Accepted ms, INTERNATIONAL JOURNAL OF CLIMATOLOGY, doi: 10.1002/joc.5867

Nonstationary peaks-over-threshold analysis of extreme precipitation events in Finland, 1961-2016

Daniele Pedretti[1], Masoud Irannezhad [2,3]

[1]Geological Survey of Finland (GTK), P.O. 96, Espoo (Finland) email: daniele.pedretti@gtk.fi

[2]School of Environmental Science and Engineering, Southern University of Science and Technology (SUSTech), Shenzhen 518055, China

[3]Water Resources and Environmental Engineering Research Unit, P.O. Box 4300, 90014 University of Oulu, Finland

Abstract: There is an urgent need to understand and predict how extreme precipitation events (EPEs) will change at high latitudes, both for local climate change adaptation plans and risk mitigation and as a potential proxy "anticipating" the impact of climate change elsewhere in the world. This paper illustrates that a combination of non-stationary modeling approaches can be adopted to evaluate trends in EPEs under uncertainty. A large database of daily rainfall events from 281 sparsely-distributed weather stations in Finland between 1961 and 2016 was analyzed. Among the tested methods, Poisson distributions provided a powerful method to evaluate the impacts of multiple physical covariates, including temperature and atmospheric circulation patterns (ACPs), on the resulting trends. The analysis demonstrates that non-stationarity is statistically valid for the majority of observations, independently of their location in the country and the season of the year. However, subsampling can severely hinder the statistical validity of the trends, which can be easily confused with random noise and therefore complicate the decision-making processes regarding long-term planning. Scaling effects have a strong impact on the estimates of non-stationary parameters, as homogenizing the data in space and time reduces the statistical validity of the trends. Trends in EPE statistics (mean, 90% and 99% percentiles) and best-fitted Generalized Pareto parameters in the tails of the distributions appear to be stronger when approaching the Polar region (Lapland) than away from it, consistent with the Arctic amplification of climate change. ACPs are key covariates in physically explaining these trends. In particular, the Arctic Oscillation (AO) and North Atlantic Oscillation (NAO) can explain statistically significant increases in extreme precipitation in Lapland, Bothnian and South regions of Finland, particularly during summer and fall seasons.

Keywords*: extreme precipitation; climate change; extreme value analysis; Generalized Pareto; non-stationarity; Poisson distribution.*

1. Introduction

Extreme weather events are expected to increase in the next decades as a consequence of global warming (*e.g.,* IPCC, 2013). Although climate change is noticeable all over the world (e.g. Fortin et al., 2017; Zandonadi et al., 2016), at the highest latitudes it may occur more rapidly than elsewhere due to a phenomenon called *Arctic amplification* (*e.g.*, Serreze and Francis, 2006). The implications of climate change for water resources at high latitudes have been observed for decades (*e.g.*, Hulme, 1995)**,** and are closely related to the change in extreme precipitation events (EPEs) (*e.g.*, Berg et al., 2013). EPEs play a key role in a variety of hydrological applications and freshwater management planning (Gilroy and McCuen, 2012; Green et al., 2011; Irannezhad et al., 2017; Madsen et al., 2014; Veijalainen et al., 2010).

Extreme value analysis (EVA), such as the peaks-over-threshold (POT) approach, provides a useful tool to analyze and predict extreme values in hydrological time series (*e.g.,* El Adlouni et al., 2008; Kim et al., 2017; Liu et al., 2017; Serinaldi and Kilsby, 2015; Silva et al., 2014). Using EVA, precipitation distributions have traditionally been evaluated under the assumption of stationarity (*e.g.*, (Papalexiou and Koutsoyiannis, 2013; Serinaldi and Kilsby, 2014). Yet, climate change can generate trends in atmospheric-related variables, invalidating the assumption of stationarity (*e.g.,* Milly et al., 2008). Although non-stationarity in hydrological time series has been investigated in a variety of studies (Berg et al., 2013; Boberg et al., 2010; Gilroy and McCuen, 2012; Khaliq et al., 2006; Liu et al., 2017; Luke et al., 2017; Maneta and Howitt, 2014; Seo et al., 2012; Silva et al., 2014; Zhang and Li, 2005), quantifying the magnitude of trends in non-stationary EPEs is challenging and uncertain.

One highlighted reason is the lack of statistical representativeness of experimental datasets (*e.g.,* Serinaldi and Kilsby, 2015). Subsampling implies that the presence of non-stationary trends could be "masked" by random statistical noise. Boberg et al. (2010) maintained that the observed change in probability density functions describing rainfall distributions was statistically significant up to a daily amount of about 90–100 mm, which corresponded in their analysis to the 99.5 percentile of their empirical distributions. When power-law models, such as the Generalized Pareto model, are used to describe EPE distributions, multiple heavily tailed models such as exponential or lognormal distributions can also reproduce the tails of the empirical distributions (*e.g.,* Clauset et al., 2009; El Adlouni et al., 2008; Pedretti and Bianchi, 2018).

A second highlighted reason is linked to the different characteristic spatial scale of models used to estimate the impact of climate change on hydrological variables. This issue is well known, for instance,

when downscaling general circulation models (GCMs) (*e.g.,* Wilby and Wigley, 1997). Local-scale changes do not necessary reflect regional-scale changes given, for instance, the influence of local topographic conditions or the effects of atmospheric circulation patterns (ACPs) (*e.g.,* Aalto et al., 2016; Bieniek and Walsh, 2017; Goovaerts, 2000; Liu et al., 2017), affecting the distribution of the statistics characterizing the regionalization of variables. As such, assessment of the non-stationarity of EPEs may require specific spatio-temporal scales to correctly identify and quantify the presence of trends.

This paper presents an analysis of trends in non-stationary EPEs in Finland by focusing on the change in POT in space and time using daily rainfall records collected by the Finnish Meteorological Institute from 281 weather stations sparsely distributed in the country between 1961 and 2016. Secondary information (daily temperature and well known ACPs) is coupled to the primary information contained in the database (rainfall depths) to examine the implication of other climatic parameters acting as time-dependent physical covariates of the primary information.

The general goal of this work is to provide a new quantitative perspective regarding the magnitude of climate change in Nordic countries in terms of extreme climatic patterns, particularly EPEs. The analysis aims at emphasizing the effects of the uncertainty deriving from the statistical fluctuation of the datasets on decision-making processes and the recommendations to provide concerning the change in the frequency of EPEs. By comparing stationary and non-stationary POT-based modeling, the analysis evaluates the implication of temporal trends in EPEs driven by different physical covariates and determines which covariate has a stronger influence on the detected trends. The ultimate aim is to emphasize the implications of the spatio-temporal variability of EPEs, and the need to account for the correct spatial and temporal discretization when estimating the presence of trends in EPEs.

2. Materials & Methods

2.1 Climatic context

Along the north-south axis (~1320 km, from latitude ~60N to ~70N), Finland experiences climatic zonation, mainly influenced by geographic settings (e.g. Aalto et al., 2016; Irannezhad et al., 2014, 2015a, 2018; Korhonen and Kuusisto, 2010; Mikkonen et al., 2015; Veijalainen et al., 2010). Temperatures and cumulative rainfall fluctuate significantly in space and time over the year. The mean monthly temperature in southern Finland spans from around 0 °C in January to 20 °C in July, whereas in northern Finland (towards Lapland), the range is from -20 °C in January to 15°C in July. The spatial and temporal gradients

in temperature are strong, especially in winter (December–March), which affects the accumulation and melting of snow. Finland has no dry season. Summers (June–September) are moderately long in southern coastal areas and shorter elsewhere. Annual precipitation in Finland naturally decreases from south to north. The mean annual precipitation on a national scale during 1911–2011 was 601 mm and mean seasonal precipitation in summer was about 209 m.

During the 20th century, the Finnish climate might have warmed by about 0.7 °C (Tuomenvirta, 2004) and by order of 1.5 °C (Heino, 2014) or over 2 °C (Mikkonen et al. 2015) since the 19th century. According to Jylhä et al. (2004), by the 2050s, the annual mean temperature is projected to increase by 2–5 °C and precipitation by 0%–30%, with respective increases by the 2080s of 2–7 °C and 5%–40%. The increase in the temperature was observed to be highest in November, December, and January, followed by the spring months, while changes in summer months were less evident. The temperature in Finland could be subject to the Arctic amplification of climate change, particularly at northernmost latitudes. Induced warming is expected to be approximately 50% higher in the country compared to the global average (Mikkonen et al., 2015). Conversely, the location of Finland between the Atlantic Ocean and continental Eurasia causes the weather to be very variable, and the temperature signal has thus been reported to be rather noisy.

2.2 Reference dataset

The analyzed dataset comprises 281 weather stations sparsely distributed across the country (Figure 1). For each station, raw information on daily precipitation, r [mm/d] and minimum, mean, and maximum temperatures (respectively t_{min}, t_{mean} and t_{max}, [ºC]) were provided by the Finnish Meteorological Institute (FMI). Although weather stations with century-long rainfall records exist in Finland (*e.g.*, Irannezhad et al. 2016) and all datasets provided by FMI are considered of very high quality, this work exclusively focuses on time series for the period 1961–2016, for the following reasons. The first reason is associated with the limited number of missing records from 1961–2016, comprising <5% of the total dataset, a value not significantly affecting the interpretation of POT statistics (*e.g.,* Pedretti and Beckie, 2015) and thus not biasing the conclusions of this work. The second reason is associated with the homogenization of the datasets (*e.g.*, Toreti et al., 2011; Venema et al., 2013; Vicente-Serrano et al., 2010). An initial statistical assessment based on the stationary POT approach (described below) of the whole dataset, comprising data between 1900 and 2016, suggested a clear breakpoint in the estimated Generalized Pareto parameters around 1961 (data not reported).

The dataset is analyzed here using different spatial and temporal criteria to emphasize the influence of the spatio-temporal variability of the datasets on the trend estimation process. The temporal variability is based on the comparison between "year-based" or year-integrated analysis and "seasonal-based" analysis, which accounts for the intra-annual seasonal variability of the datasets. The specific seasonal boundaries adopted in this work are as follows: winter, 22 December–20 March; spring, 21 March–20 June; summer, 21 June–22 September; fall, 23 September–21 December. An analysis with seasonal boundaries moved by ± 1 month generated qualitatively similar conclusions to those presented in this work.

From a spatial perspective, two scales of variability in rainfall are evaluated: a "national" scale, through which we evaluate the change in rainfall over time at the scale of the country; and a "regional" scale, through which we evaluate the change in rainfall for separate geographical regions with similar climatic patterns. The national analysis is obtained after merging the rainfall time series into a unique dataset, *de facto* eliminating any regional-scale variability. The regional analysis is obtained after splitting the dataset into four geographical regions ("Lapland", "Bothnian", "Lakes" and "South"), as shown in Figure 1. This division is consistent with regional subdivisions adopted in previous hydrological studies in Finland (Korhonen and Kuusisto 2010). "Lapland" is the northernmost region and potentially more exposed to the rebound of the climate change effects from Arctic amplification. The "South" region is opposite to the Lapland region, and therefore less exposed to Arctic rebound. In addition, it is characterized by a milder, more sea-controlled climate. The "Bothnian" region is more central than the two previous regions, but more exposed to a sea-controlled climate than the more continental "Lakes" region.

