Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-102
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 23 April 2018

Natural Hazards and Earth System Sciences
Open Access
Discussions

On the relevance of extremal dependence for spatial statistical modelling of natural hazards

Laura C. Dawkins[1*] and David B. Stephenson[1]

[1] College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK

[*] E-mail: L.C.Dawkins@exeter.ac.uk

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-102
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 23 April 2018

Natural Hazards and Earth System Sciences
Open Access
Discussions

Abstract

Natural hazard loss portfolios with exposure over a region are sensitive to the dependency between extreme values of the key hazard variable at different spatial locations. It is therefore important to correctly identify and quantify dependency to avoid poor quantification of risk.

This study demonstrates how bivariate extreme value tail dependency methods can be used together in a novel way to explore and quantify extremal dependency in spatial hazard fields. A relationship between dependency and loss is obtained by deriving how the probability distribution of conceptual loss depends on the tail dependency coefficient.

The approaches are illustrated by applying them to 6103 historical European windstorm footprints (spatial maps of 3-day maximum gust speeds). We find there is little evidence of asymptotic extremal dependency in windstorm footprints. Furthermore, empirical extremal properties and conceptual loss distributions between pairs of locations are shown to be well reproduced using Gaussian copulas but not by extremally-dependent models such as Gumbel copulas.

It is conjectured that the lack of asymptotic dependence is a generic property of turbulent flows, which may extend to other spatially continuous hazards such as heat waves and air pollution. These results motivate the potential of using Gaussian process (geostatistical) models for efficient simulation of hazard fields.

Key Words: Natural hazards; Windstorm footprint; Bivariate dependence; Reinsurance; Copulas

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-102
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 23 April 2018

Natural Hazards and Earth System Sciences
Discussions

1 Introduction

Multivariate statistical models are increasingly used to explore the spatial characteristics of natural hazards and quantify potential risk. For example, multivariate statistical models for European windstorms are used by academics and re/insurers to create catalogues of possible events, explore loss potentials, and benchmark synthetic events from atmospheric models (Bonazzi et al. 2012; Youngman and Stephenson 2016). Since natural hazards are rare events in the tail of the distribution, correctly modelling extremal dependence is very important for valid inference (Eastoe et al., 2013), which, in turn, is essential for realistically representing potential hazard losses, often occurring at multiple spatial locations.

As noted by Wadsworth et al. (2017), examples of modelling joint extremes often assume asymptotic dependence in order to accommodate asymptotically justified extreme value max-stable models. This is also common in the field of natural hazard research. Coles and Walshaw (1994) used a max-stable model for the dependence in maximum wind speeds in different directions; Blanchet et al. (2009) to model snow fall in the Swiss Alps; Huser and Davison (2013) to model extreme rainfall and Bonazzi et al. (2012) to model windstorm hazard fields at pairs of locations in Europe. Indeed, Bonazzi et al. (2012) simply base this modelling assumption on being "in line with many examples found in the literature". Therefore, it seems sensible to ask, how valid is this assumption of asymptotic dependence? And how much of an effect might a misspecification of extremal dependence have on the resulting hazard loss representation?

Contrary to the above examples, Bortot et al. (2000) provided a brief exploration of the extremal dependence between measurements of sea surge, wave height and wave period recorded off the south-west coast of England. Using pairwise scatter plots and

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-102
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 23 April 2018

empirical estimates of the Extremal Dependence coefficients, χ and $\bar{\chi}$, introduced by Coles et al. (1999), they found evidence for asymptotic independence, and hence developed a multivariate Gaussian tail model for their data, derived from the joint tail of a multivariate Gaussian distribution with margins based on univariate extreme value distributions. Bortot et al. (2000) showed that, when modelling data that are asymptotically independent, the Gaussian model is robust, has simple diagnostics, easily interpretable parameters and extends straightforwardly to higher dimensions. Similarly, Youngman and Stephenson (2016) acknowledged the possibility of asymptotic independence when developing a spatial statistical framework for simulating natural hazard events. They specified a Student's t-process to model dependence, allowing for the form of extremal dependence to be determined by the estimated degrees of freedom parameter.

In this study, we provide a more rigorous, critical approach for investigating spatial extremal dependence, which combines various bivariate extreme value modelling methods in a novel way. We apply this approach to windstorm hazard fields to explore the validity of the asymptotic dependence assumption made in previous studies, and provide a turbulence argument for the form of extremal dependence found. Furthermore, we present a comparison approach for exploring the impact of mis-specifying extremal dependence on realistically representing conceptual windstorm losses.

To critically investigate spatial extremal dependence we initially employ the Extremal Dependence coefficients of Coles et al. (1999), $\chi(p)$ and $\bar{\chi}(p)$, characterising the conditional probability of a pair of locations exceeding the same high quantile threshold, $1-p$. The upper limit of these dependence measures determines the class of extremal dependence, hence this limit is explored both empirically, as in Bortot et al. (2000), and based on a number of parametric representations. For a given pair of locations within a windstorm hazard field, we fit Gumbel and Gaussian bivariate copula dependence models, each characterising apposing extremal dependence class, and explore how well these models represent the empirical estimates of $\chi(p)$ and $\bar{\chi}(p)$. In addition, we fit the bivariate tail model of Ledford and Tawn (1996), able to characterise both classes of extremal dependence, and use the extremal dependence diagnostic approach of Ledford and Tawn (1996, 1997) to further identify extremal dependence class based on the coefficient of tail dependence parameter of this model.

The impact of extremal dependence mis-specification on conceptual loss estimation is explored by comparing how well the Gumbel and Gaussian copula dependence models are able to represent empirical conceptual joint loss. Based on existing literature in the field of windstorm modelling, a quantile threshold exceedance loss model is proposed. The copula model comparison is then made in terms of their ability to represent $\chi(p)$ and $\bar{\chi}(p)$ for the specified conceptual loss quantile threshold, and the expected conditional conceptual joint loss distributions for pairs of locations throughout Europe.

