Spectral Density Regression for Bivariate Extremes

Castro, D., and de Carvalho, M. (2016+), "Spectral Density Regression for Bivariate Extremes,"
Stochastic Environmental Research and Risk Assessment, in press (DOI: 0.1007/s00477-016-1257-z).

Stoch Environ Res Risk Assess

Spectral density regression for bivariate extremes

Daniela Castro Camilo · Miguel de Carvalho

Received: 15 May 2015

Abstract We introduce a density regression model for the spectral density of a bivariate extreme value distribution, that allows us to assess how extremal dependence can change over a covariate. Inference is performed through a double kernel estimator, which can be seen as an extension of the Nadaraya–Watson estimator where the usual scalar responses are replaced by mean--constrained densities on the unit interval. Numerical experiments with the methods illustrate their resilience in a variety of contexts of practical interest. An extreme temperature dataset is used to illustrate our methods.

Keywords Bivariate extremes values · Nonstationary extremal dependence structures · Spectral density · Statistics of extremes

1 Introduction

Extreme values play a key role in environmental research and risk assessment (see for instance Wang et al., 2014; Fernández-Ponce and Rodríguez-Griñolo, 2015; Hainy et al., 2016). Modeling nonstationarity in marginal distributions has been the focus of much recent literature in applied extreme value modelling. The simplest approach was popularized long ago by Davison and Smith (1990), and it is based on indexing the location and scale parameters of the generalized extreme value distribution by a predictor, say by considering

$$G_{(\mu_x,\sigma_x,\xi)}(y) = \exp[-\{1+\xi(y-\mu_x)/\sigma_x\}_+^{-1/\xi}]. \quad (1)$$

Daniela Castro Camilo
KAUST, Thuwal, Saudi Arabia
E-mail: Daniela.Castro@kaust.edu.sa

Miguel de Carvalho
Pontificia Universidad Católica de Chile, Santiago, Chile
E-mail: mmbbcarvalho@gmail.com

See also Coles (2001, Ch. 6), Chavez-Demoulin and Davison (2005), Eastoe and Tawn (2009), and Chavez-Demoulin et al. (2015), for related approaches.

In areas such as environmental impact assessment or financial risk management, one is often concerned in assessing how extreme outcomes of two or more variables are related, and the mathematical basis for such modeling is that of statistics of bivariate extremes. In such contexts, extremal dependence is often interpreted as a synonym of risk, and when modeling bivariate extremes we are naturally led to the bivariate extreme value distribution. It is well known that the bivariate extreme value distribution, depends on an infinite-dimensional parameter (H) (Coles, 2001, Thm 8.1), and it can be written as

$$G_H(y_1,y_2) = \exp\left\{-2\int_0^1 \max\left(\frac{w}{y_1},\frac{1-w}{y_2}\right)\mathrm{d}H(w)\right\}, \quad (2)$$

for $y_1, y_2 > 0$, where H is the so-called spectral distribution function, which is a distribution function on the unit interval obeying the moment constraint

$$\int_0^1 w\,\mathrm{d}H(w) = \frac{1}{2}. \quad (3)$$

Roughly speaking, the more mass H puts close to $1/2$ the higher the level of extremal dependence, whereas the more mass H puts close to 0 and 1 the more independent the extremes are. Since the object of interest in bivariate extremes is intrinsically nonparametric, nonparametric methods have become a natural tool for estimation. A survey on nonparametric estimation of extremal dependence can be found in Kiriliouk et al. (2015).

And how to model 'nonstationary bivariate extremes' if one must? Surprisingly, by comparison to the marginal case, approaches to modelling nonstationarity in

the extremal dependence structure have received relatively little attention. However, in many settings of applied interest, it seems natural to regard risk from a covariate-adjusted viewpoint, allowing for extremal dependence to increase/decrease according to a covariate. But to develop ideas of covariate-adjusted risk using statistics of bivariate extremes we need to allow for nonstationary extremal dependence structures, so to assess the dynamics governing extremal dependence of pairs of variables of interest.

In this paper we discuss methods for modeling nonstationary extremal dependence structures. Our approach can be regarded as an analogue to the bivariate setting of the Davison–Smith approach in (1), and it is based on indexing the parameter of the bivariate extreme value distribution (H) with a covariate, i.e. considering $\{H_x : x \in \mathcal{X} \subset \mathbb{R}\}$, and taking

$$G_{H_x}(y_1, y_2) = \exp\left\{-2\int_0^1 \max\left(\frac{w}{y_1}, \frac{1-w}{y_2}\right) \mathrm{d}H_x(w)\right\}, \tag{4}$$

for $y_1, y_2 > 0$. Obviously, the H_x—to which we refer as predictor-dependent spectral measures—will need to obey the moment constraints (3), for every x, so that G_{H_x} is a valid bivariate extreme value distribution.

A main goal of this paper is on modeling families of spectral densities indexed by a covariate, and we refer to our approach as a spectral 'density regression.' In terms of estimation, we propose a nonparametric estimator, that has connections with the Nadaraya–Watson estimator (Nadaraya, 1964; Watson, 1964). While 'density regression,' could sound like a misnomer, we underscore that similar terminology has been used on related topics for referring to contexts where the interest is in estimating a predictor-dependent family of densities; see Dunson et al. (2007). A related approach to the one discussed here has been recently proposed by de Carvalho and Davison (2014) who introduced a model for the case where several bivariate extremal distributions are linked through the action of a covariate. A challenge with their model is however that inference entails intensive constrained optimization problems. In comparison with de Carvalho and Davison (2014) approach, our model avoids the need of specifying a tilting function, it allows for straightforward extrapolation to unobserved covariate values, it allows for estimation of covariate-adjusted spectral densities (and not only spectral measures), and it is computationally straightforward. Another related approach is that of Huser and Genton (2016) who use nonstationary max-stable dependence structures to develop nonstationary models for spatial extremes in which covariates can be incorporated.

In Sect. 2 we introduce spectral density regression, propose a method for inference and estimation, and give details on computational implementation. A simulation study is conducted in Sect. 3, while an application to extreme forest temperatures is given in Sect. 4. Section 5 offers conclusions. Online supplementary materials include additional empirical reports.

