HAL open science

Spatial dependences among precipitation maxima over Belgium

S. Vannitsem, P. Naveau

To cite this version:

S. Vannitsem, P. Naveau. Spatial dependences among precipitation maxima over Belgium. Nonlinear Processes in Geophysics, 2007, 14 (5), pp.621-630. 10.5194/npg-14-621-2007 . hal-00331105

HAL Id: hal-00331105

https://hal.science/hal-00331105v1

Submitted on 18 Jun 2008

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.

Nonlin. Processes Geophys., 14, 621–630, 2007
www.nonlin-processes-geophys.net/14/621/2007/
© Author(s) 2007. This work is licensed
under a Creative Commons License.

Nonlinear Processes in Geophysics

Spatial dependences among precipitation maxima over Belgium

S. Vannitsem[1] **and P. Naveau**[2]

[1]Institut Royal Météorologique de Belgique, Belgium
[2]Laboratoire des Sciences du Climat et de l'Environnement, IPSL-CNRS, France

Received: 22 June 2007 – Revised: 31 August 2007 – Accepted: 10 September 2007 – Published: 17 September 2007

Abstract. For a wide range of applications in hydrology, the probability distribution of precipitation maxima represents a fundamental quantity to build dykes, propose flood planning policies, or more generally, to mitigate the impact of precipitation extremes. Classical Extreme Value Theory (EVT) has been applied in this context by usually assuming that precipitation maxima can be considered as Independent and Identically Distributed (IID) events, which approximately follow a Generalized Extreme Value distribution (GEV) at each recording site. In practice, weather stations records can not be considered as independent in space.

Assessing the spatial dependences among precipitation maxima provided by two Belgium measurement networks is the main goal of this work. The pairwise dependences are estimated by a variogram of order one, also called madogram, that is specially tailored to be in compliance with spatial EVT and to capture EVT bivariate structures. Our analysis of Belgium precipitation maxima indicates that the degree of dependence varies greatly according to three factors: the distance between two stations, the season (summer or winter) and the precipitation accumulation duration (hourly, daily, monthly, etc.). Increasing the duration (from one hour to 20 days) strengthens the spatial dependence. The full independence is reached after about 50 km (100 km) for summer (winter) for a duration of one hour, while for long durations only after a few hundred kilometers. In addition this dependence is always larger in winter than in summer whatever is the duration. An explanation of these properties in terms of the dynamical processes dominating during the two seasons is advanced.

Correspondence to: S. Vannitsem
(svn@oma.be)

1 Introduction

Classically, the statistical analysis of a time series of maxima is based on the well-developed Extreme Value Theory (EVT) (Coles, 2001; Fougères, 2004; Embrechts et al., 1997; Resnick, 1987; Smith, 2004). This theory demonstrates that the Generalized Extreme Value distributions (GEV) is the appropriate distribution for fitting maxima in an univariate and independent context.

Although the assumption of temporal independence seems to be a reasonable approximation in practical applications, the spatial independence seems on the other hand far from acceptable. Weather stations can be located nearby, and this spatial proximity can impact the computation of uncertainty related to the classical quantities of interest for hydrologists and flood planners. The typical example is the estimation of large return levels using maxima of precipitation recorded over a specific region in which several nearby stations are available. In this situation, the use of an erroneous hypothesis of spatial independence will lead to an underestimation of the return level confidence intervals (Buishand, 1989). Solving this problem of uncertainty underestimation is not straightforward (even the definition of the return levels in a spatial context is not straightforward). In the present paper, we do not address this general problem and we limit our investigation to the measure of spatial dependences, as a function of different parameters such as the maxima block size (one month, one year, etc.), the period of accumulation of precipitation (hours, daily, etc.) and the season.

In the history of statistics and probability, the question of how to measure the dependence between two random variables has always been of paramount importance (Ancona-Navarrete and Tawn, 2002; Hall and Tajvidi, 2000; Hsing et al., 2004; Schlather and Tawn, 2003; Smith, 2004; Tawn, 1988). For Gaussian random vectors, it is well known that the covariance matrix fully captures the dependence information. For non-Gaussian distributions, many different ways exist to

Published by Copernicus Publications on behalf of the European Geosciences Union and the American Geophysical Union.

measure the departure from independence. To illustrate this point, let us consider two continuous random variables, say (X, Y) with their joint bivariate probability distribution function, say $\mathbb{P}(X \leq x, Y \leq y)$. There exist several approaches to express the joint distribution in terms of its marginal distributions $\mathbb{P}(X \leq x)$ and $\mathbb{P}(Y \leq y)$. One is based on the following equality

$$\mathbb{P}(X \leq x, Y \leq y) = [\mathbb{P}(X \leq x) \times \mathbb{P}(Y \leq y)]^{\theta(x,y)} \quad (1)$$

where $\theta(x, y)$ is a bivariate positive function. If $\theta=1$, then full independence is reached. Accurate estimators of $\theta(x, y)$ should provide valuable information about the dependence between X and Y (Buishand, 1984). The advantage of Eq. (1) and the link between Eq. (1) and EVT will be addressed in Sect. 3.

A second and popular approach to link the joint distributions and its margins is the notion of copula (Coles, 2001) defined by $\mathbb{P}(X \leq x, Y \leq y) = C\,[\mathbb{P}(X \leq x), \mathbb{P}(Y \leq y)]$, where C is called the copula function. This is equivalent to work with uniform distributed margins on $[0, 1]$. Depending on the problem at hand, one can fit his/her favorite copula family to represent the dependence structure. In the EVT community, there has been a recent debate of the overuse of copulas. In particular, Mikosch (2006) has cautioned against copulas for a variety of theoretical and practical reasons. One issue is that the assumption that the margins are known (by stating that they are uniform) is not true in real case studies and additional errors are introduced by replacing the margin by its estimate.

Besides these two methods, there exists a variety of other approaches to estimate the bivariate structure. For example, we can assume a parametric bivariate family, e.g. bi-Gamma densities, or transform the data into a Gaussian vector, estimate the covariance matrix and then make the inverse operation, see Chapter 33 of Wackernagel (2003). But in the former approach strong distributional assumption are used, e.g. bi-Gamma densities, and for the latter, the derivation of the estimation error after the gaussian and inverse transformations can be proved very difficult.

In this paper, we will use a direct non-parametric method (see Sect. 3) consistent with EVT and for which no particular parametric class is assumed as for the copula function. It is based on the $\lambda-$madogram concept introduced by Naveau et al. (2007)[1] and Cooley et al. (2006). It is a special variogram of order one, one of the pillars of the field of geostatistics (Cressie, 1993; Wackernagel, 2003; Stein, 1999). One main advantage of working with this madogram approach is that it offers a simple and efficient connection between EVT and geostatistics. To our knowledge, this is the first case study in which such statistical advances are used to measure the spatial dependences among precipitation maxima over a large region, taking into account the influence of maxima block size, the accumulation period and the season.

