Geostatistical analysis with copula-based models of madograms, correlograms and variograms.

Fabrice Ouoba[1], Diakarya Barro[2,*], Hay Yoba Talkibing[1]

[1] *LANIBIO, UFR-SEA, Université Ouaga 1, Pr JKZ, Burkina Faso*
[2] *UFR-SEG, Université Ouaga 2, Burkina Faso*

Abstract. This paper investigates models of stochastic dependence with geostatistical tools. Specifically, we use copulas to propose new models of stochastic spatial tools such as variograms, correlograms and the madograms. Copula versions of covariograms are provided both in second stationnary and intrinsic frameworks. Moreover, some usual families of models of variograms are clarified with the corresponding parameters

1. Introduction

Spatial statistics focus on phenomena whose observations is a random process $Z = \{Z_s, s \in S\}$ indexed by a spatial set $S = \{s_1, \ldots, s_n\}$ while Z_s denotes a geographical space D. Such technics where developed first in geostatistics more specifically from the for geologists. Geostatistics are applications of probabilistic analysis methods to the study of phenomena that extends into space and present a structuration. Here, space refers to be the geographical space, but it may be the temporal axis or more abstract spaces. To quantify this structure, the geostatistical tools used are mainly the variogram, the correlogram and the madogram depending on the type of sampled data.

While modelling spatial extreme variability of an isotropic and max-stable field, Cooley et al. (2006) have introduced the F-madogram $\gamma_{_F}(h)$ defined by

$$\gamma_F(h) = \tfrac{1}{2} E\left\{ |F(Z(s)) - F(Z(s+h))| \right\}. \tag{1}$$

where h is the average value of the separating distance between the two points. This tool provides a generalization of the so called λ-madogram associated to the distribution underlying the stochastic process, such as:

$$\gamma_F(h) = \tfrac{1}{2} E\left\{ \left| [F(Z(s))]^\lambda - [F(Z(s+h))]^{1-\lambda} \right| \right\}; \lambda \in {]0,1[}. \tag{2}$$

The variogram or the semi-variogram makes it possible to determine whether the distribution or parameters studied have a structure, random or periodic. Its representation has three characteristic properties: the nugget effect, the range and the sill. The nugget effect characterizes the variability at the origin. The sill, if it exists, is characterized by the attainment of a plateau where the semi-variogram become constant with the evolution of h and the range which characterizes the limiting distance of spatial structuring.

The correlogram function is identical to the linear coefficient between a series of spatial data. It's given by:

$$\rho(s_1, s_2) = corr(Z(s_1), Z(s_2)) = \frac{cov(Z(s_1), Z(s_2))}{\sigma_{s_1} \sigma_{s_2}}. \tag{3}$$

It is possible to express graphically the correlation between two variables by mean of their separate distance h. In particular, for spatial extreme or spatial temporal phenomena, geostatistical tools such as the variogram or correlogram are not appropriate for studying the spatial structuring of data.

Typically, the madogram or variogram of the first order is used to characterize this spatial structure of the extreme data. Such as, for all separating distance h,

$$M(h) = \frac{E(|Z(s+h) - Z(x)|)}{2}. \tag{4}$$

While studying spatial models of extreme values, Barro et al.[2] have considered a set of locations $S = \{x_1, ..., x_s\} \subset \mathbb{R}^2$, where the process is observed. If $Y_{k,1}; ...; Y_{k,s}$ denote independent copies from the second-order stationary random field, for $k = 1, ..., n$, they poined out that every spatial univariate marginal laws lies in the domain of attraction of the real-value parametric Generalized extreme value (GEV) distribution, defined spatially on the subdomain:

$$S_\xi = \{x_i \in S; \sigma_i(x_i) + \xi_i(x_i)(y_i(x_i) - \mu_i(x_i)) > 0\} \subset S,$$

such as:

$$GEV(y_i(x_i)) = \begin{cases} \exp\left\{-\left[1 + \xi_i(x_i)\left(\frac{y_i(x_i) - \mu_i(x_i)}{\sigma_i(x_i)}\right)\right]^{\frac{-1}{\xi_i(x_i)}}\right\} & if \ \xi_i(x_i) \neq 0 \\ \exp\left\{-\exp\left\{-\left(\frac{y_i(x_i) - \mu_i(x_i)}{\sigma_i(x_i)}\right)\right\}\right\} & if \ \xi_i(x_i) = 0 \end{cases} ; \tag{5}$$

where the parameters $\{\mu_i(x_i) \in \mathbb{R}\}$, $\{\sigma_i(x_i) > 0\}$ and $\{\xi_i(x_i) \in \mathbb{R}\}$ are referred to as the spatial version of location, the scale and the shape parameters for the site x_i respectively.

