Stat

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.71

Empirical Bayes estimation for the conditional extreme value model

Linyin Cheng[a], Eric Gilleland[b], Matthew J. Heaton[c] and Amir AghaKouchak[a]*

Received 19 October 2014; Accepted 18 November 2014

A new estimation strategy for estimating the parameters of the Heffernan and Tawn conditional extreme value model is proposed. The technique makes use of empirical Bayes estimation for the conditional likelihood that otherwise does not have a simple closed-form expression. The approach is tested on simulations from different types of extreme dependence (and independence) structures, as well as for two real data cases consisting of precipitation analysis conditional on extreme temperature in Boulder, Colorado, and Los Angeles, California, USA. The strategy generally has good coverage when informative priors are used for one of the parameters, except for the independence case where the coverage is low until the sample size reaches about 50. Results for the precipitation and temperature data are found to be consistent with the semi-non-parametric strategy. The presented model can be potentially applied in a wide variety of science fields, especially in earth, environment and climate sciences. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: bivariate extremes; conditional extreme value model; empirical Bayes estimation

1 Introduction

Heffernan & Tawn (2004) introduced an important new methodology for modelling multivariate extreme values through a conditional distribution framework that has certain advantages over the usual multivariate extreme value analysis techniques. For example, it is based on threshold exceedences, rather than block maxima, thereby allowing more of the available data to be used for estimation. Moreover, it enables one to model cases whether they are asymptotically dependent or independent using one unified framework.

A drawback to the approach concerns the semi-non-parametric estimation procedure for the model as no simple, closed-form distribution exists, in general, without assuming a specific dependence structure for the conditional distribution. The initially proposed approach involves maximum likelihood estimation for the marginal distributions fit to data above a high threshold, Gaussian estimation for the conditional means and standard deviations, and pseudo-likelihood estimation for combining the conditional distributions into a multivariate family. A relatively complicated bootstrap procedure is employed to account for all of the levels of uncertainty in the model. While some advances have been made to this procedure, there still lacks a more fluid estimation and inference strategy for the conditional extreme value model.

[a]Center for Hydrometeorology and Remote Sensing, University of California Irvine, Irvine, CA 92697, USA
[b]Research Applications Laboratory, National Center for Atmospheric Research, PO Box 3000, Boulder, CO 80307, USA
[c]Department of Statistics, Brigham Young University, Provo, UT 84602, USA
*Email: amir.a@uci.edu

Stat

(wileyonlinelibrary.com) DOI: 10.1002/sta4.71

The ISI's Journal for the Rapid
Dissemination of Statistics Research

To circumvent these issues, we propose to couch the entire estimation procedure into a Bayesian framework. To handle the lack of a closed-form likelihood for the conditional model, empirical Bayes techniques are used, which Owen (1988) developed as a robust alternative to classical likelihood approaches or the bootstrap (cf. Owen, 1990; Mengersen et al., 2013). It was demonstrated that, for some categories of statistical models, when the likelihood function is numerically unavailable or not entirely known, empirical likelihood methods can be used to bypass simulations from the model while converging in the number of observations. Empirical likelihood has been shown in a wide range of situations to have properties analogous to a real likelihood (cf. Li, 1995; Jing, 1995; Chen, 1994; Qin & Lawless, 1994; Hall & La Scala, 1990). Although further investigation of this methodology is needed, it appears to be a valuable approach in distribution-free contexts.

It is found, here, through simulation results that the approach performs fairly well for most types of dependence structures. In some cases, strong prior information is required but is also readily obtained through some relatively simple techniques. Results are also compared with the estimation strategy described by Heffernan & Tawn (2004) using the Laplace transformation and a constrained likelihood per Keef et al. (2013a), and our proposed strategy is found to yield consistent estimated parameters.

2 Model and estimation methodology

Before describing the estimation method proposed here, it is useful to give a brief background on the Heffernan and Tawn (HT) model and previous estimation approaches (section 2.1). Further, we give an intermediate method that assumes a known extremal dependence structure and only accounts for the uncertainty in the dependence parameters of the model in section 2.2, which we use for comparison with our main approach, which is described subsequently in section 2.3.

2.1. The conditional extreme value model

The model introduced by Heffernan & Tawn (2004, henceforth, HT) is perhaps most easily summarized by following Heffernan & Resnick (2007). First, let $\mathbf{X}' = X'_1, \ldots, X'_n$ and $\mathbf{Y}' = Y'_1, \ldots, Y'_n$ represent two series of independent and identically distributed (henceforth, i.i.d.) random variables that are possibly dependent on each other. Let $\mathbf{X} = X_1 = h_X(X'_1), \ldots, X_n = h_X(X'_n)$ and $\mathbf{Y} = Y_1 = h_Y(Y'_1), \ldots, Y_n = h_Y(Y'_n)$ be suitably transformed random variables so that they have standardized marginal distributions. Heffernan & Tawn (2004) transformed the variables to the Gumbel scale, whereas Keef et al. (2013a) apply a Laplace transformation. Then, assume that there exist normalizing functions $a(Y)$ and $b(Y)$ such that for $y > 0$

$$\Pr\left\{Y - u > y, \frac{X - a(Y)}{b(Y)} \leqslant z | Y > u\right\} \longrightarrow \exp(-y)G(z) \text{ as } u \longrightarrow \infty \tag{1}$$

where G is a non-degenerate distribution. Through examining a wide class of copula dependence models, Heffernan & Tawn (2004) found that the forms for $a(Y)$ and $b(Y)$ fell into the simple class

$$a(Y) = \alpha Y \text{ and } b(Y) = Y^{\beta}. \tag{2}$$

For positively associated variables X and Y, $\alpha \in [0, 1]$ and $\beta \in (-\infty, 1)$. They also found a slightly more complicated form for negatively associated X and Y. Keef et al. (2013a) used a Laplace transformation to ensure Equation (2) was valid for either case, which then gives $\alpha \in [-1, 1]$. The parameters α and β are interdependent and control the dependence between the variables X and Y. With the Laplace transformation, $\alpha < 0$ implies negative dependence, and $\alpha > 0$ implies positive dependence. Weakly dependent X and Y are possible as $\alpha \to 0$, but it is also possible that

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.71

strong dependence exists even when $\alpha = 0$ (cf. Heffernan & Tawn, 2004). The parameter β measures the variability of the dependence with highly negative values indicating lower variability.

To briefly describe the originally proposed estimation procedure, let

$$Z = (X - a(Y))/b(Y). \tag{3}$$

Then, from Equations (1), (2) and (3),

$$X_{|Y>u} = \alpha Y + Y^{\beta} Z_{|Y>u} \tag{4}$$

where the subscripts emphasize the dependence on Y's being extreme. The key role in estimating the joint distribution function of X and Y, conditional on $Y > u$, for u large, is to know the parameters α and β and the distribution function G. Equation (4) motivates a means for estimation of the parameters α and β, which is an active area of research (see e.g. Keef et al., 2009a, 2009b; Lamb et al., 2010; Jonathan et al., 2013), and estimation of G is performed through resampling from the empirical distribution function of the "residual" vectors Z in the HT model after achieving reasonable estimates for α and β.