In applying a similar critical investigation to an alternative continuous spatial variable, such as temperature, relevant for modelling heat wave risk, a natural hazard modeller will be able to diagnose the extremal dependence class present in the spatial field and explore the sensitivity of conceptual loss to a misspecification of this class. Rather than fitting a more complex model that can accommodate both types of extremal dependence, and therefore requires the estimation of additional parameters, this diagnostic approach allows the modeller to develop a model that specifically represents this statistical property, and hence the loss potentials, correctly; either using a max-stable model to characterise asymptotic dependence, as is most common in examples in the literature, or a Gaussian, or geostatistical model to characterise asymptotic independence, shown by Bortot et al.

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-102
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 23 April 2018

(2000) to be robust and accurate in the upper tail.

The remaining paper is organised as follows. The windstorm hazard data set, used throughout this paper, is described in Section 2. Our novel approach for critically investigating spatial extremal dependence is presented and applied to the windstorm hazard data in Section 3, including a physical explanation for the form of extremal dependence identified, in Section 3.4. Section 4 describes our novel method for exploring the impact of extremal dependence on conceptual loss estimation, again applied to the windstorm data set.

2 Data

The windstorm footprint data set used in this paper is the same as that used in Dawkins et al. (2016), consisting of the 6103 windstorm events that occurred within the European domain during the 35 extended winters, October - March 1979/80 - 2013/14 (kindly provided by J. Standen and J. F. Lockwood at the Met Office).

The windstorm footprint is defined as the maximum three second wind gust speed (in ms^{-1}) at grid points in the region 15 °W to 25 °E in longitude and 35 °N to 70 °N in latitude over a 72 hour period centred on the time at which the maximum 925hPa wind speed occurred over land. The 925hPa wind speed is taken from ERA-interim reanalysis (Dee et al., 2011). The three second wind gust speed has a robust relationship with storm damage (Klawa and Ulbrich, 2003), and is commonly used in catastrophe models for risk quantification (Roberts et al., 2014). A 72 hour windstorm duration is commonly used in the insurance industry (Haylock, 2011), and is thought to capture the most damaging phase of the windstorms (Roberts et al., 2014).

These 6103 historical windstorm events have been identified using the objective track-

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-102
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 23 April 2018

ing approach of Hodges (1995) and the associated footprints are created by dynamically downscaling ERA-Interim reanalysis to a horizontal resolution of 25km using the Met Office unified model (MetUM). The wind gust speeds are calculated from wind speeds in the MetUM model, based on a simple gust parameterisation $U_{gust} = U_{10m} + C\sigma$, where U_{10m} is the wind speed at 10 metre altitude, C is a constant determined from the universal turbulence spectra and σ is the standard deviation of the horizontal wind.

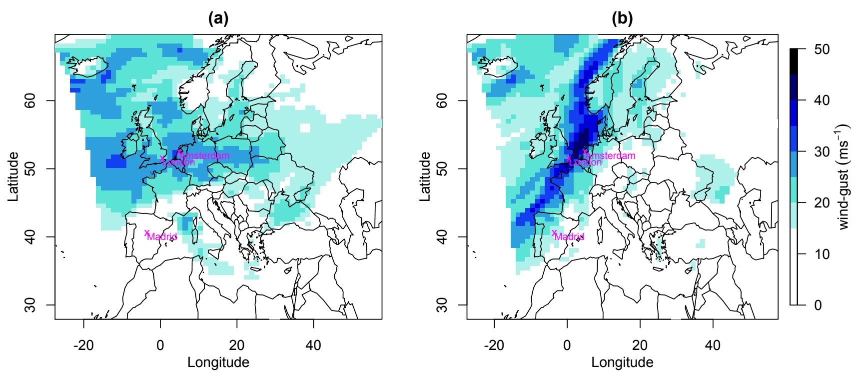

Figure 1: Hazard footprints for windstorms (a) Kyrill and (b) the Great Storm of October '87, with the location of the cities of London, Amsterdam and Madrid indicated.

Two such footprints for windstorms Kyrill ($17^{\mathrm{th}} - 19^{\mathrm{th}}$ January 2007) and the Great Storm of October '87 ($15^{\mathrm{th}} - 17^{\mathrm{th}}$ October 1987) are shown in Fig. 1. The variability in the intensity and location of extreme, damaging winds in these footprints demonstrate the potential importance of correctly modelling the spatial dependence between locations for realistically representing joint losses.

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-102
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 23 April 2018

Natural Hazards and Earth System Sciences
Open Access
Discussions

3 Extremal Dependency

We show how various bivariate extreme value modelling methods can be brought together in a novel way to critically investigate the bivariate extremal dependence property of windstorm hazard fields. This approach is illustrated based windstorm footprint wind gust speeds at two pairs of locations, London-Amsterdam and London-Madrid. These three locations are shown in Fig. 1, and these two pairings are chosen because of their contrasting separation distances and directions, and hence degrees of dependence.

3.1 Graphical summary using the empirical copula

As a motivating example, the bivariate dependence in windstorm footprint wind gust speeds for London paired with Amsterdam and Madrid are presented in Figures 2 (a) and (c) respectively. These scatter plots show a greater degree of dependence between London and Amsterdam compared to London and Madrid. Indeed, multiple windstorms have losses occurring in London and Amsterdam at the same time, when loss is associated with wind gust speeds exceeding the 99% quantile at a given location, characterised by the top right-hand corner of each plot in Fig. 2. However, does this level of dependence between London and Amsterdam necessarily suggest asymptotic dependence?