Besides daily rainfall and temperature data, this work adopts six ACPs that naturally influence the variability in precipitation across Finland, as demonstrated in previous studies (*e.g.,* Irannezhad et al., 2016, 2017):

- the Arctic Oscillation (AO),
- the East Atlantic (EA) pattern,
- the East Atlantic/West Russia (EA/WR) pattern,
- the North Atlantic Oscillation (NAO),
- the Polar/Eurasian (POL) pattern, and
- the Scandinavian (SCA) pattern.

Glantz et al. (2009) provides a comprehensive review of the influences of these ACPs on climate variability over different parts of the world. Time series of standardized monthly values of these ACPs for the period January 1950 to December 2016 were freely obtained from the website of the Climate Prediction Center (CPC) of the National Oceanic and Atmospheric Administration (NOAA), USA (http://www.cpc.ncep.noaa.gov/data/teledoc/telecontents.shtml).

2.3 Peaks-over-threshold (POT) modeling

The POT approach evaluates the distributions of records over a specific threshold (θ). Derived from the classic extreme value theory (Gumbel, 1958), POT analysis relies on the distribution F of random values r conditionally on exceeding a threshold θ, as

$$F_\theta = \frac{F(\theta + r) - F(\theta)}{1 - F(\theta)}. \quad (1)$$

A key property of this approach is that, for an increasingly larger θ, the distribution of empirical values tends to the power-law-based Generalized Pareto (GP) model (*e.g.,* Coles, 2001; Koutsoyiannis, 2004; Pujol et al., 2007; Serinaldi and Kilsby, 2014), with the cumulative density function (cdf) defined as

$$P = 1 - \left(1 + \frac{k\, r_\theta}{\sigma}\right)_+^{-1/k} \quad (2)$$

where $r_\theta = r - \theta$, $\sigma > 0$ is the scale parameter, and k is the shape parameter. This important property means that the original parent distribution describing r does not need to be known *a priori,* as it would tend to a GP distribution for $\theta \to \infty$. However, the selection of an optimal threshold is important, given that a large threshold implicitly leads to smaller empirical datasets and therefore a larger uncertainty when estimating power-law parameters from limited empirical distributions (*e.g.,* Clauset et al., 2009). Indeed, other models could be adopted when the threshold is instead small, such as the Generalized Extreme Values (GEV) model or models belonging to the Gumbel family, such as the Exponential (EXP) or Gamma model (*e.g.,* Millar, 2013; Papalexiou and Koutsoyiannis, 2013). A trade-off between an optimal threshold and an appropriate model reproducing the data must be therefore evaluated before applying POT analysis.

The theoretical behavior of the GP model as a response to the variability of k and σ for an increasing number of samples is described in detail by Pedretti and Beckie (2015). In the GP model (Equation 2), the case with $k > 0$ corresponds to the heavy-tailed or Pareto-type GP model, $k < 0$ to the bounded tail

or beta type, and $k = 0$ to the EXP model. Several studies have focused on k as a "diagnostic" metric to quantify the extremity from POT events (*e.g.*, Serinaldi and Kilsby, 2014). Indeed, similar to other power-law functions, k determines the intensity of the distribution tails, such that two distributions with the same σ contain, on average, more extreme values for $k \to 1$. As such, a distribution with $k \to 1$ indicates a heavier tailed distribution diverging from an exponential distribution. It should be noted, however, that the scale parameters also contribute to the global variability of the dataset. Conceptually similar to the effect of the standard deviation in the Gaussian model, a larger σ for a fixed k corresponds to a broader distribution of values. As such, the bivariate distribution k-σ needs to be studied when related to extreme values. For a known θ, best-fitting couples of parameters k-σ can be obtained using multiple methods (*e.g.,* Coles, 2001), among which maximum likelihood estimators (MLEs) have been shown to be the most accurate methods for estimating power-law-like distributions (*e.g.,* Clauset et al., 2009) and are thus adopted here. For the vector of discrete samples $\boldsymbol{y} = \{y_1, \dots, y_n\}$ of the r value exceeding a selected threshold θ, the likelihood (L) of the probability density functions of $\mathbf{y}$ is calculated as

$$L = \prod_{i=1}^{n} p(\boldsymbol{y}|\check{k}, \check{\sigma}) \tag{3}$$

where the tilde indicates the estimated parameters. For the GP model (Equation 2) embedding the best estimated $\check{k}, \check{\sigma}$ parameters, the log-likelihood (ℓ) is

$$\ell = -\tilde{k} \ln \tilde{\sigma} - \left(1 + \frac{1}{\tilde{k}}\right) \sum_{i=1}^{n} \ln\left(1 + \frac{\check{k}\boldsymbol{y}(\mathrm{i})}{\tilde{\sigma}}\right)_{+} . \tag{4}$$

2.4 Stationary *vs* non-stationary modeling

Although different approaches exist to evaluate the hypothesis of non-stationarity and predict the associated changes, this work adopts two different methods.

The first method consists of evaluation of the trends of extreme statistics and best-fitted parameters obtained from the application of stationary POT analysis to sequential intervals of the rainfall time series. The records of the dataset are split into subsets, each of which corresponds to $r(t, t + \Delta t)$, where Δt is the time kernel or window. For instance, if Δt = 5 years, the initial year is 1961, and using a yearly based approach, the first analysis subset is 1961–1966, the second subset is 1962–1967, and so on. In each subset, the stationary GP model is best fitted to the data, and the relevant parameters and POT statistics are stored and analyzed. The mean rainfall and two percentiles representing the extremes (P90%, P99%)

are also stored and analyzed. The trends in the new time series of estimated values are statistically validated using the Mann-Kendall (MK) method, a well-adopted approach in hydrological time series analysis (*e.g.,* Madsen et al., 2014). Time series passing the MK test are considered non-stationary, while the opposite suggests no significant difference between a stationary and non-stationary modeling approach to describe the data, thus statistically invaliding the assumption of EPE non-stationarity.

The second method is termed the Poisson-Generalized Pareto distribution (Poisson-GPD) approach and is based on the MLE methods of continuous non-stationary models and Poisson distributions (*e.g.,* Katz et al., 2002). For the GP model, the corresponding cdf is generally written as

$$P(\tau) = 1 - \left(1 + k(\tau)\frac{x(\tau)}{\sigma(\tau)}\right)_{+}^{-1/k(\tau)} \quad (5)$$

where $x(\tau) = r(\tau) - \theta(\tau)$. Equation 5 emphasizes the dependence of the GP parameters on the covariate τ, which can be either time (t) or a different variable related to r. The (log-)likelihood of Equation 5 and corresponding best-fitted parameters can be obtained numerically using different methods, including MLE, generalized MLE (GMLE), or Bayesian estimation. This work adopts the well-established *extRemes* package (Gilleland and Katz, 2016; Gilleland et al., 2013). In *extRemes*, the distribution is allowed to gradually shift (*i.e.* to generate a trend) by accounting for τ as a continuous covariate in the model formulation, such that:

$$X(\tau) = X_0 + X_1\tau + X_2(\tau^2) + \cdots + X_\alpha\,(\tau^\alpha) \quad (6)$$

where X is the generic parameter to be estimated (k, σ or θ) and α is the order of the polynomial. Following the examples in Gilleland and Katz (2016), we chose $\alpha = 2$ for θ and $\alpha = 1$ for k and σ. Since the scale parameter σ must be positive, the estimation of the log-transformed σ is adopted, such that $\log\sigma\,(\tau) = \sigma_0 + \sigma_1\tau$. Using *extRemes*, the secondary information available for this study (daily temperature and ACP time series) can be directly implemented as covariates and studied. Intra-annual or inter-annual cycles of the covariates can be explicitly accounted for by including a period function in Equation 6. For instance, a periodic function following a sine wave can be included, such that:

$$X(\tau) = X_0 + X_1\sin(2\pi\tau/T) + X_2\cos(2\pi\tau/T) \quad (7)$$

where T depends on the specific periodic function ($T = 91.31$ d for intra-annual seasonal variations).

Calculating likelihood ratios (R) is a powerful method to determine which statistical model provides the closest fit to the data and to discriminate the presence of trends in statistical variables, including climate change applications (*e.g.,* Madsen et al. 2014). Using this method, the (log-)likelihoods of alternative models are compared. The model presenting the highest (log-)likelihood is the one best representing the data among the tested models. In this work, to evaluate the assumption of non-stationarity against the assumption of stationarity for different covariates τ, we compute R as the deviation of the log-likelihood from a non-stationary GP model embedding τ (ℓ_{ns}) from the stationarity GP model (ℓ_s) as (*e.g.* Coles, 2001):

$$R = 2[\ell_{ns}(\tau) - \ell_s]. \quad (8)$$

Since the effects of subsampling can bias the reliability of the estimation of the best fitting parameters, a goodness-of-fit (GOF) test is run to evaluate the statistical validity of R *(e.g.,* Clauset et al., 2009). The GOF adopts the Kolmogorov–Smirnov (KS) test, which compares the empirical cdf and an ensemble of n_v random vectors (n_v = 1000 in this work) following the same distribution. The GOF reflects the probability (p-value) that the empirical cdf is larger than the cdf of the hypothesized distributions for a set of estimated parameters. The smaller the p-value, the more R is statistically valid; on the contrary, R may be affected by statistical fluctuations (random noise), thus statistically invaliding the assumption of EPE non-stationarity.

3. Results and Analysis

3.1 Stationary analysis

Figure 2 depicts the probability plots of national POT values using different values of θ [mm/d]. In all panels, the empirical data (circles) are obtained by merging all existing records from 1961 to 2016 from the 281 weather stations. For $\theta = 0$ (Figure 2a), the GP model describes well the data for $P < 90\%$, but underestimates the data for $P > 90\%$. The other tested parametric models (EXP, GEV, and Gamma) also fail to describe the tails. For increasing θ, the GP model tends to more accurately match the empirical data. While the Gamma model also nicely fits the data up to $P = 99.9\%$ for $\theta = 1$, as $\theta > 1$ (Figure 2b), the GP clearly outperforms the Gamma model, and closely reproduces the tails for $P > 99.9\%$. In other words, the GP model is more efficient in predicting the occurrence for rare, yet very intense events than the other models, as θ increases (Figure 2c and d). This result matches the expected behavior of the GP model (*e.g.,* Serinaldi and Kilsby, 2014). The greater validity of the GP model for increasing θ not only

applies to the national dataset, but also to time series evaluated at individual weather stations. As reported in the *Supplementary Information – S1*, the behavior of the different tested models for increasing θ was also tested for time series randomly selected from the ensemble of datasets. Qualitatively, the observations match those from the national analysis, even though the number of empirical POT values in individual time series is reduced compared to the national dataset.

Although the quality of the fitting improves for increasing thresholds, the number of POT values (n) decreases as θ grows. Recalling that the root mean square error of GP model fitting associated with subsampling increases proportionally to $\sqrt{n}$ (*e.g.* Pedretti and Beckie, 2015), the correct selection of θ becomes important in order to minimize the effects of subsampling while still ensuring good accuracy of GP fitting. As described in the *Supplementary Information – S2*, we found that θ = 2 mm/d is a valid optimal value, which is used for the remainder of the analysis (except when θ is evaluated as a non-stationary variable, as described in the main manuscript). Independence of tail values for this threshold was also tested, as also described in the *Supplementary Information – S2.*

3.2 Nonstationary analysis

3.2.1 Time-windows-based approach

Figure 3 shows the time-dependent variability of the calculated statistics and estimated GP parameters using the time-windows-based approach, according to the geographical region. We set θ = 2 mm/d as obtained from the stationary GP model analysis (see Section 3.1), and the time window kernel Δt = 5 years. Sensitivity analysis revealed that, qualitatively, the conclusions do not change using kernels of the size of 1 year or 10 years (*Supplementary Information - S3 a-b*). In Figure 3, the horizontal scale in each panel is the sequential year from the beginning of the observations. For the mean and 90% and 99% percentiles, the vertical axis reports the rainfall in [mm/d], while for k and σ, the same axis reports the corresponding best-fitted POT values. The last column is the number of samples used for the estimation. In the panels, m is the slope of the best-fitted linear regression model of the temporal trend. This value corresponds to the mean annual change in one rainfall statistic or in the GP parameter. A bold font ($\boldsymbol{m}$) indicates that the trend passes the MK test, while a normal font (m) indicates the opposite.

