Nonparametric estimation of the dependence among multivariate rainfall maxima

G. Marcon[1,*], S.A. Padoan[1], P. Naveau[2] and P. Muliere[1]

[1] *Bocconi Universtity, Department of Decision Sciences, Milan, Italy; giulia.marcon@phd.unibocconi.it, simone.padoan@unibocconi.it, pietro.muliere@unibocconi.it*
[2] *Laboratoire des Sciences du Climat et l'Environnement CNRS, Gif-sur-Yvette, France; Philippe.Naveau@lsce.ipsl.fr*
[*] *Corresponding author*

Abstract. *Multivariate analysis of extreme values has an increasing range of applications in risk analysis, especially in the fields of environmental sciences. For example, it would be of interest for hydrologists to extract relevant information hidden in complex spatial-temporal rainfall datasets. The aim of this work is to analyse the dependence structures of weekly maxima of hourly rainfall in France recorded from 1993 to 2011. Some weather stations, initially organised in clusters, are analysed in order to summarise the dependence within all groups of seven stations. However, beyond the bivariate case, the analysis of the dependence structures for moderately high dimensional problems is still challenging. Estimation methods for assessing the extremal dependence must satisfy appropriate assumptions for guaranteeing valid results. The approach used here focuses on the nonparametric estimation of the Pickands dependence function through a specific type of Bernstein polynomial representation which ensures that all required constraints are verified.*

Keywords. *Extreme values; Rainfall maxima; Pickands dependence function; Nonparametric estimation; Bernstein polynomials.*

1 Introduction

To study the stochastic behaviour of extreme precipitation is crucial for engineering design and the action planning for protecting societies against such events. Extreme rainfall can produce heavy floods causing several damages to population centers and potential loss of lives. Multivariate extreme value statistics describes the behaviour of two or more variables at extreme levels. In this context, the multivariate vector under study is the 49-dimensional vector of weekly maxima of hourly rainfall recorded at French weather stations in the Fall season in the period $1993-2011$. Thus, for each station, $n = 228$ values have been collected. Data are grouped in clusters as suggested by [1], where in each of the seven groups, seven weather stations are considered. Our aim is to analyse the dependence structures within climatologically homogeneous areas defined by such clusters.

Formally, a multivariate extreme value distribution is the joint limiting distribution of component-wise maxima of identically distributed random variables. This work aims to apply a nonparametric approach for estimating the dependence function, assuming knowledge of the marginal distributions, in the multivariate framework. Let $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,d}),\ i = 1, \ldots, n$, be a d-dimensional random variable, with continuous marginal distribution functions $F_1, \ldots, F_d$. Assume it follows a multivariate max-stable distribution [8] and that the margins are unit Fréchet, that is $\mathbb{P}(X_j \leq x) = e^{-1/x_j}$, $x_j > 0$. Then, the joint distribution of $\mathbf{X}$ can be written as

$$G(\mathbf{x}) = \exp\{-V(\mathbf{x})\}, \qquad V(\mathbf{x}) = \left(\frac{1}{x_1} + \ldots + \frac{1}{x_d}\right) A(\mathbf{w}), \tag{1}$$

where $w_j = \frac{x_j}{r}$ and $r = x_1 + \ldots + x_d$. The exponent measure function $V(\mathbf{x})$ is an homogeneous function of order -1 which describes the dependence among data, see [2]. Such function can be quite difficult to estimate, especially in a multivariate context. However, the same dependence structure can be detected by taking into account the Pickands dependence function $A(\mathbf{w})$ which is the restriction of $V(\mathbf{x})$ in the unit simplex of dimension $d-1$,

$$\mathcal{S}_{d-1} := \left\{ (w_1, \ldots, w_{d-1}) \in [0,1]^{d-1} : \sum_{j=1}^{d-1} w_j \leq 1 \right\}, \tag{2}$$

In order for $A(\mathbf{w})$ to be a valid Pickands function it must satisfy [3]:

1. continuity and convexity.
2. $A(\mathbf{w})$ has boundary codomain. It has lower limit at $\max\left(w_1, \ldots, w_{d-1}, 1 - \sum_{j \leq d-1} w_j\right) \geq 1/d$ which corresponds to the case of complete dependence; it has upper limit equal to one, for any $\mathbf{w} = (w_1, \ldots, w_{d-1}) \in \mathcal{S}_{d-1}$ which implies that each component of $\mathbf{X}$ is independent of all the others.
3. $A(\mathbf{e}_j) = A(\underline{0}) = 1$ for the extremal points of the simplex, $\mathbf{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ for $1 \leq j \leq d-1$.

Many nonparametric estimators of A have been proposed in the literature, however, most of them provide estimates which do not satisfy the properties $1-3$ above when dimension is high. The consequence is that the resulting estimate of $V(\mathbf{x})$ in (1) does not define a proper distribution $G(\mathbf{x})$.

2 Nonparametric estimation

The estimation procedure is divided into two parts, as proposed in [6]. A preliminary estimate $\hat{A}$ of A is initially obtained through the multivariate madogram. Such estimator does not necessarily satisfy the required properties. Therefore, a regularised estimator $\tilde{A}$ is constructed as the solution of a constrained optimization problem via a Bernstein polynomial representation. The multivariate madogram has the advantage of being interpretable as an L1-distance between the component-wise maximum and the component-wise mean of the rescaled cumulative distribution functions. It describes how far $\mathbf{X}$ is from the complete dependence case. The regularization procedure is based on the projection of the pilot estimate $\hat{A}$ onto the set of functions $\mathcal{A}$ which satisfy properties $1-3$. The representation of the Pickands dependence function is given through the Bernstein polynomial,

$$B_A(\mathbf{w};k) = \sum_{\ell \in L_k} \beta_\ell b_\ell(\mathbf{w};k) \tag{3}$$

