

# Dependence of maxima in space

H. Ferreira and L. Pereira

Department of Mathematics, University of Beira Interior, Covilhã, Portugal

E-mail: [helenaf@ubi.pt](mailto:helenaf@ubi.pt), [lpereira@ubi.pt](mailto:lpereira@ubi.pt)

**Abstract.** We evaluate the dependence among large values of a spatial process of maxima through a coefficient that can be applied in natural, technical and societal extreme phenomena. Its main properties are: a)  $k$  locations can be taken into account; b) it takes values in  $[0, 1]$  and higher values indicate stronger dependence; c) it is independent of the univariate marginal distributions of the random field; d) it can be related with other coefficients in the literature such as the tail dependence and the extremal coefficients; e) it agrees with the concordance property for multivariate distributions; f) it has as a particular case the variogram from geostatistics; g) it can be easily estimated. We illustrate the application of the introduced coefficient with the estimation of the degree of dependence among large values of the amount of tritium in drinking water for tree cities.

## 1. Introduction

Unpleasant extremes of variables like wind, temperature and precipitation, can affect anybody at any place. This thought is applicable also to stock markets fluctuations, given the markets' globalization.

Despite the negative connotation of extremes, who can imagine the world without extreme events like the Big bang, the extinction of some species, earthquakes or simply the hormonal chaos of passions? In short, a world where there was never large deviations from the average, the usual events or no novel opportunities?

Large fluctuations connected with phase transitions, critical phenomena or catastrophic events can be viewed as extreme or rare events. The mathematical tools for the modeling of the extremal behavior can be provided by the Theory of Extreme Values, a branch of the classical Statistics, where we find the properties of the max-stable processes and max-stable random fields.

Here we will focus on the dependence of the variables of max-stable random fields, which are the natural models for spatial extremes. They can be, for instance, good approximations for annual maxima of daily spatial rainfall and have been widely applied to real data in environmental, atmospheric and geological sciences. The Pearson correlation coefficient proves to be an inadequate measure in contexts of assessment of dependence in which the extreme values are important because it assigns equal weight to all observations when the structure of dependence between extreme observations may differ from other (Resnick (2007)).

We introduce a dependence coefficient, called multivariate variogram, to measure the dependence among extreme values of a spatial process of maxima. We justify its advantages by presenting the main properties and we illustrate its applications to real questions. In particular, we evaluate the dependence among large values of the amount of tritium (pCi/L) in drinking water, for



three cities in Alabama State (USA). The first two are relatively close to a nuclear power plant in northern Alabama and its tritium data in drinking water show greater levels in the period 1997-2013. We estimate the dependence of the annual maxima at these locations throughout the estimation of the multivariate variogram.

## 2. Definition of the multivariate variogram

Here  $\{Z(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d\}$  denotes a random field of maxima. That means that, for each choice of  $\mathbf{x}_1, \dots, \mathbf{x}_k$ , the distribution of  $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$  is a Multivariate Extreme Value (MEV) distribution  $G_{\mathbf{x}_1, \dots, \mathbf{x}_k}$ . Its tail dependence function

$$l_{\mathbf{x}_1, \dots, \mathbf{x}_k}(t_1, \dots, t_k) = \lim_{u \downarrow 0} \frac{1 - G_{\mathbf{x}_1, \dots, \mathbf{x}_k}(G_{\mathbf{x}_1}^{-1}(1 - ut_1), \dots, G_{\mathbf{x}_k}^{-1}(1 - ut_k))}{u},$$

$(t_1, \dots, t_k) \in \mathbb{R}_+^k$ , characterizes fully the dependence among its marginals distributions  $G_{\mathbf{x}_j}$ ,  $j = 1, \dots, k$  (Resnick (1987), Beirlant et al. (2004)). The superscripts  $-1$  denote the inverses of the marginal distributions.

Since this function cannot be easily inferred from data, the dependence coefficients are useful, despite the fact that one coefficient cannot preserve all the information about this function.

