Non-stationary future return levels for extreme rainfall over Extremadura (southwestern Iberian Peninsula)

Francisco Javier Acero, Sylvie Parey, Thi Thu Huong Hoang, Didier Dacunha-Castelle, José Agustín García, María Cruz Gallego

Non-stationary future return levels for extreme rainfall over Extremadura (southwestern Iberian Peninsula)

Francisco Javier Acero, Departamento de Física, Universidad de Extremadura. Instituto del Agua, Cambio Climático y Sostenibilidad. 06006 Badajoz, Spain (fjacero@unex.es).

Sylvie Parey, EDF/R&D, 6 Quai Watier, 78400 Chatou, France (sylvie.parey@edf.fr).

Thi Thu Huong Hoang, EDF/R&D, 6 Quai Watier, 78400 Chatou, France (thi-thu-huong.hoang@edf.fr).

Didier Dacunha-Castelle, Laboratoire de Mathématiques, Université Paris 11, Orsay, France (didier.dacunha-castelle@math.u-psud.fr).

José Agustín García, Departamento de Física, Universidad de Extremadura Instituto del Agua, Cambio Climático y Sostenibilidad. 06006 Badajoz, Spain (agustin@unex.es).

María Cruz Gallego, Departamento de Física, Universidad de Extremadura. Instituto del Agua, Cambio Climático y Sostenibilidad. 06006 Badajoz, Spain (maricruz@unex.es).

Corresponding author: F.J. Acero, Departamento de Física, Universidad de Extremadura, 06006 Badajoz, Spain. Phone number: +34 924489111. Fax: +34 924289651. (fjacero@unex.es)

Abstract

Based on a previous study for temperature, a new method for the calculation of non-stationary return levels for extreme rainfall is described and applied to a southwestern Spanish Region – Extremadura, using the peaks-over-threshold approach. Both all-days and rainy-days-only datasets were considered and the 20-year return levels expected in 2020 were estimated taking different trends into account: first, for all days, considering a time-dependent threshold and the trend in the scale parameter of the Generalized Pareto Distribution; and second, for rainy days only, considering the role of how the mean, variance, and number of rainy days evolve. Generally, the changes in mean, variance and number of rainy days can explain the observed trends in extremes, and their extrapolation gives more robust estimations. The results point to a decrease of future return levels in 2020 for spring and winter, but an increase for autumn.

Keywords: Extreme Value Theory, Return levels, Extreme Rainfall, Peaks Over Threshold

1. Introduction

The Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) provides likelihood forecasts about climate change, one of which is an increase in the frequency, intensity, and/or amount of heavy rainfall over most land areas (IPCC 2013). Extreme rainfall events cause severe damage to human populations through ecological disasters, destruction of infrastructure, and loss of life. In 2013, a large area in central Europe received high amounts of precipitation. Several places received as much as their normal monthly precipitation within just one or two days. The excess of precipitation resulted in high water levels in such European rivers as the Danube, Elbe, and Rhine. This event has been extensively analysed (http://cib.knmi.nl/mediawiki/index.php/Central_European_flooding_2013).

Extremes are commonly defined as rare return levels estimated using the statistical Extreme Value Theory (EVT). For example, according to the study of Van den Besselaar et al. (2013) for the period 1951-2010, a reduction in the return period of heavy 1-day and 5-day rainfall has been demonstrated from 1951-1970 to 1991-2010 in northern Europe. This decreasing trend in the return period is indicative of increasing precipitation extremes and is coherent with previous extreme precipitation studies covering the whole of Europe (Klein Tank et al. 2003; Alexander et al. 2006).

For the Iberian Peninsula (IP), recent work points to a decreasing trend in the values of extreme rainfall for winter and spring (García et al. 2007; Acero et al. 2011). For autumn, there is an increase in rare extreme precipitation events corresponding to long return periods over the southern and east-central IP (Acero et al. 2011). Other work for the IP has focused on particular regions in Spain. For instance, Beguería et al. (2011) showed for the northeast a significant decrease in extreme rainfall intensity in winter, but an increase in spring.

The present study intends to refine these results to a smaller regional scale using and comparing two approaches to compute near future 20-year return levels (RLs) for daily rainfall. The extrapolation to the near future, namely around year 2020, is based here on the identification and extrapolation of recent observed trends rather than the use of climate model results essentially because practitioners who have to estimate RLs for operational purposes often only have at their disposal observational time-series of the studied variable at the desired location. The limitations of such an approach have, however, to be borne in mind. The generally linear trends identified from observed time-series and extrapolated to the future involve both climate change and inter-annual variability signals present in the observation period considered. These are linked, and thus very difficult to separate, so that extrapolation must be limited to the very near future for which the assumption of the continuation of an unchanged trend may be reasonable. Therefore, if the values obtained are used operationally, then such estimates have to be updated regularly in order to revise any decisions that have been taken.

Following the first papers dealing with non-stationarity in natural extreme events (Parey et al. 2007, 2010) there have been recently a large amount of literature on these subjects, especially concerning hydrology (see Bayazit 2015 for a review). There are still discussions around the pertinence of explicitly taking non-stationarity into account in the estimations (Serinaldi 2015, Serinaldi and Kilsby 2015, Montanari and Koutsoyiannis 2014, Koutsoyiannis and Montanari 2015). Nevertheless, different definitions and estimations of non-stationary return levels or return periods have been proposed and compared (Rootzen and Katz 2013, Cheng et al. 2014, Obeysekera and Salas 2014, Prosdocimi et al. 2014, Read and Vogel 2015, 2016a, 2016b, Silva et al. 2015), the trends being estimated either with time or other physically based covariates. The aim here is not to discuss the precise estimation of a future return level or return period but rather to test another way of modelling non-

stationarity. The 20-year return levels are used as an illustration of the differences, and are just computed as in the stationary context with the values of the extreme value distribution parameters obtained in 2020 with the different extrapolation methods. However, the different trend modeling can be used to compute return levels or return periods according to either of the definitions proposed in the literature. Furthermore, one advantage of the proposed approach based on trends in mean and variance of the whole sample over the more classical use of trends in the extreme value distribution parameters is that these last trends are identified in smaller samples consisting of the extreme events only, and may not reflect some significant trends even though significant evolution may have been detected for the mean and/or variance. Until now, no similar work had been done for precipitation time series through a systematic study of the stationarity test consisting in testing the stationarity of the extremes once trends in mean and variance have been removed. Such a study could be of great interest for the calculation of future extreme events, especially because it allows the use of climate-model projected changes in mean and variance, which are generally more reliable than those of the extremes, as well as the use of other covariates than time.

Following Parey et al. (2010), two approaches will be compared. The first one is the commonly used application of the statistical EVT considering trends in the parameters of the distributions (Friederichs and Hense 2007, 2008; Friederichs 2010; Roth et al. 2012). Secondly, Parey et al. (2010) or Acero et al. (2014) provide another approach to estimating future temperature extremes from the generally stationary extremes of a centred and normalized variable and the changes in mean and variance of the whole dataset. This procedure was shown to take better account of changes in the mean and variance, especially the latter, than the commoner use of the trends in the parameters of the extreme value model. In this paper, the application of this second approach to rainfall time series will be studied and

86 the results compared to the extrapolation of trends in the extreme value distribution
87 parameters.

88 The organization of the paper is as follows. The general methodological framework is
89 described in Sec. 2 and its application to the special case of rainfall time series in section 3.
90 The data that were selected and analysed are described in Sec. 4. The main results are
91 presented and discussed in Sec. 5. The expected future changes in the return levels are then
92 presented in Sec. 6, and the conclusions are drawn in Sec. 7.

93 **2. Methods**

94 In this section, after reminding the conditions of application of EVT and, in particular, the
95 peaks-over-threshold (POT) method chosen here, the two approaches to investigating extreme
96 levels in a non-stationary context will be presented.

97 ***2.1 General framework***

98 We shall apply EVT to weakly dependent (quickly decreasing correlations), time series
99 with seasonality (in any characteristics as mean, variance and extreme parameters) and which
100 can be non-stationary. To handle this, we first tackle seasonalities by working separately for
101 each season. POT is usually applied to independent observations which is not the case here.
102 The exceedences generally occur in clusters, making it necessary to apply a declustering
103 procedure to identify approximately independent cluster maxima. The scheme that we used is
104 known as '*runs declustering*' (Leadbetter et al. 1989). In our work, clusters are separated by at
105 least one day with value below the threshold. Then, for each cluster, the day with the
106 maximum value is chosen, and a series of these C days (where C is the number of clusters) is
107 considered, together with their dates and the intensity of the exceedence. Then to apply POT
108 theory it is necessary to find a good balance between the two approximations made when

using EVT – the probabilistic approximation which needs to define a threshold high enough to approximate the far tail distribution as a Generalized Pareto Distribution (GPD) and the corresponding dates of these excesses as a Poisson process whose intensity is another parameter of the POT method, and the statistical estimation procedure which demands a sufficient number of values to ensure a sufficiently robust GPD and Poisson fit. The GPD has two parameters: σ the scale parameter, and ξ the shape parameter (Coles 2001). Two methods were used to check that the thresholds were reasonable (Coles 2001). First, we studied the stability of the shape parameter when fitting the GPD over a range of thresholds. And second, the mean residual life plots were considered.

It is usual to refer to extreme values in terms of quantiles or return levels. The *N*-year return level Z_N is the level expected to be exceeded once every *N* years in a stationary context. For the POT method, it is expressed, depending on the value of the shape parameter, as

$$\mathrm{Z_N} = \mathrm{u} + \sigma \log(\mathrm{N n_y I_u}) \quad \text{if } \xi = 0 \tag{1}$$

or

$$Z_N = u + \frac{\sigma}{\xi}\left[(N n_y I_u)^{\xi} - 1\right] \text{ if } \xi \neq 0 \tag{2}$$

where n_y is the number of observations per season, and I_u is the Poisson intensity which is the probability of an individual observation exceeding the threshold *u*. A likelihood ratio test is systematically applied to check whether an exponential distribution (ξ=0) can be used.

2.2 Extrapolation of trends in the GPD parameters

The first approach to computing future RLs consists in defining a time-dependent threshold to infer non-stationarity. It is based on a linear quantile regression (Koenker 2005) of the threshold chosen as a high percentile of the distribution and was also used in other works (e.g.

Roth et al. 2012). We test that the dates of exceedences over the threshold after declustering is a stationary Poisson process with a likelihood ratio test at 5% significance level (we tested $I(t)= \lambda$ versus $I(t)= \lambda + \alpha t$ using the likelihood ratio test). The GPD is fitted using the maximum likelihood criterion, and the trend of the GPD scale parameter is tested using the likelihood ratio test with a 5% significance level (Coles 2001). The confidence interval is computed here by bootstrapping in order to take the uncertainty in the trends into account. The bootstrapping procedure is detailed in the Appendix.