Our paper is organized as follows. Section 2 presents the two different networks covering the Belgian territory used in the present study, as well as the analysis of the temporal non-stationarities present in the data set. The theoretical aspects of bivariate EVT and a description of the pairwise dependence estimators among precipitation maxima are detailed in Sect. 3. In Sect. 4, we present and interpret the results obtained for the two Belgium networks. In particular the role of seasonality, of the duration and of the block size on the dependence among maxima, is investigated. The main conclusions are drawn in Sect. 5.

2 The precipitation networks

Two different networks covering the Belgian territory are currently available, the climatological and hydrometeorological networks. The former is based on daily measurements of climatologically relevant quantities such as temperature, precipitation, pressure, etc.; while the second is aimed at evaluating precipitation, temperature and humidity at a much higher rate (every 10 min) in order to compute quantities relevant for hydrological modelling like fast runoffs. In this section both networks are briefly described.

2.1 The climatological network

The climatological network started to operate in 1833. Based at that time on a few key stations, it is now composed by more than 250 stations. During its long existence, this network has experienced several changes such as displacements and withdrawals of stations. To work with an homogeneous network, we focus here on a subset of stations covering a common period from 1951 up to 2005 without substantial interruptions. We end up with 90 stations over the whole Belgian territory for which at most, 2% of missing data are present. The stars in Fig. 1 indicate their respective locations. For each station and for different accumulation periods (from 1 to 20 days), the maxima for the hydrological summer (April to September) and winter (October to March) are extracted over the whole period.

One central assumption in this study is that precipitation maxima are supposed to be stationary in time. Consequently, we need to check the validity of this assumption for our selected stations. The Mann test (Mann, 1945) which provides information on the presence of tendencies has been applied to the ensemble of maxima time series (for a duration of 1 day). The analysis reveals that the vast majority of the time series recorded for summer are stationary. For winter, the results are drastically different: about 2/3 of the stations are non-stationary at the 5% level. Similar results have been obtained by Gellens (2000), except that the number of series displaying a trend in winter was smaller (about 1/3). This probably

[1] Naveau, P., Guillou, A., Cooley, D., and Biebolt, J.: Modeling pairwise dependence of maxima in space, in review, 2007.

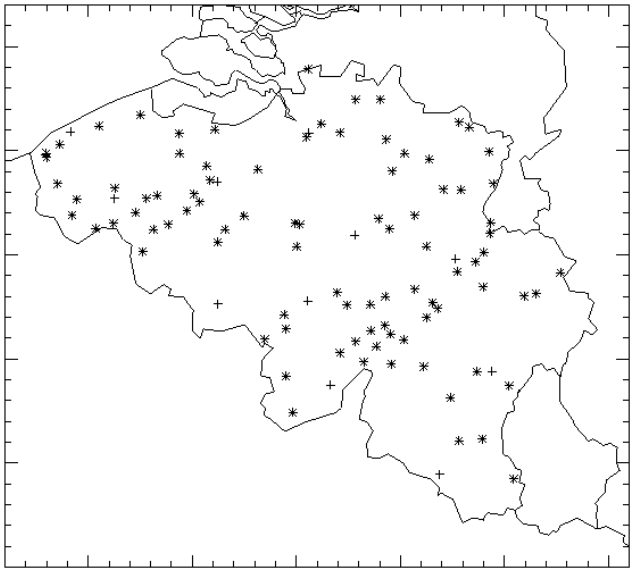

Fig. 1. Locations of the precipitation stations of the Belgian climatological (*) and hydro-meteorological (+) networks.

reflects the fact that the series are longer (by about 10 years) and the change present in the middle of the 80th can therefore be detected with more confidence. In order to perform the analysis of the spatial dependences of the extremes, we will limit ourselves to the stationary time series, namely 90 summer series and 33 winter series.

Daily precipitation maxima times series recorded during summer at three representative stations (Uccle, Uccle (Réservoir) and Spa) are displayed in Fig. 2. The two stations Uccle and Uccle (Réservoir) are very close to each other (about 2 km), while the third station, Spa, is far from the two previous ones (about 150 km). As one may expect, the maxima recorded at the two closest stations represented by crosses and white circles in Fig. 2a are very similar, while large differences are found with the values recorded at Spa (Fig. 2b). This illustrates the spatial dependences that can arise between maxima, which will be quantitatively estimated in Sect. 4.

2.2 The hydro-meteorological network

The first Hellmann-Fuess pluviograph was installed in Uccle in 1898 (Demarée, 2003). Up to 2005, this instrument has provided a continuous time series of precipitation at a very fine sampling rate (every 10 min). In 1968, 18 additional pluviographs were installed, providing a spatial coverage of the country for hydrological purposes (+ in Fig. 1). These pluviographs are nowadays progressively replaced by automatic stations.

Clearly this network is less dense than the climatological one and some stations display long periods with missing data.

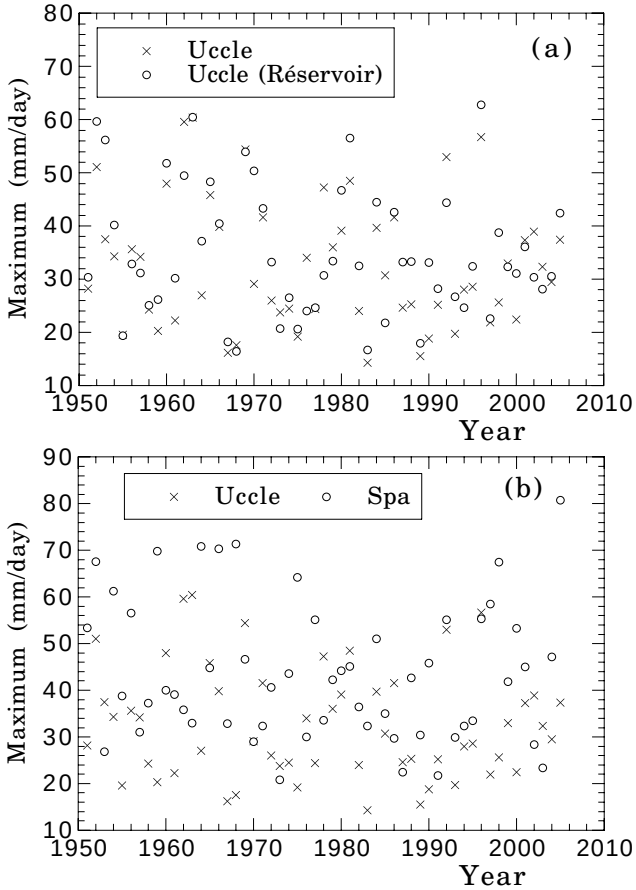

Fig. 2. Time series of summer daily precipitation maxima recorded at three stations of the climatological Belgian network. Two stations, Uccle and Uccle (Réservoir) represented by crosses and circles in panel **(a)**, are close to each other (about 2 km), while the third one, Spa (circles in panel **b**), is far from Uccle (about 150 km), see panel (b).