The major contribution of this paper is to propose new models of geostatistical dependence tools by using copula functions. Indeed, the variogram, the correlogram and the

madogram of spatial variable are modeled via the copula underlying their joint distribution. Specifically, section 2 gives the background tools of stochastic analysis that turn to be necessary, while section 3 deals with our main results, copulas versions of variogram, madogram, covariogram and correlogram via copulas.

2. Preliminaries

This section summaries definitions and properties on the copulas of multivariate joint processes dependence which turn out to be necessary for our approach. For this purpose the definition of multivariate copula is necessary. Moreover, we provide a survey of the main geostatistical tools used in this paper.

2.1. Some geostatistical tools in spatial dependence

The covariance of a random field measures the strength of the relationship witch exists between the random variables witch represents it in the different observation sites. It is defined on $\mathbb{R}^d \times \mathbb{R}^d$ in $\mathbb{R}$, $for \ \ all \ (s_1, s_2) \in \mathbb{R}^d \times \mathbb{R}^d$,

$$c(s_1, s_2) = Cov[Z(s_1), Z(s_2)] = E[Z(s_1)Z(s_2)] - m(s_1)m(s_2).$$

Since,

$$E[Z(s_1)Z(s_2)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} z_1 z_2 h(s_1, s_2) dz_1 dz_2,$$

the covariance function can still be written such as:

$$c(s_1, s_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} z_1 z_2 h(s_1, s_2) dz_1 dz_2 - m(s_1)m(s_2)$$

where $m(s_1)$ is the mean of $Z(s_1)$ and $h(s_1, s_2)$, the joint density function of $Z(s_1)$ and $Z(s_2)$. The Cauchy-Schwarz inequality links the covariance between $Z(s_1)$ and $Z(s_2)$ to the variance of $Z(s_1)$ and $Z(s_2)$,

$$Cov[Z(s_1), Z(s_2)]| \leq \sqrt{Var[Z(s_1)]Var[Z(s_2)]}.$$

The madogram of a random field, especially used in the extreme case, determines the strength of the relationship between the random variables that represents it in the different observation sites. It is set to $\mathbb{R}^d$ in $\mathbb{R}_+$ by:

$$\forall h \in \mathbb{R}^d, \quad M(h) = \frac{E(|Z(s_1 + h) - Z(s_1)|)}{2} \ ; \ \forall s_1 \in \mathbb{R}^d.$$

2.2. A Survey of Copulas Functions

Copulas functions can be used to describe the dependence of variables or for spatial interpolation. The copula were introduced by Sklar [10] in order to characterize a vector $Z = (Z_1, \ldots, Z_n)$ having given marginal laws. According to Sklar's theorem, all functions

of continuous multivariate distributions F to marginal $F_1, \ldots, F_d$, there is a copula function C such that

$$F(s_1, \ldots, s_n) = C[F_1(s_1), \ldots, F_d(s_d)].$$

Definition 1. *An n-dimensional copula is a distribution function $C_n : [0,1]^n \longrightarrow [0,1]$ satisfying the following properties.*
   i) $C(u) = 0$ if one of the coordinates of u is zero, that is

$$C_n(u_1, \ldots, u_{i-1}, 0, u_{i+1}, \ldots, u_n) = 0; \ for, \ all \ (u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n) \in [0,1]^{n-1}.$$

   ii)
$$C_n(u_1, \ldots, u_{i-1}, 1, u_{i+1}, \ldots, u_n) = C_{n-1}(u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n),$$

, an (n-1) copula for all i.
   iii) The volume V_B of any rectangle $B = [a,b] \subseteq [0,1]^n$ is positive, that is,

$$V_B([a,b]) = \Delta_{a_n}^{b_n} \Delta_{a_{n-1}}^{b_{n-1}} \ldots \Delta_{a_1}^{b_1} C(u) = \int_B dC_n(u_1, \ldots, u_n) \geq 0. \qquad (6)$$

where,

$$\Delta_{a_k}^{b_k} C(u) = C(u_1, \ldots, u_{k-1}, b_k, u_{k+1}, \ldots, u_n) - C(u_1, \ldots, u_{k-1}, a_k, u_{k+1}, \ldots, u_n) \geq 0.$$

The use of copulas in stochastic analysis whas justified by the canonical parametrization of Sklar, see Joe [9]or Nelsen [12], such that the n-dimensional copula C associated to a random vector $(X_1, \ldots, X_n)$ with cumulative distribution F and with continuous marginal $F_1, \ldots, F_n$ is given, for $(u_1, \ldots, u_n) \in [0,1]^n$ by

$$C(u_1, \ldots, u_n) = F[F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)]. \qquad (7)$$