where $k \in \mathbb{N}$ determines the order of the polynomial and the vector of coefficients $\boldsymbol{\beta}$ defines the shape of the representation, with $\beta_\ell \in [0,1]$. The polynomial basis $b_\ell(\mathbf{w};k)$, with $\mathbf{w} \in \mathcal{S}_{d-1}$, is a multivariate extension of the binomial probability function with parameters $(j;k,w)$, as in [4] and [7]. An advantage of using this approach is that Bernstein polynomials can uniformly approximate any functions [5] and since $A(\mathbf{w})$ is continuous on $\mathcal{S}_{d-1}$, it follows that the sequence of Bernstein polynomials (3) converges uniformly to A on $\mathcal{S}_{d-1}$ as k goes to infinity. Thus, the functional representation $B_A(\mathbf{w};k)$ provides an approximation of A, for any given k which becomes more accurate as k grows. The main gain of this procedure is that it regularises the estimate $\hat{A}$ by its projection $\tilde{A}$ which satisfies all constraints. Additionally, the necessary conditions are expressed as a linear constraint on the vector of the coefficients $\boldsymbol{\beta}$ in (3). Thus, the estimation method appears to be computationally feasible also in high dimensions.

A graphical representation of the improvement from the madogram to its regularised version is provided in Figure 1. The plots show the theoretical (solid line) and estimated (dashed line) Pickands functions corresponding to the trivariate symmetric logistic model with strong dependence ($\alpha = 0.3$) and sample size $n = 20$. First plot corresponds to $\hat{A}$ and second plot to $\tilde{A}$ with an order of the polynomial $k = 14$. Notice that $\hat{A}$ is not convex and therefore is not a proper Pickands function, whereas $\tilde{A}$ satisfies all the required constraints at each point of the domain.

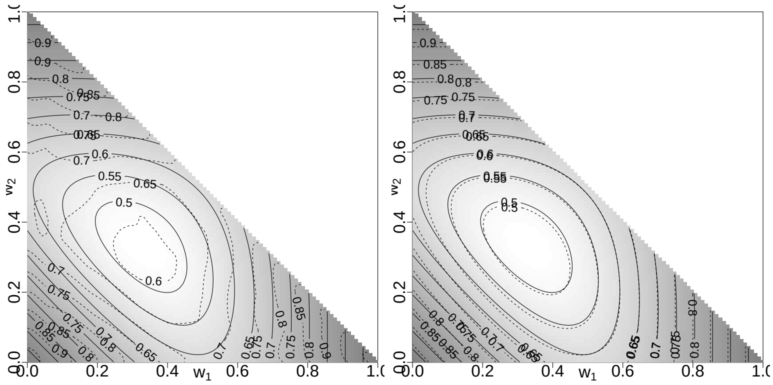

Figure 1: Theoretical (solid line) and estimated (dashed line) Pickands dependence function of the trivariate symmetric logistic model obtained by $\hat{A}$ (a) and $\tilde{A}$ (b).

3 Application: Weekly rainfall maxima

The above methodology is applied to the weakly maxima rainfall described in section 1. The Bernstein projection estimator, based on the madogram, is computed for each cluster. For seven-dimensional analysis, a graphical representation is not available. To summarise results, we consider the relation among the extremal coefficient and the Pickands function, defined by $\tilde{\theta} = d\tilde{A}(1/d,\ldots,1/d)$. In this case $d = 7$, thus $1 \leq \tilde{\theta} \leq 7$ where the lower and upper bounds represent the case of complete dependence and independence among the extremes, respectively. Extremal coefficients $\tilde{\theta}$ are displayed in Figure 2. A stronger dependence is present in the northern regions than in the south. Extreme precipitation in the south is more likely driven by localised convective storms with weak spatial dependence structures, whereas heavy rainfall is often produced by mid-latitude perturbations in Brittany or in the north of France and Paris. This confirms what would be expected from a purely climatological perspective.

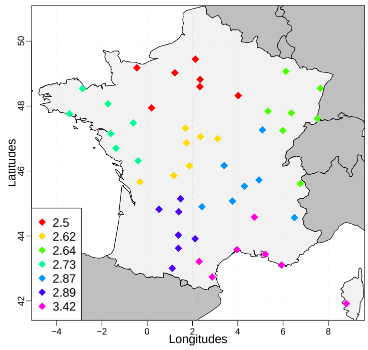

Figure 2: Map of clusters of weather stations and respective estimated extremal coefficients with $d = 7$.

References

[1] Bernard E., P. Naveau, M. Vrac and O. Mestre (2013). Clustering of maxima: Spatial dependencies among heavy rainfall in France. *Journal of Climate*.

[2] de Haan L., A. Ferreira (2006). *Extreme Value Theory: An Introduction*. Springer, Boston.

[3] Falk M., J. Hüsler, R. Reiss. (2011) *Laws of Small Numbers: Extremes and Rare Events*. Springer.

[4] Li F. J. (2011) Interpolation and convergence of bernstein-bézier coefficients. *Acta Mathematica Sinica*, English Series **27** (9), 1769-1782.

[5] Lorentz G. (1953) *Bernstein Polynominals*. Mathematical expositions. University of Toronto Press.

[6] Marcon G., S.A. Padoan, P. Naveau, P. Muliere (2014). Nonparametric estimation of the Pickands dependence function using Bernstein polynomials. *Submitted*

[7] Petrone S. (2004) On the role of mixtures in bayesian nonparametrics. *Atti della XLII Riunione Scientifica della Societá Italiana di Statistica* 1, 257-268

[8] Resnick S.I. (1987) *Extreme values, Regular Variation and Point Processes*. Berlin: Springer-Verlag