The results display considerable variation. The top-left panel in Figure 3 illustrates that in the Lapland region, the mean daily rainfall varies between 2.6 mm/d and 2.8 mm/d, with peaks of higher and lower intensities. A similar pattern is also seen in the mean rainfall across the Bothnian region, and in EPEs (99% and 99.9%) over both the Lapland and Bothnian regions. The repetition of high and low peaks

could be associated with the periodicity of the different phases of influential ACPs and/or long-term (multiyear) temperature variability, as discussed later in the text. For this specific dataset, the mean daily rainfall events decrease from 1961 to 2016 at a very modest rate of 0.002 mm/d per year. Although seemingly low, it corresponds to a mean decrease of more than 0.01 mm/d over the analyzed 55-year time window. Identical results were obtained for the Lakes region. For the Bothnian region, the linear trend is positive with a lower rate than in the previously analyzed regions. However, according to the MK test, the hypothesis of trends in the Bothnian region is null. In the South, the trend is again negative, but also does not pass the MK test.

The two percentiles associated with extreme rainfall events evaluated in these plots, 90% and 99%, are presented in the second and third columns from the left in Figure 3. In Lapland, the results indicate a non-statistically valid decrease in the 90% percentile of 0.003 mm/d and a statistically valid increase in the 99% percentile of 0.004 mm/d. In other words, between 1961 and 2016, the amount of daily rainfall during events with a return frequency of 1 in every 100 years has increased by 0.004 mm/d each year, with an overall mean increase of 0.22 mm/d. In the Bothnian region, an increase is observed for both 90% and 99% percentiles, although both trends are non-statistically valid. A more marked and statistically valid trend is observed for the 90% percentile in the Lakes region, with a mean annual decrease of 0.013 mm/d, corresponding to an overall change of 0.715 mm/d between 1961 and 2016. The statistics remain stable in the South during the considered period.

A similar approach to analyze the data can be adopted for the interpretation of the GP statistics. Only in the Lapland region do the trends for parameters k and σ pass the MK test. In particular, the shape k increases as a function of time, suggesting a higher increase in extreme events with time, while the scale σ shows the opposite behavior, suggesting a lower reduction in the variability of the events. In the South, the test is passed for k and it is slightly positive. In the other regions, these parameters do not show a statistically significant change with time.

The noteworthy regional difference in the behavior of extreme statistics in Finland can be explored more quantitatively using the likelihood ratio (R) of different heavily-tailed models. In particular, an indication of the presence of extremes can be obtained by calculating R between the heavily-tailed GP model, which was shown to closely fit the distribution tails, and the more symmetric EXP model. The higher the value of R, the more the heavily-tailed GP model outperforms the EXP model, and consequently the more extreme values exist in the dataset. For more symmetric distributions and a lower impact of extremes, $R \to 0$. Figure 4 presents the time-dependent change in R for each region compared to the whole country.

The results highlight that, in all individual regions and nationally, the likelihood ratio is always above R=1, suggesting that the GP model consistently outperforms the EXP model in space and time at all considered scales. However, R is higher in the Lapland region than elsewhere, while the South shows the lowest values. The central regions display similar behavior to the national analysis. This suggests that in Lapland, compared to the other Finnish regions, there is a relatively higher likelihood of obtaining heavily tailed power-law distributions than more symmetric exponential distributions. Because heavy tails are associated with more extreme events, the consequence is that extreme rainfall events are more likely to occur in Lapland than elsewhere in the country, particularly in the South (the farthest region from the North Pole). No specific trends are, however, observed in these data; the MK tests are not passed by any of these datasets, contrasting with the statistically valid trend observed for individual statistics and best-fitting parameters.

Since Finland's climate is affected by intra-annual seasonality (spring, summer, fall, winter), and given that key aspects associated with extreme precipitation events are also season-dependent, such as flooding events (*e.g.,* Veijalainen et al., 2010), it is worth analyzing whether a seasonal-based approach generates different results from those when using a yearly-integrated approach. Figure 5 and Figure 6 respectively illustrate the results for the statistics (mean, 90% and 99% percentiles). In all panels, the vertical axis is the m value of the specific variable. The data passing the MK test correspond to the bars marked with a thicker line, while no line emphasizing the borders of bars indicates those datasets not passing the MK test.

Figure 5 reveals that intra-annual seasonality strongly affects the trends in the statistics for all individual regions. While the implication of seasonality varies heterogeneously from region to region, the magnitude of the trends are generally higher than those found in the yearly-integrated analysis. For instance, in the Lapland region, the trends tend to be more positive in summer–fall and more negative in winter–spring. The mean and 90% and 99% percentiles show statistically significant positive trends in fall, indicating an increase in overall rainfall volumes and extreme events in this season. The 99% percentile shows a change close to $m = 0.05$ mm/d in summer and fall, suggesting that the rainfall associated with this percentile has increased on average, according to our analysis, by 2.75 mm/d in the period 1961–2016. Such a change is considered very high and, although this result represents an integrated value over an entire region, it should attract the attention of local administrations given the risks associated with an increase in extreme rainfall events (*e.g.*, flooding risks, which are directly related to rainfall events). According to our analysis, rainfall extremes measured by 90% and 99% percentiles

tend to clearly decrease in spring, with an annual rate close to 0.03 mm/d for the 99% percentile, corresponding to an overall mean decrease of 1.65 mm/d in the period 1961–2016. This result is important, as it highlights that the impacts of climate change or rainfall are strongly season dependent and will not necessarily lead to an increase in extreme values.

In the South region, the mean values behave similarly to the Lakes region, with milder changes than in Bothnian and Lapland regions. While the trends are not statistically valid for P90%, according to the MK test, the magnitude of these trends remains noticeably high (m = 0.032 in the winter is the highest value encountered among all regions), deserving attention, as it would correspond to an overall increase of 1.76 mm/d for the P90% between 1961 and 2016.

Trends in GP statistics calculated using a seasonal-based approach are illustrated in Figure 6. It can be observed that the changes in trends are more significant in the Lapland and Bothnian regions than in the Lakes and South. While there are important changes among regions, the shape parameter k (controlling the distribution tailing) generally shows a large positive increase with time during springs–summers and milder negative decreases in falls–winters. The scale parameter σ (related to the variance of distributions) generally shows the opposite behavior, with more negative trends in spring–summer than in fall–winter. Statistically valid trends evaluated using the MK approach are achieved for several analyzed time series, suggesting that non-stationarity in GP occurs in the analyzed dataset when analyzed using a regional and seasonal-based approach.

3.2.2 Poisson-GPD approach

Figure 7 illustrates the results from the application of non-stationary POT analysis using the Poisson-GPD approach. For each geographical region, the top row represents the ensemble of results (as boxplots) of the likelihood ratio R calculated using Equation 8 and obtained for each studied covariate compared to the stationary analysis, while the bottom row represents the corresponding p-value. The more $R > 0$, the more the non-stationary analysis embedding a specific τ exceeds the stationary analysis. The more $R < 0$, the more the hypothesis of data non-stationarity based on specific covariates loses validity against the hypothesis of stationarity. The R values are inversely correlated with the p-values, as shown in the *Supplementary Information - S4*. The more $p \to 0$, the more valid is the hypothesis of stationarity when compared to the hypothesis of stationarity. On the contrary, the more $p \to 1$, the more the statistical noise dominates the inference.

Given that ten different covariates were studied (temperature, ACPs, and time) and a wide combination of parameters could be tested using the Poisson-GPD modeling approach, a preliminary sensitivity test was run to highlight the parameters to be analyzed in greater detail, in the light of the goals of this analysis. Of the three GP parameters, k and σ displayed irregular regional-dependent and covariate-dependent behavior, while θ showed more consistent and systematic behavior for all regions and covariates. In particular, it was found that the use of a τ-dependent θ resulted in systematically higher R and lower p-values compared to use of a constant θ *(Supplementary Information - S5)*. Similarly, it was found that detrending the data using a sine-wave-based periodic function (Equation 7) generally provided higher R and lower p-values compared to both non-detrended and linearly-detrended data. These results are not entirely unexpected, since allowing the thresholds to depend on functions with higher degrees of freedom generates better fits and thus greater likelihoods. Based on this initial assessment, it was decided to mainly focus on $k(\tau)$ and $\sigma(\tau)$, which are the two parameters directly associated with the intensity of extreme rainfall events, setting a constant threshold $\theta(\tau) = \theta = 2$ mm/d, similar to that found in the stationary analysis (Section 3.1). All evaluated temperature and ACPs covariates are detrended using best-fitted sine-wave functions. For each covariate, in Figure 7 the blue boxplots correspond to $k(\tau)$, while the light brown boxplots correspond to $\sigma(\tau)$. The red line indicates the median, the boxes denote the 25%–75% interquartile region, and the dotted line the 5%–95% interquartile region. The outliers are marked by crosses.

In Figure 8, we note that the majority of R values (including all median values) show positive values, suggesting that the use of the non-stationary model with $k(\tau)$ and $\sigma(\tau)$ provides better fitting than the use of stationary models. This is true independently of the region or adopted covariate. As such, a first general recommendation for decision makers is to prefer non-stationary modeling approaches to stationary modeling approaches. Uncertainty, however, largely affects the results. The range of R boxplots is wide, and in many cases the p-values are found to be above $p = 0.2$, a threshold indicative of statistical validity for power-law distributions (*e.g.,* Clauset et al., 2009; Pedretti and Bianchi, 2018). Observing the median values, it is noted that the largest median values of R (and correspondingly, the lowest median p-values) are found in both the Bothnian region and Lapland when using t_{min}, besides the NAO and AO, respectively, as the covariate for $k(\tau)$ (blue colors). In the Lakes region, no median value of any covariate significantly exceeds the others. Nevertheless, in this region, the use of t_{min} and the EA/WR pattern as the covariate for $\sigma(\tau)$ resulted in the highest R value. On the contrary, in the South, t_{max} and the NAO were the dominant covariates for the largest R value using $\sigma(\tau)$.

In line with the previous conclusions based on the time-windows approach (Section 3.2.1), the analysis using Poisson-GPD suggests that uncertainty remains noticeable in the estimation of the trends and associated parameterization. On the one hand, this means that care must be taken not to overestimate the general presence of non-stationarity by simply analyzing all resulting R values. On the other hand, only statistically valid results may provide insight into the actual presence of trends and suggest which covariate dominates the non-stationarity. In this sense, we calculated and analyzed the proportion of time series exceeding $R = 2$, a threshold that is linked to the threshold p-value of 0.2 (*Supplementary Information - S4*) and thus indicates the statistical validity of the analyzed model results. For yearly-distributed data, Figure 8 displays the relative proportion of values with likelihood ratios exceeding the selected threshold, $n(R > 2)$, against the total population size by region and covariate, $n(tot)$. A line at $n(R > 2)/n(tot) = 0.2$ is used to visually assist the analysis of the figure across the multiple results.