The current estimation method from the HT model is semi-parametric and involves the following steps:

(1) Estimate the marginal distribution functions for each variable separately.
(2) Transform each variable in order that they each follow a standard marginal distribution function.
(3) Estimate the parameters of the parametric model conditional on large values of the conditioning variable.
(4) Information about G (e.g. functions such as the mean and variance) can be simulated using the empirical distribution function of the estimated standardized residuals. Back transformation can be used to put these estimates onto the original scale.

Heffernan & Tawn (2004) suggested using a hybrid, semi-parametric model for step 1 of the following form that accounts for both the extreme and non-extreme values (Coles & Tawn, 1991). To simplify notation, let $X = X_i$ and $x = x_i$ be single instances of the random variable and its realization, respectively.

$$\widehat{F}_{X'}(x') = \begin{cases} 1 - (1 - \tilde{F}_{u_{X'}}(x'))(1 + \xi_{X'}(x' - u_{X'})/\sigma_{X'})_+^{-1/\xi_{X'}} & \text{for } x' > u_{X'} \\ \tilde{F}_{X'}(x') & \text{for } x' \leqslant u_{X'} \end{cases} \tag{5}$$

where $\tilde{F}_X$ is the empirical distribution function of the X_i values, $\xi_{X'}$ is the (marginal) shape parameter and $\sigma_{X'} > 0$ is the (marginal) scale parameter (the subscript emphasizing that they are the parameters associated with the distribution function for $\mathbf{X}'$) of the generalized Pareto (GP) distribution function as a model for the upper tail of the univariate extremal excesses over a high threshold [i.e. $u_{X'}$ in Equation (5)]. Similarly, for $\mathbf{Y}'$, the Laplace transformation can be used in step 2, which is given by

$$X_i = \begin{cases} \log\left\{2\widehat{F}_{X'}(x_i')\right\} & \text{for } x_i' < \widehat{F}_{X'}^{-1}(\frac{1}{2}) \\ -\log\left\{2(1 - \widehat{F}_{X'}(x_i'))\right\} & \text{for } x_i' \geqslant \widehat{F}_{X'}^{-1}(\frac{1}{2}) \end{cases} \tag{6}$$

where $\widehat{F}_{X'}$ is estimated according to Equation (5) using maximum likelihood estimation for the GP portion and empirical estimation for $\tilde{F}_{X'}$ (similar for Y_i').

(wileyonlinelibrary.com) DOI: 10.1002/sta4.71

The ISI's Journal for the Rapid Dissemination of Statistics Research

Heffernan & Tawn (2004) used the non-linear least squares estimation in Equation (4) to estimate α and β for each X_i under the working assumption that Z follows a normal distribution function. Obviously, the assumption of a normal distribution function for Z is inappropriate as it implies that $X_{|Y>u}$ is also normally distributed, which generally is not the case. To counteract the inherent estimation bias from this approach, Keef et al. (2013a) imposed a joint constraint on the dependence parameters (α, β) in order to limit the upper quantiles of $X_{|Y>u}$ to be less than or equal to $x_{F_{|Y>u}}$, the value that would be observed under asymptotic dependence. From the estimates in step 3, Heffernan & Tawn (2004) obtain new estimates $\widehat{z}_i = (x_{i|y>u} - \widehat{a}_i(y_i))/\widehat{b}_i(y_i)$ from which simulations from $\widehat{G}$ are obtained.

Keef et al. (2013b) propose an alternative means for estimating α that does not involve Z. First, different q-th quantiles of $X|Y$ are estimated empirically from Equation (4) for different Y values that fall within two different intervals as $Y_i \in (u - \delta_u, u + s_u)$ and $Y_i \in (\nu - \delta_\nu, \nu + s_\nu)$, where the intervals are around the conditional threshold u, and $\nu > u$ within the range of the observation of Y_i. They then take the median of the estimated α values as $\hat{\alpha}$. See Keef et al. (2013b) for a more thorough description of the approach. This fast estimation procedure is used in our main method described in section 2.3 in order to garner prior information on α and, subsequently, β.

To incorporate the uncertainty inference at each stage of the estimation procedure, a bootstrap procedure is proposed by Heffernan & Tawn (2004). Although bootstrapping is a reasonable method for inference, it can be computationally expensive for some larger datasets. The strategy proposed here obviates the need to do any further simulations, as the uncertainty information can be directly obtained from the simulated posterior distribution, as well as the prior distribution and likelihood values.

Finally, it is important to note that the conditional HT approach differs from that of incorporating covariates into the parameters of a univariate extreme value distribution function in that a distribution for values of one variate is conditional on only the extreme values of another variable. Therefore, the dependence is on the processes themselves rather than indirectly through distributional parameters (cf. Gilleland et al., 2013; Jonathan et al., 2012).

2.2. Bayesian estimation under a known dependence structure

Following notation and, especially, Equation (1) shown earlier, the aim is to find

$$[\mathbf{Z}, Y \mid Y > u] = [\mathbf{Z}][Y|Y > u] \tag{7}$$

where we use the bracket $[\cdot]$ notation to denote a general distribution function.

Assuming that $[Y'|Y' > u_{Y'}]$, which can be derived through back transformation of Y, is GP leaves only the first term in Equation (7) to be determined. Following Heffernan & Tawn (2004), $Z \sim G_Z$ for some distribution function G_Z, which in the known dependence structure case has a (calculable) density $g_Z(z_1, \ldots, z_k)dz_1 \cdots dz_k$. In the examples that follow, we will assume $Z \sim N(\cdot, \cdot)$, but other choices are possible.

By Equation (4), $X = \alpha Y + Y^\beta Z$ for large values of Y. Applying the rules of transformation, the density of X is

$$g_X(x) = \left[\frac{1}{Y^\beta}\right] g_Z\left(\frac{x_1 - \alpha Y}{Y^\beta}\right) dx. \tag{8}$$

From Equation (6), recall that $\mathbf{X}$ represents a transformed variable and $\mathbf{X}'$ represents the variable on the original scale. We have that

$$\begin{aligned} X &= F^{-1}\left(F_{X'}(x')\right) \\ &= h_X(x') \end{aligned} \tag{9}$$

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.71

where $F_{X'}$ is the distribution function for X' with density $f_{X'}$, F is the common (e.g. Laplace or Gumbel) distribution function with density f and h_X is the associated transformation function. Furthermore, the conditioning variable Y is also a transformed variable $Y = F^{-1}(F_{Y'}(y')) = h_Y(y')$. Substituting these into Equation (8) and applying differentiation rules, we arrive at the likelihood for observations X' conditional on $Y' > u_{Y'}$,

$$g_{X'}(x'|Y' = y' > u_{Y'}) = g_Z\left(\frac{h_X(x') - \alpha h_Y(y')}{h_Y^{\beta}(y')}\right)\left[\frac{1}{[F^{-1}(F_{Y'}(y'))]^{\beta}} \times \frac{f_{X'}(x')}{f\left(F^{-1}\left(F_{X'}(x')\right)\right)}\right]dx', \tag{10}$$

which can be evaluated in closed form.

