Generalized madogram and pairwise dependence of maxima over two disjoint regions of a random field

Fonseca, Cecília
Instituto Politécnico da Guarda, Portugal
Guarda, Portugal
E-mail: cfONSECA@ipg.pt

Pereira, Luísa
University of Beira Interior, Department of Mathematics, Portugal
Avenida Marquês d’Ávila e Bolama
6200-001 Covilhã, Portugal
E-mail: lpereira@ubi.pt

Ferreira, Helena
University of Beira Interior, Department of Mathematics, Portugal
Marquês d’Ávila e Bolama
6200-001 Covilhã, Portugal
E-mail: helenaf@ubi.pt

Martins, Ana Paula
University of Beira Interior, Department of Mathematics, Portugal
Avenida Marquês d’Ávila e Bolama
6200-001 Covilhã, Portugal
E-mail: amartins@ubi.pt

1. Introduction

Quantifying dependence between extreme events occurring at several locations of a random field is a fundamental issue in applied spatial extreme value analysis. For max-stable processes, \( Z = \{Z_x\}_{x \in \mathbb{R}^2} \), an important measure of pairwise dependence is the \( \lambda \)-madogram defined in Naveau et al. [4] as

\[
\nu^\lambda(x_1, x_2) = \frac{1}{2} E \left[ \left| F^\lambda(Z_{x_1}) - F^{1-\lambda}(Z_{x_2}) \right| \right], \quad \lambda \in (0, 1),
\]

where \( F \) denotes the marginal distribution of \( Z \). This function resumes the dependence structure of \( (Z_{x_1}, Z_{x_2}) \).

In this paper we propose a generalization of the \( \lambda \)-madogram that enables the analysis of dependence between maxima over two disjoint regions of locations \( x = \{x_1, \ldots, x_k\} \) and \( y = \{y_1, \ldots, y_s\} \). This generalized madogram is defined as

\[
\nu^{\alpha, \beta}(x, y) = \frac{1}{2} E \left[ \left| F^{\alpha}(M(x)) - F^{\beta}(M(y)) \right| \right], \quad \alpha \geq 0, \quad \beta \geq 0,
\]

where \( M(x) = \bigvee_{i=1}^k Z_{x_i} \) and \( M(y) = \bigvee_{i=1}^s Z_{y_i} \).

Remark 1 When we take \( \beta = 1 - \alpha, \alpha \in (0, 1) \), and \( k = s = 1 \) in (2), we obtain (1).

The paper is organized as follows. Some properties of the function \( \nu^{\alpha, \beta}(x, y) \) are summarized in Section 2. In Section 3 we introduce two estimators of \( \nu^{\alpha, \beta}(x, y) \). Finally, in section 4 we define a multivariate maxima of moving maxima random field, compute its generalized madogram for some
choices of \( x \) and \( y \) and analyze the performance of the estimators of \( \nu^{\alpha,\beta}(x, y) \) through a simulation study.

2. Generalized madogram and dependence of spatial extreme events

In this section we shall assume that the margins of \( Z = \{ Z_x \}_{x \in \mathbb{R}^2} \) have a unit Fréchet distribution, \( F(x) = \exp(-x^{-1}), \ x > 0 \).

The following proposition states that \( \nu^{\alpha,\beta}(x, y) \) provides dependence information between the regions \( x \) and \( y \) through the dependence function of a multivariate extreme value distribution.

**Proposition 1** For any max-stable random field with unit Fréchet margins and for each pair of disjoint regions of locations \( x = \{ x_1, \ldots, x_k \} \) and \( y = \{ y_1, \ldots, y_s \} \) in \( \mathbb{R}^2 \), we have

\[
\nu^{\alpha,\beta}(x, y) = \frac{V_{x,y}(\alpha, \alpha, \beta, \beta)}{1 + V_{x,y}(\alpha, \alpha, \beta, \beta)} - c(\alpha, \beta)
\]

with

\[
c(\alpha, \beta) = \frac{1}{2} \left( \frac{V_x(1, 1, 1)}{\alpha + V_x(1, 1, 1)} + \frac{V_y(1, 1, 1)}{\beta + V_y(1, 1, 1)} \right),
\]

where

\[
V_{x,y}(z_1, \ldots, z_k, z_{k+1}, \ldots, z_{k+s}) = -\ln G_{x,y}(z_1, \ldots, z_k, z_{k+1}, \ldots, z_{k+s})
\]

and

\[
G_{x,y}(z_1, \ldots, z_{k+s}) = P\left( \bigcap_{i=1}^k \{ Z_{x_i} \leq z_i \} \cap \bigcap_{s=1}^s \{ Z_{y_s} \leq z_{s+k} \} \right), \quad z_i \in \mathbb{R}.
\]

