Applications of Extreme Value Theory to environmental data analysis

By GWLADYS TOULEMONDE

Institut de Mathématiques et de Modélisation de Montpellier - UMR 5149

Université Montpellier 2, Place Eugène Bataillon, 34090 Montpellier, France

gwladys.toulemonde@univ-montp2.fr

PIERRE RIBEREAU

Institut de Mathématiques et de Modélisation de Montpellier - UMR 5149

Université Montpellier 2, Place Eugène Bataillon, 34090 Montpellier, France

pierre.ribereau@univ-lyon1.fr

AND PHILIPPE NAVEAU

Laboratoire des Sciences du Climat et de l'Environnement

LSCE-IPSL-CNRS, 91191 Gif-sur-Yvette, France

philippe.naveau@lsce.ipsl.fr

SUMMARY

When analyzing extreme events, assuming independence in space and/or time may not correspond to a valid hypothesis in geosciences. The statistical modeling of such dependences

is complex and different modeling roads can be explored. In this article, some basic concepts about univariate and multivariate Extreme Value Theory will be first recalled. Then a series of examples will be treated to exemplify how this probability theory can help the practitioner to make inferences about extreme quantiles within a multivariate context.

1 Introduction - Univariate Extreme Value Theory

Extreme events are, almost by definition, rare and unexpected. Consequently it is very difficult to deal with them. Examples include the study of record droughts, annual maxima of temperature, wind and precipitation. Climate sciences is one of the main fields of applications of Extreme Value Theory (EVT) but we can also mention hydrology (e.g., Katz *et al.*, 2002), finance and assurance (e.g., Embrechts *et al.*, 1997) among others. Even if the probability of extreme events occurrence decreases rapidly, the damage caused increases rapidly and so does the cost of protection against them. The policymakers' summary of the 2007 Intergovernmental Panel on Climate Change clearly states that it *is very likely that hot extremes, heat waves, and heavy precipitation events will continue to become more frequent* and that *precipitation is highly variable spatially and temporally.*

From a probabilistic point of view, let consider a sample of n independent and identically distributed (i.i.d.) random variables (r.v.) $X_1, X_2, \ldots, X_n$ from a distribution function F. In the same way that we have the Central Limit Theorem (CLT) concerning the mean value of this sample, asymptotic results are available from EVT about the limit distribution of the rescaled sample's maximum value $X_{n,n} = \max_{i=1,\ldots,n} X_i$ as the sample size n increases. Indeed, according to the classical EVT (e.g., Embrechts *et al.*, 1997; Coles, 2001; Beirlant *et al.*, 2004; de Haan and Ferreira, 2006), the correctly rescaled sample's maximum is -under suitable conditions- asymptotically distributed according to one of the three extreme value distributions named Gumbel, Fréchet or Weibull.

More precisely if there exists sequences of constants $\{a_n\}$ and $\{b_n > 0\}$ and a non-degenerate distribution function G such that

$$\lim_{n\to+\infty} \mathbb{P}\left(\frac{X_{n,n} - a_n}{b_n} \leq x\right) = G(x)$$

then G belongs to one of the following families (with $\alpha > 0$):

I - Gumbel:

$$G(x) = \Lambda(x) = e^{-e^{-\frac{x-a}{b}}} \qquad \text{with } x \in \mathbb{R}$$

II - Fréchet:

$$G(x) = \Phi_\alpha(x) = \begin{cases} 0 & \text{if } x \leq a \\ e^{-\left(\frac{x-a}{b}\right)^{-\alpha}} & \text{if } x > a \end{cases}$$

III - Weibull:

$$G(x) = \Psi_\alpha(x) = \begin{cases} e^{-\left(-\left(\frac{x-a}{b}\right)\right)^{\alpha}} & \text{if } x \leq a \\ 1 & \text{if } x > a \end{cases}$$

A specificity of these three distributions is their max-stability property. Furthermore there are the only max-stable distributions. A distribution G is max-stable if G is invariant, up to affine transformations, i.e. up to location and scale parameters. In other words, we say that G is max-stable if there exists sequences $\{d_n\}$ and $\{c_n > 0\}$ such that, for all $n \geq 2$, the sample's maximum $X_{n,n}$ is equal in distribution to $c_n X + d_n$ with X following the same distribution G, what can be written as follows

$$G^n(x) = G\left(\frac{x - d_n}{c_n}\right).$$

From a statistical point of view, the interest of this fundamental theorem is limited. Indeed each situation corresponds to different tail behavior of the underlying distribution F. The Fréchet distribution corresponds to the limit of maxima coming from heavy tailed distributions like the Pareto distribution. The Weibull distribution is associated to distributions with a finite endpoint like the uniform distribution. The particular case of the Gumbel distribution has a special importance in EVT because it occurs as the limit of maxima from light tailed distributions for example from the Gaussian distribution. Moreover, empirically, the Gumbel distribution fits particularly well in a wide range of applications especially in atmospheric sciences.

In practice we have to adopt one of the three families but we don't have any information about F. That's why an unified approach would be very appreciated in order to characterize the limit distribution of maxima.

The previous theorem could be reformulated in an unified way using the Generalized Extreme Value (GEV) distribution. If there exists sequences of constants $\{a_n\}$ and $\{b_n > 0\}$ and a non-degenerate distribution function G such that

$$\lim_{n \to +\infty} \mathbb{P}\left(\frac{X_{n,n} - a_n}{b_n} \leq x\right) = G_{\mu,\sigma,\gamma}(x)$$

then $G_{\mu,\sigma,\gamma}$ belongs to the GEV family

$$G_{\mu,\sigma,\gamma}(x) = \exp\left(-\left[1 + \gamma\frac{x-\mu}{\sigma}\right]^{-1/\gamma}\right)$$

with $x \in \left\{z : 1 + \gamma\frac{x-\mu}{\sigma} > 0\right\}$.

It is easy to remark that $G_{\mu,\sigma,\gamma}$ merges all univariate max-stable distributions previously introduced. It depends on an essential parameter γ characterizing the shape of the F-distribution tail. Since a strictly positive γ corresponds to the Fréchet family, this case corresponds to heavy-tailed distributions. Otherwise a strictly negative γ is associated to the Weibull family. For γ tends to 0, the function $G_{\mu,\sigma,\gamma}$ tends to the Gumbel one.

Practically, in order to assess and predict extreme events, one often works with so-called block maxima, i.e. with the maximum value of the data within a certain time interval including k observations. The maximum can be assumed to be GEV distributed in the case where k is large enough. If we obtain a sufficient number of maxima and if these maxima can be considered as an i.i.d. sample, estimation values for the GEV unknown parameters can be obtained with maximum likelihood procedure or Probability Weighted Moments (PWM) for instance. The asymptotic behaviors of these estimators have been established (Smith, 1985; Hosking *et al.*, 1985; Diebolt *et al.*, 2008).

In the first considered example, we dispose of temperature daily maxima during 30 years in Colmar, a city in the east of France. We consider monthly maxima for the summer months: June, July and August (see Figure 1). By this way we avoid a seasonality in the data.

The choice of the block size denoted in the sequel by r - such as a year or a month - can be justified in many cases by geophysical considerations but this choice has actual consequences on estimation tools. In an environmental context, if we can often consider annual maxima as an i.i.d. sample, this hypothesis is stronger when we deal with monthly or weekly maxima for instance. Indeed in these latter cases we have typically a seasonality in the data so we are in a non stationary context.