Differentiating the formula (7) shows that the density function of the copula is equal to the ratio of the joint density f of F to the product of marginal densities h_i such as, for all $(u_1, \ldots, u_n) \in [0,1]^n$,

$$c(u_1, \ldots, u_n) = \frac{\partial^n C(u_1, \ldots, u_n)}{\partial u_1 \ldots \partial u_n} = \frac{f\left[F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)\right]}{f_1\left[F_1^{-1}(u_1)\right] \times \ldots \times f_n\left[F_n^{-1}(u_n)\right]}. \qquad (8)$$

3. The Main Results of the Study

Let Z be a random field in n sites $Z = \{Z(s_1), \ldots, Z(s_n)\}$. Suppose $H(s_i, s_j)$ and $h(s_i, s_j)$ are the attached distribution and density functions of $Z(.)$ with marginal distributions $F_Z(s_i)$ at the site s_i.

3.1. Modeling the Madogram and the F-Madogram via Copulas

The following result provides a relation between the F-madogram and the underlying copula function.

Theorem 1. *Let C_F be the copula underlying a stochastic process $Z = \{Z(s_1), \ldots, Z(s_n)\}$. Then, the generalized F-madogram is such that:*

$$\gamma_F(h) = \int_0^1 u \, dC_{F,h}\left(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}\right) - \frac{1}{2}\left[C(\lambda) - D_h(\lambda)\right], \tag{9}$$

where

$$C(\lambda) = \frac{1}{2}\left(\frac{\sigma_{Z_s}^2}{1 + \lambda\sigma_{Z_s}^2}\right),$$

and

$$D_h(\lambda) = \frac{1}{2}\left(\frac{\sigma_{Z_{s+h}}^2}{1 + (1-\lambda)\sigma_{Z_{s+h}}^2}\right),$$

for $\lambda \in]0; 1[$, h being the average value of the separating distance between the two points.

Proof. Consider a bivariate distribution F, satisfying the key assumption. By noting that

$$|a - b| = 2\max(a, b) - (a + b),$$

and using this relation in (2), it follows that :

$$\gamma_F(h) = \frac{E\left(2\max\left([F(Z(s))]^\lambda, [F(Z(s+h))]^{1-\lambda}\right) - [F(Z(s))]^\lambda - [F(Z(s+h))]^{1-\lambda}\right)}{2}.$$

Then,

$$\begin{aligned}
\gamma_F(h) &= E\left(\max\left([F(Z(s))]^\lambda, [F(Z(s+h))]^{1-\lambda}]\right)\right) \\
&\quad - \frac{1}{2}\left\{E\left([F(Z(s))]^\lambda\right) - E\left([F(Z(s+h))]^{1-\lambda}\right)\right\}
\end{aligned} \tag{10}$$

Furthermore,

$$F_{Z_s, Z_{s+h}, \lambda}(u) = P\left(\max\left([F(Z(s))]^\lambda, [F(Z(s+h))]^{1-\lambda}\right) \leq u\right).$$

Thus,

$$F_{Z_s, Z_{s+h}, \lambda}(u) = P\left([F(Z(s))]^\lambda \leq u, [F(Z(s+h))]^{1-\lambda} \leq u\right),$$

It yields that

$$F_{Z_s, Z_{s+h}, \lambda}(u) = P\left(F(Z(s)) \leq u^{\frac{1}{\lambda}}, F(Z(s+h)) \leq u^{\frac{1}{1-\lambda}}\right),$$

Therefore

$$F_{Z_s, Z_{s+h}, \lambda}(u) = C_{F,h}\left(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}\right) \quad for \ all \ \lambda \in]0; 1[.$$

which is equivalent to,

$$E\left(\max\left([F(Z(s))]^{\lambda}, [F(Z(s+h))]^{1-\lambda}\right)\right) = \int_0^1 u dF_{Z_s, Z_{s+h}, \lambda}(u).$$

It follows that, for all $\lambda \in]0; 1[$,

$$E\left(\max\left([F(Z(s))]^{\lambda}, [F(Z(s+h))]^{1-\lambda}\right)\right) = \int_0^1 u dC_{F,h}\left(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}\right). \tag{11}$$

Furthermore, one have

$$E\left(\max\left([F(Z(s))]^{\lambda}\right)\right) = \frac{\sigma_{Z_s}^2}{1 + \lambda \sigma_{Z_s}^2}, \tag{12}$$

and

$$E\left(\max\left([F(Z(s+h))]^{1-\lambda}\right)\right) = \frac{\sigma_{Z_{s+h}}^2}{1 + (1-\lambda)\sigma_{Z_{s+h}}^2} \quad \forall \lambda \in]0; 1[. \tag{13}$$

Using (11), (12) and (13) in (10), it follows that

$$\gamma_F(h) = \int_0^1 u dC_{F,h}\left(u^{\frac{1}{\lambda}}, u^{\frac{1}{1-\lambda}}\right) - \frac{1}{2}\left[\frac{\sigma_{Z_s}^2}{1 + \lambda \sigma_{Z_s}^2} + \frac{\sigma_{Z_{s+h}}^2}{1 + (1-\lambda)\sigma_{Z_{s+h}}^2}\right]$$

which proves the relation (9) as disserted.