2 Spectral density regression model

2.1 Bivariate statistics of extremes and K-sample setting

Let $\{(Y_{i,1}, Y_{i,2})\}_{i=1}^N$ be a sequence of independent identically distributed random vectors with unit Fréchet marginal distributions, $F_1(y) = F_2(y) = \exp(-1/y)$, for $y > 0$. The underlying theory for modeling bivariate extremes is based in the so-called Pickands' (1981) representation theorem, a convergence result which provides the limiting distribution of the componentwise standardized maximum,

$$(M_{1,N}, M_{2,N}) = N^{-1}\left(\max_{i=1,\ldots,N}\{Y_{i,1}\}, \max_{i=1,\ldots,N}\{Y_{i,2}\}\right).$$

Pickands (1981) established that

$$P(M_{1,N} \leqslant y_1, M_{2,N} \leqslant y_2) \to G_H(y_1, y_2), \tag{5}$$

as $N \to \infty$, where $y_1, y_2 > 0$, provided the limit exists and is non-degenerate; see also Coles (2001, Thm 8.1). Here G_H is the bivariate extreme value distribution defined in (2) and is in one to one correspondence with H, the spectral distribution function that is mean-constrained according to (3). The spectral measure H provides relevant information on extremal dependence, and can be used to describe the extremal dependence structure of the random vector (Y_1, Y_2). This can be understood through a pseudo-polar transformation, where we map $(Y_{1,1}, Y_{1,2}), \ldots, (Y_{N,1}, Y_{N,2})$, to pseudo-angular and radial variates

$$(W_i, R_i) = \left(\frac{Y_{i,1}}{Y_{i,1} + Y_{i,2}}, Y_{i,1} + Y_{i,2}\right), \quad i = 1, \ldots, N.$$

de Haan and Resnick (1977) showed that W_i has measure H on $[0, 1]$ conditional on $R_i \to \infty$. If W and R are general terms of the sequence $\{(W_i, R_i)\}_{i=1}^N$, the latter result tells us that when the radius R is large, the pseudo-angle W is approximately distributed according to H, and approximately independent of R. The limiting cases of the distribution H are given by asymptotic independence, whereby all mass is placed at the boundaries of $[0, 1]$, giving $G_H(y_1, y_2) = \exp\{-(y_1^{-1} + y_2^{-1})\}$, and by complete dependence, whereby all mass is placed at the centre of the interval, yielding $G_H(y_1, y_2) =$

$\exp\{-\max\left(y_1^{-1}, y_2^{-1}\right)\}$. We refer to situations where H has mass away from the vertices as asymptotic dependence, and this will be the framework of our modeling. Throughout we assume that H is absolutely continuous with spectral density $h(w) = \mathrm{d}H(w)/\mathrm{d}w$.

Given a sample $\{(Y_{1,1}, Y_{1,2}), \ldots, (Y_{N,1}, Y_{N,2})\}$, we may construct proxies for the unobservable pseudo-angles W_i by setting

$$(W_i, R_i) = \left(\frac{\widehat{Y}_{i,1}}{\widehat{Y}_{i,1} + \widehat{Y}_{i,2}}, \widehat{Y}_{i,1} + \widehat{Y}_{i,2}\right), \quad i = 1, \ldots, N.$$

where

$$\widehat{Y}_{i,1} = -1/\{\log \widehat{F}_1(Y_{i,1})\}, \quad \widehat{Y}_{i,2} = -1/\{\log \widehat{F}_2(Y_{i,2})\},$$

and where $\widehat{F}_1$ and $\widehat{F}_2$ are estimators of the marginal distribution functions F_1 and F_2. A robust choice for $\widehat{F}_1$ and $\widehat{F}_2$ is the pair of univariate empirical distribution functions, normalized by $N + 1$ rather than by N to avoid division by zero. For a high enough threshold u, the collection of angles $\{W_i : i \in I_N\}$, where $I_N = \{i = 1, \ldots, N : R_i > u\}$, can be regarded as an approximate sample from the spectral measure H. Parametric or nonparametric inference on H may then be based upon the sample $\{W_i : i \in I_N\}$.

Similarly to de Carvalho and Davison (2011, 2014), below we work under the so-called *K-sample setting for bivariate extremes*. Indeed, our applied setting of interest in Sect. 4 is one where the raw data consists of

$$\{(Y_{1,1,k}, Y_{1,2,k}), \ldots, (Y_{N_k,1,k}, Y_{N_k,2,k})\}, \quad k = 1, \ldots, K,$$

plus a covariate x_k. Applying similar principles as discussed above, the collection of angles $\mathbf{w}_k = \{W_{i,k} : i \in I_{N_k}\}$, where $I_{N_k} = \{i = 1, \ldots, N_k : R_{i,k} > u_k\}$, can be regarded as an approximate sample of $n_k = |I_{N_k}|$ pseudo-angles from the spectral measure corresponding to the kth population, H_k, for $k = 1, \ldots, K$. Thus, in the K-sample setting for bivariate extremes, data are of the type $\{(x_k, \mathbf{w}_k)\}_{k=1}^K$. Throughout, we assume that H_k is absolutely continuous with spectral density $h_k(w) = \mathrm{d}H_k(w)/\mathrm{d}w$. The combined sample size is denoted by $n = n_1 + \cdots + n_K$.

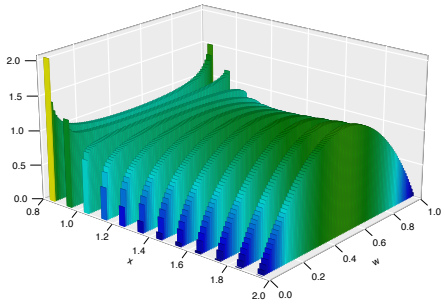
2.2 Predictor-dependent spectral measures

Formally, $\{F_x : x \in \mathcal{X}\}$ is a set of predictor-dependent (henceforth pd) probability measures if the F_x are probability measures indexed by a covariate $x \in \mathcal{X} \subseteq \mathbb{R}$. Analogously, we say that the family

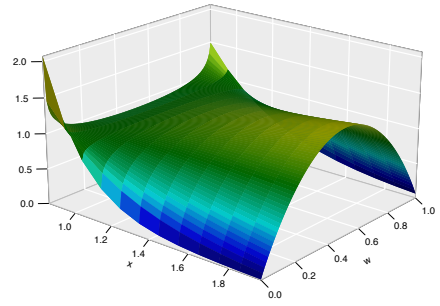
$$\{H_x : x \in \mathcal{X}\}$$

is a set of *pd spectral measures* if

$$\int_0^1 \mathrm{d}H_x(w) = 1, \quad \int_0^1 w\,\mathrm{d}H_x(w) = \frac{1}{2}, \quad x \in \mathcal{X}. \tag{6}$$



(a)



(b)

Figure 1. (a) Histograms for set of pseudo-angles and (b) interpolated spectral surface. Both figures were generated from $h_x(w) = \beta(w; x, x)$ with x taking values in a grid between 0.85 and 2.