Coming back to our application, when the GEV distribution ($G_{\mu,\sigma,\gamma}$) is fitted by maximum likelihood directly on the sample of $k = 90$ monthly maxima on summer (3*30 years), we obtain $\widehat{\mu} = 31.502$, $\widehat{\sigma} = 2.456$ and $\widehat{\gamma} = -0.256$. The corresponding 95%-Confidence intervals are $[30.943; 32.060]$, $[2.058; 2.853]$ and $[-0.383; -0.128]$. The quantile plot in Figure 2 is close to the unit diagonal indicating a good fit of data (*Empirical*) by a GEV distribution (*Model*). The results on the GEV fitting must be taken with care. Indeed, taking the maxima only on 30 measures may lead to unexpected results (see de Haan and Ferreira, 2006). Here, the limiting distribution of the maxima may be the Gumbel one even if the estimation of γ is

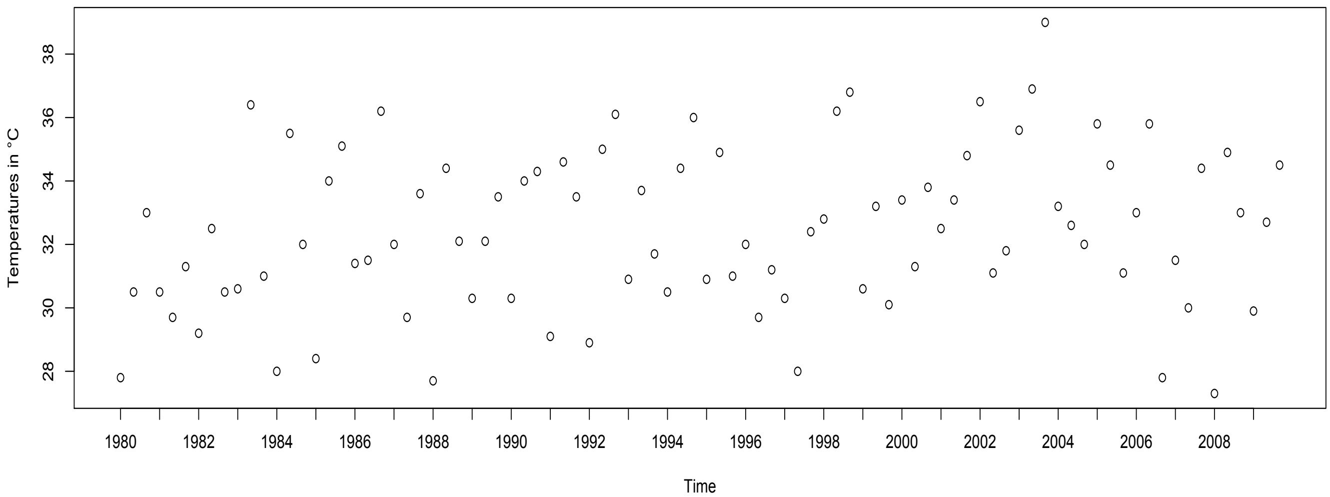

Figure 1: The y-axis corresponds to temperature maxima (in $^\circ C$) for the months of June, July and August from 1980 to 2009 (x-axis) recorded at Colmar (France).

negative (for example, if the original variables are gaussian distributed, the effective gamma approximately $-1/(2\log k)$ where k is the block size).

A question of interest in this kind of application concerns the estimation of extreme quan-

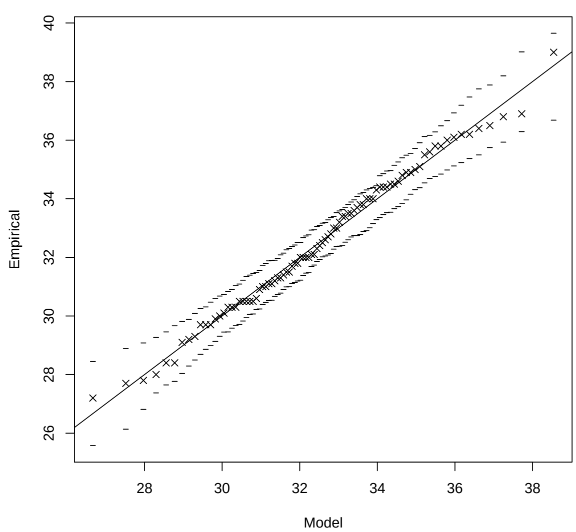

Figure 2: Quantile Plot for the GEV distribution - Colmar data.

tiles of the maxima distribution on a period (a block). In other words, we are looking for $z_{\frac{1}{p}}$ such that $\mathbb{P}\left(X_{r,r} \leq z_{\frac{1}{p}}\right) = 1 - p$.

Since we have $\mathbb{P}\left(X_{k,k} \leq z_{\frac{1}{p}}\right) \approx G_{\mu,\sigma,\gamma}\left(z_{\frac{1}{p}}\right)$, we obtain

$$\begin{aligned} z_{\frac{1}{p}} &\approx \mu - \frac{\sigma}{\gamma}\left[1 - (-\log(1-p))^{-\gamma}\right] \quad \text{if } \gamma \neq 0 \\ &\approx \mu - \sigma \log(-\log(1-p)) \quad \text{if } \gamma = 0. \end{aligned}$$

The quantity z_T is called return level associated to a return period $T = \frac{1}{p}$. The level z_T is expected to be exceeded on average once every $T = \frac{1}{p}$ blocks (e.g. months).

We use the following estimator $\widehat{z}_T$:

$$\begin{aligned} \widehat{z}_T &= \widehat{\mu} - \frac{\widehat{\sigma}}{\widehat{\gamma}}\left[1 - (-\log(1-p))^{-\widehat{\gamma}}\right] \quad \text{if } \gamma \neq 0 \\ &= \widehat{\mu} - \widehat{\sigma} \log(-\log(1-p)) \quad \text{if } \gamma = 0. \end{aligned}$$

Starting from the asymptotic behavior of the GEV parameters vector estimator, it is possible to deduce thanks to the δ-method (e.g., van der Vaart, 1988) the asymptotic behavior of $\widehat{z}_T$ leading us to associated confidence intervals.

Coming back to our example, we would like to compute the return level associated to the return period $T = 3 \times 50$ months corresponding to 50 years as we consider only 3 months a year. We easily obtain a return level $\widehat{z}_{150} = 38.43$ and a 95%-Confidence interval $[37.107; 39.758]$. That means that considering only summer periods, the temperature $38.43°C$ is expected to be exceeded on average once every 50 years. So thanks to EVT we are able to estimate a return level corresponding to 50 years whereas we only consider data since 30 years.

We have supposed that maxima constitute an i.i.d. sample which is reasonable with regards of Figure 1. A likelihood ratio test indicates no trend in our data. To support this aim we fit a GEV distribution with a linear trend in the localisation parameter ($G_{\mu(t),\sigma,\gamma}$). We obtain with the likelihood procedure $\widehat{\mu}(t) = 30.635 + 0.059\,t$, $\widehat{\sigma} = 2.439$ and $\widehat{\gamma} = -0.281$. The corresponding likelihood ratio test statistic is $D = 2 \times (209.82 - 207.91) = 3.82$. This value is small when compared to the χ^2_1 distribution, suggesting that the simple model (without trend) is adequate for these data. Other estimation procedures have been developed in the non-stationary context. For example Maraun *et al.* (2010) have proposed various models to describe the influence of covariates (possible non linearities in the covariates and seasonality) on UK daily precipitation extremes. In the same way, Ribereau *et al.* (2008) extend the PWM method in order to take into account temporal covariates and provide accurate GEV-based return levels. This technique is particularly adapted for small sample sizes and permits for example to consider seasonality in data.

If block sizes are sufficiently large and even if the stationary hypothesis is not always satisfied, the independence one remains very often valid. On the other hand, even if the series could be considered as stationary, the independence hypothesis is not always satisfied if we consider too small block size like daily maxima.