Let $Z = \{Z(x), x \in \mathbb{R}^d\}$ be a regular random field defined on $\mathbb{R}^d$. It is a well known that the **madogram** associated to the random field Z is the function M, mapping $\mathbb{R}^d$ to $\mathbb{R}^+$ such as:

$$\forall h \in \mathbb{R}^d, M(h) = \frac{E(|Z(x+h) - Z(x)|)}{2}, \quad x \in \mathbb{R}^d. \tag{14}$$

Proposition 1. *(Copula-based madogram)*
If $Z(.)$ is a stationary random field of order two then, the relation between its madogram and the copula function bivariate is given by:

$$M(h) = \int_0^1 F_Z^{-1}(u) dC_h(u, u) - \mu,$$

where $\mu = E(Z(x+h)) = E(Z(x))$, $C_h(.,.)$ being the jointed copula function which describes the dependence structure between two remote sites of h.

The following figure provide a representation of the joint copula of these two variables.
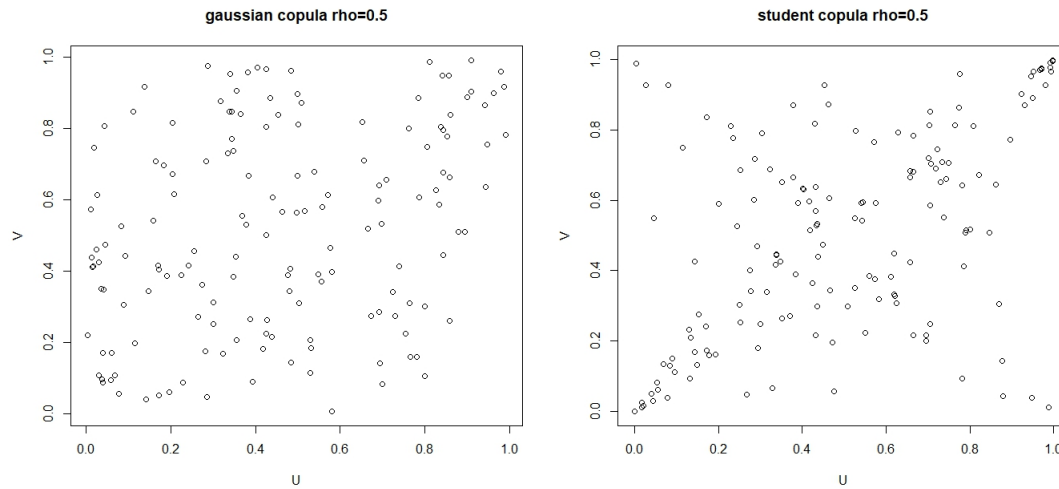
Figure 1: Graph of a joint distribution of copula

Proof. Recall that $|a - b| = 2 \max(a, b) - (a + b)$. Using this relation in (14), it's comes that *for all* $h, x \in \mathbb{R}^d$

$$M(h) = \frac{E\left(2 \max[Z(x + h), Z(x)] - Z(x + h) - Z(x)\right)}{2}. \tag{15}$$

Since $E(.)$ is a linear application, the relation (15) gives

$$M(h) = \frac{E\left(2 \max[Z(x + h), Z(x)]\right) - E\left(Z(x + h)\right) - E\left(Z(x)\right)}{2}.$$

Thus,

$$M(h) = E\left(\max[Z(x + h), Z(x)]\right) - \mu. \tag{16}$$

Furthermore,

$$P\left(\max[Z(x + h), Z(x)] \leq z\right) = P\left(Z(x + h) \leq z, Z(x) \leq z\right),$$

which is equivalent to:

$$P\left(\max[Z(x + h), Z(x)] \leq z\right) = C_h\left(P\left(Z(x + h) \leq z\right), P\left(Z(x) \leq z\right)\right)$$

Thus,

$$P\left(\max[Z(x + h), Z(x)] \leq z\right) = C_h\left(F_Z(z), F_Z(z)\right),$$

so,

$$E\left(\max[Z(x + h), Z(x)]\right) = \int_0^1 F_Z^{-1}(u) dC_h(u, u).$$

Substituting this expression in the equation (16) and taking into account that $z = F_Z^{-1}(u)$, we obtain the result:

$$M(h) = \int_0^1 F_Z^{-1}(u) dC_h(u, u) - \mu.$$

Which proves the assertion.