Pd spectral measures allow us to assess how extremal dependence evolves over a certain covariate x, i.e., they allow us to model nonstationary extremal dependence structures; further details on pd spectral measures can be found in de Carvalho (2016, Sect. 2.3).

Suppose H_x is absolutely continuous for all $x \in \mathcal{X}$. We define the pd spectral density as $h_x(w) = \mathrm{d}H_x(w)/\mathrm{d}w$, and following de Carvalho (2016) we refer to the set

$$\{h_x(w) : w \in [0, 1], x \in \mathcal{X}\}$$

as the *spectral surface*. Spectral surfaces can be readily constructed from parametric models for the spectral density; see, for instance, Coles (2001, Sect. 8.2.1). Examples of spectral surfaces can be found in Figs. 1(b), 3, and 6.

By using pd spectral measures we are essentially indexing the parameter of the bivariate extreme value distribution (H) with a covariate, and thus the approach in (4) can be regarded as an analogue of the Davison–Smith paradigm in (1), but for the bivariate setting.

In practice we need to obtain estimates which obey the marginal moment constraint, and which define a density on the unit interval, for all $x \in \mathcal{X}$. It is challenging to construct nonparametric estimators able to yield valid pd spectral densities. Indeed, any such estimator, $\widehat{h}_x$, needs to obey the moment constraint, i.e., $\int_0^1 w\widehat{h}_x(w)\,\mathrm{d}w = 1/2$, for all $x \in \mathcal{X}$. In the next section we introduce one such estimator.

2.3 Double kernel estimator

Figure 1 resumes key ideas underlying the construction of our estimator. Figure 1(a) shows histograms for sets of pseudo-angles generated from $h_x(w) = \beta(w; x, x)$ with x taking values in a grid between 0.85 and 2. For each value of x in the grid (x_k), we would like to estimate the associated spectral density, and then interpolate for unobserved values of x, as shown in Fig. 1(b).

Suppose that we have a method to compute $\widetilde{h}_k$, the spectral density estimates at every x_k; in Fig. 1(a) $\widetilde{h}_k$ would correspond to the histogram estimates, but for reasons that will become obvious below we will not work with these estimates. Estimation of the pd spectral density on the basis of data available on the K-sample setting, $\{(x_k, \mathbf{w}_k)\}_{k=1}^K$, entails two challenges:

1. Although we want to estimate h_x at every $x \in \mathcal{X}$, we only have data at $x_1, \ldots, x_K$.
2. We need to impose to $\widehat{h}_x$ and $\widetilde{h}_k$ the corresponding moment constraints.

To estimate the spectral surface, h_x, we propose the estimator

$$\widehat{h}_x(w) = \frac{\sum_{k=1}^K \mathbb{K}_b(x - x_k)\widetilde{h}_k(w)}{\sum_{k=1}^K \mathbb{K}_b(x - x_k)}, \tag{7}$$

for $w \in (0, 1)$, where $\mathbb{K}_b$ is a kernel density estimator and $b > 0$ is a bandwidth parameter controling smoothing in the x-direction. The estimator in (7) is similar to the well-known Nadaraya–Watson estimator (Nadaraya, 1964; Watson, 1964), but here—contrary to the usual nonparametric regression setting—the responses are spectral densities, and hence infinite-dimensional objects; further details on kernel regression can be found in Wand and Jones (1994, Ch. 5). If the spectral density estimates at every x_k are such that

$$\int_0^1 \widetilde{h}_k(w)\,\mathrm{d}w = 1, \quad \int_0^1 w\widetilde{h}_k(w)\,\mathrm{d}w = 1/2,$$

for $k = 1, \ldots, K$, then

$$\int_0^1 \widehat{h}_x(w)\,\mathrm{d}w = \frac{\sum_{k=1}^K \mathbb{K}_b(x - x_k)\int_0^1 \widetilde{h}_k(w)\,\mathrm{d}w}{\sum_{k=1}^K \mathbb{K}_b(x - x_k)} = 1,$$

$$\int_0^1 w\widehat{h}_x(w)\,\mathrm{d}w = \frac{\sum_{k=1}^K \mathbb{K}_b(x - x_k)\int_0^1 w\widetilde{h}_k(w)\,\mathrm{d}w}{\sum_{k=1}^K \mathbb{K}_b(x - x_k)} = 1/2,$$

for all $x \in \mathcal{X}$. Put differently, valid spectral surfaces can be obtained from our estimator in (7) if at every x_k we estimate a valid spectral density, $\widetilde{h}_k$, i.e. a density on $[0, 1]$ obeying the moment constraint; an histogram estimate $\widetilde{h}_k$ would not however obey the moment constraint, and would lead to a nonsmooth estimate. To ensure that each spectral density estimate, $\widetilde{h}_k$, obeys the normalization and marginal moment constraints, we use the smooth Euclidean likelihood estimator (de Carvalho et al., 2013), which for a sample of n_k pseudo-angles is defined as

$$\widetilde{h}_k(w) = \sum_{i=1}^{n_k} \widetilde{p}_{i,k}\beta(w; W_{i,k}\nu, (1 - W_{i,k})\nu), \tag{8}$$

for $w \in (0, 1)$, where

$$\widetilde{p}_{i,k} = \frac{1}{n_k}\{1 - (\overline{W}_k - 1/2)S_k^{-2}(W_{i,k} - \overline{W}_k)\}, \tag{9}$$

for $i = 1, \ldots, n_k$ and $k = 1, \ldots, K$. Here, β denotes the Beta density and $\overline{W}_k$ and S_k^2 denote the sample mean and sample variance of $W_{1,k}, \ldots,$
$W_{n_k,k}$, that is,

$$\overline{W}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} W_{i,k}, \quad S_k^2 = \frac{1}{n_k}\sum_{i=1}^{n_k}(W_{i,k} - \overline{W}_k)^2.$$

The parameter $\nu > 0$ in (8) is a concentration parameter, responsible for controlling the amount of smoothing, in the w-direction. A method for parameter selection using cross-validation is discussed in Sect. 2.4. The estimator in (9) can be understood as an empirical likelihood-based kernel density estimator (Chen, 1997); the weights in (9) differ from the usual $1/n_k$ appearing in kernel density estimation, as they are obtained through an empirical likelihood-based method, in order to produce estimates which obey the moment constraint. Specifically, the $\widetilde{p}_{i,k}$ in (9) are Euclidean likelihood weights (Owen, 2001, pp. 63–66), i.e., are the solution to the optimization problem

$$\begin{aligned} \max_{\mathbf{p}_k \in \mathbb{R}^{n_k}} \; & -\tfrac{1}{2}\textstyle\sum_{i=1}^{n_k}(n_k\, p_{i,k} - 1)^2 \\ \text{s.t. } & \textstyle\sum_{i=1}^{n_k} p_{i,k} = 1 \\ & \textstyle\sum_{i=1}^{n_k} W_{i,k} p_{i,k} = 1/2. \end{aligned} \tag{10}$$

Finally, our estimator in (7) can be rewritten as a double kernel estimator

$$\widehat{h}_x(w) = \frac{\sum_{k=1}^K \sum_{i=1}^{n_k} \widetilde{p}_{i,k}\mathbb{K}_b(x - x_k)\beta(w; W_{i,k}\nu, (1 - W_{i,k})\nu)}{\sum_{k=1}^K \mathbb{K}_b(x - x_k)}. \tag{11}$$

Next we provide details on practical aspects and computing.

