List of all relevant changes

The following major changes have been made to the manuscript:
(R1, R2 and R3 indicate suggestions by reviewer 1, 2 and 3, respectively).

1. The title of the manuscript was changed (R1).
2. All computations were made from the beginning using an informative prior for the shape parameter (R1, R3). Fig. 1-8, S7-S10 and Tables 1, 2, 6 were updated. No major changes were observed in spatial patterns. Shape parameters were constrained to reasonable limits.
3. Table 4 with summary statistics of the shape parameter was added (R1).
4. The median flood quantiles were used as point estimate instead of the modal estimate used before (R1). Figures 5, 6, 7 and Table 6 were updated. Results are similar to those in the initial manuscript but uncertainty bounds in Fig. 7 are less skewed.
5. A probability of exceedance 0.02 instead of 0.01 was used for the calculation of flood quantiles (R1, R3). Figures 5, 6, 7 and Table 6 were updated. Spatial patterns are similar to those in the initial manuscript. The absolute values of percent level differences in flood quantiles were decreased in comparison to a probability 0.01.
6. Table S1 was added in the supplementary material, showing correlations between seasonal climate indices (R2).
7. Fig. 8, comparing streamflow quantiles for the classical and a conditional model for a single station, was added (R2).
8. Fig. S5 was added in the supplementary material. This figure shows the evolution in time of the climate indices (R2), along with their decadal-scale variability (R1).
9. Fig. S6 showing histograms of the seasonal climate indices was added in the supplementary material (R2).
10. Extensive changes in discussion and conclusions were made (R1, R2, and R3).

In addition to the above major changes, numerous smaller changes have been made and are shown in the marked manuscript (R1, R2, and R3).

Climate influences on flood probabilities across Europe~~Do climate-informed extreme value statistics improve the estimation of flood probabilities in Europe?~~

Eva Steirou[1], Lars Gerlitz[1], Heiko Apel[1], Xun Sun[2, 3], Bruno Merz[1, 4]

[1]Section Hydrology, GFZ German Research Center for Geosciences, Potsdam, 14473, Germany
[2]Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, 200241, Shanghai, China
[3]Columbia Water Center, Earth Institute, Columbia University, New York, NY 10027, USA
[4]Institute of Environmental Science and Geography, University of Potsdam, Potsdam, 14476, Germany

Correspondence to: Eva Steirou (esteirou@gfz-potsdam.de), Xun Sun (xs2226@columbia.edu)

Abstract. The link between streamflow extremes and climatology has been widely studied during the last decades. However, a study investigating the effect of large-scale circulation variations on the distribution of seasonal discharge extremes at the European level is missing. Here we fit a climate-informed Generalized Extreme Value distribution (GEV) to about 600 streamflow records in Europe for each of the standard seasons, i.e. to winter, spring, summer and autumn maxima, and compare it with the classical GEV with parameters invariant in time. The study adopts a Bayesian framework and covers the period 1950 to 2016. Five indices with proven influence on the European climate are examined independently as covariates, namely the North Atlantic Oscillation (NAO), the East Atlantic pattern (EA), the East Atlantic / West Russian pattern (EA/WR), the Scandinavia pattern (SCA) and the Polar-Eurasian pattern (POL).

It is found that for a high percentage of stations the climate-informed model is preferred to the classical model~~, a result that provides evidence towards an improvement of the estimation of flood probabilities~~. Particularly for NAO during winter, a strong influence on streamflow extremes is detected for large parts of Europe (preferred to the classical GEV for ~~44~~46% of the stations). Climate-informed fits are characterized by spatial coherence and form patterns that resemble relations between the climate indices and seasonal precipitation, suggesting a prominent role of the considered circulation modes for flood generation. For certain regions, such as Northwest Scandinavia and the British Isles, yearly variations of the mean seasonal climate indices result in considerably different extreme value distributions and thus in highly different flood estimates for individual years that ~~. Plots of extreme streamflow with a probability of exceedance of 0.01 indicate that the deviation between the classical and climate-informed analysis concerns single years but~~ can also persist for longer time periods.

1. Introduction

The understanding of extreme streamflow is a key issue for infrastructure design, flood risk management and (re-) insurance, and the estimation of flood probabilities has been in the focus of the scientific debate during recent decades. Traditionally, streamflow has been analyzed~~sed~~ with regard to associated hydro-climatic processes acting at the catchment scale. During recent years many studies have additionally focused on the link between local streamflow and larger-scale climate mechanisms, extending beyond the catchment boundaries (Merz et al., 2014). An early example can be found in Hirschboeck (1988), who provides a detailed explanation of relationships between floods and synoptic patterns in the USA. Large-scale atmospheric patterns acting at global or continental scales have been shown to significantly influence flood magnitude and frequency at the local and regional scale. Regional in this context refers to the joint consideration of several gauges. For example, Kiem et al. (2003) stratified a regional flood index in Australia according to quantiles of the El Niño/Southern Oscillation (ENSO) index and showed that La Niña events are associated with a distinctly higher flood risk compared with El Niño events. Ward et al. (2014) found that peak discharges are strongly influenced by ENSO for a large fraction of catchments across the globe. Delgado et al. (2012) detected a dependence between the variance of the annual maximum flow at stations along the Mekong River and the intensity of the Western Pacific monsoon.

This perception of climate-influenced extremes has been incorporated in flood frequency analysis by including climatic variables as covariates of extreme value distribution parameters. It is therefore assumed that the probability density function (pdf) of streamflow is not constant in time but it is conditioned on external variables. This framework, usually called nonstationary, can be particularly useful for hydro-climatic studies since the influence of the climatic phenomena on the distribution of the hydrological target variable, such as extreme streamflow, can be considered (Sun et al., 2014). This means that the whole distribution as well as certain parts of the target variable distribution, such as the tails, can be assessed including the influence of the large scale climate phenomenon, and used for flood risk management~~, engineering design~~ or reinsurance purposes. This conditional or nonstationary frequency analysis has been popularized in the field of hydrology and flood research during recent years. Different covariate types have been examined for their influence on flood extremes, e.g. time (e.g. Delgado et al., 2010; Sun et al., 2015), snow cover indices (Kwon et al., 2008), reservoir indices (López and Francés, 2013; Silva et al., 2017), population measures (Villarini et al., 2009) and large-scale atmospheric and oceanic fields and indices (Delgado et al., 2014; Renard and Lall, 2014). A review of nonstationary approaches for local frequency analyses is given by Khaliq et al. (2006), while some of their limitations are discussed by Koutsoyiannis and Montanari (2015) and Serinaldi and Kilsby (2015) and ~~(~~Serinaldi et al. (~~,~~2018).

In this study, we focus on the European continent and the relation between streamflow extremes and the large-scale atmospheric circulation. The European climate is mainly influenced by pressure patterns acting at the broader region covering Europe~~,~~ and the northern Atlantic. In particular, five circulation modes have been shown to significantly modify the moisture fluxes into the European domain: the North Atlantic Oscillation (NAO), the East Atlantic (EA), the East Atlantic/Western Russia (EA/WR), the Scandinavia (SCA) and the Polar/Eurasia (POL) patterns (Bartolini et al., 2010; Casanueva et al., 2014; Rust et al., 2015; Steirou et al., 2017). These patterns represent the first five pressure modes north of 50°, derived by means of a rotated principle component analysis of monthly mean 500hPa geopotential height fields (Barnston and Livezey, 1987). The modes indicate the position and magnitude of large-scale atmospheric waves and thus control the strength and location of the northern hemispheric Jetstream. All modes are characterized by a particular pattern of large-scale winds and moisture fluxes and strongly affect near-surface climate conditions over vast parts of the northern hemisphere. Particularly NAO has been shown to significantly influence the European winter climate: its positive state has been linked to ~~with~~ positive (negative) anomalies of moisture fluxes, cyclone passages and precipitation over northern (southern) Europe ~~during its positive state~~ (Hurrell and Deser, 2009; Wibig, 1999). A seasonal shift of the NAO pressure ~~centres~~centers and moisture fluxes towards north during summer has been detected (Hurrell and Deser, 2009). EA, often referred to as a southward shifted NAO, is characterized by distinctly defined geopotential height anomalies and an associated influence on westerly moisture fluxes and local climate conditions over Great Britain (Comas-Bru and McDermott, 2014; Moore and Renfrew, 2012). EA/WR features two ~~centres~~centers of action over Central Europe and Central Russia. During its positive state, a planetary ridge is located over north-western Europe, ~~which~~ and this reduces the advection of moist air masses (Krichak and Alpert, 2005). SCA is particularly active over northern Europe and triggers atmospheric blocking during its positive phase (Bueh and Nakamura, 2007). POL represents the strength of the pressure gradient between the polar regions and the mid-latitudes and thus controls the westerly circulation, particularly over northern Europe (Claud et al., 2007). Correlation maps, demonstrating links between these circulation modes and seasonal precipitation and temperature, are included in the Supplementary Material (Fig. S1-S4).

Apart from Northern Hemisphere modes, the El Niño-Southern Oscillation (ENSO) has been suggested to influence the European hydrology. Significant relations have been found with precipitation and different discharge indices (Guimarães Nobre et al., 2017; Mariotti et al., 2002; Steirou et al., 2017). However, in contrast to the above described circulation modes, ENSO does not shape the European climate and hydrology directly, but rather indirectly through the regulation of the phase of other large-scale modes, such as the EA (Iglesias et al., 2014). Other patterns acting at a smaller scale, such as the

Mediterranean Oscillation (MO) and the Western Mediterranean Oscillation (WMO), have also been related with hydrological variables in Europe (Criado-Aldeanueva and Soto-Navarro, 2013; Dünkeloh and Jacobeit, 2003; Martin-Vide and Lopez-Bustins, 2006). However, such modes seem to have limited importance at the continental scale.

While the relation between European hydrology and large-scale circulation has attracted much attention and has been widely studied, only few studies have adopted a conditional flood frequency framework for the investigation of climate-flood interactions. Villarini et al. (2012) conducted a frequency analysis of annual maximum and peak-over-threshold discharge in Austria with NAO as a covariate. López and Francés (2013) examined maximum annual flows in Spain conditioned on the principal components of four winter climate modes: NAO, AO, MO and WMO. Still, a comprehensive study on streamflow extremes at the European scale has not ~~yet~~ been conducted.

Thus, this study aims at a large-scale investigation of circulation-streamflow interactions for the entire European continent by adopting a flood frequency framework. We examine seasonal streamflow maxima from more than 600 gauges covering the entire European continent and particularly investigate the influence of the five major pressure modes that directly affect the European climate: NAO, EA, EA/WR, SCA and POL. In order to quantify the effect of important hydro-climatological processes for the streamflow regimes, we investigate contemporaneous relationships only, without considering any time lags. We identify regions with a consistent influence of each particular circulation index in order to explain the spatial coherence of flood frequency. The analysis is conducted at a seasonal scale in order to better account for the intra-annual variations of the circulation characteristics and the associated seasonal shift of climate-streamflow relationships. A Bayesian framework is adopted for the flood frequency analysis because of its advantages concerning the quantification and interpretation of uncertainty. Furthermore, prior information about hydrologic extremes exists in ~~l~~the literature and can be used for inference.

2. Data and Methods

2.1 Streamflow data and circulation indices

The time period of our analysis is from 1950 to 2016, defined by the overlap between streamflow data and circulation indices. Daily streamflow data for the European continent were received from GRDC (Global Runoff Data Centre~~, 2017~~). From this dataset, gauges with record lengths of at least 50 years after 1950 and with a catchment area larger than 200 km^2 were selected. Small catchments are not considered, as they may be more prone to local phenomena, which could blur the large-scale atmospheric influence. In total, 649 stations covering North and Central Europe with the exception of Poland are considered. Due to the underrepresentation of Southern Europe, additional data from other sources satisfying the above mentioned criteria are included in the analysis. Five time series with monthly maximum discharges were obtained for Spain and one station with daily discharge was provided for Portugal. For details about these additional stations the reader is referred to Mediero et al. (2014, 2015)~~, respectively~~. Finally, one record with daily streamflow data was provided for Pontelagoscuro in Italy (Domeneghetti, 2017, personal communication). For each station, the maximum value of mean daily streamflow is derived for the four standard boreal seasons: winter (DJF), spring, (MAM), summer (JJA) and autumn (SON). Seasons with more than 20% missing values are not considered. Overall 586 records in winter, 604 in spring, 599 in summer and 597 for the autumn season are utilized for the analysis.

Time series of monthly circulation indices for the period 1950-2016 were retrieved from the Climate Prediction Center (CPC) of the National Oceanic and Atmospheric Administration (NOAA), (http://www.cpc.ncep.noaa.gov/data/teledoc/telecontents.shtml). We make use of the five indices mentioned in the introduction, namely, the NAO, EA, EA/WR, SCA and POL patterns. Seasonal mean climate indices are used for the adjustment of the extreme value distribution, however, we also examine whether the results differ if monthly values (in accordance with the observed flood date) are considered as covariate. The time series of the seasonal indices, along with their running mean for a 10-year window, are shown in Fig. S5. Histograms showing the distribution of mean circulation indices for each season are provided in Fig. S6.

2.2 Flood frequency analysis – Competing models

The GEV with parameters invariant in time and with parameters conditioned on the climate indices are fitted to the seasonal maximum streamflow data. For the climate-informed models the condition of independent and identically distributed observations of the classical GEV is relaxed to include parameters conditioned on time-varying covariates (Katz et al., 2002). For the two types of models we use the terms "classical model" instead of stationary model and "climate-informed model" rather than "nonstationary model"~~, respectively~~. It has been suggested that if covariates have a stochastic structure and no deterministic component, the resulting distribution is not truly nonstationary (Montanari and Koutsoyiannis, 2014; van Montfort and van Putten, 2002; Serinaldi and Kilsby, 2015). As our climate covariates have no distinguishable deterministic component (~~analysed, but~~ not shown), it is consequently not clear if they result in nonstationary models.

Here e~~E~~ach streamflow gauge is handled independently and site-specific parameters are derived. Let $Y(t)$ denote a streamflow observation at time t and $\boldsymbol{Y} = (Y(t_1), Y(t_2), \ldots, Y(t_n))$ denote the vector of streamflow observations at a specific site. Then ~~F~~for the classical case the model is given as:

$$Y(t) \sim GEV(\boldsymbol{\theta}) \tag{1}$$

where ~~***Y* is the vector of streamflow observations at a specific site and**~~ $\boldsymbol{\theta}$ is the vector of length m of (time-invariant) distribution parameters. The classical GEV comprises $m = 3$ parameters~~,~~: ~~namely~~ a location parameter μ, a scale parameter σ and a shape parameter ξ.

In the Bayesian framework, the posterior pdf of the parameter vector is computed as follows, based on ~~the~~ Bayes theorem:

$$f(\boldsymbol{\theta}|\boldsymbol{Y}) \propto f(\boldsymbol{Y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) \tag{2}$$

where f($\boldsymbol{\theta}$) is the prior pdf of distribution ~~regression~~ parameters and f($\boldsymbol{Y}|\boldsymbol{\theta}$) is the likelihood function:

$$f(\boldsymbol{Y}|\boldsymbol{\theta}) = \prod_t f(Y(t)|\boldsymbol{\theta}) \tag{3}$$

For the climate-informed distribution, parameters are assumed to be a function h_i of the vector of time-varying climate covariates $\boldsymbol{x}(t)$. In the general case, Eq. (1) takes the form:

$$Y(t) \sim GEV(\boldsymbol{\theta}(t)) \tag{4}$$

with $\boldsymbol{\theta}(t) = (\theta_1(t), \theta_2(t), \ldots, \theta_m(t))$ the collection of m distribution parameters at time t, and

$$\theta_i(t) = h_i(\boldsymbol{x}(t); \boldsymbol{\beta_i}) \qquad i = \{1,2,\ldots,m\} \tag{5}$$

Here $\boldsymbol{\beta_i}$ is the vector of (internal) parameters used in function h_i (not to be confused with parameters θ_i).

The climate-informed GEV is a generalization of the classical GEV. The likelihood function is then defined as:

$$f(\boldsymbol{Y}|\boldsymbol{\theta}) = \prod_t f(Y(t)|\theta(t)) = \prod_t f(Y(t)|h_1(\cancel{\boldsymbol{x_t}}\boldsymbol{x}(t), \boldsymbol{\beta_1}), h_2(\boldsymbol{x}(t)\cancel{\boldsymbol{x_t}}, \boldsymbol{\beta_2}), \ldots, h_m(\boldsymbol{x}(t)\cancel{\boldsymbol{x_t}}, \boldsymbol{\beta_m})) \tag{6}$$

The function h_i, linking the distribution parameters with climate covariates, is derived by means of a linear regression. ~~Due to the brevity of observational records, we only examine conditional extreme value distributions with a time varying location parameter. A preliminary analysis considering the effect of a covariate on both the location and scale parameter did not improve the results (not shown).~~ The shape parameter is assumed to be constant as its estimation includes large uncertainties, even under the assumption of stationarity (Coles, 2001, Papalexiou and Koutsoyiannis, 2013; Silva et al., 2017). A preliminary analysis considering the effect of a covariate on both the location and scale parameter (cf. section 2.3 below) did not provide very different results than those for a covariate on the location parameter only (not shown). Consequently and for reasons of parsimony, we examine only conditional extreme value distributions with a time-varying location parameter.

~~We derive c~~Conditional distributions of only one covariate at a time are derived, since we are interested in the separate effect of each individual climate index on flood quantiles.

Based on the above mentioned assumptions concerning model structure and the form of the function h_i, Eq. (5) can be simplified to:

$$\mu(t) = \mu_0 + \mu_1 x(t) \tag{7}$$

where $\mu(t)$ is the varying location parameter, μ_0 the location intercept, μ_1 the location slope and $x(t)$ the single covariate examined.

Consequently, the conditional GEV comprises four parameters: scale~~-~~ and shape parameters, and intercept μ_0 and slope μ_1 for the location parameter. Since five different climate covariates $x(t)$ are investigated, we construct six different models (one classical and five conditional) for each station and season. The posterior pdf of parameters in Eq. (2) for both the classical and conditional model is estimated using a No-U-Turn Sampler (NUTS) - Hamiltonian Monte Carlo (HMC) approach ~~(Hoffman and Gelman, 2014),~~ implemented in Rstan, the R interface to Stan (Stan Development Team, 2017). NUTS is an extension to HMC, a Markov chain Monte Carlo (MCMC) algorithm that avoids the random walk ~~u~~behavior and sensitivity to correlated parameters which characterize many MCMC methods (Hoffman and Gelman, 2014). Stan is a state-of-the-art platform for statistical modelling and high-performance statistical computation.

For all covariates and seasons, models are fitted independently. No posterior distributions from the classical approach are used as priors for the climate-informed case. ~~For the fitting of the distribution we use n~~For all models, non-informative uniform priors are used for the location parameter (for both intercept and slope) and for the scale parameter, since no prior information is available. For the shape parameter an informative normal distribution with mean 0.093 and standard deviation 0.12 is used. This distribution is adopted from a global study of extreme rainfall by ~~(~~Papalexiou and Koutsoyiannis~~,~~ (2013), which, to our knowledge, summarizes an analysis of shape parameters using the largest number of stations with hydrological data worldwide. Although rainfall extremes may be characterized by slightly different shape parameter than those of streamflow, our informative prior is very close to the “geophysical prior” of Martins and Stedinger (2000), which is often used to restrict the range of shape parameters based on previous hydrological experience (Renard et al. 2013). The latter prior was not preferred because it is bounded to the interval (-0.5, 0.5), while the distribution of Papalexiou and Koutsoyiannis (2013) allows more extreme shape values with a low probability.

Five chains of 14,000 simulations, with the first half discarded as warmup period, are run for all parameters. Convergence is investigated by the potential scale reduction statistic~~,~~ $\hat{R}$ (Gelman and Rubin, 1992). Following Gelman (1996), we assume convergence for values of $\hat{R}$ below 1.2. Thinning is applied to the post-warm up simulations to remove autocorrelation. Every tenth value from all chains is kept, leading to a final sample of 3,500 simulations for all ~~five conditioned~~each models ~~for every~~and season.

2.3 Model selection

We apply a two-step methodology to select the optimal model among the classical and conditional competitors.

First, we assess if the covariates have a significant effect on our extreme streamflow models by examining the posterior distribution of the slope μ_1 of the location parameters (Eq. 7). Conditional models are considered as significant if the zero value is not included in the 90% posterior interval of the slope parameter (and thus not by means of a significance test). A second criterion is additionally adopted in order to select the distribution with the best performance by taking into consideration that complex models with more parameters tend to fit the data better. The Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) is chosen for model selection. The DIC was preferred against two more common tools, the Akaike Information Criterion (AIC; Akaike, 1974) and the Bayesian Information Criterion (BIC; Schwarz, 1978), because it is based on the posterior distribution of the model parameters and thus includes parameter uncertainties, while the AIC and BIC are based on maximum likelihood estimates of parameters.

The deviance, used for the calculation of the DIC, is defined as:

$$D(\boldsymbol{\theta}) = -2log(f(\cancel{y}\boldsymbol{Y}|\boldsymbol{\theta})) \tag{8}$$

where $\boldsymbol{\theta}$ is the parameter vector. The DIC is then given by the following equation:

$$DIC = \overline{D} + p_D \tag{9}$$

where $\overline{D}$ is the expectation of the deviance with respect to the posterior distribution, and $p_D = \overline{D} \cancel{+} - D(\overline{\boldsymbol{\theta}})$ is the effective number of parameters (penalty for model complexity, following Spiegelhalter et al., 2002). $\overline{\boldsymbol{\theta}}$ is a vector of the expectation of parameters $\boldsymbol{\theta}$. Models with smaller DIC values are preferred.

Conditional models satisfying both criteria are preferred to the classical model. The model comparison is performed in two steps: first, for each station and season, each climate-informed competitor is pairwise compared to the classical GEV. Subsequently, the model with the overall best performance is identified.

2.4 Conditional flood quantiles

In the classical or stationary approach one can define the n-year return level as the high quantile of the examined variable for which the probability of exceedance is 1/n. In this case, the same probability of exceedance is assigned to same events in different years. The concept of return period can then be introduced as the reciprocal of the probability of exceedance of a specific value or return level of the examined variable (Cooley, 2013). In engineering practice, return period is often used to communicate risk and is understood either as the expected time interval at which the examined variable exceeds a certain threshold for the first time (average occurrence interval) or as the average of the time intervals between two exceedances of a given threshold (average recurrence interval) (Volpi et al., 2015). When the parameters of the distribution vary in time, as in the nonstationary or conditional frequency analysis, a different probability of exceedance is assigned to different years. In this case, the concept of return period becomes less straightforward to define. Thus, communicating risk by means of probabilities makes more sense (Cooley, 2013). Instead of the classical return levels the term “effective” return levels has been introduced (Gilleland and Katz, 2016) which represents the quantiles of the conditioned distribution under consideration of a particular value of the covariate during a given year.

Here we assess whether the consideration of climatic drivers leads to a significant alteration of flood “effective” return levels or conditional quantiles in individual years. ~~Such an assessment allows to evaluate the applicability of the conditional extreme value analysis for engineering purposes. We quantify d~~Differences of flood quantiles during years with high and medium values of the considered circulation indices are quantified. Since the model is linear, the effect of high and low covariate values on the extreme value distribution quantiles is approximately symmetric~~al~~ (it would be symmetric if the seasonal indices had a symmetric distribution around zero – see Fig. S6) and thus low covariate values are not considered. The 95th and 50th quantile of the considered climate index are chosen as high and medium index values, respectively. Index quantiles are calculated for the entire period 1950-2016.

~~Since a Bayesian framework is used, a certain covariate value does not correspond to a single flood quantile for a given probability of exceedance *p*. In our case, f~~From the No-U-Turn sampling ~~and~~ after thinning, 3,500 post-warm up sets of parameters are obtained, each corresponding to a flood quantile (for a given probability of exceedance ~~*p*~~). The median value of all 3,500 flood quantiles is chosen as a point estimate. The median estimate was preferred to the maximum a posteriori (MAP) estimate because it is more representative of the posterior distribution. ~~From all possible sets of parameters we choose the set corresponding to the maximum likelihood for the calculation of flood quantiles. Parameter uncertainty is not taken into consideration. It should be noticed that Monte Carlo Markov Chain (MCMC) methods, such as the No-U-Turn sampling, are not optimization methods and the maximum likelihood estimate from the sampling may slightly differ from the estimate from the standard maximum likelihood approach. A refinement of the modal estimate can be succeeded with the use of an optimization method to get closer to the posterior mode (Renard et al., 2013). In our case, because of the large size of posterior samples, the two methods were found to converge.~~

Based on this approach, the percent relative difference Y_p of the two flood quantiles for a particular probability of exceedance p, corresponding to the high and medium climate index quantiles, respectively, is calculated as follows:

$$Y_p = \frac{y_{p,h} - y_{p,m}}{y_{p,m}} (\%) \qquad (10)$$

where $y_{p,h}$ is a flood quantile for the probability p, incorporating a high value of the considered climate index (95th quantile). $y_{p,m}$ is the quantile value for the same probability p under consideration of the medium (50th quantile) climate index. The analysis is performed for probability~~ies~~ of exceedance of 0.02 (corresponding to the 50-year return period of the classical case) ~~and 0.01~~.

2.5 Uncertainty analysis

In the previous chapters an automatic methodology for the choice of an adequate model and a discussion of flood quantiles for different covariate values is presented. However, a visual comparison of point estimates and uncertainty intervals of the classical and conditional models can be useful, since it illustrates the differences but also the plausibility and possible drawbacks of the competing models. For this reason, we plot the time series of flood quantiles for a probability of exceedance of 0.~~01~~ 02 for selected gauges and covariates based on both the classical and the climate-informed extreme value distribution. As discussed in the previous section, the median flood quantile for a probability of exceedance of 0.02 is chosen as point estimate (median quantile curve). ~~Point estimates are derived for the set of parameters corresponding to the maximum likelihood (modal curve).~~ Uncertainty of flood quantiles is quantified by means of posterior or credibility intervals, which are the Bayesian equivalent to frequentist confidence intervals, although there exist differences ~~considering~~ in the interpretation of the two types ~~exist~~ (Renard et al. 2013, Gelman et al., 2013).