In Lapland (Figure 9a), the new analysis suggests that time has a comparable impact to the EA and SCA patterns, as about 20% of the observations display $R = 2$ or a higher value. For the Bothnian region (Figure 9b), on the other hand, time is statistically the most relevant covariate, while the EA, NAO, and t_{min} are the most relevant physical covariates using the same criterion. For the Lakes region (Figure 9c), time has a somewhat surprisingly smaller influence than t_{min} and the EA pattern, while all remaining covariates have smaller effects. In the South (Figure 9d), time is also statistically the most relevant covariate, while among the physical covariates, the EA and NAO patterns are the most significant ones, similar to what is observed for the Bothnian region.

A similar analysis was repeated to evaluate the number of time series exceeding $R > 2$ using the seasonal time series (Figure 9). “Time” is no longer as dominant as for the annual-based analyses, while physical covariates gain relevance, stressing the importance of performing seasonal- and regional-based analyses compared to annual and national analyses. For instance, the most influential covariate in Lapland is the SCA pattern in spring, while the POL pattern is the most influential in fall. During summer and winter, several covariates present the same validity. In the Bothnian region, the EA and NAO are the most relevant physical covariates in summer, AO and NAO in fall, and EA/WR and POL patterns dominate in winter. The Lakes region shows more homogeneous behavior among all the seasons, while in the South, only the summer seasons present noticeable peaks corresponding to the physical covariates, in particular the SCA pattern. The latter has about 50% of the values exceeding the imposed threshold and is the highest value found in the whole database. This is in sharp contrast to winter in the South, where no values exceeding $R > 2$ are found for SCA.

4. Discussion

The results obtained in this work are consistent with previous analyses of the Finnish climate and provide a new perspective to illustrate the expected behavior of EPEs in Nordic countries.

In the South and Bothnian regions in Finland, the changes in P90% and P99% are statistically valid and noticeable trends were obtained. Similarly, Irannezhad et al. (2017) reported significant ($p < 0.05$) increases in EPE intensity in Finland during 1961–2011, particularly across the south, southwest, and western coastal areas. Given that the majority of the Finnish population lives in the coastal areas of the country, the results should not be ignored in decision-making processes by local administrations.

In Lapland, trends in extreme patterns are generally more statistically valid and the trends present higher slopes than elsewhere in Finland. Irannezhad et al. (2016, 2017) also found significant decreasing (increasing) trends for very light precipitation intensity (frequency) in northern Finland, covering Lapland, compared to the central and southern parts of the country. Such results could also be related to the fact that the total amount of precipitation in Lapland is lower than in other parts of Finland (during 1911–2011, the average cumulative rainfall in Lapland was about 420–500 mm/y *vs* the country mean of about 600 mm/y, *e.g.,* Irannezhad et al., 2014). As such, small variations in EPEs may be more substantially reflected in Lapland than elsewhere in the country.

For some Finnish regions, temperature and ACPs generate higher R values than the variable "time", providing further evidence that the trends in hydrological variables, and specifically EPEs, are linked to trends in the other physical parameters studied in this work and traditionally associated with climate change. In the Lakes region, the EA/WR pattern substantially influenced extreme precipitation, consistent with the analysis by Irannezhad et al. (2016, 2017) at the Kajaani weather station in the same region. The EA/WR pattern generally describes the meridional circulation across Finland, which typically weakens the influences of westerly airflow (Krichak and Alpert, 2005). The positive (negative) phase of the EA/WR pattern reflects negative (positive) pressure anomalies over the west and southwest of Russia (Krichak and Alpert, 2005). Hence, the positive (negative) EA/WR phase is naturally in accordance with the anomalous northwesterly and northerly (southeasterly and southerly) winds (Krichak and Alpert, 2005), resulting in drier (wetter) weather than the normal climate across central Finland. Such a signature of the EA/WR pattern is particularly seen during the summer season (Irannezhad et al., 2015b), in which the highest daily precipitation intensity principally occurs. Similarly, increases in summertime EPEs over the Lakes region determined in this study may reflect the modes of

the EA/WR pattern, which have mostly been negative in recent decades (Vicente-Serrano and López-Moreno, 2006).

EPEs showed the most significant relationships with the AO and NAO over Lapland, but only with the NAO across the Bothnian and South regions. In agreement with this observation, Irannezhad et al. (2016, 2017) identified the AO and NAO indices as the most important ACPs positively influencing annual maximum one-day precipitation and very heavy precipitation, respectively, over northern Finland, covering Lapland. Positive phases of NAO and AO correspond to strong westerly airflows (Thompson and Wallace, 1998) and the prevailing of wetter weather than normal over Finland (*e.g.,* Irannezhad et al., 2014), particularly during the cold season. Serreze et al. (2000) concluded that the NAO could be considered a major component of the AO, which principally expresses annular mode of atmospheric circulation in the Northern Hemisphere (Thompson et al., 2000). Hence, observed increasing trends in both the NAO and AO in recent decades (Wang et al., 2005) could explain statistically significant increases in EPEs in the Lapland, Bothnian and South regions of Finland, particularly seen during the summer and fall seasons. Although the AO index is mainly a winter season pattern, such intensifications of EPEs in warm seasons may refer to the recent declines in pressure over northern regions during April–September (Serreze et al., 1997).

The SCA and POL patterns are the most influential ACPs for variations in EPEs across the Lapland region in northern Finland during the spring and fall seasons, respectively. Similarly, Irannezhad et al. (2014) reported the SCA pattern as the most significant ACP negatively affecting spring precipitation at the Sodankylä station in the Lapland region during 1950–2011. The key action center of the SCA pattern is located across Scandinavia and a large segment of the Arctic Ocean in northern Siberia. With the opposite sign of anomalies, two other centers of this ACP are located in the northeastern Atlantic (Western Europe) and in Mongolia throughout western China. The positive (negative) SCA pattern describes a high (low) pressure system across the Scandinavian peninsula, bringing drier (wetter) air to Finland (Bueh and Nakamura, 2007). On the other hand, Irannezhad et al. (2017) concluded that very and extremely wet precipitation days (R95p and R99p, respectively) during the fall season in northern parts of Finland were negatively associated with the POL patterns in recent decades. The main action center of this ACP is located at the North Pole, while two other separate centers with opposite signs are located across Europe and northeastern China (Barnston and Livezey, 1987). Accordingly, the POL pattern describes major changes in the circumpolar circulation power, exposing systematic variations in mid-latitude airflow across a wide area of Europe and Asia. The positive (negative) phase of the Polar

pattern manifests a strong (weak) circumpolar vortex naturally associated with a drier (wetter) climate over northern Scandinavia.

Conclusion

The combination of multiple methods consistently demonstrates the presence of non-stationary trends in the analyzed dataset, and consequently the need to adopt non-stationary modeling to evaluate the corresponding statistics. However, the results highlight the importance of accounting for the correct spatio-temporal scales to achieve an accurate detection of the trends in the observations, which can be easily masked by statistical noise. Noteworthy regional differences in the behavior of extreme statistics are observed in Finland, as confirmed in particular by the use of log-likelihood ratios and *p*-value statistics. The comparison with an assessment based on nationally integrated datasets suggests that spatial homogenization reduces the statistical validity of the trends, with potential direct implications for the downscaling method. This aspect is left open for future development.

The trends in EPEs and the corresponding statistics and GP parameters are also emphasized when adopting a seasonal-based clustering of the data, as compared with year-round integrated analysis. In some cases, we found an increase in extreme values up to 2.76 mm/d over the analyzed period of 1961–2016, such as rainfall events in Lapland occurring with a probability of 99% (i.e. very extreme events with rainfall close to 20 mm/d or more). In the same regions during the spring, a strong decrease was recorded in the same statistics of about 1.76 mm/d. The prevention and mitigation of flooding risks (*e.g.,* Jylhä et al., 2004; Korhonen and Kuusisto, 2010; Veijalainen et al., 2010) and other hydrogeological-related decision-making processes are known to be season-dependent in Finland, suggesting that a seasonal approach is preferred to a yearly integrated one to predict the trends in EPEs.

The change in EPEs is more marked in Lapland than elsewhere in the country. This is consistent with a more influential effect of Arctic amplification in Nordic regions, widely described in the literature but not always easy to quantify. Indeed, the change in magnitude of GP parameters describing the extremes is generally observed to be milder when approaching the South (*i.e.* farther away from the North Pole), drawing attention to the probability of return of extreme events in the Arctic area.

ACPs are key covariates for evaluating the change in trends in estimated GP parameters. Intra-annual seasonality and regional heterogeneity strongly affect the role of each ACP in the resulting GP parameters in all individual regions, further emphasizing the need to adopt the appropriate spatiotemporal scales when analyzing EPEs.

The most influential covariate in Lapland is the SCA pattern in spring, while the POL pattern is the most influential in fall. During summer and winter, several covariates present the same validity. In agreement with previous studies, EPEs showed significant relationships with the AO and NAO over Lapland, but only with the NAO across the Bothnian and South regions.

The increasing trends in both the NAO and AO observed in recent decades (Wang et al., 2005) could explain the statistically significant increases in EPEs in the Lapland, Bothnian, and South regions of Finland, particularly seen during the summer and fall seasons. As such, these ACPs should be closely monitored in order to properly determine the future trends in GP parameters and consequently make the correct decisions when using extreme value analysis.

Acknowledgements

Masoud Irannezhad was partially supported by the Finnish Cultural Foundation and Ma- ja vesitekniikan tuki r.y. The authors would like to acknowledge the Editor, Dr Sergey Gulev, and three anonymous reviewers, whose comments have contributed to increasing the quality of our manuscript. The authors thank Dr Roy Siddall for support with the language revision.

References

Aalto, J., Pirinen, P., and Jylhä, K. (2016). New gridded daily climatology of Finland: Permutation-based uncertainty estimates and temporal trends in climate. J. Geophys. Res. Atmospheres *121*, 3807–3823.

Barnston, A.G., and Livezey, R.E. (1987). Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns. Mon. Weather Rev. *115*, 1083–1126.

Berg, P., Moseley, C., and Haerter, J.O. (2013). Strong increase in convective precipitation in response to higher temperatures. Nat. Geosci. *6*, 181.

Bieniek, P.A., and Walsh, J.E. (2017). Atmospheric circulation patterns associated with monthly and daily temperature and precipitation extremes in Alaska. Int. J. Climatol. *37*, 208–217.

Boberg, F., Berg, P., Thejll, P., Gutowski, W.J., and Christensen, J.H. (2010). Improved confidence in climate change projections of precipitation further evaluated using daily statistics from ENSEMBLES models. Clim. Dyn. *35*, 1509–1520.

Bueh, C., and Nakamura, H. (2007). Scandinavian pattern and its climatic impact. Q. J. R. Meteorol. Soc. J. Atmospheric Sci. Appl. Meteorol. Phys. Oceanogr. *133*, 2117–2131.

Clauset, A., Shalizi, C.R., and Newman, M.E.J. (2009). Power-law distribution in empirical data. SIAM Rev. *51*, 661–703.

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

El Adlouni, S., Bobée, B., and Ouarda, T.B.M.J. (2008). On the tails of extreme event distributions in hydrology. J. Hydrol. *355*, 16–33.

Fortin, G., Acquaotta, F., and Fratianni, S. (2017). The evolution of temperature extremes in the Gaspé Peninsula, Quebec, Canada (1974–2013). Theor. Appl. Climatol. *130*, 163–172.