In this approach based only on extremes, once linear trends have been identified in the threshold and in the logarithm of the GPD scale parameter, they can be extrapolated to infer those parameters in 2020 and compute the corresponding 20-year RL.

2.3 Extrapolation of trends in mean and variance

Another idea proposed by Parey et al. (2010) is to use trends in the main characteristics of the whole distribution rather than trends in extreme values only. The idea is that in this kind of situation the non-stationarity of extremes is in a statistical framework mainly explained by that of the mean and the variance. To do so, it is necessary to find a simple enough transformation of the whole dataset in order to get a process with stationary extremes. In the following this process is named the residual process and is defined later.

Parey et al. (2010, 2013) and Acero et al. (2014) have shown that, for temperature, stationarity of the extremes can be obtained by removing non-parametric temporal evolutions of the mean and the standard deviation from the original time series. Non-parametric temporal evolutions are chosen in order to capture in the same nonlinear (in general) trend, all the non-stationarity. For instance, in climate, the effect of climate change is often considered as linear at least on short periods and inter-annual variability signals such as the North Atlantic Oscillation (NAO) cannot, in general, be associated with a linear trend. The global

signal is highly non linear. The trend estimation is done by local regression (LOESS) with an optimal smoothing parameter given by a modified partitioned cross-validation procedure (Parey et al. 2013). The stationarity of the extremes of the obtained residuals is then subjected to a test which checks whether the parameters of the extreme value distribution fitted to the extremes of the residuals can be considered constant. The principle of the test can be summarized as follows:

1. Compute a nonparametric trend for the mean of the observed time series $X(t)$ using LOESS: $\hat{m}(t)$

2. Compute the variance as $var(t) = (X(t) - \hat{m}(t))^2$ and its nonparametric trend $\hat{s}^2(t)$ using the same LOESS.

3. Compute $\mathrm{Y}(t) = \frac{X(t) - \hat{m}(t)}{\hat{s}(t)}$, where $Y(t)$ is the residual process.

4. Estimate $\hat{I}(t)$ and $\hat{\sigma}(t)$, the frequency of the Poisson and the scale parameter of the GPD respectively in the two following ways:

    a. As constant in time: $\hat{I}_0$, $\hat{\sigma}_0$

    b. As nonparametric functions of time $\hat{I}(t)$ using Kernel density for the intensity of the Poisson and $\hat{\sigma}(t)$ using cubic splines for the scale parameter of the GPD

    c. And their distances: $\Delta I = \left(\int_{t \in D} \left(\hat{I}(t) - \hat{I}_0\right) dt\right)^{1/2}$, $\Delta\sigma = \left(\int_{t \in D} (\hat{\sigma}(t) - \hat{\sigma}_0)\, dt\right)^{1/2}$, D being the number of days

5. Compute 500 samples of the same number of exceedences with the constant parameters $\hat{I}_0$ and $\hat{\sigma}_0$ and the 500 distances between the parameters estimated as

constant or time varying from these 500 samples. We get an estimate of the distribution of $\Delta\sigma$ and ΔI in the stationary case.

6. Situate the distances ΔI and $\Delta\sigma$ in the distribution of distances previously obtained from a stationary distribution to accept or reject the hypothesis

7. Remark: In the same way, we can do the same simulation choosing $\hat{I}(t)$ and $\hat{\sigma}(t)$ instead of $\hat{I}_0$ and $\hat{\sigma}_0$ and thus we can compute the power of the test of stationarity of extremes.

8. Remark: We cannot get a residual process with stationarity properties if we take linear trends instead of the non-parametric ones.

2.4 Summary

In sum, two different approaches were taken to calculating near future RLs:

- **M1:** A linear threshold is taken, and, as the objective is to study the temporal change in extremes, the GPD parameters are allowed to vary with time according to the following widely accepted trend model: $\xi(t) = \xi$ and $\log \sigma(t) = \sigma_0 + \sigma_1 * t$. Once the trend in $\sigma(t)$ is known (and significant according to a likelihood ratio test at 5%), its linear extrapolation to 2020 is used to calculate the 20-year RLs in that year (Z20-f1).
- **M2:** A residual process is constructed whose extremes can be considered as stationary (a test is applied to check for this). Then, to calculate the 20-year RLs in 2020 (Z20-f2), the daily mean and standard deviation in that year are estimated by linear extrapolation of the linear trends estimated from observations.

Finally, to draw maps of the spatial distribution of trends and return levels, some parameters are spatially interpolated by a kriging procedure.

3. Application to rainfall time series

Precipitation is a complex variable in that it conflates two distinct processes – occurrence (rain or no rain) and rainfall (quantity of rain when it rains). Therefore, when precipitation is averaged for all the days in a season, the result is a mix of the two processes, and a change in mean may be due either to changes in the quantity of rain, a change in the distribution of rainy days, or both. This is why we preferred to consider rainy days only to compute the mean and the variance. For the application of classical POT, this separation does not really matter, because values over the chosen threshold correspond only to rainy days. The only difference is in the computation of the frequency of threshold crossing, estimated as the number of threshold exceedences divided by the total number of days: the total number of days is either the length of the season or the number of rainy days in the season. The two are linked, however. If u is the selected high threshold, n_u the mean seasonal number of events exceeding u, n_y the number of days per season, and n_R the mean number of rainy days per season, then:

$$\frac{n_u}{n_y} = \frac{n_u}{n_R}\frac{n_R}{n_y} \tag{3}$$

Now, to study extreme values with the POT approach using a GPD, it is first necessary to select a threshold u. The threshold chosen was different for the all-days case (98th percentile of the daily rainfall time series) and the rainy-days-only case (95th percentile of the non-zero values of the time series) and for each gauge. The two methods described in the previous section each confirmed that the threshold values for both the all-days and the rainy-days-only precipitation time series could indeed be considered suitable. Finally, it is necessary to ensure independence of the values above the selected threshold. Since many rainfall events over the IP are due to frontal systems, one can expect consecutive days with high rainfall amounts

exceeding the threshold. We apply the declusterization procedure defined in the previous section. For each season, the length of the new series ranged from 40 to 70.

For the second approach, we work with the rainy-days-only time series denoted $X_R(t)$, and we want to define a stationary time series $Y_R(t)$. Then the stationarity of the extremes of $Y_R(t)$ has to be tested using the previously described procedure.

If the stationarity of the extremes of $Y_R(t)$ cannot be rejected, how can the return level be estimated? Indeed, if v is the threshold for Y_R, then $Y_R > v$ is equivalent to $X_R > w$, with $w = s_R * v + m_R$, and m_R and s_R being the mean and standard deviation in the desired time period, respectively. Then

$$\xi_w = \xi_v = \xi$$

$$\sigma_w = \sigma_v * s_R \qquad (4)$$

$$I_{Rw} = I_v$$

where σ_w and I_{Rw} are the corresponding values of the parameters over the threshold w, and σ_v and I_v are those obtained for Y_R over the threshold v. Then, if I_w is the mean frequency of exceedence of threshold w for all days, not only rainy days, one has $n_y * Iw = I_{Rw} * n_R$. Thus the N-year return level Z_N becomes:

$Z_N = w + \frac{\sigma_w}{\xi}\left[\left(N n_y I_w\right)^{\xi} - 1\right]$, with I_w being computed as $I_{Rw} * n_R / n_y$, n_R/n_y being the mean proportion of rainy days and n_R the mean number of rainy days in the desired time period, or

$$Z_N = w + \frac{\sigma_w}{\xi}\left[\left(N n_R I_{Rw}\right)^{\xi} - 1\right] \qquad (5)$$

240 Now, to compute the 20-year RL in 2020, one has to estimate values of the mean and standard
241 deviation in that year. This was done by extrapolation from the observations, using linear
242 regression in order to allow comparison with the first approach. Also, linear regression for
243 the daily values (mean) and the daily standard deviation computed as $(X_R(t)- \hat{m}_R(t))^2$ was
244 used to calculate the observed trends in mean and standard deviation for rainy days. In
245 practice, the future mean and standard deviation are estimated for the year 2020 from the
246 linear extrapolations. Thus, if m_{Rf} and s_{Rf} are these estimated mean and standard deviation
247 values in the future period, then $w=s_{Rf}*v+m_{Rf}$ as was stated above. Lastly, a linear trend is
248 fitted to the number of rainy days and extrapolated to 2020 to obtain the future expected
249 number of rainy days. The RLs are then calculated using these estimated future quantities,
250 and the corresponding confidence intervals are constructed by the bootstrap procedure
251 described in Parey et al. (2010) (see the Appendix).

252 **4. Data**

253 The study area was the Extremadura Region, in the southwest of the IP (Fig. 1), with a total
254 area of 41 635 km². There is a contrasting orography: the extensive areas of the Rivers Tagus
255 and Guadiana depressions have altitudes under 400 m a.s.l, while the Region's highest peak is
256 over 2400 m a.s.l. Three main mountain ranges in Extremadura lead to a complex distribution
257 of rainfall over the study area.

258 To detect trends in time series of extreme values requires highly reliable data. Reliability
259 is usually achieved by carefully selecting time series which involve no changes in location,
260 instrument type, or measuring procedure. In the present study, an additional criterion was that
261 only records with no missing data were selected.

262 The time series were taken from an extensive database of daily rainfall time series
263 provided by the Spanish National Meteorology Agency (AEMET). The set of series had to

264 cover the orographic diversity of the Extremadura Region, leaving no large areas without
265 coverage because the distribution of rainfall over the Region is very complex due mainly to
266 the effects of altitude.

267 The final choice was a set of 72 homogeneous daily rainfall time series corresponding to
268 gauges as regularly spaced as possible over Extremadura. Their locations are shown in Fig. 1.
269 The study period was 1961 to 2010. There are no gauges in the mountainous areas of
270 Extremadura due to the absence of population and difficult accessibility. The altitudes of the
271 chosen gauges range from 185 to 796 m a.s.l.

272 Data homogeneity was checked using the R-based program RHTestV3, developed at the
273 Climate Research Branch of the Meteorological Service of Canada, and available from the
274 ETCCDMI Website (http://etccdi.pacificclimate.org/soft). This program is capable of
275 identifying multiple step changes at documented or undocumented change-points. It is based
276 on a two-phase regression model with a common linear trend (Wang 2003). Homogeneity
277 was tested on the monthly time series. This analysis, together with the metadata of the
278 stations, showed that none of the 72 time series had change-points significant at 5%, with all
279 of them being homogeneous in the cited period of study.