3. Results

3.1 Spatial patterns of competing models

For all seasonal indices climate-informed models are preferred over the classical distribution for a large number of stations; ~~Percentages~~ percentages of preferred models (based on both the DIC and the significance of the slope of the location parameter) are shown in Table 1 and spatial patterns are mapped in Fig. 1-2. The climate-informed fits form spatial clusters that resemble the correlations between the climate indices and average seasonal precipitation (Fig. S1-S4), while a relation with the correlations of seasonal mean temperature is not straightforward. Particularly for NAO a dipole pattern is evident in winter, with a positive influence on extreme discharge in northern and Central Europe and a negative relationship south of the Alps (Fig. 1). The intra-annual shift of the NAO pressure ~~centres~~centers is well captured. The positive influence of NAO on flood magnitudes during summer is only detected for northern Scandinavia (Fig. 2). Similar dipole structures, resembling the correlations with seasonal mean precipitation, are found for other indices. However, there are some deviations from the precipitation patterns. For example, contradicting results are found in Scandinavia during spring and summer for the SCA index. Scandinavian rivers usually have small catchments and are particularly fed by snowmelt in spring, subsequently in this area, both temperature and precipitation are~~may be~~ important for runoff generation. An opposite sign between correlations with precipitation and the slope of the location parameter can also be found during autumn in north-eastern Germany for the EA index.

NAO is the covariate with the highest number of significant fits in winter (~~44~~46%) and autumn (~~32~~31%) and EA in spring (~~33~~32%) and summer (~~22~~18%). High percentages of preferred climate-informed models are also found for EA and SCA in winter, which is the season where most indices are characterized by their strongest influence on the European climate (Table 1). Worst overall results are found for EA/WR in spring (~~4~~3%) and POL in summer (7%). It can be argued that these two latter cases could occur solely by chance or due to spatial correlation of nearby flood time series;~~,~~ however, results are coherent in space and cover large regions, which suggests a real influence of the circulation modes on the location parameter of the extreme value distributions, restricted though to certain sub-regions of Europe.

Similar spatial patterns are obtained from the same analysis if monthly covariates during the month of the seasonal discharge peaks are examined (Fig ~~S5~~S7-~~S6~~S8). Clusters of stations with positive or negative slopes of the location parameter agree with those for seasonal indices, however in most cases the percentages of preferred fits are lower for the monthly covariates, with ~~the~~ EA/WR in spring being an exception. In particular, the role of NAO in winter and autumn and of EA during the rest of the seasons is less pronounced in the monthly-scale analysis. NAO and SCA are the covariates with the highest number of preferred fits in spring and EA during the rest of the seasons, together with EA/WR in summer (Table 2). Regarding the spatial patterns of preferred fits, deviations from those for seasonal covariates can be found for EA/WR, SCA and POL during spring and summer.

For all indices examined, a percentage of stations between ~~6~~ 5 and 13%, depending on the season and the covariate, are characterized by lower DIC for the climate-informed model although the slope of the location parameter is not ~~statistically~~ significant (illustrated as yellow points in Fig. 1 and 2). Only a few~~No~~ station records, up to three per season and index (not shown in Fig. 1 and 2), are characterized by higher DIC value for the climate-informed model without showing a significant slope. These results indicate that DIC is a weaker criterion for model selection than the slope significance at 10% level.

In order to illustrate the spatial structure of best models, the preferred model (classical or climate-informed) is mapped in Fig. 3 and 4 for each station for seasonal covariates. Spatial patterns do not resemble the pattern of significant fits for separate indices (Fig. 1, 2), since the influence of the selected climate modes on flood frequencies is overlapping for some regions and some of the indices are correlated for particular seasons (Table S1). Winter (summer) is the season with the highest (lowest) overall percentage of preferred climate-informed models: 77% and ~~45~~38%, respectively.

In winter, NAO is the most influential climate mode, being preferred over the other modes for ~~27~~28% of the gauges. The largest influence of NAO on flood frequencies is detected in Central Europe, Great Britain, ~~and~~ parts of Scandinavia and the Iberian Peninsula (Fig. 3). The ~~same~~ first three regions show also a high fraction of SCA-influenced models, which points towards a joint effect of NAO and SCA during winter. The two indices are significantly correlated during this season (Table S1). EA is identified as the best covariate in winter for Great Britain. In spring an expansion of the EA influence towards Central Europe is detected. The NAO influence is shifted to the south during the transition seasons (spring and autumn) and is completely dissolved in summer. Patterns for SCA are heterogeneous throughout the year. The same results but for monthly covariates are shown in Fig. S7 and S8. Spatial patterns resemble those for seasonal covariates. Percentages of preferred climate-informed models are included in Tables 1 and 2.

3.2 Conditional quantiles and uncertainty analysis

In the previous section it is shown~~,~~ that~~,~~ models with monthly covariates do not outperform those with seasonal covariates for most indices and seasons. Hence, quantiles of climate indices are calculated at the seasonal scale only (Table 3). Figures 5 and 6 show the relative differences of seasonal flood quantiles for a probability of exceedance of 0.~~01~~ 02 between a (hypothetical) year with a climate index value equal to the 95th index quantile and a year with an index value equal to the median. ~~The patterns for probabilities of exceedance of 0.02 are very similar (not shown).~~

For a probability of exceedance of 0.~~01~~02, relative differences higher than 20% and up to ~~28~~22% are detected in winter for NAO~~, EA and SCA~~. For the rest of the seasons, maximum relative differences are lower than 20% with highest values for EA/WR in autumn (marginally below 20%). In spring and summer the highest value is considerably lower, ~~about~~ between 11-13% for NAO~~, EA,~~ and SCA ~~and POL~~ in spring and EA and SCA in summer.

A difference of 5-10% is quite common for NAO in winter. For example, a station ~~record~~ with a positive slope of the location parameter and a probability of exceedance of 0.~~01~~ 02 for a maximum seasonal discharge value of 600 m^3/s during years characterized by a medium NAO index~~,~~ has an effective return level between 630 and 660 m^3/s during years with a highly positive NAO state. Particularly for Great Britain and Scandinavia, high relative differences, positive or negative, ~~for different indices~~ are found in winter for different indices. Differences of extreme discharge higher than 10% are characteristic for variations of the EA index in south-eastern Britain and for EA/WR in Norway. Some stations with high differences are also found in Norway and northern Britain for NAO and SCA in spring. Summer is characterized by low relative differences, below 5% for most stations. On the contrary, in autumn clusters of stations with medium to high differences, positive or negative (higher than 5% and locally exceeding 10%), are found in Scandinavia for NAO and EA/WR, in entire northern Europe for EA and in the Alpine region, ~~and~~ southern Great Britain and Norway for SCA.

The high ~~values of~~ relative differences of flood quantiles could partly reflect differences in catchment size or unreasonable posterior values of the shape parameter. A~~Such a~~ link with catchment size was, however, not found (not shown). Posterior shapes for all seasons and indices were further analyzed. Summary statistics of the median shape from the posterior distribution of each fitted model are given in Table 4. Little deviation is observed for different models (classical or climate-informed) during the same season but some inter-season variation is present. No unreasonable values are observed, thus we assume that the use of an informative prior distribution for shape adequately restricts the posterior distributions to~~in~~ reasonable limits.

The results for three selected gauges with high relative differences $Y_{0.0\cancel{1}2}$ are presented in detail. The selected stations cover different characteristic combinations with regard to the investigated season and the considered covariate. The time series of discharge values with a probability of exceedance of 0.~~01~~02 are illustrated for the classical case and the climate-informed case for the three indices with the lowest DIC (Fig. 7). Conditional quantiles are calculated o~~i~~n a year-to-year basis, based on the observed values of the selected climate indices. ~~based on the conditional distribution considering the three indices with the lowest DIC (Fig. 7).~~ Details about the streamflow gauges and the climate-informed fits are given in Table ~~4~~ 5 and ~~Table 5~~6, respectively.

Results show~~,~~ that the conditional and unconditional point estimates and uncertainty bounds can differ considerably, particularly for models with a high relative difference $Y_{0.0\cancel{1}2}$ and a low DIC (subplots A1, B1 and C1 in Fig. 7). Obviously results from the conditional ~~uncertainty bounds~~models vary with time. ~~Remarkably, modal values of conditional and unconditional models also diverge.~~ For example, for the station Asbro 3 in Sweden, strongly different results are obtained by the classical and the NAO-conditional model in winter, particularly for the period 1960-1970, which was dominated by negative NAO conditions and reduced winter precipitation amounts over Northern Europe. The same applies for the station Teston in Great Britain during the period 1960-1980, if EA is considered as a covariate. These results show that the climate-informed models can modulate the estimated flood risk for single years or longer periods and thus substantially deviate from the estimation based on the classical distributions.

For models characterized by small relative differences or insignificant slopes of the location parameter (subplots A3, B3 and C3), conditional uncertainty bounds tend to converge to a straight line resembling the classical case.

The classical case is theoretically a subcase of the climate-informed model. However, the two models are fitted independently and the two intervals do not always overlap.

The uncertainty bounds of the climate-informed fits can be narrower or wider than those of the classical model. They are also ~~remarkably~~ asymmetric~~al~~, ~~in~~ contrary to uncertainty bounds that result from a method using a normal approximation. Asymmetric~~al~~ intervals are associated with the shape parameter of the GEV and are not uncommon (see for example Zeng et al., 2017). The range of uncertainty bounds reflects an interplay between model complexity and the additional information provided by the more complex models. In ~~f~~Fig. 7, uncertainty bounds are narrower in the case of the "best" conditional models (e.g. subplot A1). Uncertainty increases when extrapolations are made towards high and low index values. This can be more easily observed in Fig. 8. For the classical case, the range is about 94 m^3/s. For the climate-informed case and NAO = 0 (close to its median value) the range is around 70 m^3/s. The range increases to 74 m^3/s for NAO = 1 and to 80 m^3/s for the most extreme observed NAO value (NAO = -2.1). For a NAO value around 3/-3 the range of uncertainty bounds reaches that of the classical model.

4. Discussion and conclusions

This study explored ~~if~~ whether a climate-informed flood frequency analysis provides insights and can improve~~s~~ the estimation of flood probabilities at the European scale. A site-specific model using a Bayesian framework was developed, and five Euro-Atlantic circulation modes were investigated as potential covariates: the North Atlantic Oscillation (NAO), the East Atlantic pattern (EA), the Scandinavia pattern (SCA) and the Polar/Eurasian pattern (POL). Streamflow was ~~analysed~~analyzed at a seasonal time scale in order to account for the variable influence of the circulation modes on the European climate during different seasons of the year. Covariates were averaged and examined at both seasonal and monthly scales, contemporaneous to the season or month of the seasonal streamflow maxima, respectively.

The developed climate-informed models were compared to the classical GEV with time-invariant parameters. For most seasons and covariates investigated, the climate-informed models were preferred over the classical GEV for a high ~~number~~ percentage of stations (around 20% on average), with best results found in winter for NAO and EA, in spring for EA and in autumn for NAO (Table 1). Results were shown to be coherent in space, indicating that certain regions are influenced by particular circulation modes (Fig. 1-4). In winter 77% of the stations were found to be influenced by one of the climate modes which indicates a high potential for an improvement of flood probability estimations by including climate information into extreme value statistics. On the contrary, less than half of the stations examined were significantly affected by at least one of the five large-scale indices during summer season, indicating a rather convective and non-predictable precipitation regime (Table 1).

Based on the variability of the circulation indices, we identified regions that are characterized by preferred climate-informed fits and by steep slopes of the location parameter. For models with significant slopes, variations of the climate indices lead to highly varying flood quantile estimations for the same probability of exceedance. ~~The results indicate, that the inclusion of climate information into the extreme value analysis leads to highly varying flood quantile estimations for different probabilities of exceedance.~~ Particularly for ~~N~~northwest Scandinavia and the British Isles, variations of the climate indices result in considerably different extreme value distributions and thus highly different flood estimates for individual years (Fig. 5-6). This difference in estimates could be partly a result of unreasonable posterior values of the shape parameter, however, the use of an informative prior distribution for shape adequately restricts the posterior distributions to~~in~~ reasonable limits. Plots of extreme streamflow under consideration of a probability of exceedance of 0.~~01~~ 02 indicate that the deviation~~-~~ between the classical and climate-informed analysis concerns not only single years but can also persist for longer time periods (Fig. 7), which reflects the decadal-scale variability of NAO and other large-scale circulation indices (Fig. S5).

Although the circulation indices examined are characterized by a high intra-seasonal variability, the seasonally averaged indices provided in most cases better fits compared with monthly values (Tables 1-2). This should be emphasized, since extreme precipitation events are most likely stronger related ~~with~~ to monthly circulation states, which better represent the moisture fluxes into the target domain. On the contrary, the catchment wetness before the flood event is likely to be influenced by the seasonal mean circulation and the associated precipitation sums. Hence, our result suggests that the skill of climate informed extreme values distributions is to a significant part a consequence of the important link between catchment wetness and flooding. Thus we assume, in line with recent studies (Blöschl et al., 2017; Merz et al., 2018; Schröter et al., 2015) ~~(Blöschl et al., 2017; Merz et al., 2018)~~, that in many regions of Europe, catchment wetness plays an important role for flood generation.

For the selection of the best model among the classical and climate-informed two criteria were adopted: the DIC and the significance of the slope of the location parameter μ_1. For all indices and seasons, the DIC ~~favoured~~favored the climate-informed models over the classical distribution for a larger number of stations compared to the slope significance. DIC has received some criticism for not adequately ~~penalising~~penalizing complex models and tending to choose overfitted models (Silva et al., 2017; Spiegelhalter et al., 2014). Our results show that at least compared to the slope significance, DIC is a weaker criterion for model selection. A criterion comprising a higher penalty term for model complexity could alternatively be adopted. A more conservative version of DIC has been proposed by Ando (2011) but is not commonly used until today (Silva et al., 2017).

The described methodology can be complemented in several ways.

(a) ~~Number of covariates~~Regional framework

In this study, a local, site-specific flood frequency model was developed. This model allowed to identify spatial coherence in relations between streamflow extremes and large-scale atmospheric patterns. However, a shortcoming of this methodology is the high uncertainty of streamflow estimates for high probabilities of exceedance (corresponding for example to the 100- or 200-year flood). Instead of a local framework, a regional framework can be alternatively implemented. The latter, by considering all available streamflow information in a region, decreases uncertainty and offers the possibility of improving streamflow quantile estimation.

(b) Alternative models

A linear relationship was assumed between streamflow extremes and the large-scale atmospheric indices. This is ~~of course~~ a simplification of reality and some relations may be over- or underestimated due to existing non-linearities in the climate-streamflow system. ~~Alternative m~~More complex, particularly non-linear models ~~are~~would ~~this~~ also be possible candidates for describing the relation between climate indexes and flood probabilities. ~~,~~ H~~h~~owever, with increasing~~as~~ model complexity ~~increases~~, the chances for model overfitting also increase~~caution concerning overfitting is necessary~~. ~~Here~~In this study we assumed ~~A~~a symmetrica~~l~~ influence of the positive and negative phases of the climate indices ~~on the extreme value distribution has been assumed in this study~~. However, an asymmetrica~~l~~ relation may better describe the effect of certain climate modes on streamflow extremes. For example, Sun et al. (2014) used an asymmetric piecewise-linear regression to account for the different effects of El Niño and La Niña on rainfall extremes in Southeast Queensland, Australia. Furthermore, ~~in this study~~ we also assumed a varying location parameter and ~~stable~~constant scale parameter. A constant coefficient of variation as in ~~(~~Serago and Vogel (~~,~~ 2018) would also be possible and as parsimonious as our model. In this case, a varying scale parameter linked to the location parameter would need to be implemented.

Number of covariates

~~(a)~~(c)

Single covariate models were developed, focusing on the separate effect of each individual climate mode. The methodology can be extended to a model considering several covariates at the same time. In that case, dependencies between the covariates, if existent, should be taken into consideration. López and Francés (2013) overcame this problem by using the principal components of climatic indices as covariates for the flood frequency analysis. This, however, increases the model complexity considere~~e~~ably and thus the chances of model overfitting. This needs to be considered in developing models with multiple covariates.

~~(b) Symmetric and asymmetric influence~~

~~(c) A symmetrical influence of the positive and negative phases of the climate indices on the extreme value distribution has been assumed in this study. However, an asymmetrical relation may better describe the effect of certain climate modes on streamflow extremes. For example, Sun et al. (2014) used an asymmetric piecewise-linear regression to account for the different effects of El Niño and La Niña on rainfall extremes in Southeast Queensland, Australia.~~

(d) Contemporaneous and lagged relationships

In this paper we considered contemporaneous relationships between streamflow extremes and pressure modes that directly shape the European climate and hydrology. However, lagged relationships may ~~also~~ prove more useful for flood risk management and the (re-)insurance industry, since they would allow forecasts of temporal variable flood quantiles for the following month or season. The contemporaneous streamflow-covariate setup presented here can be used, together with a seasonal prediction of indices, for an ahead-season forecast of streamflow quantiles. In this case covariate uncertainty must be additionally considered. A second possibility is ~~Therefore, we plan~~ to operate the presented model in a forecast mode under consideration of different time lags between selected covariates and observed streamflow maxima. Our results suggest that catchment wetness has an important role in shaping seasonal maximum streamflow. ~~Thus i~~In a follow up study, we will systematically test the skill of various predictor variables, describing both the climate and catchment state, in forecasting runoff extremes in Europe.

Data availability

The GRDC discharge dataset was obtained from The Global Runoff Data Centre, 56068 Koblenz, Germany (https://www.bafg.de/GRDC/EN, last access in October 2017) and is available upon request.~~,~~ Time series of monthly circulation indices were retrieved from the Climate Prediction Center (CPC) of the National Oceanic and Atmospheric Administration (NOAA) and can be accessed through http://www.cpc.ncep.noaa.gov/data/teledoc/telecontents.shtml. Additional discharge data from Spain and Portugal were provided upon request by Luis Mediero and for Pontelagoscuro, Italy by Alessio Domeneghetti. Gridded pressure data were extracted from the NCEP/NCAR Reanalysis dataset and are provided through http://www.esrl.noaa.gov/psd/. Gridded temperature and precipitation data were extracted from the CRU TS3.24 dataset from the climatic research unit (CRU, https://crudata.uea.ac.uk/cru/data/hrg/) of the University of East Anglia.

Author contributions

B.M. conceived the original idea, and all co-authors designed the overall study. E.S. developed the model code with contributions from X.S, performed the analysis, and prepared the manuscript. All co-authors contributed to the interpretation of the results and writing of the manuscript.

Acknowledgements

The authors are grateful to the three reviewers, Alberto Viglione, Elena Volpi and Francesco Marra for their helpful comments and suggestions that substantially improved the manuscript. ~~The GRDC discharge dataset was obtained from The Global Runoff Data Centre, 56068 Koblenz, Germany.~~ Alessio Domeneghetti is thanked for providing unpublished discharge data from Italy and Luis Mediero for providing discharge data from Spain and Portugal. Daniel Beiter is thanked for his support in coding and parallel computing. ~~Dr.~~ Xun Sun is supported by the National Key R&D Program of China (No. 2017YFE0100700) and Shanghai Pujiang Program (No. 17PJ1402500). This study was conducted in the frame of the projects “Conditional flood frequency analysis: exploring the link of flood frequency to catchment state and climate variations” and “The link of flood frequency to catchment state and climate variations”, two joined research initiatives between AXA Global P&C and GFZ, Potsdam. The authors wish to acknowledge the AXA Research Fund for ~~the~~ financial support.

References

Akaike, H.: New look at statistical-model identification, IEEE Trans. Automat. Control, AC 19 (6), 716–723, 1974.

Barnston, A. G. and Livezey, R. E.: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns, Mon. Weather Rev., 115, 1083–1126, 1987.

Bartolini, E., Claps, P. and D'Odorico, P.: Connecting European snow cover variability with large scale atmospheric patterns, Adv. Geosci., 26, 93–97, doi:10.5194/adgeo-26-93-2010, 2010.

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., Aronica, G. T., Bilibashi, A., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Fiala, K., Frolova, N., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Rogger, M., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Viglione, A., Volpi, E., Wilson, D., Zaimi, K. and Živković, N.: Changing climate shifts timing of European floods, Science (80-.)., 357(6351), 588–590, doi:10.1126/science.aan2506, 2017.

Bueh, C. and Nakamura, H.: Scandinavian pattern and its climatic impact, Q. J. R. Meteorol. Soc., 133(629), 2117–2131, doi:10.1002/qj.173, 2007.

Casanueva, A., Rodríguez-Puebla, C., Frías, M. D. and González-Reviriego, N.: Variability of extreme precipitation over Europe and its relationships with teleconnection patterns, Hydrol. Earth Syst. Sci., 18(2), 709–725, doi:10.5194/hess-18-709-2014, 2014.

Claud, C., Duchiron, B. and Terray, P.: Associations between large-scale atmospheric circulation and polar low developments over the North Atlantic during winter, J. Geophys. Res. Atmos., 112(12), 1–16, doi:10.1029/2006JD008251, 2007.

Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer.

Comas-Bru, L. and McDermott, F.: Impacts of the EA and SCA patterns on the European twentieth century NAO-winter climate relationship, Q. J. R. Meteorol. Soc., 140(679), 354–363, doi:10.1002/qj.2158, 2014.

Cooley, D.: Return periods and return levels under climate change, in Extremes in a Changing Climate, pp. 97–114, Springer, Amster- dam, Netherlands, 2013.

Criado-Aldeanueva, F. and Soto-Navarro, F. J.: The Mediterranean Oscillation Teleconnection Index: Station-Based versus Principal Component Paradigms, Adv. Meteorol., 2013, 1–10, doi:10.1155/2013/738501, 2013.

Delgado, J. M., Apel, H. and Merz, B.: Flood trends and variability in the Mekong river, Hydrol. Earth Syst. Sci., 14(3), 407–418, doi:10.5194/hess-14-407-2010, 2010.

Delgado, J. M., Merz, B. and Apel, H.: A climate-flood link for the lower Mekong River, Hydrol. Earth Syst. Sci., 16(5), 1533–1541, doi:10.5194/hess-16-1533-2012, 2012.

Delgado, J. M., Merz, B. and Apel, H.: Projecting flood hazard under climate change: An alternative approach to model chains, Nat. Hazards Earth Syst. Sci., 14(6), 1579–1589, doi:10.5194/nhess-14-1579-2014, 2014.

Dünkeloh, A. and Jacobeit, J.: Circulation dynamics of Mediterranean precipitation variability 1948-98, Int. J. Climatol., 23(15), 1843–1866, doi:10.1002/joc.973, 2003.

Gelman, A. and Rubin, D. B.: Inference from Iterative Simulation Using Multiple Sequences, Stat. Sci., 7(4), 457–472, doi:10.1214/ss/1177011136, 1992.

Gelman A.: Inference and monitoring convergence. In: Gilks W. R., Richardson S, Spiegelhalter D. J. (eds) Markov chain Monte Carlo in practice. Chapman & Hall, New York, pp 131–143, 1996.

Gelman A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A., Rubin D. B.: Bayesian Data Analysis. 3rd edition. Chapman & Hall/CRC, London, 2013.

Gilleland, E. and Katz, R. W.: extRemes 2.0: An Extreme Value Analysis Package in *R*, J. Stat. Softw., 72(8), doi:10.18637/jss.v072.i08, 2016.

Guimarães Nobre, G., Jongman, B., Aerts, J. and Ward, P. J.: The role of climate variability in extreme floods in Europe, Environ. Res. Lett., 12(8), 084012, doi:10.1088/1748-9326/aa7c22, 2017.

Hirschboeck, K. K.: Flood hydroclimatology. In V. R. Baker (Ed.), Flood geomorphology (pp. 27-49). Wiley-Interscience, 1988.

Hoffman M. D., Gelman A.: The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J. Mach. Learn. Res. 15: 1593–1623, 2014.

Hurrell, J. W. and Deser, C.: North Atlantic climate variability: The role of the North Atlantic Oscillation, J. Mar. Syst., 78(1), 28–41, doi:10.1016/j.jmarsys.2008.11.026, 2009.

Iglesias, I., Lorenzo, M. N. and Taboada, J. J.: Seasonal Predictability of the East Atlantic Pattern from Sea Surface Temperatures, edited by J. M. Dias, PLoS One, 9(1), e86439, doi:10.1371/journal.pone.0086439, 2014.

Katz, R. W., Parlange, M. B. and Naveau, P.: Statistics of extremes in hydrology, Adv. Water Resour., 25(8–12), 1287–1304, doi:10.1016/S0309-1708(02)00056-8, 2002.

Khaliq, M. N., Ouarda, T. B. M. J., Ondo, J. C., Gachon, P. and Bobée, B.: Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review, J. Hydrol., 329(3–4), 534–552, doi:10.1016/j.jhydrol.2006.03.004, 2006.

Kiem, A. S., Franks, S. W. and Kuczera, G.: Multi-decadal variability of flood risk, Geophys. Res. Lett., 30(2), 1035, doi:10.1029/2002GL015992, 2003.

Koutsoyiannis, D. and Montanari, A.: Negligent killing of scientific concepts: the stationarity case, Hydrol. Sci. J., 60(7–8), 1174–1183, doi:10.1080/02626667.2014.959959, 2015.

Krichak, S. O. and Alpert, P.: Decadal trends in the east Atlantic-west Russia pattern and Mediterranean precipitation, Int. J. Climatol., 25(2), 183–192, doi:10.1002/joc.1124, 2005.

Kwon, H.-H., Brown, C. and Lall, U.: Climate informed flood frequency analysis and prediction in Montana using hierarchical Bayesian modeling, Geophys. Res. Lett., 35(5), L05404, doi:10.1029/2007GL032220, 2008.

López, J. and Francés, F.: Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates, Hydrol. Earth Syst. Sci., 17(8), 3189–3203, doi:10.5194/hess-17-3189-2013, 2013.

Mariotti, A., Zeng, N. and Lau, K.-M.: Euro-Mediterranean rainfall and ENSO—a seasonally varying relationship, Geophys. Res. Lett., 29(12), 1621, doi:10.1029/2001GL014248, 2002.

Martin-Vide, J. and Lopez-Bustins, J.-A.: The Western Mediterranean Oscillation and rainfall in the Iberian Peninsula, Int. J. Climatol., 26(11), 1455–1475, doi:10.1002/joc.1388, 2006.

Martins, E. S. and Stedinger, J. R.: Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, Water Resour. Res., 36(3), 737–744, doi:10.1029/1999WR900330, 2000.

Mediero, L., Santillá~~a´~~n, D., Garrote, L. and Granados, A.: Detection and attribution of trends in magnitude, frequency and timing of floods in Spain, J. Hydrol., 517, 1072–1088, doi:10.1016/j.jhydrol.2014.06.040, 2014.

Mediero, L., Kjeldsen, T. R., Macdonald, N., Kohnova, S., Merz, B., Vorogushyn, S., Wilson, D., Alburquerque, T., Blöschl, G., Bogdanowicz, E., Castellarin, A., Hall, J., Kobold, M., Kriauciuniene, J., Lang, M., Madsen, H., Onuşluel Gül, G., Perdigão, R. A. P., Roald, L. A., Salinas, J. L., Toumazis, A. D., Veijalainen, N. and Þórarinsson, Ó.: Identification of coherent flood regions across Europe by using the longest streamflow records, J. Hydrol., 528, 341–360, doi:10.1016/j.jhydrol.2015.06.016, 2015.