Gilleland, E., and Katz, R.W. (2016). extRemes 2.0: An Extreme Value Analysis Package in R. J. Stat. Softw. *72*.

Gilleland, E., Ribatet, M., and Stephenson, A.G. (2013). A software review for extreme value analysis. Extremes *16*, 103–119.

Gilroy, K.L., and McCuen, R.H. (2012). A nonstationary flood frequency analysis method to adjust for future climate change and urbanization. J. Hydrol. *414–415*, 40–48.

Glantz, M.H., Katz, R.W., and Nicholls, N. (Neville) (2009). Teleconnections linking worldwide climate anomalies (Cambridge University Press).

Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J. Hydrol. *228*, 113–129.

Green, T.R., Taniguchi, M., Kooi, H., Gurdak, J.J., Allen, D.M., Hiscock, K.M., Treidel, H., and Aureli, A. (2011). Beneath the surface of global change: Impacts of climate change on groundwater. J. Hydrol. *405*, 532–560.

Gumbel, E.J. (1958). Statistics of Extremes (Columbia University Press).

Heino, R. (2014). Climatic changes in Finland during the last hundred years. Fenn. - Int. J. Geogr. *150*.

Hulme, M. (1995). Estimating global changes in precipitation. Weather *50*, 34–42.

IPCC (2013). Climate Change 2013 - The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, New York, USA. ISBN 978-1-107-05799-1 (British library).

Irannezhad, M., Marttila, H., and Kløve, B. (2014). Long-term variations and trends in precipitation in Finland. Int. J. Climatol. *34*, 3139–3153.

Irannezhad, M., Chen, D., and Kløve, B. (2015a). Interannual variations and trends in surface air temperature in Finland in relation to atmospheric circulation patterns, 1961–2011. Int. J. Climatol. *35*, 3078–3092.

Irannezhad, M., Haghighi, A.T., Chen, D., and Kløve, B. (2015b). Variability in dryness and wetness in central Finland and the role of teleconnection patterns. Theor. Appl. Climatol. *122*, 471–486.

Irannezhad, M., Marttila, H., Chen, D., and Kløve, B. (2016). Century-long variability and trends in daily precipitation characteristics at three Finnish stations. Adv. Clim. Change Res. *7*, 54–69.

Irannezhad, M., Chen, D., Kløve, B., and Moradkhani, H. (2017). Analysing the variability and trends of precipitation extremes in Finland and their connection to atmospheric circulation patterns. Int. J. Climatol. *37 (Suppl. 1)*, 1053–1066.

Irannezhad, M., Moradkhani, H., and Kløve, B. (2018). Spatio-temporal variability and trends in extreme temperature events in Finland over the recent decades: influence of Northern Hemisphere teleconnection patterns. Adv Meteo *7169840*.

Jylhä, K., Tuomenvirta, H., and Ruosteenoja, K. (2004). Climate change projections for Finland during the 21st century. Boreal Environ. Res. *9*, 127–152.

Katz, R.W., Parlange, M.B., and Naveau, P. (2002). Statistics of extremes in hydrology. Adv. Water Resour. *25*, 1287–1304.

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

Kim, H., Kim, S., Shin, H., and Heo, J.-H. (2017). Appropriate model selection methods for nonstationary generalized extreme value models. J. Hydrol. *547*, 557–574.

Korhonen, J., and Kuusisto, E. (2010). Long-term changes in the discharge regime in Finland. Hydrol. Res. *41*, 253–268.

Koutsoyiannis, D. (2004). Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation / Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: I. Recherche théorique. Hydrol. Sci. J. *49*, 575–590.

Krichak, S.O., and Alpert, P. (2005). Decadal trends in the east Atlantic–west Russia pattern and Mediterranean precipitation. Int. J. Climatol. *25*, 183–192.

Liu, S., Huang, S., Huang, Q., Xie, Y., Leng, G., Luan, J., Song, X., Wei, X., and Li, X. (2017). Identification of the non-stationarity of extreme precipitation events and correlations with large-scale ocean-atmospheric circulation patterns: A case study in the Wei River Basin, China. J. Hydrol. *548*, 184–195.

Luke, A., Vrugt, J.A., AghaKouchak, A., Matthew, R., and Sanders, B.F. (2017). Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States. Water Resour. Res. *53*, 5469–5494.

Madsen, H., Lawrence, D., Lang, M., Martinkova, M., and Kjeldsen, T.R. (2014). Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. J. Hydrol. *519*, *Part D*, 3634–3650.

Maneta, M.P., and Howitt, R. (2014). Stochastic calibration and learning in nonstationary hydroeconomic models. Water Resour. Res. *50*, 3976–3993.

Mikkonen, S., Laine, M., Mäkelä, H.M., Gregow, H., Tuomenvirta, H., Lahtinen, M., and Laaksonen, A. (2015). Trends in the average temperature in Finland, 1847–2013. Stoch. Environ. Res. Risk Assess. *29*, 1521–1529.

Millar, R. (2013). A statistical approach for deriving project design rainfall. (Melbourne, Australia: The Australasian Institute of Mining and Metallurgy), pp. 273–276.

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., and Stouffer, R.J. (2008). Stationarity Is Dead: Whither Water Management? Science *319*, 573–574.

Moseley, C., Haerter, J.O., and Berg, P. (2013). Strong increase in convective precipitation in response to higher temperatures. Nat. Geosci. *6*, 181.

Papalexiou, S.M., and Koutsoyiannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall. Water Resour. Res. *49*, 187–201.

Pedretti, D., and Beckie, R.D. (2015). Stochastic evaluation of simple pairing approaches to reconstruct incomplete rainfall time series. Stoch. Environ. Res. Risk Assess. 1–14.

Pedretti, D., and Bianchi, M. (2018). Reproducing tailing in breakthrough curves: Are statistical models equally representative and predictive? Adv. Water Resour. *113*, 236–248.

Pujol, N., Neppel, L., and Sabatier, R. (2007). Approche régionale pour la détection de tendances dans des séries de précipitations de la région méditerranéenne française. Comptes Rendus Geosci. *339*, 651–658.

Seo, L., Kim, T.-W., Choi, M., and Kwon, H.-H. (2012). Constructing rainfall depth-frequency curves considering a linear trend in rainfall observations. Stoch. Environ. Res. Risk Assess. *26*, 419–427.

Serinaldi, F., and Kilsby, C.G. (2014). Rainfall extremes: Toward reconciliation after the battle of distributions. Water Resour. Res. *50*, 336–352.

Serinaldi, F., and Kilsby, C.G. (2015). Stationarity is undead: Uncertainty dominates the distribution of extremes. Adv. Water Resour. *77*, 17–36.

Serreze, M.C., and Francis, J.A. (2006). The Arctic Amplification Debate. Clim. Change *76*, 241–264.

Serreze, M.C., Carse, F., Barry, R.G., and Rogers, J.C. (1997). Icelandic low cyclone activity: Climatological features, linkages with the NAO, and relationships with recent changes in the Northern Hemisphere circulation. J. Clim. *10*, 453–464.

Serreze, M.C., Walsh, J.E., Chapin, F.S., Osterkamp, T., Dyurgerov, M., Romanovsky, V., Oechel, W.C., Morison, J., Zhang, T., and Barry, R.G. (2000). Observational evidence of recent change in the northern high-latitude environment. Clim. Change *46*, 159–207.

Silva, A.T., Portela, M.M., and Naghettini, M. (2014). On peaks-over-threshold modeling of floods with zero-inflated Poisson arrivals under stationarity and nonstationarity. Stoch. Environ. Res. Risk Assess. *28*, 1587–1599.

Thompson, D.W.J., and Wallace, J.M. (1998). The Arctic oscillation signature in the wintertime geopotential height and temperature fields. Geophys. Res. Lett. *25*, 1297–1300.

Thompson, D.W., Wallace, J.M., and Hegerl, G.C. (2000). Annular modes in the extratropical circulation. Part II: Trends. J. Clim. *13*, 1018–1036.

Toreti, A., Kuglitsch, F.G., Xoplaki, E., and Luterbacher, J. (2011). A Novel Approach for the Detection of Inhomogeneities Affecting Climate Time Series. J. Appl. Meteorol. Climatol. *51*, 317–326.

Tuomenvirta, H. (2004). Reliable estimation of climaric variations in Finland. PhD thesis. University of Helsinki.

Veijalainen, N., Lotsari, E., Alho, P., Vehviläinen, B., and Käyhkö, J. (2010). National scale assessment of climate change impacts on flooding in Finland. J. Hydrol. *391*, 333–350.

Venema, V.K., Mestre, O., Aguilar, E., Auer, I., Guijarro, J.A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., and Zahradnicek, P. (2013). Benchmarking homogenization algorithms for monthly data. In AIP Conference Proceedings, (AIP), pp. 1060–1065.

Vicente-Serrano, S.M., and López-Moreno, J.I. (2006). The influence of atmospheric circulation at different spatial scales on winter drought variability through a semi-arid climatic gradient in northeast Spain. Int. J. Climatol. *26*, 1427–1453.

Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., García-Vera, M.A., and Stepanek, P. (2010). A complete daily precipitation database for northeast Spain: reconstruction, quality control, and homogeneity. Int. J. Climatol. *30*, 1146–1163.

Wang, D., Wang, C., Yang, X., and Lu, J. (2005). Winter Northern Hemisphere surface air temperature variability associated with the Arctic Oscillation and North Atlantic Oscillation. Geophys. Res. Lett. *32*.

Wilby, R.L., and Wigley, T.M.L. (1997). Downscaling general circulation model output: a review of methods and limitations. Prog. Phys. Geogr. Earth Environ. *21*, 530–548.

Zandonadi, L., Acquaotta, F., Fratianni, S., and Zavattini, J.A. (2016). Changes in precipitation extremes in Brazil (Paraná River Basin). Theor. Appl. Climatol. *123*, 741–756.

Zhang, Y.-K., and Li, Z. (2005). Temporal scaling of hydraulic head fluctuations: Nonstationary spectral analyses and numerical simulations. Water Resour. Res. *41*, W07031.