280 For this study of precipitation extremes over Extremadura, in view of the highly seasonal
281 nature of the IP rainfall, each season was studied separately. The working definition of the
282 seasons was: winter – December, January, and February; spring – March, April, and May; and
283 autumn – September, October, and November. The summer months were not considered due
284 to the lack of sufficient rainy days in most parts of Extremadura. As an illustration, Figure 2
285 shows the spatial distribution of the threshold u (the 95th percentile of rainy day rainfall
286 amounts) for the precipitation over Extremadura. The patterns of the thresholds for the three
287 seasons considered are quite similar, with higher values in the north and northeast

corresponding to the more mountainous areas, and lower in the mid-south, corresponding to the least rainy area. Autumn is the season with the highest values of the threshold, closely followed by winter. The lowest values are for spring. There is thus a clear difference in the rainfall distributions for each season, justifying the separation.

5. Results

This section presents the main results of calculating the 20-year RLs in 2020. First, we shall deal with the preliminary results for the shape parameter in a stationary context. Its value is zero for most of the observatories for all the seasons considered according to the likelihood ratio test at a 95% confidence level.

5.1. Stationarity test

In order to check the hypothesis that the non-parametric temporal evolutions are essentially linked to the evolutions of the mean and variance, the previously described methodological approach was used to test for the stationarity of the extremes of the standardized residuals computed from the rainy-day time series. Figure 3 shows the distribution of the gauges that verified this stationarity at a 90% confidence level either totally or partially.

The stationarity test was quite well satisfied for all three seasons considered. In autumn, 68 (94%) gauges satisfied the test for the scale parameter and the Poisson intensity, while 4 satisfied it only for the Poisson intensity. In winter, 62 (86%) gauges satisfied the test for the scale parameter and the Poisson intensity, and 10 satisfied it only for the Poisson intensity. Finally, in spring, 70 (97%) gauges satisfied the test for the scale parameter and the Poisson intensity, 1 gauge satisfied it for the Poisson intensity only, and 1 for the scale parameter only. When stationarity was rejected for the scale parameter, we systematically tested for a linear trend in the scale parameter of the standardized variable using the likelihood ratio test as done

311 in method M1. In each case, we found that the remaining non-parametric temporal evolution
312 did not represent a significant increasing or decreasing linear trend. This means that
313 interannual variability was the main factor leading to rejection of stationarity.

314 ***5.2. 20-year return levels in 2020***

315 *5.2.1. Trends in mean, variance, number of rainy days, and POT parameters*

316 To present the values and the significance of the trends, Figure 4 shows for each season
317 considered the spatial distribution of the linear trends in the number of rainy days and in the
318 mean and standard deviation of the rainfall on those days. Blue means negative and red
319 positive. Black triangles mean a trend that is significant at the 10% level, and open triangles
320 mean the trend is not significant (at the 10% significance level according to a Mann-Kendall
321 test). Upward pointing triangles mean positive, and downward negative. The size of the
322 triangles represents the value of the trend for each gauge. Table 1 summarizes the results.

323 In autumn, the mean rainfall shows a decreasing trend for Extremadura as a whole, with 49
324 negative trends, 22 of them being significant. Although there are 23 positive trends, they are
325 not appreciable in the figure because their values are too low. The trend in standard deviation
326 is clearly positive over most of Extremadura. The number of significant trends of either sign
327 in this statistical moment is low however – only 7 of the 50 positive trends and 2 of the 22
328 negative trends. Finally, the number of rainy days shows a positive behaviour, with 93% of
329 the observatories having a positive trend, 29 of them being significant. The impact of these
330 opposite behaviours on the estimated future return level is of interest.

331 In winter, mean and variance both show negative trends for the whole area studied. For the
332 mean, there are 48 significant negative trends of the 68 total, and, for the variance, there are
333 37 significant negative trends of the 60 total. None of the low number of positive trends is

significant for either mean or variance. The number of rainy days shows a different pattern in winter, with mainly positive trends for a great part of Extremadura, and 89% of these trends being significant.

Spring shows a behaviour similar to that of winter for the three variables. The mean and variance trends are mainly negative for Extremadura as a whole, except in the northwest where there are positive trends in the variance. The number of rainy days shows a mainly positive behaviour with 16 of the 50 observatories showing a significant positive trend.

Figure 5 shows the spatial distribution of the trends in the time-varying threshold used in method M1 and in the GPD scale parameter σ for each season considered. One sees that the threshold trends, and to a lesser extent σ, appear closer to the trends in standard deviation than to those in the mean. In particular, it seems that where the standard deviation trends are largest there is an additional trend in σ.

5.2.2. Comparison of the 20-year RLs obtained for 2020

The following step is to compare the future RLs in the year 2020 obtained for all days by extrapolating the trends in the threshold and the GPD scale parameter using method M1 (Z20-f1) with the 2020 RLs obtained by extrapolating the linear trends in the daily mean and standard deviation of the amount of rain of rainy days and the number of rainy days using method M2 (Z20-f2). Figure 6 shows the spatial distribution of these two sets of future 20-year RLs in 2020 for each season considered. One observes in the figure that there are especially differences in winter and spring. Figure 7 shows the spatial distribution of the observatories with Z20-f2 inside the Z20-f1 CI (which was here estimated by bootstrapping to take the uncertainty of the trend into account). First, it has to be mentioned that the Z20-f1 CIs are larger than those of Z20-f2, as shown in Figure 8 for the three seasons considered. This clearly illustrates the greater uncertainty in the estimation of trends from a smaller

sample in method M1. Then, overall, for by far the greater part of the observatories, Z20-f2 lies inside the Z20-f1 CI for all three seasons considered – all of the observatories in autumn, 69 in winter, and 71 in spring. Thus, although Z20-f2 is generally lower than Z20-f1, when the uncertainties are taken into account and the CIs overlap, the two values are equally possible.

Looking in greater depth at the exceptions, one finds two different cases concerning the trends:

- Trends in mean and variance are identified, but not in the scale parameter. In these cases, although there might be a trend in the threshold, it is very low. This is the case for one gauge in spring for which σ is constant but there is a significant decreasing trend in the mean, leading to Z20-f2 being significantly lower than Z20-f1.
- A trend is identified in the scale parameter but not in the mean and variance. For all such cases, the trend in the scale parameter is found to be very sensitive to slight changes in the threshold, leading to different results for the RL. For the San Vicente de Alcántara gauge in particular (westernmost red point in Fig. 7-winter), with the 98th percentile as threshold, one finds a significant trend in the scale parameter, and Z20-f1=111.11[75.88;151.35]. But, with the 98.5th percentile as threshold, one finds no significant trend in the scale parameter, and Z20-f1=78.35[48.73;107.97].

Finally, there are very special cases for which neither method seems to be well adapted. This may be so when there is an isolated maximum much greater than the rest of the exceedences, or there is a high frequency of exceedences located in just a short section of the whole time series. Figure 9 illustrates this behaviour for Torrejoncillo in winter (northern red point in Fig. 7-winter). From top to bottom, it shows the exceedences distribution, the temporal evolutions of the scale parameter, and a set of figures with the temporal evolutions and linear trends of mean, standard deviation, and number of rainy days. The non-parametric

evolution of the scale parameter is estimated using cubic splines with a smoothing parameter obtained by cross-validation since the extremes are independent, as stated before. As can be seen, the exceedences are mainly located in the centre of the time series where they are the highest in value. At the beginning of the period, the values are lower, and there are only a few exceedences at the end of the period. In view of this distribution, one can understand the second plot which shows the temporal evolution of the scale parameter. Nevertheless, the red line shows the great rise in the trend of the scale parameter – recall that the trend is linear for $\log(\sigma)$ – implied by the need for a parametric form seems to make no sense, or at least seems exaggerated such an important increasing trend. Moreover, considering the bottom three plots, one observes that, although the mean and standard deviation increase at the same time as does the scale parameter, their trends are smoother, so that this approach is unable to take the isolated maximum of this case into account properly. This leads to different results for the 20-year RL: Z20-f1=169.31[94.00;335.74] and Z20-f2=79.30[66.31;98.92], with the latter no longer lying within the CI of the former. In this case, it seems that all extremes do not have the same distribution, some being clearly out of the range of the others. Such cases are really difficult to handle with classical EVT.

6. Expected Changes in Return Levels

In view of the comparative analysis above, the new approach using an extrapolation of the linear trends in mean, standard deviation, and number of rainy days to compute the 20-year RLs in 2020 seems at least coherent, and even better suited to some cases than the first method, with smaller confidence intervals. It was therefore applied to study the possible changes in future RLs relative to the present values.

Figure 10 shows the spatial distribution of the 20-year RLs in 2020 obtained with method M2 (Z20-f2) according to whether they lie or do not lie inside the CI of the present 20-year

RLs obtained with the same method (Z20-p2). In general terms, for the three seasons considered, there are more gauges with Z20-f2 outside the Z20-p2 CI – in particular, 67% for autumn, 76% for winter, and 72% for spring. For these gauges, the direction of the change is different in autumn from that in spring and winter. In autumn, there are 31 gauges with 20-year RLs in 2020 higher than those of the present, while in winter and spring there are 47 and 44 gauges, respectively, with future RLs lower than those of the present.

Figure 11 shows the spatial distributions of the 20-year RLs. The present RLs are on the left, the future ones in the centre, and the differences between the two on the right, with blue (red) meaning decreasing (increasing) values of the 20-year RLs in 2020. The main decrease in the 20-year RLs is in winter for the overall study area. The gauges with a significant decreasing trend in variance systematically show a decrease in RL. This is reflected in a major decrease in the extreme rainfall events in winter. Spring also shows a general decrease for the study area as a whole, but less than in winter, and more notable high values in the north. The behaviour in autumn is different. Much of the Extremadura Region shows an increase in the future RLs, particularly notable in the mid-north where the increase is greatest. The areas of increase closely match those of increases in the standard deviation shown in Fig. 4. This underlines the role that changes in variance play in changes in extremes. There is a decrease in the east of the Region, as well as a slight decrease in the northeast. In sum, if the observed trends actually continue linearly in the future, and considering the maps in the centre for the future RLs in the three seasons considered, autumn will clearly become the season with the greatest extreme rainfall events because of the major decrease in winter.

7. Conclusions

We have described an EVT study calculating non-stationary RLs of extreme rainfall in 2020 for Extremadura (SW Spain) using a set of complete daily rainfall time series from 72 gauges

for the period 1961-2010. For each time series, the exceedences over either a fixed or time-varying threshold *u* were subjected to a 'runs declustering' procedure. The resulting extreme rainfall data time series were then fitted with a flexible GPD, and the occurrence dates with a Poisson process in order to calculate the RLs.