Several dependence coefficients have been considered in order to resume the dependence among the marginals of  $G_{\mathbf{x}_1, \dots, \mathbf{x}_k}$ : extremal coefficients, tail dependence coefficients and madogram (Li (2009), Smith (1990), Schlather, M. and Tawn, J. A. (2003), Cooley et al. (2006), Naveau et al. (2009)).

Next definition introduces our proposal and then we explore its advantages and relations with the previous coefficients. In the following,  $\bigvee_{j=1}^k a_j$  and  $\bigwedge_{j=1}^k a_j$  denote the maximum and the minimum of  $\{a_1, \dots, a_k\}$ , respectively.

**Definition 2.1.** *The variogram of  $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$  is the coefficient*

$$v(\mathbf{x}_1, \dots, \mathbf{x}_k) = 1 - \frac{k+1}{k-1} E \left( \bigvee_{j=1}^k G_{\mathbf{x}_j}(Z(\mathbf{x}_j)) - \bigwedge_{j=1}^k G_{\mathbf{x}_j}(Z(\mathbf{x}_j)) \right).$$

We remark that

$$v(\mathbf{x}_1, \dots, \mathbf{x}_k) = 1 - \frac{k+1}{k-1} E \left( \bigvee_{\{i,j\} \subset \{1, \dots, k\}} |G_{\mathbf{x}_i}(Z(\mathbf{x}_i)) - G_{\mathbf{x}_j}(Z(\mathbf{x}_j))| \right).$$

In particular, for two locations  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , it holds  $v(\mathbf{x}_i, \mathbf{x}_j) = 1 - 6\nu(|\mathbf{x}_i - \mathbf{x}_j|)$ , where

$$\nu(|\mathbf{x}_i - \mathbf{x}_j|) = \frac{1}{2} E(|G_{\mathbf{x}_i}(Z(\mathbf{x}_i)) - G_{\mathbf{x}_j}(Z(\mathbf{x}_j))|)$$

is the first-order variogram or madogram (Matheron (1987), Cooley et al. (2006)). Therefore, the previous definition extends the rescaled bivariate variogram, in order to consider several locations.

### 3. Main properties

We state in this section some properties of the multivariate variogram and its proofs are presented in the appendix. The first proposition is a technical result that relates the introduced coefficient with the previous coefficients in the literature.

**Proposition 3.1.** *The coefficient  $v(\mathbf{x}_1, \dots, \mathbf{x}_k)$  is independent of the univariate distributions  $G_{\mathbf{x}_j}$ ,  $j = 1, \dots, k$  and it holds*

$$v(\mathbf{x}_1, \dots, \mathbf{x}_k) = 1 - \frac{k+1}{k-1} \left( \frac{l_{\mathbf{x}}(1, \dots, 1)}{1 + l_{\mathbf{x}}(1, \dots, 1)} - \sum_{\emptyset \neq I \subseteq \{1, \dots, k\}} (-1)^{|I|+1} \frac{l_{\mathbf{x}_I}(1, \dots, 1)_I}{1 + l_{\mathbf{x}_I}(1, \dots, 1)_I} \right),$$

where  $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_k)$  and  $\mathbf{x}_I$  denotes the sub-vector of  $\mathbf{x}$  with indices in  $I$ .

The above proposition enhances that

- a) the tail dependence function  $l_{\mathbf{x}_1, \dots, \mathbf{x}_k}$  of  $G_{\mathbf{x}_1, \dots, \mathbf{x}_k}$  determines  $v(\mathbf{x}_1, \dots, \mathbf{x}_k)$ ;
- b) the extremal coefficients  $\epsilon_{\mathbf{x}_I} = l_{\mathbf{x}_I}(1, \dots, 1)_I$ , for  $I \subseteq \{1, \dots, k\}$  determine  $v(\mathbf{x}_1, \dots, \mathbf{x}_k)$ ;
- c) when applied for the particular case of  $k = 2$ , we obtain  $\nu(|\mathbf{x}_i - \mathbf{x}_j|) = \frac{l_{\mathbf{x}_i, \mathbf{x}_j}(1, 1)}{1 + l_{\mathbf{x}_i, \mathbf{x}_j}(1, 1)} - \frac{1}{2}$ , which is the equation (14) in Cooley et al. (2006) for the relation between the bidimensional variogram and the extremal coefficient.

