



# A non-parametric entropy-based approach to detect changes in climate extremes

Philippe Naveau, Armelle Guillou, Théo Rietsch

## ► To cite this version:

Philippe Naveau, Armelle Guillou, Théo Rietsch. A non-parametric entropy-based approach to detect changes in climate extremes. Journal of the Royal Statistical Society: Series B, 2014, 76 (5), 10.1111/rssb.12058 . hal-01312930

**HAL Id: hal-01312930**

**<https://hal.science/hal-01312930>**

Submitted on 11 May 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A non-parametric entropy-based approach to detect changes in climate extremes

Philippe Naveau

*Laboratoire des Sciences du Climat et de l'Environnement, LSCE-IPSL-CNRS, Gif-sur-Yvette, France.*

Armelle Guillou

*Université de Strasbourg et CNRS, Institut de Recherche Mathématique Avancée, UMR 7501, Strasbourg, France.*

Théo Rietsch

*Université de Strasbourg et CNRS, Institut de Recherche Mathématique Avancée, UMR 7501, Strasbourg, and Laboratoire des Sciences du Climat et de l'Environnement, LSCE-IPSL-CNRS, Gif-sur-Yvette, France.*

**Abstract.** This paper focuses primarily on temperature extremes measured at 24 european stations with at least 90 years of data. Here, the term extremes refers to rare excesses of daily maxima and minima. As mean temperatures in this region have been warming over the last century, it is mechanical that this positive shift can be detected also in extremes. After removing this warming trend, we focus on the question of determining if other changes are still detectable in such extreme events. As we do not want to hypothesize any parametric form of such possible changes, we propose a new non-parametric estimator based on the Kullback-Leibler divergence tailored for extreme events. The properties of our estimator are studied theoretically and tested with a simulation study. Our approach is also applied to seasonal extremes of daily maxima and minima for our 24 selected stations.

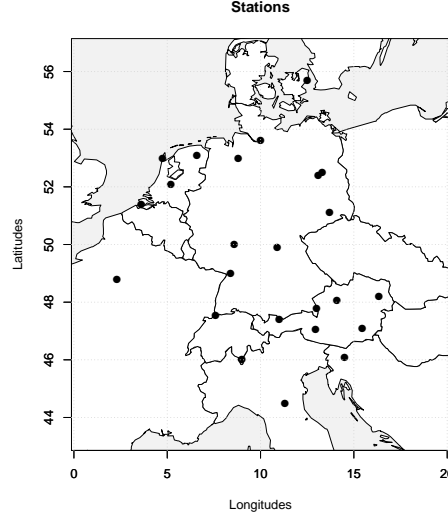
## 1. Introduction

In a global warming context, climatologists, flood planners, insurers and risk modellers have been increasingly interested in determining whether the upper tail distribution of some meteorological quantity has changed over time at some specific places (Zwiers et al., 2011; Kharin et al., 2007). As a motivating example, we focus on 24 weather stations that have at least 90 years of daily maxima and minima temperature measurements, see Table 1 and black dots in Figure 1.

A typical inquiry in impact studies is to wonder if high temperatures over the current climatology (i.e. the last thirty years) significantly differ from the ones measured during previous time periods. As a positive shift in the mean behaviour of temperatures has been observed (e.g., see Figure 1 in Abarca-Del-Rio and Mestre, 2006), such a warming automatically translates into higher absolute temperatures (e.g., Shaby and Reich, 2013; Jarušková and Rencová, 2008; Dupuis, 2012). This mean behaviour being removed, is it still possible to detect changes in high extreme temperatures over the last century? This important climatological question can be related to the investigation of Hoang et al. (2009) who wondered if the trends in extremes are only due to trends in mean and variance of the whole dataset. Our scope is different here. We neither aim at identifying smooth trends in extremes nor at linking changes between variances and upper quantiles. Our objective differs in the sense that we only want to determine if there is a change in extremes distributions. To explore such a question, we would like to assume very few distributional hypotheses, i.e. not imposing a specific parametric density, and to propose a fast statistical non-parametric approach

**Table 1.** Characteristics of 24 weather stations from the European Climate Assessment & Dataset project <http://eca.knmi.nl/dailydata/predefinedseries.php>. The heights are expressed in meters.

<b>Austria</b>						
Station name	Latitude	Longitude	Height	First year	Last year	Missing years
Kremsmunster	+48:03:00	+014:07:59	383	1876	2011	-
Graz	+47:04:59	+015:27:00	366	1894	2011	2
Salzburg	+47:48:00	+013:00:00	437	1874	2011	5
Sonnblick	+47:03:00	+012:57:00	3106	1887	2011	-
Wien	+48:13:59	+016:21:00	199	1856	2011	2
<b>Denmark</b>						
Station name	Latitude	Longitude	Height	First year	Last year	Missing years
Koebenhavn	+55:40:59	+012:31:59	9	1874	2011	-
<b>France</b>						
Station name	Latitude	Longitude	Height	First year	Last year	Missing years
Montsouris	+48:49:00	+002:19:59	77	1900	2010	-
<b>Germany</b>						
Station name	Latitude	Longitude	Height	First year	Last year	Missing years
Bamberg	+49:52:31	+010:55:18	240	1879	2011	-
Berlin	+52:27:50	+013:18:06	51	1876	2011	1
Bremen	+53:02:47	+008:47:57	4	1890	2011	1
Dresden	+51:07:00	+013:40:59	246	1917	2011	-
Frankfurt	+50:02:47	+008:35:54	112	1870	2011	1
Hamburg	+53:38:06	+009:59:24	11	1891	2011	-
Karlsruhe	+49:02:21	+008:21:54	112	1876	2011	2
Potsdam	+52:22:59	+013:04:00	100	1893	2011	-
Zugspitze	+47:25:00	+010:58:59	2960	1901	2011	1
<b>Italy</b>						
Station name	Latitude	Longitude	Height	First year	Last year	Missing years
Bologna	+44:30:00	+011:20:45	53	1814	2010	-
<b>Netherlands</b>						
Station name	Latitude	Longitude	Height	First year	Last year	Missing years
De Bilt	+52:05:56	+005:10:46	2	1906	2011	-
Den Helder	+52:58:00	+004:45:00	4	1901	2011	-
Eelde	+53:07:24	+006:35:04	5	1907	2011	-
Vlissingen	+51:26:29	+003:35:44	8	1906	2011	2
<b>Slovenia</b>						
Station name	Latitude	Longitude	Height	First year	Last year	Missing years
Ljubjana	+46:03:56	+014:31:01	299	1900	2011	5
<b>Switzerland</b>						
Station name	Latitude	Longitude	Height	First year	Last year	Missing years
Basel	+47:33:00	+007:34:59	316	1901	2011	-
Lugano	+46:00:00	+008:58:00	300	1901	2011	-



**Figure 1.** Weather station locations described in Table 1 (source: ECA&D database).

that could be implemented to large datasets. Although no large outputs from Global climate models will be treated here, we keep in mind computational issues when proposing the statistical tools developed therein.

One popular approach used in climatology consists in building a series of so-called extreme weather indicators and in studying their temporal variabilities in terms of frequency and intensity (e.g., Alexander et al., 2006; Frich et al., 2002). A limit of working with such indices is that they often focus on the “moderate extremes” (90% quantile or below), but not on upper extremes (above the 95% quantile). In addition, their statistical properties have rarely been derived. A more statistically oriented approach to analyze large extremes is to take advantage of Extreme Value Theory (EVT). According to Fisher and Tippet’s theorem (1928), if  $(X_1, \dots, X_n)$  is an independent and identically distributed sample of random variables and if there exists two sequences  $a_n > 0$  and  $b_n \in \mathbb{R}$  and a non degenerate distribution  $G_{\mu, \sigma, \xi}$  such that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \max_{i=1, \dots, n} \frac{X_i - b_n}{a_n} \leq x \right) = G_{\mu, \sigma, \xi}(x),$$

then  $G_{\mu, \sigma, \xi}$  belongs to the class of distributions

$$G_{\mu, \sigma, \xi}(x) = \begin{cases} \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]_+^{-\frac{1}{\xi}} \right\} & \text{if } \xi \neq 0 \\ \exp \left\{ - \exp \left[ - \left( \frac{x - \mu}{\sigma} \right) \right]_+ \right\} & \text{if } \xi = 0 \end{cases}$$

which is called the Generalized Extreme Value distributions (GEV). The shape parameter, often called  $\xi$  in environmental sciences, is of primary importance. It characterizes the GEV tail behaviour (see Coles, 2001; Beirlant et al., 2004, for more details). If one is ready to assume that temperature maxima for a given block size (day, month, season, etc) are approximately GEV distributed, then it is possible to study changes in the GEV parameters themselves (e.g. Fowler and Kilsby, 2003; Kharin et al., 2007).

This GEV based method is attractive because it takes advantage of EVT and hypothesis testing can be clearly implemented. For example, Jarušková and Rencová (2008) studied test statistics for detecting changes in annual maxima and minima temperature series measured at five meteorological sites—Bruxelles, Cadiz, Milan, St. Petersburg, and Stockholm. Three limitations of such GEV-based approaches can be identified. It is tailored for maxima and this discards all but one observation per block, e.g. one value out of 365 days for annual maxima. It imposes a fixed GEV form that may be too restrictive for small blocks (the GEV being a limiting distribution). Not one but three parameters have to be studied to detect changes in a time series. Regarding the first limitation, a classical solution in EVT (e.g., Coles, 2001) is to work with excesses above a high threshold instead of block maxima. The tail (survival) distribution of such excesses is usually modelled by a Generalized Pareto (GP) tail (Pickands, 1975)

$$\bar{H}_{\sigma,\xi}(y) = \begin{cases} (1 + \xi \frac{y}{\sigma})^{-1/\xi}, & \text{if } \xi \neq 0, \\ \exp(-\frac{y}{\sigma}), & \text{if } \xi = 0, \end{cases}$$

where the scale parameter  $\sigma$  is positive and  $y \geq 0$  if the shape parameter  $\xi$  is positive, and  $y \in [0, -\sigma/\xi[$  when  $\xi < 0$ . Still, the two other limitations remain for the GP model.