As an example, Figure 3 represents a series of daily maxima of CO2 (in part per million), a greenhouse gas recorded from 1981 to 2002 in Gif sur Yvette, a city of France. The trend in the series has been removed in order to consider a stationary series.

This example illustrates the connection between light-tailed maxima and the Gumbel distribution. Indeed, if we fit a Gumbel distribution on the daily CO2 maxima, we see on the Gumbel Quantile Quantile (QQ-)plot (see left part of Figure 4) that it is quite reasonable to suppose that these data are Gumbel distributed (with $\widehat{\mu} = -0.44$ and $\widehat{\sigma} = 0.76$). But in this practical case the length of our observations is too short to study yearly maxima, or even

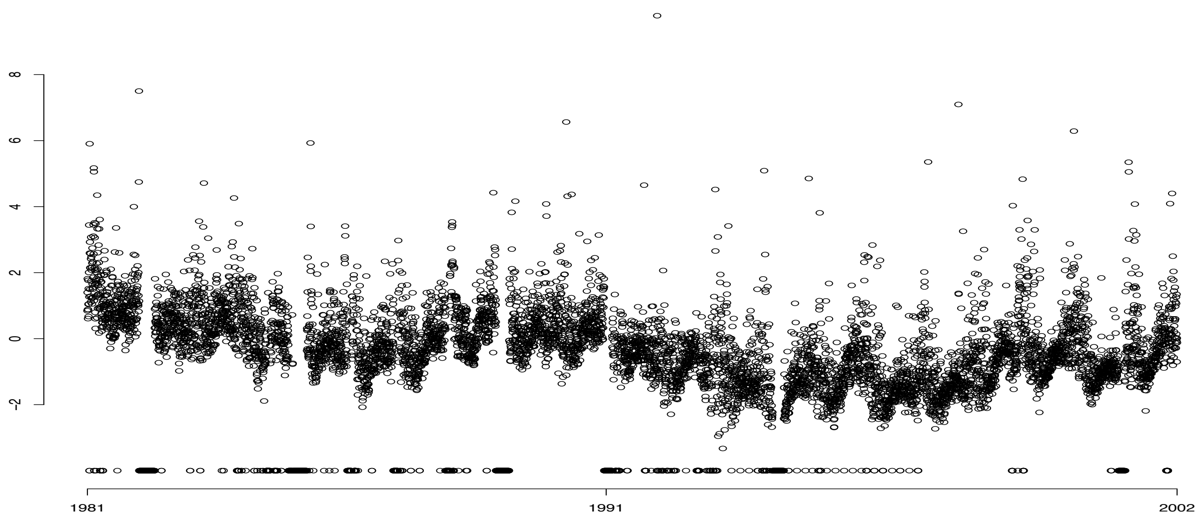

Figure 3: Series of daily maxima of carbon dioxide.

monthly maxima. As a consequence, when studying daily maxima it is natural to observe some day-to-day memories. The scatter plot of successive values (see right part of Figure 4) confirms this short-term temporal dependence.

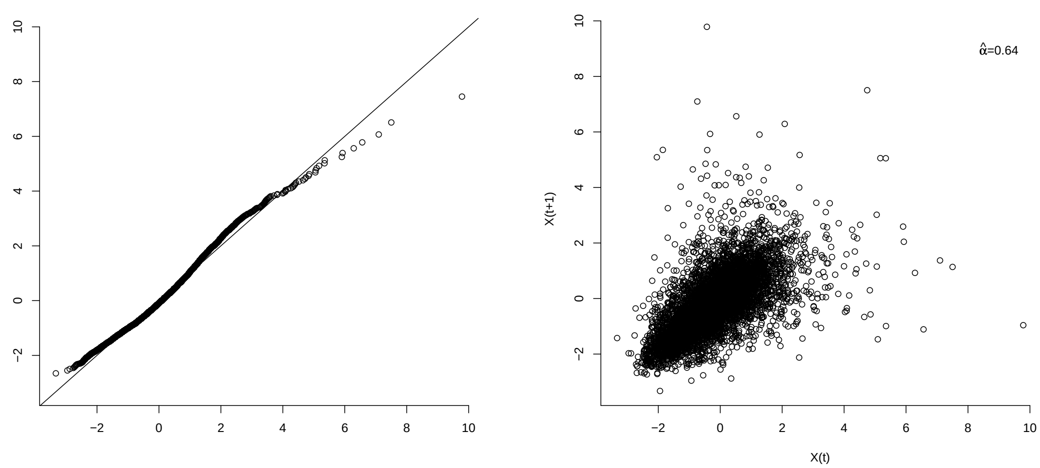

Figure 4: Gumbel QQplot (left) and scatter plot of successive values (right) corresponding to daily maxima of carbon dioxide.

A very simple approach in the time series analysis would be to consider linear autoregressive (AR) models. Classical hypotheses are noise Gaussianity and model linearity. But in an extreme value context, if X_t is a maximum, then one expects X_t to follow a GEV distribution and it is impossible to satisfy this distributional constraint with a gaussian additive AR process.

That's why Toulemonde *et al.* (2010) have proposed a linear AR process adapted to the Gumbel distribution. This model is based on an additive relationship between Gumbel and positive α-stable variables[1] (see for example Crowder, 1989; Hougaard, 1986; Tawn, 1990; Fougères *et al.*, 2009) and is defined as follows.

First, let consider $S_{t,\alpha}$ being i.i.d. positive asymmetric α-stable r.v. defined by its Laplace transform for any $t \in \mathbb{Z}$

$$\mathbb{E}(\exp(-uS_{t,\alpha})) = \exp(-u^{\alpha}), \text{ for all } u \geq 0 \text{ and for } \alpha \in]0,1[. \tag{1}$$

Let $\{X_t, t \in \mathbb{Z}\}$ be a stochastic process defined by the recursive relationship

$$X_t = \alpha\ X_{t-1} + \alpha\ \sigma\ \log S_{t,\alpha} \tag{2}$$

where $\sigma > 0$. It has been proved first that Equation (2) has a unique strictly stationary solution,

$$X_t = \sigma \sum_{j=0}^{\infty} \alpha^{j+1} \log S_{t-j,\alpha} \tag{3}$$

and secondly that X_t follows a Gumbel distribution with parameters $(0, \sigma)$.

This model is a linear AR model and has consequently the associated advantages such that their conceptual simplicity and their flexibility for modeling quasi-periodic phenomena (e.g. sunspots time series) and short-term dependencies (e.g. day-to-day memories in weather systems). Moreover, whereas one drawback of current linear AR models is that they are

[1] A random variable S is said to be stable if for all non-negative real numbers c_1, c_2, there exists a positive real a and a real b such that $c_1S_1 + c_2S_2$ is equal in distribution to $aS + b$ where S_1, S_2 are i.i.d. copies of S.

unable to represent the distributional behavior of maxima, a key point of parameterization (2) is that X_t follows a Gumbel distribution. In other words this process is suitable for maxima data coming from light-tailed distributions. Even if our process is specific to the Gumbel distribution, it is possible to extend it for maxima coming from bounded or heavy tailed distribution. Nevertheless this leads to a process which is not additive anymore.

Coming back to our example, identifying the temporal structure among the largest CO_2 measurements is of primary interest for the atmospheric chemist because this can help to predict future maxima of CO_2 at a specific location. As an illustration, Figure 5 presents one-step previsions of CO_2 daily maxima on the year 2001.