Let the $n \times 2$ variable (X, Y) represent the wind gust speeds associated with the $n = 6103$ windstorm events at any given pair of locations within the European domain, e.g. London and Amsterdam. The bivariate relationship between X and Y can be represented by two components, the marginal distributions of each variable, and their joint dependence. The dependence component of the relationships shown in Fig. 2 (a) and (c) can therefore be isolated by, for each location, transforming wind gust speeds associated with each of the windstorm events, e.g. X_i for $i = 1, ..., n$, to uniform margins

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-102
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 23 April 2018

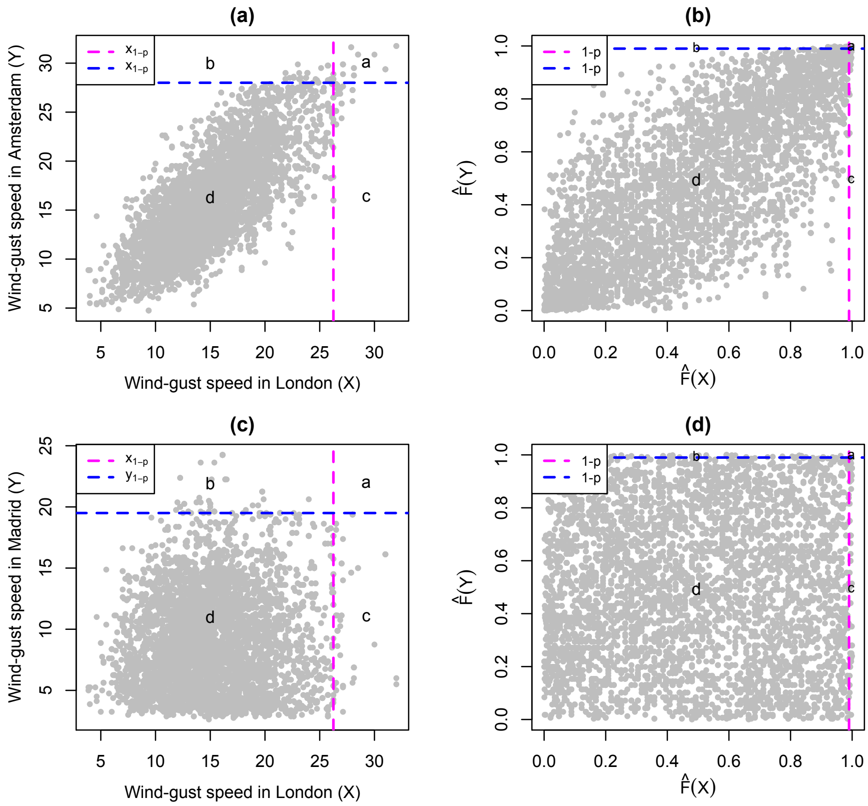

Figure 2: Scatter plot/empirical copulas comparing historical windstorm footprint wind gust speeds (ms^{-1}) in London paired with (a)/(b) Amsterdam and (c)/(d) Madrid. Dashed lines show the 99% quantile of wind gust speed at each location, and labels a-d represent the number of points in each section of each plot, related to being above or below these high quantile thresholds.

using the estimator of the empirical distribution function ($\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}_{X_j \leq X_i}$), shown in Fig. 2 (b) and (d) respectively. This is known as the empirical copula.

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-102
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 23 April 2018

Natural Hazards and Earth System Sciences
Discussions
Open Access

3.2 Measures for quantifying extremal dependence

The degree of conditional dependence between locations, at a specified high quantile threshold, $1 - p$, can then be explored, based on the empirical copula, using the Extremal Dependence Coefficients, $\chi(p)$ and $\bar{\chi}(p)$, introduced by Coles et al. (1999), and the asymptotic limit of these measures, as $p \to 0$, classifies the class of bivariate extremal dependence as either *asymptotically dependent* or *asymptotically independent*. These measures are defined as,

$$\chi(p) = \Pr(Y > y_{1-p}|X > x_{1-p}) = \frac{\Pr(Y > y_{1-p}, X > x_{1-p})}{p}, \tag{1}$$

where x_{1-p} and y_{1-p} are the $(1-p)^{th}$ quantiles of X and Y respectively, $0 \leq \chi(p) < 1$ for all $0 \leq (1-p) \geq 1$, and,

$$\bar{\chi}(p) = \frac{2\log(\Pr(X > x_{1-p}))}{\log(\Pr(X > x_{1-p}, Y > y_{1-p}))} - 1 = \frac{2\log(p)}{\log(\chi(p)p)} - 1 = \frac{\log(p) - \log(\chi(p))}{\log(p) + \log(\chi(p))}, \tag{2}$$

where $-1 \leq \bar{\chi}(p) < 1$ for all $0 \leq (1-p) \leq 1$.

Hence, if $\lim_{p\to 0} \chi(p) = \chi(0) > 0$, $\lim_{p\to 0} \bar{\chi}(p) = \bar{\chi}(0) = 1$, and the pair (X, Y) are said to be asymptotically dependent with strength $\chi(0)$. If instead $\chi(0) = 0$, and hence, $\bar{\chi}(0) < 1$, the pair are said to be asymptotically independent, and the non-vanishing measure $\bar{\chi}(0)$ represents the strength of this non-asymptotic dependence.

As an initial empirical exploration of bivariate extremal dependence class between variables, these conditional probability measures can be calculated empirically over a range of quantile thresholds, as shown in Fig. 3 for windstorm footprint wind gust speeds in London paired with Amsterdam and Madrid. These empirical estimates are calculated as functions of the counts (a,b,c,d) in Fig. 2, as defined in Table 1. Based on these em-

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-102
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 23 April 2018