To obtain the result, we just transform the definition of \( \nu^{\alpha,\beta}(x, y) \) through the relation \( |a - b| = 2(a \lor b) - (a + b) \).

**Remark 2** If \( Z \) has unit Fréchet margins, the dependence function \( V \) is homogeneous of order -1, i.e., \( V(\alpha u_1, \ldots, \alpha u_k) = \alpha^{-1} V(u_1, \ldots, u_k) \), and therefore

\[
\epsilon_x = V_x(1, 1, 1),
\]

where \( \epsilon_x \) is the extremal coefficient defined in Schlather and Tawn [5], which measures the extremal dependence between the variables indexed in the region \( x \). So, \( c(\alpha, \beta) \) considers the dependence in each of the regions \( x \) and \( y \) through the extremal coefficients of vectors with margins \( Z_{x_1}, \ldots, Z_{x_k} \) and \( Z_{y_1}, \ldots, Z_{y_s} \).

In the following proposition we establish some properties of the generalized madogram.

**Proposition 2** Let \( x = \{ x_1, \ldots, x_k \} \) and \( y = \{ y_1, \ldots, y_s \} \) be disjoint regions of \( \mathbb{R}^2 \). We have, for each \( \alpha, \beta \in \mathbb{R}^+_0 \),

1. \( 0 \leq \nu^{\alpha,\beta}(x, y) \leq \frac{1}{2} \);
2. \( \nu^{0,0}(x, y) = 0 \);
3. \( \nu^{0,\beta}(x, y) = \frac{\beta}{2(\beta + \epsilon_x)} \);
4. \( \nu^{\alpha,0}(x, y) = \frac{\alpha}{2(\alpha + \epsilon_x)} \);
5. \( \nu^{\alpha,\beta}(x, y) = \frac{c(x_1, \ldots, x_k, y_1, \ldots, y_s)}{\alpha + \epsilon_x + \beta - \epsilon_x} - \frac{1}{2} \left( \frac{\epsilon_x}{\alpha + \epsilon_x} + \frac{\epsilon_y}{\alpha + \epsilon_y} \right) \).
Remark 3 The function $\nu^{\alpha,\beta}(x, y)$ can also be related with the dependence coefficients considered in Ferreira [3] as follows:

$$
\nu^{\alpha,\beta}(x, y) = \frac{\epsilon_y \epsilon_1(x, y)}{\alpha + \epsilon_y \epsilon_1(x, y)} - c(\alpha, \alpha) = \frac{(\epsilon_y + \epsilon_x) \epsilon_2(x, y)}{\alpha + (\epsilon_y + \epsilon_x) \epsilon_2(x, y)} - c(\alpha, \alpha), \quad \alpha > 0,
$$

where $\epsilon_1(x, y) = \frac{\epsilon(x_1, \ldots, x_k, y_1, \ldots, y_s)}{\epsilon_y}$ and $\epsilon_2(x, y) = \frac{\epsilon(x_1, \ldots, x_k, y_1, \ldots, y_s)}{\epsilon_x + \epsilon_y}$. These coefficients evaluate the strength of dependence between the events $\{M(x) \leq u\}$ and $\{M(y) \leq u\}$.

Remark 4 If the variables $M(x)$ and $M(y)$ are independent then

$$
\nu^{\alpha,\beta}(x, y) = \frac{\epsilon_x + \epsilon_y}{\alpha + \epsilon_x + \epsilon_y} - c(\alpha, \alpha),
$$

whereas if the variables are totally dependent

$$
\nu^{\alpha,\beta}(x, y) = \frac{\epsilon_x + \epsilon_y}{2\alpha + \epsilon_x + \epsilon_y} - c(\alpha, \alpha).
$$

Remark 5 The relation in 5. of Proposition 2. suggest that estimators for $\nu^{\alpha,\beta}(x, y)$ can be considered from those available for $\epsilon_x$ and $\epsilon_y$ (Schlather and Tawn [5]).