In this paper, we move away from imposing a fixed parametric density form. Having no parametric density at our disposal obviously implies that it is impossible to monitor changes in parameters (e.g. Grigga and Tawn, 2012). Another strategy has to be followed to compare the distributional differences between extremes. In information theory (e.g., Burnham and Anderson, 1998), it is a common endeavour to compare the probability densities of two time periods by computing the entropy (Kullback Leibler directed divergence)

$$I(f;g) = \mathbb{E}_f \left\{ \log \left( \frac{f(\mathbf{X})}{g(\mathbf{X})} \right) \right\},$$

where  $\mathbf{X}$  represents the random vector with density  $f$  and  $g$  another density. Although not a true distance, this expectation provides relevant information on how close  $g$  is from  $f$ . Kullback (1968) coined the term “directed divergence” to distinguish it from the divergence defined by

$$D(f;g) = I(f;g) + I(g;f),$$

which is symmetric relative to  $f$  and  $g$ . We will follow this terminology. Working with the entropy presents a lot of advantages. It is a notion shared by many different communities : physics, climatology, statistics, computer sciences and so on. It is a concise one-dimensional summary. It is clearly related to model selection and hypothesis testing (e.g. Burnham and Anderson, 1998). For example, the popular Akaike criterion (Akaike, 1974) can be viewed throughout the Kullback Leibler divergence lenses. For some distributions, explicit expressions of the divergence can be derived. This is the case if  $g$  and  $f$  correspond to two Gaussian multivariate densities (e.g., Penny and Roberts, 2000). In terms of extremes, if  $g$  and  $f$  represent two univariate GP densities with identical scale parameters and two different positive shape parameters,  $\xi_f$  and  $\xi_g$ , then we can write

$$I(f;g) = -1 - \xi_f + \text{sign}(\xi_f - \xi_g) \left( 1 + \frac{1}{\xi_g} \right) \left| \frac{\xi_f}{\xi_g} - 1 \right|^{-1/\xi_f} \int_{\frac{\xi_g}{\xi_f}}^1 t^{-1/\xi_f} |1 - t|^{1/\xi_f - 1} dt. \quad (1)$$

Although we will not assume that excesses follow explicitly a GP in our method, Equation (1) will be used in our simulations as a test case.

In this paper, our goal is to provide and study a non-parametric estimator of the divergence for large excesses. Fundamental features in an extreme value analysis are captured by the tail behaviour. In its original form, the divergence is not expressed in terms of tails but in function of probability densities. One important aspect of this work is to propose an approximation of the divergence in terms of the tail distributions, see Section 2. This leads to a new non-parametric divergence estimator tailored for excesses. Its properties will be studied in Section 3. The last section is dedicated to the analysis of a few simulations and of temperature extremes recorded at the stations plotted in Figure 1. All proofs are deferred to the Appendix.

## 2. Entropy for excesses

Our main interest resides in the upper tail behaviour and a first task is to make the divergence definition relevant to this extreme values context. This is done by replacing the densities  $f$  and  $g$  by densities of excesses above a certain high threshold  $u$ . We need a few notations to describe precisely this adapted divergence.

**DEFINITION 1.** *Let  $X$  and  $Y$  be two absolutely continuous random variables with density  $f$ , resp.  $g$ , and tail  $\bar{F}(x) = \mathbb{P}(X > x)$ , resp.  $\bar{G}(y) = \mathbb{P}(Y > y)$ . Denote the random variable above the threshold  $u$  as  $X_u = [X|X > u]$  with density  $f_u(x) = f(x)/\bar{F}(u)$  and tail  $\bar{F}_u(x) = \bar{F}(x)/\bar{F}(u)$  for  $x \in (u, x_F)$  where  $x_F$  is the upper endpoint of  $F$ . The same type of notations can be used for  $Y_u = [Y|Y > u]$ . A suitable version of the directed divergence for extreme values is then*

$$I(f_u; g_u) = \mathbb{E}_{f_u} \left\{ \log \left( \frac{f_u(X_u)}{g_u(X_u)} \right) \right\} = \frac{1}{\bar{F}(u)} \int_u^{x_F} \log \left( \frac{f_u(x)}{g_u(x)} \right) f(x) dx.$$

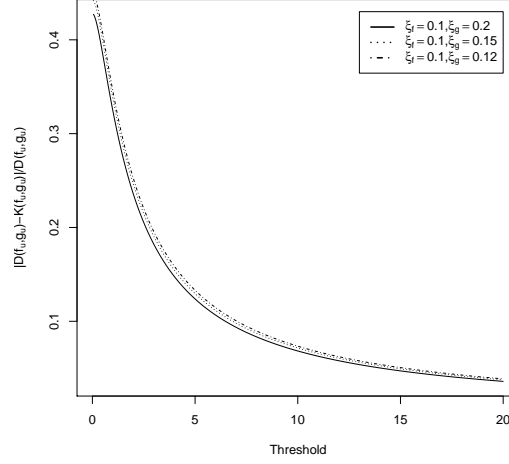
**ASSUMPTION 1.** *We always assume in this paper that the densities  $f$  and  $g$  are chosen such that the two directed divergences  $I(f_u; g_u)$  and  $I(g_u; f_u)$  are finite and in particular, both upper endpoints are equal to  $\tau = x_G = x_F$  in order to compute the ratios  $\frac{f_u(x)}{g_u(x)}$  and  $\frac{g_u(x)}{f_u(x)}$  for large  $x > u$ .*

If those assumptions are not satisfied in practice, this does not necessarily stop us from answering our climatological question: deciding if current temperature extremes over central Europe differ from past ones. If the difference  $|x_F - x_G|$  is large, then the divergence is infinite and there is no need to develop complex statistical procedures to detect differences between current and past extremes. If the difference  $|x_F - x_G|$  becomes smaller and smaller, it is more and more difficult to determine if the divergence is infinite from a given finite sample. For the limiting case,  $x_F = x_G$ , the divergence is finite and our estimate almost surely converges, see THEOREM 1. This case corresponds to our main assumption and it is particularly relevant about temperature extremes over central Europe because it is physically possible that an upper temperature bound exists for this region, (e.g., Shaby and Reich, 2013; Jarušková and Rencová, 2008).

As previously mentioned, we would like to express the divergence in terms of survival functions which are more adapted to extremes than densities. The next proposition reaches this goal by providing an approximation of the divergence in function of  $\bar{F}$  and  $\bar{G}$ .

**PROPOSITION 1.** *If*

$$\lim_{u \rightarrow \tau} \int_u^\tau \left( \log \frac{f(x)}{\bar{F}(x)} - \log \frac{g(x)}{\bar{G}(x)} \right) (f_u(x) - g_u(x)) dx = 0 \quad (2)$$



**Figure 2.** Relative error between  $K(f_u; g_u)$  and  $D(f_u; g_u)$  in function of different thresholds, see Proposition 1 when  $f$  and  $g$  correspond to two GP densities with a unit scale parameter and different shape parameters  $\xi_f = 0.1$  and  $\xi_g = 0.2, 0.15$  and  $0.12$ .

then the divergence  $D(f_u; g_u) = I(f_u; g_u) + I(g_u; f_u)$  is equivalent, as  $u \uparrow \tau$ , to the quantity

$$K(f_u; g_u) = -L(f_u; g_u) - L(g_u; f_u)$$

where

$$L(f_u; g_u) = \mathbb{E}_f \left\{ \log \frac{\bar{G}(X)}{\bar{G}(u)} \middle| X > u \right\} + 1. \quad (3)$$

For the special case of GP densities, we can explicitly compute from (1) the true divergence  $D(f_u; g_u)$  and its approximation  $K(f_u; g_u)$ . In Figure 2, the relative error  $|K(f_u; g_u) - D(f_u; g_u)| / D(f_u; g_u)$  when  $f$  and  $g$  correspond to two GP densities with a unit scale parameter and  $\xi_f = 0.1$  is displayed for different threshold values (x-axis). The solid, dotted and dash-dotted lines correspond to shape parameters  $\xi_g = 0.2, 0.15$  and  $0.12$ , respectively.

As the threshold increases, the relative error between  $K(f_u; g_u)$  and  $D(f_u; g_u)$  rapidly becomes small. The difference between  $\xi_f = 0.1$  and  $\xi_g$  does not play an important role.

The idea behind condition (2) is the following one. If  $\log \left( \frac{f(x) \cdot \bar{G}(x)}{\bar{F}(x) \cdot g(x)} \right)$  tends to a constant rapidly enough, then the integral  $\int_u^\tau \text{cst} \times (f_u(x) - g_u(x)) dx$  equals zero because  $\int_u^\tau f_u(x) dx = \int_u^\tau g_u(x) dx = 1$ . Is condition (2) satisfied for a large class of densities? The coming two subsections answer positively to this inquiry.

### 2.1. Checking condition (2)

In EVT, three types of tail behaviour (heavy, light and bounded) are possible and correspond to the GP sign of  $\xi$ , positive, null and negative, respectively. Those three cases have been extensively studied and they have been called the three domains of attraction, Fréchet, Gumbel and Weibull (e.g., see Chapter 2 of Embrechts et al., 1997). The next two propositions focus on the validity of

condition (2) for tails belonging to the Fréchet and Weibull domains of attraction, respectively. The Gumbel case that contains a lot of classical densities like the Gamma and Gaussian ones is more complex to deal with and we opt for a different approach based on stochastic ordering to check condition (2) for those types of densities.