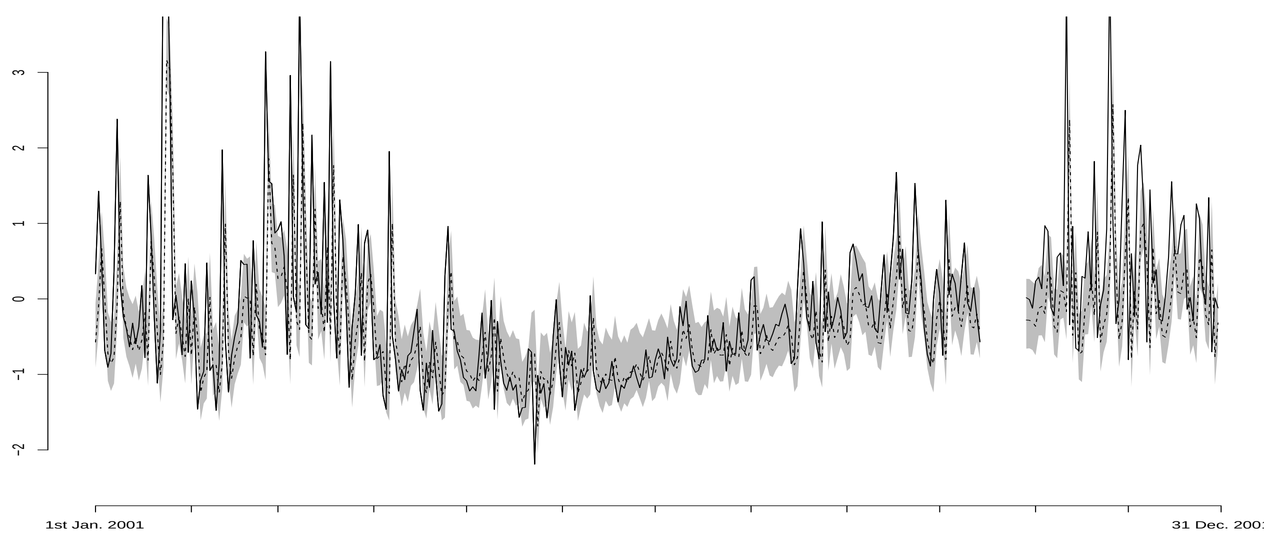

Figure 5: One-day previsions of daily maxima of CO_2 (y-axis) on the year 2001 (x-axis). The black line corresponds to the observed series and the dotted line corresponds to the estimated series (median). The grey area is delimited by the first and third empirical quartiles.

This method, presented in Toulemonde et al. (2010), is exemplified in their paper on daily maxima series of two other greenhouse gas, the methane (CH_4) and the oxide nitrous (N_2O) recorded at LSCE (Laboratoire des Sciences du Climat et de l'Environnement) in Gif-sur-

Yvette (France). Since the beginning of 2007, the LSCE has proceeded to record daily maxima of CH_4 but has stopped the regular recordings of N_2O daily maxima. That's why in a recent paper, Toulemonde et al. (2012) proposed a method adapted to maxima from light-tailed distribution able to reconstruct a hidden series. In this state-space context, they take advantage of particle filtering methods. In an extreme adapted model, they compute optimal weights for the use of the auxiliary filter (Pitt and Shepard, 1999) and they denote this filter by APF-Opt. Basing on observations of CH_4 and N_2O from 2002 to the middle of 2006, they obtain similar results than those presented in Figure 6 for the reconstruction of N_2O daily maxima.

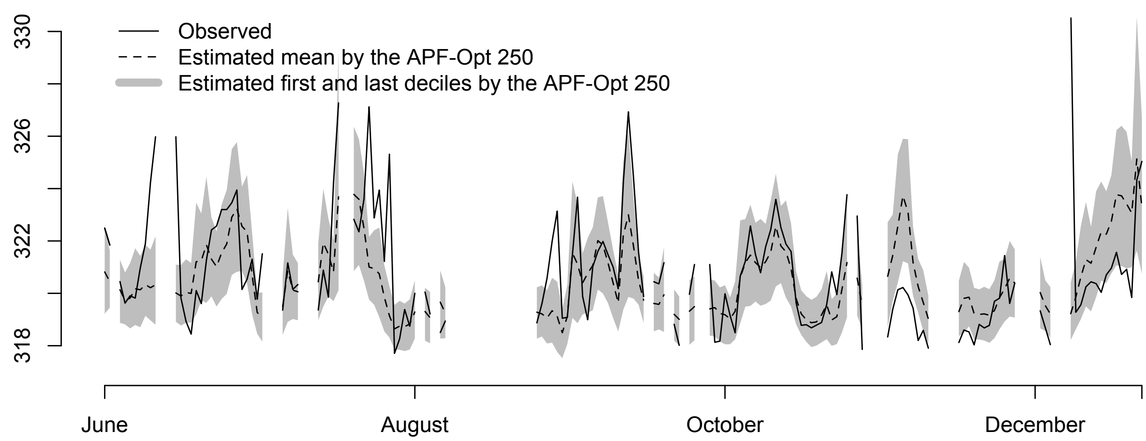

Figure 6: Mean values of particles from APF-Opt and punctual empirical $IC_{80\%}$ with 250 particles for the series of N_2O daily maxima in Gif-sur-Yvette from June to December 2006.

Concerning inference procedure, as usually in the block maxima approach, only the maxima are used. To remove this drawback, another technique consists in modeling exceedances above a given threshold u. In the so-called Peaks-over-Threshold (PoT) approach, the distributions of these exceedances are also characterized by asymptotic results.

Let $X_1, \ldots, X_n$ a sample of n i.i.d. r.v. from a distribution function F. We consider the N_u

of them which are over the threshold u. The exceedance Y_i corresponding to the variable X_i is defined by $X_i - u$ if $X_i > u$.

The distribution function F_u of an exceedance Y over a threshold u is given for $y > 0$ by

$$\begin{aligned} F_u(y) &= \mathbb{P}(Y \leq y | X > u) = \mathbb{P}(X - u \leq y | X > u) \\ &= \frac{\mathbb{P}(u < X \leq u + y)}{\mathbb{P}(X > u)} = \frac{F(u+y) - F(u)}{1 - F(u)}. \end{aligned}$$

If the threshold is sufficiently high, we can approximate this quantity by the distribution function of the generalized Pareto distribution $H_{\gamma,\sigma}(y)$. We defined its survival function as follows

$$\begin{aligned} \overline{H}_{\gamma,\sigma}(y) = &\quad \left(1 + \gamma \tfrac{y}{\sigma}\right)^{-1/\gamma} \quad \text{if } \gamma \neq 0 \\ = &\quad \exp\left(-\tfrac{y}{\sigma}\right) \quad \text{otherwise.} \end{aligned}$$

This function is defined on $\mathbb{R}^+$ if $\gamma \geq 0$ or on $[0; -\sigma/\gamma[$ if $\gamma < 0$ where $\sigma > 0$ is a scale parameter and $\gamma \in \mathbb{R}$ a shape parameter.

The famous theorem of Pickands (1975) establishes the following equivalence:

$$\lim_{n \to +\infty} \mathbb{P}\left(\frac{X_{n,n} - a_n}{b_n} \leq x\right) = G_{\mu,\sigma,\gamma}(x)$$

if and only if

$$\lim_{u \to x_F} \sup_{y \in [0; x_F - u]} \left|\overline{F}_u(y) - \overline{H}_{\gamma,\sigma(u)}(y)\right| = 0.$$

It is important to mention that the shape parameter in the block maxima approach coincides with the shape parameter in the PoT approach.