Two approaches to computing future rainfall RLs with POT were studied. In the first, trends in the extremes considering all the days were identified, taking into account a time-varying threshold based on a linear quantile regression and, when appropriate, a trend in the GPD scale parameter. Then, in the second, we calculated the RLs considering only the rainy days, examining the impact of evolutions of the mean and variance and of the number of rainy days. In this second case, we applied a novel adaptation of a stationarity test to rainfall that had been designed and used for temperature time series, finding that it was indeed satisfied for the majority of the gauges for all three seasons considered.

The principal objective of the present work was to compare estimates of the 20-year RLs expected in 2020 using the aforementioned two methods. The main conclusions that we can draw are:

- Generally, the two approaches give comparable results for the future RLs, but there are some exceptions. These are mainly due to the sensitivity to the threshold of the identification of the trend in the scale parameter, and may sometimes lead to unrealistic results. The use of the mean and variance constitutes a more robust approach when the identification of a trend in the GPD scale parameter is difficult and very sensitive to the threshold choice. It also leads to reduced CIs.
- There are special cases for which both approaches seem to fail. They give different values for the future RLs, but probably neither of them is reliable.
- The future evolution of the RLs varies from season to season. There are decreases in winter and spring, and increases in autumn. The evolution of the variance was seen to

play a major role in the estimation of the extremes since the increases in autumn closely matched the increases in the variance. There was relatively little evolution in the number of rainy days, and it had the correspondingly least impact. These results showed a decrease in extreme rainfall events in the near future (2020), with the central-eastern part of the Extremadura Region showing the greatest decrease for both winter and spring. In autumn however, the increasing trends in both mean and variance lead to the opposite behaviour – an increase in extreme rainfall events, with a wide area showing changes in the 20-year RLs in 2020 that range from +5 to +15 mm.

The present results for the RLs in Extremadura are consistent with previous findings covering Spain showing a decrease in spring and winter (Goodess and Jones 2002; Rodrigo and Trigo 2007; García et al. 2007; Acero et al. 2011) and an increase in autumn (García et al. 2007; Acero et al. 2011).

The negative trends in this southwestern part of the IP could be related to the prevailing positive phase of the North Atlantic Oscillation (NAO) during the last few decades. Since this oscillation is known to influence rainfall over the southwest of the IP (Rodríguez-Puebla et al. 1998; García et al. 2002; Trigo et al. 2004), it could be the cause of the decreasing winter rainfall for this region. As mentioned above, extrapolating observed trends does not allow the two signals to be separated, and complementary analyses with the aid of climate simulations might be necessary to better understand the impact of climate change on extreme rainfall in the region.

The results for spring and autumn are in agreement with those reported by Fernández-Montes et al. (2014) for the relationship between extreme precipitation days and circulation types. They find a decrease in extreme precipitation days in the west of the Iberian Peninsula mainly due to a decrease in the frequency of cyclonic southwesterly flow. But in autumn,

extreme precipitation becomes more frequent (as in the present study) due to the northwesterly flow.

It was important to carry out this type of study for a small region and to consider different seasons so as to better understand the possible evolution of extreme rainfall events. Also, the procedure that was newly tested in this work was found to be reasonable for the estimation of future extremes, opening up the possibility of using the evolution of mean and variance as projected by climate models to anticipate possible changes in a more distant future.

There are two situations for which the application of EVT does not seem appropriate: first, for the summer season in the study area because it includes very few rainfall events, and second, when a time series presents exceedence values or exceedence frequencies well above (or below) the other values. These cases will be further analysed in future work in order to investigate other possible ways of inferring rare levels in such cases

Acknowledgements

Thanks are due to the Spanish Weather Service (Agencia Estatal de Meteorología: www.aemet.es) for providing the daily rainfall time series used in this study.

Funding

This work was partially supported by the Gobierno de Extremadura – FEDER Funds (IB13049) and Junta de Extremadura-FEDER Funds (GR15137).

References

Acero, F. J., García J. A., Gallego, M.C., 2011. Peaks-over-threshold study of trends in extreme rainfall over the Iberian Peninsula. *Journal of Climate*, 24, 1089–1105.

501 Acero, F. J., García, J.A., Gallego, M.C., Parey, S., Dacunha-Castelle, D., 2014. Trends in
502 summer extreme temperatures over the Iberian Peninsula using nonurban station data.
503 *Journal of Geophysical Research - Atmosphere*, 119, 39–53, doi:10.1002/2013JD020590

504 Alexander, L. V., et al. 2006. Global observed changes in daily climate extremes of
505 temperature and precipitation. *Journal of Geophysical Research* ,111, D05109,
506 doi:10.1029/2005JD006290.

507 Bayazit, M., 2015. Nonstationarity of hydrological records and recent trends in trend analysis:
508 a state-of-the-art review. *Environmental Processes*, 2, 527-542.doi: 10.1007/s40710-015-
509 0081-7

510 Beguería, S., Angulo-Martínez, M., Vicente-Serrano, S.M., López-Moreno, J.I., El-Kenawy
511 A., 2011. Assessing trends in extreme precipitation events intensity and magnitude using
512 non-stationary peaks-over-threshold analysis: a case study in northeast Spain from 1930
513 to 2006. *International Journal of Climatology*, 31, 2102–2114. doi: 10.1002/joc.2218

514 Cheng, L., AghaKouchak, A., Gilleland, E., Katz, R. W., 2014. Non-stationary extreme value
515 analysis in a changing climate. *Climatic Change*, 127, 353-369

516 Coles, S. (2001), An Introduction to Statistical Modelling of Extreme Values, 208 pp.,
517 Springer-Verlag, London.

518 Fernández-Montes, S., Seubert, S., Rodrigo, F.S., Rasilla Álvarez, S.F., Hertig, E., Esteban,
519 P., Philip, A., 2014. Circulation types and extreme precipitation days in the Iberian
520 Peninsula in the transition seasons: Spatial links and temporal changes. *Atmospheric*
521 *Research*, 138, 41-58.

522 Friederichs, P., Hense, A., 2007. Statistical downscaling of extreme precipitation events using
523 censored quantile regression. *Monthly Weather Review*, 135, 2365–2378

524 Friederichs, P., Hense, A., 2008. A probabilistic forecast approach for daily precipitation
525 totals. *Weather Forecast*, 23, 659–673.

526 Friederichs, P., 2010. Statistical downscaling of extreme precipitation events using extreme
527 value theory. *Extremes* 13, 109-132.

528 García, J. A., Serrano, A., Gallego, M.C., 2002. A spectral analysis of Iberian Peninsula
529 monthly rainfall. *Theoretical and Applied Climatology*., 71, 77–95.

530 García, J. A., Gallego, M.C., Serrano, A., Vaquero, J.M., 2007. Trends in block-seasonal
531 extreme rainfall over the Iberian Peninsula in the second half of the twentieth century.
532 *Journal of Climate*, 20, 113–120.

533 Goodess, C. M., Jones, P.D, 2002: Links between circulation and changes in the
534 characteristics of Iberian rainfall. *International Journal of Climatology*, 22, 1593–1615.

535 IPCC (Intergovernmental Panel on Climate Change), 2013: Climate Change 2013: The
536 Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report
537 of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner,
538 M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)].
539 Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA,
540 1535 pp, doi:10.1017/CBO9781107415324.

541 Klein Tank, A. M. G., Können, G.P., 2003. Trends in indices of daily temperature and
542 precipitation extremes in Europe, 1946–99. *Journal of Climate*, 16, 3665–3680.

543 Koenker, R., (2005) Quantile Regression. Econometric Society Monographs, vol. 38.
544 Cambridge University Press, Cambridge.

545 Koutsoyiannis, D., Montanari, A., 2015. Negligent killing of scientific concepts: the
546 stationarity case. *Hydrological Sciences Journal*, 60(7-8), 1174-1183.

547 Leadbetter, M. R., Weissman, I., Haan, L.D., Rootzen, H., 1989. On clustering of high values
548 in statistically stationary series, Tech. Rep. 253, Centre for Stochastic Processes,
549 University of North Carolina, Chapel Hill.

550 Montanari, A., Koutsoyiannis, D., 2014. Modeling and mitigating natural hazards:
551 Stationarity is immortal!. *Water Resources Research*, 50, 9748-9756.
552 doi:10.1002/2014WR016092.

553 Obeysekera, J., Salas, J. D., 2014. Quantifying the uncertainty of design floods under
554 nonstationary conditions. *Journal of Hydrologic Engineering*, 19(7), 1438-1446.

555 Parey, S., Malek, F., Laurent, C., Dacunha-Castelle, D., 2007, Trends and climate evolutions:
556 statistical approach for very high temperatures in France. *Climatic Change*, 81, 331–352.

557 Parey, S., Hoang, R.T.H., Dacunha-Castelle, D., 2010. Different ways to compute temperature
558 return levels in the climate change context. *Environmetrics*, 21, 698–718.

559 Parey, S., Hoang, R.T.H., Dacunha-Castelle, D., 2013. The importance of mean and variance
560 in predicting changes in temperature extremes. *Journal of Geophysical Research*, 118,
561 8285–8296, doi:10.1002/jgrd.50629

562 Prosdocimi, I., Kjeldsen, T. R., Svensson, C., 2014. Non-stationarity in annual and seasonal
563 series of peak flow and precipitation in the UK. *Natural Hazards and Earth System*
564 *Sciences*, 14, 1125–1144. doi:10.5194/nhess-14-1125-2014

565 Read, L. K., Vogel, R. M., 2015. Reliability, return periods, and risk under nonstationarity.
566 *Water Resources Research*, 51, 6381-6398. doi:10.1002/2015WR017089.

567 Read, L. K., Vogel, R. M., 2016a. Hazard function theory for nonstationary natural hazards.
568 *Natural Hazards and Earth System Sciences*,16, 915–925. doi:10.5194/nhess-16-915-
569 2016

570 Read, L. K., Vogel, R. M. 2016b. Hazard function analysis for flood planning under
571 nonstationarity. *Water Resources Research*, 52, 4116–4131.doi:10.1002/2015WR018370

572 Rodrigo, F. S., Trigo, R.M., 2007. Trends in daily rainfall in the Iberian Peninsula from 1951
573 to 2002. *International Journal of Climatology*, 27, 513–529.

574 Rodríguez-Puebla, C., Encinas, A.H., Nieto, S., Alonso, S., 1998. A 30-year (1964–1994)
575 daily rainfall data base for the Spanish Mediterranean regions. *International Journal of*
576 *Climatology*, 18, 541–560.