**PROPOSITION 2.** *Suppose that the random variables  $X$  and  $Y$  belong to the Fréchet max-domain of attraction, i.e.  $\bar{F}$  and  $\bar{G}$  are regularly varying,*

$$\lim_{t \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\alpha} \text{ and } \lim_{t \rightarrow \infty} \frac{\bar{G}(tx)}{\bar{G}(t)} = x^{-\beta},$$

*for all  $x > 0$  and some  $\alpha > 0$  and  $\beta > 0$ . We also impose the following classical second order condition (see e.g. de Haan and Stadtmüller, 1996)*

$$\lim_{t \rightarrow \infty} \frac{\frac{\bar{F}(tx)}{\bar{F}(t)} - x^{-\alpha}}{q_F(t)} = x^{-\alpha} \frac{x^\rho - 1}{\rho} \text{ and } \lim_{t \rightarrow \infty} \frac{\frac{\bar{G}(tx)}{\bar{G}(t)} - x^{-\beta}}{q_G(t)} = x^{-\beta} \frac{x^\eta - 1}{\eta},$$

*for some  $\rho < 0$  and  $\eta < 0$  and some functions  $q_F \neq 0$  and  $q_G \neq 0$ . If the functions*

$$B(x) = \frac{xf(x)}{\bar{F}(x)} - \alpha \text{ and } C(x) = \frac{xg(x)}{\bar{G}(x)} - \beta,$$

*are eventually monotone, then condition (2) is satisfied.*

**PROPOSITION 3.** *Suppose that the random variables  $X$  and  $Y$  belong to the Weibull max-domain of attraction and have the same finite upper endpoints. Condition (2) is satisfied if the assumptions of Proposition 2 hold for the tail functions  $\bar{F}_*(x) := \bar{F}(\tau - x^{-1})$  and  $\bar{G}_*(x) := \bar{G}(\tau - x^{-1})$ .*

To treat the most classical densities belonging to the Gumbel domain of attraction, we need to recall the definition of the asymptotic stochastic ordering, (e.g., see Shaked and Shanthikumar, 1994). Having  $X_u \geq_{st} Y_u$  for large  $u$  means  $\mathbb{P}(X_u > t) \geq \mathbb{P}(Y_u > t)$ , for all  $t > u$  and for large  $u$ .

**PROPOSITION 4.** *Suppose that  $X_u \geq_{st} Y_u$  for large  $u$  and define*

$$\alpha(x) = \log \left( \frac{f(x)}{\bar{F}(x)} \right) - \log \left( \frac{g(x)}{\bar{G}(x)} \right).$$

*If  $\mathbb{E}(X_u)$ ,  $\mathbb{E}(\alpha(X_u))$  and  $\mathbb{E}(\alpha(Y_u))$  are finite and the derivative  $\alpha'(\cdot)$  is monotone and goes to zero as  $x \uparrow \tau$ , then (2) is satisfied.*

The proof of this proposition relies on a probabilistic version of the mean value theorem, (e.g. di Crescenzo, 1999). Applying this proposition can be straightforward in some important cases. For example, suppose that  $X$  and  $Y$  follow a standard exponential and standard normal distributions, respectively. We have  $\mathbb{E}(X_u) = 1 + u$ ,  $\mathbb{E}(Y_u) = \frac{\phi(u)}{\Phi(u)}$  and  $\alpha'(x) = x - \frac{\phi(x)}{\Phi(x)}$  which is monotone and goes to zero as  $x \uparrow \infty$ . For large  $u$ ,  $X_u \geq_{st} Y_u$ . Hence, condition (2) is satisfied.

Overall, condition (2) appears to be satisfied for most practical cases. In a nutshell, this condition tells us that our approximation can be used whenever the two densities of interest are comparable in their upper tails. For the few cases for which this condition is not satisfied, the discrepancy between the two tails is likely to be large and they can be easily handled for our climatological applications by just watching the two plotted time series under investigation and deduce that they are different.



### 3. Estimation of the divergence

In terms of inference, the key element given by Proposition 1 is captured by Equation (3). This expectation only depends on  $\bar{G}(X)$  and consequently, one can easily plug in an empirical estimator of  $\bar{G}$  to infer (3). More precisely, suppose that we have at our disposal two independent samples of size  $n$  and  $m$ ,  $\mathbf{X} = (X_1, \dots, X_n)^T$  and  $\mathbf{Y} = (Y_1, \dots, Y_m)^T$ . In our climate example, this could correspond to temperatures before 1980 and after 1980. To estimate  $\bar{G}(t)$ , we denote  $\bar{G}_m(t) = \sum_{j=1}^m \mathbb{1}_{\{Y_j > t\}}/m$  the classical empirical tail. To avoid taking the logarithm of zero in  $\log \bar{G}(X)$ , we slightly modify it by introducing

$$\tilde{\bar{G}}_m(t) := 1 - \frac{1}{m+1} \sum_{j=1}^m \mathbb{1}_{\{Y_j \leq t\}} = \frac{m}{m+1} \bar{G}_m(t) + \frac{1}{m+1}.$$

Our estimator of (3) is then simply defined by

$$\hat{L}(f_u; g_u) = 1 + \frac{1}{N_n} \sum_{i=1}^n \log \left( \frac{\tilde{\bar{G}}_m(X_i \vee u)}{\tilde{\bar{G}}_m(u)} \right), \text{ and } \hat{K}(f_u; g_u) = -\hat{L}(f_u; g_u) - \hat{L}(g_u; f_u), \quad (4)$$

where  $N_n$  represents the number of data points above the threshold  $u$  in the sample  $\mathbf{X}$ . To avoid dividing by zero when calculating  $\hat{L}(f_u; g_u)$ , we use the convention  $0/0 = 0$  whenever  $N_n$  is equal to zero in (4). The estimator  $\hat{L}(f_u; g_u)$  is non-parametric and it has the advantage of being extremely fast to compute. Its asymptotic properties need to be derived. One non trivial element for this theoretical task comes from the mixing of the two samples in (4) that makes the random variables  $\bar{G}_m(X_i \vee u)$  dependent.

**THEOREM 1.** *Assume that  $F$  and  $G$  are continuous. Let  $u < \tau$  fixed and suppose that the means*

$$\mathbb{E}_f \left( \log \left( \frac{\bar{G}(X \vee u)}{\bar{G}(u)} \right)^2 \right) \text{ and } \mathbb{E}_g \left( \log \left( \frac{\bar{F}(Y \vee u)}{\bar{F}(u)} \right)^2 \right)$$

*are finite,  $\frac{n}{m} \rightarrow c \in (0, \infty)$  and that there exists two non increasing sequences of positive numbers,  $k_n/n$  and  $\ell_m/m$ , satisfying*

$$k_n \geq \max \left( \log n, 8n\bar{F} \left( \bar{G}^{\leftarrow} \left( \frac{\ell_m}{m} \right) \right) \right), \quad \frac{k_n}{n} \log n \rightarrow 0 \quad \text{and} \quad \frac{\ell_m}{\log \log m} \rightarrow \infty.$$

*Then we have*

$$\hat{L}(f_u; g_u) - L(f_u; g_u) = o(1) \text{ a.s.}$$

*and*

$$\hat{K}(f_u; g_u) - K(f_u; g_u) = o(1) \text{ a.s.}$$

This theorem requires the existence of four sequences  $k_n^{(1)}, \ell_m^{(1)}$  and  $k_m^{(2)}, \ell_n^{(2)}$ . In the specific case of strict Pareto tails  $\bar{F}(x) = x^{-\alpha}$  and  $\bar{G}(x) = x^{-\beta}$  with  $\alpha, \beta > 0$ , a possible choice for these sequences is

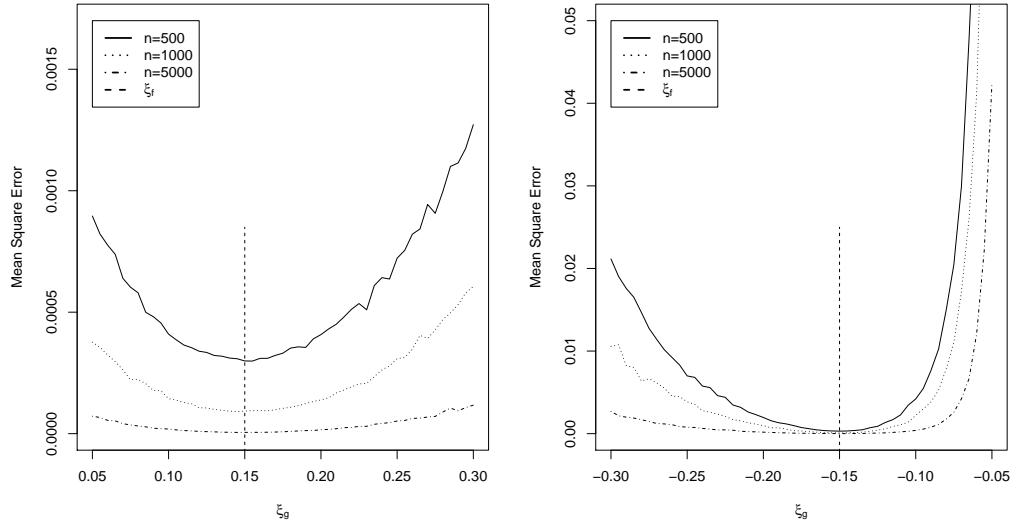
$$k_n^{(1)} = \frac{n}{\log n \log \log n} \text{ and } \ell_m^{(1)} = \frac{m}{(8 \log m \log \log m)^{\beta/\alpha}}$$

$$k_m^{(2)} = \frac{m}{\log m \log \log m} \text{ and } \ell_n^{(2)} = \frac{n}{(8 \log n \log \log n)^{\alpha/\beta}}.$$