Coming back to extreme quantiles estimation, we have for $x_p \geq u$, the following approximation

$$
\begin{aligned}
p = \overline{F}(x_p) &\approx \overline{F}(u)\left[1+\gamma\frac{x_p-u}{\sigma}\right]^{-\frac{1}{\gamma}} \quad \text{if } \gamma \neq 0 \\
&\approx \overline{F}(u)\exp\left(-\frac{x_p-u}{\sigma}\right) \quad \text{if } \gamma = 0
\end{aligned}
$$

which implies

$$
\begin{aligned}
x_p = \overline{F}^{-1}(p) &\approx u + \frac{\sigma}{\gamma}\left(\left[\frac{p}{\overline{F}(u)}\right]^{-\gamma} - 1\right) \quad \text{if } \gamma \neq 0 \\
&\approx u - \sigma\log\left(\frac{p}{\overline{F}(u)}\right) \quad \text{if } \gamma = 0.
\end{aligned}
$$

By construction, x_p is the return level associated to the $\frac{1}{p}-$observation. In other words this level x_p is expected to be exceeded on average once every $\frac{1}{p}$-observations.

If we are interested on a return level associated to the a return period T blocks (e.g. months) denoted by z_T and supposing we have r observations per block (e.g. month), we will consider $p = \frac{1}{rT}$.

Finally, a PoT estimator $\widehat{z}_T$ of the return level T-block with r observations per block is given by

$$
\begin{aligned}
\widehat{z}_T = \widehat{\overline{F}}^{-1}\left(\frac{1}{rT}\right) &= u + \frac{\widehat{\sigma}}{\widehat{\gamma}}\left(\left[\frac{n}{rTN_u}\right]^{-\widehat{\gamma}} - 1\right) \quad \text{if } \gamma \neq 0 \\
&= u - \widehat{\sigma}\log\left(\frac{n}{rTN_u}\right) \quad \text{if } \gamma = 0.
\end{aligned}
$$

The choice of the threshold is difficult and is clearly a question of trade-off between bias and variance for the estimation of the parameters (γ, σ). There is no perfect solution but only tools based for example on the mean Excess Function (MEF) which help us to make a choice. Again, as in the block maxima approach, the GPD approximation for excesses is asymptotic and results obtained on finite sample size should be considered carefully.

2 Multivariate Approach

The probabilistic foundations for the statistical study of multivariate extremes are well developed. Since the classical work of Resnick (1987), many books (see for example Beirlant *et al.*, 2004; de Haan and Ferreira, 2006 and Resnick, 2007) have paid considerable attention to this particular case.

We will focus here on analog of block maxima results discussed in previous sections for univariate extremes. Suppose that $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ is a sequence of i.i.d. random vectors with same common distribution function F. Examples of such variables are Maximum and Minimum temperatures or precipitations in two locations. As in the univariate case, the characterization of the behavior of extremes in a multivariate context is based on the block maxima. Denote

$$M_{x,n} = \max_{i=1,\dots,n} \{X_i\} \qquad \text{and } M_{y,n} = \max_{i=1,\dots,n} \{Y_i\}$$

$$\mathbf{M_n} = (M_{x,n}, M_{y,n}) \, .$$

$\mathbf{M_n}$ is the vector of component-wise maxima. Note that the maximum of the X_i can occur at a different time than the one of the Y_i, so M_n does not necessarily correspond to an observed vector. The multivariate EVT begins with the study of the M_n behavior. If $\mathbf{z} = (z_1, z_2)$ and

$\mathbf{M_n} \leq \mathbf{z}$ meaning that $M_{x,n} \leq z_1$ and $M_{y,n} \leq z_2$, we have

$$\mathbb{P}(\mathbf{M_n} \leq \mathbf{z}) = F(\mathbf{z})^n.$$

We assume that there exists two rescaling sequences of vector $\{\mathbf{a}_n\}$ and $\{\mathbf{b}_n\}$ where $a_{n,j} > 0$ and $b_{n,j} \in \mathbb{R}$ for $j = 1, 2$ and G a distribution function with non-degenerate margins such that

$$F^n(\mathbf{a}_n\mathbf{z} + \mathbf{b}_n) \to G(\mathbf{z})$$

where $\mathbf{a}_n\mathbf{z}+\mathbf{b}_n = (a_{n,1}z_1+b_{n,1}, a_{n,2}z_2+b_{n,2})$. If such sequences exist, G is a bivariate extreme value distribution. With the same notations, another consequence is that

$$G^k(\mathbf{a}_k\mathbf{z} + \mathbf{b}_k) = G(\mathbf{z})$$

that is G is max stable. The problem of the limiting distribution G is partially solved considering separately $(X_i)_i$ and $(Y_i)_i$ since the univariate EVT can be applied directly.

2.1 Characterization theorem

As the margins of the limiting distribution are GEV distributed, that means we can get easier representation by assuming that the margins are known. Simple representations arise when assuming that both X_i and Y_i are unit Fréchet distributed with distribution function $G_{1,1,1}(z) = \exp(-1/z)$.

Once the margins transform to unit Frechet, we should consider the re-scaled vector

$$\mathbf{M_n^*} = \left(M_{x,n}^*, M_{y,n}^*\right) = \left(M_{x,n}/n, M_{y,n}/n\right),$$

in order to obtain standard univariate results for each margin.

THEOREM 1 *Let (X_i, Y_i) be i.i.d. random vectors with unit Frechet marginals and define $\left(M_{x,n}^*, M_{y,n}^*\right)$ as previous. If :*

$$\mathbb{P}\left(M_{x,n}^* \leq x, M_{y,n}^* \leq y\right) \longrightarrow G(x,y)$$

where G is a non degenerated distribution, then G is of the form :

$$G(x,y) = \exp\left(-V(x,y)\right), \quad x > 0, y > 0$$

where

$$V(x,y) = 2\int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) dH(\omega)$$

and H is a distribution on $[0,1]$ verifying the following mean constraint :

$$\int_0^1 \omega dH(\omega) = 1/2. \tag{4}$$

For example, if H is such that :

$$H(\omega) = \begin{cases} 1/2 & \text{if } \omega\text{=}0 \text{ or } 1 \\ 0 & \text{else} \end{cases}$$

the corresponding bivariate extreme value distribution is

$$G(x,y) = \exp\left\{-(x^{-1} + y^{-1})\right\}$$

for $x > 0$ and $y > 0$. This distribution is a product of a function of x and another of y and therefore corresponds to the independence. Another interesting case of distribution H is a measure that place mass equal to 1 in 0.5. In that case, the bivariate extreme value distribution is

$$G(x, y) = \exp\left(-\max\left\{x^{-1}, y^{-1}\right\}\right)$$

for $x > 0$ and $y > 0$. It is the special case of variables X and Y which are unit Fréchet distributed and perfectly dependent i.e. $X = Y$ a.s.

As any function verifying the mean constraint defines a bivariate extreme distribution, there is a one-to-one relation between the set of bivariate extreme distributions with unit Fréchet margins and the set of distributions on $[0, 1]$ satisfying (4). So any parametric family for H satisfying (4) defines a class of bivariate extreme value distribution.

One classical family is the logistic one. In that case, we have :

$$G(x, y) = \exp\left(-\left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}\right)$$

for $x > 0$ and $y > 0$ and for $\alpha \in (0, 1)$. The popularity of this family is its simplicity and its flexibility. Indeed, as $\alpha \to 1$, it is easy to check that we get the independence. In contrast, if $\alpha \to 0$ we get the perfect dependence. So the logistic family covers all levels of dependence but the model is limited since the variables x and y are exchangeable because of the symmetry of the density h. In order to avoid this limitation, there exists two generalizations of this model.

The first one is the asymmetric logistic family for which we have

$$G(x, y) = \exp\left\{-\left(1 - t_1\right) x^{-1} - \left(1 - t_2\right) y^{-1} - \left[\left(\frac{x}{t_1}\right)^{-1/\alpha} + \left(\frac{y}{t_2}\right)^{-1/\alpha}\right]^{\alpha}\right\}$$

where $0 < \alpha \leq 1$ and $0 \leq t_1, t_2 \leq 1$. The parameter α controls the dependence while t_1 and t_2 control asymmetry. When $t_1 = t_2 = 1$ we get the logistic family.