We will now see that our dependence coefficient increases with respect to the multivariate concordance ordering. More concordance among the margins of the vector means that extreme values are more likely. In this case, we expect a smaller discrepancy between the maximum and the minimum of the components of the vector and thus a higher value for the coefficient.

The random vector  $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$  is more concordant than another  $k$ -dimensional random vector  $(Z(\mathbf{y}_1), \dots, Z(\mathbf{y}_k))$  (Joe (1997)) if, for each  $(u_1, \dots, u_k) \in \mathbb{R}^k$ , it holds that

$$P \left( \bigcap_{i=1}^k \{Z(\mathbf{x}_i) > u_i\} \right) \geq P \left( \bigcap_{i=1}^k \{Z(\mathbf{y}_i) > u_i\} \right)$$

and

$$P \left( \bigcap_{i=1}^k \{Z(\mathbf{x}_i) \leq u_i\} \right) \geq P \left( \bigcap_{i=1}^k \{Z(\mathbf{y}_i) \leq u_i\} \right).$$

Next proposition states that for sets of locations  $(\mathbf{x}_1, \dots, \mathbf{x}_k)$  with more concordant values for  $Z(\mathbf{x}_i)$ ,  $i = 1, \dots, k$ , we find higher values for the coefficient  $v(\mathbf{x}_1, \dots, \mathbf{x}_k)$ .

**Proposition 3.2.** *If  $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$  is more concordant than  $(Z(\mathbf{y}_1), \dots, Z(\mathbf{y}_k))$  then  $v(\mathbf{x}_1, \dots, \mathbf{x}_k) \geq v(\mathbf{y}_1, \dots, \mathbf{y}_k)$ .*

If  $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$  has MEV distribution then it is associated and then (Esary et al. (1967))

$$P \left( \bigcap_{i=1}^k \{Z(\mathbf{x}_i) > u_i\} \right) \geq \prod_{i=1}^k P(Z(\mathbf{x}_i) > u_i)$$

and

$$P \left( \bigcap_{i=1}^k \{Z(\mathbf{x}_i) \leq u_i\} \right) \geq \prod_{i=1}^k P(Z(\mathbf{x}_i) \leq u_i).$$

In other words, a vector  $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$  with MEV distribution is more concordant than the corresponding vector  $(\hat{Z}(\mathbf{x}_1), \dots, \hat{Z}(\mathbf{x}_k))$  with independent margins.

Otherwise, for the above multivariate concordance ordering the maximal distribution is the Frechét upper bound, corresponding to margins totally dependent. We have then the following consequence of the above proposition.

**Proposition 3.3.** *a) If  $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$  has independent margins then  $v(\mathbf{x}_1, \dots, \mathbf{x}_k) = 0$ .  
b) If  $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$  has totally dependent margins then  $v(\mathbf{x}_1, \dots, \mathbf{x}_k) = 1$ .  
c)  $0 \leq v(\mathbf{x}_1, \dots, \mathbf{x}_k) \leq 1$ .*