## 4. Applications

### 4.1. Simulations

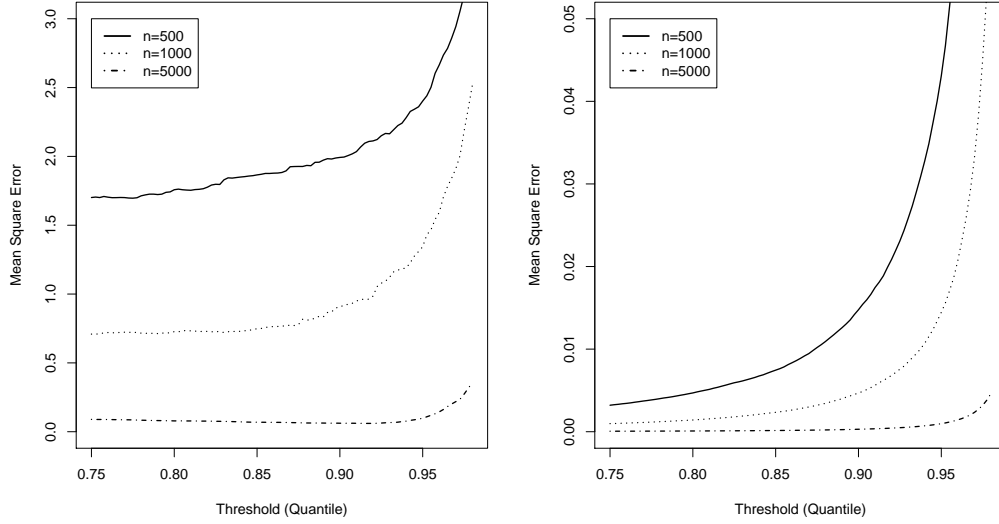
To compute  $\hat{K}(f_u; g_u)$  in (4) in our simulation study, we need to generate excesses from two different densities. A first choice is to choose two unit scale parameter GP densities with different shape parameters because we have the explicit expressions of  $L(f; g)$ , and  $K(f; g)$  for such distributions (see Equation (1)). To explore the Fréchet case, we arbitrarily set  $\xi_f = 0.15$ . This corresponds to a typical value for daily precipitation extremes (e.g., see Table 1 in Katz et al., 2002). Concerning the shape parameter for  $g$ , it varies from  $\xi_g = 0.05$  to  $0.3$ . For each value of  $\xi_f = 0.15$  and  $\xi_g$ , two samples with  $m = n$  are simulated and our estimator  $\hat{K}(f_u; g_u)$  defined from (4) can be calculated. We repeat this experiment 500 times for three different sample sizes  $n \in \{500, 1000, 5000\}$ . The classical Mean Square Error (MSE) can be inferred from the 500 estimates of  $\hat{K}(f_u; g_u)$ . The resulting MSE is plotted in the left panel of Figure 3. As expected, the MSE decreases as the sample size increases. The estimation of  $K(f; g)$  improves whenever the two shape parameters are close to each other.



**Figure 3.** Mean Square Error of  $\hat{K}(f_u; g_u)$  based on (4) and computed from 500 simulations of two GP distributed samples of sizes  $n \in \{500, 1000, 5000\}$  and  $n = m$ . The left and right panels correspond to a Fréchet and Weibull cases,  $\xi_f = 0.15$  and  $\xi_f = -0.15$ , respectively. The x-axis corresponds to different shape parameter values of  $\xi_g$ .

The same type of conclusions can be drawn for the Weibull domain. For this case, we set  $\xi_f = -0.15$ . Shape parameter values for temperature extremes usually belong to the interval  $[-0.3, -0.1]$  (e.g., see tables 6 and 7 in Jarušková and Rencová, 2008). In our simulations,  $\xi_g$  varies from  $-0.3$  to  $-0.05$  in the right panel of Figure 3. Overall, those MSE are small, but the two compared densities are GP distributed. To move away from this ideal situation, we keep the same GP density for  $g(\cdot)$  with  $\xi_g = 0.1$  (or  $\xi_g = -0.1$  for the Weibull case) but  $f$  corresponds to a Burr with survival function  $\bar{F}(x) = \left(\frac{1}{1+x^\tau}\right)^\lambda$  for  $x > 0$  or to a reverse Burr defined by  $\bar{F}(x) = \left(\frac{1}{1+(1-x)^{-\tau}}\right)^\lambda$  for  $x < 1$ . We fix

$\lambda = .5$  and  $\tau = 20$  in order to have  $\xi = \frac{1}{\lambda\tau} = 0.1$  for the Burr and  $\xi = -\frac{1}{\lambda\tau} = -0.1$  for the Reverse Burr. This design allows to assess the impact of the threshold choice that is represented in terms of quantiles on the x-axis of Figure 4.



**Figure 4.** Mean Square Error of  $\hat{K}(f_u; g_u)$  based on (4) and computed from 500 simulations with sizes  $n \in \{500, 1000, 5000\}$  and  $n = m$ . The left panel compares a GP distribution with  $\xi_f = 0.1$  and a Burr distribution. The right panel compares a GP distribution with  $\xi_f = -0.1$  and a reverse Burr distribution. The x-axis corresponds to different thresholds (expressed as the mean of the 95% quantiles of the  $\mathbf{X}$ 's and  $\mathbf{Y}$ 's samples).

If the sample size is large, the MSE remains constant over a wide range of thresholds. For small sample sizes, special caution have to be applied for the heavy tail case. The right panel of Figure 4 dealing with the Weibull case clearly shows a rapid increase of the MSE for  $n = 500$ .

Concerning our main question on how to decide if two time series have different extremes, Table 2 displays the number of false positive (grey columns) and negative at a confidence level of 95% for four different GP distributions situations: the so-called Weibull-Weibull case ( $\xi_f = -.1$  and  $\xi_g < 0$ ), the Fréchet-Fréchet case ( $\xi_f = .1$  and  $\xi_g > 0$ ), the Gumbel-Weibull case ( $\xi_f = 0$  and  $\xi_g < 0$ ) and the Gumbel-Fréchet case ( $\xi_f = 0$  and  $\xi_g > 0$ ). The scale parameter  $\sigma$  is taken as  $-\xi$  in the Weibull-Weibull case (to make sure that the endpoints of the two distributions are the same) and set to one in all other cases. To see the influence of the sample sizes,  $m = n$  can take five values (see the first column of Table 2). For each row of  $n$  and each column of  $\xi_g$ , the divergence  $\hat{K}(f_u; g_u)$  between the two samples is computed. To derive significance levels, we use a random permutation procedure (e.g., Davis et al., 2012). By randomly permuting observations between two samples, the type I error under the null of no distributional difference can be controlled. Repeating 200 times this type of resampling leads to 200 values of  $\hat{K}_{H_0}(f_u; g_u)$  for which the 95% quantile level can be inferred. The original  $\hat{K}(f_u; g_u)$  can be compared to this quantile and a decision can then be made. The bold values in Table 2 correspond to the number of wrong decisions made by the test based on the divergence. As we have implemented our procedure on 1000 replicas, we expect in average to count 50 false positives at the 95% level, i.e. the grey column should contain a number close to 50.

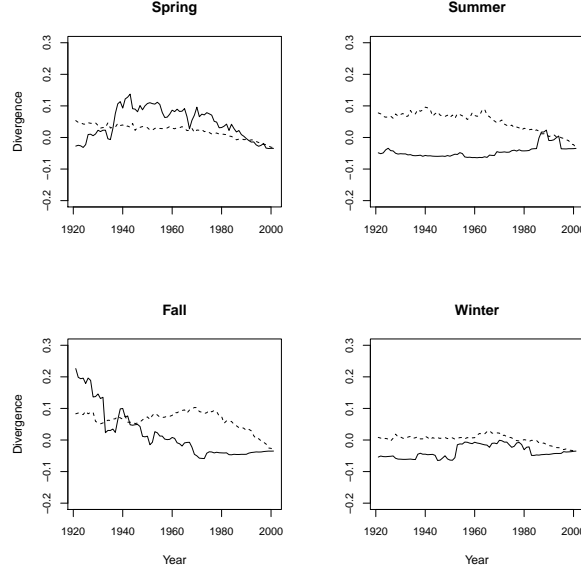
Outside of the grey columns, a value near zero indicates a good performance.

To benchmark our approach, we have also computed the classical non-parametric Kolmogorov-Smirnov, Wilcoxon rank sum and Wilcoxon signed rank tests (non bold values in Table 2) and they can be compared to our divergence based approach. Table 2 can teach us a few lessons. The Kolmogorov-Smirnov and Wilcoxon rank sum tests are overall worse than the others, especially the Wilcoxon rank sum. The Wilcoxon signed rank test and our approach provide similar results for the Weibull-Weibull case. With a sample size of 200, one can distinguish  $\xi_f = -.1$  from  $\xi_g \in \{-.2, -.05\}$ , but not from  $\xi_g \in \{-.15, -.08\}$ , those values being too close to  $-.1$ . For larger sample sizes, both approaches work well. The story is very different for the heavy tail case (Fréchet-Fréchet). One cannot expect differentiating  $\xi_f = .1$  from any of the values of  $\xi_g$  of the table for small and moderate sample sizes. For  $n = 10,000$ , our divergence estimate is able to identify a difference when  $\xi_g \in \{.05, 0.15, .2\}$ . This is not the case for the Wilcoxon signed rank test which is only able to detect a difference when  $\xi_g = 0.2$ . Finally, if we are in the Gumbel case ( $\xi_f = 0$ ), the statistic  $\hat{K}(f_u; g_u)$  works adequately for  $n = 200$  if  $\xi_g \leq -.3$ . It is also the case for  $n = 500$  and  $\xi_g \geq .2$ . In comparison, the Wilcoxon test has a much smaller validity range.