As $\alpha \to 1$ or as t_1 or t_2 equal to 0, we get the independence, while if $\alpha \to 0$ and $t_1 = t_2 = 1$, we have the perfect dependence.

The second generalization is the bilogistic model defined by :

$$G(x,y) = \exp\left(-x^{-1}q^{1-\alpha} - y^{-1}(1-q)^{1-\beta}\right)$$

where $q = q(x, y, \alpha, \beta)$ is the solution of the following equation:

$$(1-\alpha)x^{-1}(1-q)^{\beta} - (1-\beta)y^{-1}q^{\alpha} = 0,$$

and $0 < \alpha, \beta < 1$. When $\alpha = \beta$, the bilogistic family reduces to the logistic class. The complete dependence arises when $\alpha = \beta$ tends to 0 while independence is obtained when $\alpha = \beta$ tends to 1 or when one of the two parameters is fixed and the other tends to 1.

2.2 *Other representations of bivariate extremes*

We can obtain other kind of representations of bivariate extreme value distribution. For example, the following theorem presents a point process approach. Let the set E denote here $E = [0,\infty]^2 \backslash \{0\}$, $\|.\|$ any norm of $\mathbb{R}^2$ and $B \subset E$ the associated unit sphere.

THEOREM 2 *The following assertions are equivalent :*

- *G is a bivariate extreme value distribution with unit Fréchet margins as mentioned in Theorem 1.*
- *There exists a non homogeneous Poisson process on $[0,\infty) \times E$ with intensity Λ defined*

for $t > 0$ *by* $\Lambda([0,t] \times B) = t\mu^*(B)$*, where for all* $A \subset B$ *and* $r > 0$*,*

$$\mu^* \left(\mathbf{x} \in E \ : \ \|\mathbf{x}\| > r; \frac{\mathbf{x}}{\|\mathbf{x}\|} \in A \right) = 2\frac{H(A)}{r}, \tag{5}$$

where $\mathbf{x} = (x_1, x_2)$ *and* H *is a finite measure such that (4) holds and*

$$G(\mathbf{x}) = \exp\left(-\mu^* \left\{(x_1, \infty) \times (x_2, \infty)\right\}\right).$$

This last representation gives an interesting interpretation of the distribution H. Let $\|\mathbf{x}\| = x_1 + x_2$, the transformation used in (5) : $\mathbf{x} = (x_1, x_2) \to (x_1 + x_2, x_1/(x_1 + x_2)) = (r, \omega)$ is a transformation from cartesian to pseudo-polar coordinates, in which r is a measure of distance from origin and ω measures angle on a $[0,1]$ scale. It is easy to check that the case $\omega = 0$ corresponds to the case $x_1 = 0$ and $\omega = 1$ to the case $x_2 = 0$. Equation (5) implies that the measure μ^* is a product measure of a simple function of the radial component and a measure H of the angular component. In other words, the angular spread is determined by H and is independent of the radial distance.

Interpretation in the case that H is differentiable with density h is easier: since ω measures the direction of the extremes, $h(\omega)$ measures the relative frequency of extremes in this direction. With this representation, it is easy to understand what was previously mentioned : when $h(\omega)$ is large for values of ω close to 0 and 1 , we tend to the independence because large values of x correspond to small values of y and vice versa. On the contrary, if the dependence is very strong, large values of x will correspond to large values of y, so $h(\omega)$ will be very large for values of ω close to $1/2$ and small elsewhere.

In dimension 2, there exists a popular approach of the bivariate extreme value distribution based on the dependence function.

THEOREM 3 (Representation using Pickands dependence function) *G is an extreme bivariate distribution with unit Fréchet marginals if and only if*

$$G(x,y) = \exp\left(-\left(\frac{1}{x}+\frac{1}{y}\right)A\left(\frac{x}{x+y}\right)\right)$$

where $A(.)$ is a convex function on $[0,1]$ in $[1/2,1]$ such that

$$\max(t, 1-t) \leq A(t) \leq 1$$

for all t in $[0,1]$ and

$$A(0) = A(1) = 1$$

$$-1 \leq A'(0) \leq 0 \quad \textit{and} \quad 0 \leq A'(1) \leq 1$$

$$A''(t) \geq 0.$$

The form of the Pickands function provides important informations on the dependence between marginals :

- if $A(t) = 1$ we get the independence in the extremes,
- if $A(t) = \max(t, 1-t)$, we have the complete and perfect dependence.

Obviously, there is a link between the function A and the measure H :

$$A(u) = 2\int_0^1 \max\left(u(1-\omega), (1-u)\omega\right) dH(\omega).$$

Other methods can be implemented to model the bivariate behavior between two consecutive maxima. The copula approaches (e.g., Joe, 1997; Gudendorf and Segers, 2010) allow to construct bivariate distributions under the assumption that all marginals are identified.

We can also mention Naveau *et al.* (2009) and Bacro *et al.* (2010) who take advantage of bivariate EVT, i.e. they choose and estimate a bivariate extremal dependence function.

2.3 Inference and estimation

There exists a wide range of bivariate extreme distribution families. For example, in the `R` package `evd` (Stephenson, 2002), there are 8 different classes.

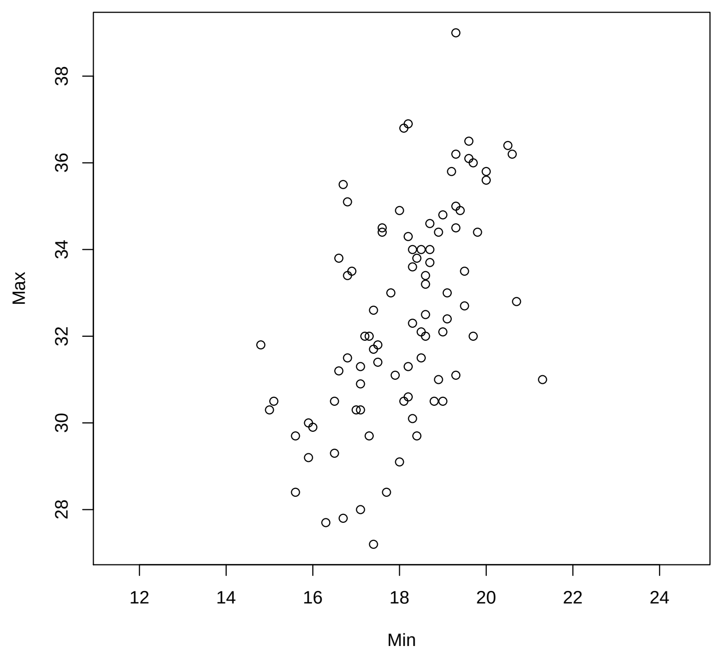

Figure 7: Summer Maxima of minimum daily temperature (x) and maximum daily temperature (y) of the Colmar data

The Figure 7 shows the summer maximum of minimum daily temperature against the corresponding maximum daily temperature. Obviously, there seems to be a trend for large values of the minimum temperature to correspond with large maximum temperature.