Natural Hazards and Earth System Sciences
Open Access
Discussions

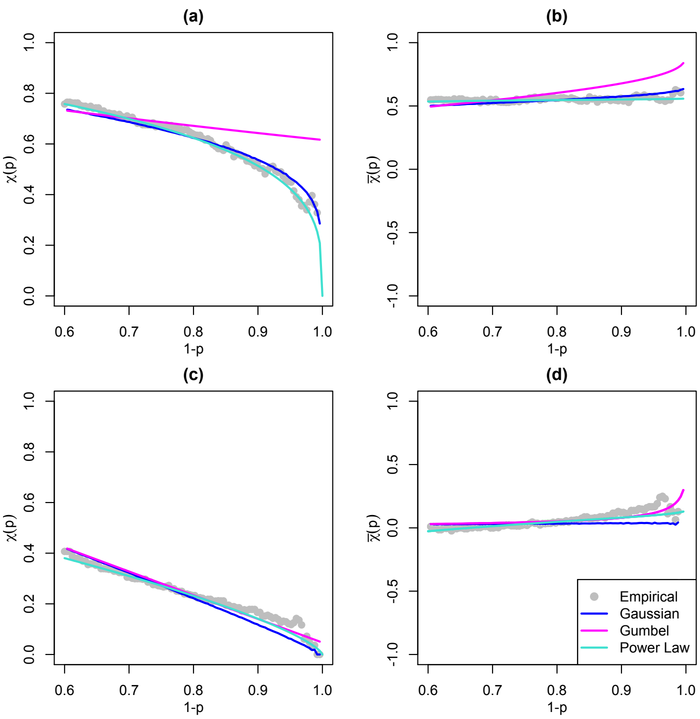

Figure 3: Dependence measures, $\chi(p)/\bar{\chi}(p)$, for $p \in [0, 0.4]$, for windstorm footprint wind gust speeds in London paired with (a)/(b) Amsterdam and (c)/(d) Madrid, calculated empirically and based on the Gaussian, Gumbel and Power Law bivariate dependence functions, as defined in Table 1.

pirical estimates, for both pairs of locations, $\chi(p) \to 0$ and $\bar{\chi}(p) < 1$ as $p \to 0$, suggesting asymptotic independence. In addition, Fig. 3 presents three parametric representations of these conditional probability measures, used to better explore the asymptotic limits, $\chi(0)$ and $\bar{\chi}(0)$. These parametric forms are also shown in Table 1 and are discussed further in Section 3.3.

3.3 Estimation of the tail dependence coefficient

The rarity of extreme events within this historical data set makes it impossible to empirically quantify the asymptotic limits $\chi(0)$ and $\bar{\chi}(0)$, necessary for extremal dependence class identification. To overcome this, Ledford and Tawn (1996) developed a bivariate tail model, able to characterise both classes of extremal dependence, which when fit to a bivariate random variable can be used to model the asymptotic limit of the conditional probability measures and specify the class of extremal dependence.

As in Ledford and Tawn (1996), let Z_1 and Z_2 denote X and Y transformed to unit Fréchet margins respectively, that is $\Pr(Z_1 \le z) = \Pr(Z_2 \le z) = \exp(-1/z)$. Then the joint survivor function for Z_1 and Z_2, above some large quantile threshold z_{1-p}, takes the form,

$$\Pr(Z_1 > z_{1-p}, Z_2 > z_{1-p}) \sim \mathcal{L}(z_{1-p}) p^{1/\eta}, \tag{3}$$

where $p = \Pr(Z_1 > z_{1-p}) = \Pr(Z_2 > z_{1-p})$, $\frac{1}{2} \le \eta \le 1$ is a constant and $\mathcal{L}(z_{1-p})$ is a slowly varying function as $p \to 0$. Based on this power law model, as shown by Coles et al. (1999),

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-102
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 23 April 2018

Natural Hazards and Earth System Sciences
Open Access
Discussions

$$\chi(p) \sim \mathcal{L}(z_{1-p})p^{1/\eta-1},$$

$$\bar{\chi}(p) = \frac{2\log(p)}{\log(\mathcal{L}(z_{1-p})) + \frac{1}{\eta}\log(p)} - 1,$$

$$\to 2\eta - 1 \quad \text{as } p \to 0.$$

Hence, the parameter η, named the coefficient of tail dependence by Ledford and Tawn (1996), characterises the nature of the asymptotic dependence. When $\eta = 1$, $\chi(0) = \mathcal{L}(z_{1-p}$ and $\bar{\chi}(0) = 1$, hence the pair (X, Y) are asymptotically dependent of degree $\mathcal{L}(z_{1-p})$. Alternatively, if $\eta < 1$, $\chi(0) = 0$ and $\bar{\chi}(0) = 2\eta - 1)$, and the pair are asymptotically independent with non-asymptotic dependence of degree $2\eta - 1$.

For a given pair, e.g. wind gust speeds in London and Amsterdam, the Ledford and Tawn (1996) is fit to the joint survivor function along the diagonal, equivalent to the univariate distribution of $T = \min\{Z_1, Z_2\}$, known as the structure variable, which has length n. Using the stable two parameter Poisson process representation of T, presented by Ferro (2007), who employed the Ledford and Tawn (1996) model for the verification of extreme weather forecasts, the exceedance of T above a high threshold w has the form,

$$\Pr(T > t) = \frac{1}{n}\exp\left[-\left(\frac{t-\alpha}{\eta}\right)\right] \quad \text{for all } t \geq w, \tag{4}$$

with location parameter α and scale parameter $0 < \eta \leq 1$, equivalent to η in Eqn. (3), estimated by maximum likelihood (Ferro, 2007).

We fit this model to the pairs London-Amsterdam and London-Madrid, requiring the specification of the high threshold, w. This threshold must be high enough that this asymptotic model is valid, but low enough that enough data are used to estimate the parameters. Here, the 85% quantile of the structural variable T is selected, based on stability plots of quantile thresholds w against model parameter estimates obtained by fitting the model to exceedances of w. Based on this choice of w, $\eta = 0.78 < 1$ for London-Amsterdam and $\eta = 0.58 < 1$ for London-Madrid, indicating asymptotic independence for both pairs of locations. Figure 3 shows how these fitted models represent the asymptotic limit of the conditional dependence measures $\chi(p)$ and $\bar{\chi}(p)$ as $p \to 0$, the Poisson process form of which are presented in Table 1, referred to as the Power Law model.

In addition, alternative parametric bivariate dependence models, known as the Gaussian and Gumbel copulas, characterising apposing extremal dependence class, can be used to model the pair (X, Y), and hence $\chi(p)$ and $\bar{\chi}(p)$, for comparison, also shown in Fig. 3. The representation of $\chi(p)$ and $\bar{\chi}(p)$ in the limit $p \to 0$ for these apposing models then gives further indication of the extremal dependence class.