In summary, besides telling us that classical tests do not perform well but for Weibull-Weibull case, Table 2 emphasizes the difficulty of identifying small changes in non-negative shape parameters. For such a task, very large sample sizes are needed. Concerning our temperatures application, previous studies (e.g., Jarušková and Rencová, 2008) showed that the shape parameter of daily maxima is either below or equal to zero, i.e. we are in a Weibull-Weibull or a Gumbel-Weibull case. In the coming application section, we will deal with  $3 \times 30 \times 90 = 8,100$  daily measurements per season (a season has three months, a month around thirty days and we have about 90 years of data for most stations). We will compare periods of 30 years and work with a 95% quantile threshold which will provide approximately 130 extremes per season for each period. According to Table 2, this will enable us to explore the Weibull-Weibull case and the Gumbel-Weibull case since the results given by the divergence-based test are acceptable in these two cases for this kind of sample size.

**Table 2.** Number of false positive and negative out of 1000 replicas of two samples of sizes  $n = m$  for a 95% level. The grey columns count the number of false positive (wrongly rejecting that  $f$  and  $g$  are equal).  $f$  and  $g$  correspond to a  $GP(0, -\xi_f, \xi_f)$  and  $GP(0, -\xi_g, \xi_g)$  density, respectively in the first row of the table and to a  $GP(0, 1, \xi_f)$  and  $GP(0, 1, \xi_g)$  density, respectively in the last three rows. The bold values correspond to the number of wrong decisions obtained with the divergence approach, and the three other columns correspond, from the left to the right, to the classical Kolmogorov-Smirnov, Wilcoxon rank sum, and Wilcoxon signed rank tests, respectively.

Weibull-Weibull case		$\xi_f = -.1$																		
$n \backslash \xi_g$		-.2	-.15		-.1		-.08		-.05											
50	262	168	138	107	691	605	563	550	26	41	48	54	889	845	841	816	241	153	123	93
100	42	16	11	3	445	318	266	233	50	70	69	58	824	739	719	680	24	15	11	5
200	0	0	0	0	110	56	32	35	30	49	47	45	627	530	476	469	0	0	0	0
500	0	0	0	0	0	0	0	0	52	57	51	62	189	131	95	94	0	0	0	0
1000	0	0	0	0	0	0	0	0	42	48	57	50	17	10	4	5	0	0	0	0
Fréchet-Fréchet case		$\xi_f = .1$																		
$n \backslash \xi_g$		.05	.08		.1		.15		.2											
10	986	955	953	946	991	952	960	939	15	41	44	49	986	953	951	954	990	957	963	950
100	973	944	945	935	960	944	950	941	43	59	59	49	953	942	943	930	947	925	925	899
1000	937	923	901	740	946	957	955	926	45	34	34	49	943	931	908	773	854	835	740	285
10000	648	701	470	9	911	903	859	597	41	54	54	55	688	717	503	16	12	182	14	0
Gumbel-Weibull case		$\xi_f = 0$																		
$n \backslash \xi_g$		-.5	-.4		-.3		-.2		0											
50	804	806	680	155	868	847	757	377	921	893	847	628	949	934	912	799	28	45	50	48
100	513	635	390	3	727	730	556	37	896	853	743	231	933	906	862	599	38	53	45	50
200	63	332	104	0	348	534	280	0	709	722	502	12	888	870	777	252	37	48	52	60
500	0	28	0	0	1	159	19	0	134	424	138	0	616	712	490	4	58	55	52	53
1000	0	1	0	0	0	12	0	0	0	139	7	0	229	452	178	0	36	55	44	54
Gumbel-Fréchet case		$\xi_f = 0$																		
$n \backslash \xi_g$		0	.2		.3		.4		.5											
50	33	57	58	63	943	931	921	853	945	924	898	746	919	861	794	589	885	854	755	473
100	36	54	62	74	923	900	860	695	907	866	794	470	844	813	690	253	733	708	525	100
200	30	53	52	57	902	886	794	412	807	768	626	120	606	634	389	18	392	498	236	1
500	43	40	48	55	706	737	541	52	335	505	222	0	101	306	69	0	13	152	14	0
1000	45	40	51	59	473	541	262	1	34	213	31	0	0	71	0	0	0	10	0	0



**Figure 5.** Paris weather station: evolution of the divergence estimator (black curve),  $\hat{K}(f_u; g_u)$ , in function of the years  $[1900+t, 1929+t]$  with  $t \in \{1, \dots, 80\}$ . The reference period is the current climatology,  $[1981, 2010]$ . The dotted line represents the 95% significant level obtained by a random permutation procedure.

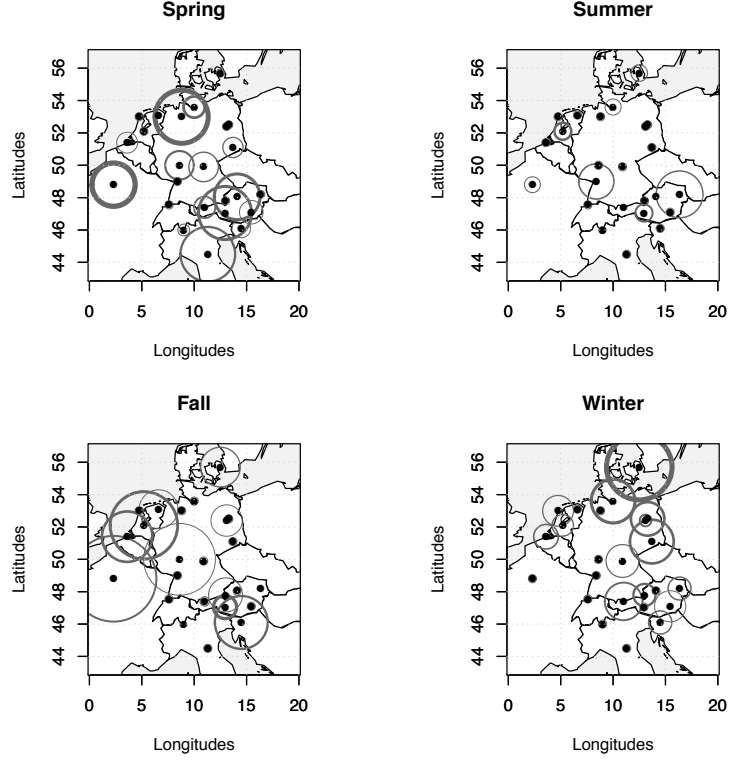
#### 4.2. Extreme temperatures

In geosciences, the yardstick period called a climatology is made of 30 years. So, we would like to know how temperature maxima climatologies have varied over different 30 year periods. To reach this goal, for any  $t \in \{1, \dots, 80\}$ , we compare the period  $[1900+t, 1929+t]$  with the current climatology  $[1981, 2010]$ . All our daily maxima and minima come from the ECA&D database (European Climate Assessment & Dataset project <http://eca.knmi.nl/dailydata/predefinedseries.php>). This database contains thousands of stations over Europe, but most measurement records are very short or incomplete and consequently, not adapted to the question of detecting changes in extremes. In this context, we only study stations that have at least 90 years of data, i.e. the black dots in Figure 1. As previously mentioned in the Introduction section, a smooth seasonal trend was removed in order to discard warming trends due to mean temperature changes. This was done by applying a classical smoothing spline with years as covariate for each season and station (R-package *mgcv*). The resulting trends appear to be coherent with mean temperature behaviour observed at the national and north-hemispheric levels (e.g., see Figure 5 in Abarca-Del-Rio and Mestre, 2006): an overall warming trend with local changes around 1940 and around 1970.

As a threshold needs to be chosen, we set it as the mean of the 95% quantiles of the two climatologies of interest. We first focus on one single location, the Montsouris station in Paris, where daily maxima of temperatures have been recorded for at least one century. Figure 5 displays the estimated  $\hat{K}(f_u; g_u)$  on the y-axis and years on the x-axis with  $t \in \{1, \dots, 80\}$ . We are going to use this example to explain how the grey circles on figures 6 and 7 have been obtained.

Similarly to our simulation study, a random permutation procedure with 200 replicas is run to derive the 95% confidence level. One slight difference with our simulation study is that instead of resampling days we have randomly resampled years in order to take care of serial temporal

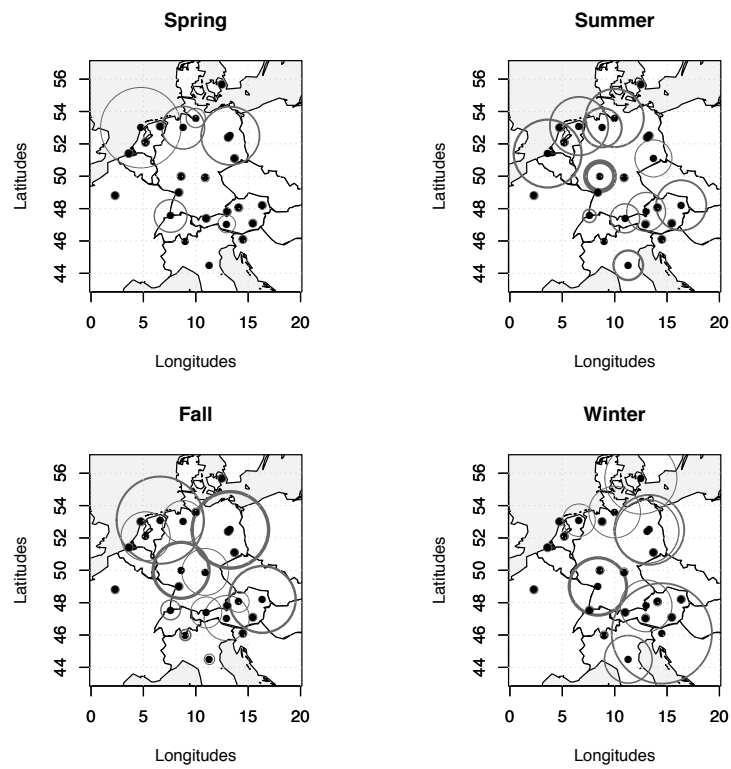
## Maxima



**Figure 6.** The black dots represent the 24 locations described in Table 1 and come from the ECA&D database. The way the circles are built is explained in Section 4.2.

correlations (it is unlikely that daily maxima are dependent from year to year). From Figure 5, the Fall season in Paris appears to be significantly different at the beginning of the 20th century than today. To quantify this information, it is easy to compute how long and how much the divergence is significantly positive. More precisely, we count the number of years for which  $\hat{K}(f_u; g_u)$  resides above the dotted line, and we sum up the divergence during those years (divided by the total number of years). Those two statistics can be derived for each station and for each season. In figures 6 and 7, the circle width and diameter correspond to the number of significant years and to the cumulative divergence over those years, respectively. For example, temperature maxima at the Montsouris station in Figure 5 often appear significantly different during Spring time (the border of the circle is thick in Figure 6) but the corresponding divergences are not very high on average. On the contrary, there are very few significant years during the Fall season, but the corresponding divergences are much higher (larger diameters with thinner border in Figure 6). This spatial representation tends to indicate that there are geographical and seasonal differences. For daily maxima, few locations witnessed significant changes in Summer. In contrast, the Winter season appears to have witnessed extremes changes during the last century. This is also true for the Spring season, but to a lesser degree. Daily minima divergences plotted in Figure 7 basically follows an opposite pattern, the Summer and Fall seasons appear to have undergone the most detectable changes.