577 Rootzén, H., Katz, R. W., 2013. Design life level: quantifying risk in a changing climate.
578 *Water Resources Research*, 49, 5964-5972. doi:10.1002/wrcr.20425

579 Roth, M., Buishand, T.A., Jongbloed, G., Klein Tank, A.M.G., van Zanten, J.H., 2012. A
580 regional peaks-over-threshold model in a nonstationary climate, *Water Resources*
581 *Research*, 48, W11533, doi:10.1029/2012WR012214.

582 Serinaldi, F., 2015. Dismissing return periods!. *Stochastic Environmental Research and Risk*
583 *Assessment*, 29, 1179–1189. doi:10.1007/s00477-014-0916-1

584 Serinaldi, F., Kilsby, C. G., 2015. Stationarity is undead: Uncertainty dominates the
585 distribution of extremes. *Advances in Water Resources*, 77, 17-36.

586 Silva, A. T., Naghettini, M., Portela, M. M., 2016. On some aspects of peaks-over-threshold
587 modeling of floods under nonstationarity using climate covariates. *Stochastic*
588 *Environmental Research and Risk Assessment*, 30, 207-224. doi: 10.1007/s00477-015-
589 1072-y

590 Trigo, R.M, Pozo-Vázquez, D., Osborn, T.J., Castro-Diez, Y., Gamiz-Fortis, S., Esteban-
591 Parra, M.J., 2004. North Atlantic Oscillation influence on precipitation, river flow and
592 water resources in the Iberian Peninsula. *International Journal of Climatology*, 24: 925–
593 944, DOI: 10.1002/joc.1048.

van den Besselaar, E. J. M., Klein Tank, A.M.G., Buishand, T.A., 2013. Trends in European precipitation extremes over 1951–2010. *International Journal of Climatology*, 33: 2682–2689. doi: 10.1002/joc.3619

Wang, X. L., 2003. Detection of undocumented changepoints: A revision on the two-phase regression model. *Journal of Climate*, 16, 3383–3385.

Appendix: Bootstrap procedures for confidence intervals

Trends in the parameters of the GPD:

A distribution of 20-year RLs in 2020 is computed by executing the following steps 500 times:

- Simulate a random number of days with rainfall over the linear threshold *u(t)* by a Poisson process with the observed intensity.
- Compute the corresponding rainfall amounts for these exceedence dates as $X(t) = u(t) + \sigma(t)\varepsilon(t)$, with $\varepsilon(t)$ being the residuals of the observed exceedences, determined by sampling among the $\varepsilon(t)$ with replacement.
- Identify trends in the scale parameter of this new sample.
- Extrapolate the trends to derive a future 20-year RL.

Trends in mean, variance, and number of rainy days:

Again, a distribution of the future 20-year RLs is constructed by 500 re-samplings in the following steps:

- Simulate mean seasonal numbers of rainy days as the observed numbers + a randomly chosen shift from the linear trend.
- Compute the corresponding rainfall amounts as $\tilde{Y}_R \hat{s}_R + \hat{m}_R$ where $\tilde{Y}_R$ is obtained by block bootstrapping Y_R values (with a block of length 10, to manage temporal dependency).

- Compute the 20-year RL of $\tilde{Y}_R$ and the future mean, variance, and number of rainy days by extrapolating linear trends fitted to the new sample.
- Derive the future 20-year RL.

======= TABLE 1 =======

Table 1. For the three seasons considered, the number of positive or negative trends in the mean, variance, and number of rainy days, with the number of significant trends of each sign in parentheses.

	Mean		Variance		Number of rainy days	
	+	-	+	-	+	-
Autumn	23 (4)	49 (22)	50 (7)	22 (2)	67 (29)	5 (0)
Winter	4 (0)	68 (48)	12 (0)	60 (37)	46 (41)	26 (7)
Spring	10 (1)	62 (38)	20 (1)	52 (11)	50 (16)	22 (2)

======= FIGURE CAPTIONS =======

Figure 1. Location of the study area (Extremadura) in the Iberian Peninsula, and the spatial distribution of the gauges used with the correspondent altitude showed in the scale.

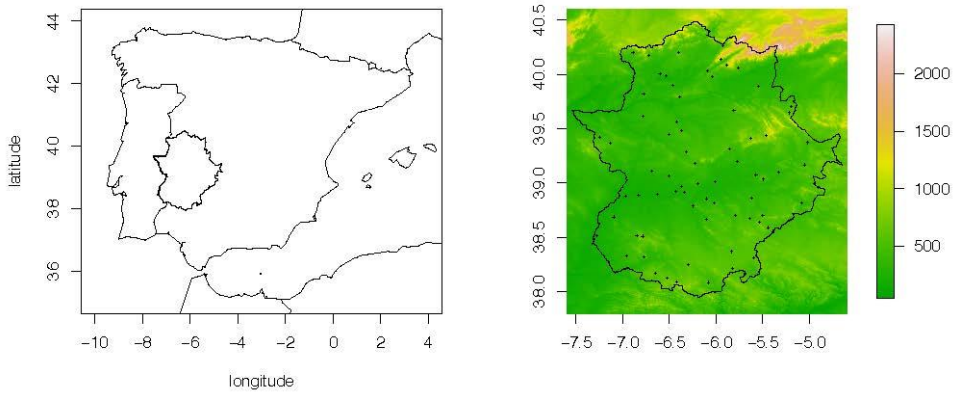

Figure 2. Spatial distribution of the 95th percentile of rainy-day rainfall amounts (mm) used as threshold.

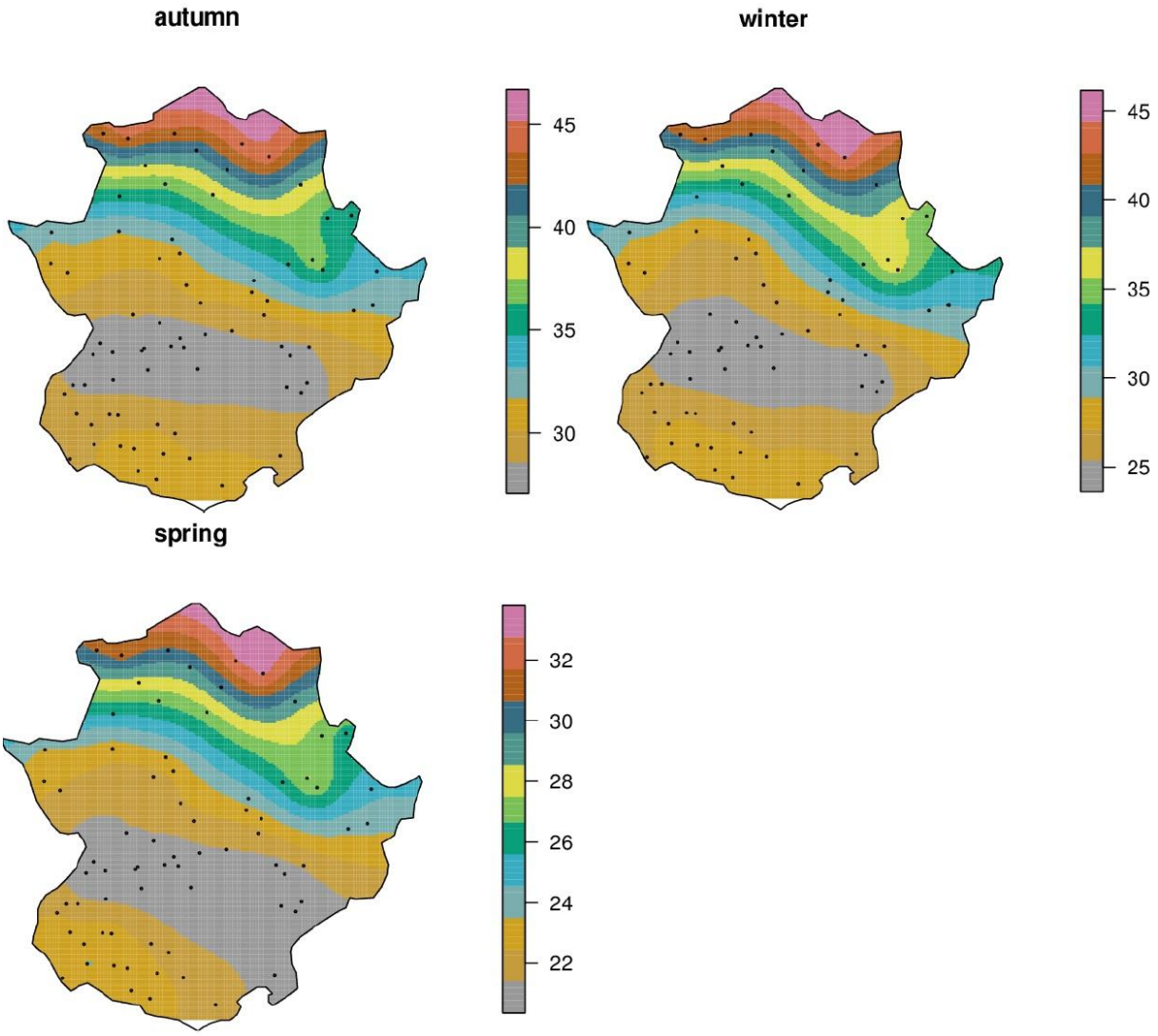

Figure 3. Spatial distribution of the observatories that satisfy the stationarity of the extremes of the residuals computed from the rainy-day time series.