3.2. Modeling the covariogram via copulas

The variogram allows to measure the linear dependence between the random variables of a field. For two given sites, the associated variogram is given by

$$\vartheta(s_i, s_j) = Var(Z(s_i) - Z(s_j)) = Var(Z(s_i)) + Var(Z(s_j)) - 2\hat{c}(s_i, s_j). \qquad (17)$$

These tools do not take into account the extreme data observed in the different observation sites. However, the copula function makes it possible to model the extreme data and to detect any nonlinear link between different observation sites. So, it is necessary to express the variogram and the covariogram via the copula to allow the model to take into account the spatial structure even in case of extremes data. The model could also be able to detect the presence of some nonlinear dependence.

Theorem 2. *Let $Z = \{Z(s_1), \ldots, Z(s_n)\}$ be a stochastic process with variogram given by (14). Then, the covariogram $\hat{c}(s_i, s_j)$ and the copula function are linked by the relation.*

$$\vartheta(s_i, s_j) = \sigma_Z^2(s_i) + \sigma_Z^2(s_j) - 2\hat{c}(s_i, s_j) \qquad (18)$$

where

$$\hat{c}(s_i, s_j) = \int_0^1 \int_0^1 F_Z^{-1}(u) F_Z^{-1}(v) c(u, v) du dv - m_i m_j.$$

The quantity m_i being the mean of $Z(s_i)$; $c(u, v)$ the copula density function attached to $Z(s_i)$ and $Z(s_j)$.

  Proof. Let $u_i = F(Z(s_i)) = F_Z(s_i)$. It's follow that:

$$z_i = F_Z^{-1}(u_i) \Longrightarrow du_i = f_Z(z_i) dz_i,$$

where $f_Z(s_i)$ is the density function of the variable $Z(s_i)$.
So, it comes that

$$dz_i dz_j = \frac{1}{f_Z(z_i) f_Z(z_j)} du_i du_j.$$

  Similarly, according to Sklar's theorem [10], it follow that, for all couple of sites $(s_i, s_j) \in S^2$

$$H(s_i, s_j) = C(F_Z(s_i), F_Z(s_j)).$$

Moreover,

$$c(u, v) = \frac{f\left[H_1^{-1}(u), H_n^{-1}(u_n)\right]}{f_1\left[F_1^{-1}(u_1)\right] \times \ldots \times F_n\left[F_n^{-1}(u_n)\right]}. \qquad (19)$$

The relation between the joint density function $h(s_i, s_j)$ and the joint density function of the copula $c(u, v)$ is given by:

$$h(s_i, s_j) = c(u, v) f_Z(z_i) f_Z(z_j).$$

Using these expressions in the covariogram expression, it follows that

$$\hat{c}(s_i, s_j) = \int_0^1 \int_0^1 F_Z^{-1}(u) F_Z^{-1}(v) c(u, v) du dv - m_i m_j.$$

By using the last relation in the variogram expression, it follows that

$$\vartheta(s_i, s_j) = \sigma_Z^2(s_i) + \sigma_Z^2(s_j) - 2 \int_0^1 \int_0^1 F_Z^{-1}(u) F_Z^{-1}(v) c(u, v) du dv - 2 m_i m_j.$$

The following figure provide a representation of some theoretical variogram
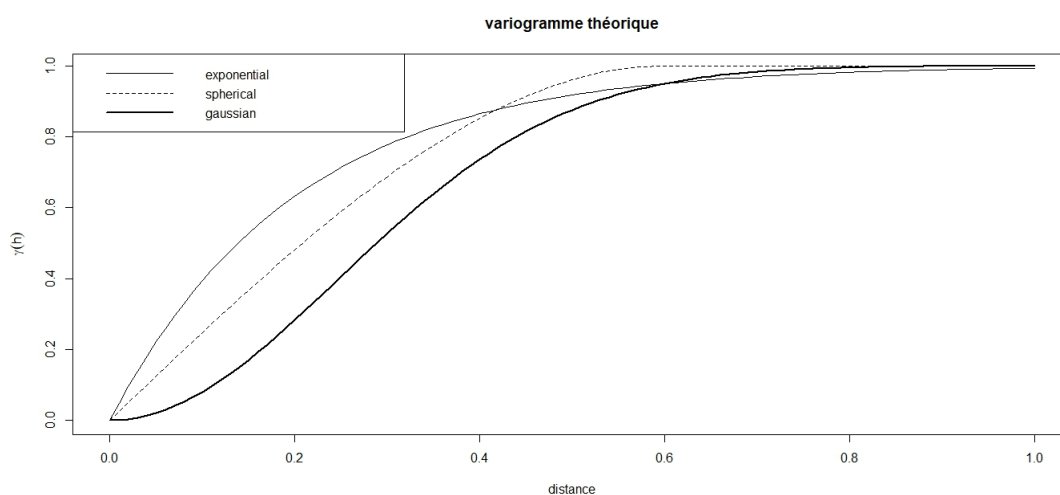

Figure 2: Graph of some theoretical variograms

3.3. Modeling the correlogram via copulas

The correlogram measures the spatial dependence between two sites s_i and s_j for all i and j. The following result gives a relation between the correlogram and the copula function.