Merz, B., Aerts, J., Arnbjerg-Nielsen, K., Baldi, M., Becker, A., Bichet, A., Blöschl, G., Bouwer, L. M., Brauer, A., Cioffi, F., Delgado, J. M., Gocht, M., Guzzetti, F., Harrigan, S., Hirschboeck, K., Kilsby, C., Kron, W., Kwon, H.-H., Lall, U., Merz, R., Nissen, K., Salvatti, P., Swierczynski, T., Ulbrich, U., Viglione, A., Ward, P. J., Weiler, M., Wilhelm, B. and Nied, M.: Floods and climate: emerging perspectives for flood risk assessment and management, Nat. Hazards Earth Syst. Sci., 14(7), 1921–1942, doi:10.5194/nhess-14-1921-2014, 2014.

Merz, B., Dung, N. V., Apel, H., Gerlitz, L., Schröter, K., Steirou, E. and Vorogushyn, S.: Spatial coherence of flood-rich and flood-poor periods across Germany, J. Hydrol., 559, 813–826, doi:10.1016/j.jhydrol.2018.02.082, 2018.

Montanari, A. and Koutsoyiannis, D.: Modeling and mitigating natural hazards: Stationary is immortal, Water Resour. Res., 9748–9756, doi:10.1002/2014WR016092.Received, 2014.
van Montfort, M. A. J. and van Putten, B.: A comment on modelling extremes : Links between Multi-Component Extreme Value and General Extreme Value distributions, J. Hydrol. (New Zealand), 41(2), 197–202, 2002.
Moore, G. W. K. and Renfrew, I. A.: Cold European winters: interplay between the NAO and the East Atlantic mode, Atmos. Sci. Lett., 13(1), 1–8, doi:10.1002/asl.356, 2012.
Papalexiou, S. M. and Koutsoyiannis, D.: Battle of extreme value distributions: A global survey on extreme daily rainfall, Water Resour. Res., 49(1), 187–201, doi:10.1029/2012WR012557, 2013.
Renard, B. and Lall, U.: Regional frequency analysis conditioned on large-scale atmospheric or oceanic fields, Water Resour. Res., 50(12), 9536–9554, doi:10.1002/2014WR016277, 2014.
Renard, B., Sun, X., Lang, M.: Bayesian methods for non-stationary extreme value analysis. In: AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., Sorooshian, S. (Eds.), Extremes in a Changing Climate: Detection, Analysis and Uncertainty. Water Science and Technology Library, Springer, Netherlands, pp. 39–95, 2013.
Rust, H. W., Richling, A., Bissolli, P. and Ulbrich, U.: Linking teleconnection patterns to European temperature – a multiple linear regression model, Meteorol. Zeitschrift, 24(4), 411–423, doi:10.1127/metz/2015/0642, 2015.
Schröter, K., Kunz, M., Elmer, F., Mühr, B. and Merz, B.: What made the June 2013 flood in Germany an exceptional event? A hydro-meteorological evaluation, Hydrol. Earth Syst. Sci., 19(1), 309–327, doi:10.5194/hess-19-309-2015, 2015.
Schwarz, G.: Estimating the dimension of a model. Ann. Stat. 6 (2), 461–464, 1978.
Serago, J. M. and Vogel, R. M.: Parsimonious nonstationary flood frequency analysis, Adv. Water Resour., 112(November 2017), 1–16, doi:10.1016/j.advwatres.2017.11.026, 2018.
Serinaldi, F. and Kilsby, C. G.: Stationarity is undead: Uncertainty dominates the distribution of extremes, Adv. Water Resour., 77, 17–36, doi:10.1016/j.advwatres.2014.12.013, 2015a.
Serinaldi, F. and Kilsby, C. G.: Stationarity is undead: Uncertainty dominates the distribution of extremes, Adv. Water Resour., 77, 17–36, doi:10.1016/j.advwatres.2014.12.013, 2015b.
Serinaldi, F., Kilsby, C. G. and Lombardo, F.: Untenable nonstationarity: An assessment of the fitness for purpose of trend tests in hydrology, Adv. Water Resour., 111, 132–155, doi:10.1016/J.ADVWATRES.2017.10.015, 2018.
Silva, A. T., Portela, M. M., Naghettini, M. and Fernandes, W.: A Bayesian peaks-over-threshold analysis of floods in the Itajaí-açu River under stationarity and nonstationarity, Stoch. Environ. Res. Risk Assess., 31(1), 185–204, doi:10.1007/s00477-015-1184-4, 2017.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A.: Bayesian Measures of Model Complexity anf Fit, J. R. Stat. Soc. Ser. B (Statistical Methodol., 64(4), 583–639, doi:10.1111/1467-9868.00353, 2002.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Linde, A. Van Der: The deviance information criterion: 12 years on (with discussion), J. R. Stat. Soc. Ser. B, 64, 485–493, 2014.
Stan Development Team: RStan: the R interface to Stan. R package version 2.16.2. http://mc-stan.org, 2017.
Steirou, E., Gerlitz, L., Apel, H. and Merz, B.: Links between large-scale circulation patterns and streamflow in Central Europe: A review, J. Hydrol., 549, doi:10.1016/j.jhydrol.2017.04.003, 2017.
Sun, X., Thyer, M., Renard, B. and Lang, M.: A general regional frequency analysis framework for quantifying local-scale climate effects: A case study of ENSO effects on Southeast Queensland rainfall, J. Hydrol., 512, 53–68, doi:10.1016/j.jhydrol.2014.02.025, 2014.
Sun, X., Lall, U., Merz, B. and Dung, N. V.: Hierarchical Bayesian clustering for nonstationary flood frequency analysis: Application to trends of annual maximum flow in Germany, Water Resour. Res., 51(8), 6586–6601, doi:10.1002/2015WR017117, 2015.
Villarini, G., Smith, J. A., Serinaldi, F., Bales, J., Bates, P. D. and Krajewski, W. F.: Flood frequency analysis for

nonstationary annual peak records in an urban drainage basin, Adv. Water Resour., 32(8), 1255–1266, doi:10.1016/j.advwatres.2009.05.003, 2009.

Villarini, G., Smith, J. A., Serinaldi, F., Ntelekos, A. A. and Schwarz, U.: Analyses of extreme flooding in Austria over the period 1951-2006, Int. J. Climatol., 32(8), 1178–1192, doi:10.1002/joc.2331, 2012.

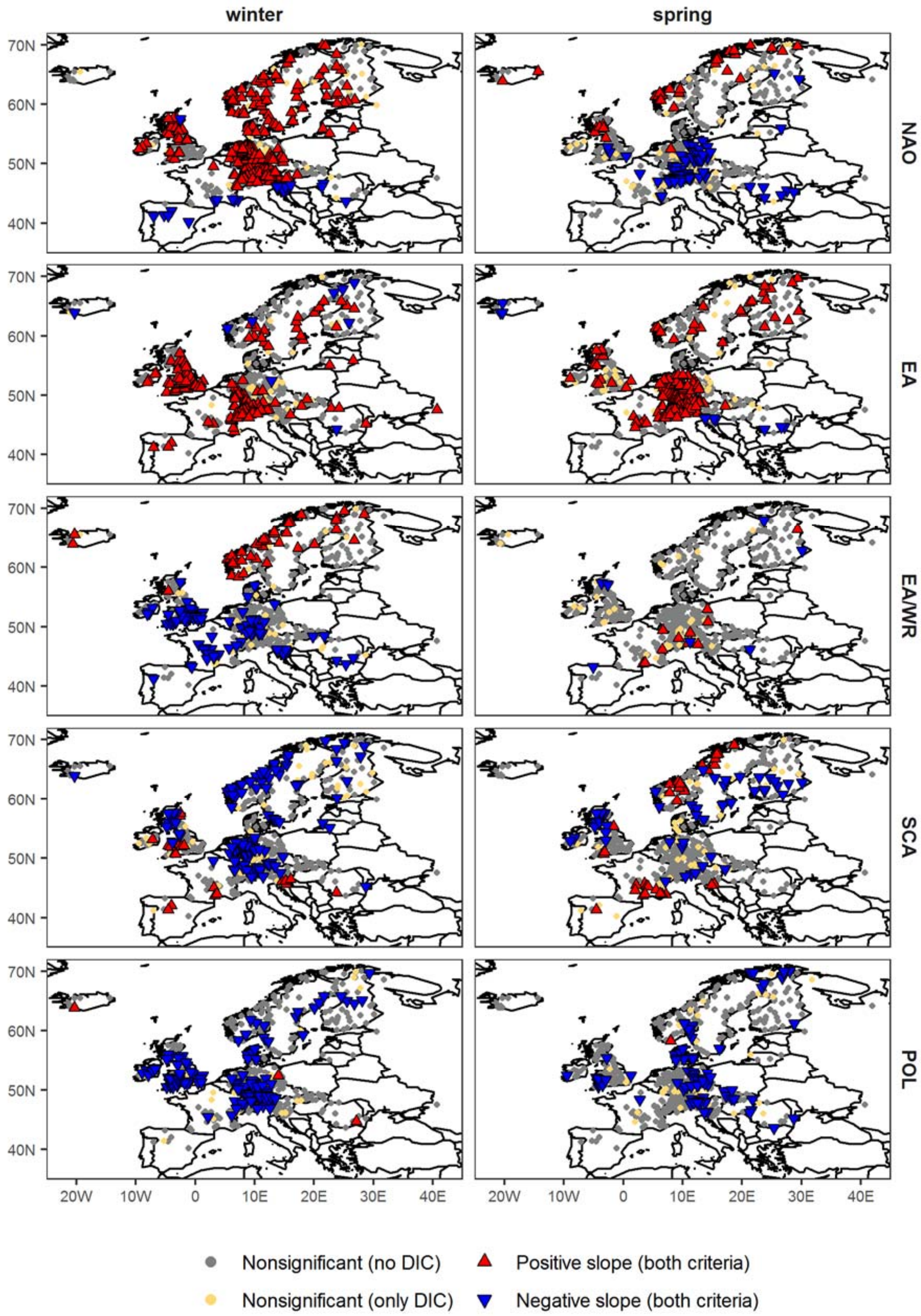
Volpi, E., Fiori, A., Grimaldi, S., Lombardo, F. and Koutsoyiannis, D.: One hundred years of return period: Strengths and limitations, Water Resour. Res., 51(10), 8570–8585, doi:10.1002/2015WR017820, 2015.

Ward, P. J., Eisner, S., Flörke, M., Dettinger, M. D. and Kummu, M.: Annual flood sensitivities to El Niño–Southern Oscillation at the global scale, Hydrol. Earth Syst. Sci., 18(1), 47–66, doi:10.5194/hess-18-47-2014, 2014.

Wibig, J.: Precipitation in Europe in relation to circulation patterns at the 500 hPa level, Int. J. Climatol., 19(3), 253–269, doi:10.1002/(SICI)1097-0088(19990315)19:3<253::AID-JOC366>3.0.CO;2-0, 1999.

Zeng, H., Sun, X., Lall, U. and Feng, P.: Nonstationary extreme flood/rainfall frequency analysis informed by large-scale oceanic fields for Xidayang Reservoir in North China, Int. J. Climatol., 37(10), 3810–3820, doi:10.1002/joc.4955, 2017.

660 **Figures**



winter
spring
NAO
EA
EA/WR
SCA
POL
70N
60N
50N
40N
20W
10W
0
10E
20E
30E
40E
Nonsignificant (no DIC)
Nonsignificant (only DIC)
Positive slope (both criteria)
Negative slope (both criteria)

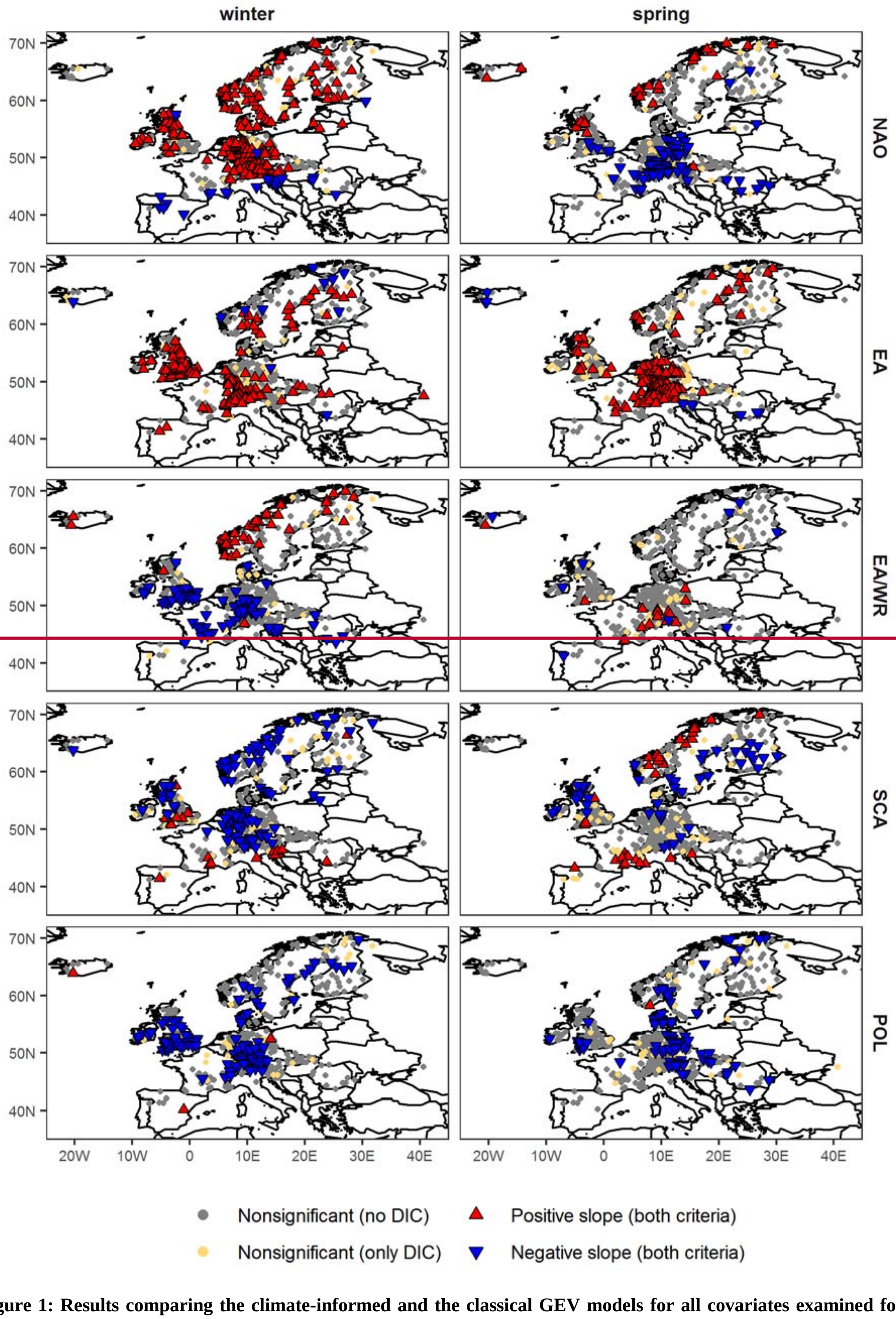
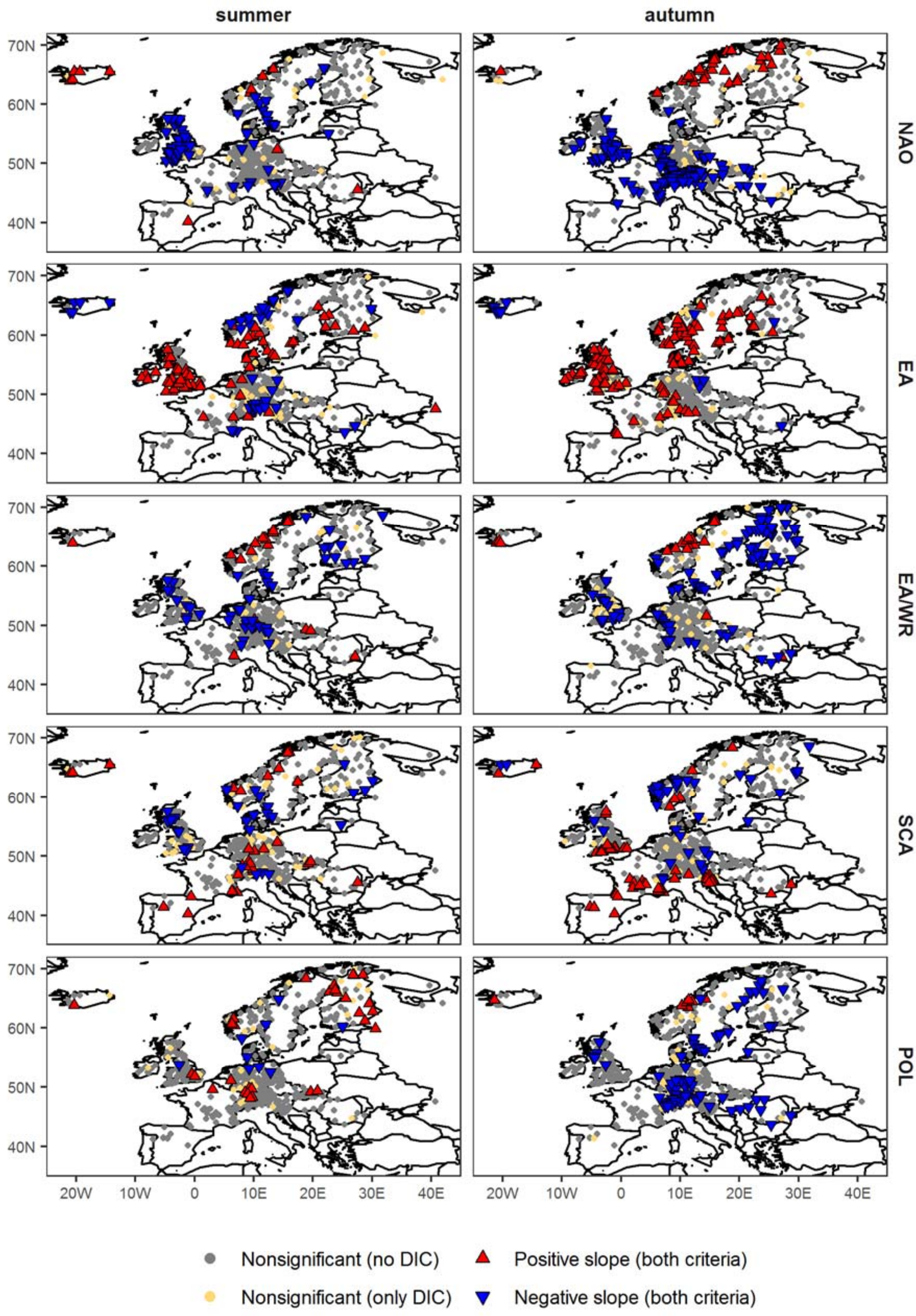


Figure 1: Results comparing the climate-informed and the classical GEV models for all covariates examined for the winter and spring season. Nonsignificant models preferred only by the DIC (yellow points) are plotted on top of stations for which climate-informed models were not chosen by any of the two criteria (grey points). Preferred climate-informed models chosen by both criteria (blue/red triangles) are illustrated on top of the other models so that they can be better distinguished.


summer
autumn
NAO
EA
EAWR
SCA
POL
70N
60N
50N
40N
20W
10W
0
10E
20E
30E
40E
Nonsignificant (no DIC)
Nonsignificant (only DIC)
Positive slope (both criteria)
Negative slope (both criteria)

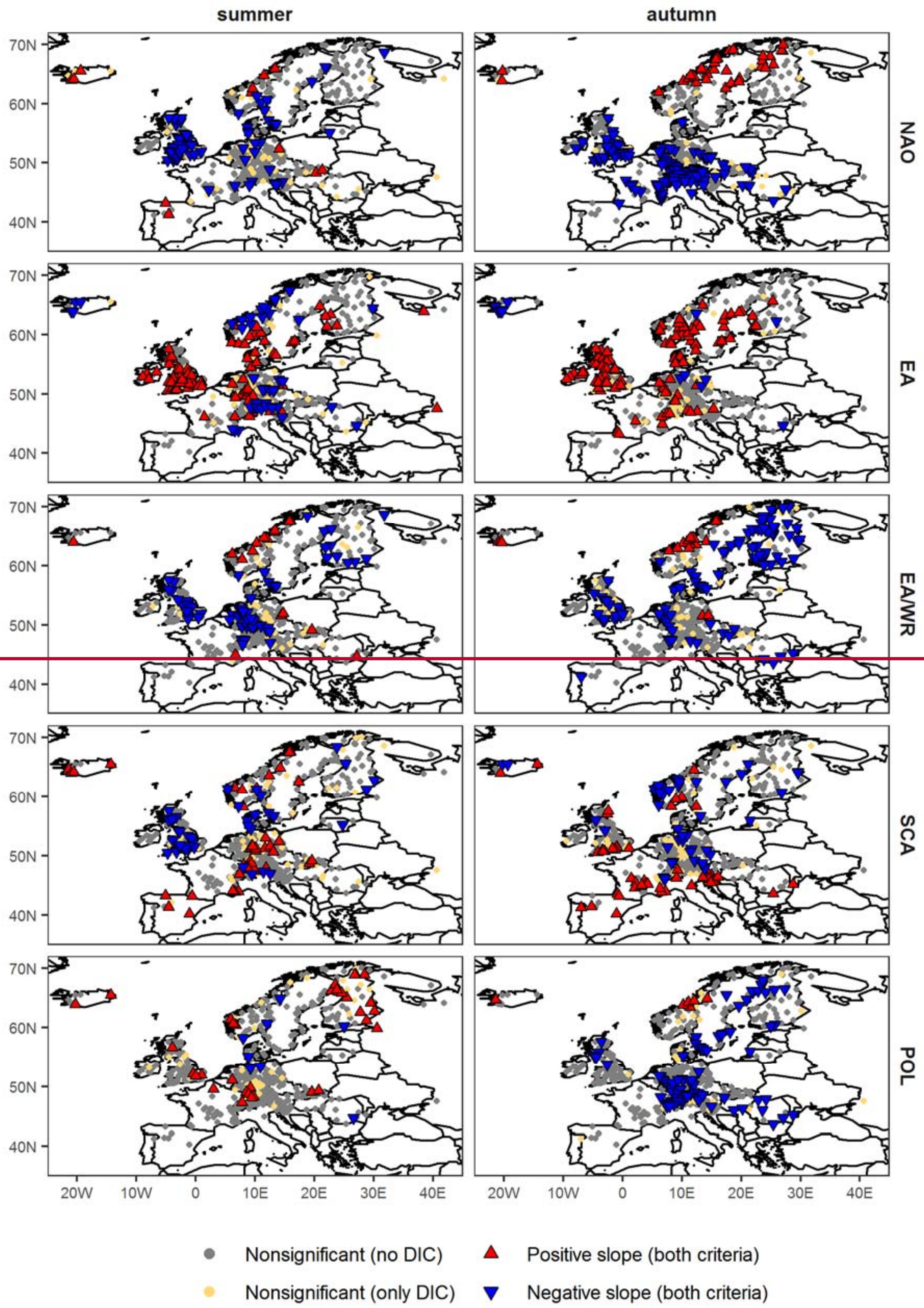
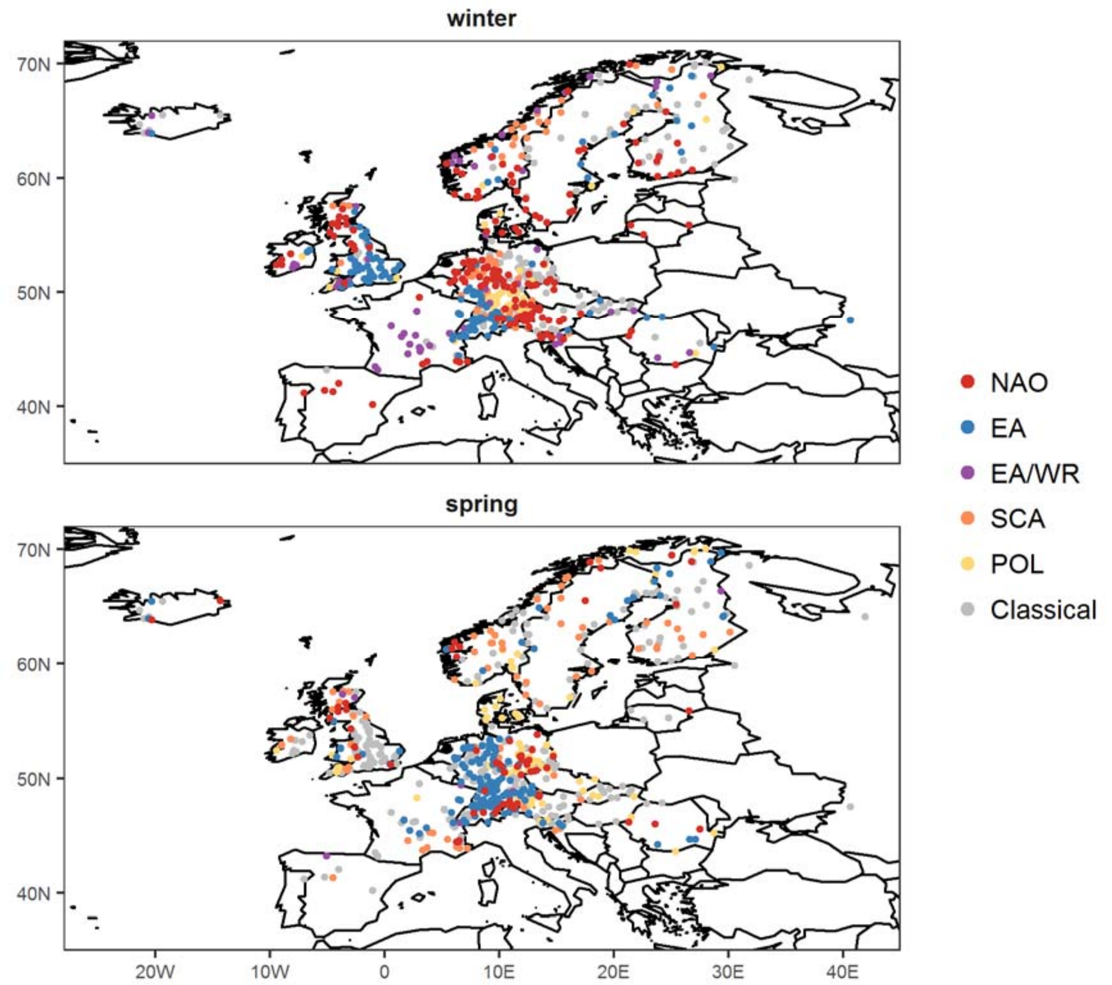