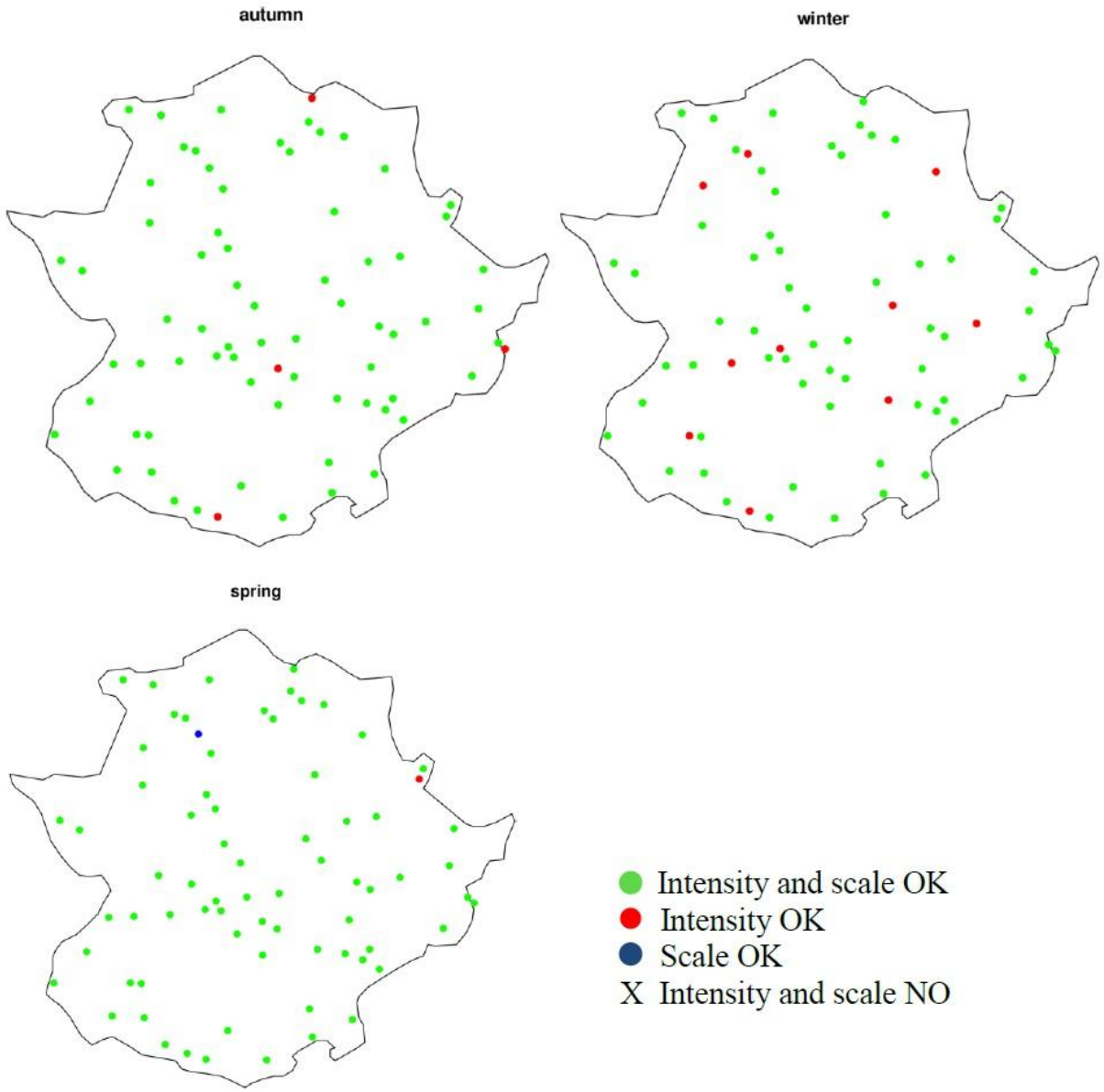

Figure 4. Spatial distribution of the linear trends in mean, standard deviation, and number of rainy days for the three seasons considered. Solid triangles mean a trend significant at the 10% level. Open triangles mean a non-significant trend (according to a Mann-Kendall test at 10%). Upward triangles mean a positive trend, and downward negative. The size of the triangles represents the value of the trend for each gauge.

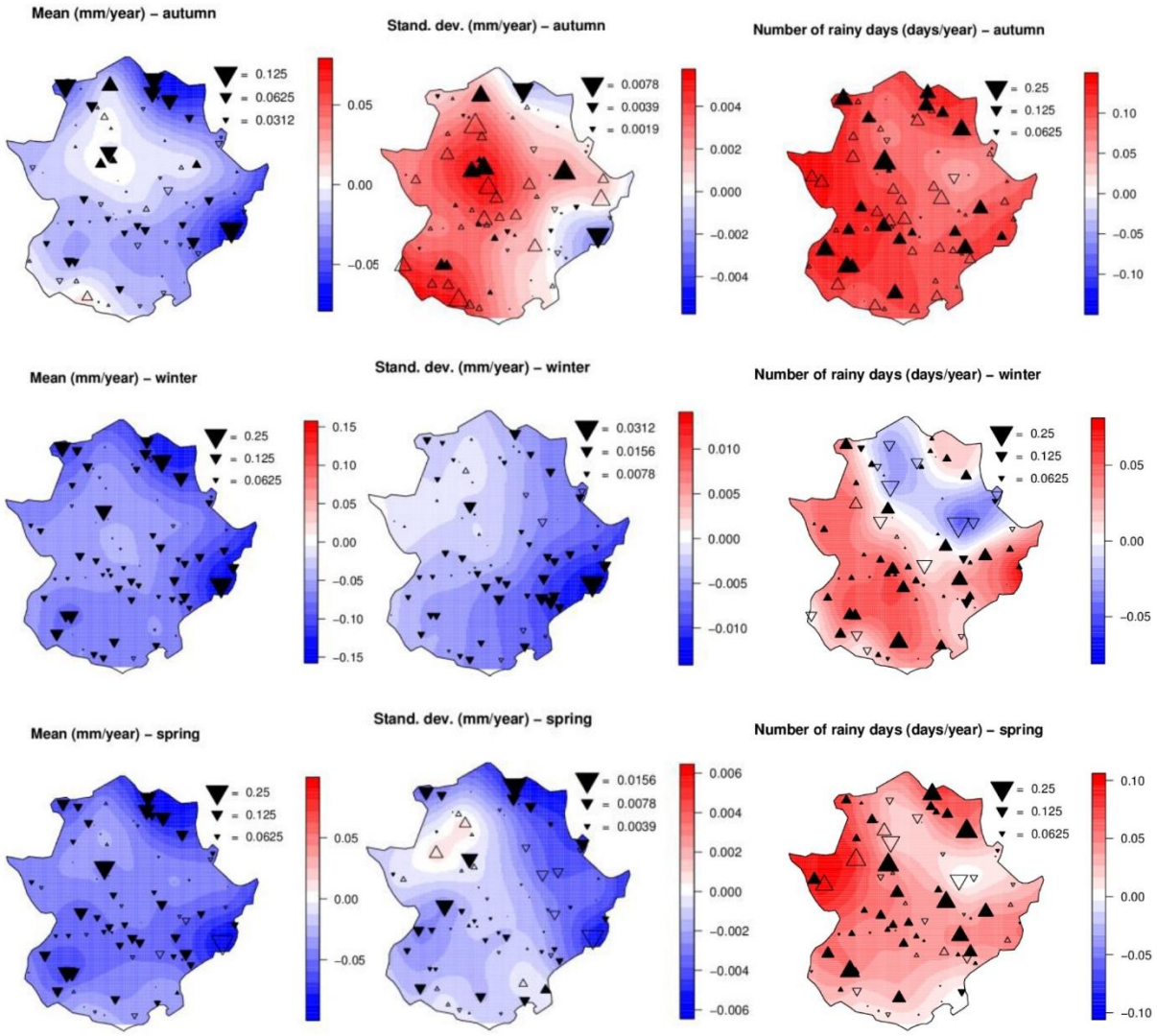

Figure 5. Linear trends in the time-varying threshold used in method M1 (left) and in the GPD scale parameter (right) for the three seasons considered. Upward triangles mean a positive trend, and downward negative. The size of the triangles represents the value of the trend for each gauge.

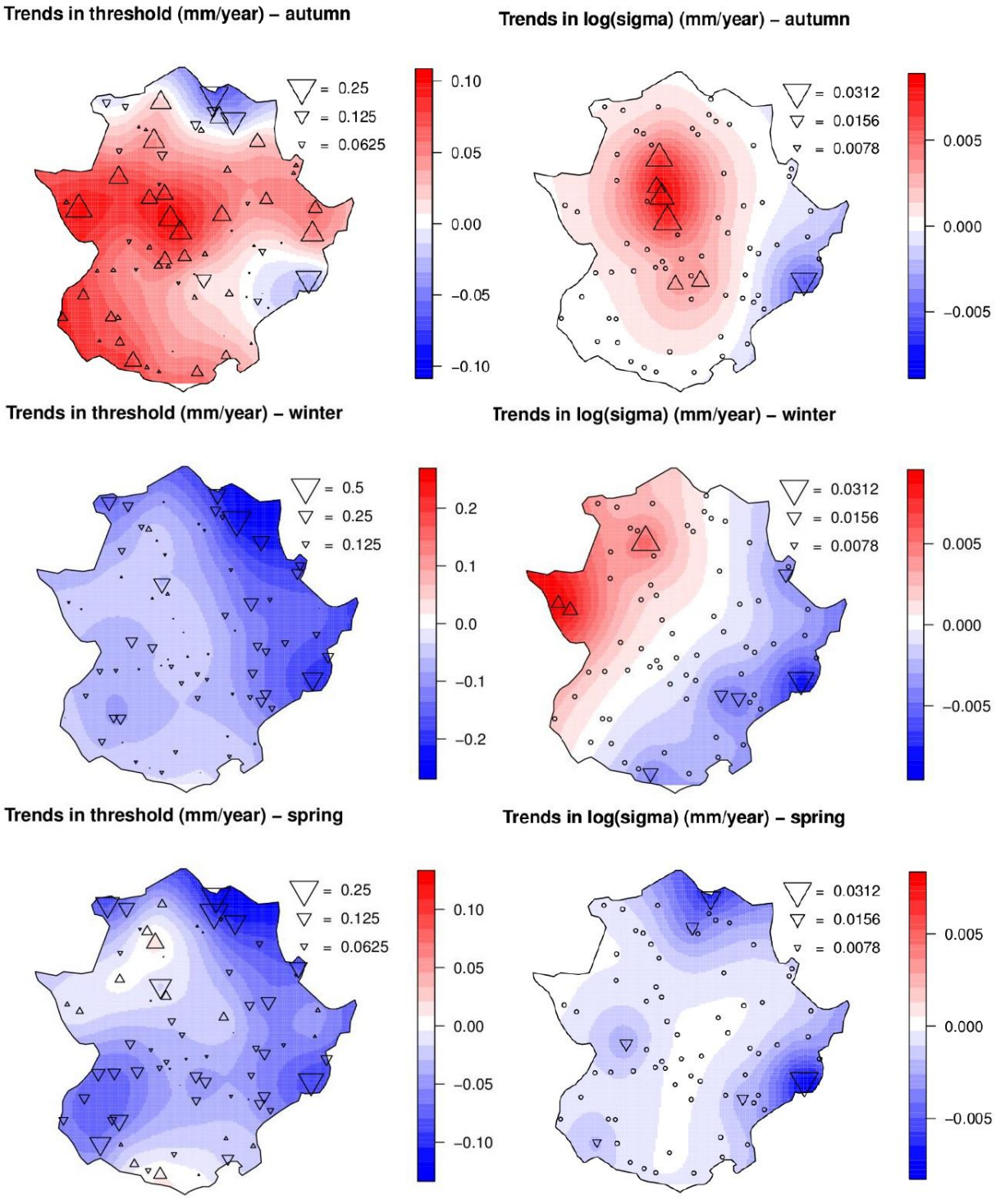

Figure 6. Spatial distribution of the 20-year RLs (Z20) in mm for the future climate in 2020, calculated from the all-day time series (left) using method M1 and from the rainy-days-only time series (right) using method M2.