The Gumbel bivariate copula model characterises asymptotic dependence with the degree of dependence quantified by parameter r. For each pair of locations, this parameter is estimated via maximum likelihood using the `copula` R package. The Gaussian bivariate model characterises asymptotic independence with dependence parameter ρ, here, for each pair of locations, represented by the Spearman's rank correlation coefficient. The parametric forms of $\chi(p)$ and $\bar{\chi}(p)$ for these two apposing models are shown in Table 1. In Fig. 3, the Gumbel model is calculated as in Table 1, however, since the closed form definition for the Gaussian model in Table 1 only holds for the limit $p \to 0$, for this model $\chi(p)$ and $\bar{\chi}(p)$ are estimated as the median of 1000 parametric bootstrap simulations from the associated bivariate normal distribution.

For both pairs of locations in Fig. 3, the Power Law model characterises asymptotic independence, since $\chi(0) = 0$ and $\bar{\chi}(0) < 1$. Asymptotic independence is further supported by the success of the Gaussian, and failure of the Gumbel, dependence models in

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-102
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 23 April 2018

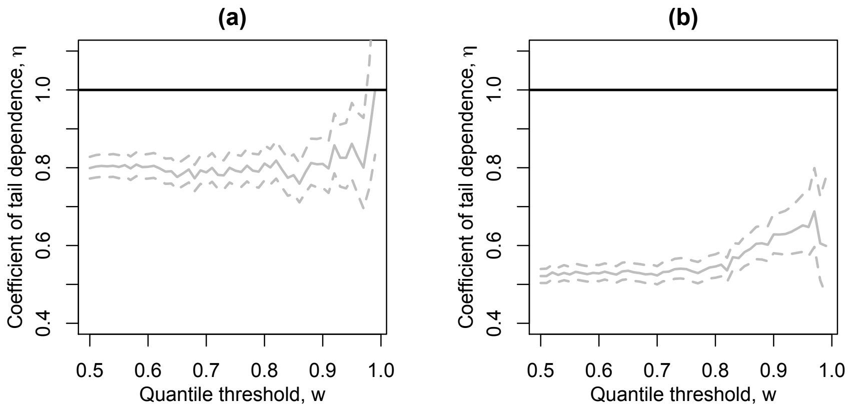

Figure 4: Diagnostic plots of maximum likelihood estimates (solid) and 95% profile likelihood confidence intervals (dashed) of η, in Eqn. (4), for threshold w in the range of the $0.5-1$ quantile of T, for London paired with (a) Amsterdam and (b) Madrid.

capturing the empirical bivariate dependence structure in the upper limit, $p \to 0$. The Gumbel model overestimates the conditional probability of joint extremes due to a misspecification of asymptotic dependence, while the Gaussian model matches closely with the empirical estimates and the Power Law model.

The Power Law model shown in Fig. 3 is based on using a high threshold, w, equal to the 85% quantile of the structural variable T. The resulting estimate of η, and hence the identification of extremal dependence class, depends on this threshold choice, hence the sensitivity of this diagnosis to the choice of threshold must be explored before a conclusion can be reached. As in Ledford and Tawn (1996, 1997), here this is done by observing the proportion of time $\eta = 1$ is within the profile likelihood confidence interval for η, when estimated over a range of values of w. The pair (X, Y) are said to be asymptotically dependent if $\eta = 1$ is contained within these confidence intervals for a majority of the range of w, and asymptotically independent otherwise. This exploration is presented for London paired with Amsterdam and Madrid in Fig. 4, confirming the diagnosis of asymptotic independence for both pairs, based on this criterion. Indeed, the same conclusion is reached when applying this methodology to additional pairs of land locations within the European domain, including neighbouring locations.

3.4 Why are wind gust speeds asymptotically independent?

It is of interest to ask whether there are fundamental fluid dynamical reasons for why wind gust speeds should be asymptotically independent at different spatial locations.

One approach to answering this question is to consider extremal dependence in turbulent flows. The atmospheric flow in storm track regions is highly chaotic and irregular and is therefore turbulent rather than smoothly varying laminar flow (see Held 1999; and references therein). Furthermore, over short enough spatial distances, the horizontal flow in a storm may be considered to be stationary in space and directionally invariant, in other words, homogeneous isotropic turbulence.

As originally proposed by Von Kármán (1937), turbulent wind fields can be efficiently and realistically simulated using stochastic processes (Mann, 1998). This approach is widely used for many applications such as testing loads on new aircraft designs. The basic assumption in homogeneous turbulence is that the Cartesian velocity components are independent Gaussian processes, each with a prescribed spatial covariance function. In the special case of isotropic turbulence, the spatial covariance functions are identical for each velocity component. Hence, for 2-dimensional windstorm gusts, the wind gust speed at spatial location, s, is given by $X(s) = \sqrt{u^2 + v^2}$, where $u = u(s)$ and $v = v(s)$ are independent Gaussian processes having identical covariance functions.

So what can be deduced about the extremal dependence class of wind speeds from such turbulence models? Firstly, since the individual velocity components are bivariate normal, they are asymptotically independent at different locations e.g. $u_1 = u(s_1)$ and $u_2 = u(s_2)$ are asymptotically independent when s_1 differs from s_2, and likewise for $v(s)$. Furthermore, the square of each velocity component is also asymptotically independent. This can be proven by noting that $\Pr(u_1^2 > t^2) = 2\Pr(u_1 > t)$ and $\Pr(u_1^2 > t^2, u_2^2 > t^2) \leq 4\chi_{max}\Pr(u_1 > t)$ where

$$\chi_{max} = \frac{\max(\Pr(u_1 > t, u_2 > t), \Pr(u_1 > t, u_2 \leq -t))}{\Pr(u_1 > t)},$$

and for bivariate normal velocity components $\chi_{\max} \to 0$ as the threshold $t \to \infty$. The squared wind speeds at pairs of locations are sums of two such independent components, $(X^2, Y^2) = (u_1^2 + v_1^2, u_2^2 + v_2^2)$, and so it would be surprising if somehow this pair were not also asymptotically independent.