Theorem 3. *Let $Z = \{Z(s_i), \ldots, Z(s_j)\}$ be a stochastic process on a geostatistical domain S. For two sites $s_i, s_j \in S$, the correlogram is given, via the associated copula by*

$$\rho(s_i, s_j) = \int_0^1 \int_0^1 \frac{F_Z^{-1}(u)}{\sigma_Z(s_i)} \frac{F_Z^{-1}(v)}{\sigma_Z(s_j)} c(u, v) du dv - \frac{m_i}{\sigma_Z(s_i)} \frac{m_j}{\sigma_Z(s_j)}; \tag{20}$$

where m_i denotes the mean of $Z(s_i)$, $c(u, v) = \frac{\partial^2 C(u,v)}{\partial u \partial v}$ being the density function of the copula C and $\sigma_Z(s_i)$ and the standard error of $Z(s_i)$ and $Z(s_j)$.

Proof. By definition, the correlogram of a random field in two observation sites is given by:

$$\rho(s_i, s_j) = \frac{\hat{c}(s_i, s_j)}{\sigma_Z(s_i)\sigma_Z(s_j)} = \frac{1}{\sigma_Z(s_i)\sigma_Z(s_j)} \times \hat{c}(s_i, s_j).$$

Using the result of precedent theorem (theorem 3), we get (20).

3.4. Stationnary framework for covariance modeling

In spatial context, the stationarity describes in a way, a form of spatial homogeneity of regionalization. From a mathematical point of view, stationarity hypothesies consists in assuming that the probabilistic properties of a set of values do not depend on the absolute position of the associated sites, but only on their separation.

Under the assumption of the second order stationarity of the random field $Z(.)$, the mean function deviates a constant and the covariance depends only on the distance separating the sites. So,

$$E(Z(s_i)) = \mu \quad \forall i = \overline{1,n} \;\; and \;\; \hat{c}(s_i, s_j) = \hat{c}(s_i - s_j) = \hat{c}(h_{ij}).$$

Previous relationships can be written differently. The following result gives a relationship between the covariogram and the copula function in second-order stationarity case.

Corollary 1. *In a second order stationarity framework the covariogram is given by*

$$\hat{c}(s_i, s_j) = \hat{c}(h_{ij}) = \int_0^1 \int_0^1 F_Z^{-1}(u) F_Z^{-1}(v) c_{h_{ij}}(u, v) du dv - \mu^2, \tag{21}$$

where $h_{ij} = |s_i - s_j|$ and $c_{h_{ij}}(u, v)$ the jointed copula density function of the two localized variables at two remote sites of h_{ij}.

Proof. Under the assumption of two order stationarity, the result of theorem (Theorem 3) gives the relation (21).

Similarly, under the assumption of second-order stationaity, the correlogram and the variogram are expressed as a function af the copula by the relation,

Proposition 2. *Assuming that $h_{ij} = s_i - s_j$ is the average distance and using the relation (21), then, the correlogram is such as*

$$\rho(h_{ij}) = \int_0^1 \int_0^1 \frac{F_Z^{-1}(u) F_Z^{-1}(v)}{\hat{c}(0_{\mathbb{R}^d})} c_{h_{ij}}(u, v) du dv - \frac{\mu^2}{\hat{c}(0_{\mathbb{R}^d})}. \tag{22}$$

and

$$\vartheta(h_{ij}) = 2 \left[\hat{c}(0_{\mathbb{R}^d}) + \mu^2 - \int_0^1 \int_0^1 F_Z^{-1}(u) F_Z^{-1}(v) c_{h_{ij}}(u, v) du dv \right]. \tag{23}$$

Proof. Under the hypothesis of two order stationary and using the relation (21), the result of theorem (Theorem 3) gives the result (22) and (23).