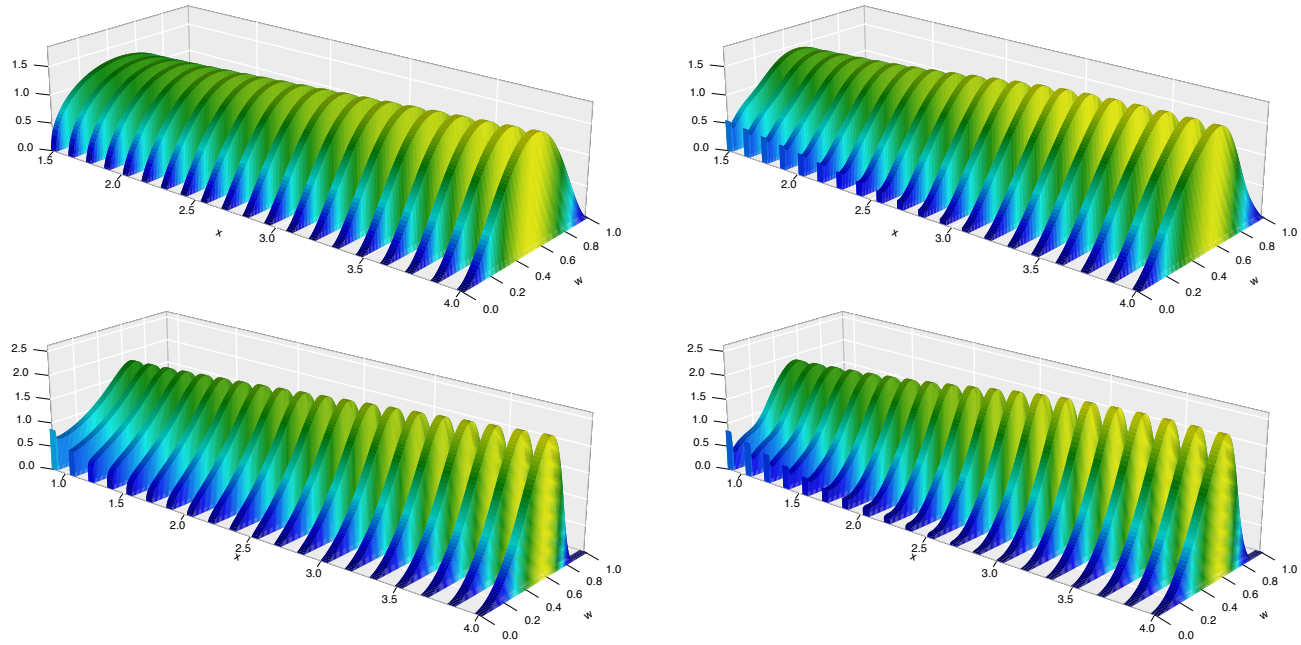


Figure 2. True (left) cross sections of the spectral surface and corresponding estimates (right) from the symmetric (top) and asymmetric (bottom) Dirichlet predictor-dependent models, for $K = 20$ values of the predictor and Configuration 1.

2.4 Details on implementation

We select the tuning parameters via leave-one-out cross-validation for each parameter separately. Specifically, for the concentration parameter ν we choose

$$\nu^\star = \arg\min_{\nu>0} \sum_{k=1}^{K}\sum_{i=1}^{n_k} -\log\{\widetilde{h}_{-i}(W_{i,k})\}, \tag{12}$$

where

$$\widetilde{h}_{-i}(w) = \sum_{j\neq i} \widetilde{p}_{j,k}\beta(w; W_{j,k}\nu, (1-W_{j,k}\nu)),$$

whilst for the bandwidth b we select

$$b^\star = \arg\min_{b>0} \int_0^1 \sum_{k=1}^{K} \{\widetilde{h}_k(w) - \widehat{h}_{-k}(w)\}^2 \mathrm{d}w, \tag{13}$$

with

$$\widehat{h}_{-k}(w) = \frac{\sum_{j\neq k} \mathbb{K}_b(x_j - x_k)\widetilde{h}_j(w)}{\sum_{j\neq k} \mathbb{K}_b(x_j - x_k)}.$$

In principle, $\mathbb{K}_b$ should be a symmetric and unimodal density. While there are many kernel functions that verify these basic requirements, it is well known that the choice of the kernel has little impact on the corresponding estimators; see Wand and Jones (1994, Ch. 2) and references therein. In practice, we use a normal kernel.

Next, we give computational details on how to implement the double kernel estimator using `ksmooth` from the `R` package `stats` (R Core Team, 2014).

Pseudocode for double kernel estimator

1. Compute $\nu^\star$ and $b^\star$ using `optim` according to (12) and (13), respectively.
2. Construct a grid $\{w_j\}_{j=1}^J \in (0,1)$ and compute $\widetilde{h}_k(w_j)$ according to (8).
3. `for` $j = 1, \ldots, J$, `do`:
   Compute $\widehat{h}_x(w_j)$ using `ksmooth` with data
   $$\{(x_k, \widetilde{h}_k(w_j))\}_{k=1}^K.$$