## Minima



**Figure 7.** Same as Figure 6, but for daily minima.



## 5. Discussions

Recently, there have been a series of articles dealing with temperature extremes over Europe (e.g., Shaby and Reich, 2013; Jarušková and Rencová, 2008) and it is natural to wonder if our results differ from those past studies. The two main differences are the object of study and the variable of interest. Here, the latter corresponds to seasonal excesses obtained after removing a trend and the former focuses on determining if current excesses are different from the past ones (albeit the warming trend present in mean temperatures). Shaby and Reich (2013) aimed at a different objective. They solely focused on yearly maxima (not seasonal component) and took advantage of a flexible spatial max-stable model to pool information from around 1,000 stations (most of the sites appear after 1950 and this puts a stronger weight on the last 50 years). They found that “the distribution of extreme high temperatures seems to have shifted to the right (indicating warmer temperatures) in central Europe and Italy, while the distribution seems to have shifted to the left in Denmark, large swaths of Eastern Europe, and small pockets of Western Europe.” It is not clear, if those shifts are due to changes in their GEV location parameters or to other alterations of the overall distribution shape. Hence, our study provides a complementary view by zooming on second order characteristics. Modifications of the distribution shape could have potentially dire straits consequences. We lack the spatial component for two reasons. Practically, stations with a very long record are very few and it is difficult to infer a reasonable spatial structure. Theoretically, statistical estimation techniques for excesses processes (e.g., Ferreira and de Haan, 2012) are still very rare, especially if we want to stay within a non-parametric framework. Future developments are needed to explore this theoretical and applied question.

## Acknowledgments

Part of this work has been supported by the EU-FP7 ACQWA project ([www.acqwa.ch](http://www.acqwa.ch)) under Contract Nr 212250, by the PEPER-GIS project, by the ANR MOPERA project, by the ANR McSim project and the MIRACCLE-GICC project. The authors thank the European Climate Assessment & Dataset project that freely provides daily climatological data. The authors are very grateful to the associate editor and the reviewers for their careful reading of the paper and their comments that lead to significant improvements of the initial draft.

## Appendix

### Proof of Proposition 1

For  $x > u$ , using the decomposition

$$\log \left( \frac{f_u(x)}{g_u(x)} \right) = \log \left( \frac{f(x)}{\bar{F}(x)} \right) + \log \left( \frac{\bar{F}(x)}{\bar{F}(u)} \right) + \log \left( \frac{\bar{G}(u)}{\bar{G}(x)} \right) + \log \left( \frac{\bar{G}(x)}{g(x)} \right),$$

together with the fact that  $\mathbb{E}_{f_u} \{ \log \bar{F}_u(X_u) \} = -1$ , the Kullback Leibler distance  $D(f_u; g_u)$  can be rewritten as

$$\begin{aligned} D(f_u; g_u) &= \int_u^\tau \left( \log \frac{f(x)}{\bar{F}(x)} - \log \frac{g(x)}{\bar{G}(x)} \right) (f_u(x) - g_u(x)) dx - 2 \\ &\quad - \mathbb{E}_f \left\{ \log \frac{\bar{G}(X)}{\bar{G}(u)} \middle| X > u \right\} - \mathbb{E}_g \left\{ \log \frac{\bar{F}(Y)}{\bar{F}(u)} \middle| Y > u \right\}. \end{aligned}$$

Using (2), Proposition 1 follows.

### Proof of Proposition 2

Let  $(u_n)_{n \in \mathbb{N}}$  be a sequence tending to infinity. We want to prove that

$$\int h_{u_n}(x) dx := \int \left( \log \frac{xf(x)}{\alpha \bar{F}(x)} - \log \frac{xg(x)}{\beta \bar{G}(x)} \right) \frac{f(x)}{\bar{F}(u_n)} \mathbb{1}_{\{x > u_n\}} dx \longrightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Combining the remark following the proof of Theorem 1 in de Haan (1996) with the second order condition stated in Proposition 2 and the fact that  $B(\cdot)$  and  $C(\cdot)$  are eventually monotone, we deduce that these functions are of constant sign for large values of  $x$ , go to 0, and that their absolute value is regularly varying with index  $\rho$  (respectively  $\eta$ ).

Thus

$$\begin{cases} \frac{xf(x)}{\alpha \bar{F}(x)} = 1 + x^\rho L_\rho(x) \\ \frac{xg(x)}{\beta \bar{G}(x)} = 1 + x^\eta L_\eta(x) \end{cases} \quad (5)$$

where  $L_\rho(\cdot)$  and  $L_\eta(\cdot)$  are two slowly varying functions. It is then clear that for  $x > u_n$

$$h_{u_n}(x) \longrightarrow 0.$$

Now, remark that for a sufficiently large sequence  $u_n$  we have the bound

$$|h_{u_n}(x)| \leq x^\zeta L_\zeta(x) \frac{f(x)}{\bar{F}(x)} \mathbb{1}_{\{x > 1\}} \leq C x^{\zeta-1} L_\zeta(x) \mathbb{1}_{\{x > 1\}}$$

where  $\zeta < 0$  and  $C$  a suitable constant. Thus this bound is integrable. Condition (2) follows by the dominated convergence theorem.

### Proof of Proposition 3

Similar to the proof of Proposition 2 with  $\bar{F}$  and  $\bar{G}$  replaced by  $\bar{F}_*$  and  $\bar{G}_*$  respectively.

### Proof of Proposition 4

Note that

$$\Delta(u) := \int_u^\tau \alpha(x) (f_u(x) - g_u(x)) dx = \mathbb{E}[\alpha(X_u)] - \mathbb{E}[\alpha(Y_u)].$$

Moreover, the stochastic ordering implies that if  $\mathbb{E}(X_u)$  and  $\mathbb{E}(Y_u)$  exist, then we have the inequality

$$\mathbb{E}(X_u) \geq \mathbb{E}(Y_u).$$

Thus an application of a probabilistic version of the mean value theorem leads to

$$\Delta(u) = \mathbb{E}[\alpha'(Z_u)] \{\mathbb{E}(X_u) - \mathbb{E}(Y_u)\}, \quad (6)$$

where  $Z_u$  corresponds to a non-negative random variable with density

$$f_{Z_u}(z) = \frac{\mathbb{P}[X > z \mid X > u] - \mathbb{P}[Y > z \mid Y > u]}{\mathbb{E}(X_u) - \mathbb{E}(Y_u)}, \quad \forall z > u$$

(see Theorem 4.1 in di Crescenzo, 1999). Note that we implicitly assume that  $\mathbb{E}(X_u) > \mathbb{E}(Y_u)$ . Otherwise Proposition 4 is trivial since  $\mathbb{E}(X_u) = \mathbb{E}(Y_u)$  combined with  $X_u \geq_{st} Y_u$  implies that  $X_u = Y_u$  in distribution.

To conclude the proof, we only need to show that  $\mathbb{E}[\alpha'(Z_u)] \rightarrow 0$  as  $u \rightarrow \tau$ . Since  $\alpha'(\cdot)$  is monotone and tends to 0 at  $\tau$ ,  $|\alpha'(\cdot)|$  is decreasing. Thus

$$|\mathbb{E}[\alpha'(Z_u)]| = \left| \int_u^\tau \alpha'(x) f_{Z_u}(x) dx \right| \leq \int_u^\tau |\alpha'(x) f_{Z_u}(x)| dx \leq |\alpha'(u)| \rightarrow 0.$$

The proof of Proposition 4 is then achieved.