Despite these drawbacks, the hydro-meteorological network provides important precipitation information for short durations that cannot be obtained from the climatological network.

For this network, the stationarity has also been investigated, indicating that for very short durations (typically 1 h), the data are stationary for both seasons.

2.3 The climatological information gathered from the networks

Measurements provided by both networks have been extensively used for climatological analyses (e.g. Sneyers, 1975; Dupriez and Demarée, 1988, 1989; Sneyers et al., 1989; Gellens, 2000) and modelling (e.g. Schmitt et al., 1998). In particular, R. Sneyers and his collaborators have made a lot of efforts to provide homogeneous time series for the climatological network and to give a detailed description of

the statistical properties of the data (see for instance Sneyers, 1975).

One important question in this context is to know whether the maxima of this (limited) data set are indeed sampled from a GEV distribution, which is in principle only valid for maxima selected over a large number of recorded events. In Dupriez and Demarée (1988, 1989) a detailed analysis of the appropriateness of the GEV distributions for a set of maxima time series was performed. It was shown that for most of the time series, the hypothesis that annual maxima of precipitation accumulated over periods from 1 to 30 days follows a GEV distribution, cannot be rejected. This test is particularly important for long duration maxima since they are selected over a relatively small number of recorded events during each year or season. A more recent study has reached similar conclusions (Gellens, 2003). In the following, we will therefore adopt the assumption that the distribution of precipitation maxima falls into the asymptotic domain of convergence of the GEV distribution, even for long duration events.

3 Theoretical background

Let us define precipitation maxima as follows

$$M_i(t) = \max\{Z_{i,1}(t), \dots Z_{i,m}(t)\}, \tag{2}$$

where m denotes the number of periods of duration t recorded over a time window, T, and i=1, ..., n with n corresponding to the number of stations. The random variable $Z_{i,j}(t)$ represents the precipitation amount temporally indexed by j, accumulated over the duration t. For instance, if $M_i(t)$ represents monthly maxima of hourly precipitation, T is equal to 1 month, t to one hour and m=24×30. Note that maxima are usually selected over annual or seasonal windows, but can also be defined over longer time windows, i.e. on a multi-annual basis.

The marginal distribution $\mathbb{P}(M_i(t) \leq x)$ is denoted by $F_i(x)$ and its pairwise distribution between two stations labeled i and k by

$$F_{ik}(x, y) = \mathbb{P}(M_i(t) \leq x, M_k(t) \leq y).$$

Conceptually, it is easy to go beyond the bivariate case, but the estimation of multivariate distributions from a limited number of observations is difficult without adding very strong distributional assumptions. As most authors in geostatistics have done in the past, we restrict our attention on estimating the bivariate structure of random vectors. Besides its computational feasibility, this pairwise framework already offers novel and practical information about the spatial properties of precipitation maxima.

Univariate EVT has been very popular amongst hydrologists for many decades (e.g. Katz et al., 2002), and it has also become a standard tool in the climate community in recent years (e.g. Kharin et al., 2007). However some of the latest advances in EVT obtained in the mathematical field of probability and statistics are not widely circulating among climatologists and hydrologists. For example, applied and theoretical statisticians (e.g. Capéraà et al., 1997; de Haan and Pereira, 2006; Hall and Tajvidi, 2000; Pickands, 1981), have proposed and explored many directions to infer the characteristics of the bivariate structure between maxima. Several parametric families have also been studied (e.g. Tawn, 1988, 1990; Joe, 1993; Coles et al., 1999). A very popular model is the logistic one proposed by Gumbel (1960).

One rather old but still fundamental result (e.g. Gumbel, 1960; Resnick, 1987; Fougères, 2004; Beirlant et al., 2004) states that, under mild conditions, all pairwise maxima vectors, $(M_i(t), M_k(t))$, asymptotically converge to a bivariate non-degenerate distribution $G_{ik}(.,.)$ called the "bivariate Extreme Value distribution" and defined by

$$G_{ik}(x, y) = \exp\left[-V_{ik}\left(\frac{-1}{\ln G_i(x)}, \frac{-1}{\ln G_k(y)}\right)\right],$$
$$\text{where } V_{ik}(x, y) = 2\int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) dH_{ik}(w), \tag{3}$$

$H_{ik}(.)$ corresponds to any distribution function on $[0, 1]$ such that $\int_0^1 w\, dH_{ik}(w)$=0.5, and $G_i(.)$ and $G_k(.)$ are the marginal GEV distributions. The function $V_{ik}(.,.)$ is called the "pairwise extremal dependence function" and it is the one that captures the bivariate dependence structure. To clarify the meaning of Eq. (3), we can look at the special case where y=x and the margins are equally distributed, i.e. G_i=G_k. This setup gives

$$\begin{aligned} G_{ik}(x, x) &= \exp\left[V_{ik}(1, 1) \ln G_i(x)\right], \\ &= \left[G_i(x) \times G_k(x)\right]^{V_{ik}(1,1)/2}, \end{aligned}$$

because $G_i(x)$=$G_k(x)$. Here we can recognize the form of Eq. (1) with $\theta(x, x)$=$V_{ik}(1, 1)/2$, for more details see the recent book by Beirlant et al. (2004). The scalar $V_{ik}(1, 1)$ is called the "extremal coefficient", equal to two for the independent case, and to one for the full dependent case. Otherwise, its value varies between one and two depending on the degree of dependence.

Two important features of Eq. (3) are that the margins have no influence on the dependence structure and that $H_{ik}(.)$ can be any distribution function on $[0, 1]$ such that $\int_0^1 w\, dH_{ik}(w)$=0.5. In other words, this implies that there is an infinite number of ways to define a bivariate structure for sample pairwise maxima (although the margins have a known parametric distribution, the univariate GEV). Consequently, fixing a parametric form for $V_{ik}(x, y)$ is equivalent to impose a very strong hypothesis on the bivariate structure.