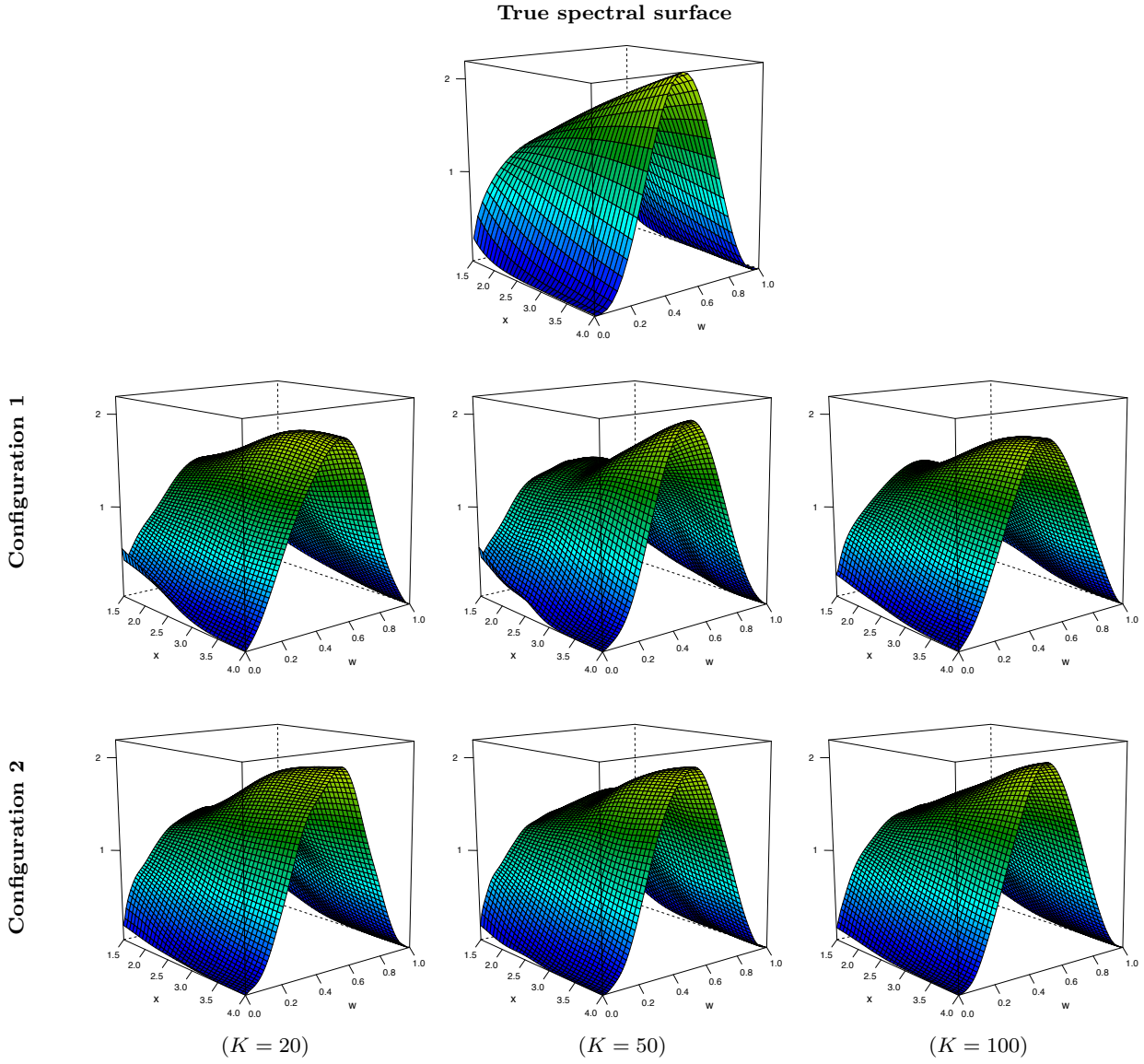


Figure 3. On the top: true spectral surface from the symmetric Dirichlet predictor-dependent model detailed in Sect. 3.1, followed by spectral surface estimates for Configurations 1 (above) and 2 (below).

3 Simulation study

3.1 Models, configurations, and preliminary experiments

We construct samples of pseudo-angles $\{\mathbf{w}_k\}_{k=1}^K$ from the Dirichlet spectral surface, a covariate-adjusted extension of the Dirichlet model (Coles and Tawn, 1991), based on the pd spectral density

$$h_x(w) = \frac{a_x b_x \Gamma(a_x + b_x + 1)(a_x w)^{a_x - 1}\{b_x(1-w)\}^{b_x - 1}}{2\Gamma(a_x)\Gamma(b_x)\{a_x w + b_x(1-w)\}^{a_x + b_x + 1}}. \quad (14)$$

Here $a_x : \mathcal{X} \mapsto (0, \infty)$, $b_x : \mathcal{X} \mapsto (0, \infty)$, and $\Gamma(t) = \int_0^\infty x^{t-1} \mathrm{e}^{-x}\, \mathrm{d}x$. The values of the parameters in (14) are chosen to produce two scenarios: a symmetric Dirichlet spectral surface with $(a_x, b_x) = (x, x)$, where $x \in \mathcal{X}_{\text{sDir}} = [1.5, 4]$; an asymmetric Dirichlet spectral surface with $(a_x, b_x) = (x, 100)$, where $x \in \mathcal{X}_{\text{aDir}} \in [0.9, 4]$. For each of the two scenarios, we consider $K \in \{20, 50, 100\}$, and for every K, the values for $\{n_k\}_{k=1}^K$ are chosen as follows:

- Configuration 1: $n_k = 50$ for $k = 1, \ldots, K$.
- Configuration 2: $n_k = 500$ for $k = 1, \ldots, K$.

This gives rise to six different simulation schemes for each of the two predictor-dependent models.

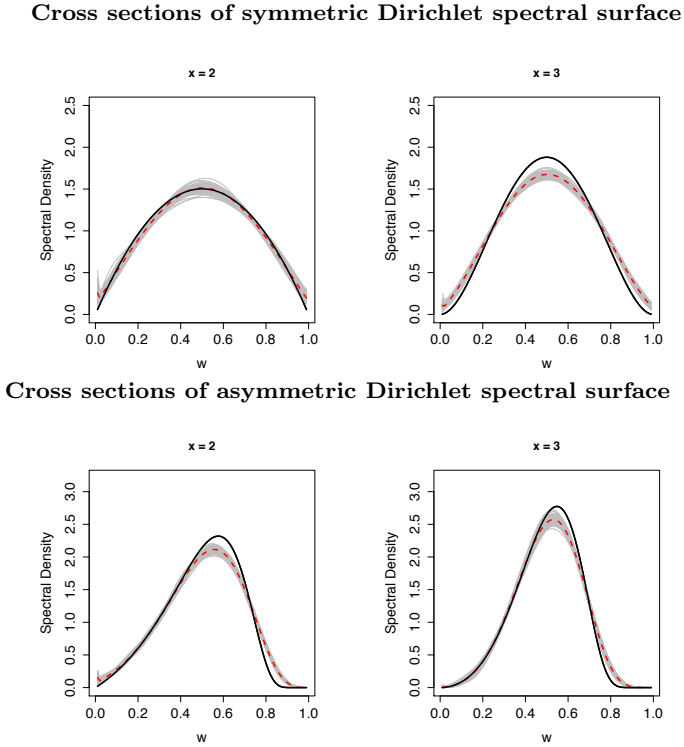


Figure 4. Trajectories of 100 randomly selected estimates of cross sections of the spectral surface (gray lines), using $K = 100$ and Configuration 1, as well as their corresponding true values (solid line) and Monte Carlo means (dashed line).