Proposition 3. *The relationship between the variogram and the covariogram is such than,*

$$\hat{c}(h_{ij}) = \hat{c}(0_{\mathbb{R}^d}) - \frac{1}{2}\vartheta(h_{ij}).$$

Proof. Let us consider the relations (21) and (23). It is known that:

$$\hat{c}(s_i, s_j) = \hat{c}(h_{ij}) = \int_0^1 \int_0^1 F_Z^{-1}(u)F_Z^{-1}(v)c_{h_{ij}}(u, v)dudv - \mu^2,$$

and

$$\vartheta(h_{ij}) = 2\left[\hat{c}(0_{\mathbb{R}^d}) + \mu^2 - \int_0^1 \int_0^1 F_Z^{-1}(u)F_Z^{-1}(v)c_{h_{ij}}(u, v)dudv\right].$$

Adding the two relations above, we get:

$$2\hat{c}(h_{ij}) + \vartheta(h_{ij}) = 2\hat{c}(0_{\mathbb{R}^d}).$$

So

$$\hat{c}(h_{ij}) = \hat{c}(0_{\mathbb{R}^d}) - \frac{1}{2}\vartheta(h_{ij}).$$

3.5. Covariogram in Stationnary intrinsic framework

Consider an intrinsic random field $Z(.)$ without drift, that is, the average of the increments is zero and the variance of the increments is the variogram. The intrinsic hypothesis is written:

$$E[Z(x + h) - Z(x)] = 0,$$

and

$$Var[Z(x + h) - Z(x)] = E\left\{[Z(x + h) - Z(x)]^2\right\} = 2\gamma(h). \qquad (24)$$

By integrating the intrinsic hypothesis, we obtain the following result.

Theorem 4. *Let $Z(.)$ An intrinsic random field without drift. It follows that the covariance is such that*

$$\hat{c}(s_i, s_j) = \hat{c}(h_{ij}) = \int_0^1 \int_0^1 F_Z^{-1}(u)F_Z^{-1}(v)c_{h_{ij}}(u, v)dudv - \mu^2, \qquad (25)$$

while the correlogram is:

$$\rho(h_{ij}) = \int_0^1 \int_0^1 \frac{F_Z^{-1}(u)F_Z^{-1}(v)}{\hat{c}(0_{\mathbb{R}^d})}c_{h_{ij}}(u, v)dudv - \frac{\mu^2}{\hat{c}(0_{\mathbb{R}^d})}.$$

And the variogram

$$\vartheta(h_{ij}) = 2\left[\hat{c}(0_{\mathbb{R}^d}) + \mu^2 - \int_0^1 \int_0^1 F_Z^{-1}(u)F_Z^{-1}(v)c_{h_{ij}}(u,v)dudv\right].$$

Where μ denote the mean, $\hat{c}(0_{\mathbb{R}^d})$ the variance and $c_{h_{ij}}(u,v)$ the density copula function.

Proof. By considering again the relation (24) we obtain:

$$\gamma(h) = \frac{1}{2}E\left\{[Z(x+h) - Z(x)]^2\right\}. \tag{26}$$

Now, it comes that

$$E\left\{[Z(x+h) - Z(x)]^2\right\} = E([Z(x+h)]^2) - 2E(Z(x+h)Z(x)) + E([Z(x)]^2).$$

So, by taking into account the density of the copula,

$$E\left\{[Z(x+h) - Z(x)]^2\right\} = \hat{c}(0_{\mathbb{R}^d}) + \mu^2 - 2\int_0^1 \int_0^1 F_Z^{-1}(u)F_Z^{-1}(v)c_{h_{ij}}(u,v)dudv + \hat{c}(0_{\mathbb{R}^d}) + \mu^2.$$

Furthermore, we have,

$$E\left\{[Z(x+h) - Z(x)]^2\right\} = 2\hat{c}(0_{\mathbb{R}^d}) + 2\mu^2 + -2\int_0^1 \int_0^1 F_Z^{-1}(u)F_Z^{-1}(v)c_{h_{ij}}(u,v)dudv.$$

Using this last relation in the equation (26), we get finaly,

$$\vartheta(h) = 2\left[\hat{c}(0_{\mathbb{R}^d}) + \mu^2 - \int_0^1 \int_0^1 F_Z^{-1}(u)F_Z^{-1}(v)c_{h_{ij}}(u,v)dudv\right]. \tag{27}$$

The variogram expression is similar in the intrinsic and second-order case, we deduce that the covariance and correlogram expressions remain unchanged.