The main drawback of this approach is that G_Z is known and has a calculable density g_Z. As stated earlier, we will assume that G_Z is the Gaussian distribution but, admittedly, this is an incorrect assumption when considering extremes. However, because Heffernan & Tawn (2004) use the Gaussian distribution to obtain initial estimate of α and β, we feel that this is a close alternative for comparison. The proposed empirical Bayes estimation procedure (section 2.3) will circumvent the need to assume a closed form for G_Z.

2.3. Empirical Bayes estimation

It is desired to obtain estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{G}$. However, because G does not have a simple, closed-form expression and we only need to simulate from G, only the estimates $\hat{\alpha}$ and $\hat{\beta}$ are required. That is, $p(\alpha, \beta) = [\alpha, \beta | X, Y > u_Y]$ is sought. However, it is important to glean uncertainty information from every component of the model, including the marginal distributions of the untransformed variables X' and Y'. To do so, we consider X and Z to be two distinct, if they are highly dependent, random variables, despite that, in the HT model, one is completely determined by the other. That is, we modify Equation (4) to be (dropping the $|Y > u_Y$ notation)

$$X = \alpha Y + Y^{\beta} Z + \varepsilon, \tag{11}$$

where ε is i.i.d. noise. In fact, because $\hat{G}$ is desired to be found, $\mathbf{Z}$ could be considered as additional unknown parameters to be sought. However, such a scheme would require estimation of many parameters, which would be highly inefficient and unnecessary. Instead, we look on ε as a nuisance term, giving it zero or unity as its prior depending on whether resulting values for X are within its range or not.

Employing the previous modification and from Equation (1), and for simplicity of notation, allowing $\vec{\theta}$ to represent all of the marginal distribution parameters for X' and Y', and $\vec{\eta}$ to represent any hyperparameters, we have that

$$\begin{aligned}\left[\alpha, \beta, \vec{\theta} \mid \mathbf{Z}, \mathbf{X}, \mathbf{Y} > u_Y\right] &\propto \left[\mathbf{Z}, \mathbf{X}, \mathbf{Y} \mid \alpha, \beta, \vec{\theta}, \mathbf{Y} > u_Y\right]\left[\alpha, \beta, \vec{\theta} \mid \vec{\eta}\right][\vec{\eta}] \\ &= \left[\mathbf{X}, \mathbf{Y} \mid \mathbf{Z}, \alpha, \beta, \vec{\theta}, \mathbf{Y} > u_Y\right][\mathbf{Z} \mid \alpha, \beta, \mathbf{Y} > u_Y] x \left[\alpha, \beta, \vec{\theta} \mid \vec{\eta}\right][\vec{\eta}] \\ &= \left[\mathbf{X} \mid \mathbf{Y}, \mathbf{Z}, \alpha, \beta, \vec{\theta}\right]\left[\mathbf{Y} \mid \mathbf{Z}, \alpha, \beta, \vec{\theta}, \mathbf{Y} > u_Y\right] x [\mathbf{Z} \mid \alpha, \beta, \mathbf{Y} > u_Y]\left[\alpha, \beta, \vec{\theta} \mid \vec{\eta}\right][\vec{\eta}],\end{aligned} \tag{12}$$

where the last line is arrived at by assuming conditional independence and by Equation (1).

The first two terms $[\mathbf{X} \mid \ldots]$ and $[\mathbf{Y} \mid \ldots]$ are simply the marginal distributions defined by Equation (5) and the GP distribution function (recalling that a back transformation involving the conditioning arguments is first required). For simplicity, here, α, β and $\vec{\theta}$ are taken to be independent, and $\vec{\eta}$ is taken to be unity as long as the values fall within their support.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.71

The ISI's Journal for the Rapid Dissemination of Statistics Research

In theory, the prior knowledge on parameters does not depend on the observations Y and should therefore be specified without using observations but rather by using any external source of knowledge (Renard et al., 2013). Thus, we suppose that prior information of parameters should not produce much effect on the simulation results. However, we found that in this specific problem, prior information does matter. Therefore, relatively informative priors are preferred here.

To obtain prior information on α, we employ the fast estimation method introduced by Keef et al. (2013b) and briefly described in section 2.1. An initial estimate for β is achieved through a linear regression on the log-transformed Equation (3); namely,

$$\hat{\beta}\log(Y) + \log(Z) = \log(X - \hat{\alpha}(Y)) \text{ given } Y > u \tag{13}$$

with Z some random residual, the informative initial of $\hat{\beta}$ is obtained and, subsequently, we have some knowledge about the prior distributions for α and β.

All that remains is the primary term of interest, $G = [\mathbf{Z} \mid \alpha, \beta, \mathbf{Y} > u_Y]$. Because G has no simple closed-form expression in general, we employ empirical Bayesian estimation for this term. Empirical likelihood provides likelihood ratio statistics for parameters by profiling a non-parametric likelihood; the approach is analogous to that used for parametric models (Qin & Lawless, 1994). Owen (1990) showed that for d-variate i.i.d. random vectors $\mathbf{Y}$ (each variate as $y_1, ..., y_n$), with an unknown distribution density f, mean μ_θ and variance σ_θ^2, the approach applies to quite general parameters $\theta(f)$, where θ is the parameter associated with f. Rather than defining the likelihood from the density f as usual, the empirical likelihood method starts by defining the parameters of interest θ as functionals of f, for instance, as moments of f, and then profiles a non-parametric likelihood (Mengersen et al., 2013). More precisely, given a set of constraints of the form

$$E_f[\,h(\mathbf{Y}, \theta)] = 0 \tag{14}$$

where the dimension of h sets the number of constraints unequivocally defining θ, the empirical likelihood is defined as

$$L(\theta|\mathbf{y}) = \max_p \prod_{i=1}^{n} p_i \tag{15}$$

for p in the set

$$p \in [\,0, 1]^n\,, \sum p_i = 1, \sum{}_i p_i h(y_i, \theta) = 0 \tag{16}$$

where $p_1, \ldots, p_n$ are non-negative real numbers summing to unity. The validation of the empirical likelihood approximation is also provided by Owen (1988, 1990). He has proven that, under mild conditions, if θ satisfies Equation (14), then $-2\log(\frac{L(\theta|\mathbf{y})}{n^{-n}}) \to \chi_d^2$ in distribution when $n \to \infty$ and notes that n^{-n} is the maximum of $L(\theta|\mathbf{y})$.