Figure 2: Same as Fig. 1 but for the summer and autumn season.



winter
spring
70N
60N
50N
40N
20W
10W
0
10E
20E
30E
40E
NAO
EA
EA/WR
SCA
POL
Classical

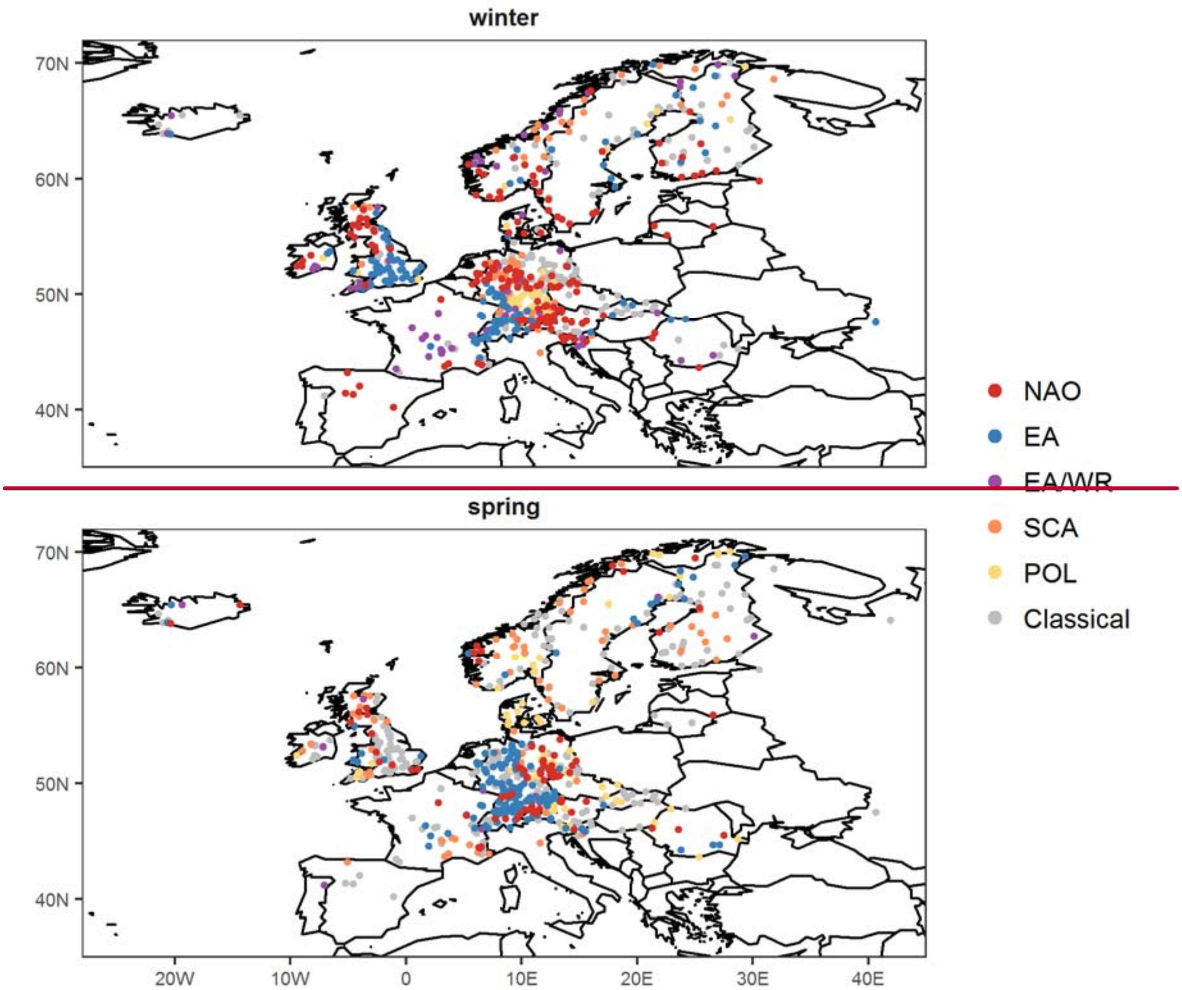
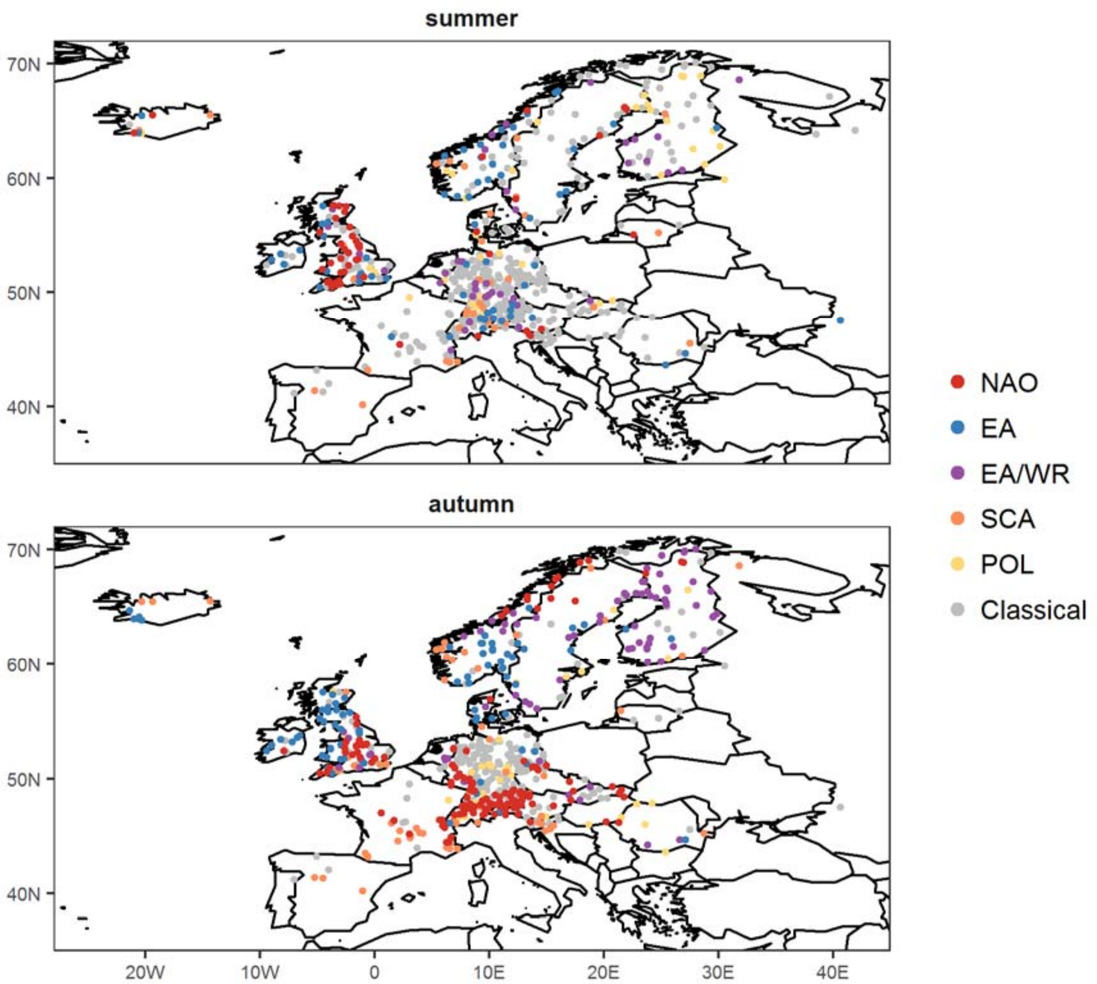


Figure 3: Best overall models among the five climate-informed and classical GEV tested for the winter and spring season. Mean seasonal covariates are examined.


summer
autumn
70N
60N
50N
40N
20W
10W
0
10E
20E
30E
40E
NAO
EA
EA/WR
SCA
POL
Classical

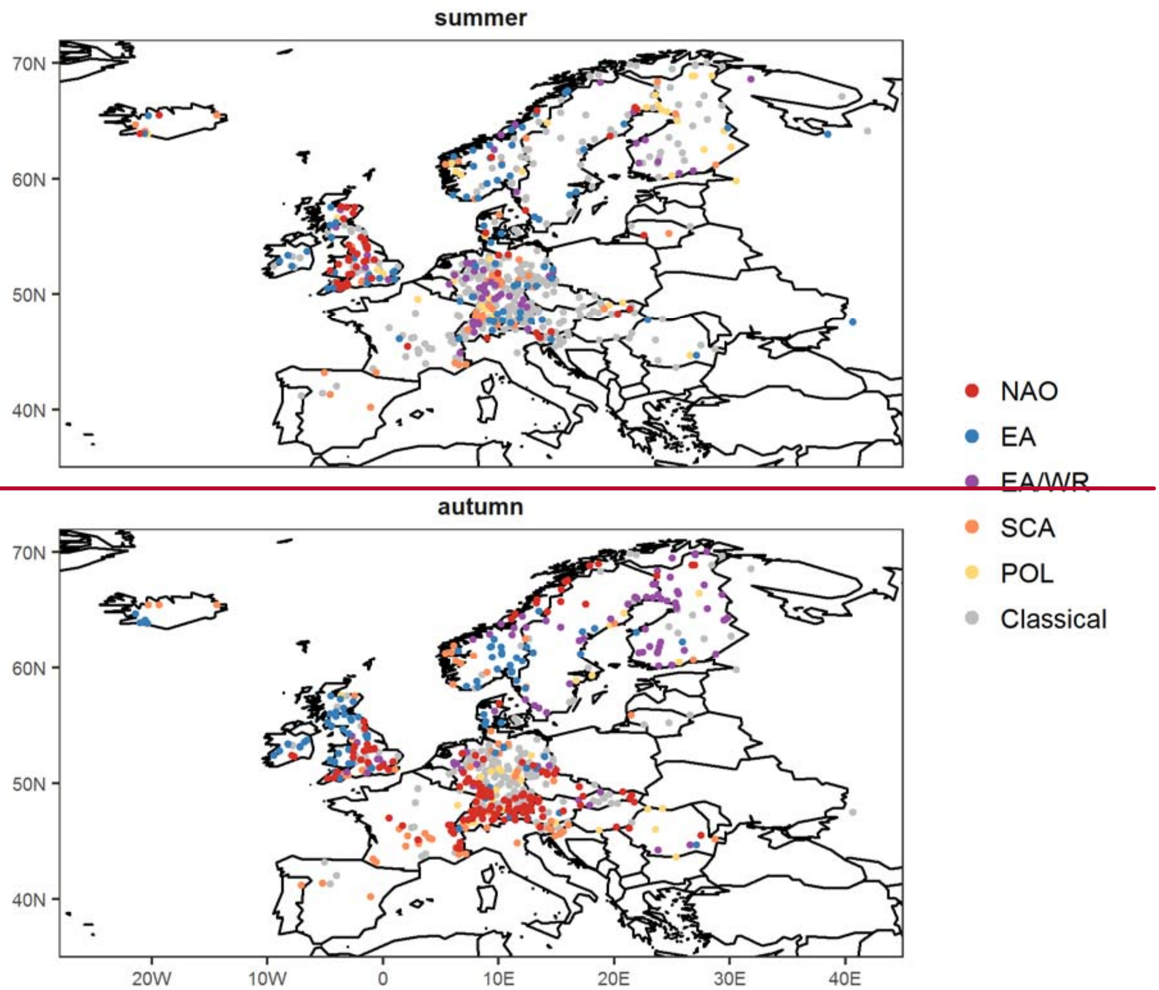
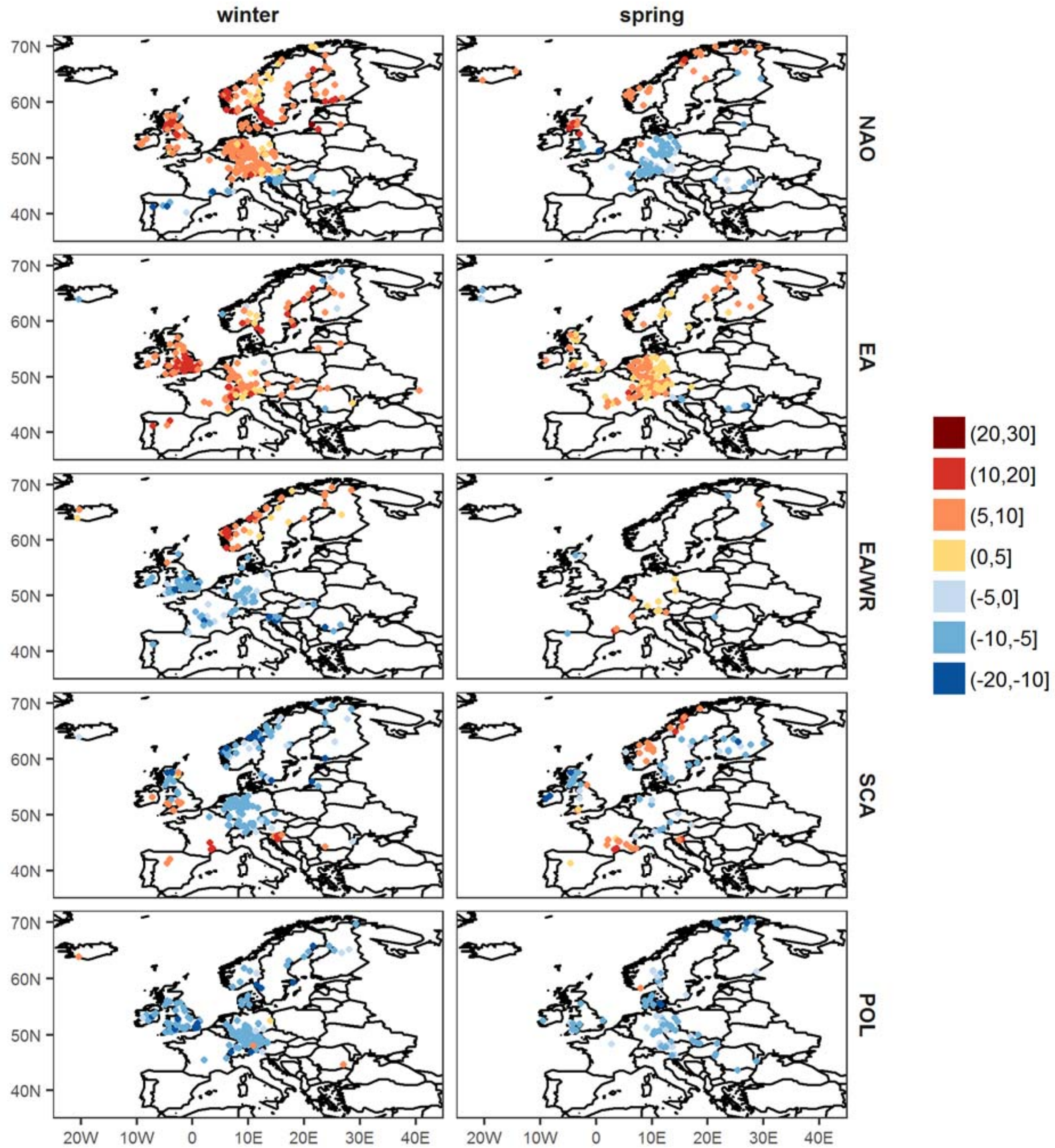


Figure 4: Same as Fig. 3 but for the summer and autumn season.


winter
spring
NAO
EA
EA/WR
SCA
POL
70N
60N
50N
40N
20W
10W
0
10E
20E
30E
40E
(20,30]
(10,20]
(5,10]
(0,5]
(-5,0]
(-10,-5]
(-20,-10]

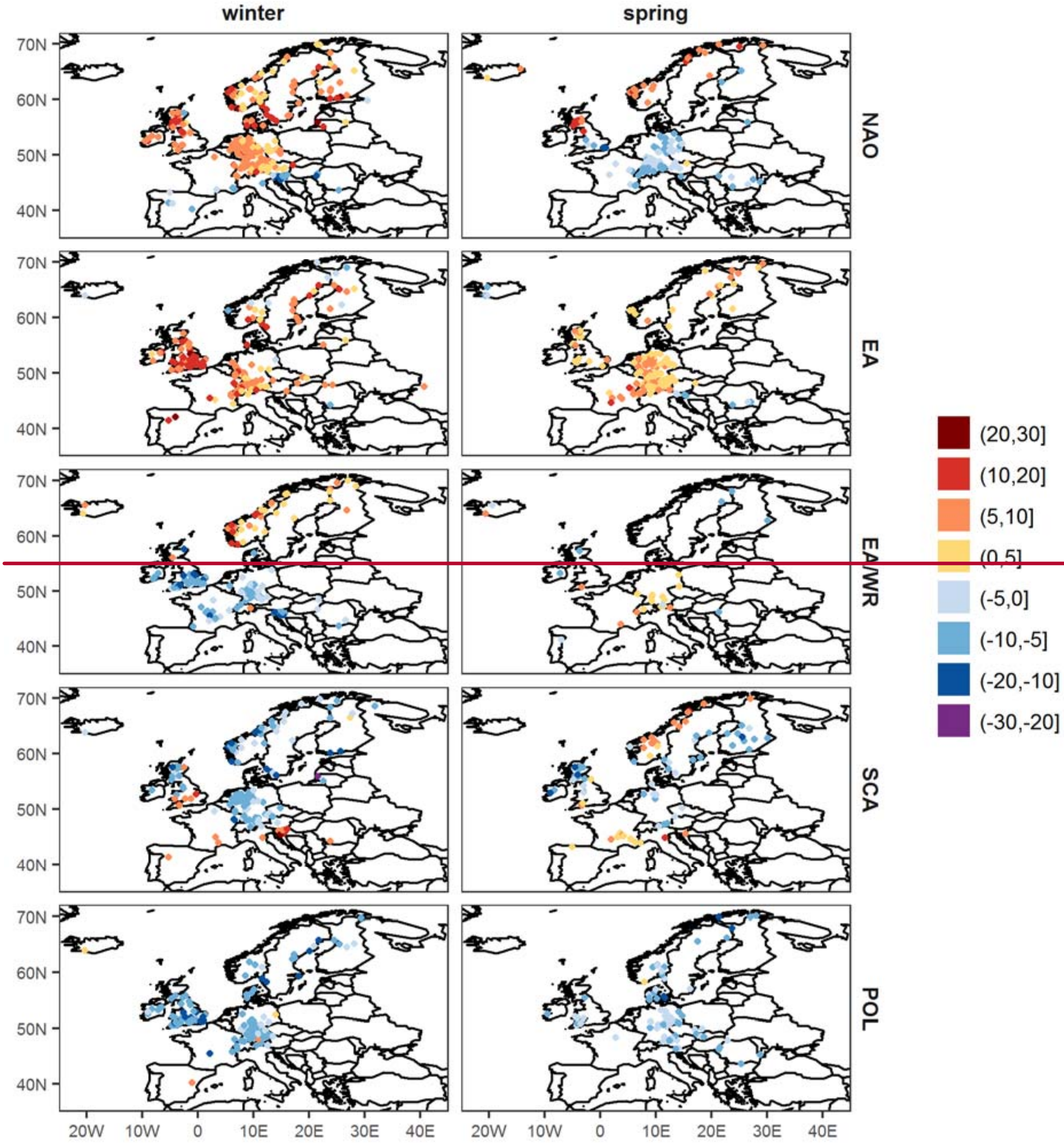
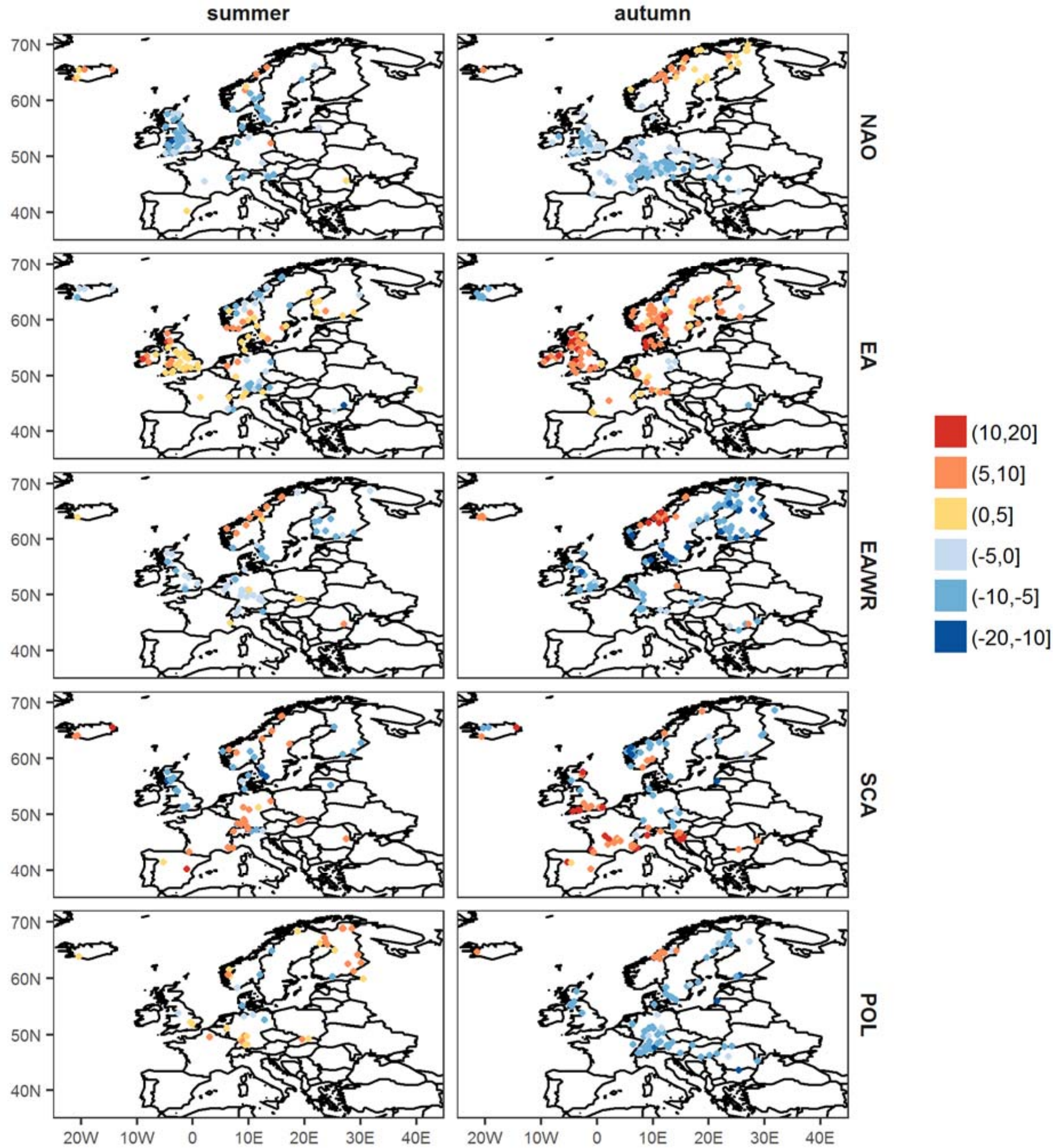
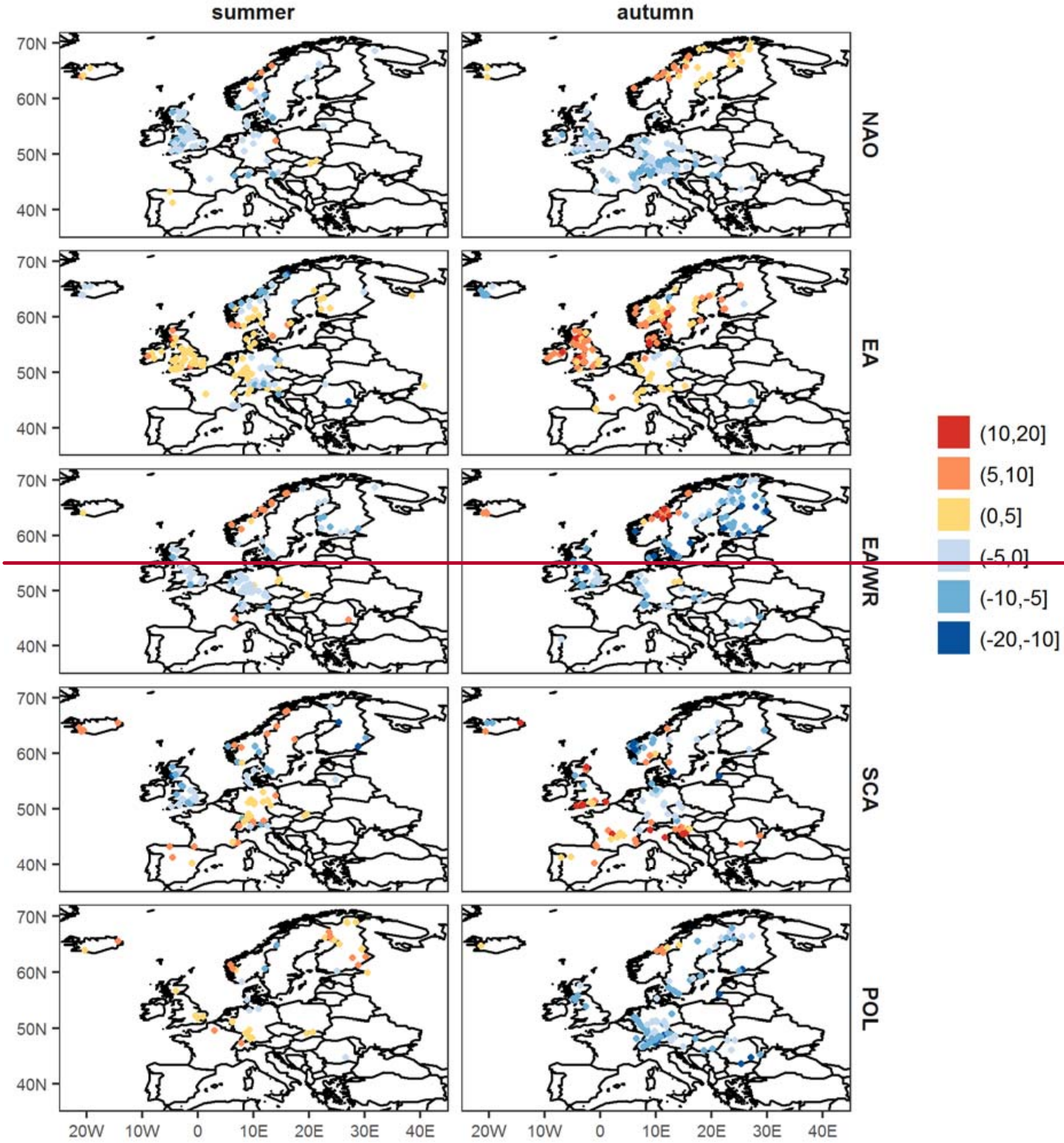
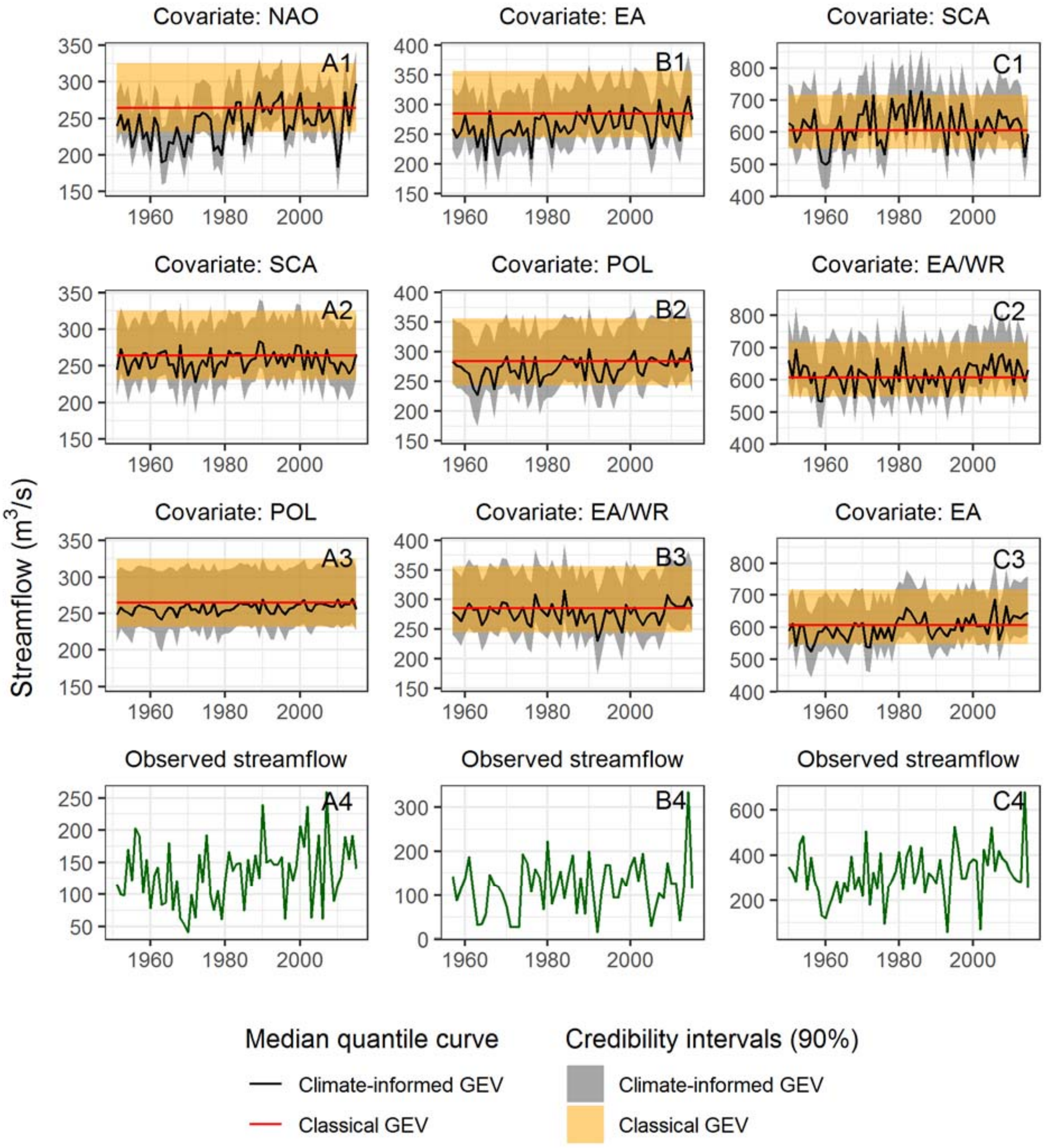