Unfortunately a proof of asymptotic independence between (X^2, Y^2) (and hence (X, Y)) remains elusive. However, the conjecture can be tested by numerical simulation. By simulating velocities from bivariate normal distributions, we have found no evidence of extremal dependence in wind speeds even when each velocity component is highly correlated. Figure 5 shows an example obtained by simulating a million wind speeds at two locations where the u and v velocity components are independent standard normal variates each with correlation of 0.9 between locations (i.e. the correlation between u_1 and u_2 is 0.9). The squared wind speeds at each location are chi-squared distributed with 2 degrees of freedom but are not independent: there is positive association clearly visible in Fig. 5(a). To assess extremal dependence, Fig. 5(b) shows how the joint exceedance probability, $\Pr(X^2 > t^2, Y^2 > t^2)$, and the marginal exceedance probability, $\Pr(X^2 > t^2) = \Pr(Y^2 > t^2)$, behave as threshold t^2 is varied. As the threshold is increased

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-102
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 23 April 2018

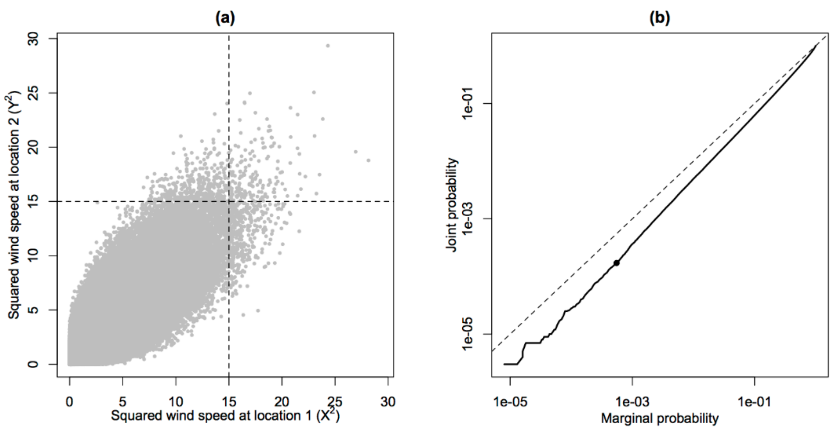

Figure 5: Simulation of wind speeds at two sites having highly correlated velocities (see main text for details): (a) scatter plot of squared wind speeds at the two sites (1000 points randomly sampled out of the million); (b) joint versus marginal exceedance probabilities (on logarithmic axes). The dot shows an example obtained by counting the fraction of points in the upper right and the right hand quadrants of (a). The curve has a steeper slope than the dashed line (equal probabilities denoting complete dependence) suggesting asymptotic independence.

the joint probability drops to zero faster than the marginal exceedance probability (the curve in Fig. 5(b) is steeper than the dashed line), which suggests that the ratio, the conditional probability of exceedance, equivalent to χ in Eqn. (1), will tend to zero in the asymptotic limit.

4 Loss Distributions

To explore the importance of correctly modelled extremal dependence on the distribution of a conceptual loss, we first define a conceptual loss function and then use it to compare how well the Gumbel and Gaussian copula dependence models, characterising apposing extremal dependence class, are able to represent empirical conceptual losses.

In the absence of insurance industry exposure and vulnerability information, we define conceptual windstorm loss as a function of the footprint wind gust speeds, similar to many examples in the literature (see Dawkins et al. (2016) for a review). Following the conclusions of Roberts et al. (2014) and Dawkins et al. (2016), we propose a threshold exceedance conceptual loss function over land locations in Europe. Roberts et al. (2014) and Dawkins et al. (2016) showed that this form of loss function is more representative of extreme windstorm loss than the commonly used, more complex loss function of Klawa and Ulbrich (2003), which includes additional terms for cubed wind gust speed magnitude and population density. Roberts et al. (2014) used an exceedance threshold of 25ms^{-1} while Dawkins et al. (2016) used a threshold of 20ms^{-1}, in line with the loss threshold used by German insurance companies (Klawa and Ulbrich, 2003). Here, however, similar to Klawa and Ulbrich (2003), we propose a locally varying wind gust speed quantile threshold, accounting for local adaptation to varying wind intensity.

Figure 6 shows that the 99% quantile of windstorm footprint wind gust speed is in

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-102
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 23 April 2018

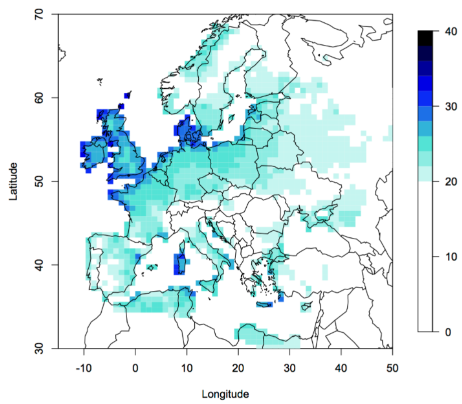

Figure 6: The 99% quantile of windstorm footprint wind gust speeds (ms[1]) at land locations in Europe, used as the threshold above which windstorm conceptual insured losses occur.

excess of the commonly used 20ms^{-1} loss threshold for most land locations in Europe, with a higher loss threshold used in regions where stronger winds occur. Hence, we define our bivariate conceptual loss function for the pair (X, Y) as,

$$L(X, Y) = H(X - x_{0.99}) + H(Y - y_{0.99}),$$

where $H(m)$ is a Heaviside function: $H(m) = 1$ if $m > 0$ and $H(m) = 0$ otherwise. Hence, for a given pair of locations the conceptual loss can take the values 0, 1 or 2 depending on the joint exceedance of X and Y above their respective 99% quantiles, $x_{0.99}$ and $y_{0.99}$, equivalent to falling within the four sections in Fig. 2.