In general, the basic idea in this approach is to maximize the empirical likelihood [see Equation (15)] subject to constraints provided by Equation (14), which reflect the characteristics of the quantity of interest. For instance, in the one-dimensional case when $\theta = E_f[\mathbf{Y}]$, the empirical likelihood in θ is the maximum of the product $(p_1, \ldots, p_n)$ under the constraint $p_1 y_1 + \cdots + p_n y_n = \theta$. Solving Equation (15) is based on the Newton–Lagrange algorithm, and more are derived with details in the study of Mengersen et al. (2013), Qin & Lawless (1994) and Owen (1990, 1988). Because of its ability to conduct a non-parametric inference without knowledge of higher order moments of the distribution while implicitly taking them into consideration (according to Chen & Cui, 2003), when applying to the conditional likelihood estimation in this study, the first,

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.71

$$E[\mathbf{Y}-\mu_\theta]=0, \quad (17)$$

and second,

$$E[(\mathbf{Y}-\mu_\theta)^2-\sigma_\theta^2]=0, \quad (18)$$

statistical moments of "residual" vectors Z [in Equation (3)] and conditional vectors of $X|Y$ in the HT model are used as sufficient constraints to estimate the empirical likelihood of G. And if Z has the mean vector μ and vector of standard deviation σ, the respective conditional mean and standard deviation vectors of $X|Y$, for $Y>u$, are $\alpha Y+\mu Y^\beta$ and σY^β, respectively (see Keef et al., 2013b). Now we have all the components in the Bayesian framework [see Equation (12)] to derive the empirical Bayes estimation.

To estimate the parameters using Bayesian inference, a large number of realizations are generated from the parameters' posterior distributions, using the differential evolution Markov chain (DE-MC; ter Braak, 2006). The DE-MC utilizes the genetic algorithm differential evolution (Storn & Price, 1997) for global optimization over real parameter space with Markov chain Monte Carlo (MCMC) approach (ter Braak, 2006; Gilks et al., 1996). The advantages of simplicity, speed of calculation and convergence make DE-MC favorable over the conventional MCMC (ter Braak, 2006). In this model, for example, five Markov Chains are constructed in parallel and are allowed to learn from each other by generating candidate draws based on two random parent Markov chains such that the equilibrium distribution is the target posterior distribution (see ter Braak, 2006; Gelman & Shirley, 2011). Meanwhile, the uncertainty of each parameter in Equation (12) is estimated. This approach has been used in extreme value analysis (Cheng & AghaKouchak, 2014; Renard et al., 2013).

3 Simulation experiment

In this study, the results of the conditional extreme value analysis (henceforth, conditional EVA) simulated by the proposed empirical Bayes estimation approach are compared with the originally proposed estimation strategy using the R (R Core Team, 2013) package `texmex` developed by Southworth & Heffernan (2010), which allows for the constrained estimation of the dependence parameters with the Laplace transformation on the marginal variables following Keef et al. (2013a).

3.1. Background

Ledford (1996) identified four classes of extremal dependence. The first class is that of asymptotically dependent distributions. The other three classes comprise distributions with asymptotically independent dependence structures exhibiting positive extremal dependence, near extremal independence and negative extremal dependence for a d-dimensional variable $\mathbf{Y}$. These three classes correspond respectively to joint extremes of $\mathbf{Y}$ occurring more often than, approximately as often as or less often than joint extremes if all components of the variable were independent (see Ledford & Tawn, 1996, for more details). In this study, the focus is to clarify the performance of interpreting dependence structure [see Equation (2)] by the empirical Bayes estimation approach for the different types of dependence derived in detail in Section 8 of Heffernan & Tawn (2004). We choose the following extremal dependent types to be simulated, which are also described by Keef et al. (2013a).

(1) Independence. Here, $\alpha=\beta=0$ and G factorizes into Laplace distribution functions.
(2) Asymptotic dependence. Here, $\alpha=\pm 1$ (with $+1$ indicating positive dependence and -1 implying negative dependence) and $\beta=0$, and G takes a range of forms.
(3) Asymptotic independence. Variable X is asymptotically independent of variable Y if $-1<\alpha<1$.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.71

The ISI's Journal for the Rapid Dissemination of Statistics Research

Table I. Dependence models for bivariate extreme value distributions used in this study.

Dependence models	Negative logistic (nlog)	Logistic (log)
Formula	$V(x,y) = \frac{1}{x} + \frac{1}{y} - (x^{\gamma} + y^{\gamma})^{-\frac{1}{\gamma}}$	$V(x,y) = (x^{-1/\gamma} + y^{-1/\gamma})^{\gamma}$
Pickands' dependence	$A(\omega) = 1 - [(1-\omega)^{-\gamma} + (\omega)^{-\gamma}]^{-\frac{1}{\gamma}}$	$A : [0,1] \to [0,1]; \omega \mapsto [(1-\omega)^{\frac{1}{\gamma}} + (\omega)^{\frac{1}{\gamma}}]^{\gamma}$
Independence	$\gamma \to 0$	$\gamma = 1$
Total dependence	$\gamma \to +\infty$	$\gamma \to 0$

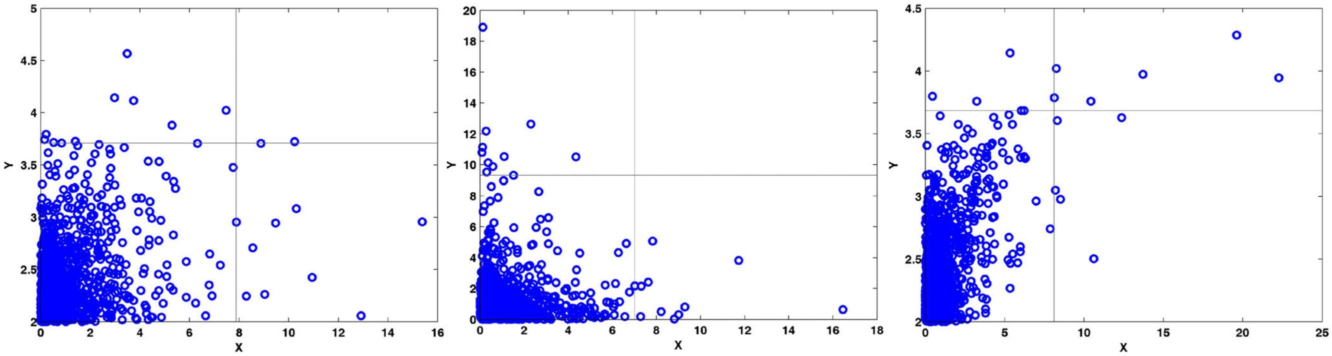

Figure 1. Scatter plot of randomly generated data of sample size 1000. The marginal threshold level, corresponding to the 0.99 quantile, is shown in grey lines. Data are shown on original scales. Asymptotic independence (left), independence (middle) and asymptotic dependence (right).

