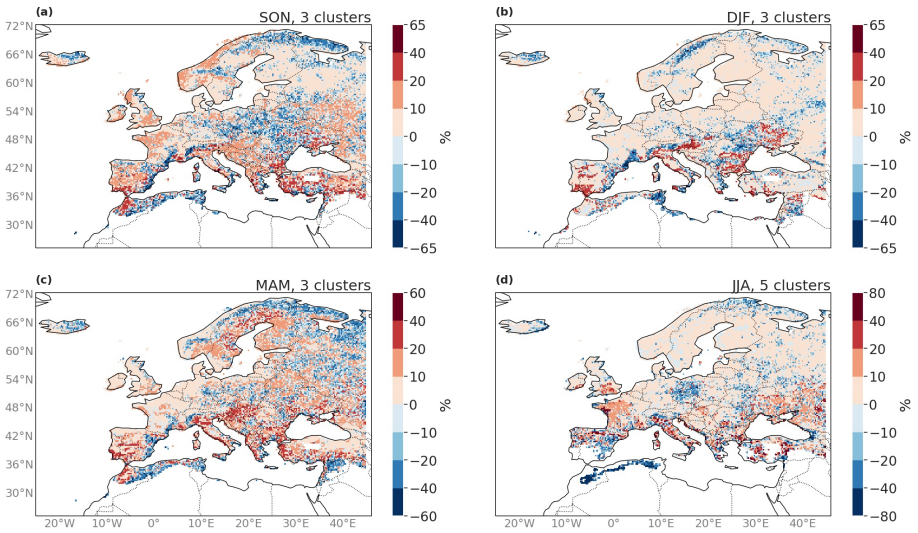

Figure 11: Relative difference between the 100-year return levels computed with the regional fitting and the local fitting