The probability mass function of the conceptual loss function can easily be obtained by considering the joint probability of (X, Y) in each of the quadrants shown in Fig. 2:

$$\Pr(L(X,Y)=2)=\chi(p)p,$$

$$\Pr(L(X,Y)=1)=2(1-\chi(p))p,$$

$$\Pr(L(X,Y)=0)=1+p(\chi(p)-2).$$

From this, it is straightforward to derive the following 1st and 2nd moments:

$$\begin{aligned}\mathbb{E}(L(X,Y)) &= 2\chi(p)p+2(1-\chi(p))p=2p,\\ \mathbb{E}(L(X,Y)|L(X)=1) &= 2\chi(p)p+(1-\chi(p))p=p(\chi(p)+1),\\ \mathrm{Var}(L(X,Y)) &= 2(1+\chi(p))p-4p^2=2p(1+2\chi(p)-2p).\end{aligned} \tag{5}$$

Although the expected loss does not depend on $\chi(p)$, the conditional expectation and the variance of loss both depend on $\chi(p)$ as well as the marginal probability of exceedance p. The conditional expectation and the variance increase considerably as χ goes from 0 to 1.

We will now illustrate this by comparison of different extremal dependence class models. For London paired with each land location (grid cell) in Europe, we fit both the Gumbel and Gaussian copula dependence models, and explore how well these two models represent empirical conceptual losses. This is done in two ways.

Firstly by comparing the empirical, Gaussian and Gumbel estimates for $\chi(p)$ and $\bar{\chi}(p)$ for $p=0.01$ (Table 1), i.e. representing measures of the conditional probability of a loss occurring in Y (e.g. Amsterdam), given a loss has occurred in X (e.g. London), shown in Fig. 7. Secondly, as presented in Fig. 8, by comparing the distribution, for all land locations in the European domain, of the expected conditional joint loss with London, given a loss has occurred in London (Eqn. 5), again calculated empirically and using the

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-102
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 23 April 2018

Natural Hazards and Earth System Sciences
Open Access
Discussions

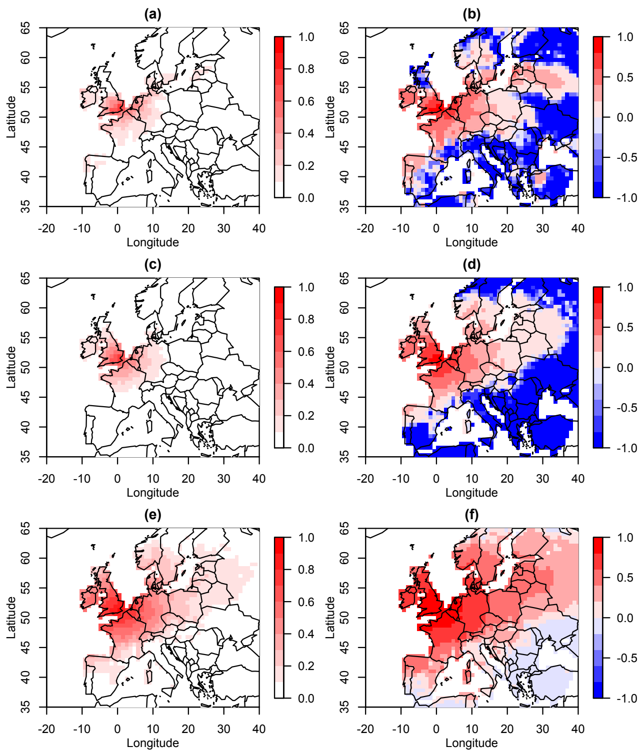

Figure 7: (a)/(b) Empirical, (c)/(d) Gaussian, and (e)/(f) Gumbel estimates of $\chi(0.01)/\bar{\chi}(0.01)$, representing measures of the conditional probability of a conceptual loss occurring at each land location in the European domain, given a conceptual loss has occurred in London.

two apposing dependence models.

Figure 7 identifies a large over estimation in both $\chi(0.01)$ and $\bar{\chi}(0.01)$ for the Gumbel dependence model, when compared with the empirical estimates, due to a misspecification of asymptotic dependence between locations. The Gaussian model, on the other hand, which correctly represents the identified asymptotic independence between locations, provides a good representation of the empirical conditional loss measures, with the spatial range of $\chi(0.01)$ >0 and areas of positive and negative $\bar{\chi}(0.01)$ in general

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-102
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 23 April 2018

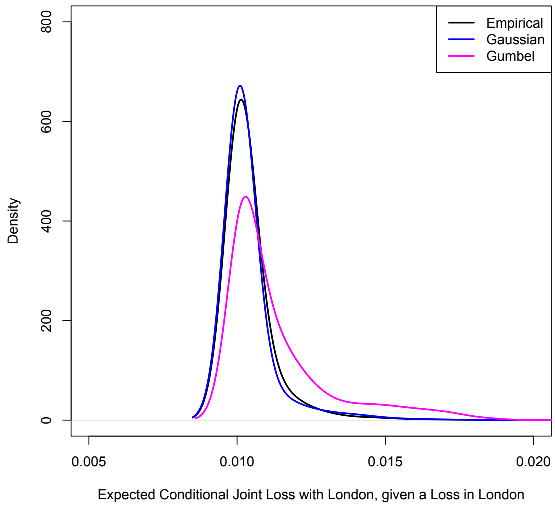

Figure 8: For all land locations in the European domain, the expected conditional joint loss with London, given a loss has occurred in London (Eqn. 5), calculated empirically and using the Gaussian and Gumbel copula models.

agreement with the empirical values.

Figure 8, further illustrate the importance of correctly specifying extremal dependence class when representing loss. When a conceptual loss occurs in London, the Gumbel dependence model over estimates the expected conditional joint loss with other European land locations, while conversely, the Gaussian model provides a very good estimate of the empirical expected conditional joint loss distribution.

5 Conclusion

This study has shown how to explore and identify extremal dependence in hazard fields using extremal dependence coefficients, $\chi(p)$ and $\bar{\chi}(p)$, and estimates of the tail dependence parameter Ledford and Tawn (1996). These measures have been compared to what one would expect from Gaussian and Gumbel copulas.