The simulated data for analyzing these three types are randomly generated from bivariate extreme value distributions using the R (R Core Team, 2013) package `POT` (Ribatet, 2006). Mainly two types of models are used as listed in Table I. The experiment is particularly designed to see the performance in estimating dependence parameters α and β with different exceedence sizes based on the proposed empirical Bayesian approach. Initially, sample sizes of 1000, 3000, 5000 and 10 000 are generated, and the high threshold of 99% quantile is selected for all cases to keep independence between Z and Y; thus, the exceedence sizes of 10, 30, 50 and 100 are left to be experimented on (see an example in Figure 1 and Laplace-transformed results in Figure 2). All simulation cases are repeated for 800 trials, to see the percentage of times (over 800 trials) that the true parameter(s) fell within the estimated 95% credible interval (henceforth, 95% CI) and to explore how the sample sizes (i.e. 10, 30, 50 and 100) might affect estimation inference. Ideally, the percentage of fall-in times for the true parameter(s) should be approximately 0.95 when considering a 95% CI. Another aspect of this experiment also compares simulations given vague priors (uniform distributions with wide support and random initials) for dependence parameters with relative informative ones to check the consequential effect from initials and priors.

3.2. Simulation results

The selected three different forms of dependence structures described in section 3.1 (Figures 1 and 2) are tested with the proposed empirical Bayes estimation approach. For the simulation experiment, with exceedence sizes of 10, 30, 50 and 100, the percentage of times that the true parameter(s) fell within the estimated 95% CI is shown for parameters α and β individually, as well as for when both parameters fell within the bounds simultaneously in Tables II and III. Table II summarizes the simulation results associated with *vague priors* for dependence parameters,

Stat

The ISI's Journal for the Rapid
Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.71

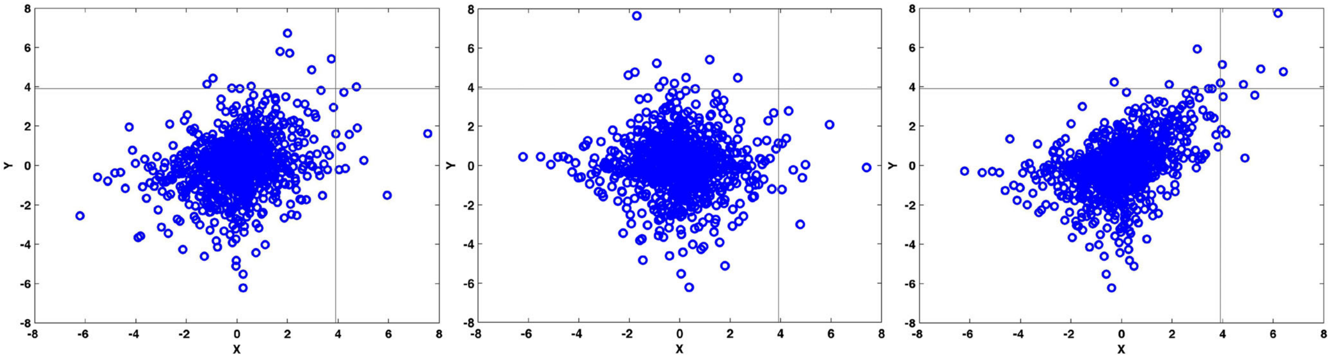

Figure 2. Scatter plot of randomly generated data of sample size 1000. The marginal threshold level, corresponding to the 0.99 quantile, is shown in grey lines. Data are shown on Laplace transformed scales. Asymptotic independence (left), independence (middle) and asymptotic dependence (right).

Table II. Results from fitting the conditional extreme value model to simulated data using the empirical Bayes estimation approach with *vague priors* proposed here. For each (exceedence) sample size, the percentage (of 800 trials) given is the percentage of times that the *true* parameter(s) fell within the estimated 95% CIs. Results are shown for parameters individually, as well as for when both parameters fell within the bounds simultaneously.

	Exceedence size	10	30	50	100
Asymptotic independence	α	0.796	0.666	0.745	0.701
	β	0.821	0.883	0.938	0.874
	Both	0.661	0.588	0.713	0.630
Independence	α	0.800	0.639	0.773	0.695
	β	0.499	0.409	0.396	0.349
	Both	0.433	0.301	0.336	0.296
Asymptotic dependence	α	0.674	0.518	0.585	0.278
	β	0.685	0.674	0.633	0.484
	Both	0.439	0.331	0.345	0.135

while Table III displays the results having used *informative priors*. In both tables, it is clear that the results improve, if only slightly, with increasing sample sizes, except for the asymptotic dependence case, where the estimation performs relatively poorly for the parameter α. This inefficiency may be caused by the fact that it is a special case of estimating a single point in a continuous parameter space (at least for one type of exact dependence).

By comparing Tables II and III, with *informative priors*, the performance of hit percentage [the true parameter(s) fell within the estimated 95% CI in over 800 trials] for either α or β or both simultaneously is much better. For example, in the independence case, with exceedences of 50, Table III shows that the individual parameter as well as both simultaneously has the hit percentage over 0.9, while in Table II, the inference performance is 0.5 on average. That is, over 800 trials, the true parameter(s) fell within the estimated 95% CI with *informative priors* over 720 times, while it fell in the interval approximately only 400 times, on average, when using *vague priors*. In some cases, such as in the asymptotic independence case with exceedences over 10

Stat

(wileyonlinelibrary.com) DOI: 10.1002/sta4.71

The ISI's Journal for the Rapid
Dissemination of Statistics Research

Table III. Results from fitting the conditional extreme value model to simulated data using the empirical Bayes estimation approach with *informative priors* proposed here. For each (exceedence) sample size, the percentage (of 800 trials) given is the percentage of times that the *true* parameter(s) fell within the estimated 95% CIs. Results are shown for parameters individually, as well as for when both parameters fell within the bounds simultaneously.

	Exceedence size	10	30	50	100
Asymptotic independence	α	0.941	0.708	0.946	0.923
	β	0.941	0.991	0.960	0.993
	Both	0.890	0.704	0.914	0.915
Independence	α	0.965	0.975	0.961	0.940
	β	0.781	0.786	0.934	0.958
	Both	0.764	0.768	0.901	0.905
Asymptotic dependence	α	0.844	0.585	0.741	0.995
	β	0.956	0.929	0.980	0.636
	Both	0.823	0.560	0.730	0.634

Table IV. Results from fitting the conditional extreme value model to simulated data using the Bayesian approach with known dependence structure and *informative priors*. For each (exceedence) sample size, the percentage (of 800 trials) given is the percentage of times that the *true* parameter(s) fell within the estimated 95% CIs. Results are shown for parameters individually, as well as for when both parameters fell within the bounds simultaneously.