Figure 5: Percent relative difference of the streamflow for an exceedance probability of 0.~~01~~ 02 between a (hypothetical) year with a climate index value equal to the 95th quantile and a year with an index value equal to the median index. Results are shown for winter and spring and seasonal mean covariates.


summer
autumn
NAO
EA
EA/WR
SCA
POL
70N
60N
50N
40N
20W
10W
0
10E
20E
30E
40E
(10,20]
(5,10]
(0,5]
(-5,0]
(-10,-5]
(-20,-10]



685 **Figure 6: Same as Fig. 5 but for the summer and autumn season.**


Covariate: NAO
A1
Covariate: EA
B1
Covariate: SCA
C1
Covariate: SCA
A2
Covariate: POL
B2
Covariate: EA/WR
C2
Covariate: POL
A3
Covariate: EA/WR
B3
Covariate: EA
C3
Observed streamflow
A4
Observed streamflow
B4
Observed streamflow
C4
Streamflow (m3/s)
1960
1980
2000
Median quantile curve
Climate-informed GEV
Classical GEV
Credibility intervals (90%)
Climate-informed GEV
Classical GEV

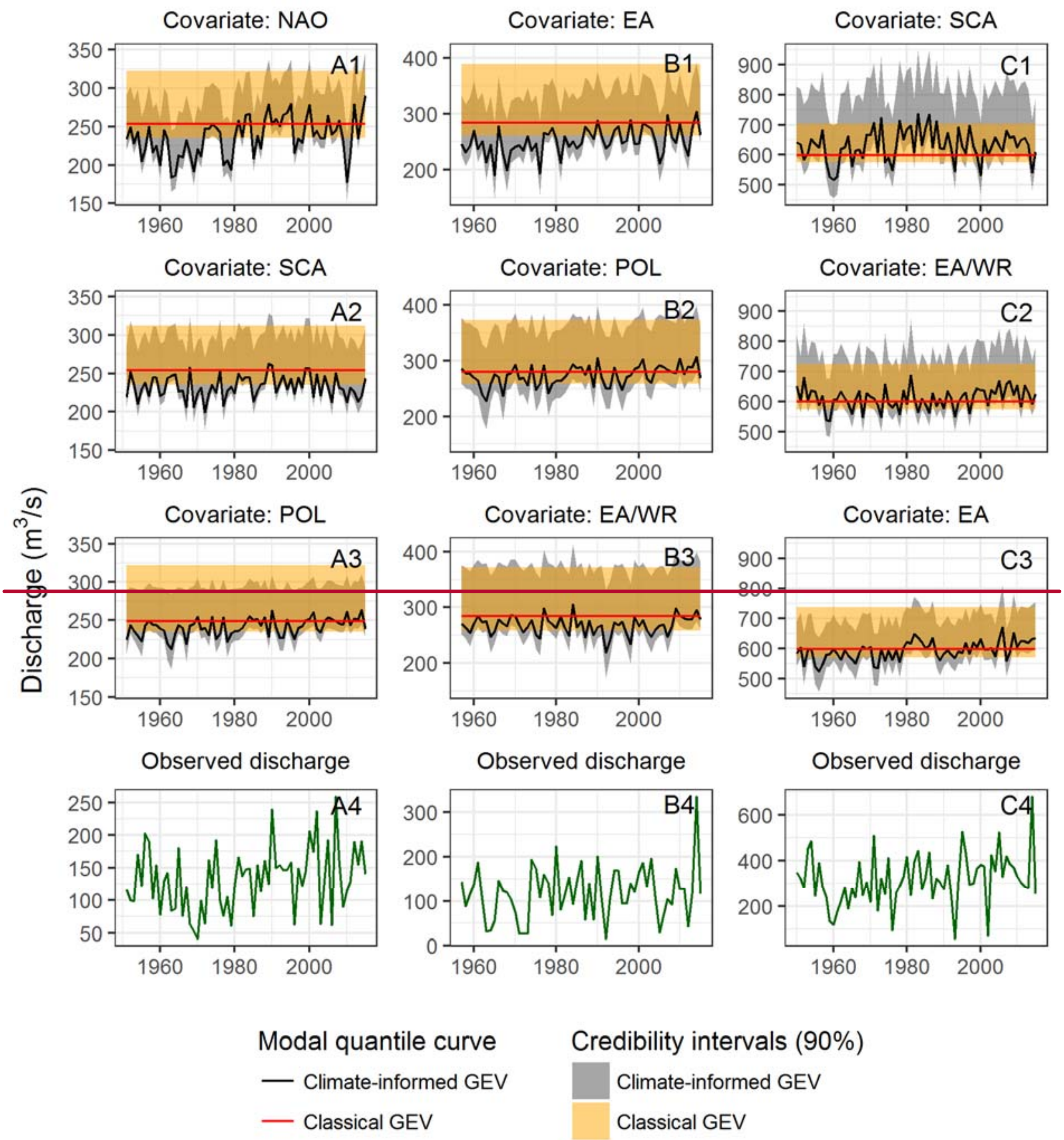


Figure 7: Annual maximum discharge time series (lower panel: 4) and climate-informed quantiles (upper panels: 1-3) with credibility intervals for an exceedance probability 0.02 and ~~P(Y>y=0.01)~~ for three selected gauges (Table ~~4~~5, ~~5~~6). Climate-informed quantiles are compared with those of the classical GEV. The three best climate-informed models based on the DIC are shown for each site, with increasing DIC from top to bottom.

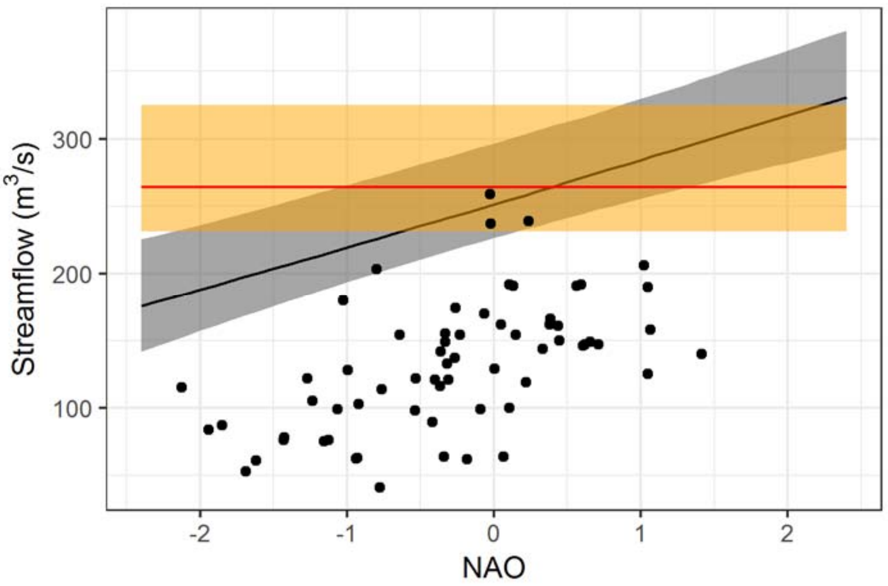


Figure 8: Comparison of climate-informed and classical streamflow quantiles for station A (see Fig. 7 and Tables 5, 6 for more details), an exceedance probability 0.02 and NAO as covariate of the climate-informed model. The legend is the same as in Fig. 7. Observed streamflow is indicated with black dots.

Tables

Table 1: Percentage of stations with climate-informed fits preferred to the classical GEV model. Indicated is the result of the pairwise comparison of each covariate with the classical model and the percentage of preferred fits for each covariate when all models are compared (in brackets). Results are shown per season and for mean seasonal covariates.

Index	Winter	Spring	Summer	Autumn
NAO	~~44~~ 46 (~~27~~28)	~~21~~ 19 (~~12~~10)	~~14~~ 13 (~~9~~8)	~~32~~ 31 (24)
EA	~~30~~ 29 (~~21~~20)	~~30~~ 32 (~~25~~26)	~~22~~ 18 (~~13~~11)	~~20~~ 19 (~~14~~13)
EA/WR	~~23~~ 24 (~~11~~10)	~~4~~ 3 (~~2~~1)	~~14~~ 10 (~~9~~7)	~~21~~ 20 (~~14~~13)
SCA	~~25~~ 26 (11)	~~13~~ 14 (~~11~~12)	~~15~~ 10 (~~8~~6)	~~16~~ 15 (~~12~~11)
POL	~~23~~ 24 (~~8~~9)	14 (~~10~~11)	7 (6)	~~15~~ 13 (5)
All indices	(77)	(~~61~~60)	(~~45~~38)	(~~68~~66)

Table 2: Same as Table 1 but for monthly covariates at the same month as the seasonal streamflow extremes.

Index	Winter	Spring	Summer	Autumn
NAO	~~21~~ 33 (~~16~~28)	~~17~~ 16 (12)	~~15~~ 13 (~~11~~9)	~~16~~ 13 (~~10~~8)
EA	~~22~~ 27 (~~17~~18)	15 (12)	~~16~~ 15 (~~11~~10)	22 (18)
EA/WR	~~9~~ 26 (~~6~~14)	14 (~~10~~9)	~~16~~ 13 (~~13~~11)	~~20~~ 17 (~~13~~12)
SCA	~~16~~ 15 (~~10~~8)	17 (12)	~~13~~ 12 (~~10~~9)	~~14~~ 13 (9)
POL	~~13~~ 7 (~~9~~4)	16 (~~12~~13)	~~8~~ 6 (~~6~~4)	~~11~~ 10 (7)
All indices	(~~42~~72)	(~~59~~58)	(~~50~~43)	(~~56~~54)

Table 3: Seasonal quantiles of the five climate indices: median and in the parenthesis the 95th quantile are provided.

Index	Winter	Spring	Summer	Autumn
NAO	-0.26 (1.04)	-0.15 (0.84)	0.02 (1.14)	0.17 (0.96)
EA	-0.37 (1.07)	-0.13 (0.70)	-0.07 (0.80)	-0.19 (0.69)
EA/WR	-0.19 (0.78)	-0.04 (0.78)	0.15 (1.23)	0.11 (1.29)
SCA	0.21 (1.25)	0.05 (0.90)	0.09 (1.33)	0.21 (1.44)
POL	0.11 (1.44)	0.07 (0.90)	-0.11 (0.94)	-0.02 (0.91)

Table 4: Summary statistics of median posterior shape parameter of all stations examined. Statistics are taken over all models for one season. In the parenthesis the maximum deviation of all the models fitted (classical and climate-informed) is provided.

	Winter	**Spring**	**Summer**	**Autumn**
Min	-0.420 (0.072)	-0.365 (0.057)	-0.303 (0.058)	-0.303 (0.074)
Q5	-0.137 (0.013)	-0.104 (0.007)	-0.055 (0.009)	-0.057 (-0.010)
Q25	-0.008 (0.007)	0.002 (0.010)	0.062 (-0.005)	0.045 (-0.06)
Median	0.062 (0.006)	0.066 (0.006)	0.165 (0.005)	0.127 (-0.008)
Mean	0.063 (0.003)	0.066 (0.005)	0.165 (0.002)	0.125 (-0.008)
Q75	0.133 (-0.006)	0.133 (-0.005)	0.271 (-0.005)	0.200 (-0.010)
Q95	0.262 (-0.014)	0.226 (-0.009)	0.385 (-0.006)	0.316 (0.009)
Max	0.461 (-0.053)	0.381 (-0.019)	0.537 (-0.020)	0.527 (-0.031)

Table ~~4~~5: General information about selected sites shown in Fig. 7. Ref. code is the number of the subplot of Fig. 7.

Ref. code	**Station name**	**Country**	**GRDC No**	**Latitude**	**Longitude**	**Catchment size (km^2)**
A	ASBRO 3	Sweden	6233100	57.240	12.310	2160.2
B	TESTON	United Kingdom	6607851	51.251	0.447	1256.1
C	BULKEN	Norway	6731200	60.630	6.280	1102.0

Table ~~5~~6: Climate-informed results as shown in Fig. 7. Ref. code is the number of the subplot of Fig. 7. Mean seasonal covariates for the same season as streamflow extremes are examined. dDIC is the difference from the DIC value of the classical distribution. $Y_{0.0\cancel{1}2}$ is the percent relative difference of streamflow with exceedance probability ~~P(Y>y=~~0.~~01~~02~~)~~ for the 95th quantile of the covariate ($y_{0.02\cancel{1},h}$) and of streamflow with exceedance probability 0.02 ~~P(Y>y=0.01)~~ for the median ($y_{0.02\cancel{1},m}$). The sign of the slope is reported if it is significantly different than zero at the 10% significance level.

Ref. code	**Season**	**Covariate**	**dDIC**	**Slope**	**$Y_{0.0\cancel{1}2}$ [%]**	**$y_{0.0\cancel{1}2,m}$ [m^3/s]**	**$y_{0.02\cancel{1},h}$ [m^3/s]**
A	Winter	NAO	-22.~~6~~9	Positive	~~19~~17.7	~~239~~242.~~5~~7	~~281~~285.~~3~~6
		SCA	-~~5.0~~4.3	Negative	-~~9.2~~7.9	~~236~~255.~~2~~3	~~211~~235.~~9~~2
		POL	-0.5	Nonsignificant	-	-	-
B	Winter	EA	-9.~~6~~4	Positive	~~16.1~~14.7	260.54~~7.9~~	299.0~~1.7~~

		POL	-~~5.8~~4.6	Negative	-12.0~~3.5~~	27~~6~~3.~~5~~8	241.0~~29.3~~
		EA/WR	-4.~~0~~2	Negative	-~~7.8~~8.5	~~269.2~~276.8	~~246.6~~253.3
C	Autumn	SCA	-~~14.3~~15.5	Negative	-15.48	~~643.2~~613.7	~~547.0~~516.8
		EA/WR	-5.9~~3~~	Negative	-11.2~~0.9~~	615.4~~02.8~~	546.0~~39.5~~
		EA	-3.6~~5~~	Positive	7.7~~1~~	601.6~~593.9~~	648.0~~636.8~~

Supplementary Material

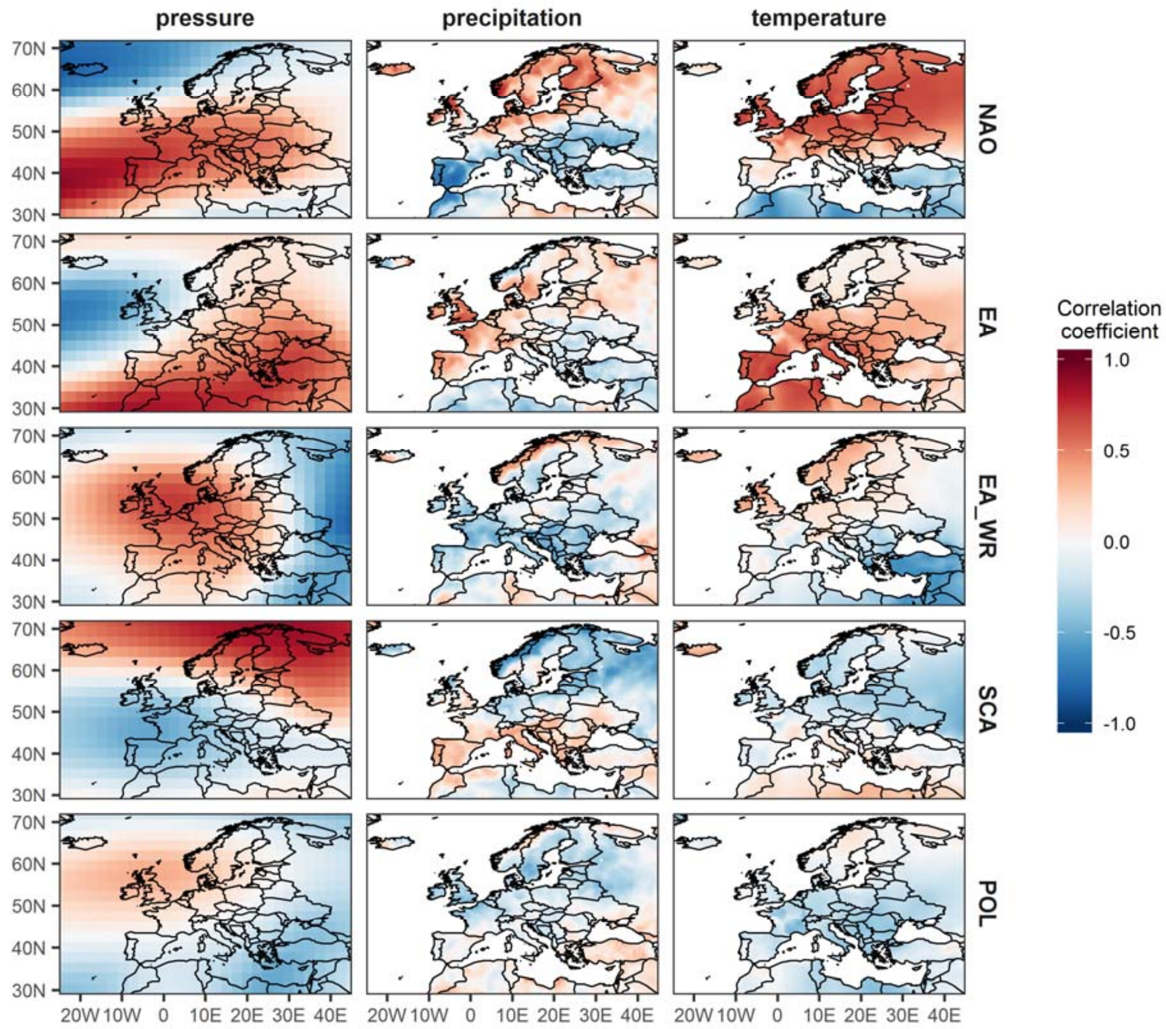


Figure S1: Linear correlations between the five circulation indices examined and mean seasonal gridded pressure (left), precipitation (middle) and temperature (right). All variables are averaged over the winter season for the period 1952-2015 (year attributed to respective January). Gridded pressure data were retrieved from the NCEP/NCAR Reanalysis dataset (Kalnay et al., 1996) and provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/. Temperature and precipitation data were extracted from the gridded data set CRU TS3.24 from the climatic research unit (CRU, https://crudata.uea.ac.uk/cru/data/hrg/) of the University of East Anglia (Harris et al., 2014). Details about the circulation indices are given in chapter 2.1. The figure was modified after Steirou et al. (2017).

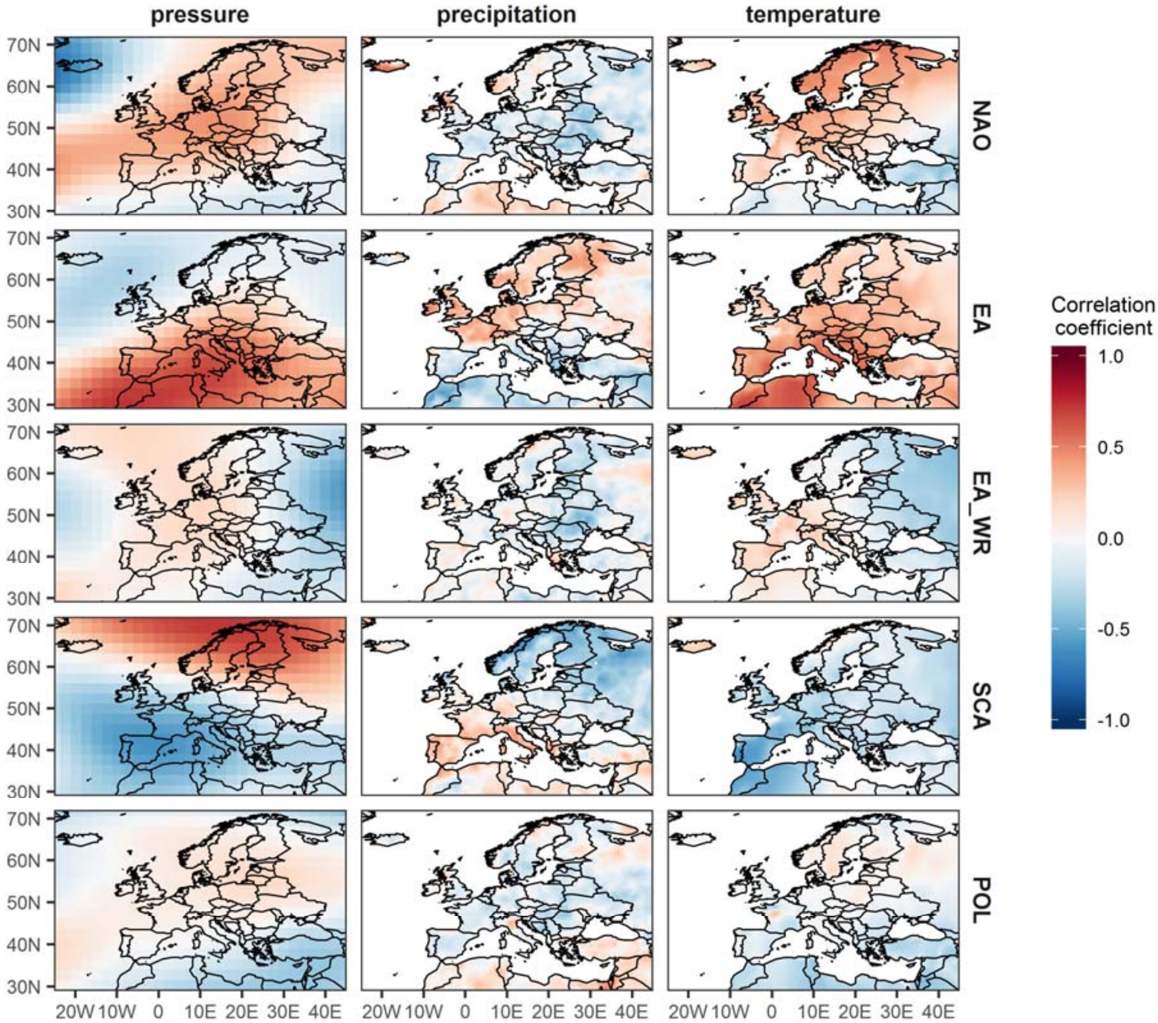


Figure S2: Same as Fig. 1 but for the spring season and the period 1951-2015.