We start with a single-run experiment. Figure 2 shows true and estimated spectral densities from the symmetric (top) and asymmetric (bottom) Dirichlet models described above, for $K = 20$ values of the predictor and Configuration 1. Spectral density estimates were computed using the smooth Euclidean estimator in (8). If $\mathbb{K}_b$ is chosen as a normal kernel with standard deviation b, and if we smooth over all the predictor space using the double kernel estimator for both configurations and $K \in \{20, 50, 100\}$, we obtain what is shown in Fig. 3. Figure 3 corresponds to the symmetric Dirichlet spectral surface, where extremal dependence increases as a function of the predictor. The analogue of Fig. 3 for the asymmetric Dirichlet spectral surface is displayed in the Supplementary Materials.

The single-run experiment in Figure 3 allows us to illustrate strengths and limitations of the double kernel estimator. Pointwise estimation is troublesome at the edge of the predictor space, due to boundary bias of $\mathbb{K}_b$ which is well-known issue for many kernel-based estimators on bounded domains (Hardle, 1990, Sect. 4.4, and references therein). The double kernel estimator seems to have more difficulties estimating the asymmetric spectral surface, probably due to the need to recover a more complicated surface. In spite of these limitations, our estimator recovers satisfactorily the shape of the true spectral surface, and thus is able to reproduce satisfactorily the evolution of extremal dependence over the predictor. Another interesting aspect is that the performance of the estimator seems to be more sensitive to changes in n_k (the number of pseudo-angles for every value of the predictor) rather than changes in K (the number of predictor values).

3.2 Simulation results

Cross sections of the spectral surface give rise to (valid) spectral densities. To assess the precision of the estimates of such cross sections, we display in Fig. 4 trajectories of 100 estimates of these cross sections along with their Monte Carlo means, for $K = 100$ and Configuration 1 detailed above. These trajectories allow us to illustrate the performance of our estimator under different dependence dynamics. The top panel of Fig. 4 displays the results for the symmetric Dirichlet spectral densities, where we can see the limitations due to boundary bias that were discussed in Sect. 3.1, mostly for $x = 3$. The same plot shows that extremal dependence is underestimated by the simulations, whereas it is slightly overestimated for $x = 2$. The asymmetric Dirichlet spectral densities, presented in the bottom panel of Figure 4, display less dispersed estimates than their symmetric counterparts, and the asymmetry does not seem to be a major issue. All in all, the estimator shows a positive performance in recovering the different shapes of the cross sections of the spectral surfaces, and Monte Carlo means produce reasonable estimates.

Table 1. Mean integrated absolute error estimates of the spectral surface computed over 1000 samples for the data-generating configurations discussed in Sect. 3.1.

Model	Conf.	$K = 20$	$K = 50$	$K = 100$
Symmetric	1	0.287	0.279	0.260
Dirichlet	2	0.219	0.131	0.129
Asymmetric	1	0.499	0.438	0.416
Dirichlet	2	0.221	0.182	0.168

In Table 1 we present the mean integrated absolute error computed from 1000 samples for the data-generating configurations discussed in Sect. 3.1. Results are coherent with what we already anticipated from the single-run experiment in Fig. 3; under the same configuration for K, increasing the number of pseudo-angles yields great improvements in the performance of the estimator. On the other hand, if we fix the number of pseudo-angles and increase K, gains are not as significant. Overall, simulations confirm that the double kernel

estimator produces reasonable estimates of the spectral surface. Monte Carlo mean spectral surfaces are reported in the Supplementary Materials.

4 Extreme forest temperature illustration

4.1 Data description and preprocessing

The data were gathered from the Long-term Forest Ecosystem Research database maintained by LWF (Langfristige Waldökosystem-Forschung), and consist of daily average air temperatures under the forest canopy and in a nearby open field at 14 monitoring stations in Switzerland. At each site, data are recorded at two meteorological stations close to each other and with comparable topographic characteristics (Ferrez et al., 2011, p. 3), one under the forest cover and one in an open field; the sample period ranges 1997 to 2007. The location and altitude (above sea level) of the monitoring stations is detailed in Table 2. Our aim is to assess the dynamics governing dependence between extremely high temperatures in the open and under the canopy as a function of the altitude, and this is motivated by previous experiments conducted by Ferrez et al. (2011, p. 999), who suggested that extremal dependence between temperatures under the canopy and in the open field could be linked to altitude. The raw data consist of two series of air temperature per site, measured in circular metal shelters two meters above ground every 10 minutes. The number of observations differ from one station to the other, but comprise a total of 38.923 observations.

We use the same preprocessing steps as in Ferrez et al. (2011); in particular, we take daily maxima of the residual series resulting from removal of the annual cycle in both location and scale, i.e., we subtract a periodic mean and divide by a periodic standard deviation. Following Section 2.2, after transforming the bivariate sample at each station to unit Fréchet margins, we threshold the pseudo-radius ($R_{i,k}$) at the 98% empirical quantile (u_k) of each population, reducing the initial 38.923 observations to a total of 785 pseudo-angles. The number of pseudo-angles for each station (n_k) is detailed in Table 2.

Table 2. Locations of monitoring stations, along with the corresponding number of pseudo-angles and altitude.

Location	n_k	Altitude (m)
Beatenberg	57	1500
Bettlachstock	54	1150
Celerina	53	1890
Chironico	45	1350
Isone	45	1200
Jussy	62	500
Lausanne	63	800
Nationalpark	59	1900
Neunkirch	41	600
Novaggio	65	950
Othmarsingen	57	490
Schänis	58	750
Visp	64	700
Vordemwald	61	480

4.2 Altitude-adjusted extremal dependence

We first estimate extremal dependence of temperatures in the open and under the canopy in every site. Figure 5 shows spectral density estimates using the smooth Euclidean likelihood estimator in (8). As it can be observed in Fig. 5, different levels of extremal dependence seem to be observed over different altitudes. Particularly, altitudes between 1400 m and 1650 m present more dispersion—as can be seen for example in the spectral density estimate corresponding to Beatenberg—but, in general, strong dependence of extreme temperatures is noted in all sites. One could wonder whether there could be something particularly different about the features of Beatenberg compared to those of Chironico (1350 m) and Celerina (1890 m). Beatenberg, Chironico, and Celerina have similar slopes (from 33–35), equal soil types (Podzolic), and a similar management system (high forest) (Ferrez et al., 2011, Table 2). Beatenberg and Chironico are also mainly composed of the same species (Spruces), while Celerina is mainly composed of Larch and Arolla Pine. Thus, while other important characteristics—beyond altitude—could be driving the shape of the spectral densities in Fig. 5, from the point of view of the characteristics above Beatenberg is similar to Chironico and not that different from Celerina.