#### 4. Estimation

Let  $(Z^{(i)}(\mathbf{x}_1), \dots, Z^{(i)}(\mathbf{x}_k))$ ,  $i = 1, \dots, n$ , independent copies of  $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$  and  $\hat{G}_{\mathbf{x}_j}$  the empirical distribution function provided by  $Z^{(i)}(\mathbf{x}_j)$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, k$ . The natural estimator for the variogram  $v(\mathbf{x}_1, \dots, \mathbf{x}_k)$  is its empirical counterpart

$$\hat{v}(\mathbf{x}_1, \dots, \mathbf{x}_k) = 1 - \frac{k+1}{k-1} \times \frac{1}{n} \sum_{i=1}^n \left( \bigvee_{j=1}^k \hat{G}_{\mathbf{x}_j}(Z^{(i)}(\mathbf{x}_j)) - \bigwedge_{j=1}^k \hat{G}_{\mathbf{x}_j}(Z^{(i)}(\mathbf{x}_j)) \right).$$

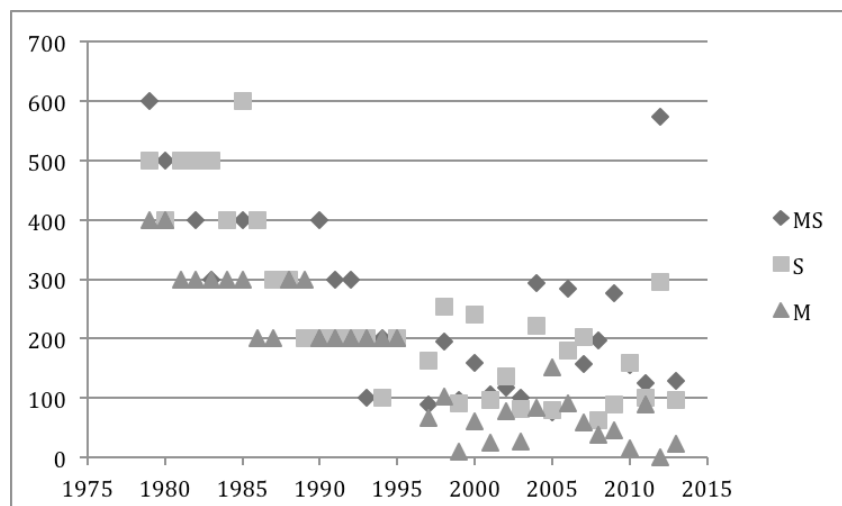
The estimator is strongly consistent and its asymptotic normality can be deduced from the Theorem 6 in Fermanian et al. (2004).

#### 5. Application

We compute the estimates for the variogram of the amount  $Z(\mathbf{x})$  of tritium (pCi/L) in drinking water, for Muscle Shoals, Scottsboro and Montgomery, three locations  $\mathbf{x}$  in Alabama State (USA).

The first two cities are relatively close to a nuclear power plant in northern Alabama (Browns Ferry reactors at Decatur): Muscle Shoals is about 20 miles west of the plant, Scottsboro is about 60 miles east of the plant and both are situated on the banks of the Tennessee River. Montgomery is far from any nuclear plant.

Measurements (quarterly) can be accessed at <http://oaspub.epa.gov/enviro/erams>. The annual maximum values are summarized in the following graph where one can observe greater concordance between the values of Muscle Shoals and Scottsboro.



The tritium data in drinking water show both Muscle Shoals and Scottsboro have greater levels than Montgomery in the period 1997-2013, and we evaluate the dependence of the annual maxima at these locations throughout the estimated variogram  $\hat{v}$ .

| location $\mathbf{x}$ | MS                                        | S                                        | M                                         |
|-----------------------|-------------------------------------------|------------------------------------------|-------------------------------------------|
| MS                    | $\hat{v}(Z(\text{MS}), Z(\text{MS}))=0$   | $\hat{v}(Z(\text{MS}), Z(\text{S}))=0,4$ | $\hat{v}(Z(\text{MS}), Z(\text{M}))=0,13$ |
| S                     | $\hat{v}(Z(\text{S}), Z(\text{MS}))=0,4$  | $\hat{v}(Z(\text{S}), Z(\text{S}))=0$    | $\hat{v}(Z(\text{S}), Z(\text{M}))=0,22$  |
| M                     | $\hat{v}(Z(\text{M}), Z(\text{MS}))=0,13$ | $\hat{v}(Z(\text{M}), Z(\text{S}))=0,22$ | $\hat{v}(Z(\text{M}), Z(\text{M}))=0$     |