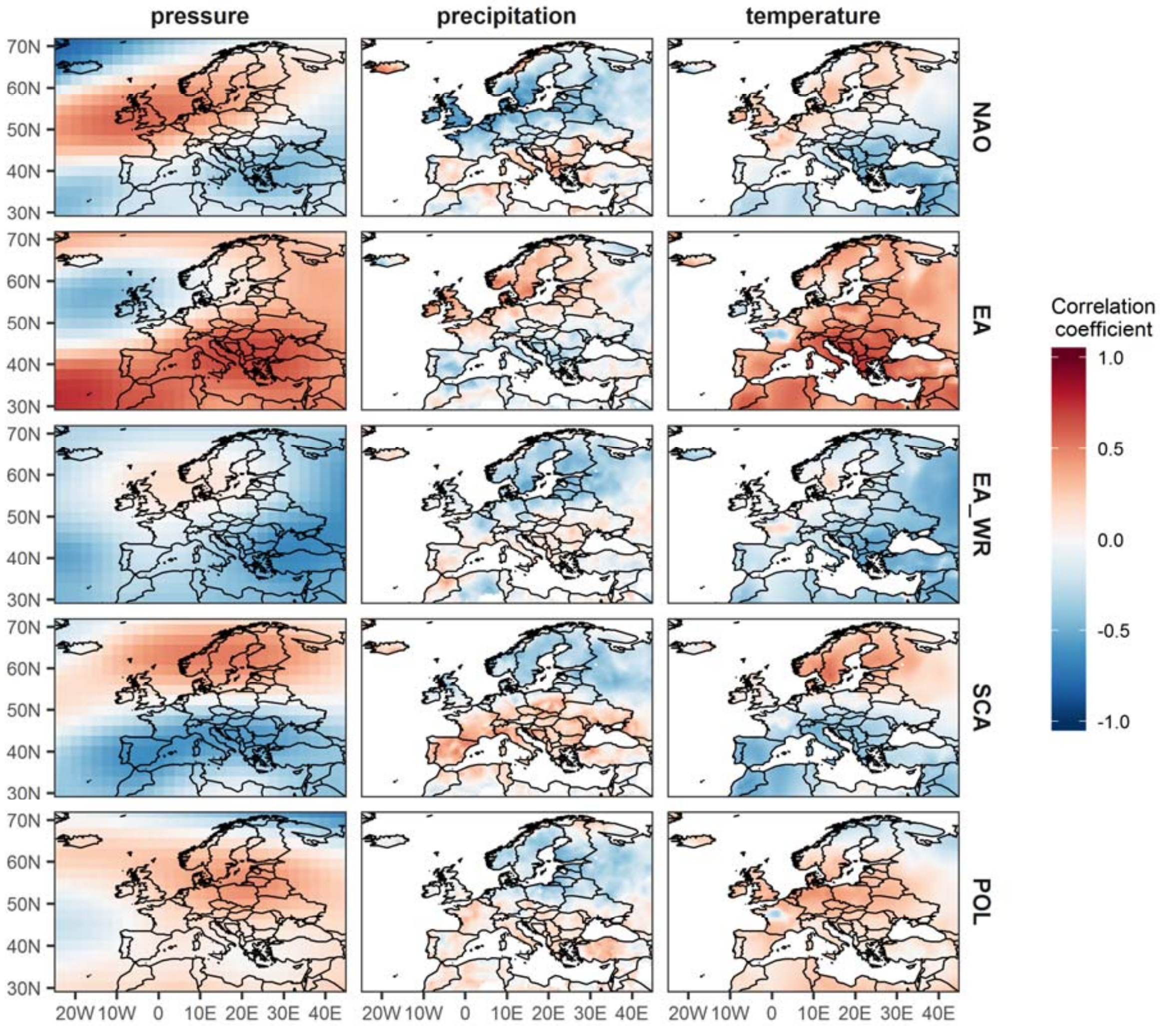


Figure S3: Same as Fig. 1 but for the summer season and the period 1951-2015.

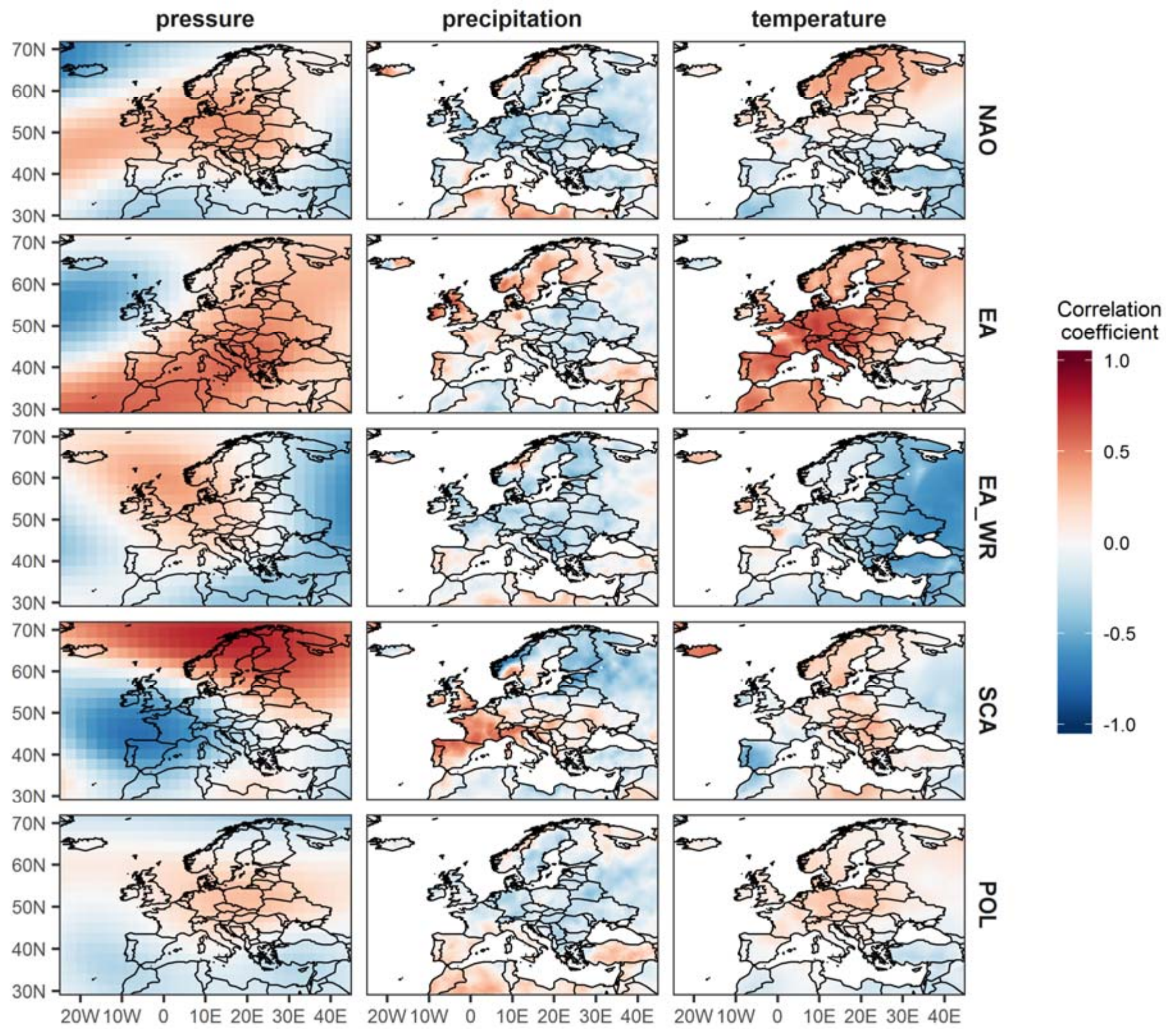


Figure S4: Same as Fig. 1 but for the autumn season and the period 1951-2015.

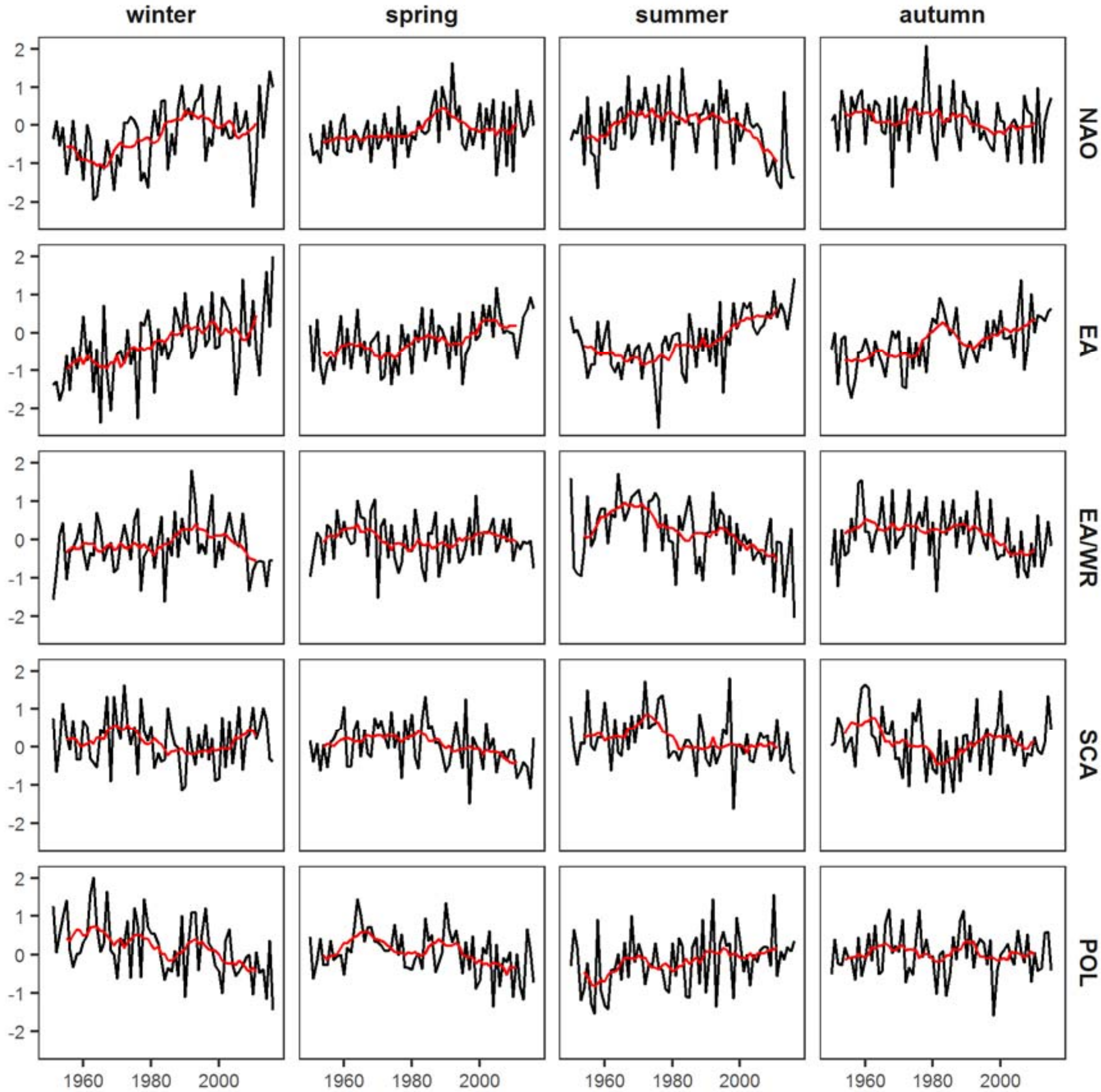


Figure S5: Mean seasonal indices examined (black curves). The red lines indicate a moving average with a window of 10 years.

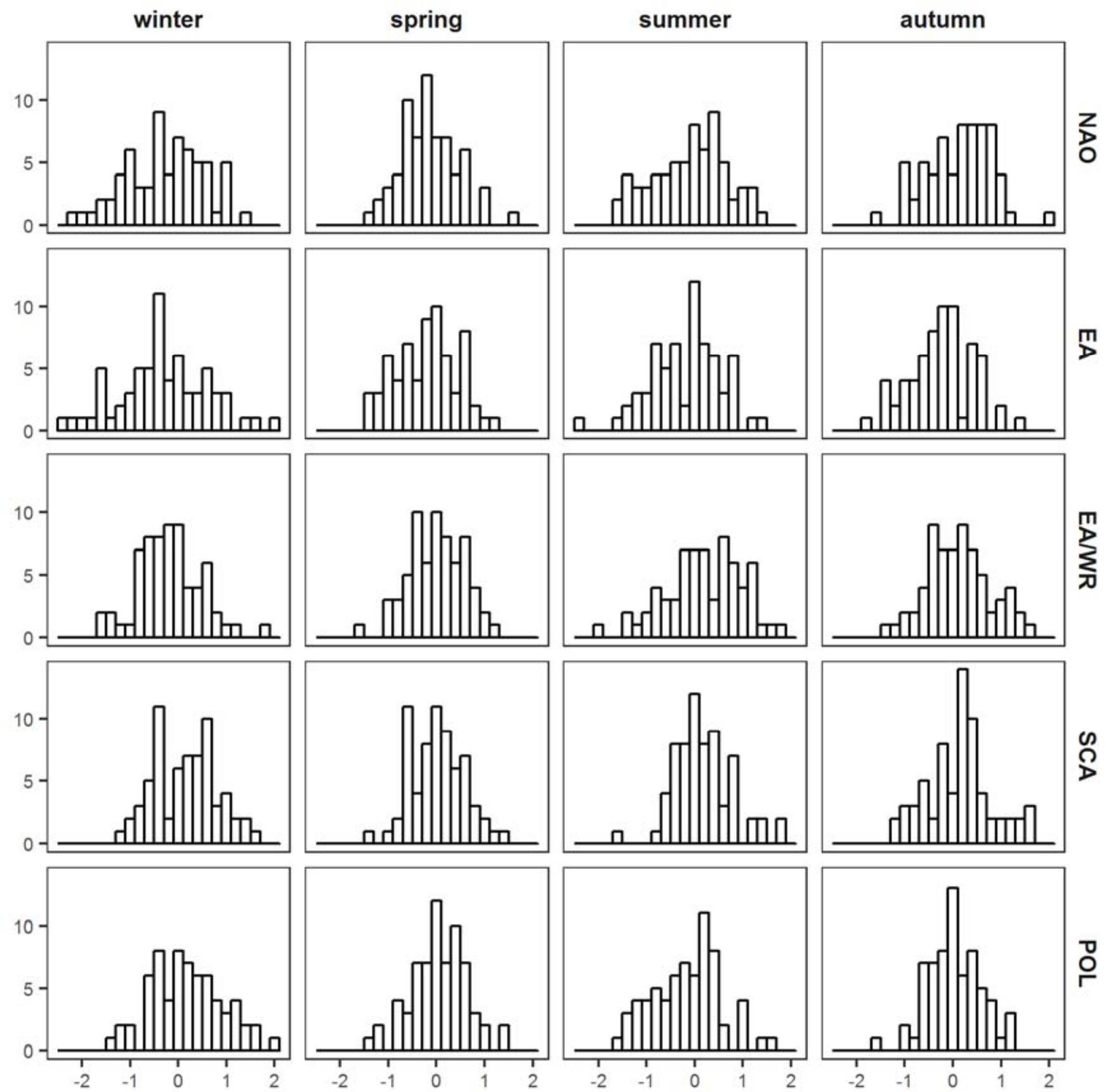
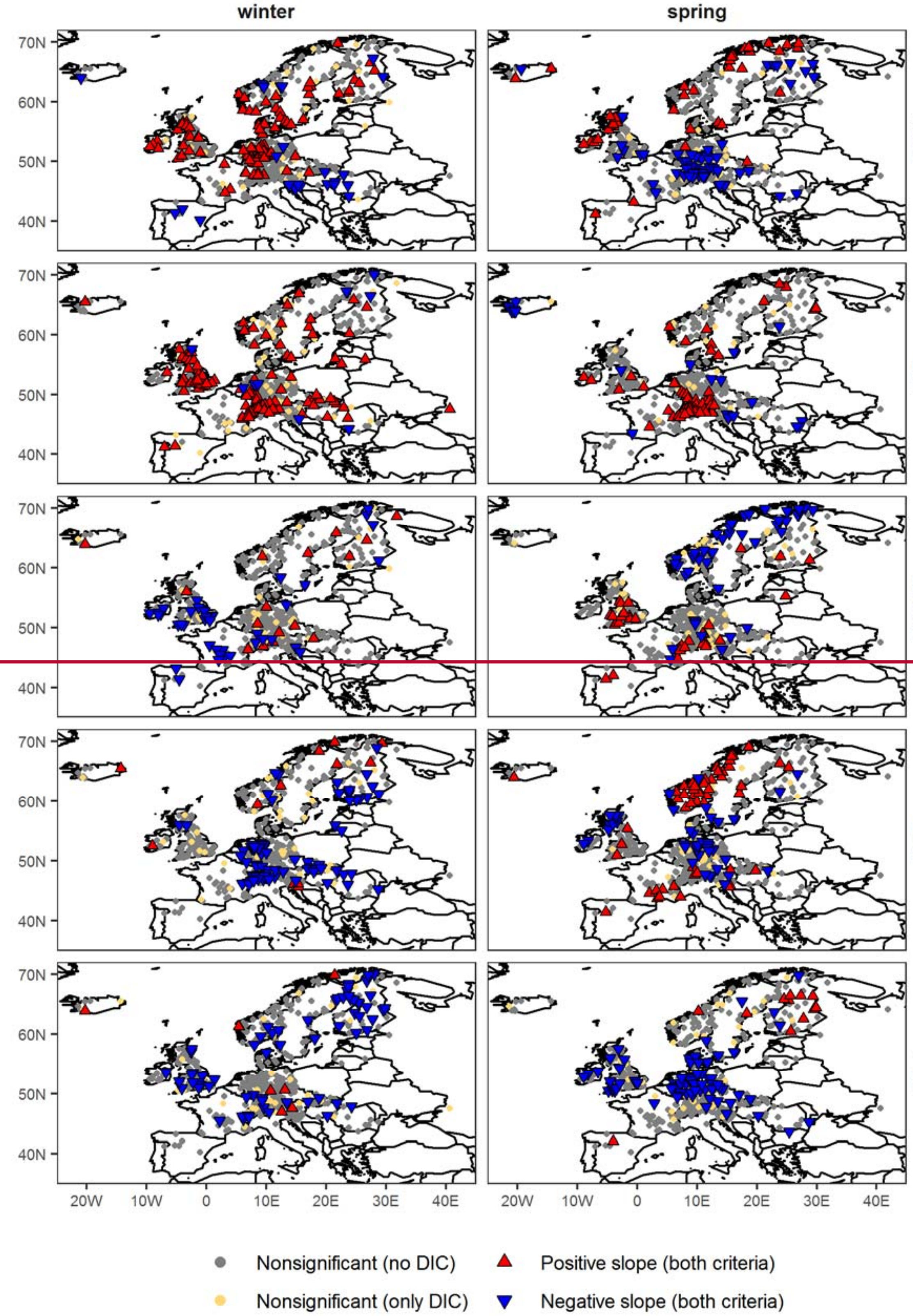


Figure S6: Histograms of mean seasonal indices examined.


winter
spring
NAO
EA
EA/WR
SCA
POL
70N
60N
50N
40N
20W
10W
0
10E
20E
30E
40E
Nonsignificant (no DIC)
Nonsignificant (only DIC)
Positive slope (both criteria)
Negative slope (both criteria)

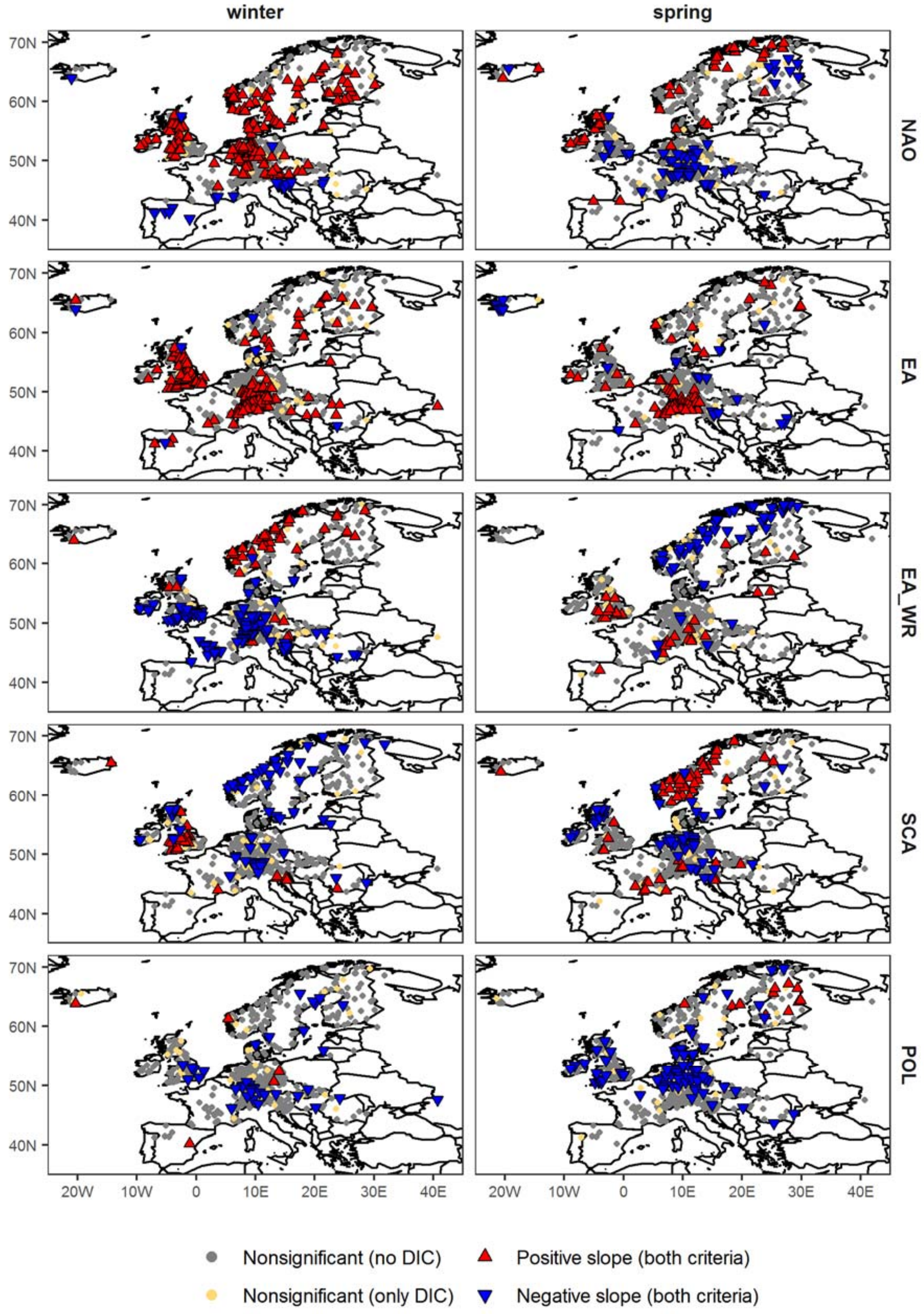
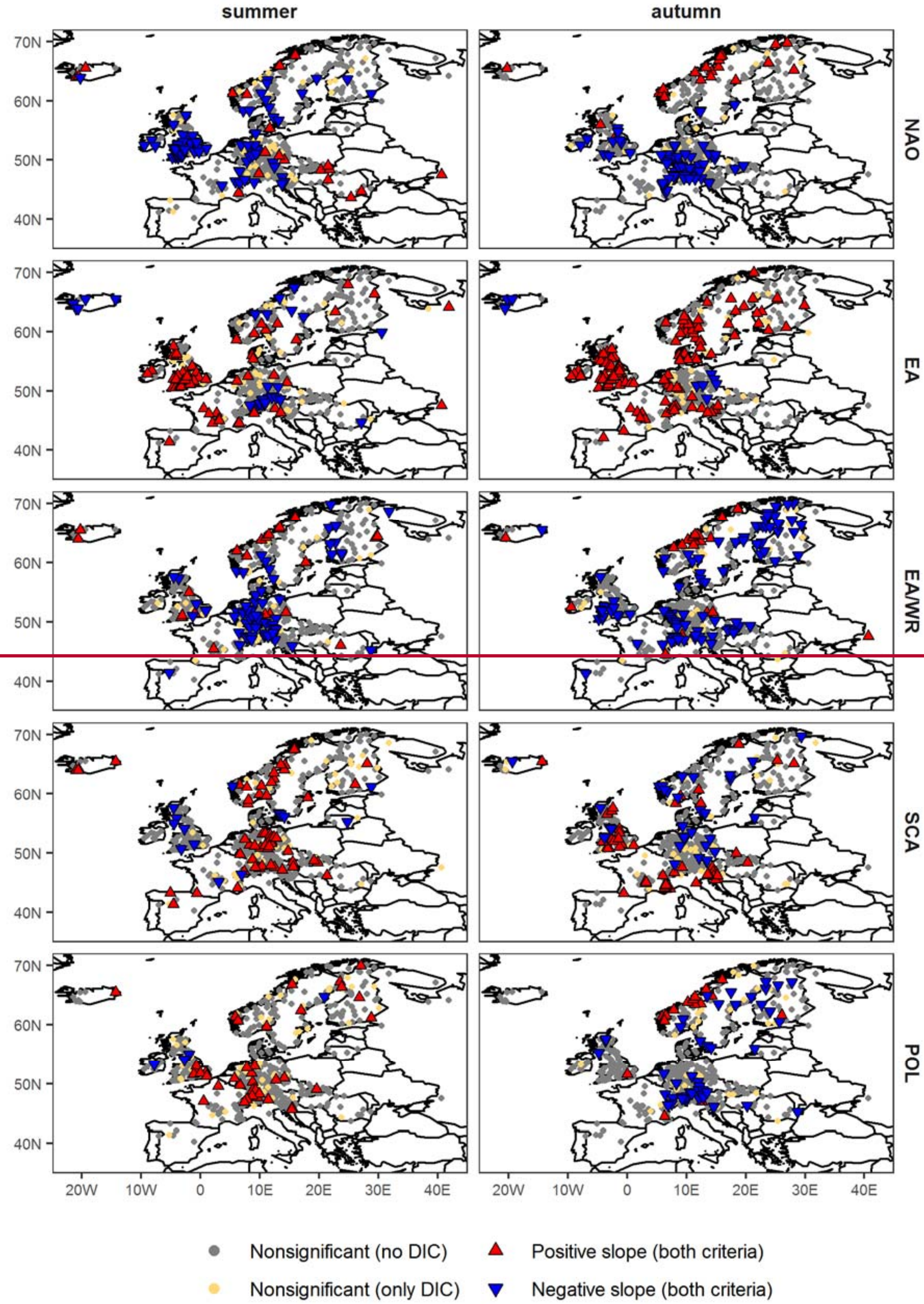