These methods have revealed strong evidence of asymptotic independence in windstorm footprint hazard fields, contrary to what has been assumed in previous studies such as Bonazzi et al. (2012). A reason for this lack of asymptotic dependency has been proposed based on arguments from turbulence theory. It is shown that mis-specification of the dependency (e.g. by using a Gumbel copula) leads to severe over-estimation of the probability of joint losses.

These results provide justification that spatial representation and simulation of windstorm hazard fields can be represented by a Gaussian geostatistical model, such as that developed in Chapter 5 of Dawkins (2016), rather than a max-stable, asymptotically dependent model.

Acknowledgements

Laura C. Dawkins was supported by the Natural Environment Research Council (Consortium on Risk in the Environment: Diagnostics, Integration, Benchmarking, Learning and Elicitation (CREDIBLE project); NE/J017043/1).

References

Blanchet, J., Marty, C., and Lehning, M. (2009). Extreme value statistics of snowfall in the Swiss Alpine region. *Water Resources Research*, 45.

Bonazzi, A., Cusack, S., Mitas, C., and Jewson, S. (2012). The spatial structure of European wind storms as characterized by bivariate extreme-value copulas. *Nat. Hazards Earth Syst. Sci.*, 12:1769–1782.

Bortot, P., Coles, S., and Tawn, J. (2000). The multivariate gaussian tail model: an application to oceanographic data. *Applied Statistics*, 49:31–49.

Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analysis. *Extremes*, 2(4):339–365.

Coles, S. G. and Walshaw, D. (1994). Directional modelling of extreme wind speeds. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43(1):139–157.

Dawkins, L. C. (2016). *Statistical modelling of European windstorm footprints to explore hazard characteristics and insured loss*. PhD thesis, University of Exeter, College of Engineering Mathematics and Physical Sciences.

Dawkins, L. C., Stephenson, D. B., Lockwood, J. F., and Maisey, P. E. (2016). The 21st century decline in damaging european windstorms. *Natural Hazards and Earth System Science*, 16:1999–2007.

Dee, D., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. N., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J. N., and Vitart, F. (2011). The ERA interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Roy. Meteorol. Soc.*, 137:553–597.

Eastoe, E. F., Koukoulas, S., and Jonathan, P. (2013). Statistical measures of extremal dependence illustrated using measured sea surface elevations from a neighbourhood of coastal locations. *Ocean Engineering*, 62:68–77.

Ferro, C. A. T. (2007). A probability model for veryfiying deterministic forecasts of extreme events. *Weather Forecasting*, 22:1089–1100.

Haylock, M. (2011). European extra-tropical storm damage risk from a multi-model ensemble of dynamically-downscaled global climate models. *Natural Hazards and Earth System Science*, 11:2847–2857.

Held, I. M. (1999). The macroturbulence of the troposphere. *Tellus A: Dynamic Meteorology and Oceanography*, 51A-B:59–70.

Hodges, K. I. (1995). Feature tracking on the unit sphere. *Monthly Weather Review*, 123:3458–3465.

Huser, R. and Davison, A. C. (2013). Space–time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Methodology)*, 76(2):439–461.

Klawa, M. and Ulbrich, U. (2003). A model for the estimation of storm losses and the identification of severe winter storms in Germany. *Natural Hazards and Earth System Sciences*, 3:725–732.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society*, 59(2):475–499.

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-102
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 23 April 2018

Natural Hazards and Earth System Sciences
Open Access
Discussions

Mann, J. (1998). Wind field simulation. *Probabilistic engineering mechanics*, 13(4):269–282.

Roberts, J. F., Dawkins, L., Youngman, B., Champion, A., Shaffrey, L., Thornton, H., Stevenson, D. B., Hodges, K. I., and Stringer, M. (2014). The XWS open access catalogue of extreme windstorms in Europe from 1979 to 2012. *Natural Hazards and Earth System Science*, 14:2487–2501.

Von Kármán, T. (1937). The fundamentals of the statistical theory of turbulence. *Journal of the Aeronautical Sciences*, 4(4):131–138.

Wadsworth, J. L., Tawn, J. A., Davison, A. C., and Elton, D. M. (2017). Modelling across extremal dependence classes. *Journal of the Royal Statistical Society: Series B (Methodology)*, 79(1):149–175.

Youngman, B. D. and Stephenson, D. B. (2016). A geostatistical extreme-value framework for fast simulation of natural hazard events. *Proceedings of the Royal Society A*, 472(20150855).

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-102
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 23 April 2018

Table 1: Empirical and Parametric forms for extremal dependence measures $\chi(p)$ and $\bar{\chi}(p)$.

	$\chi(p)$	$\bar{\chi}(p)$
Empirical	$\frac{a}{a+c}$	$\frac{2\log(a+c)/n}{\log(a/n)} - 1$
Power Law	$\frac{1}{n}\exp\left(\frac{\alpha}{\eta}\right)p^{\frac{1}{\eta}-1}$	$\frac{2\log(p)}{\log\left(\frac{1}{n}\exp\left(\frac{\alpha}{\eta}\right)\right)+\frac{1}{\eta}\log(p)} - 1$
Gumbel	$\sim 2 - \frac{(2\log(1-p)^r)^{\frac{1}{r}}}{\log(1-p)} = 2 - 2^{\frac{1}{r}}$ (Coles et al., 1999)	$\frac{2\log(p)}{\log(2p(1-p)^2)} - 1$
Gaussian	$\bar{F}(1-p,1-p)/p$, where $\bar{F}(1-p,1-p) = Pr(X > x_{1-p}, Y > y_{1-p}) \sim (1+\rho)^{\frac{3}{2}}(1-\rho)^{\frac{1}{2}}(4\pi)^{-\frac{\rho}{1+\rho}}(-\log(p))^{\frac{\rho}{1+\rho}}p^{\frac{2}{1+\rho}}$ as $p \to 0$ (Coles et al., 1999)	$\frac{2\log(p)}{\log(\bar{F}(1-p,1-p))} - 1$