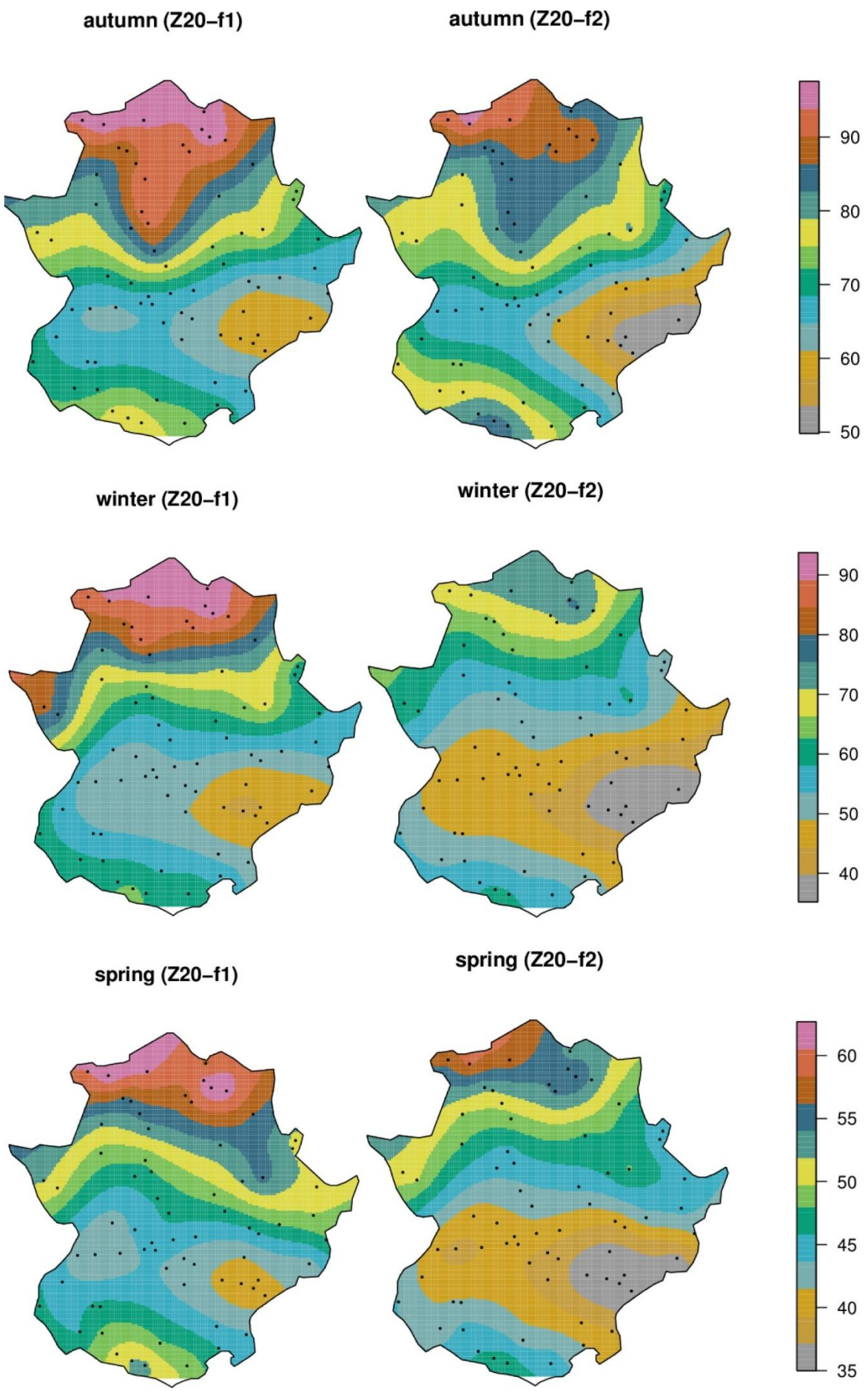
autumn (Z20–f1)
autumn (Z20–f2)
90
80
70
60
50
winter (Z20–f1)
winter (Z20–f2)
90
80
70
60
50
40
spring (Z20–f1)
spring (Z20–f2)
60
55
50
45
40
35

Figure 7. Spatial distribution of the 20-year RLs in 2020 obtained through the stationarity test (Z20-f2) that lie or do not lie inside the CI of the 20-year RLs obtained through extrapolation of the scale parameter (Z20-f1).

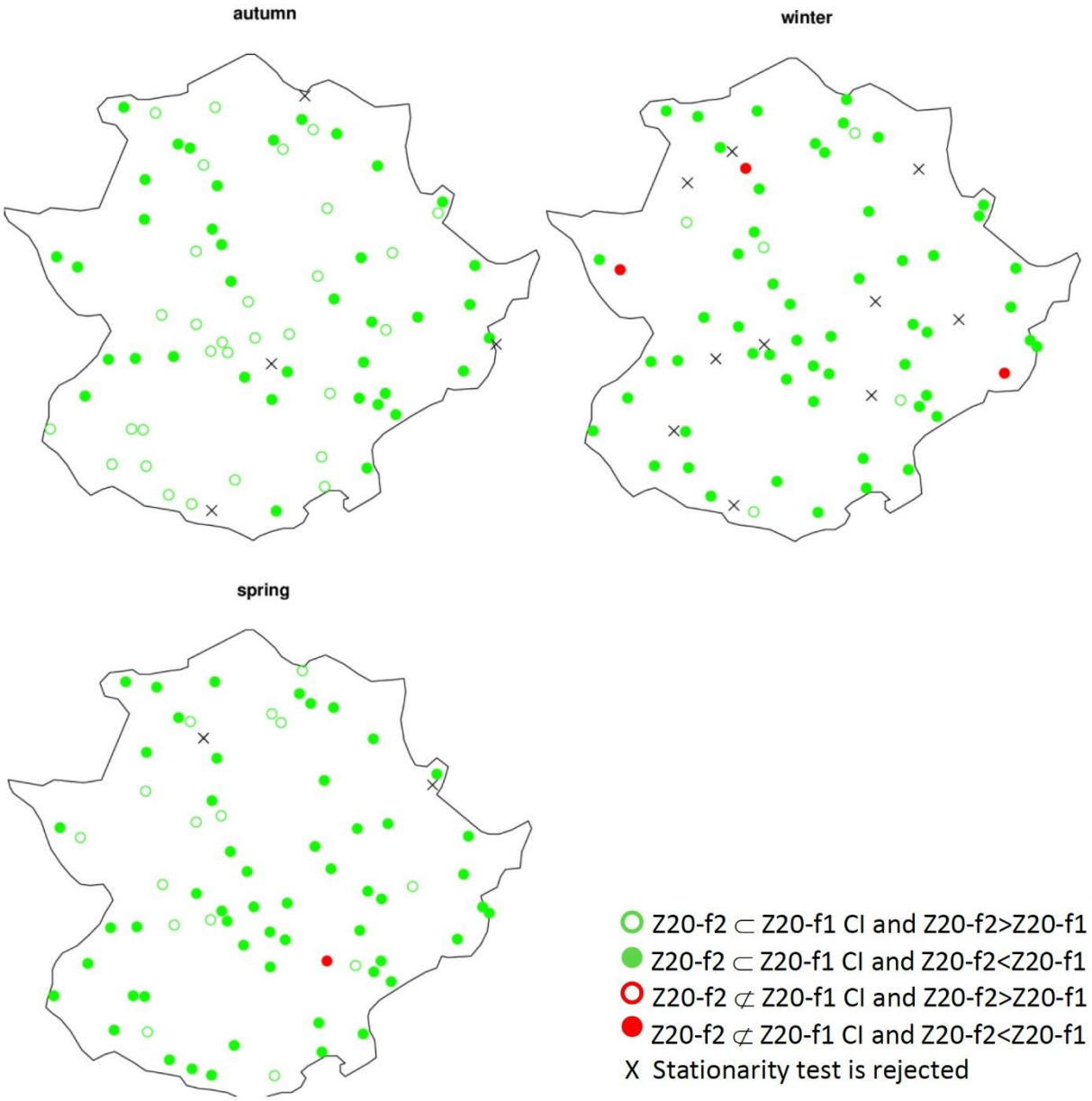

Figure 8. Spatial distribution of the width of the 20-year RL (Z20) confidence intervals (mm) in 2020, calculated from the all-day time series (left) using method M1 and from the rainy-days-only time series (right) using method M2.