In Fig. 6 we present the spectral surface estimate computed through our double kernel estimator in (11). In a similar way that a regression line provides a graphical summary of the association between response and covariate, the spectral surface provides a graphical summary of the association between extremal dependence and a covariate. All in all, we see substantial changes in the extremal behavior, and it is not possible to identify a pattern that indicates a monotone evolution of extremal dependence. Indeed, extremal dependence seems to decrease with altitude (for moderate altitudes), and then to increase again on higher altitudes, but to levels of extremal dependence which are lower than those for moderate altitudes. Our spectral surface in Fig. 6 provides evidence reasonably compatible with the findings in Ferrez et al. (2011, p. 1000) who claim that:

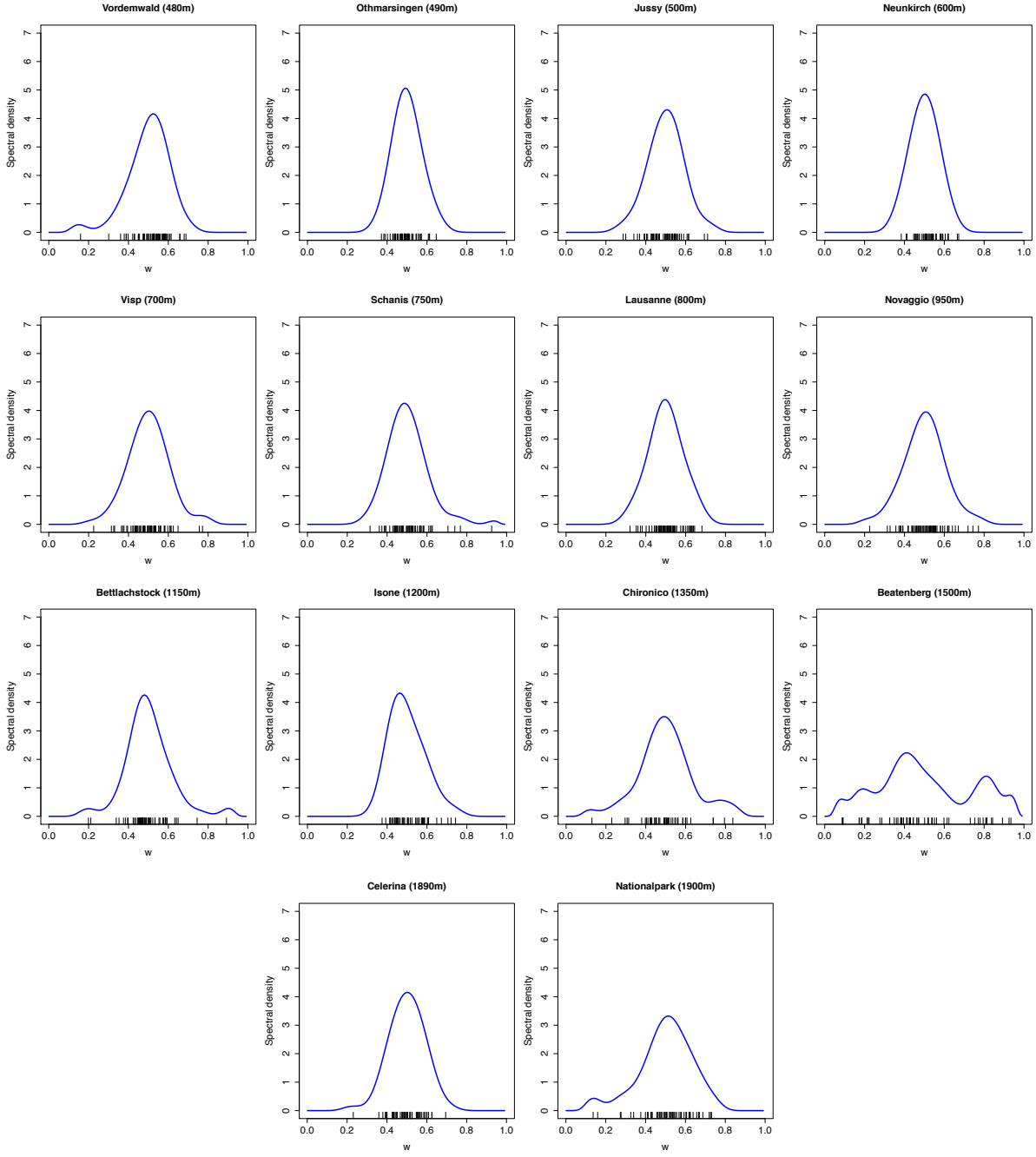


Figure 5. Spectral density estimates using the smoothed Euclidean likelihood estimator in (8) in increasing order of altitude. The associated pseudo-angles are plotted using a rug at the bottom of every plot.

> "Thus at high altitudes [...] extreme maximum temperatures under cover depend less on those in the open than at lower altitudes, indicating that the forest cover sheltering effect seems to be more efficient at higher altitudes."

Similarly to Ferrez et al. (2011), we have however no physical interpretation on the reason justifying these dynamics. Without Beatenberg we would recover similar dynamics to those in Ferrez et al. (2011) but the kink in Fig. 6 would be gone (results available from the authors).

In agreement with Fig. 5, the spectral surface in Fig. 6 also reveals strong dependence for almost all altitudes, with the exception of values between and 1400 m and 1650 m, where we can see more dispersion and consequently a decrease in the extremal dependence. In addition to this, the spectral surface in Fig. 6 provides a

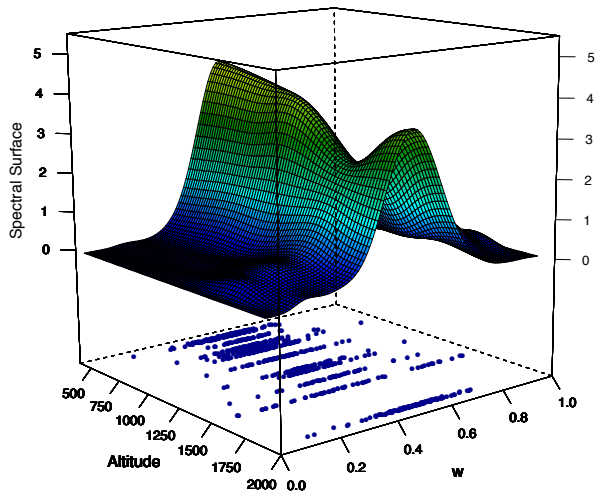


Figure 6. Spectral surface estimate using the double kernel estimator in (11); pseudo-angles are overlaid on the bottom of the box.

more complete portrait of 'altitude-adjusted' extremal dependence of these temperature data, than the spectral density estimates in Fig. 5, and for example it allows us to extrapolate into unobserved altitudes. Still, other covariates, beyond altitude, could be playing a role on determining the shapes of the spectral densities in Fig. 5—and thus the expression 'altitude-adjusted' should be here interpreted with this in mind.