### Proof of Theorem 1

We will use the notation

$$\widetilde{G}(t) := 1 - \frac{n}{n+1}G(t) = \frac{n}{n+1}\overline{G}(t) + \frac{1}{n+1}.$$

We start by decomposing the difference  $\widehat{L}(f_u; g_u) - L(f_u; g_u)$  into 6 terms :

$$\begin{aligned} \widehat{L}(f_u; g_u) - L(f_u; g_u) &= \frac{1}{N_n} \sum_{i=1}^n \log \left( \frac{\widetilde{G}_m(X_i \vee u)}{\widetilde{G}_m(u)} \right) - \mathbb{E}_f \left( \log \frac{\overline{G}(X)}{\overline{G}(u)} \middle| X > u \right) \\ &= \frac{1}{N_n} \sum_{i=1}^n \log \left( \frac{\widetilde{G}(X_i \vee u)}{\widetilde{G}(u)} \right) - \mathbb{E}_f \left( \log \frac{\overline{G}(X)}{\overline{G}(u)} \middle| X > u \right) \\ &\quad + \frac{1}{N_n} \sum_{i=1}^n \log \left( \frac{\widetilde{G}_m(X_i \vee u)}{\widetilde{G}(X_i \vee u)} \right) - \frac{n}{N_n} \log \left( \frac{\widetilde{G}_m(u)}{\widetilde{G}(u)} \right) \\ &= \frac{1}{\overline{F}_n(u)} \left[ \frac{1}{n} \sum_{i=1}^n \log \left( \frac{\widetilde{G}(X_i \vee u)}{\widetilde{G}(u)} \right) - \mathbb{E}_f \left( \log \frac{\overline{G}(X \vee u)}{\overline{G}(u)} \right) \right] \\ &\quad + \left( \frac{1}{\overline{F}_n(u)} - \frac{1}{\overline{F}(u)} \right) \mathbb{E}_f \left( \log \frac{\overline{G}(X \vee u)}{\overline{G}(u)} \right) \\ &\quad + \frac{1}{N_n} \sum_{i=1}^{n-k_n} \frac{\widetilde{G}_m(X_{i,n} \vee u) - \widetilde{G}(X_{i,n} \vee u)}{\widetilde{G}(X_{i,n} \vee u)} \\ &\quad + \frac{1}{N_n} \sum_{i=1}^{n-k_n} \left[ \log \left( 1 + \frac{\widetilde{G}_m(X_{i,n} \vee u) - \widetilde{G}(X_{i,n} \vee u)}{\widetilde{G}(X_{i,n} \vee u)} \right) - \frac{\widetilde{G}_m(X_{i,n} \vee u) - \widetilde{G}(X_{i,n} \vee u)}{\widetilde{G}(X_{i,n} \vee u)} \right] \\ &\quad + \frac{1}{N_n} \sum_{i=n-k_n+1}^n \log \left( \frac{\widetilde{G}_m(X_{i,n} \vee u)}{\widetilde{G}(X_{i,n} \vee u)} \right) \\ &\quad - \frac{n}{N_n} \log \left( \frac{\widetilde{G}_m(u)}{\widetilde{G}(u)} \right) \\ &=: \sum_{\ell=1}^6 Q_{\ell,n,m}. \end{aligned}$$

We study each term separately.

**Term  $Q_{1,n,m}$ .**

$$\begin{aligned}
 Q_{1,n,m} &= \frac{1}{\bar{F}_n(u)} \left[ \frac{1}{n} \sum_{i=1}^n \log \left( \frac{\frac{n}{n+1} \bar{G}(X_i \vee u) + \frac{1}{n+1}}{\frac{n}{n+1} \bar{G}(u) + \frac{1}{n+1}} \right) - \mathbb{E}_f \left( \log \frac{\bar{G}(X \vee u)}{\bar{G}(u)} \right) \right] \\
 &= \frac{1}{\bar{F}_n(u)} \left[ \frac{1}{n} \sum_{i=1}^n \log \left( \frac{\bar{G}(X_i \vee u) + \frac{1}{n}}{\bar{G}(u) + \frac{1}{n}} \right) - \mathbb{E}_f \left( \log \frac{\bar{G}(X \vee u)}{\bar{G}(u)} \right) \right] \\
 &= \frac{1}{\bar{F}_n(u)} \left[ \frac{1}{n} \sum_{i=1}^n \left\{ \log \left( \frac{\bar{G}(X_i \vee u) + \frac{1}{n}}{\bar{G}(u) + \frac{1}{n}} \right) - \mathbb{E}_f \left( \log \frac{\bar{G}(X \vee u) + \frac{1}{n}}{\bar{G}(u) + \frac{1}{n}} \right) \right\} \right] \\
 &\quad + \frac{1}{\bar{F}_n(u)} \mathbb{E}_f \left( \log \left( \frac{\bar{G}(X \vee u) + \frac{1}{n}}{\bar{G}(u) + \frac{1}{n}} \frac{\bar{G}(u)}{\bar{G}(X \vee u)} \right) \right) \\
 &=: Q_{1,n,m}^{(1)} + Q_{1,n,m}^{(2)}.
 \end{aligned}$$

Denote by

$$Z_i^{(n)} := \log \left( \frac{\bar{G}(X_i \vee u) + \frac{1}{n}}{\bar{G}(u) + \frac{1}{n}} \right) - \mathbb{E}_f \left( \log \frac{\bar{G}(X \vee u) + \frac{1}{n}}{\bar{G}(u) + \frac{1}{n}} \right).$$

Clearly  $Z_i^{(n)}$  is an array of centred random variables that are identically distributed and rowwise independent. Thus, according to the strong law for arrays (see e.g. Chow and Teicher, 1978, p. 393), we have

$$\frac{1}{n} \sum_{i=1}^n Z_i^{(n)} = o(1) \text{ a.s.}$$

as soon as  $\mathbb{E}_f(Z_i^{(1)})^2 < \infty$ . Now remark that  $\log \left( \frac{\bar{G}(u) + \frac{1}{n}}{\bar{G}(X \vee u) + \frac{1}{n}} \right)$  is a positive increasing function of  $n$ , thus

$$0 \leq \log \left( \frac{\bar{G}(u) + 1}{\bar{G}(X \vee u) + 1} \right) \leq \log \left( \frac{\bar{G}(u)}{\bar{G}(X \vee u)} \right) \implies \mathbb{E}_f(Z_i^{(1)})^2 \leq \mathbb{E}_f \left( \log^2 \left( \frac{\bar{G}(u)}{\bar{G}(X \vee u)} \right) \right) < \infty$$

by assumption. Consequently

$$Q_{1,n,m}^{(1)} = o(1) \text{ a.s.}$$

Now

$$Q_{1,n,m}^{(2)} = -\frac{1}{\bar{F}_n(u)} \left\{ \mathbb{E}_f \left( \log \left( \frac{\bar{G}(u) + \frac{1}{n}}{\bar{G}(X \vee u) + \frac{1}{n}} \right) \right) + \mathbb{E}_f \left( \log \left( \frac{\bar{G}(X \vee u)}{\bar{G}(u)} \right) \right) \right\}.$$

Using again the fact that  $\log \left( \frac{\bar{G}(u) + \frac{1}{n}}{\bar{G}(X \vee u) + \frac{1}{n}} \right)$  is a positive increasing function of  $n$ , by the dominated convergence theorem, we deduce that

$$\lim_n \mathbb{E}_f \left( \log \left( \frac{\bar{G}(u) + \frac{1}{n}}{\bar{G}(X \vee u) + \frac{1}{n}} \right) \right) = \mathbb{E}_f \left( \lim_n \left( \log \left( \frac{\bar{G}(u) + \frac{1}{n}}{\bar{G}(X \vee u) + \frac{1}{n}} \right) \right) \right) = \mathbb{E}_f \left( \log \frac{\bar{G}(u)}{\bar{G}(X \vee u)} \right).$$

This implies that

$$Q_{1,n,m}^{(2)} = o(1) \text{ a.s.},$$

and thus

$$Q_{1,n,m} = o(1) \text{ a.s.}$$

**Term  $Q_{2,n,m}$ .**

By the strong law of large numbers, we have

$$Q_{2,n,m} = o(1) \text{ a.s.}$$

**Term  $Q_{3,n,m}$ .**

We need to use the sequences  $k_n$  and  $\ell_m$  to treat this term. Since  $k_n \geq \log n$ ,  $F^{\leftarrow}(1 - k_n/(8n))$  is eventually a.s. larger than  $F_n^{\leftarrow}(1 - k_n/n)$ , hence than  $X_{n-k_n,n}$ .

Here is a quick way to see this : If we set  $T_n = F^{\leftarrow}(1 - p\epsilon_n)$  for some  $0 < p < 1$  and let ‘Bin( $r, q$ )’ stand for a *binomial* ( $r, q$ ) *random variable*, then the properties of quantile functions imply that  $\mathbb{P}\{F_n^{\leftarrow}(1 - \epsilon_n) > F^{\leftarrow}(1 - p\epsilon_n)\} \leq \mathbb{P}\{\text{Bin}(n, p\epsilon_n) > n\epsilon_n\}$ , which is dominated by  $(\frac{enp\epsilon_n}{n\epsilon_n})^{n\epsilon_n} = (ep)^{n\epsilon_n} = n^{n\epsilon_n \log(ep)/\log(n)}$  (Giné and Zinn, 1984, Remark 4.7); if  $p = 1/8$  and  $n\epsilon_n \geq \log n$  then the series  $\sum (ep)^{n\epsilon_n}$  converges.

Thus, if we rewrite

$$\left| \frac{\tilde{G}_m(t) - \tilde{G}(t)}{\tilde{G}(t)} \right| = \frac{m}{n} \frac{n+1}{m+1} \frac{\overline{G}(t)}{\overline{G}(t) + \frac{1}{n}} \left| \frac{\overline{G}_m(t) - \overline{G}(t)}{\overline{G}(t)} - \frac{1 - \frac{n}{m} \overline{G}(t)}{n+1} \frac{\overline{G}(t)}{\overline{G}(t)} \right| \quad (7)$$

for  $n$  and  $m$  sufficiently large, we have

$$|Q_{3,n,m}| \leq \frac{2}{N_n} \sum_{i=1}^{n-k_n} \left| \frac{\overline{G}_m(X_{i,n} \vee u) - \overline{G}(X_{i,n} \vee u)}{\overline{G}(X_{i,n} \vee u)} \right| + \frac{2}{n+1} \left| 1 - \frac{n}{m} \right| \frac{1}{N_n} \sum_{i=1}^{n-k_n} \frac{1}{\overline{G}(X_{i,n} \vee u)}.$$

Remark now that

$$\begin{aligned} \frac{1}{N_n} \sum_{i=1}^{n-k_n} \frac{1}{\overline{G}(X_{i,n} \vee u)} &\leq \frac{1}{F_n(u)} \frac{n-k_n}{n} \frac{1}{\overline{G}(X_{n-k_n,n} \vee u)} \\ &\leq \frac{1}{F_n(u)} \frac{n-k_n}{n} \left\{ \frac{1}{\overline{G}(X_{n-k_n,n})} + \frac{1}{\overline{G}(u)} \right\} \\ &\leq \frac{1}{F_n(u)} \frac{n-k_n}{n} \left\{ \frac{m}{\ell_m} + \frac{1}{\overline{G}(u)} \right\} \text{ a.s.} \end{aligned}$$