	Exceedence size	10	30	50	100
Asymptotic independence	α	0.920	0.914	0.923	0.923
	β	0.845	0.871	0.911	0.943
	Both	0.801	0.832	0.843	0.877
Independence	α	0.932	0.912	0.926	0.953
	β	0.784	0.793	0.842	0.878
	Both	0.729	0.723	0.802	0.882
Asymptotic dependence	α	0.675	0.698	0.712	0.734
	β	0.687	0.656	0.766	0.865
	Both	0.598	0.668	0.700	0.699

(e.g. if exceedence size is 50 and 100, then the percentage for both is around 0.914 and 0.915) or in the independence simulation with larger sample size (e.g. if there are the same exceedence sizes, the percentage is about 0.961 for α and is approximately 0.958 for β), the percentage even reaches over the ideal situation, which is around 0.95. The identifiability issue with β may be because the model at some point is actually $X = \alpha Y + Y^{\beta}(Z - \nu_Z)/\zeta_Z$, where ν_Z and ζ_Z are the mean and standard deviation vectors of Z, respectively, and as noticed, it is not possible to differentiate β from ν_Z and ζ_Z.

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.71

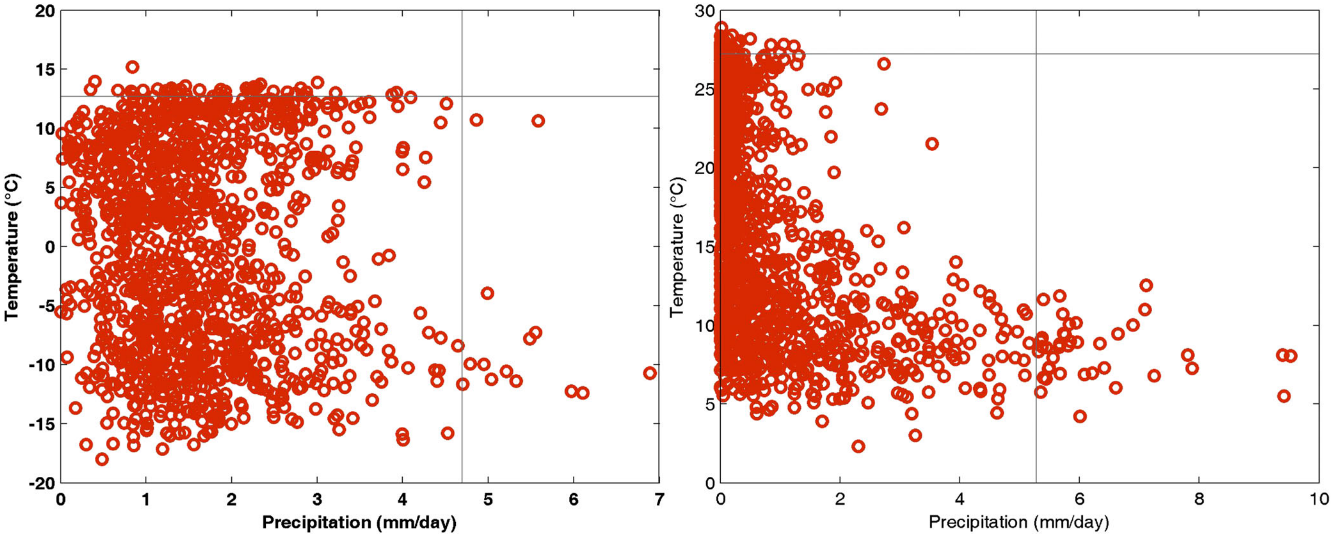

Figure 3. Scatter plot of precipitation and temperature data at Boulder, Colorado, USA (left), and Los Angeles, California, USA (right). The marginal threshold level for temperature, corresponding to the 0.99 quantile at both locations; quantile of 0.99 at Boulder and quantile of 0.97 at Los Angeles for the precipitation are shown in grey lines.

To compare the empirical Bayes approach to an approach with known dependence structure, Table IV shows the results using informative prior distributions. Comparing Tables III and IV, the known dependence structure approach is comparable for the asymptotic independence and independence cases. Clearly, however, for the asymptotic dependence case, assuming a known dependence structure is sub-optimal.

We feel that it is reasonable to obtain prior information for β so that this issue is not a major concern. Overall, simulation results show that the proposed approach performs fairly well for most types of dependence structures, but strong and reasonable prior information for β is generally necessary. And the benefit from increasing samples is not so obvious as using *informative priors*.

4 Temperature and precipitation test case

4.1. Background

In this study, the proposed empirical Bayes estimation approach is further applied on two real data cases, and results are compared with the (slightly modified) HT estimation strategy as implemented by the R (R Core Team, 2013) package `texmex` (Southworth & Heffernan, 2010). Monthly observations of precipitation and temperature from the Climatic Research Unit (CRU; New et al., 2000; Mitchell & Jones, 2005) regridded in a common 2×2-degree spatial resolution, are used to provide the test case data. Historical monthly precipitation and temperature records are available from 1901 to 2009. CRU observations have been validated and used in numerous studies of historical climate variability (e.g. Tanarhte et al., 2012; Hao et al., 2013).

To identify extreme conditions, two grid points in the central (latitude 40.02° N, longitude 105.27° W) and western (latitude 34.05° N, longitude 118.24° W) USA are selected for conditional extreme dependence structure analysis in Figure 3. The two locations are close to urban areas in Boulder, Colorado, and Los Angeles, California, where long-term observation stations have been available. In both cases, we consider the precipitation conditional on the temperatures

(wileyonlinelibrary.com) DOI: 10.1002/sta4.71

The ISI's Journal for the Rapid Dissemination of Statistics Research

Table V. Comparison results from fitting the conditional extreme value model to real data using the empirical Bayes estimation approach and `texmex`.

	Exceedence	α	β	σ_1	ξ_1	σ_2	ξ_2
Los Angeles							
Empirical Bayes	40	–0.065	–0.559	0.733	–0.367	0.820	0.193
	Std. dev.	0.070	0.214	0.028	0.043	0.044	0.046
`texmex`	40	–0.071	–0.568	0.688	–0.338	0.849	0.140
Boulder							
Empirical Bayes	39	–0.230	–0.270	0.683	–0.143	0.890	–0.251
	Std. dev.	0.074	0.334	0.033	0.027	0.012	0.060
`texmex`	39	–0.249	–0.284	0.648	–0.149	0.780	–0.143

being extreme, that is, precip | temp $> u$, for u large. The marginal threshold level (i.e. u) of temperature at the two locations is corresponding to the 0.97 quantile. For the upper tail of precipitation data, the threshold level is taken to be the 0.97 quantile for Los Angeles and the 0.99 quantile for Boulder. The proposed empirical Bayes approach is applied to infer the dependence structure parameters, α describing the dependence strength between precipitation and extremal temperature and β outlining the dependence variability, along with the scale and shape parameters for the upper tail of precipitation and temperature distributions.