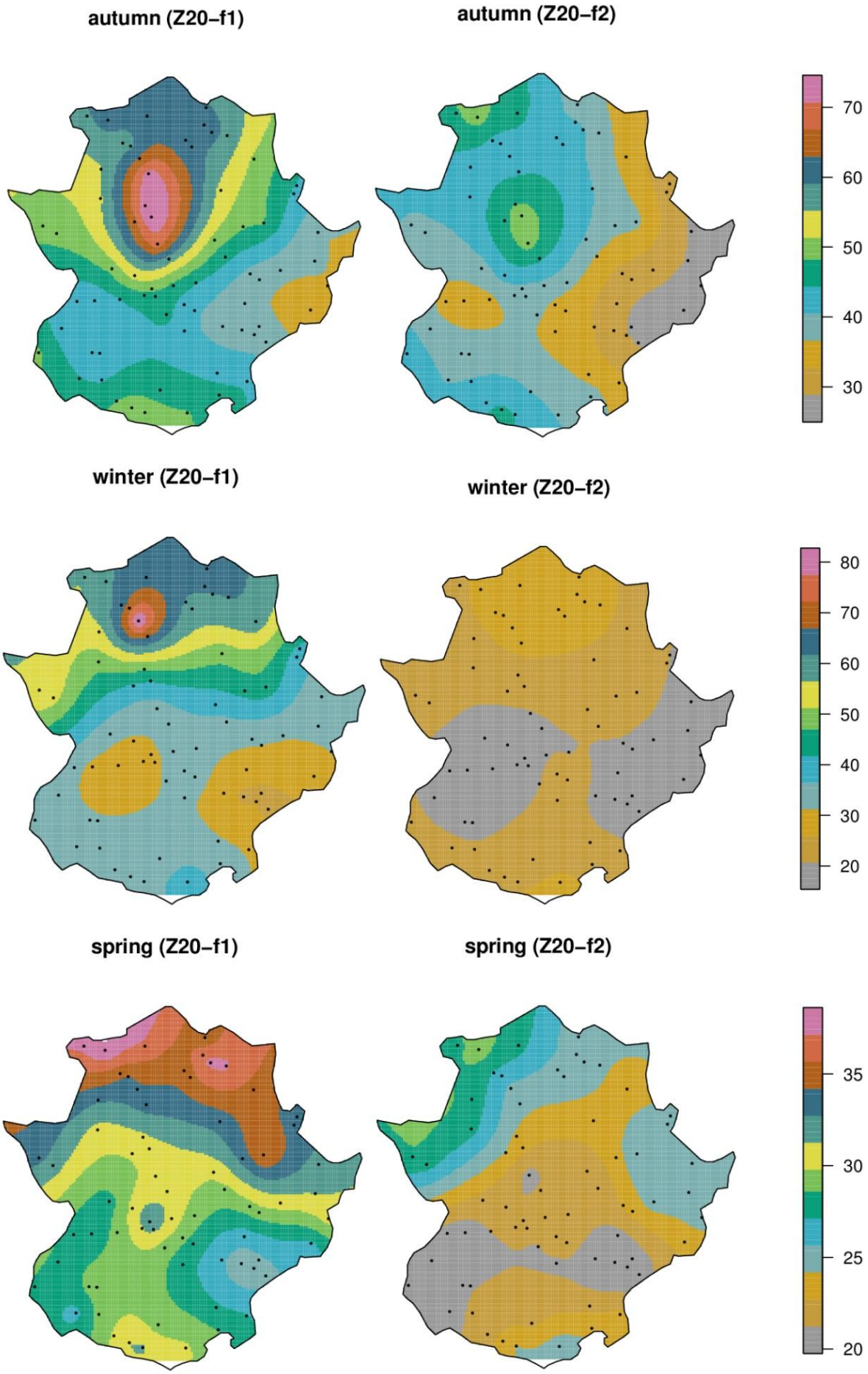
autumn (Z20–f1)
autumn (Z20–f2)
70
60
50
40
30
winter (Z20–f1)
winter (Z20–f2)
80
70
60
50
40
30
20
spring (Z20–f1)
spring (Z20–f2)
35
30
25
20

Figure 9. For Torrejoncillo gauge in winter (northernmost red point in Fig. 9-winter): (top) exceedences (dots) of the time-varying 98th percentile threshold (line); (centre) trend in the scale parameter (non-parametric in black, parametric in red); (bottom) a set of three plots for the mean, standard deviation, and number of rainy days (seasonal values in black, linear trends in red).

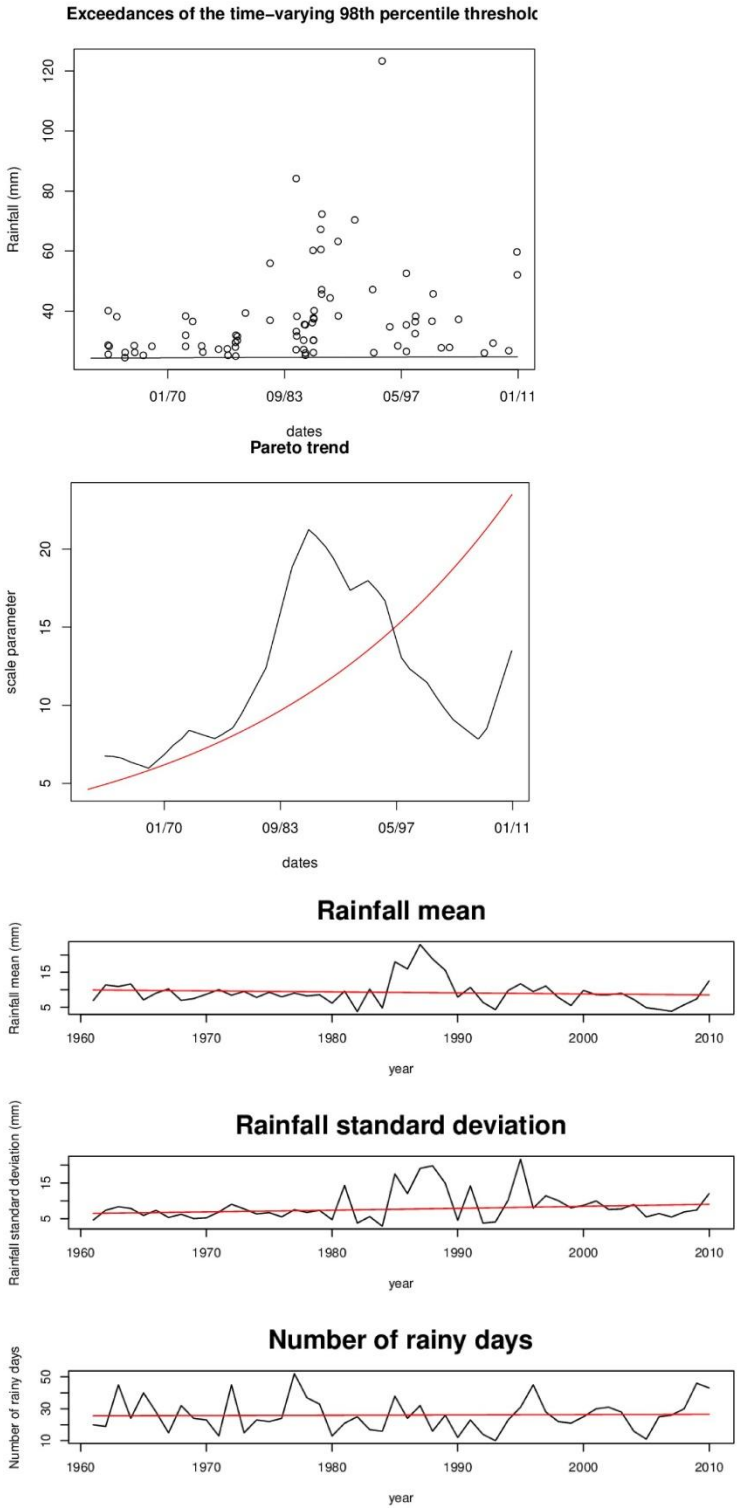

Figure 10. Spatial distribution of the 20-year RLs in 2020 obtained through the stationarity test (Z20-f2) that lie or do not lie inside the CI of the present 20-year RLs obtained with the same method (Z20-p2).

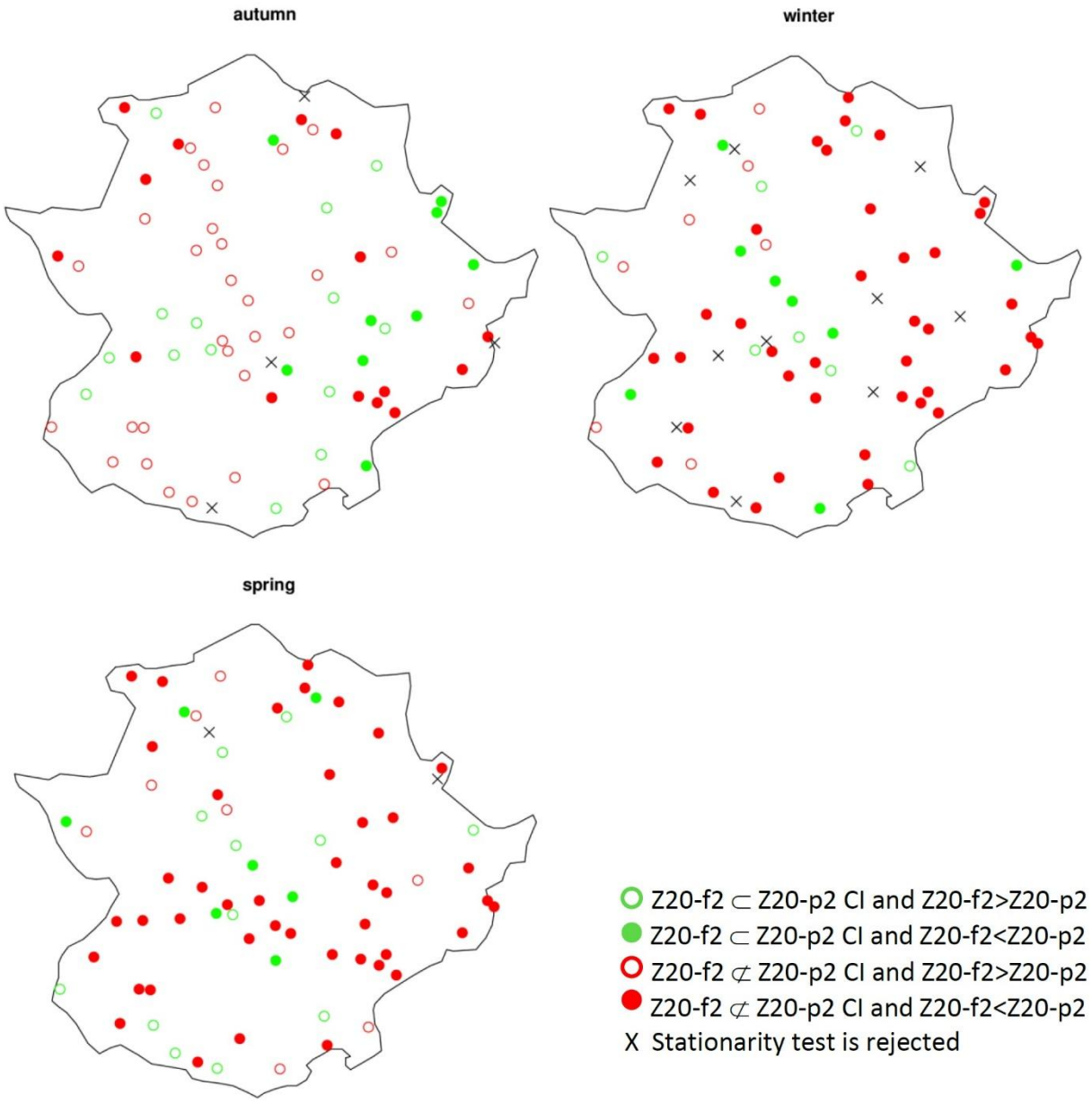

Figure 11. Spatial distribution of the 20-year RLs (mm) for each season considered for the present time (left) and future time (centre), and the differences between the present and the future cases (right).