3. Estimating the generalized madogram

Let $(Z_{x1}^{(1)}, \ldots, Z_{xk}^{(1)})$ and $(Z_{y1}^{(1)}, \ldots, Z_{yk}^{(1)})$, $t = 1, \ldots, T$, be independent replications of $(Z_{x1}, \ldots, Z_{xk})$ and $(Z_{y1}, \ldots, Z_{yk})$, respectively. Hence $\{M_t(x) = \sqrt{\sum_{i=1}^k Z_{xi}^{(t)}}, \quad t = 1, \ldots, T\}$ and $\{M_t(y) = \sqrt{\sum_{i=1}^k Z_{yi}^{(t)}}, \quad t = 1, \ldots, T\}$ are random samples of $M(x)$ and $M(y)$, respectively.

If the marginal distribution $F$ of $Z$ is known then a natural estimator for the generalized madogram is given by

$$
\widehat{\nu}^{\alpha,\beta}(x, y) = \frac{1}{2T} \sum_{t=1}^T \left| F^{\alpha}(M_t(x)) - F^{\beta}(M_t(y)) \right|, \quad \alpha \geq 0, \quad \beta \geq 0.
$$

When $F$ is unknown it can be estimated by the empirical distribution function and in this case we obtain the following estimator for $\nu^{\alpha,\beta}(x, y)$:

$$
\widehat{\nu}^{\alpha,\beta}(x, y) = \frac{1}{2T} \sum_{t=1}^T \left| \hat{F}_{KT}(M_t(x)) - \hat{G}_{ST}(M_t(y)) \right|, \quad \alpha \geq 0, \quad \beta \geq 0,
$$

where

$$
\hat{F}_{KT}(u) = \frac{1}{kT} \sum_{i=1}^k \sum_{j=1}^T \mathbb{I}_{\{M_j(x_i) \leq u\}} \quad \text{and} \quad \hat{G}_{ST}(u) = \frac{1}{sT} \sum_{i=1}^s \sum_{j=1}^T \mathbb{I}_{\{M_j(y_i) \leq u\}}.
$$

Theoretical properties of this estimator can be derived in the framework of general rank order statistics of extreme events (Fermanian et al. [2], Van Der Vaart and Wellner [7]).

4. An M4 random field

It is well known that the class of max-stable processes called multivariate maxima of moving maxima processes or simply M4 processes, introduced by Smith and Weissman [6], is particularly well adapted to modeling the extreme behaviour of several time series.

To illustrate the generalized madogram given in (2) we will now define an M4 random field.
Let's consider that the distribution of $(Z_{x_1}, \ldots, Z_{x_p})$ is characterized by the copula

$$C(u_{x_1}, \ldots, u_{x_p}) = \prod_{i=1}^{+\infty} \prod_{m=-\infty}^{+\infty} u_{x_i}^{0m}, \quad u_{x_i} \in [0,1], \quad i = 1, \ldots, p,$$

where, for each $x \in \mathbb{Z}^2$, $\{a_{lmx}\}_{m,n}^{+\infty}$ are non-negative constants such that $\sum_{m=-\infty}^{+\infty} \sum_{l=1}^{+\infty} a_{lmx} = 1$. This random field $Z$ is max-stable, since, for each $t > 0$, the copula (4) satisfies

$$C^{t}(u_{x_1}, \ldots, u_{x_p}) = C(u_{x_1}^{t}, \ldots, u_{x_p}^{t}),$$

for any locations $x_1, \ldots, x_p$.

As the M4 process considered in Weissman and Smith [6], we can consider that for each location $x$, $Z_x$ is a moving maxima of variables $X_{l,n}$, i.e.,

$$Z_x = \max_{l \geq 1} \max_{-\infty < m < +\infty} a_{lmx}X_{l,1-m}, \quad x \in \mathbb{Z}^2,$$

where $\{X_{l,n}\}_{l \geq 1, n \in \mathbb{Z}}$ is a family of independent unit Fréchet random variables. The dependence structure of $(Z_{x_1}, \ldots, Z_{x_p})$ is regulated by the signature patterns $a_{lmx}$ and is given by (4).