4.2. Test case results

Using the CRU precipitation and temperature monthly data (see Figure 3), analyzing precip | temp $> u$, for u large, the dependence structure controlling parameters, α and β, and distribution parameters of σ and ξ fitted with the GP distribution function for each variable at two locations (Boulder and Los Angeles) are presented in Table V. The parameters derived by the empirical Bayes estimation approach and the HT model are simulated with *informative priors* for the α and β. In both locations, the dependence relationship of precipitation given extreme temperature tends to show asymptotic independence, and in Los Angeles, it is near extremal independence, while in Boulder, it is towards negative extremal dependence. Looking at the parameter β, it appears that the variability of the dependence is relatively lower in Los Angeles than in Boulder. Parameters σ_1 and ξ_1 describe the scale and shape of temperature data, which indicate a bounded upper tail distribution, while σ_2 and ξ_2 stand for the precipitation distribution. In Los Angeles, the precipitation distribution shows a heavy tail property (indicated by $\xi_2 > 0$), while in Boulder, it shows a bounded upper tail (see $\xi_2 < 0$). Table V also compares the results using the empirical Bayes estimation approach and the R (R Core Team, 2013) package `texmex` for the HT method. In general, from the table, we can see that all the parameters, including dependence structure parameters α and β, and GP distribution function parameters σ and ξ, inferred by the two approaches, are consistent with each other.

5 Summary, conclusions and discussion

The conditional EVA approach introduced by Heffernan & Tawn (2004) is an important new methodology for modelling multivariate extreme values through a conditional distribution framework. Although this approach does not require a priori knowledge of the dependence structure nor that the variables be simultaneously extreme, a difficulty for estimating the parameters is that no simple, closed-form distribution exists in general for G. Therefore, in the original

(wileyonlinelibrary.com) DOI: 10.1002/sta4.71

approach, several estimation methods and constraints are mixed together to evaluate the G distribution function and counteract the inherent bias. This disadvantage motivates the development of the empirical Bayes estimation approach proposed in this study.

Simulations are employed to reproduce known dependence structures with 800 repeated trials for each of three types of dependence and sample sizes and to test how well the estimation procedure performs. Simulations show generally good coverage of credible intervals. However, the parameter β is relatively hard to infer precisely, and sometimes, it is not unique, so strong prior information for β is generally necessary, which is the primary holdback in this approach and might require further refinement.

The identifiability issue with β may result from ignoring the mean (ν_Z) and standard deviation (ζ_Z) vectors of Z, which are difficult to differentiate from β. To possibly solve this issue, one may include ν_Z and ζ_Z as parameters of interest in the empirical likelihood estimation by imposing other prior knowledge for those parameters, but still, the identifiability problem might not disappear. Another possibility is that one may explore and include higher order moments (other than first and second moments in this study) of the conditional distribution as empirical likelihood constraints, which may also be difficult to identify without any additional assumption because the conditional distribution is generally, numerically unknown.

As for the inefficiency resulting from trying to estimate a single point in a continuous parameter space, a possible extension for the estimation model would be to use a reversible jump Markov chain to include probability terms for those particular cases, namely $\alpha = 0, \beta = 0, \alpha = \pm 1$ and $\beta = 1$, analogously as proposed in the univariate setting by Stephenson & Tawn (2004) for the shape parameter, $\xi = 0$ versus $\xi \neq 0$. Such a scheme might help with the estimation in the asymptotic dependence case and will be explored in a future study.

Additionally, precipitation data conditional on having extreme temperature is also analysed and compared with the (slightly modified) estimation strategy proposed by Heffernan & Tawn (2004). Resulting parameter estimates are found to be consistent with those from the originally proposed estimation strategy from Heffernan & Tawn (2004).

The presented model can be potentially applied in a wide variety of science fields including finance, earth science, environmental science and biology. Particularly, this model can be used for assessing spatial climatic extremes (e.g. Gilleland et al., 2013). A myriad of papers shows that climatic extremes have been changing and are projected to change in the future (e.g. Wehner, 2013; AghaKouchak et al., 2013; Field et al., 2012; Schubert & Lim, 2013; Easterling et al., 2000; Alexander et al., 2006; Cheng et al., 2014). Even concurrent extremes (e.g. joint precipitation and temperature extremes) have been reported to have increased/changed over time (Hao et al., 2013). The proposed methodology allows assessing one extreme variable conditioned on another, and hence, we expect it to be a useful tool for conditional extreme value analysis.

Acknowledgements

The first author acknowledges the financial support from the National Center for Atmospheric Research Advanced Study Program Graduate Student Visitor Program. Support for the second author was provided by the Weather and Climate Impacts Assessment Program (WCIAS, http://www.assessment.ucar.edu/) and by the National Science Foundation through Earth System Modeling grant number AGS-1243030. The National Center for Atmospheric Research and the WCIAS Program is sponsored by the National Science Foundation. The first and fourth authors acknowledge the financial support from the Environmental Sciences Division of the Army Research Office (Award No. W911NF-14-1-0684) and the National Science Foundation (Award No. DMS 1331611). The authors would also like to thank Holger Rootzen for giving valuable feedback on an earlier version of this manuscript.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.71

References

AghaKouchak, A, Easterling, D, Hsu, K, Schubert, S & Sorooshian, S (2013), *Extremes in a Changing Climate*, *Springer*, Netherlands, Dordrecht.

Alexander, L, Zhang, X, Peterson, T, Caesar, J, Gleason, B, Klein Tank, A, Haylock, M, Collins, D, Trewin, B, Rahimzadeh, F, Tagipour, A, Ambenje, P, Rupa Kumar, K, Revadekar, J & Griffiths, G (2006), 'Global observed changes in daily climate extremes of temperature', *Journal of Geophysical Research*, **111**, D05109.

Chen, SX (1994), 'Comparing empirical likelihood and bootstrap hypothesis tests', *Journal of Multivariate Analysis*, **51**(2), 277–293.

Chen, SX & Cui, H (2003), 'An extended empirical likelihood for generalized linear models', *Statistica Sinica*, **13**(1), 69–82.

Cheng, L & AghaKouchak, A (2014), 'Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate', *Scientific Reports*, **4**, 7093.

Cheng, L, AghaKouchak, A, Gilleland, E & Katz, R (2014), 'Non-stationary extreme value analysis in a changing climate', *Climatic Change*, **127**(2), 353–369. DOI: 10.1038/srep07093.