Figures

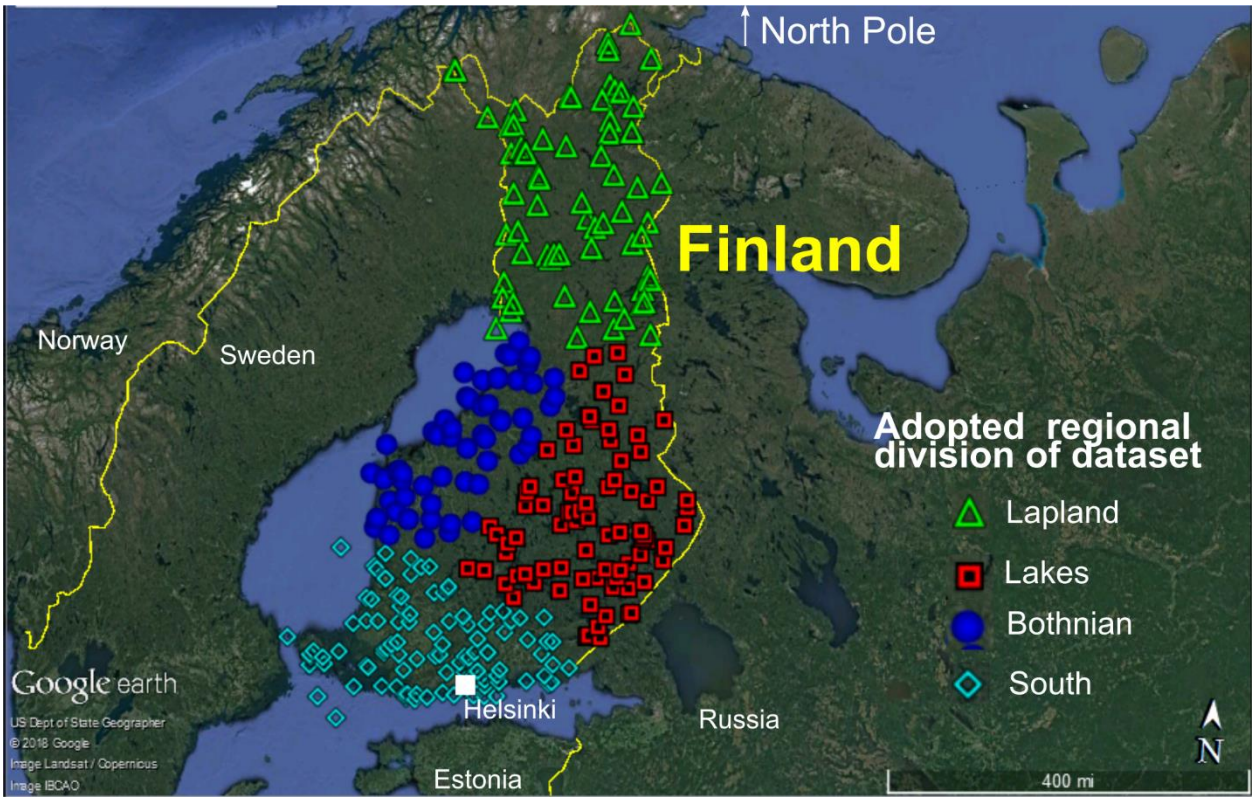

Figure 1 Location of the weather stations analyzed in this work for the analysis of the extreme precipitation events. The subdivision by regions is based from previous works in Finnish contexts. Image taken from Google Earth Pro®, accessed on May 2nd, 2018 at the Geological Survey of Finland.

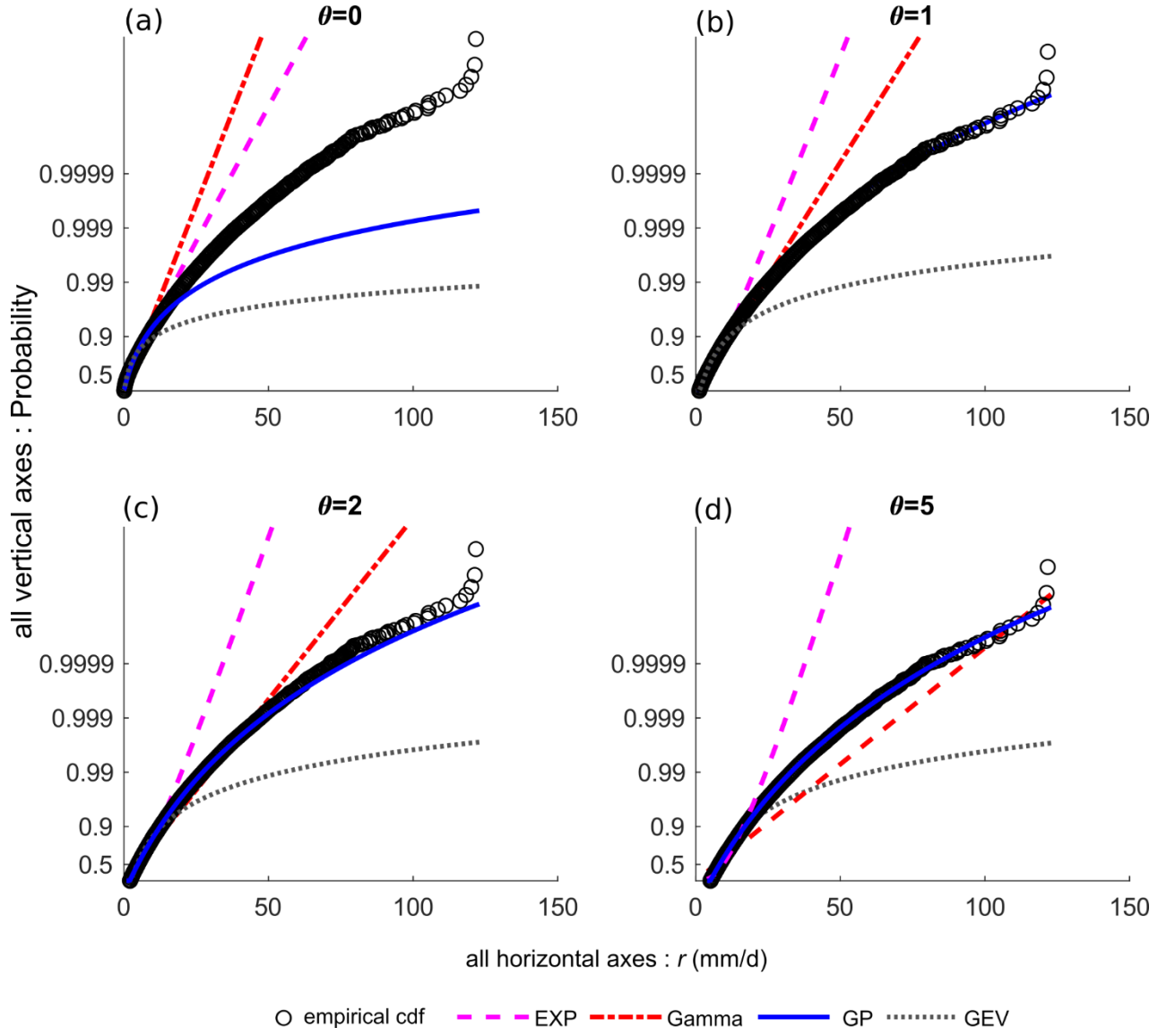

Figure 2 Probability plots of the nation-wide stationary peaks-over-threshold values for different tested thresholds θ [mm/d] (panel a=0 mm/d, panel b=1 mm/d, panel c=2 mm/d, panel d=5 mm/d). The lines represent the best-fitted models following exponential (EXP), Gamma, Generalized Pareto (GP) and Generalized Extreme Values (GEV) distributions. The GP model clearly outperforms the other models as θ grows.

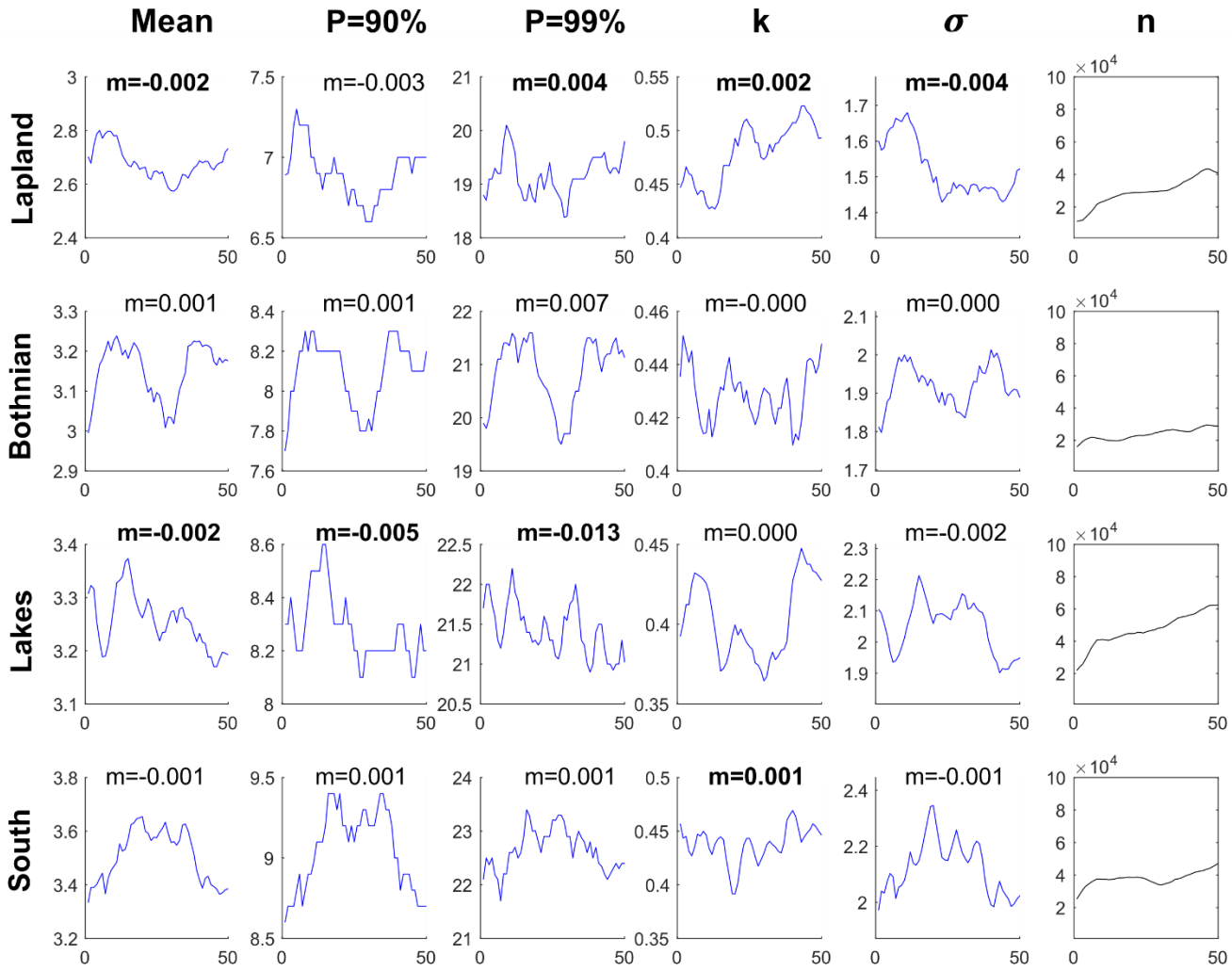

Figure 3 Non-stationary yearly-based analysis of extreme rainfall events exceeding a threshold $\theta = 2$ mm/d with a moving time window $\Delta\tau = 5$ years, divided by regions. In all plots, the vertical axis represents the cumulative daily precipitation [mm] and the horizontal axis the time [years] since 1961. m = best-fitted regression slope (trend) of the data. Bold m indicates that the trend passes the Mann-Kendall test. The last column represents the total number of records used for the calculation.