For each pair of regions $x = \{x_1, \ldots, x_k\}$ and $y = \{y_{k+1}, \ldots, y_{k+s}\}$ we have

$$V_{x,y}(z_1, \ldots, z_k, z_{k+1}, \ldots, z_{k+s}) = -\ln C(e^{-z_1}, \ldots, e^{-z_{k+s}})$$

$$= \sum_{i=1}^{+\infty} \sum_{m=-\infty}^{+\infty} \sum_{i=1}^{k+s} z_i^{-1} a_{lmx_i}, \quad z_i \in \mathbb{R}, \quad i = 1, \ldots, k + s,$$

and consequently, for $\alpha > 0$ and $\beta > 0$ we obtain

$$V_{x,y}(\alpha, \ldots, \alpha, \beta, \ldots, \beta, \ldots) = \sum_{i=1}^{+\infty} \sum_{m=-\infty}^{+\infty} \sum_{i=1}^{k+s} \left( \alpha^{-1} a_{lmx_i} \lor \beta^{-1} a_{lmx_i}, \beta^{-1} a_{lmx_i} \right),$$

To illustrate the computation of the generalized madogram we shall consider, in what follows, examples with a finite number of signature patterns ($1 \leq l \leq L$) and a finite range of sequential dependencies ($M_1 \leq m \leq M_2$).

**Example 4.1** Let's consider that for each location $x \in \mathbb{Z}^2$ with even coordinates we have $a_{11x} = a_{12x} = \frac{1}{2}$ and otherwise $a_{11x} = \frac{1}{4} = 1 - a_{12x}$. The values of $(a_{11x}, a_{12x})$ determine the moving pattern or signature pattern of the random field, which in this case corresponds to one pattern ($L = 1$).

For the disjoint regions of locations $x = \{(2, 1), (2, 2)\}$ and $y = \{(3, 3), (3, 4)\}$ we have

$$V_{x,y}(\alpha, \alpha, \beta, \ldots) = \frac{3}{4}(\alpha^{-1} \lor \beta^{-1}) + \frac{3}{4}(\alpha^{-1} \lor \beta^{-1})$$

and therefore, the generalized madogram in this pair of locations is given by

$$\nu^{\alpha, \beta}(x, y) = \frac{1}{4}(2\alpha^{-1} \lor \beta^{-1}) + \frac{3}{4}(\alpha^{-1} \lor \beta^{-1}) = \frac{1}{2} \left( \frac{5}{4} \alpha + \frac{1}{4} \beta \right), \quad \alpha > 0, \beta > 0.$$

**Example 4.2** Let's now assume that for each location $x = (i, j) \in \mathbb{Z}^2$, $a_{11x} = \frac{1}{4} = 1 - a_{12x}$ if $i \leq j$ and $a_{11x} = \frac{2}{3} = 1 - a_{12x}$ if $i > j$.

As in the previous example the M4 random field generated by these sequences has a single signature pattern. Considering now two disjoint regions of locations with different size, $x = \{(1, 1)\}$ and $y = \{(3, 2), (3, 3), (4, 3)\}$, we obtain

$$V_{x,y}(\alpha, \beta, \beta, \beta) = \frac{1}{4}(\alpha^{-1} \lor 3\beta^{-1}) + \frac{3}{4}(\alpha^{-1} \lor \beta^{-1})$$
Int. Statistical Inst.: Proc. 58th World Statistical Congress, 2011, Dublin (Session CPS020)

and consequently

\[ \nu^{\alpha, \beta}(x, y) = \frac{1}{4}(\alpha^{-1} \lor 3\beta^{-1}) + \frac{3}{4}(\alpha^{-1} \lor \beta^{-1}) - \frac{1}{2} \left( \frac{1}{\alpha + 1} + \frac{3}{\beta + \frac{3}{2}} \right), \quad \alpha > 0, \ \beta > 0. \]

**Example 4.3** As stated in Zhang and Smith [8], in a real data generating process it is unrealistic to assume that a single signature pattern would be sufficient to describe the shape of the process every time it exceeds some high threshold. Hence, we shall now consider one example with two signature patterns \((L = 2)\).