Figure ~~S5~~S7: Same as Fig. 1 but for the monthly indices at the month of the season maximum streamflow.


summer
autumn
NAO
EA
EA/WR
SCA
POL
70N
60N
50N
40N
20W
10W
0
10E
20E
30E
40E
Nonsignificant (no DIC)
Nonsignificant (only DIC)
Positive slope (both criteria)
Negative slope (both criteria)

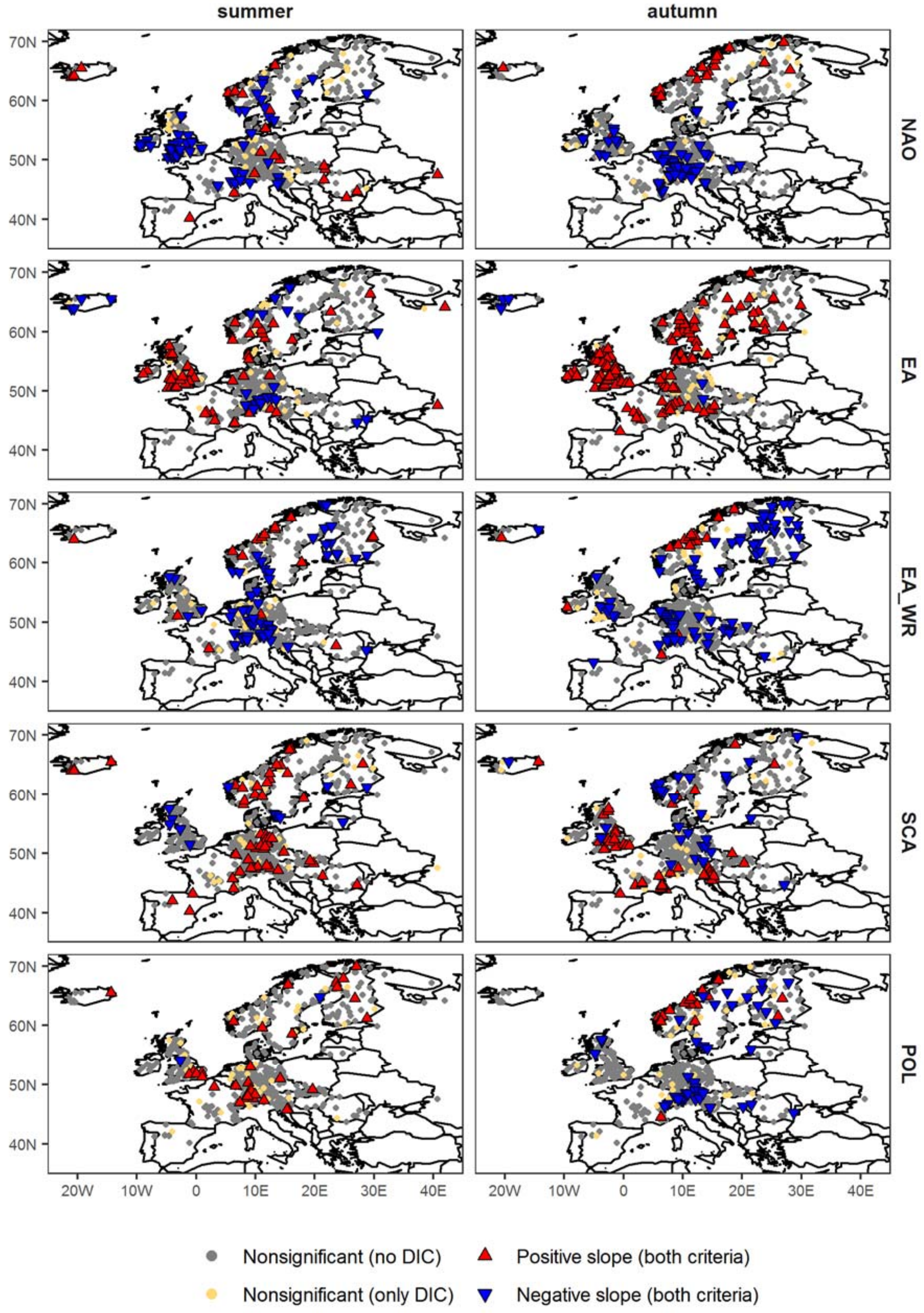
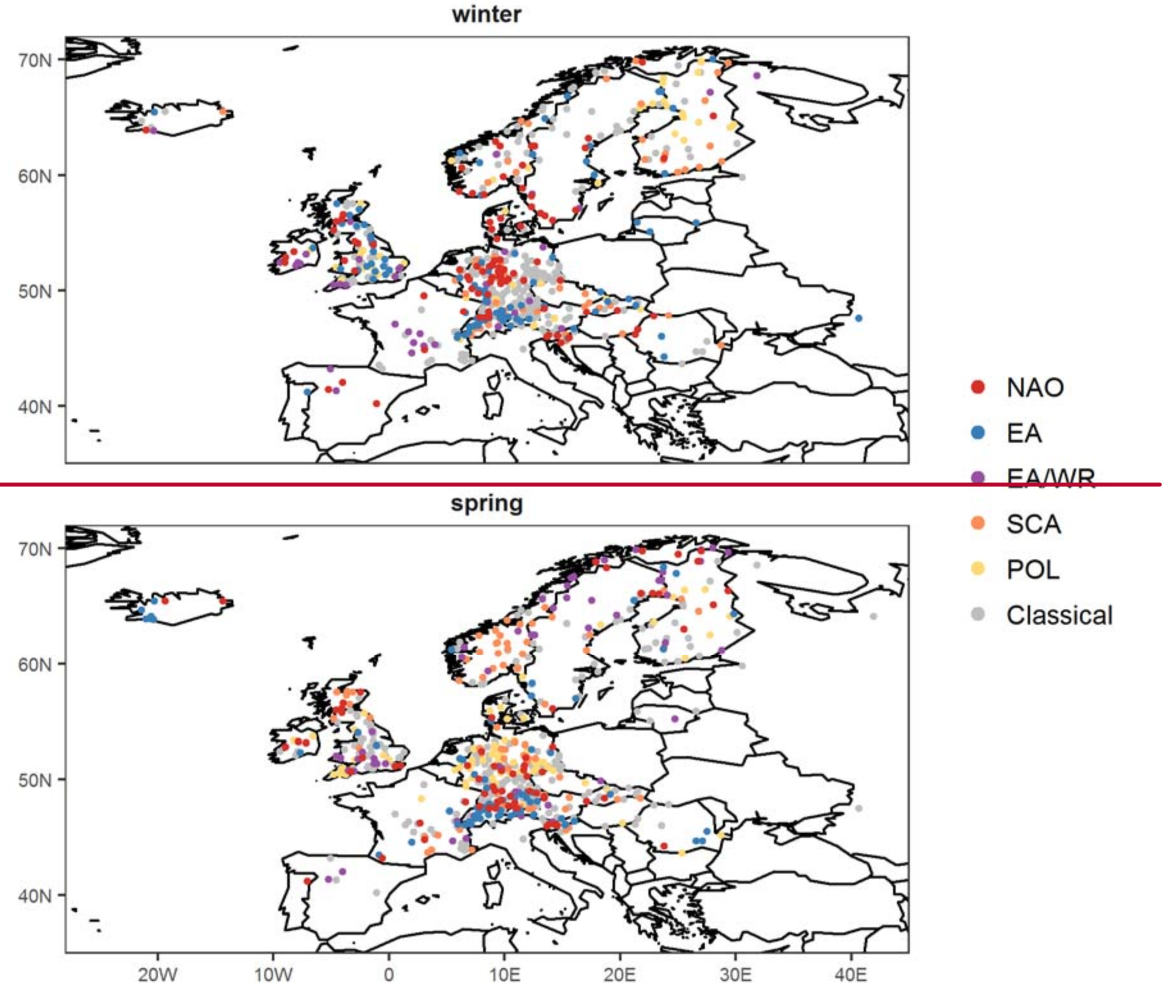


Figure ~~S6~~S8: Same as Fig. ~~S5~~ S7 but for the summer and autumn season.



winter
spring
70N
60N
50N
40N
20W
10W
0
10E
20E
30E
40E
NAO
EA
EA/WR
SCA
POL
Classical

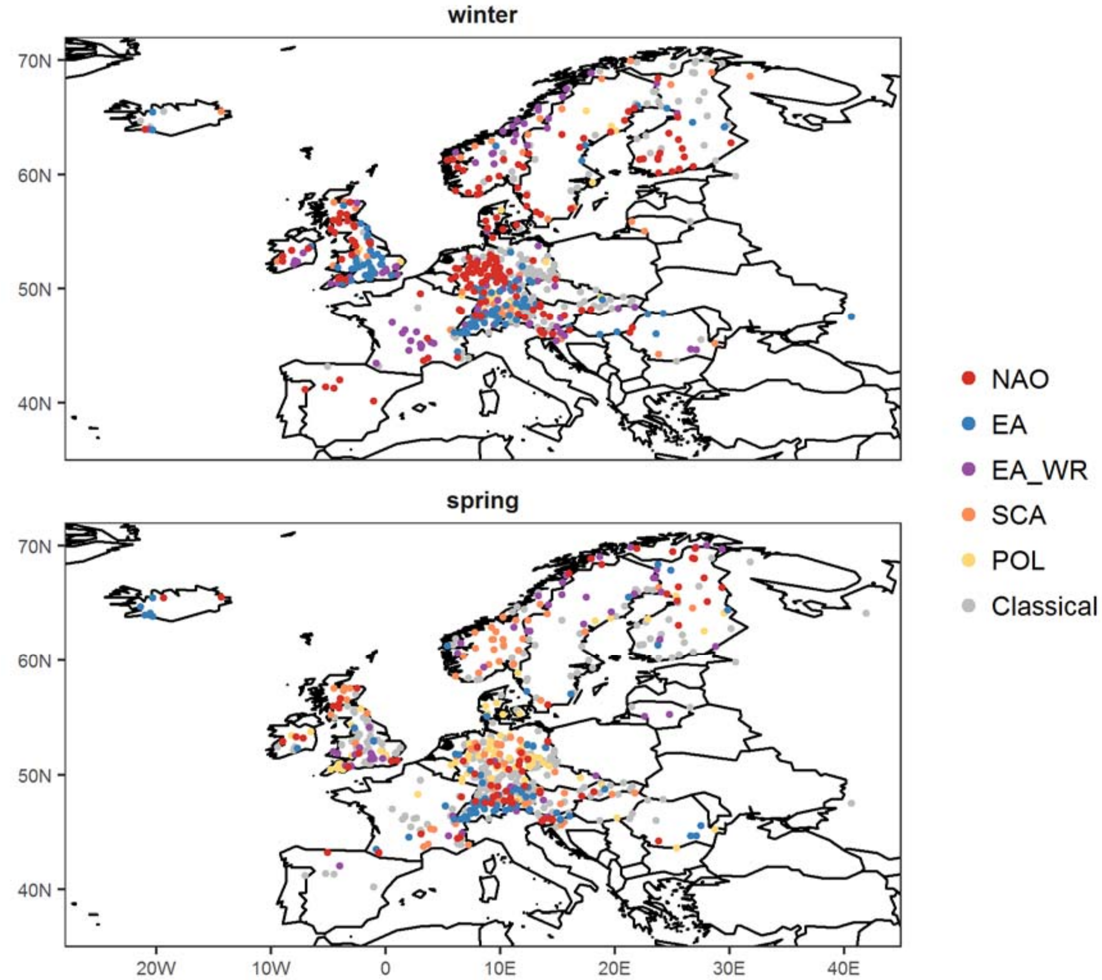
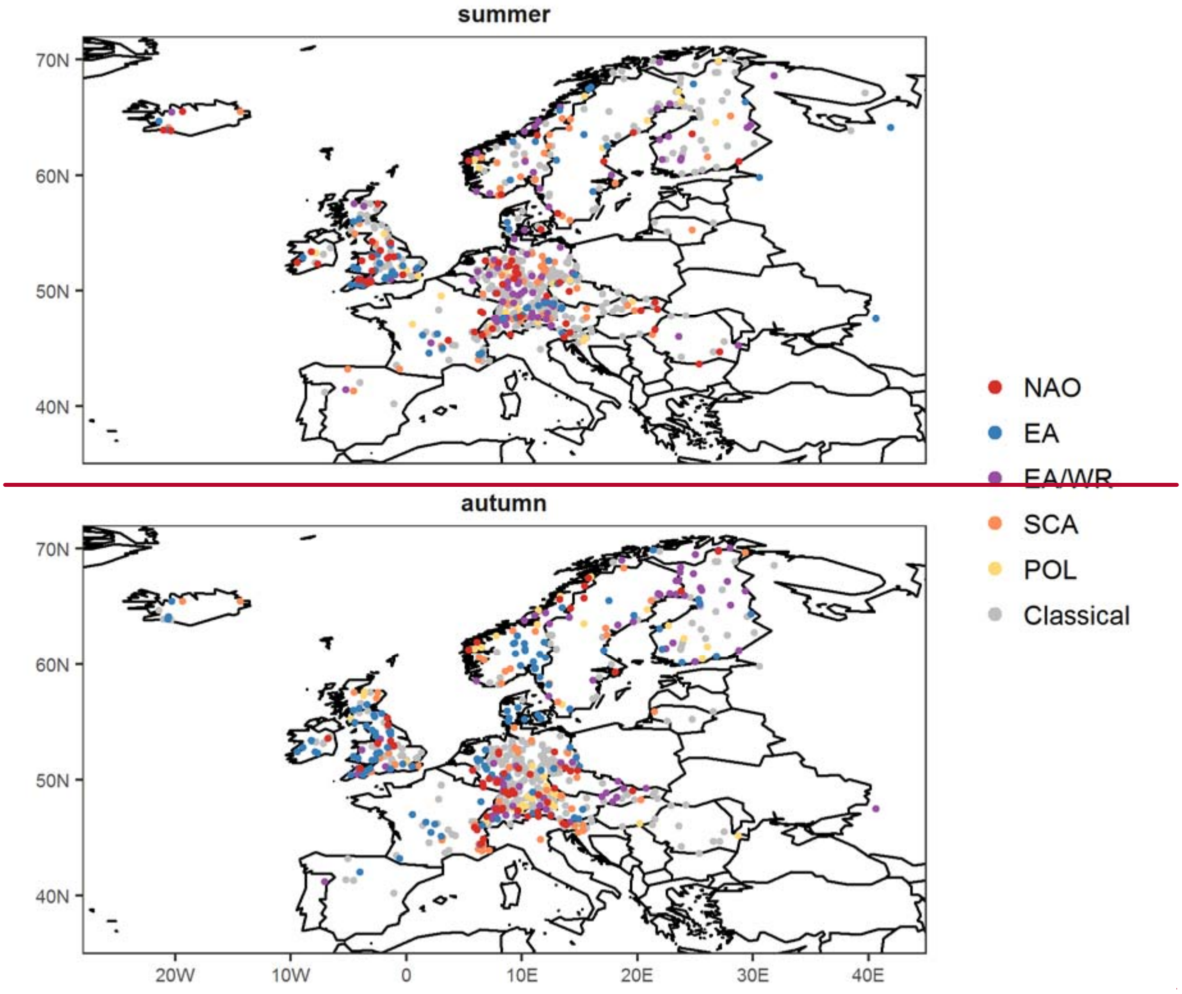


Figure ~~S7~~S9: Same as Fig. 3 but for the monthly indices at the month of the season maximum streamflow.


summer
autumn
70N
60N
50N
40N
20W
10W
0
10E
20E
30E
40E
NAO
EA
EA/WR
SCA
POL
Classical

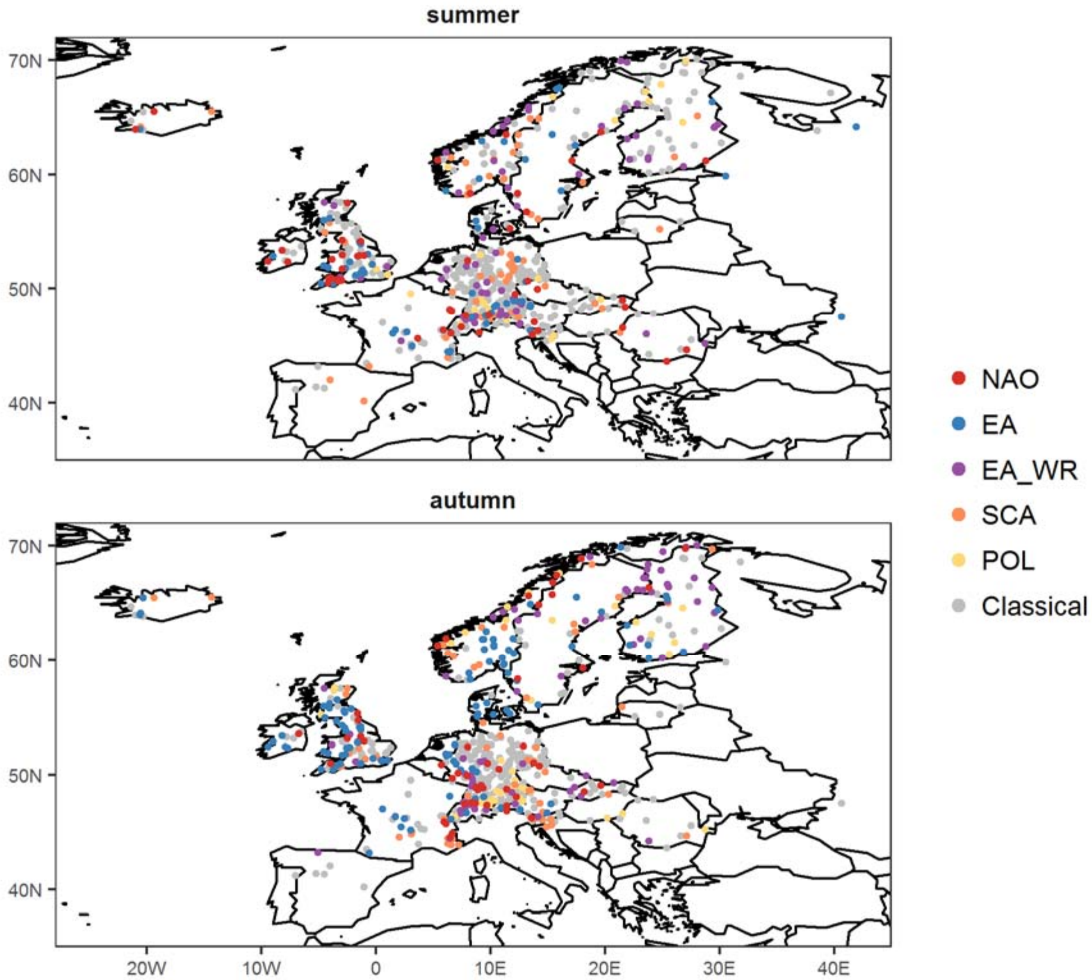


Figure ~~S8~~S10: Same as Fig. 4 but for the monthly indices at the month of the season maximum streamflow.

Table S1: Pearson correlation coefficient between the indices examined. Statistically significant results at the 5% level are highlighted.

Indices	**Winter**	**Spring**	**Summer**	**Autumn**
NAO - EA	0.126	0.026	-0.336*	-0.207
NAO – EA/WR	0.085	0.176	0.252*	0.155
NAO - SCA	-0.314*	0.065	0.253*	0.022

NAO - POL	-0.235	-0.029	-0.069	0.144
EA – EA/WR	-0.028	-0.003	-0.446*	-0.095
EA – SCA	0.023	-0.166	-0.343*	-0.040

References for Supplementary Material

Harris, I., Jones, P. D., Osborn, T. J. and Lister, D. H.: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset, Int. J. Climatol., 34(3), 623–642, doi:10.1002/joc.3711, 2014.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R. and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bull. Am. Meteorol. Soc., 77(3), 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Steirou, E., Gerlitz, L., Apel, H. and Merz, B.: Links between large-scale circulation patterns and streamflow in Central Europe: A review, J. Hydrol., 549, doi:10.1016/j.jhydrol.2017.04.003, 2017.

Reply to Referee #1 Alberto Viglione:

This paper presents a European data-based analysis of the correlation between a number of atmospheric indices and flood exceedance probabilities at the sub-annual timescale. The novelty of the paper is related to the extent of the study region, i.e., all of Europe. The outcomes are interesting because of the coherent spatial patterns of the identified correlations in climatically different parts of Europe. The paper is well written, properly concise and clear. I believe it can become a very valuable entry for HESS. However, as always, some improvements are possible. My main comments/criticisms/suggestions are the following:

Response: We would like to thank Alberto Viglione for his comments. The points he raised are constructive and addressing them definitely improved our manuscript.

General comment 1

- Title: I do not think that it is possible to easily answer this question, it never is when dealing with extreme value statistics in the real world. Actually, while I consider very interesting the analysis of the correlation between the atmospheric indices and the parameters of flood exceedance probabilities, I am less convinced about the accuracy of flood frequency estimation provided here. The reason is that, in engineering hydrology, I think nobody would fit locally a GEV distribution using a likelihood-based method with no information on its shape parameter. Regional analysis is normally used to improve quantile estimation for high return periods (say 100-years) which is not performed here. I would agree that the paper provides an indication that there is potential for improving flood frequency estimation by including atmospheric dynamics in our models, but I guess there is much more to do to actually improve the existing regional models in use. Maybe this is what the Authors meant but, to me, the title is a bit misleading. In my opinion, a title that focuses on the identified correlations between atmospheric indices and local floods would be better.

Response: We found the comment very important and we adopted the reviewer's recommendation concerning the title of our manuscript. The title was changed to "Climate influences on flood probabilities across Europe" that focuses on the spatial aspect of our analysis and on streamflow-climate interactions. In addition, we reduced the extrapolation towards high return periods and we calculated streamflow quantiles for a probability of exceedance 0.02 (50-year return period). The common time period of streamflow data and circulation indices is between 50-70 years, so the extrapolation and possible uncertainty from the absence of a regionalization framework is considerably reduced. Furthermore, an informative prior distribution was used for the shape parameter, in order to constrain the shape parameter from adopting unreasonably high or low values and to improve the GEV fits (see also our reply to general comment 2).
Finally, a comment was added in the discussion about the possibility of improving quantile estimation by using a regionalization framework (Lines 384-389 of the revised manuscript): "In this study, a local, site-specific flood frequency model was developed. This model allowed to identify spatial coherence in relations between streamflow extremes and large-scale atmospheric patterns. However, a shortcoming of this methodology is the high uncertainty of streamflow estimates for high probabilities of exceedance (corresponding for example to the 100- or 200-year flood). Instead of a local framework, a regional framework can be alternatively implemented. The latter, by considering all available streamflow information in a region, decreases uncertainty and offers the possibility of improving streamflow quantile estimation".

General comment 2

- It is always strange, to me, to see studies that use Bayesian inference without using prior information, especially when some very useful prior information is out there. For example, for the GEV shape parameter of the stationary model (but also for the nonstationary one) I recommend using (at least) the “geophysical” prior in Martins and Stedinger (2000). They demonstrate that without a prior on the shape parameter of the GEV, maximum-likelihood estimates (and therefore presumably also Bayesian estimates) for hydrological samples are much less accurate than those obtained with other methods (e.g., L-moments). The “geophysical” prior in Martins and Stedinger (2000) provides “common sense” limits, but even better constraints could be obtained through regional analysis, of course.

Response: We found this comment very useful and we repeated the analysis with an informative prior for the shape parameter. The following paragraph was added in the “Flood frequency analysis – Competing models” section (lines 169-175):
“For the shape parameter an informative normal distribution with mean 0.093 and standard deviation 0.12 is used. This distribution is adopted from a global study of extreme rainfall by Papalexiou and Koutsoyiannis (2013), which, to our knowledge, summarizes an analysis of shape parameters using the largest number of stations with hydrological data worldwide. Although rainfall extremes may be characterized by slightly different shape parameter than those of streamflow, our informative prior is very close to the “geophysical prior” of Martins and Stedinger (2000), which is often used to restrict the range of shape parameters based on previous hydrological experience (Renard et al. 2013). The latter prior was not preferred because it is bounded to the interval (-0.5, 0.5), while the distribution of Papalexiou and Koutsoyiannis (2013) allows more extreme shape values with a low probability”.

General comment 3
- The motivation for assuming that only the location parameter varies in time (through its relationship to the covariates) should be discussed more in detail. Considering the proposed model, with the scale and shape parameters fixed, implies that the variance of the flood series does not change over time (e.g., if the mean annual flood peak increases of 5 m3/s, also the 100-yr flood increases of 5 m3/s, and so all other quantiles).
Is this a reasonable assumption? For example, Serago and Vogel (2018) strongly criticize it and propose to use models with the coefficient of variation of the flood series constant in time, since that is consistent with observations in many studies (see the cited literature there). Using a model where CV is constant would be as parsimonious as the one used here and, according to Serago and Vogel (2018), more justified. I suspect that using this other assumption would not invalidate the spatial patterns that are shown in Figures 1 to 4, but would result in very different values in Figures 5 to 7.

Response: The reviewer is right, indeed both the location and scale parameter are expected to change based on the climate state. However, we tested for significance of varying scale parameter by running the model with both location and scale variable. This preliminary study showed only very few cases with significant slopes of the scale parameter. For this reason and for reasons of parsimony, we decided to keep the scale parameter stable and to condition only the location parameter on the climate indices. A comment on this issue was added in the “Flood frequency analysis – Competing models” section (lines 149-152):
“A preliminary analysis considering the effect of a covariate on both the location and scale parameter (cf. section 2.3 below) did not provide very different results than those for a covariate on the location parameter only (not shown). Consequently and for reasons of parsimony, we examine only conditional extreme value distributions with a time-varying location parameter”.
The model with constant coefficient of variation (CV) is an interesting alternative to the model that we present in our manuscript. However, investigating additionally this model would lead to a different and considerably extended manuscript. We feel that such a change is beyond the scope of this paper. The possibility of this model is discussed in the discussion and conclusions section (lines 399-401):
“A constant coefficient of variation as in Serago and Vogel (2018) would also be possible and as parsimonious as our model. In this case, a varying scale parameter linked to the location parameter would need to be implemented”.

General comment 4
- One issue I would also suggest to discuss is the uncertainty in the covariates. The model used here assumes that the covariates are exactly known. If the uncertainty in their knowledge would also be included, would the flood quantile estimates still be more precise than for the classical GEV model?