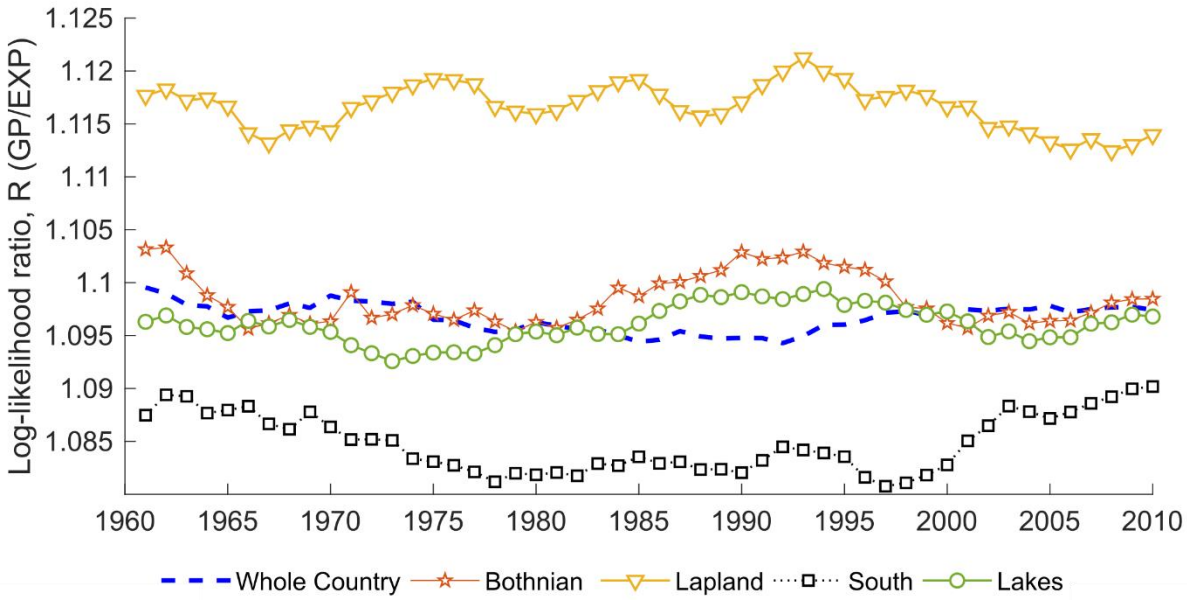

Figure 4 Time-dependent log-likelihood ratios between the exponential (EXP) and the Generalized Pareto (GP) models best fitting the POTs values, by region. The corresponding time-dependent ratio for the whole country is reported for comparison purposes. The threshold is θ=2 [mm/d] for all datasets, the time window is Δt=5 [years].

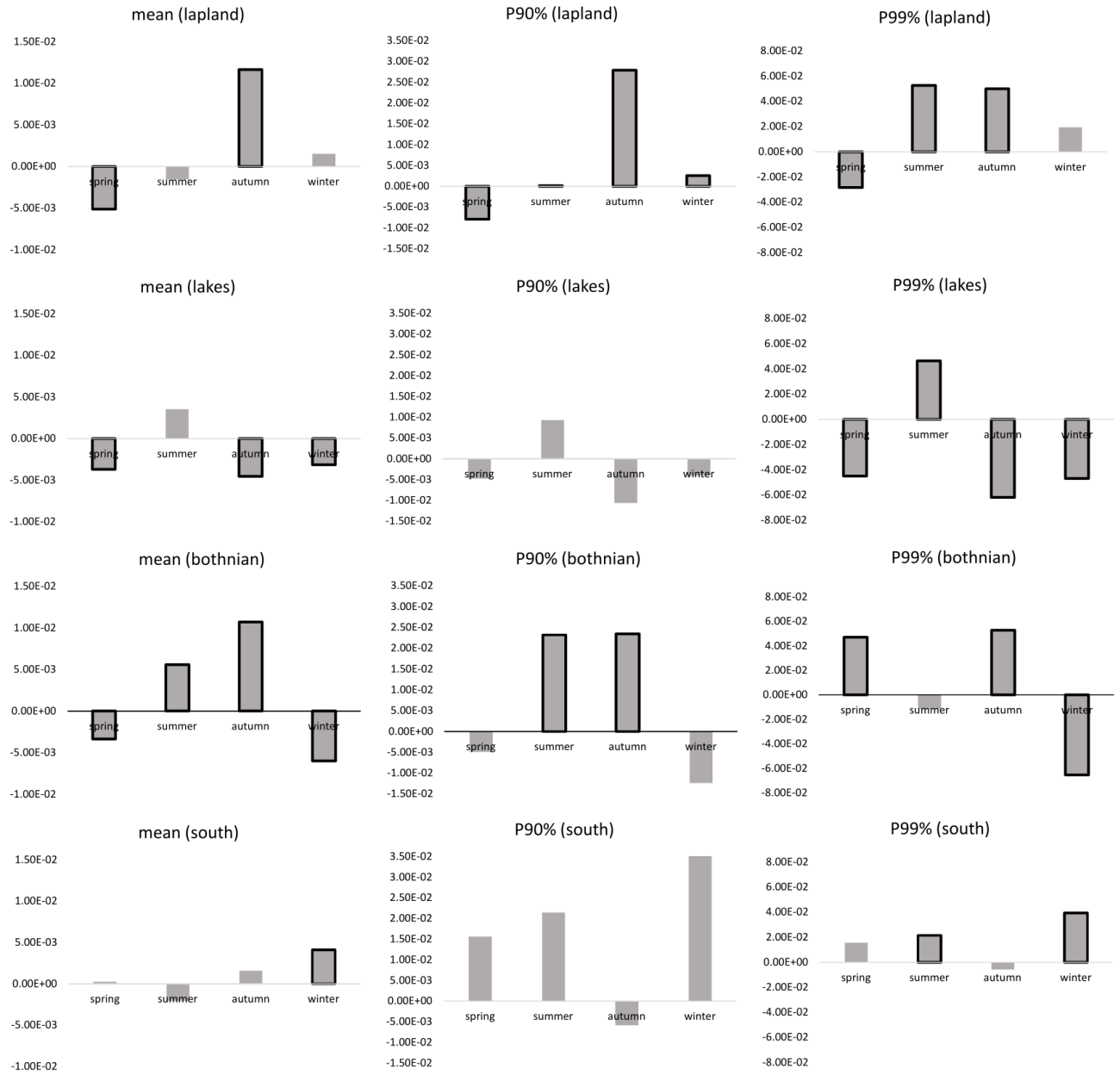

Figure 5 trends in the statistics (mean, 90% and 99% percentiles) calculated using the time-windows-, seasonal- and regional-based analysis of extreme rainfall events exceeding a threshold θ=2 [mm/d] and with time window $\Delta\tau$=5 [years]. The vertical axes are the best-fitted linear regression slopes (m) of each statistic. The data passing the Mann-Kendall test are marked with thicker bar borders.

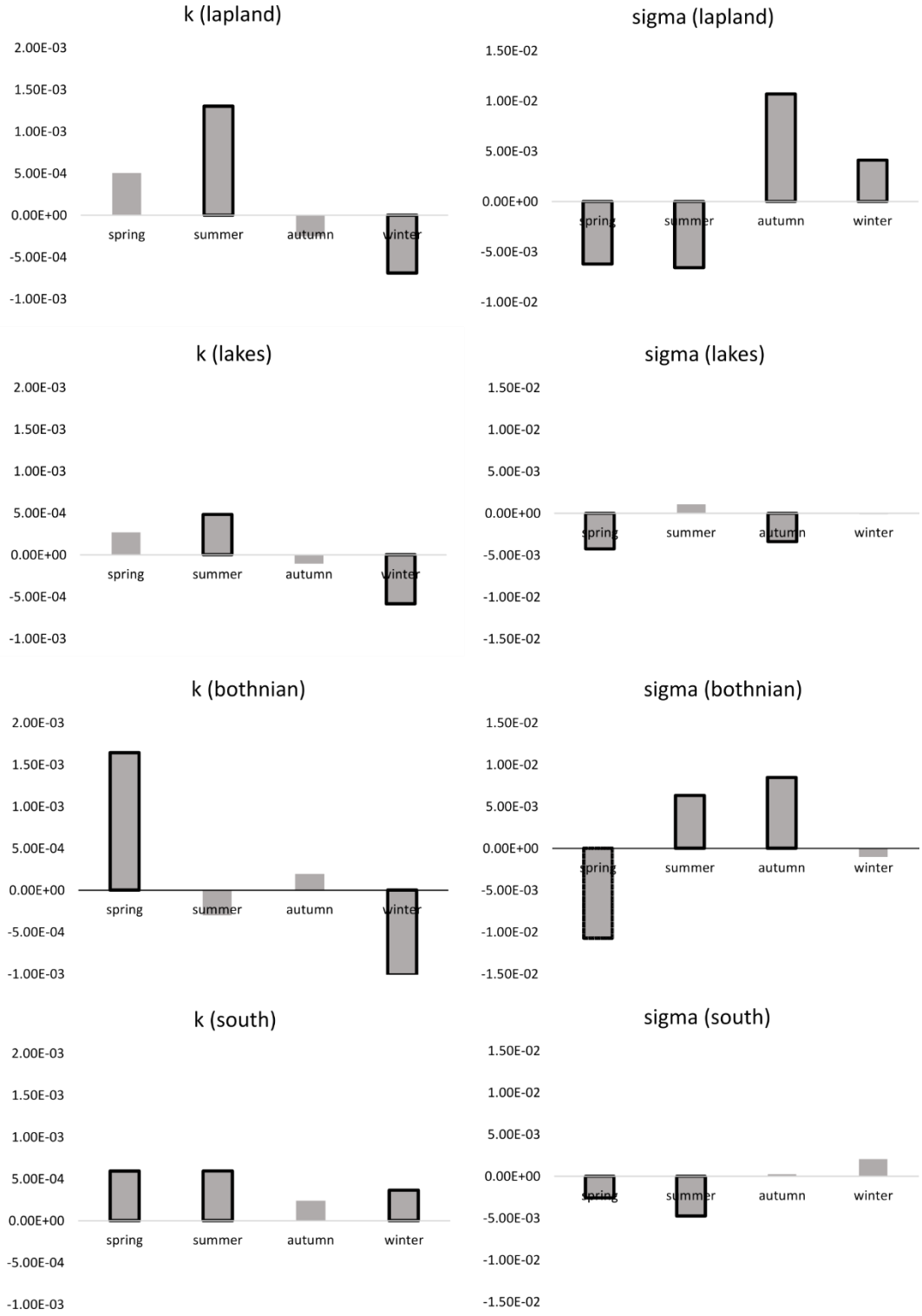

Figure 6 Trends in the best-fitted GP models (shape k and scale σ) calculated using the time-windows-, seasonal- and regional-based analysis of extreme rainfall events exceeding a threshold θ=2 [mm/d] and with time window $\Delta\tau$=5 [years]. The vertical axes are the best-fitted linear regression slopes (m) of each statistic. The data passing the Mann-Kendall test are marked with thicker bar borders.