Consequently, since  $\frac{n}{m} \rightarrow c \in (0, \infty)$ :

$$\begin{aligned} |Q_{3,n,m}| &\leq \frac{2}{N_n} \sum_{i=1}^{n-k_n} \left| \frac{\overline{G}_m(X_{i,n} \vee u) - \overline{G}(X_{i,n} \vee u)}{\overline{G}(X_{i,n} \vee u)} \right| + o(1) \text{ a.s.} \\ &\leq \frac{n-k_n}{n} \frac{2}{F_n(u)} \sup_{t \leq X_{n-k_n,n} \vee u} \left| \frac{\overline{G}_m(t) - \overline{G}(t)}{\overline{G}(t)} \right| + o(1) \text{ a.s.} \\ &\leq \frac{n-k_n}{n} \frac{2}{F_n(u)} \max \left( \sup_{t \leq X_{n-k_n,n}} \left| \frac{\overline{G}_m(t) - \overline{G}(t)}{\overline{G}(t)} \right|, \sup_{t \leq u} \left| \frac{\overline{G}_m(t) - \overline{G}(t)}{\overline{G}(t)} \right| \right) + o(1) \text{ a.s.} \end{aligned}$$

Thus, by our choice of sequences  $k_n$  and  $\ell_m$ , we have

$$\begin{aligned} \sup_{t \leq X_{n-k_n, n}} \left| \frac{\bar{G}_m(t) - \bar{G}(t)}{\bar{G}(t)} \right| &\leq \sup_{t \leq F^{\leftarrow}(1-k_n/(8n))} \left| \frac{\bar{G}_m(t) - \bar{G}(t)}{\bar{G}(t)} \right| \leq \sup_{t \leq \bar{G}^{\leftarrow}(\ell_m/m)} \left| \frac{\bar{G}_m(t) - \bar{G}(t)}{\bar{G}(t)} \right| \\ &= \sup_{t \geq \frac{\ell_m}{m}} \left| \frac{\mathcal{U}_m(t) - t}{t} \right| = o(1) \text{ a.s.} \end{aligned}$$

where  $\mathcal{U}_m$  denotes the empirical distribution function of  $m$  uniform  $(0, 1)$  random variables (see Corollary 1 in Wellner, 1978).

Also  $\forall T < \tau$ , we have:

$$\sup_{t \leq T} \left| \frac{\bar{G}_m(t) - \bar{G}(t)}{\bar{G}(t)} \right| = o(1) \text{ a.s.}$$

which leads to

$$|Q_{3,n,m}| = o(1) \text{ a.s.}$$

**Term  $Q_{4,n,m}$ .**

Now, following the lines of proof of the term  $Q_{3,n,m}$ , for  $i = 1, \dots, k_n$ , we have

$$\left| \frac{\tilde{\bar{G}}_m(X_{i,n} \vee u) - \tilde{\bar{G}}(X_{i,n} \vee u)}{\tilde{\bar{G}}(X_{i,n} \vee u)} \right| = o(1) \text{ a.s.}$$

and thus using the inequality  $\forall x \geq -\frac{1}{2}, |\log(1+x) - x| \leq x^2$ , we deduce that

$$|Q_{4,n,m}| \leq \frac{1}{N_n} \sum_{i=1}^{n-k_n} \left| \frac{\tilde{\bar{G}}_m(X_{i,n} \vee u) - \tilde{\bar{G}}(X_{i,n} \vee u)}{\tilde{\bar{G}}(X_{i,n} \vee u)} \right|^2 = o(1) \text{ a.s.}$$

**Term  $Q_{5,n,m}$ .**

This term can be rewritten as :

$$Q_{5,n,m} = \frac{1}{N_n} \sum_{i=n-k_n+1}^n \log \left( \frac{\frac{1}{m+1} + \frac{m}{m+1} \bar{G}_m(X_{i,n} \vee u)}{\frac{1}{n+1} + \frac{n}{n+1} \bar{G}(X_{i,n} \vee u)} \right).$$

Remark that

$$\frac{1}{m+1} \leq \frac{\frac{1}{m+1} + \frac{m}{m+1} \bar{G}_m(X_{i,n} \vee u)}{\frac{1}{n+1} + \frac{n}{n+1} \bar{G}(X_{i,n} \vee u)} \leq n+1$$

which implies that

$$|Q_{5,n,m}| \leq \frac{k_n}{n} \frac{1}{\bar{F}_n(u)} \max(\log(n+1), \log(m+1)) = o(1) \text{ a.s.}$$

**Term  $Q_{6,n,m}$ .**

Finally, remark that

$$Q_{6,n,m} = -\frac{1}{\bar{F}_n(u)} \log \left( \frac{\frac{1}{m+1} + \frac{m}{m+1} \bar{G}_m(u)}{\frac{1}{n+1} + \frac{n}{n+1} \bar{G}(u)} \right) = -\frac{1}{\bar{F}_n(u)} \log \left( \frac{m}{n} \frac{n+1}{m+1} \frac{\frac{1}{m} + \bar{G}_m(u)}{\frac{1}{n} + \bar{G}(u)} \right) = o(1) \text{ a.s.}$$

Combining all these results, Theorem 1 follows since

$$\hat{K}(f_u; g_u) - K(f_u; g_u) = -\left(\hat{L}(f_u; g_u) - L(f_u; g_u)\right) - \left(\hat{L}(g_u; f_u) - L(g_u; f_u)\right).$$

## References

- Abarca-Del-Rio, R. and O. Mestre (2006). Decadal to secular time scales variability in temperature measurements over france. *Geophysical Research Letters* 33, 1–4.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6), 716–723.
- Alexander, L. V., X. Zhang, T. Peterson, J. Cesar, B. Gleason, A. M. G. K. Tank, M. Haylock, D. Collins, B. Trewin, F. Rahimzadeh, A. Tagipour, K. R. Kumar, J. Revadekar, G. Griffiths, L. Vincent, D. B. Stephenson, J. Burn, E. Aguilar, M. Brunet, M. Taylor, M. New, P. Zhai, M. Rusticucci, and J. L. Vazquez-Aguirre (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research* 111.
- Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of extremes: Theory and applications*. John Wiley & Sons: New York.
- Burnham, K. P. and D. R. Anderson (1998). *Model Selection and Inference : A Practicle Information-Theoretical Approach*. Springer-Verlag.
- Chow, Y. and H. Teicher (1978). *Probability theory: independence, interchangeability, martingales*. Springer.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*.
- Davis, R., T. Mikosch, and I. Cribben (2012). Towards estimating extremal serial dependence via the bootstrapped extremogram. *Journal of Econometrics* 170, 142–152.
- de Haan, L. (1996). Von mises-type conditions in second order regular variation. *Journal of Mathematical Analysis and Applications* 197, 400–410.
- de Haan, L. and U. Stadtmüller (1996). Generalized regular variation of second order. *Journal of the Australian Mathematical Society* 61, 381–395.
- di Crescenzo, A. (1999). A probabilistic analogue of the mean value theorem and its applications to reliability theory. *Journal of Applied Probability* 36, 706–719.
- Dupuis, D. J. (2012). Modeling waves of extreme temperature: The changing tails of four cities. *Journal of the American Statistical Association* 107, 24–39.
- Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag.
- Ferreira, A. and L. de Haan (2012). The generalized pareto process; with application. *arXiv*.
- Fisher, R. A. and L. H. C. Tippett (1928). Limiting forms of the frequency distribution of the largest or the smallest member of a sample. *Proceedings of the Cambridge Philosophical Society* 24, 180–190.
- Fowler, H. J. and C. G. Kilsby (2003). A regional frequency analysis of united kingdom extreme rainfall from 1961 to 2000. *International Journal of Climatology* 23, 1313–1334.

- Frich, P., L. V. Alexander, P. Della-Marta, B. Gleason, M. Haylock, A. M. G. K. Tank, and T. Peterson (2002, January). Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climate Research* 19, 193–212.
- Giné, E. and J. Zinn (1984). Some limit theorems for empirical processes. *The Annals of Probability* 12(4), 929–989.
- Grigga, O. and J. Tawn (2012). Threshold models for river flow extremes. *Environmetrics* 23, 295–305.
- Hoang, T. H., S. Parey, and D. Dacunha-Castelle (2009). Multidimensional trends: The example of temperature. *European Physical Journal (Special Topics)* 174, 113–124.
- Jarušková, D. and M. Rencová (2008). Analysis of annual maximal and minimal temperatures for some european cities by change point methods. *Environmetrics* 19(3), 221–233.
- Katz, R., M. Parlange, and P. Naveau (2002). Extremes in hydrology. *Advances in Water Resources* 25, 1287–1304.
- Kharin, V. V., F. W. Zwiers, X. Zhang, and G. C. Hegerl (2007). Changes in temperature and precipitation extremes in the ipcc ensemble of global coupled model simulations. *Journal of Climate* 20, 1419–1444.
- Kullback, S. (1968). *Information theory and statistics*. Dover books on mathematics. Dover Publications.
- Penny, W. and S. Roberts (2000, November). Variational bayes for 1-dimensional mixture models. Technical report, Department of Engineering Science, Oxford University.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics* 3, 119–131.
- Shaby, B. and B. Reich (2013). Bayesian spatial extreme value analysis to assess the changing risk of concurrent high temperatures across large portions of european cropland. *Environmetrics In press*.
- Shaked, M. and J. G. Shanthikumar (1994). *Stochastic orders and their applications*. Academic Press (Boston).
- Wellner, J. A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 45, 73–88.
- Zwiers, F., X. Zhang, and Y. Feng (2011). Anthropogenic influence on long return period daily temperature extremes at regional scales. *Journal of Climate* 24, 881–892.