Let us assume that for each location \(x = (i, j)\) we have \(a_{11x} = a_{12x} = a_{13x} = \frac{1}{12},\)
\(a_{21x} = a_{22x} = a_{23x} = \frac{1}{4}\) if both coordinates are odd and \(a_{11x} = \frac{1}{18},\) \(a_{12x} = \frac{1}{9},\) \(a_{13x} = \frac{1}{6},\)
\(a_{21x} = a_{22x} = a_{23x} = \frac{2}{9}\) otherwise. Now the values of \((a_{11x}, a_{12x}, a_{13x})\) and \((a_{21x}, a_{22x}, a_{23x})\) define the two signature patterns of the random field.

For two disjoint regions \(x = \{(2, 1), (2, 2)\}\) and \(y = \{(2, 3), (3, 3)\}\) we now have

\[ V_{x,y}(\alpha, \alpha, \beta, \beta) = \left( \alpha^{-1} \frac{1}{18} \lor \beta^{-1} \frac{1}{12} \right) + \frac{9}{6} (\alpha^{-1} \lor \beta^{-1}) + \frac{1}{6} (\alpha^{-1} \lor \beta^{-1}) + \left( \alpha^{-1} \frac{3}{4} \lor \beta^{-1} \frac{3}{4} \right)\]

and consequently

\[ \nu^{\alpha, \beta}(x, y) = \frac{\left( \alpha^{-1} \frac{1}{18} \lor \beta^{-1} \frac{1}{12} \right) + \frac{\alpha^{-1} \lor \beta^{-1}}{9} + \frac{\alpha^{-1} \lor \beta^{-1}}{6} + \frac{2\alpha^{-1} \lor \beta^{-1}}{3} + \frac{3\beta^{-1}}{4}}{1 + \frac{\alpha^{-1} \lor \beta^{-1}}{12} + \frac{\alpha^{-1} \lor \beta^{-1}}{6} + \frac{\alpha^{-1} \lor \beta^{-1}}{9} + \frac{2\alpha^{-1} \lor \beta^{-1}}{3} + \frac{3\beta^{-1}}{4}} \]

\[ - \frac{1}{2} \left( \frac{1}{\alpha + 1} + \frac{10}{\beta + \frac{3}{2}} \right), \quad \alpha > 0, \ \beta > 0. \]

These examples will be used in the following simulation studies to assess the performance of the estimator given in (3). The figures 1., 2. and 3. show the simulation results obtained by generating 50 replications of 100 independently and identically distributed max-stable M4 random fields in the three situations previously presented, with \(\alpha\) and \(\beta\) taking values in \(\{0.01, k \times 0.5 : k = 1, \ldots, 40\}\).

The performance of the estimator \(\hat{\nu}^{\alpha, \beta}(x, y)\) is given by the estimated mean square error.

The estimator \(\hat{\nu}^{\alpha, \beta}(x, y)\) gives estimates quite close to the theoretical value. The largest mean square errors are obtained when \(\alpha = \beta\).

![Figure 1: Simulation results obtained with Example 4.1 \((x = \{(2, 1), (2, 2)\}, y = \{(3, 3), (3, 4)\})\) for the true values of the generalized madogram \((\nu^{\alpha, \beta}(x, y))\), the estimated values \((\hat{\nu}^{\alpha, \beta}(x, y))\) and the estimated mean squared error \((MSE)\).](image-url)
Figure 2: Simulation results obtained with Example 4.2 ($x = \{(1,1)\}, y = \{(3,2),(3,3),(4,3)\}$) for the true values of the generalized madogram ($\nu^{\alpha,\beta}(x, y)$), the estimated values ($\hat{\nu}^{\alpha,\beta}(x, y)$) and the estimated mean squared error ($MSE$).

Figure 3: Simulation results obtained with Example 4.3 ($x = \{(2,1),(2,2)\}, y = \{(2,3),(3,3)\}$) for the true values of the generalized madogram ($\nu^{\alpha,\beta}(x, y)$), the estimated values ($\hat{\nu}^{\alpha,\beta}(x, y)$) and the estimated mean squared error ($MSE$).

Acknowledgments: This research was supported by the research unit "Centro de Matemática" of the University of Beira Interior and the research project PTDC/MAT/108575/2008 through the Foundation for Science and Technology (FCT) co-financed by FEDER/COMPETE.

References