In order to better visualize the dependence in our data, it is possible to proceed to a transfor-

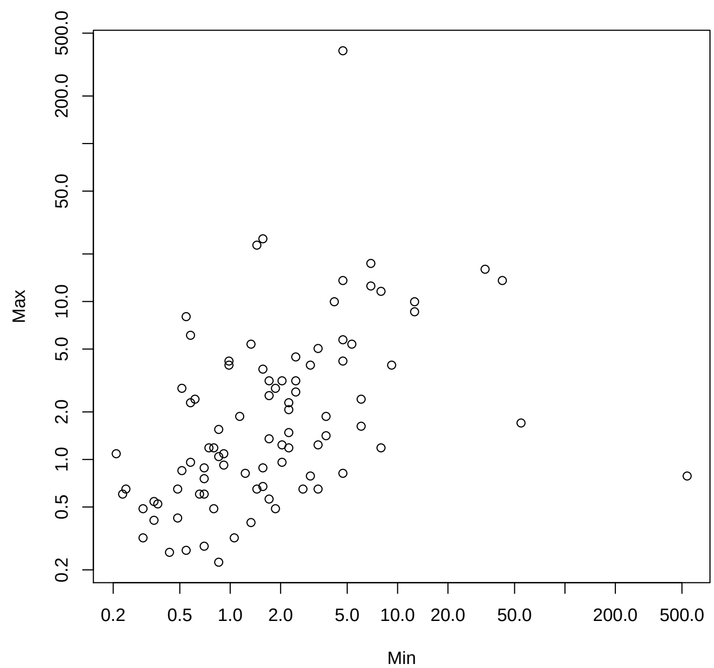

Figure 8: Summer Monthly Maxima of minimum daily temperature (x) and maximum daily temperature (y) with a Fréchet margins transformation (logarithm scale).

	μ_x	σ_x	γ_x	μ_y	σ_y	γ_y	α
Estimates	17.59	1.37	-0.28	31.56	2.41	-0.21	0.63
Standard Error	0.16	0.11	0.06	0.28	0.19	0.07	0.06

Table 1: Results of fitting a logistic bivariate extreme value distribution to the Colmar data. Values given are maximum likelihood estimates of the GEV parameters of both margins and maximum likelihood estimates of α.

mation of the margins. A usual choice consists in transforming the two margins distributions into unit Fréchet distribution. If we fit a GEV distribution to the x_i, we get the following estimates

$$\hat{\mu} = 17.62, \hat{\sigma} = 1.38, \hat{\gamma} = -0.33$$

while on the y_i the estimates are

$$\hat{\mu} = 31.59, \hat{\sigma} = 2.41, \hat{\gamma} = -0.25.$$

Here again, the approximation is made on a finite sample and may lead to errors on the dependence structure estimation of the model. This unit Fréchet margins transformation leads to the representation in Figure 8. The sample seems symmetric so a logistic model could be appropriated. Even if, for sake of simplicity, we have presented in Section 2.1 the logistic model with common unit Fréchet margins, practically, it is straightforward to estimate jointly the six GEV margins parameters and the dependence parameter α using maximum likelihood. The Table 1 represents the corresponding results.

The value of the dependence parameter α estimation is equal to 0.63 with an asymptotic confidence interval of $[0.51, 0.75]$ which corresponds to the first impression one can have looking at Figure 8 : reasonably weak level of dependence but significatively different from independence. The maximized log-likelihood is equal to -344.27.

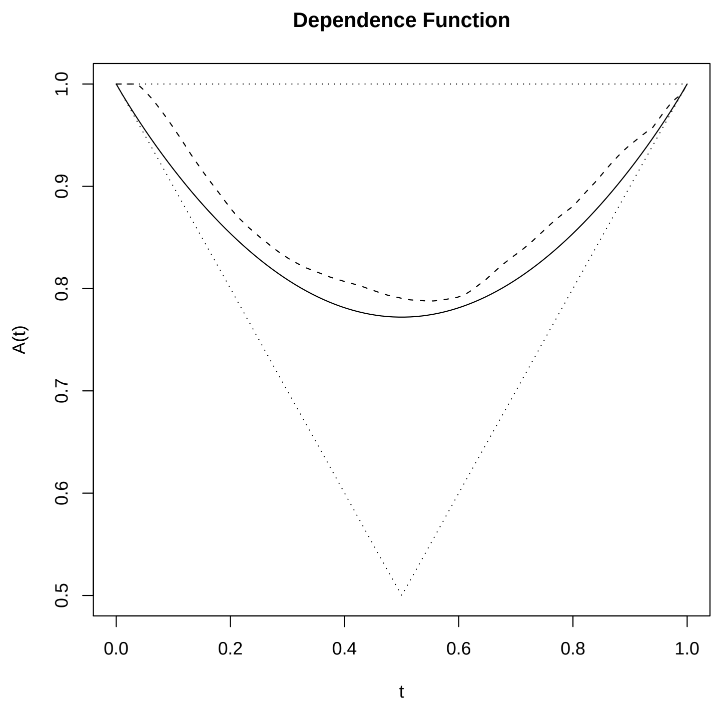

Figure 9: Nonparametric estimate of the Pickands dependence function for the Colmar data (dashed line) and the Pickands dependence function corresponding to the fitted logistic model (full line).

	μ_x	σ_x	γ_x	μ_y	σ_y	γ_y	α	β
Estimates	17.59	1.37	-0.28	31.54	2.40	-0.20	0.64	0.60
Standard Error	0.14	0.10	0.05	0.23	0.17	0.06	0.09	0.09

Table 2: Results of fitting a bilogistic bivariate extreme value distribution to the Colmar data. Values given are maximum likelihood estimates of the GEV parameters of both margins and maximum likelihood estimates of α.

According to the Theorem 3, the Pickands dependence function is a convex function and their theoritical borders are represented with dotted lines in Figure 9 and 10. The Pickands dependence function corresponding to the fitted logistic model (Table 1) is represented in Figure 9 with the full line. A non-parametric estimate of the function is also presented with the dashed line. This figure seems to indicate that the fitting of the data may be improved by using a more complex model. That's why we represent in Figure 10 the Pickands dependence function corresponding to the fitted bilogistic model (Table 2).

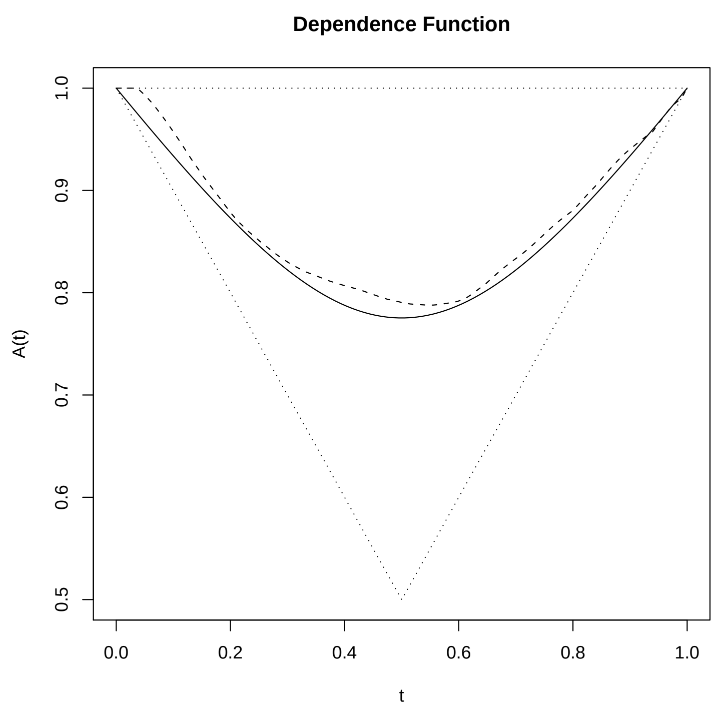

Figure 10: Nonparametric estimate of the Pickands dependence function for the Colmar data (dashed line) and the Pickands dependence function corresponding to the fitted bilogistic model (full line).