Coles, SG & Tawn, JA (1991), 'Modelling extreme multivariate events', *Journal of the Royal Statistical Society: Series B (Statistical Methodological)*, **53**, 377–392.

Easterling, D, Meehl, G, Parmesan, C, Changnon, S, Karl, T & Mearns, L (2000), 'Climate extremes: observations, modeling, and impacts', *Science*, **289**(5487), 2068–2074. DOI: 10.1126/science.289.5487.2068.

Field, CB, Barros, V, Stocker, TF, Qin, D, Dokken, DJ, Ebi, KL, Mastrandrea, MD, Mach, KJ, Plattner, G-K, Allen, SK, Tignor, M & Midgley, PM (eds) (2012), *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, *Cambridge University Press*, Cambridge.

Gelman, A & Shirley, K (2011), 'Inference from simulations and monitoring convergence', Handbook of Markov Chain Monte Carlo in Gelman, SA, Jones, G & Meng, X-L (eds), Chapman and Hall/CRC, USA, 163–174.

Gilks, WR, Richardson, S & Spiegelhalter, DJ (1996), 'Introducing Markov chain Monte Carlo', Markov chain Monte Carlo in practice in Gilks, WR, Richardson, S & Spiegelharter, DJ (eds), Chapman and Hall, New York, 1–19.

Gilleland, E & Brown, BG (2013), 'Spatial extreme value analysis to project extremes of large-scale indicators for severe weather', *Environmetrics*, **24**(6), 418–432.

Hall, P & La Scala, B (1990), 'Methodology and algorithms of empirical likelihood', *International Statistical Review/Revue Internationale de Statistique*, **58**(2), 109–127.

Hao, Z, AghaKouchak, A & Phillips, T (2013), 'Changes in concurrent monthly precipitation and temperature extremes', *Environmental Research Letters*, **8**, 034014.

Heffernan, JE (2007), 'Limit laws for random vectors with an extreme component', *The Annals of Applied Probability*, **17**(2), 537–571. DOI: 10.1088/1748-9326/8/3/034014.

Heffernan, JE & Tawn, JA (2004), 'A conditional approach for multivariate extreme values (with discussion)', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **66**(3), 497–546.

Jing, BY (1995), 'Two-sample empirical likelihood method', *Statistics & Probability Letters*, **24**(4), 315–319.

The ISI's Journal for the Rapid
Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.71

Jonathan, P, Ewans, K & Flynn, J (2012), 'Joint modelling of vertical profiles of large ocean currents', *Ocean Engineering*, **42**, 195–204.

Jonathan, P, Ewans, K & Randell, D (2013), 'Joint modelling of extreme ocean environments incorporating covariate effects', *Coastal Engineering*, **79**, 22–31.

Keef, C, Papastathopoulos, I & Tawn, JA (2013a), 'Estimation of the conditional distribution of a multivariate variable given that one of its components is large: additional constraints for the Heffernan and Tawn model', *Journal of Multivariate Analysis*, **115**, 396–404.

Keef, C, Svensson, C & Tawn, JA (2009a), 'Spatial dependence in extreme river flows and precipitation for Great Britain', *Journal of Hydrology*, **378**(3–4), 240–252.

Keef, C, Tawn, J & Svensson, C (2009b), 'Spatial risk assessment for extreme river flows', *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **58**(5), 601–618.

Keef, C, Tawn, JA & Lamb, R (2013b), 'Estimating the probability of widespread flood events', *Environmetrics*, **24**(1), 13–21.

Lamb, R, Keef, C, Tawn, J, Laeger, S, Meadowcroft, I, Surendran, S, Dunning, P & Batstone, C (2010), 'A new method to assess the risk of local and widespread flooding on rivers and coasts', *Journal of Flood Risk Management*, **3**(4), 323–336.

Ledford, AW (1996), 'Statistics for near independence in multivariate extreme values', *Biometrika*, **83**(1), 169–187.

Li, G (1995), 'Nonparametric likelihood ratio estimation of probabilities for truncated data', *Journal of the American Statistical Association*, **90**(431), 997–1003.

Mengersen, KL, Pudlo, P & Robert, CP (2013), 'Bayesian computation via empirical likelihood', *Proceedings of the National Academy of Sciences*, **110**(4), 1321–1326.

Mitchell, T & Jones, P (2005), 'An improved method of constructing a database of monthly climate observations and associated high-resolution grids', *International Journal of Climatology*, **25**(6), 693–712.

New, M, Hulme, M & Jones, P (2000), 'Representing twentieth-century space-time climate variability. Part II: development of 1901-96 monthly grids of terrestrial surface climate', *Journal of Climate*, **13**(13), 2217–2238.

Owen, A (1990), 'Empirical likelihood ratio confidence regions', *The Annals of Statistics*, **18**(1), 90–120.

Owen, AB (1988), 'Empirical likelihood ratio confidence intervals for a single functional', *Biometrika*, **75**(2), 237–249.

Qin, J & Lawless, J (1994), 'Empirical likelihood and general estimating equations', *The Annals of Statistics*, **22**(1), 300–325.

R Core Team (2013), *R: A Language and Environment for Statistical Computing*, *R Foundation for Statistical Computing*, Vienna, Austria.

Renard, B, Sun, X & Lang, M (2013), Bayesian methods for non-stationary extreme value analysis, in *Extremes in a Changing Climate*, *Springer*. DOI: 10.1007/978-94-007-4479-0 3.

Ribatet, MA (2006), *A User's Guide to the POT Package (Version 1.0)*.

Schubert, S & Lim, YK (2013), Climate variability and weather extremes: model-simulated and historical data, in *Extremes in a Changing Climate*, *Springer*. DOI: 10.1007/978-94-007-4479-0 9.

Stat

(wileyonlinelibrary.com) DOI: 10.1002/sta4.71

The ISI's Journal for the Rapid
Dissemination of Statistics Research

Southworth, H & Heffernan, J (2010), texmex: Threshold exceedences and multivariate extremes, *R package version* **1**.

Stephenson, A & Tawn, J (2004), 'Bayesian inference for extremes: accounting for the three extremal types', *Extremes*, **7**(4), 291–307.

Storn, R & Price, K (1997), 'Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces', *Journal of Global Optimization*, **11**(4), 341–359.

Tanarhte, M, Hadjinicolaou, P & Lelieveld, J (2012), 'Intercomparison of temperature and precipitation data sets based on observations in the Mediterranean and the Middle East', *Journal of Geophysical Research*, **117**(D12), D12102.

ter Braak, C (2006), 'A Markov chain Monte Carlo version of the genetic algorithm differential evolution: easy Bayesian computing for real parameter spaces', *Statistics and Computing*, **16**(3), 239–249.

Wehner, M (2013), Methods of projecting future changes in extremes, in *Extremes in a Changing Climate*, *Springer*. doi: 10.1007/978-94-007-4479-0 8.