To estimate the extremal coefficient $V_{ik}(1, 1)$, we follow the non-parametric method proposed by Cooley et al. (2006) who showed that the extremal coefficient can be directly estimated from the *madogram* defined by

$$\nu_{ik} = \frac{1}{2}\mathbb{E}(|F_i(M_i(t)) - F_k(M_k(t))|), \tag{4}$$

where $\mathbb{E}(.)$ denotes the expectation. Note that ν_{ik} and $F_i(x)$ like the random variable $M_i(t)$ also depends on t. But, to keep the notations as transparent as possible, we drop this temporal information. For independent maxima, ν_{ik} is equal to 1/6, and smaller for dependent events, i.e. $\nu_{ik}<1/6$ if there is positive association. Compared to past non-parametric estimators of $V_{ik}(1,1)$, the performances of the estimators proposed by Capéraà et al. (1997) and Hall and Tajvidi (2000) are similar to the ones obtained with the madogram. But while the relationship between the madogram (that represents a first-order variogram) and geostatistics is obvious, the link of these past estimators with the field of geostatistics is not clear. Hence, the simplicity of defining and computing the madogram within a spatial setup represents one of our main reasons for preferring the madogram.

Because bivariate EVT allows us to assume that the bivariate vector $(M_i(t), M_k(t))$ satisfies Eq. (3), it is possible to show (Cooley et al., 2006) that

$$V_{ik}(1,1) = \frac{0.5+\nu_{ik}}{0.5-\nu_{ik}}. \tag{5}$$

This equality implies that if one can estimate ν_{ik}, then a direct estimate of $V_{ik}(1,1)$ can be derived. Fortunately, it is simple to find a naive estimator of ν_{ik} that does not depend on the form of the margins

$$\hat{\nu}_{ik} = \frac{1}{2L}\sum_{l=1}^{L} |\mathbb{F}_i(M_{i,l}(t)) - \mathbb{F}_k(M_{k,l}(t))|.$$

where L corresponds to the number of pairs of maxima under study, $(M_{i,l}(t), M_{k,l}(t))_{l=1,\ldots,L}$, and $\mathbb{F}_i(.)$ represents the empirical distribution of the random variable $M_i(t)$, i.e.

$$\mathbb{F}_i(x) = \frac{1}{L}\sum_{l=1}^{L} \mathbb{1}_{\{M_{i,l}(t)\le x\}},$$

with $\mathbb{1}_A$ corresponding to the indicator function (equals to one if A is true and to zero otherwise). A limitation of this approach is that Eq. (5) only provides an estimate of the extremal coefficient $V_{ik}(1,1)$, but not of the full pairwise extremal dependence function $V_{ik}(x,y)$ with x not necessarily equal to y. To solve this issue, one can build on the madogram concept by adding an extra dimension to Eq. (4). Such an extension has been recently developed by Naveau et al. (2007)[1] who defined the "λ-madogram" by

$$\nu_{ik}(\lambda) = \frac{1}{2}\mathbb{E}(|F_i^{\lambda}(M_i(t)) - F_k^{1-\lambda}(M_k(t))|), \forall\lambda \in [0,1]. \tag{6}$$

From this definition, it is possible to show the following relationship between $\nu_{ik}(\lambda)$ and the complete pairwise extremal dependence function $V_{ik}(\lambda, 1-\lambda)$

$$\nu_{ik}(\lambda) = \frac{V_{ik}(\lambda,1-\lambda)}{1+V_{ik}(\lambda,1-\lambda)} - c(\lambda), \tag{7}$$

where $c(\lambda)=\frac{3}{2(1+\lambda)(2-\lambda)}$ and $\lambda\in(0,1)$. Concerning Eq. (7) three comments are in order. First, because $V_{ik}(x,y)=\frac{1}{x+y}V_{ik}(\lambda,1-\lambda)$ with $\lambda=\frac{x}{x+y}$, the λ-madogram fully characterizes the dependence function $V_{ik}(x,y)$ for any $x,y\in\mathbb{R}^2$. Second, the extremal coefficient $V_{ik}(1,1)$ equals $V_{ik}(1/2,1/2)/2$. This means that the λ-madogram like the madogram can provide the extremal coefficient $V_{ik}(1,1)$. Third, the $\lambda-$madogram satisfies $\nu_{ik}(0)=\nu_{ik}(1)=0.25$. This latter remark allows us to propose the following estimator for $\nu_{ik}(\lambda)$

$$\begin{aligned}\hat{\nu}_{ik}(\lambda) = &\frac{1}{2L}\sum_{l=1}^{L}|\mathbb{F}_i^{\lambda}(M_{i,l}(t)) - \mathbb{F}_k^{1-\lambda}(M_{k,l}(t))| \\ &-\frac{\lambda}{2L}\sum_{l=1}^{L}\left(1-\mathbb{F}_i^{\lambda}(M_{i,l}(t))\right) \\ &-\frac{1-\lambda}{2L}\sum_{l=1}^{L}\left(1-\mathbb{F}_k^{1-\lambda}(M_{k,l}(t))\right) \\ &+\frac{1}{2}\frac{1-\lambda+\lambda^2}{(2-\lambda)(1+\lambda)}.\end{aligned} \tag{8}$$

Although more complex, this definition has the advantage to force $\hat{\nu}_{ik}(\lambda)$ to satisfy the limiting conditions $\mathbb{E}(\hat{\nu}_{ik}(0))=\mathbb{E}(\hat{\nu}_{ik}(1))=0.25$.

Before closing this theoretical section, we would like to emphasize the inherent assumptions made in this paper. We assume implicitely that each year (or each season), the field of maxima is isotropic and homogeneous in space, and stationary in time. In addition, the bivariate vector $(M_i(t), M_k(t))$ drawn for such processes belongs to the class of bivariate Extreme Value distributions whose bivariate structure is defined by $V_{ik}(x,y)$. The assumptions of homogeneity and isotropy may not be entirely valid, but a complete full description taking into account inhomogeneity and anisotropy would necessitate a very large amount of data. A large number of synthetic data can be simulated by numerical models (see for instance Vannitsem, 2007), but they are very rarely available when dealing with real measurements. In addition, the assumptions needed to apply bivariate EVT may not always hold. For instance, Ancona-Navarrete and Tawn (2002) found that the dependence structure of 1-day extreme rainfall in south-east England resembles that of a bivariate Gaussian random variable rather than that of a bivariate EV distribution. While interpreting the results and making conclusions about precipitation maxima in Belgium, one has to keep in mind these assumptions.