The estimate for the dependence among the three locations presents a low value which can be explained by the weak dependence for the pairs  $(Z(\text{MS}), Z(\text{M}))$  and  $(Z(\text{S}), Z(\text{M}))$ . The estimated dependence for  $(Z(\text{MS}), Z(\text{S}))$  presents greater value as expected from the locations of these two cities.

## 6. Appendix: Proofs

*Proof of Proposition 3.1* Let  $\mathbf{Z}$  be a random vector with MEV distribution  $G$ , unit Fréchet  $F$  margins and tail dependence function  $l$ . It is sufficient to prove that if  $\mathbf{Y}$  has MEV distribution with marginal distributions  $F_j$ ,  $j = 1, \dots, d$ , and the same copula as  $\mathbf{Z}$ , then it holds that

$$E \left( \bigvee_{j=1}^k F_j(Y_j) \right) = E \left( \bigvee_{j=1}^k F(Z_j) \right) = \frac{l(1, \dots, 1)}{1 + l(1, \dots, 1)}.$$

We first deduce the distribution of  $\bigvee_{j=1}^k F_j(Y_j)$ . Denoting the copula of  $\mathbf{Z}$  by  $C_{\mathbf{Z}}$ , we have, for each  $u \in [0, 1]$ ,

$$P \left( \bigvee_{j=1}^k F_j(Y_j) \leq u \right) = C_{\mathbf{Z}}(u, \dots, u) = G \left( -\frac{1}{\log u}, \dots, -\frac{1}{\log u} \right) = \exp(-l((-\log u), \dots, (-\log u))) = u^{l(1, \dots, 1)}.$$

Then

$$E \left( \bigvee_{j=1}^k F_j(Y_j) \right) = \int_0^1 u^{l(1, \dots, 1)} l(1, \dots, 1) du = \frac{l(1, \dots, 1)}{1 + l(1, \dots, 1)}.$$

The result then follows by relating the minimum and maximum through

$$\bigwedge_{j=1}^k F(X_j) = \sum_{\emptyset \neq I \subseteq \{1, \dots, k\}} (-1)^{|I|+1} \bigvee_{j \in I} F(X_j).$$

*Proof of Proposition 3.2* The concordant order implies (Shaked and Shanthikumar (2007)) that

$$E \left( f \left( \bigwedge_{i=1}^k Z(\mathbf{x}_i) \right) \right) \geq E \left( f \left( \bigwedge_{i=1}^k Z(\mathbf{y}_i) \right) \right)$$

and

$$E \left( f \left( \bigvee_{i=1}^k Z(\mathbf{x}_i) \right) \right) \leq E \left( f \left( \bigvee_{i=1}^k Z(\mathbf{y}_i) \right) \right),$$

for all non-decreasing functions  $f$ . The result then follows by taking, without loss of generality,  $G_{\mathbf{x}_j} = G = f$ .

*Proof of Proposition 3.3* For  $\mathbf{Z} = (Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$  let

$$R(\mathbf{Z}, \mathbf{x}) = E \left( \bigvee_{j=1}^k G_{\mathbf{x}_j}(Z(\mathbf{x}_j)) - \bigwedge_{j=1}^k G_{\mathbf{x}_j}(Z(\mathbf{x}_j)) \right).$$