Since the logistic model is a subset of the bilogistic model, we can apply a deviance test to choose the model. For the bilogistic model the maximized log-likelihood is equal to -344.24. The Deviance statistics is then equal to 0.06 which is very small respect to the 95% quantile of the χ_1^2 distribution. The benefit brought by the asymetric logistic model is not sufficient. This is expected because of the estimation of α and β, the case $\alpha = \beta$ corresponding to the logistic model.

Finally, the Figure 11 represents the α quantile curves for the fitted logistic model for $\alpha = 0.7$, 0.8 and 0.9 generalizing quantiles to the bivariate case.

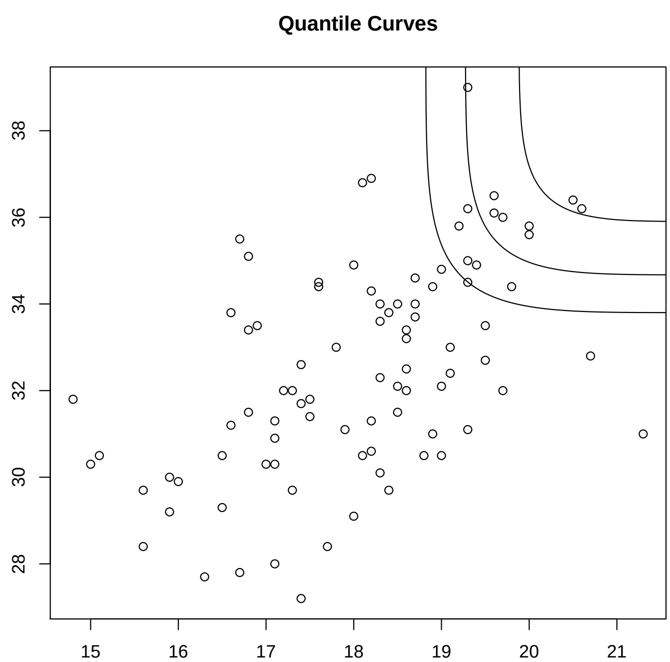

Figure 11: α quantile curves for the fitted logistic model for $\alpha = 0.7$, 0.8 and 0.9 for the Colmar data.

Conclusion

Through a series of extreme data analysis, univariate and multivariate basic concepts in EVT have been presented in an environmental context. These concepts can be extended to the spatial case through max-stable fields, see for instance de Haan (1984) and Smith (1990). Inference on such processes can be obtained using composite likelihood as described in Lindsay (1988) and Varin (2008) (see Padoan *et al.* (2010) for an application in an extreme value context with an illustration on US precipitation extremes and Blanchet and Davison (2011) for an illustration on annual maximum snow depth).

Acknowledgment

The authors are grateful to the referee for their constructive comments from which the paper has benefitted a lot. Part of this work has been supported by the EU-FP7 "ACQWA" Project (www.acqwa.ch) under Contract Nr 212250, by the PEPER-GIS project, by the ANR-MOPERA project, by the ANR-McSim project, by the MIRACCLE-GICC project and by the *Chaire d'excellence "Generali - Actuariat responsable : gestion des risques naturels et changements climatiques"*.

References

Bacro JN, Bel L, Lantuéjoul C. 2010. Testing the independence of maxima : from bivariate vectors to spatial extreme fields. *Extreme*, **13**:155-175.

Beirlant J, Goegebeur Y, Segers J, Teugels J. 2004. *Statistics of Extremes: Theory and Applications*; Wiley Series in Probability and Statistics.

Blanchet J, Davison A. 2011. Spatial Modelling of extreme snow depth. *The Annals of Applied Statistics*, **5**(3):1699-1725.

Coles S. 2001. *An introduction to statistical modeling of extreme values*; Springer Series in Statistics, Springer-Verlag, London.

Crowder MJ. 1989. A multivariate distribution with Weibull components. *J. Roy. Statist. Soc. B*, **51**:93-108.

Diebolt J, Guillou A, Naveau P, Ribereau P. 2008. Improving Probability Weighted Moment Methods for the Generalized Extreme Value Distribution. *REVSTAT*, **6**:33-50.

Embrechts P, Klüppelberg C, Mikosch T. 1997. *Modelling Extremal Events for Insurance and Finance*; volume 33 of Applications of Mathematics, Springer-Verlag, Berlin.

Fougères A-L, Nolan JP, Rootzén H. 2009. Models for dependent extremes using stable mixtures. *Scand. J. Statist.*, **36**:42-59.

Gudendorf G, Segers J. 2010. Extreme-value copulas. *In: Copula theory and its applications (Warsaw, 2009) (editor: P. Jaworski);* Lecture Notes in Statistics - Proceedings, Springer-Verlag, Berlin, 127-146.

de Haan L. 1984. A Spectral Representation for Max-Stable Processes. *The Annals of Probability*, **12**:1194-1204.

de Haan L, Ferreira A. 2006. *Extreme Value Theory: An Introduction*; Springer Series in Operations Research.

Hosking JRM, Wallis JR, Wood EF. 1985. Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments. *Technometrics*, **27**:251-261.

Hougaard P. 1986. A class of multivariate failure time distributions. *Biometrika*, **73**:671-678.

Joe H. 1997. *Multivariate models and dependence concepts*; Monographs on Statistics and Applied Probability, 73; Chapman & Hall, London.

Katz R, Parlange M, Naveau P. 2002. Extremes in hydrology. *Advances in Water Resour.*, **25**:1287-1304.

Lindsay BG. 1988. Composite Likelihood Methods. *Contemporary Mathematics*, **80**:221-239.

Maraun D, Rust HW, Osborn TJ. 2010. Synoptic airflow and UK daily precipitation extremes. Development and validation of a vector generalized model. *Extremes*, **13**:133-153.

Naveau P, Guillou A, Cooley D, Diebolt J. 2009. Modeling Pairwise Dependence of Maxima in Space. *Biometrika*, **96**:1:17.

Padoan SA, Ribatet M, Sisson SA. 2010. Likelihood-based inference for max-stable processes. *J. Amer. Statist. Assoc.*, **105**:263-277.

Pickands J. 1975. Statistical inference using extreme order statistics. *Ann. Statist.*, **3**:119-131.

Pitt MK, Shepard N. 1999. Filtering via simulation: Auxiliary particle filters. *J. Am. Statis. Assoc.*, **94**:590-599.

Resnick SI. 1987. *Extreme Values, Regular Variation and Point Processes*; Springer Verlag, Nw York.

Resnick SI. 2007. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Operations Research and Financial Engineering, Springer, New-York..

Ribereau P, Guillou A, Naveau P. 2008. Estimating return levels from maxima of non-stationary random sequences using the Generalized PWM method. *Nonlin. Processes Geophys.*, **15**:1033-1039.

Smith RL. 1985. Maximum likelihood estimation in a class of non-regular cases. *Biometrika*, **72**:67-90.

Smith, RL. 1990. *Max-Stable Processes and Spatial Extremes.* Unpublished manuscript, University of North California.

Stephenson, AG. 2002. evd: Extreme Value Distributions. *R News*, **2**:31-32. URL: http://CRAN.R-project.org/doc/Rnews/

Tawn JA. 1990. Modelling multivariate extreme value distributions. *Biometrika*, **77**:245-253.

Toulemonde G, Guillou A, Naveau P, Vrac M, Chevallier F. 2010. Autoregressive models for maxima and their applications to CH_4 and N_2O. *Environmetrics*, **21**:189-207.

Toulemonde G, Guillou A, Naveau P. 2012. Particle filtering for Gumbel-distributed daily maxima of methane and nitrous oxide. *Environmetrics*, published online, doi : 10.1002/env.2192.

Van der Vaart AW. 1998. *Asymptotic Statistics*; Cambridge University Press.

Varin C. 2008. On Composite Marginal Likelihoods. *Advances in Statistical Analysis*, **92**:1–28.