4 Estimating and interpreting pairwise dependences among Belgium precipitation maxima

Many empirical statistical studies of the spatial properties among maxima have been limited to the regionalization or interpolation of the parameters of the GEV univariate distributions (e.g. Cooley et al., 2007; Coles and Casson, 1999,

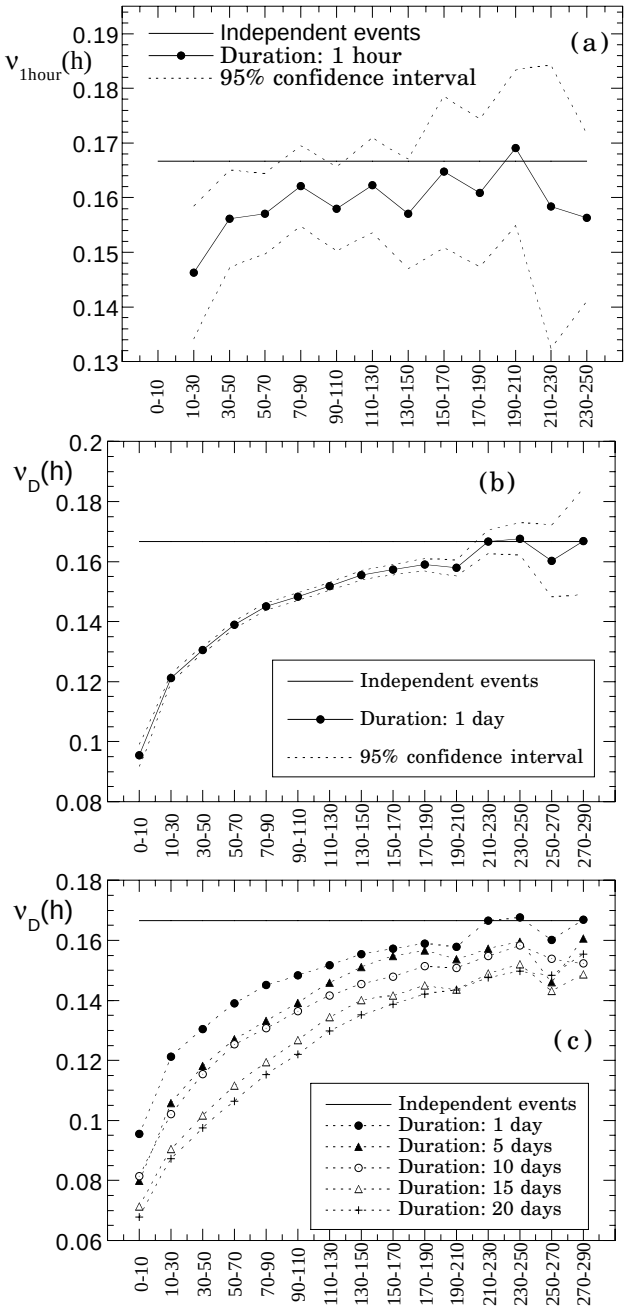

Fig. 3. Variations of the madogram as a function of the distance h given in kilometers, Eq. (4), for summer precipitation maxima. Panel **(a)** is obtained from the hydro-meteorological network data (see Sect. 2.2) with a duration of one hour. Panels **(b)** and **(c)** are obtained using the climatological network data (see Sect. 2.1) with a duration of one day (panel b) and for a series of different durations (panel c). The dashed lines in panels (a) and (b) refer to the 95% confidence interval of the madogram estimation and the horizontal continuous line to the independent case.

and references therein). Conditionally on the GEV parameters the observations are assumed to be IID and consequently the spatial structure is only captured by the GEV parameters. Hence this latent/Bayesian approach does not take full advantage of multivariate EVT which brings a more profound and detailed understanding of the properties of multivariate extremes. In particular the spatial dependences among maxima constitute a central problem, as already revealed in Buishand (1984, 1989) and Coles and Tawn (1990).

For Belgium, Gellens (2000) showed that spatial correlations are present among maxima for durations from 1 day to 30 days. Defining a (somewhat arbitrary) mean decorrelation length, he found for winter a decorrelation length varying between 150 km and 400 km for durations from 1 to 30 days; while for summer, a mean decorrelation length varying between 40 to 200 km. These results already suggest a strong difference between summer and winter and for short and long precipitation durations. This question is now reassessed based on the techniques described in Sect. 3.

Before presenting the result, we must indicate how the madogram is effectively computed. As presented in Sect. 2, the length of the time series is relatively short. This implies that the statistics for one pair (i,k) of stations is quite small. In order to get a relatively good estimate of the madogram, one must aggregate the results obtained at several pairs of sites. We have therefore defined a certain number of distance bins for which the madogram is evaluated based on the ensemble of pairs of stations whose distance falls into these bins.

Figures 3a and 4a display the variation of the madogram (full dots) as a function of the distance for summer and winter precipitation maxima recorded with the hydrometeorological network. The duration is fixed here to 1 h. The dashed lines represent the 95% confidence interval which has been computed based on the asymptotic probability distribution of the madogram estimator (see Cooley et al., 2006). The horizontal line refers to the value corresponding to independent events. Obviously, the spatial dependences for this duration is quite small as indicated by the closeness of the curve to the independent event case (continuous horizontal line). In addition the horizontal line falls within the confidence interval for distances larger than 50 km in summer and 100 km in winter, suggesting that the distance for which the complete independence is reached is larger in winter than in summer.

This last result is consistent with the properties of the maxima arising in winter and summer, respectively. In summer precipitation maxima are usually associated with the convective thunderstorm activity whose lifetime and spatial scales are relatively small. While in winter they are usually associated with atmospheric situations arising from the intense baroclinic activity prevailing during this season, giving rise to large-scale stratiform precipitation structures.

Figures 3b and 4b display the madogram as well as the 95% confidence interval for the stations of the climatological

network and for a duration of 1 day. Note that for winter, we have only kept stations which do not display a tendency, implying that the confidence interval for the madogram is larger in winter. For both seasons, the result suggests that the independence is reached only after about 200 km for maxima of daily precipitation. For shorter distances, the dependence is stronger in winter than in summer, as expected from the very nature of the dynamics prevailing in winter or summer as already discussed above. The present results differ from classical correlation analyses which indicates a highly different decorrelation distance in winter and summer for this duration (Gellens, 2000). This emphasizes the necessity to use appropriate techniques for evaluating dependences for extreme values that are on one hand originating from a nonlinear process and on the other, obtained through a highly nonlinear operation, Eq. (2).