5 Final remarks

We propose a density regression model that allows us to assess the changes in the extremal dependence structure over the values of a discrete predictor. This is a first step to tackle the gap between the developments in non-stationary marginal distributions and bivariate distributions. We perform inference by introducing a nonparametric double kernel estimator that smooths in two steps: first in the pseudo-angles direction using the Euclidean likelihood estimator of de Carvalho et al. (2013), and then in the covariate direction through an approach similar to the Nadaraya–Watson estimator, but where responses are spectral density estimates. Extensions can be found in Castro (2015). This paper treats the bivariate case only. While smoothing along covariates should work similarly as in the bivariate setting—in the sense that an analog of Eq. (7) can still be of use in the D-dimensional setting—a resilient extension of the smooth Euclidean likelihood estimator of de Carvalho et al. (2013) to the D-dimensional setting is nontrivial and it requires further investigation.

Our model is related to the spectral density ratio model of de Carvalho and Davison (2014) in the sense that covariates can be incorporated, but rather than just linking extremal distributions, our model assesses directly the evolution of extremal dependence over a predictor. Furthermore, implementation of our estimator is straightforward, and inference is computationally convenient.

We illustrate our methods in a temperature data application where altitude is considered a variable of interest. Results suggest an impact of the altitude on the extremal dependence of temperatures under the forest and on a open field, illustrating the need to consider covariate-adjusted extreme value dependence structures.

Acknowledgements We thank the Editor, Associate Editor, and Reviewers for helpful comments and suggestions on an earlier draft of this article. We extend our thanks to Vanda Inácio de Carvalho, Rodrigo Herrera, Raphael Huser, and Jenny Wadsworth for discussions, and to Edgardo Dörner for computational support. This research was partially supported by the Fondecyt project 11121186.

References

Castro, D (2015) Bivariate extremes: Modeling, smoothing, and regression. PhD thesis. Pontificia Universidad Católica de Chile.

Chavez-Demoulin, V, Davison, AC (2005) Generalized additive modelling of sample extremes. J R Stat Soc C, 54(1):207–222.

Chavez-Demoulin, V, Embrechts, P, Hofert, M (2015) An extreme value approach for modeling operational risk losses depending on covariates. J Risk Insur (DOI: 10.1111/jori.12059).

Chen, SX (1997) Empirical likelihood-based kernel density estimation. Australian J Stat 39(1):47–56.

Coles, SG (2001) An Introduction to the Statistical Modeling of Extreme Values. London: Springer.

Coles, SG, Tawn, JA (1991) Modelling extreme multivariate events. J R Stat Soc B 53(2):377–392.

Davison, AC, Smith, RL (1990) Models for exceedances over high thresholds (with Discussion). J R Stat Soc B 52(3):393–442.

de Carvalho, M (2016, to appear) Statistics of extremes: Challenges and opportunities. *Extreme Values in Finance: A Handbook of Extreme Value Theory and its Applications*. Ed F Longin. Hoboken: Wiley.

de Carvalho, M., and Davison, A. C. (2011) Semiparametric estimation for K-sample multivariate extremes. In: Proceedings 58th Int Stat Inst, pp. 2961–2969.

de Carvalho, M, Davison, AC (2014) Spectral density ratio models for multivariate extremes. J Am Stat Assoc 109(506):764–776.

de Carvalho, M, Oumow, B, Segers, J, Warchoł, M (2013) A Euclidean likelihood estimator for bivariate tail dependence. Comm Stat Theo Meth 42(7):1176–1192.

de Haan, L, Resnick, SI (1977) Limit theory for multivariate sample extremes. Zeitsch Wahr verw Geb 40(4):317–377.

Dunson, DB, Pillai, N, Park, JH (2007) Bayesian density regression. J R Stat Soc B 69(2):163–183.

Eastoe, EF, Tawn, JA (2009) Modelling non-stationary extremes with application to surface level ozone. J R Stat. Soc C 58(1):25–45.

Fernández-Ponce, JM, Rodríguez-Griñolo, MR (2015) Testing exponentiality against NBUE distributions with an application in environmental extremes. Stoc Environ Res Risk Assess 29(3): 679–692.

Ferrez, J, Davison, AC, Rebetez, M (2011) Extreme temperature analysis under forest cover compared to an open field. Agric For Met 151(7):992–1001.

Hainy, M, Müller, WG, Wagner, H (2016) Likelihood-free simulation-based optimal design with an application to spatial extremes. Stoch Environ Res Risk Assess (DOI 10.1007/s00477-015-1067-8).

Hardle, W (1990) Applied Nonparametric Regression. Cambridge, UK: Cambridge University Press.

Huser, R, Genton, MG (2016) Non-stationary dependence structures for spatial extremes. J Agric Biol Env Stat (to appear).

Kiriliouk, A, Segers, J, Warchoł, M (2015) Nonparametric estimation of extremal dependence. In: Extreme Value Modelling and Risk Analysis: Methods and Applications, Eds D Dey, J Yan. Boca Raton, FL: Chapman and Hall/CRC.

Nadaraya, EA (1964) On estimating regression. Theo Probab Appl 9(1):141–142.

Owen, A (2001) Empirical Likelihood. Boca Raton, FL: Chapman & Hall.

Pickands, J (1981) Multivariate extreme value distributions. Proceedings 43rd Session Int Stat Inst 859–878.

R Core Team (2014) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Wand, MP, Jones MC (1994) Kernel Smoothing. Boca Raton, FL: Chapman & Hall/CRC.

Wang, H, Chen, Y, Weihong, L (2014) Hydrological extreme variability in the headwater of Tarim River: Links with atmospheric teleconnection and regional climate. Stoch Environ Res Risk Assess 28(2):443–453.

Watson, GS (1964) Smooth regression analysis. Sankhyā 26(4):359–372.