Response: Thank you for this comment. In our manuscript we investigate only contemporaneous relationships between climate indices and flood peaks and do not focus on prediction. Our goal is mainly to identify spatial patterns of these relationships. For this reason we assume that covariates are exactly known. Of course if one wants to use the current model in a predictive mode, the uncertainty in the covariates must be additionally considered. A comment on this issue was added in the discussion (lines 413-415):
"The contemporaneous streamflow-covariate setup presented here can be used, together with a seasonal prediction of indices, for an ahead-season forecast of streamflow quantiles. In this case covariate uncertainty must be additionally considered".

Additional detailed comments:
Line 20: I would expect that the improvement of estimation of flood probabilities is conditional on how well the covariates can be predicted.

Response: The reviewer is right. This sentence was omitted in the revised manuscript.

Line 71: I am a little confused by the positive-negative anomalies vs. Northernsouthern Europe because the sentence terminates with "during its positive state".
Maybe a rephrasing could help.

Response: In the new version a change was made from "Particularly NAO has been shown to significantly influence the European winter climate with positive (negative) anomalies of moisture fluxes, cyclone passages and precipitation over northern (southern) Europe during its positive state" to "Particularly NAO has been shown to significantly influence the European winter climate: its positive state has been linked to positive (negative) anomalies of moisture fluxes, cyclone passages and precipitation over northern (southern) Europe".

Line 107: The motivation of using Bayesian inference because of the quantification of uncertainty sounds a bit weak. The quantification of uncertainties is possible also with other methods than Bayesian, which is instead usually selected when subjective preferences or prior information is available (at least by us... statisticians have more profound reasons).

Response: The reviewer is right. The reasons for choosing a Bayesian framework were better highlighted. The following sentences were added (lines 95-97):
"A Bayesian framework is adopted for the flood frequency analysis because of its advantages concerning the quantification and interpretation of uncertainty. Furthermore, prior information about hydrologic extremes exists in the literature and can be used for inference".

Line 150: The climate covariates are assumed exactly known in the method. Would it be possible to account for the fact that they are stochastic variables as well? I do not ask to change the method but maybe some discussion could be dedicated to this issue (see main comments).

Response: This point was answered in our reply to general comment 4.

Line 158: The motivation for assuming that only the location parameter vary in time, i.e., the brevity of records, is not very convincing. The Authors should discuss it more (see main comments).

Response: This point was answered in our reply to general comment 3.

Line 174: Since Bayesian inference is done here, there is no reason why priors should not be used. For the GEV shape parameter of the stationary model I recommend to use (at least) the “geophysical” prior in Martins and Stedinger (2000) (see main comments).

Response: This point was answered in our reply to general comment 2.

Line 174: Which non-informative priors are used? Not all of them would result in the same inference. For repeatability, they should be stated.

Response: We added a description of the prior distributions used: uniform priors for the location and scale parameters and a normal informative prior for the shape parameter (see also our reply to general comment 2.

Line 184: I would also look at the posterior distribution of the slope parameter and do the same as the Authors do here. I would just add a sentence to state that this is not a significance test (which has no meaning in Bayesian statistics).

Response: The comment was adopted. The following sentence was added (line 186): “Conditional models are considered as significant if the zero value is not included in the 90% posterior interval of the slope parameter (and thus not by means of a significance test)”.

Line 216: I worry that, for engineering purposes, the estimates of 100-yr floods through GEV without accounting for regional information is not to be recommended.

Response: This point was answered in our reply to general comment 1.

Line 221: Starting the sentence with “Since a Bayesian framework is used” is confusing because it sounds like saying that uncertainties cannot be quantified with other methods too.

Response: The sentence was omitted in the revised manuscript.

Line 224 (and elsewhere): I would use the wording “posterior mode” instead of “maximum likelihood” because they may not be the same (it depends on the type of noninformative priors that are used). Bayesian posterior predictive distribution of flood peak quantiles or their posterior mean could have also been used. Is there a reason for choosing the posterior mode?

Response: The posterior mode was initially used in order to make results comparable with those of frequentist approaches. In the revised manuscript, the posterior median of flood peak quantiles was instead used because it is more representative of the posterior distribution.

Line 258: Is there any (even speculative) reason for the contradicting patterns for Scandinavia?

Response: In the revised manuscript a possible reason for these contradicting patterns is shortly discussed (lines 254-256).
“Scandinavian rivers usually have small catchments and are particularly fed by snowmelt in spring, subsequently in this area, both temperature and precipitation are important for runoff generation”.

Line 265: I also believe that the coherence in space is indicative of a real signal, however spatial correlation of the flood time-series could be a nuisance here, meaning that one sees the same dynamics in many sites because the same floods are occurring there (and therefore they should count as one site only). Since the spatial patterns are here over very large regions, the spatial correlation of the flood time-series cannot alone be responsible for it. However, I would suggest mentioning the problem.

Response: We agree that spatial correlation of floods plays a role for the detected coherence particularly for smaller regions, i.e. nearby gauges. We added a comment on the spatial correlation of the flood time-series (lines 261-264).
"It can be argued that these two latter cases could occur solely by chance or due to spatial correlation of nearby flood time series, however, results are coherent in space and cover large regions, which suggests a real influence of the circulation modes on the location parameter of the extreme value distributions, restricted though to certain sub-regions of Europe".

Line 299: One curiosity. Since the proposed model has constant variance (and the dependence of small and large floods on the covariate is the same in terms of the difference in m3/s) I suspect the relative difference to be affected by catchment area (meaning by the average flow in the river). Is it the case? Of course, since the model is fitted independently to every site, the differences in fitted shape parameters will make this relationship noisier.

Response: Since the difference between streamflow quantiles for high and medium covariate is normalized by the streamflow quantile for medium covariate we were not expecting that the catchment size plays a role in the percent relative differences. We assume that the high relative differences are due to a stronger influence of the climatic indices.

Section 3.2: I wonder how much the relative differences calculated here are due to the slope of the regression for the location parameter vs. the estimated shape parameters. Since no priors are used, I suspect that the posterior distributions of the shape parameter can be wide (spanning unreasonable values) and widely different between sites. Maybe a figure/table that also informs the reader about the obtained shape parameters would be useful.

Response: We added Table 4 with summary statistics of the shape parameter.

Figure 7: Shouldn't be the classical GEV the same within each column? The credible bounds look different. Have I missed something?

Response: We are sorry, this was a typo error noticed by the reviewer and was corrected.

Lines 335: The asymmetry of the credible bounds around the posterior mode is very well expected. If the posterior predictive distribution (or posterior mean) would have been used, that would have lied much more in the center of the bounds.

Response: The posterior median is used now. Indeed credible bounds are less asymmetric.

Line 352: I would add here a brief discussion on the predictability of the covariates since that is needed to make use of the model for prediction.

Response: This point was answered in our reply to general comment 4.

Line 357: I don't get the meaning of "...leads to highly varying flood quantile estimations for different probabilities of exceedance". Is the sentence referring to variations in time? Or space? Or between models with different covariates? And, finally, the variability for "different probabilities of exceedance" of flood peaks at one site exists in terms of relative differences. In term of difference in m3/s, there is no variability at all, since in the models only the location parameter can vary. Maybe I just misunderstood. A rephrasing could help.

Response: The variability concerns flood quantiles for the same station and probability of exceedance and for different values of the climate indices. It was rephrased to:
“For models with significant slopes, variations of the climate indices lead to highly varying flood quantile estimations for the same probability of exceedance”.

Line 358: Related to my previous comment on line 299: is it because the catchments in North-West Scandinavia and Britain are smaller than the others? Or is it because of unreasonably large shape parameters of the GEV?

Response: The highly varying results in this area are in our opinion the result of a more important influence of the circulation indices. No influence of the catchment size was found.

Line 363: It is for me hard to see the decadal-scale variability in Figure 7. Maybe that could be shown in the figure.

Response: In the revised manuscript Fig. S5 was added in the supplementary material showing the evolution in time of the climate indices and their decadal-scale variability. We think that this will help the readers better interpret Fig. 7.

Line 402: As an additional challenge (on top of the three that the Authors have listed) I would add the fact that now covariates are assumed perfectly known and should be instead treated as stochastic variables, I think.

Response: This point was answered in our reply to general comment 4.

Reply to Referee #2 Elena Volpi:

The manuscript investigates the effectiveness of performing climate-informed extreme value analysis for flood probability estimation at the European scale. More specifically, the Authors analyze the effects of large-scale circulation patterns on seasonal extreme distributions by accounting for the relationship between extreme probabilities and climatic indices. As stated by the Authors, climatic indices are considered in recent literature works to justify or explain a non-stationary behavior depicted by extreme events.
In this regard, the innovative contribution of this paper is to perform a large-scale analysis, at a spatial scale that is "comparable" to that of the climatic indices considered in the work aiming at defining the conditional probability distribution of extreme flood events and proving coherent spatial patterns.
The manuscript is well written and organized; the methodology is almost well described, even if additional details could be included to help for reader understanding, and conclusions are well supported by results. Finally, within the Conclusion Section a detailed list of the limitations of the study is provided. Summarizing, the topic is of interest for the scientific community and the manuscript deserves to be considered for publication in this Journal. I have some comments about the work that are listed in the following paragraph; I hope that they will be helpful for manuscript improvement.

Response: We would like to thank Elena Volpi for her comments. In the revised manuscript we followed most of the reviewer's recommendations, since this definitely improved our study. Below, we provide justification for some suggestions that we did not follow. We saw from the comments of the reviewer that some parts of our study needed a more detailed explanation. In our revised manuscript we provide these additional details.

General comment 1
1. The Authors hint in the Introduction Section at the nonstationary framework incorporating climatic indices into flood frequency analysis, but they do not make a clear distinction between periodicity (or cyclo-stationarity) and trends (in the mean or variance). For the sake of clarity, this could be discussed from the very beginning of the manuscript (e.g. at line 49). Are the Authors assuming stationarity which is a "prerequisite to make inference from data", as discussed in detail by the cited papers by Koutsoyiannis and Montanari (2015) and Serinaldi and Kilsby (2015)?
General comment 2
2. At line 136 the Authors define the model driven by climatic indices as "climatic informed model", justifying this choice based on the fact that "if covariates have a stochastic structure and no deterministic component, the resulting distribution is not truly nonstationary". I do agree on this, as the Authors states at line 135 that the climatic indices are stochastic process not showing clear trends. But, it is expected they are characterized by persistence and/or periodicity. A detailed description of the stochastic behavior of the climatic indices is missing in the manuscript, while they are clearly described from a physical point of view (lines 59-90). E.g. which are the relevant time-period and is the period covered by observations long enough to catch climatic indices periodicities?

Response to general comments 1 and 2:
We want to thank the reviewer for these interesting comments. In our manuscript we acknowledge the issue that models conditional on time-varying covariates with a stochastic structure can be stationary, even if the probability density function changes in consequent years. However, we feel that addressing the issue of stationarity/nonstationarity (and thus ergodicity) and the stochastic structure of the covariates in adequate detail would change considerably the focus of our manuscript and we prefer not to make this addition. For this reason we chose the term "classical" and "climate-informed models" and we do not refer to stationary/nonstationary models. We will consider going more in this direction in our future work.

General comment 3

3. Even if the aim of the work is to find results at the European scale, I would suggest the Authors to add a figure showing results for a single station, as an illustrative example to explain the methodology and the rationale behind it (e.g. the structure of the climatic informed GEV). Similar to figure 7, it could be of interest to show the evolution in time of the climatic indices (see comments 2) and the performance of classical GEV and climatic-informed GEV, especially for quantile extrapolation, with uncertainty bounds.

Response: We adopted the recommendation of the reviewer and, additionally to the results for three specific stations shown in Fig. 7, we show in Fig. 8 the performance (point estimates and uncertainty bounds) of the classical GEV and the climatic-informed GEV when plotted against the covariate values (here the extrapolation towards more extreme index values and not lower probabilities of exceedance is shown). We feel that this further clarifies our methodology. We did not opt to include an additional figure explaining the climate-informed model because, as we also state in the introduction, during the last years there have been many studies applying such a conditional framework to single or a few stations. We feel there is enough published material explaining this methodology. The suggestion to show the evolution in time of the climate indices was adopted and these are illustrated in Fig. S5 of the supplementary material.

4. If I understand correctly, conditional models preferred o classical GEV in Table 1 are those respecting both criteria (minimum value of DIC and significantly different from zero coefficient of linear variation with the climatic indices); this could be highlighted in the result section from the beginning of the section. The number of times (stations) each conditional model is preferred with respect to classical GEV is not so high, being in the best case the 44% and on average at about 20%. The use of two criteria does not seem to affect this result much (as in lines 276-280); hence, the evidence of the climatic informed model does not appear to be very strong, even if clear spatial patterns emerge. The latter is the more relevant result, based on my opinion, and this should be stressed in the abstract and conclusion sections.

Response: We adopted this suggestion and in the revised manuscript we highlighted the selection criteria in the results section (line 246). Furthermore, in the discussion and conclusions section we highlighted that the effect of each index independently affects on average a 20% of the database (line 351). However, the number of stations affected by at least one index significantly is much higher, especially in winter. We feel that this is a result that indicates a real influence of the circulation indices to the streamflow extremes.

5. Since spatial patterns are influenced by correlations among climatic indices (that are illustrated in the supplementary material as spatial maps), I suggest the Authors to report in the manuscript a table summarizing cross-correlations among the indices (even if they are not an exhaustive measure of the underling complex physical phenomena).

Response: We adopted this suggestion and we added Table S1 in the Supplementary material summarising linear correlations between the seasonal indices.

6. Lines 276-279. DIC is a measure of model evidence; even if the climatic informed model has a smaller value of DIC with respect to classical GEV, the difference among the two values is probably not enough to results in a “strong evidence” of the first model compared to the second one. See, e.g., Kass and Raftery (1995) where two different interpretations of the Bayes factor are provided.

• Kass, R. E., Raftery, A. E. (1995). Bayes factors. Journal of the american statistical association, 90(430), 773-795.

Response: Indeed when information criteria are used for model comparison, the difference of two values does not always provide “strong evidence” for model choice. Here since we are using two criteria, the DIC and slope significance, we feel that the evidence for model selection.

7. Figure 7 compares conditional (climate informed) and unconditional quantiles considering p=0.01 for three stations. It should be clearly stated that conditional quantiles are computed in this case based on the observed values of the climatic indices year by year.

Response: We adopted this suggestion. A sentence stating this year by year quantile calculation was added (Lines 319-320): “Conditional quantiles are calculated on a year-to-year basis, based on the observed values of the selected climate indices”.

8. As the climate informed models have a larger number of parameters (one more in this case) to be estimated based on data, it is expected that their uncertainty bounds are larger than those provided by classical GEV. In other words, nonstationarity flood frequency analysis adds an additional component of uncertainty if the model between parameters and covariates is estimated from data and not fully a-priori defined based on additional physical information (Serinaldi and Kilsby, 2015). However, this is not what emerges from figure 7. This issue should be clarified.

Response: This was a very helpful comment. We realised that Fig. 7 and the conditional/unconditional uncertainty bounds needed to be further discussed. Fig. 7 does not contradict the findings and discussion of Serinaldi and Kilsby (2015). The range of the uncertainty bounds is an interplay between the model complexity and the additional information provided by the more complex models. In general, more complex models not providing extra information are expected to lead to an increase in uncertainty. More complex models providing “adequate” additional information are expected to lead to decreased uncertainty.
In order to explain this relation better, we added Fig. 8 comparing streamflow quantiles for the classical and a conditional model for a single station. Point estimates and credibility intervals of streamflow quantiles for a probability of exceedance 0.02 are plotted versus NAO values. The following paragraph was added (Lines 334-340):
“The range of uncertainty bounds reflects an interplay between model complexity and the additional information provided by the more complex models. In Fig. 7, uncertainty bounds are narrower in the case of the “best” conditional models (e.g. subplot A1). Uncertainty increases when extrapolations are made towards high and low index values. This can be more easily observed in Fig. 8. For the classical case, the range is about 94 m^3/s. For the climate-informed case and NAO = 0 (close to its median value) the range is around 70 m^3/s. The range increases to 74 m^3/s for NAO = 1 and to 80 m^3/s for the most extreme observed NAO value (NAO = -2.1). For a NAO value around 3/-3 the range of uncertainty bounds reaches that of the classical model”.

9. Lines 329-330. This should be true if the climate indices can be accurately predicted. The issue should be discussed further since it is closely related to the implications of the results presented in the paper for practical applications. Furthermore, I’m asking myself if the improvement in flood quantile estimate at the local scale thanks to climate indices is really significant from a practical point of view given the large uncertainty that characterizes all the estimates (fig. 7); I would like to read a comment on this from the Authors.

Response: A comment was added in the discussion explaining the effect of the index uncertainty in the limitations of the study. Indeed if one wants to predict the indices in order to use them for the estimation of streamflow quantiles, uncertainties will be higher. In lines 413-415 we state: “The contemporaneous streamflow-covariate setup presented here can be used, together with a seasonal prediction of indices, for an ahead-season forecast of streamflow quantiles. In this case covariate uncertainty must be additionally considered”.
Concerning the uncertainty of flood quantiles

I our opinion the use of the climate-informed model is significant from a practical point of view, especially when point estimates of flood quantiles and their uncertainty bounds strongly diverge between the classical and climate-informed models (as e.g. in Fig. 7, 8). However, for low exceedance probabilities, uncertainties are large and this can be improved with the use of a regional framework. We discuss this issue in lines 384-389 of the discussion and conclusions section: “In this study, a local, site-specific flood frequency model was developed. This model allowed to identify spatial coherence in relations between streamflow extremes and large-scale atmospheric patterns. However, a shortcoming of this methodology is the high uncertainty of streamflow estimates for high probabilities of exceedance (corresponding for example to the 100- or 200-year flood). Instead of a local framework, a regional framework can be alternatively implemented. The latter, by considering all available streamflow information in a region, decreases uncertainty and offers the possibility of improving streamflow quantile estimation”.

10. Line 58. The Authors could also consider the recent paper from Serinaldi et al. (2018) discussing limitations of nonstationary detection based on trend tests.
• Serinaldi, F., Kilsby, C. G., Lombardo, F. (2018). Untenable nonstationarity: An assessment of the fitness for purpose of trend tests in hydrology.

Response: Thanks for this very interesting paper. We included it in the discussion of limitations of conditional / nonstationary models.

11. Line 71. A reference is needed.

Response: It was added.

12. Please definite t after eq. (3).

Response: It was defined (Line 130).

13. Line 173. Please define what is meant by non-informative priors in this case. If the non-informative prior is a uniform distribution, its support (range of variability of the random parameter) could have effects of posterior distribution and evidence estimation.

Response: A description of the prior distributions was added. In the revised manuscript uniform priors were used for the location and scale parameters and a normal informative prior for the shape parameter.

14. Eq. (8). y or Y ?

Response: We replaced y with Y.

15. Line 194. Please define $\bar{\theta}$.

Response: The definition was added.

16. Line 219. Are the Authors assuming a Gaussian (marginal) distribution for climatic indices? The assumptions on those variables and their stochastic behaviour are not clear (see also general comments 2 and 3).

Response: We are currently not making an assumption about the marginal distribution of climate indices. Figure S6 was added in the supplementary material showing the histograms of all seasonal indices.

Reply to Referee #3 Francesco Marra:

This study presents a methodology to assess and quantify the impact of climatic covariates in the estimation of time-dependent flood probabilities. The method is tested on a wide sample of catchments in Europe. The paper is clearly and concisely written and the topic is of interest for the readers of this journal.
In my opinion, the study should be seen as an additional step in the efforts of the hydrological community towards a better understanding and quantification of flood risk and, as well underlined by the authors, the spatial consistency of the results indicates some degree of significance in the adopted model. However, further steps are required before the suggested method can be effectively applied in practice. Both the reviewers before me pointed out very interesting comments, many of which I happened to share. I come last so I'll try not to overlap.

Response: We would like to thank Francesco Marra for his comments. The referee has shared very valid concerns that we hope to address in the revised version of our manuscript.

General comment 1
In general, my main concern derives from the GEV approach that requires a number of hypotheses and, if not integrated within a regionalization framework, is prone to extremely large uncertainties.

Response: The concern of the reviewer is valid. Indeed, a regional framework is commonly used in order to extrapolate inference to higher return periods. Here, focusing more on identifying coherent patterns in space, we used a local framework, which is able to recognise significant influence of certain indices to the extreme streamflow quantiles in certain regions of Europe. In order to reduce uncertainty, we constrained our analysis to the 50-year return period which covers the data length: in our study the overlapping period between climate indices and streamflow time series is between 50 and 70 years. Furthermore, in order to improve GEV fits, prior information for the shape parameter was included in the analysis (see also our response to main comment 2 of Alberto Viglione).

In addition to the references recommended by Elena Volpi, I may suggest the reading of Marani and Ignaccolo (2015), that provide a different perspective on extreme value analysis and the GEV approach (potentially for nonstationary extremes) that deserves attention.

Response: Thanks for this very interesting paper. We will consider it for our future work.

To conclude, I think the paper definitely deserves publication, but some more discussion and comments on the adopted methods are required. Potentially, some additional analyses could be of help. Below my detailed comments.

• Is the use of climate-informed models contradicting the identical-distribution assumption behind the use of GEV? This perhaps needs to be discussed.

Response: The condition of independent and identically distributed observations of GEV can be relaxed to include parameters conditioned on time-varying covariates. This can be achieved by converting the original data to generalized “residuals” that are identically distributed (Katz et al., 2002). A comment on this issue was added in the revised manuscript (Lines 123-124): “For the climate-informed models the condition of independent and identically distributed observations of the classical GEV is relaxed to include parameters conditioned on time-varying covariates (Katz et al., 2002)”.

• The inclusion of climate information in the model raises the number of parameters to be estimated to 4. Is there a risk of overfitting?

Response: Our climate-informed models have one more parameter compared to the classical GEV, derived by physical reasoning (the climate indices influence the climate and hydrology of the European area). According to e.g. Katz et al. (2002), such models are reasonable. Furthermore, our conditional models are compared to the classical GEV by means of the DIC, which penalizes model complexity. We are thus confident that the possibility of overfitting is minimized.

• How the authors explain that the linear model applied to the scale parameters (rather than location) provides similar results? Shouldn't the two parameters be related one another since the location is related the mean and the scale to the variance of the annual maxima? Is it correct to change one of them and keep the other fixed?

Response: The reviewer is right, indeed both the location and scale parameter are expected to change based on the climate state. However, we tested for significance of varying scale parameter by running the model with both location and scale variable. This preliminary study showed only few cases with significant slopes of the scale parameter. For this reason and for reasons of parsimony, we decided to keep the scale parameter stable and to condition only the location parameter on the climate indices (see Lines 149-152). In the discussion and conclusions sections we further comment on the possibility of a model with a constant coefficient of variation as an alternative to our model (Lines 398-401):
"Furthermore, we also assumed a varying location parameter and constant scale parameter. A constant coefficient of variation as in Serago and Vogel (2018) would also be possible and as parsimonious as our model. In this case, a varying scale parameter linked to the location parameter would need to be implemented".

• The GEV approach is highly sensitive to the shape parameter, which is prone to large estimation uncertainty when derived from short data records (50 years), particularly when using the maximum likelihood method. Why not using an Lmoments estimation method? Could the inclusion of prior information on the GEV shape parameter improve the accuracy of the results? Perhaps this aspect should be addressed in the study to check consistency in the significant indices (in the end shape and scale are then used as prior information in the estimation of the climate-informed model parameters).

Response: Thanks for this interesting comment. The Bayesian approach was chosen because of its advantages concerning quantification of uncertainty. A second reason is the possibility to include prior information for model fitting (see Lines 95-97). The choice of a likelihood-based method offers additionally a straightforward way of including covariates in the frequency analysis. Indeed inclusion of prior information on the GEV shape parameter improves the accuracy of the results. In our revised manuscript we used an informative prior for our classical and climate-informed GEV. Please see our reply to general comment 2 of Alberto Viglione for more details on the prior distribution chosen. The posterior distributions of the parameters of the classical GEV are not currently used as priors for the climate-informed case. Each model is fitted independently. A comment on this issue was added in Lines 168-169.

• A linear model to relate climatic indices and GEV location parameter is chosen. Clearly, more complex models are not recommended due to the limited data sample and overfitting problems, but this represents a simplification of reality. How can this affect the results? This should be discussed.

Response: In our revised manuscript we extended our discussion on alternative models to point out more clearly the limitations because of choosing a linear model (Lines 391-398):
"A linear relationship was assumed between streamflow extremes and the large-scale atmospheric indices. This is a simplification of reality and some relations may be over- or underestimated due to existing non-linearities in the climate-streamflow system. More complex, particularly non-linear models would also be possible candidates for describing the

relation between climate indexes and flood probabilities. However, with increasing model complexity, the chances for model overfitting also increase. In this study we assumed a symmetric influence of the positive and negative phases of the climate indices. However, an asymmetric relation may better describe the effect of certain climate modes on streamflow extremes. For example, Sun et al. (2014) used an asymmetric piecewise-linear regression to account for the different effects of El Niño and La Niña on rainfall extremes in Southeast Queensland, Australia".

• What do the authors recommend for situations in which more than one climatic index is significant?

Response: A multi-linear model for such cases is also possible. For example, in our study there seems to be a significant influence of both NAO and SCA in winter in Central Europe. We comment on this issue in our discussion (Lines 403-408):
"Single covariate models were developed, focusing on the separate effect of each individual climate mode. The methodology can be extended to a model considering several covariates at the same time. In that case, dependencies between the covariates, if existent, should be taken into consideration. López and Francés (2013) overcame this problem by using the principal components of climatic indices as covariates for the flood frequency analysis. This, however, increases the model complexity considerably and thus the chances of model overfitting. This needs to be considered in developing models with multiple covariates".

Minor comments:
- Lines 24-28 in the abstract are not easy to read, I suggest to rephrase them;

Response: We rephrased these sentences. In the revised manuscript we state: "For certain regions, such as Northwest Scandinavia and the British Isles, yearly variations of the mean seasonal climate indices result in considerably different extreme value distributions and thus in highly different flood estimates for individual years that can also persist for longer time periods".

- Introduction: the proposed method is of interest for (re-)insurance applications and for flood risk management. I think the design applications are not interested since year-by-year variability is not relevant

Response: Indeed the proposed framework makes more sense for (re-)insurance purposes and flood risk management. We omitted "engineering design" in line 44.

- 169-172: please provide more details for readers not familiar with the technique;

Response: More details on the No-U-Turn Sampler-Hamiltonian Monte Carlo approach. In lines 163-165 we state: "NUTS is an extension to HMC, a Markov chain Monte Carlo (MCMC) algorithm that avoids the random walk behavior and sensitivity to correlated parameters which characterize many MCMC methods (Hoffman and Gelman, 2014)".