For longer durations, the madogram monotonically decreases as illustrated in Figs. 3c and 4c, but substantial differences are found in winter and summer. In winter, the madogram progressively decreases for durations from 1 to 5 days. For 5 to 15 days, it does not change substantially and starts again to decrease beyond durations of 15 days. To understand this variation of the madogram as a function of the duration, one must recall that the dominant structures which govern the dynamics at midlatitude during winter are the baroclinic waves whose typical time scale is 5 days. One can therefore suspect that the plateau for intermediate durations (from 5 to 15 days) reflects the fact that once this duration is reached, the spatio-temporal structure of the maxima is associated with the prevalence of the baroclinic waves during this season. The subsequent decrease beyond 15 days reflects the convergence toward the coherent large space scale organisation of the seasonal variations.

For summer, the picture is substantially different. The decrease is quite progressive from 1 to 20 days, reflecting that the baroclinic waves are much less dominant during this season. It is also very interesting to note that for space scales larger than 200 km, the madogram does not vary much and is very close to the curve corresponding to independent events. This further suggests that the space scales of the processes are quite limited, as expected by the predominantly convective nature of the dynamics during this season. The comparison of Figs. 3c and 4c further reveals that the spatial coherence of the multi-day extremes in winter is very large (much beyond 250 km), while in summer, this coherence is much more limited.

Other techniques like the one proposed in Coles (2001) provide a similar picture for the spatial dependence of the maxima (not shown).

Up to now, the analysis has revealed a large dependence of the madogram as a function of the duration. In the problem of the spatial dependences of the extremes, another parameter is of interest, the time window T. One can wonder whether by changing T one can modify the properties of the spatial dependences. In Fig. 5 the time window T has been increased

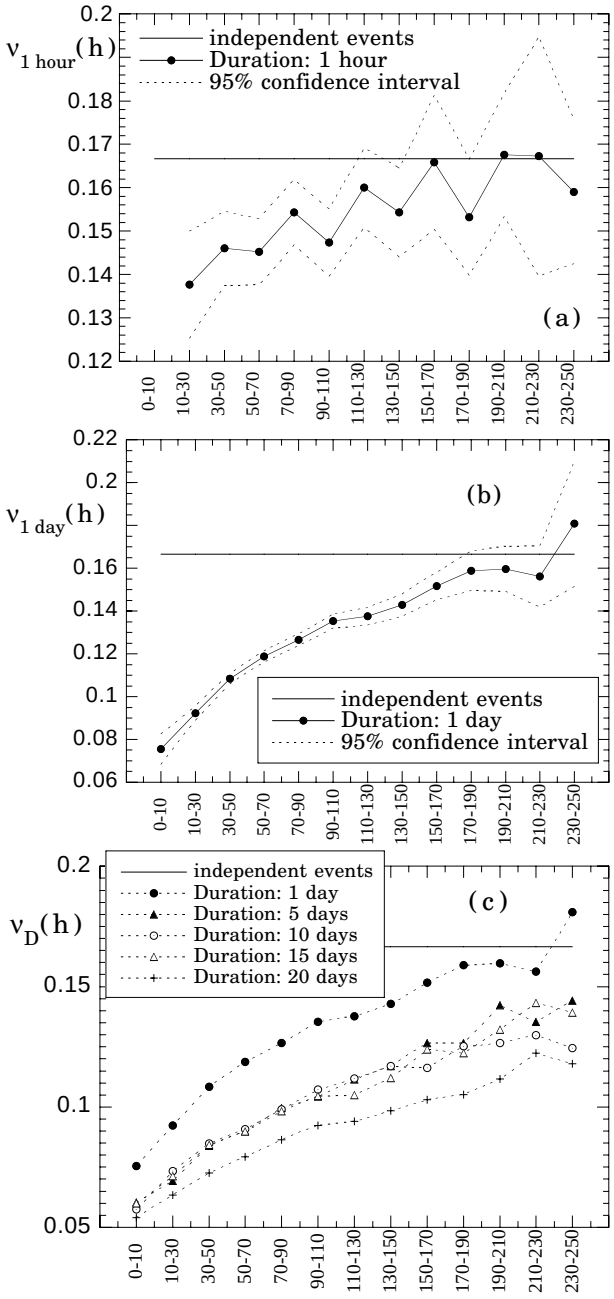

Fig. 4. As in Fig. 3 but for winter precipitation.

from 1 summer to 2 and 3 summers. Although the uncertainty on the madogram is larger for longer time windows (less pairs are available), the results do not seem to reveal any systematic increase or decrease of the dependence, suggesting that the spatial dependences found for one-summer time window is similar for longer windows. In other words, this dependence measure does not seem to converge toward the independent case for larger windows.

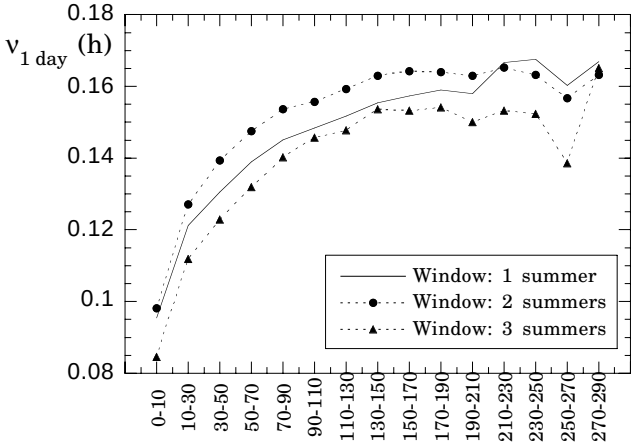

Fig. 5. Variation of the madogram as a function of the distance (given in kilometers) for summer precipitation maxima selected over different periods, T, from one to three consecutive summers. The climatological network is used.

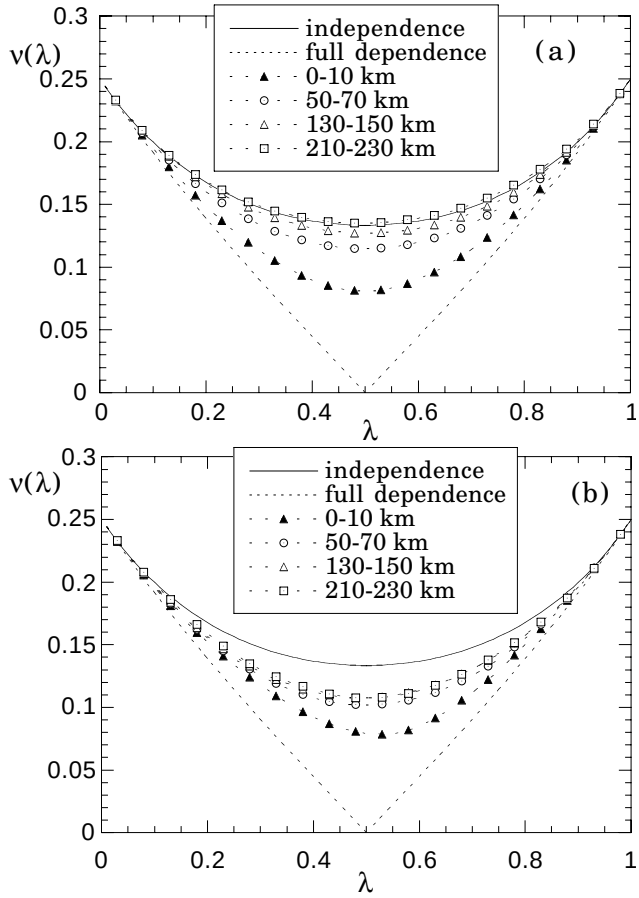

Fig. 6. The λ-madogram defined by Eq. (6) is plotted for several distance intervals for summer **(a)** and winter **(b)** daily precipitation maxima coming from the climatological network. The continuous line refers to the independent case and the dashed line to the full dependence case.