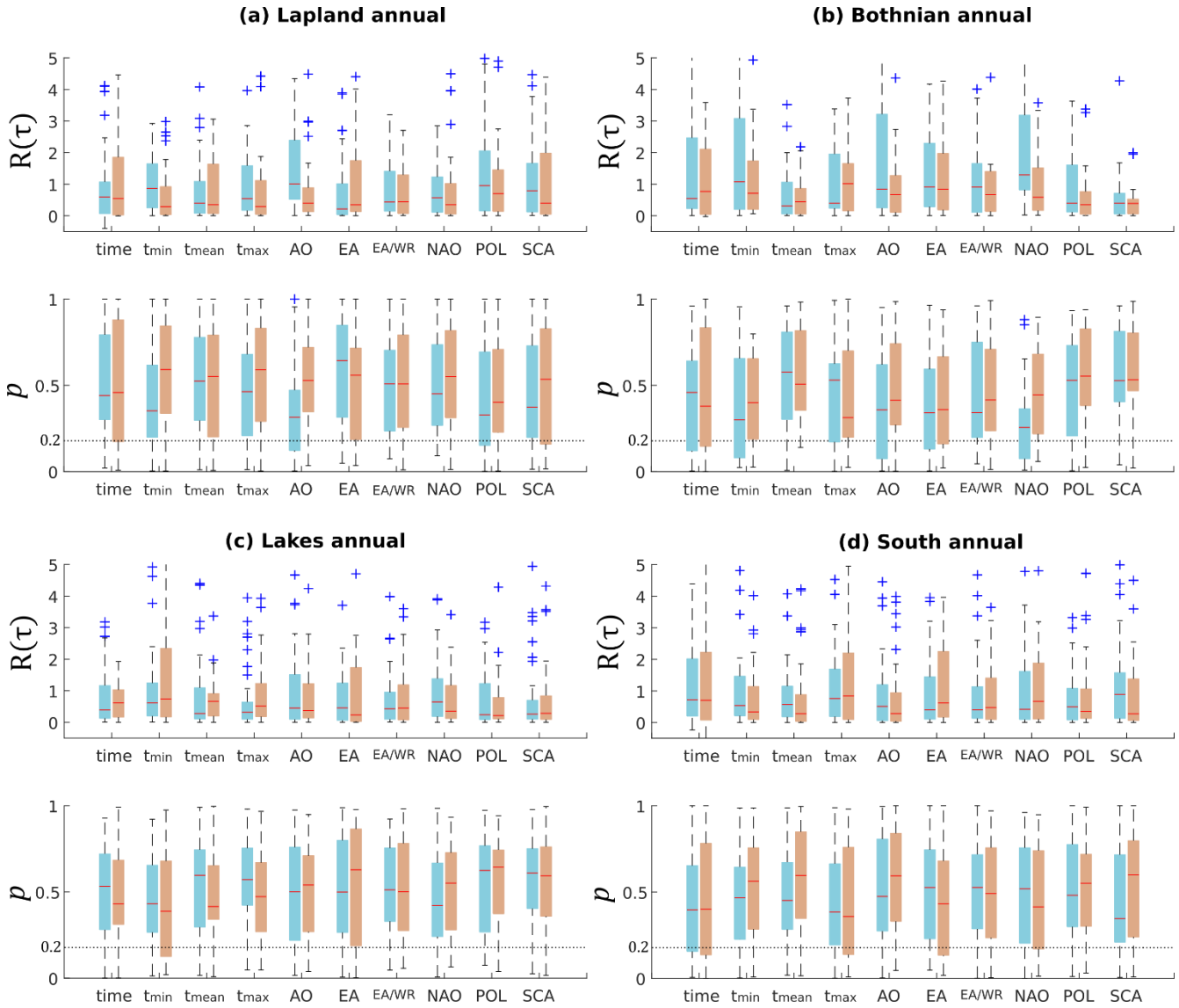

Figure 7 Comparison by region of log-likelihood ratios (R) and p-values for different covariates adopted to evaluate the non-stationarity of rainfall extremes using the Poisson-based approach adopting Generalize Pareto (GP) model. The red lines are the median values, while the crosses are the outliers. The blue boxplots correspond to the shape parameter $k(\tau)$ of the GP model, the brown colors to the scale parameter $\sigma(\tau)$ of the GP model. In the horizontal axes, "time" = days; "t_{min}", "t_{mean}", "t_{max}" = daily temperature measured along with the rainfall depth in each station; "AO"= Arctic Oscillation ACP; "EA" = East Atlantic ACP; "EA/WR"= East Atlantic/West Russia ACP; "NAO"= North-Atlantic Oscillation ACP; "POL"=Polar/Eurasian ACP; SCA=Scandinavia ACP. All covariates except time have been de-trended using best-fitted sin-wave periodicity.

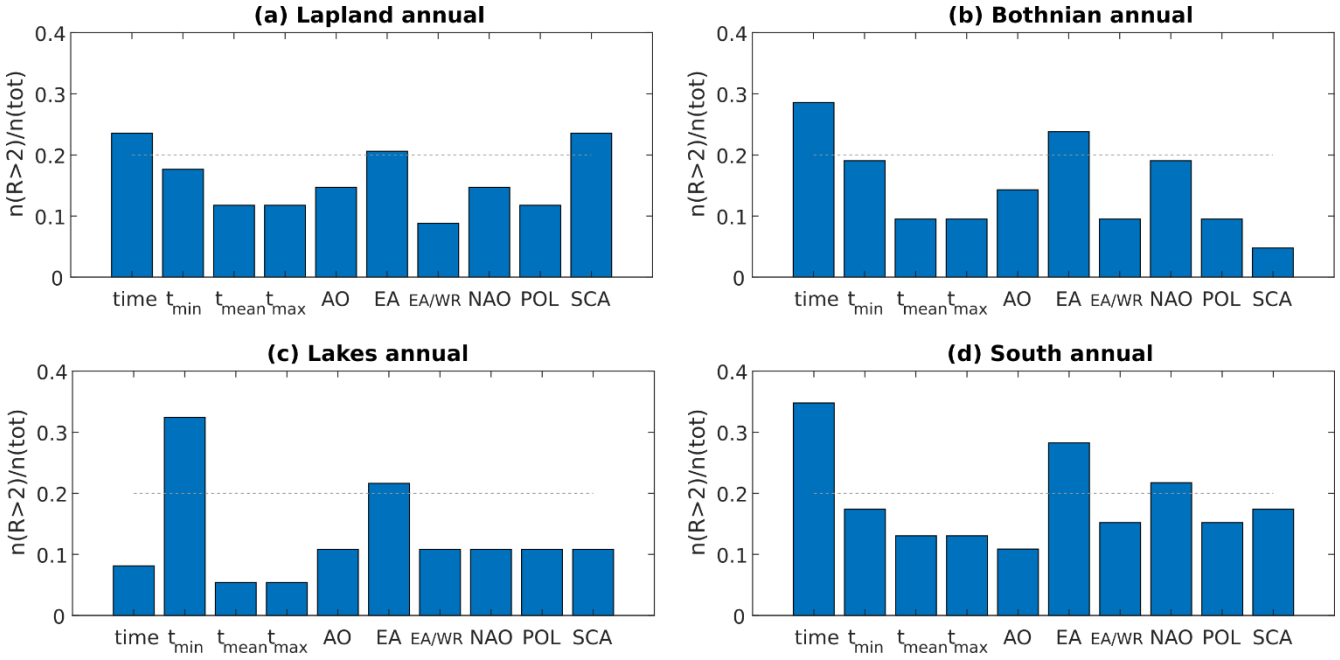

Figure 8 Relative proportion of log-likelihood ratios values passing the threshold R=2, evaluated for the annual values by region. *n(R>2)* = number of passing values; *n(tot)* = total size of the population. In the horizontal axes, "time" = days; "t_{min}", "t_{mean}", "t_{max}" = daily temperature measured along with the rainfall depth in each station; "AO"= Arctic Oscillation ACP; "EA" = East Atlantic ACP; "EA/WR"= East Atlantic/West Russia ACP; "NAO"= North-Atlantic Oscillation ACP; "POL"=Polar/Eurasian ACP; SCA=Scandinavia ACP. All covariates except time have been de-trended using best-fitted sin-wave periodicity.

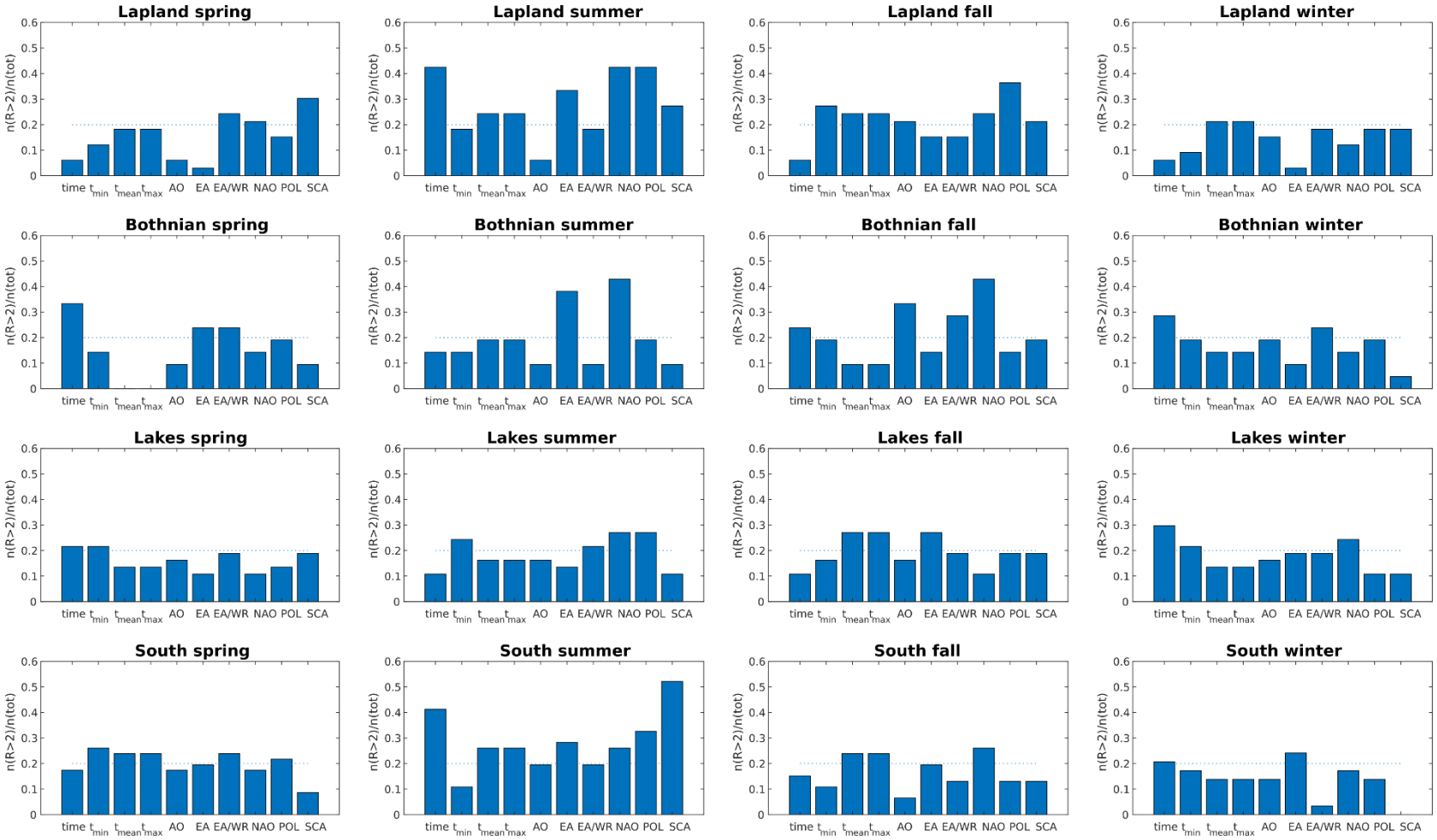

Figure 9 Relative proportion of log-likelihood ratios values passing the threshold $R = 2$, evaluated for the seasonal values by region. $n(R > 2)$=number of passing values; $n(tot)$=total size of the population. In the horizontal axes, "time" = days; "t_{min}", "t_{mean}", "t_{max}" = daily temperature measured along with the rainfall depth in each station; "AO"= Arctic Oscillation ACP; "EA" = East Atlantic ACP; "EA/WR"= East Atlantic/West Russia ACP; "NAO"= North-Atlantic Oscillation ACP; "POL"=Polar/Eurasian ACP; SCA=Scandinavia ACP. All covariates except time have been de-trended using best-fitted sin-wave periodicity.