We will prove that if  $\bar{\mathbf{Z}} = (\bar{Z}(\mathbf{x}_1), \dots, \bar{Z}(\mathbf{x}_k))$  has totally dependent margins then it holds that  $R(\bar{\mathbf{Z}}, \mathbf{x}) = 0$  and if  $\hat{\mathbf{Z}} = (\hat{Z}(\mathbf{x}_1), \dots, \hat{Z}(\mathbf{x}_k))$  has independent margins then it holds that  $R(\hat{\mathbf{Z}}, \mathbf{x}) = \frac{k-1}{k+1}$ . We need to apply the following equalities to obtain these results:

$$\sum_{\emptyset \neq I \subseteq \{1, \dots, k\}} (-1)^{|I|+1} = 1, \quad \sum_{i=1}^{k+1} (-1)^{i+1} \binom{k+1}{i} i = 0, \quad \sum_{i=0}^{k+1} (-1)^i \binom{k+1}{i} = 0.$$

From the proposition 3.1 we have  $R(\bar{\mathbf{Z}}, \mathbf{x}) = \frac{1}{2} - \sum_{\emptyset \neq I \subseteq \{1, \dots, k\}} (-1)^{|I|+1} \frac{1}{2} = 0$  and

$$\begin{aligned} R(\hat{\mathbf{Z}}, \mathbf{x}) &= \frac{k}{k+1} - \sum_{\emptyset \neq I \subseteq \{1, \dots, k\}} (-1)^{|I|+1} \frac{|I|}{1+|I|} = \\ &= \frac{k}{k+1} - \sum_{j=1}^k (-1)^{j+1} \binom{k}{j} \frac{j}{j+1} = \frac{k}{k+1} + \frac{1}{k+1} \sum_{j=1}^k (-1)^{j+2} \binom{k+1}{j+1} j = \\ &= \frac{k}{k+1} + \frac{1}{k+1} \sum_{j=1}^k (-1)^{j+2} \binom{k+1}{j+1} (j+1) - \frac{1}{k+1} \sum_{j=1}^k (-1)^{j+2} \binom{k+1}{j+1} = \\ &= \frac{k}{k+1} + \frac{1}{k+1} \left( 0 - \binom{k+1}{1} \right) + \frac{1}{k+1} \left( 0 - \binom{k+1}{0} + \binom{k+1}{1} \right) = \frac{k-1}{k+1}. \end{aligned}$$

The statement in c) follows from a), b) and the previous proposition.

**Acknowledgments:** We acknowledge the valuable suggestions and corrections from the referee. This work was partially supported by the research project PEst-OE/MAT/UI0212/2014.

## References

- [1] Beirlant J, Goegebeur Y, Segers J and Teugels J 2004 *Statistics of Extremes: Theory and Application*. John Wiley.
- [2] Cooley K, Naveau P and Poncet P 2006 Variograms for max-stable random fields. *Lecture Notes in Statistics* **187** 373-390, New York, NY: Springer.
- [3] Esary J D, Proschan F and Walkup D W 1967 Association of Random Variables, with Applications. *The Annals of Mathematical Statistics* **38** 1466-74.
- [4] Fermanian J D, Radulović D and Wegkamp M 2004 Weak convergence of empirical copula processes Bernoulli **5**, 847-60.

- [5] Joe H. 1997 *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.
- [6] Li H 2009 Orthant tail dependence of multivariate extreme value distributions *J. Multivariate Anal.*, **100** 243-56.
- [7] Matheron G 1987 Suffit-il, pour une covariance, d'être de type positif? *Sci. Terre. Info. Geol.* **26** 51-66.
- [8] Naveau P, Guillou A, Cooley D and Diebolt J 2009 Modelling pairwise dependence of maxima in space, *Biometrika* **96** 1-17.
- [9] Resnick S 1987 *Extreme Values, Regular Variation and Point Processes* Springer-Verlag, New York.
- [10] Resnick S 2007 *Heavy-Tail Phenomena* Springer-Verlag, New York.
- [11] Shaked M and Shanthikumar J.G. 2007 *Stochastic orders*. Springer-Verlag, New York.
- [12] Smith R.L. 1990 Max-stable processes and spatial extremes. Preprint, Univ. North Carolina, USA.