Let us now turn to the λ-madogram. Figure 6a displays the λ-madogram as a function of λ for different distance intervals, 0–10 km, 50–70 km, 130–150 km and 210–230 km for Summer. The continuous line corresponds to the case for which the extremes are independent, while the dashed line to the full dependence. The λ-madogram estimates fall between these two asymptotic solutions and, as expected, they progressively converge to the independent solution for increasing distances. This phenomenon takes place independently of the value of λ suggesting a similar convergence toward independence.

For winter (Fig. 6b), the results are similar, except for the very slow convergence toward independence whatever is the value of λ. This result is consistent with what was found for the madogram during this season (Fig. 4c).

5 Conclusions

In this work, the spatial dependences among maxima have been explored using the ($\lambda-$)madogram, in the context of two Belgium precipitation networks. This technique is a non-parametric approach recently developed in order to provide a direct evaluation of the pairwise extremal dependence function.

Using this approach, valuable information about the spatial properties of precipitation maxima during the summer and winter seasons in Belgium, were extracted. First, for very short durations (typicall one hour), the madogram of precipitation maxima for both seasons indicates a weak spatial dependence, but larger in winter than in summer. This feature already reflects the nature of the dynamical processes prevailing during summer or during winter: In summer precipitation maxima are mainly associated with the thunderstorm activity whose space and time scales are small; for winter, the precipitation field usually displays very large space and time scales, related to the baroclinic activity of the flow.

For daily precipitation maxima the dependence is much larger for both seasons, with a similar space scale for complete independence (about 200 km). Still the dependence at shorter distances is larger in winter than in summer.

For longer durations (up to 20 days), the dependence progressively increases. It is interesting to note that in winter, a plateau is reached for the madogram for all distances (between 5 to 15 days) corresponding to the typical time scales of the baroclinc waves. This saturation phase is supposed to be associated with the prevalence of these structures in winter. In summer, this plateau is not apparent but the decrease at very large distance is much less significant than in winter. This probably reflects the convective nature of the dynamics from which the maxima are extracted, inducing a very weak

dependence for distances larger than 200 km even for long durations.

Finally, the λ-madogram that captures the full dependence structure between pairwise maxima first confirms the results obtained above. But more importantly it suggests that the convergence toward independence is similar whatever are the values of λ.

This analysis can be extended is several ways. First the analysis of a much larger (homogeneous) region than Belgium can improve the quality of the madogram estimation, in particular for short durations from 1 to 12 h relevant for rapid and intense precipitation events such as flash floods. Another potential extension is to use the dependence information derived in this paper to build networks of recording stations for which maxima are independent, in order to apply for instance the station-year method allowing for an evalutation of maxima of long return periods (Buishand, 1984).

Appendix A

Sketch of the proof of Eq. (7)

First, classical probability theory tells us that the random variable $F_i(M_i)$ follows an uniform distribution on [0, 1]. Hence, it is easy to show that we have for the uniform distribution,

$$\mathbb{E}(F_i^{\alpha}(M_i)) = \frac{1}{1+\alpha}, \text{ for any } \alpha > 0. \tag{A1}$$

Because $|a-b|/2 = \max(a,b) - (a+b)/2$, we can write

$$\begin{aligned}\nu_{ik}(\lambda) &= \mathbb{E}(\max\left[F_i^{\lambda}(M_i), F_k^{1-\lambda}(M_k)\right]) \\ &\quad -\frac{1}{2}\mathbb{E}(F_i^{\lambda}(M_i)) - \frac{1}{2}\mathbb{E}(F_k^{1-\lambda}(M_k)), \\ &= A - \frac{1}{2}B - \frac{1}{2}C,\end{aligned}$$

with obvious notations for A, B and C. Applying Eq. (A1) allows us to derive that $B = 1/(1+\lambda)$ and $C = 1/(2-\lambda)$. Hence, $\frac{B+C}{2} = \frac{3}{2(1+\lambda)(2-\lambda)}$. Concerning A, we introduce the random variable $M = \max\left[F^{\lambda}(M_i), F^{1-\lambda}(M_k)\right]$. We also assume without loss of generality that $F_i = F_k$ are unit-Fréchet, i.e. $F_i(x) = \exp(-1/x)$. With this notation, we can write

$$\begin{aligned}\mathbb{P}(M \le u) &= \mathbb{P}\left(M_i \le \lambda F_i^{-1}(u), M_k \le (1-\lambda)F_k^{-1}(u)\right), \\ &= \exp\left(-\frac{1}{F_i^{-1}(u)} V_{ik}(\lambda, 1-\lambda)\right), \\ &\quad \text{by definition of } V_{ik}, \\ &= \left[\mathbb{P}\left(M_i \le F_i^{-1}(u)\right)\right]^{V_{ik}(\lambda,1-\lambda)} \\ &\quad \text{, since } F \text{ unit-Fréchet,} \\ &= u^{V_{ik}(\lambda,1-\lambda)}.\end{aligned}$$

It follows that

$$A = \frac{V_{ik}(\lambda, 1-\lambda)}{1 + V_{ik}(\lambda, 1-\lambda)}.$$

Hence, the equality

$$\nu_{ik}(\lambda) = \frac{1}{2}\mathbb{E}|F_i^{\lambda}(M_i(t)) - F_k^{1-\lambda}(M_k(t))|$$

follows.

Acknowledgements. Discussions with G. R. Demarée and the help of S. Derasse in handling the data are appreciated. This work was supported by the european E2-C2 grant, the National Science Foundation (grant: NSF-GMC (ATM-0327936)), by The Weather and Climate Impact Assessment Science Initiative at the National Center for Atmospheric Research (NCAR), the ANR-AssimilEx project and the Belgian Federal Science Policy Program under contract MO/34/017. The authors would also like to credit the contributors of the R project.

Edited by: P. Yiou
Reviewed by: three anonymous referees

References

Ancona-Navarrete, M. A. and Tawn, J. A.: Diagnostics for pairwise extremal dependence in spatial processes, Extremes, 5, 271–285, 2002.

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

Buishand, A.: Bivariate extreme-value data and the station-year method, J. Hydrol., 69, 77–95, 1984.

Buishand, A.: Bivariate extreme-value data and the station-year method, Statistica Neerlandica, 43, 1–30, 1989.

Capéraà, P., Fougères, A. L., and Genest, C.: A nonparametric estimation procedure for bivariate extreme value copulas, Biometrika, 84, 567–577, 1997.

Chilès, J.-P. and Delfiner, P.: Geostatistics: Modeling Spatial Uncertainty, John Wiley & Sons Inc., New York, A Wiley-Interscience Publication, 1999.

Coles, S. and Casson, E. A.: Spatial regression for extremes, Extremes, 1, 339–365, 1999.

Coles, S. G., Heffernan, J. E., and Tawn, J. A.: Dependence measures for extreme value analyses, Extremes, 2, 339–365, 1999.

Coles, S. G.: An introduction to statistical modeling of extreme values, Springer Series in Statistics, 208 p, 2001.

Coles, S. G. and Tawn, J. A.: Statistics of coastal flood prevention, Phil. Trans. R. Soc. Lond. A, 332, 457, 1990.

Cooley D., Naveau, P., and Poncet, P.: Variograms for spatial max-stable random fields, Chapter of the book Statistics for dependent data (Lecture Notes In Statistics, Springer), 2006.

Cooley, D., Nychka, D., and Naveau, P.: Bayesian Spatial Modeling of Extreme Precipitation Return Levels, J. Am. Statist. Assoc., in press, 2007.

Cressie, N. A. C.: Statistics for Spatial Data, John Wiley & Sons Inc., New York, 1993.

Davis, R. and Resnick, S.: Prediction of stationary max-stable processes, Ann. Applied Probab., 3, 497–525, 1993.

de Haan, L. and Pereira, T.: Spatial extremes: the stationary case, Ann. Statist., 34, 1, 146–168, 2006.

Demarée, G. R.: Le pluviographe centenaire du plateau d'Uccle: son histoire, ses données et ses applications, La Houille Blanche, 4, 1–8, 2003.

Dupriez, G. L. and Demarée, G. R.: Totaux pluviométriques sur des périodes continues de 1 à 30 jours. Vol. I: Analyse de 11 séries pluviométriques de plus de 80 ans, Miscelanea Série A, 8, 21 pp, 1988.

Dupriez, G. L. and Demarée, G. R.: Totaux pluviométriques sur des périodes continues de 1 à 30 jours. Vol. I: Analyse de 11 séries pluviométriques d'au moins 30 ans, Miscelanea Série A, 9, 27 pp., 1989.

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

Fougères, A.: Multivariate extremes, in: Extreme Values in Finance, Telecommunications and the Environment, edited by: Finkenstadt, B. and Rootzen, H., 373–388, Chapman and Hall CRC Press, London, 2004.

Gellens, D.: Trend and correlation analysis of k-day extreme precipitation over Belgium, Theor. Appl. Climatol., 66, 117–129, 2000.

Gellens, D.: Etude des précipitations extrêmes. Etablissement des fractiles et des périodes de retour dévénements pluviométriques, PhD thesis, ULB, 242 pp, 2003.

Gumbel, E. J.: Distributions des valeurs extrêmes en plusieurs dimensions, Publ. Inst. Statist. Univ. Paris, 9, 171–173, 1960.

Hall, P. and Tajvidi, N.: Distribution and dependence function estimation for bivariate extreme-value distributions, Bernoulli, 6, 835–844, 2000.

Heffernan, J. E. and Tawn, J. A.: A conditional approach for multivariate extreme values, J. Roy. Statist. Soc. Ser. B, 66, 497–546, 2004.

Hsing, T., Kluppelberg, C., and Kuhn, G.: Dependence estimation and vizualisation in multivariate extremes with applications to financial data, Extremes, 7, 99–121, 2004.

Joe, H.: Parametric family of multivariate distributions with given margins, J. Multivariate Anal., 46, 262–282, 1993.

Katz, R., Parlange, M., and Naveau, P.: Extremes in hydrology, Adv. Water Resour., 25, 1287–1304, 2002.

Kharin, V. V., Zwiers, F. W., Zhang, X., and Hegerl, G. C.: Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations, J. Climate, 20, 1419–1444, 2007.

Mann, H. B.: Nonparametric test against trend, Econometrica, 13, 245–259, 1945.

Mikosch, T.: Copulas: Tales and facts, Extremes, 9, 3–20, 2006.

Pickands III, J.: Multivariate extreme value distributions, Bull. Inst. Internat. Statist., 49, 859–878, 1981.

Resnick, S.: Extreme Values, Regular Variation, and Point Processes, Springer-Verlag, New York, 1987.

Schlather, M.: Models for stationary max-stable random fields, Extremes, 5, 33–44, 2002.

Schlather, M. and Tawn, J.: A dependence measure for multivariate and spatial extreme values: Properties and inference, Biometrika, 90, 139–156, 2003.

Schmitt, F., Vannitsem, S., and Barbosa, A.: Modeling of rainfall time series using two-state renewal processes and multifractals, J. Geophys. Res., 103, 23 181–23 193, 1998.

Smith, R.: Statistics of extremes, with applications in environment, insurance and finance, in: Extreme Values in Finance, Telecommunications and the Environment, edited by: Finkenstadt, B. and Rootzen, H., 1–78, Chapman and Hall CRC Press, London, 2004.

Sneyers, R.: Sur l'analyse statistique des séries d'observations, OMM, Note technique, 143, 192 p., 1975.

Sneyers, R., Vandiepenbeeck, M., and Vanlierde, R.: Principal component analysis of Belgian rainfall, Theor. Appl. Climatol., 39, 199–204, 1989.

Stein, M. L.: Interpolation of Spatial Data. Some theory for Kriging, Springer-Verlag, New York, 1999.

Tawn, J. A.: Bivariate extreme value theory: Models and estimation, Biometrika, 75, 397–415, 1988.

Tawn, J. A.: Modelling multivariate extreme value distributions, Biometrika, 77, 245–253, 1990.

Vannitsem, S.: Statistical properties of the temperature maxima in an intermediate order Quasi-Geostrophic model, Tellus, 59A, 80–95, 2007.

Wackernagel, H.: Multivariate Geostatistics. An Introduction with Applications, Springer, Heidelberg, third edition, 2003.