Remark 1. *The variogram is related to the covariogram by the relation:*

$$\hat{c}(s_i, s_j) = \hat{c}(h_{ij}) = \int_0^1 \int_0^1 F_Z^{-1}(u)F_Z^{-1}(v)c_{h_{ij}}(u,v)dudv - \mu^2. \tag{28}$$

and

$$\vartheta(h_{ij}) = 2\left[\hat{c}(0_{\mathbb{R}^d}) + \mu^2 - \int_0^1 \int_0^1 F_Z^{-1}(u)F_Z^{-1}(v)c_{h_{ij}}(u,v)dudv\right]. \tag{29}$$

Adding the two relations above, we get finaly,

$$2\hat{c}(h_{ij}) + \vartheta(h_{ij}) = 2\hat{c}(0_{\mathbb{R}^d})$$

and

$$\hat{c}(h_{ij}) = \hat{c}(0_{\mathbb{R}^d}) - \frac{1}{2}\vartheta(h_{ij}).$$

3.6. Mains families of models of variograms

In this section we summary some of the most usefull models of variagrams in spatial modeling

a) Families with bearing or transition models

	Families or Modeles of variograms	Parameter
1	· Pure nugget effect model $\gamma(h_{ij}) = \begin{cases} 0 & \text{for } i = j \\ \hat{c} & \forall i \neq j \end{cases}$ · Meaning: reflects an absence of spatial structuring, due at the presence of an undetectable micro-structure experimentally.	
2	· Spherical model of parameter range a and sill $\hat{c}$ (valid in $\mathbb{R}^d$, $d \leqslant 3$) $\gamma(\parallel h_{ij} \parallel) = \begin{cases} \hat{c}(7\frac{\parallel h_{ij}\parallel^2}{a^2} - \frac{35}{4}\frac{\parallel h_{ij}\parallel^3}{a^3} + \frac{7}{2}\frac{\parallel h_{ij}\parallel^5}{a^5} - \frac{3}{4}\frac{\parallel h_{ij}\parallel^7}{a^7}) & \text{pour } 0 \leqslant \parallel h_{ij} \parallel \leqslant a \\ \hat{c} \ \text{pour} \parallel h_{ij} \parallel \geqslant a \end{cases}$ · Meaning: the presence of an undetectable micro-structure experimentally.	a
3	Gaussian model of parameter a and sill $\hat{c}$ $\gamma(\parallel h_{ij} \parallel) = \hat{c}(1 - exp(-\frac{\parallel h_{ij} \parallel^2}{a^2}))$ · Meaning: the sill is reached asymptotically and the practical range can be taken equal to $a\sqrt{3}$	$a\sqrt{3}$

b) Bessel and Polynomial models of variogram

Models includes mainly two models: the k modified model characterized

|		
1	· Bessel modified model $\gamma(\| h_{ij} \|) = \hat{c}\left\{1 - \dfrac{1}{2^{\alpha-1}\Gamma(\alpha)}(\dfrac{\| h_{ij} \|}{a})^{\alpha}K_{\alpha}(\dfrac{\|h_{ij}\|}{a})\right\}$ characterized by sill $\hat{c}$, scale factor a and parameter α, and Γ being the function of Euler interpolating the factorial $K(\alpha)$. $K_{\alpha}(u) = \dfrac{\pi\left(\Sigma_{k=0}^{\infty}\dfrac{1}{k!\Gamma(-\alpha+k+1)}(\dfrac{u}{2})^{2k-\alpha} - \Sigma_{k=0}^{\infty}\dfrac{1}{k!\Gamma(\alpha+k+1)}(\dfrac{u}{2})^{2k+\alpha}\right)}{2\sin(\alpha\pi)}$	*with* $\alpha > 0$
2	· The Bessel J model $\gamma(\| h_{ij} \|) = \hat{c}\left\{1 - (\dfrac{\| h_{ij} \|}{2a})^{-\alpha}\Gamma(\alpha+1)J_{\alpha}(\dfrac{\| h_{ij} \|}{a})\right\}$; characterized by sill $\hat{c}$, scale factor a and parameter α, and J_{α} being the Bessel function of the first kind of order α, $J_{\alpha}(u) = (\dfrac{u}{2})^{\alpha}\Sigma_{k=0}^{\infty}\dfrac{(-1)^{k}}{k!\Gamma(\alpha+k+1)}(\dfrac{u}{2})^{2k}..$	· $\alpha = -1/2$ for cosine model · $\alpha = 1/2$ for sinus model
3	· Polynomial Models with a scope and threshold $\hat{c}$ $\gamma(\| h_{ij} \|) = \begin{cases} \hat{c}\left(\dfrac{35}{12}\dfrac{\|h_{ij}\|}{a} - \dfrac{35}{8}\dfrac{\|h_{ij}\|^{3}}{a^{3}} + \dfrac{7}{2}\dfrac{\| h_{ij} \|^{5}}{a^{5}} - \dfrac{25}{24}\dfrac{\|h_{ij}\|^{7}}{a^{7}}\right) \text{ pour } 0 \leqslant \| h_{ij} \| \leqslant a \\ \hat{c} \qquad\qquad\qquad \text{pour } \| h_{ij} \| \geqslant a \end{cases}$	on $\mathbb{R}^{d}$ for $d \leqslant 3$
4	· An other Polynomial Models with a scope and threshold $\hat{c}$ $\gamma(\| h_{ij} \|) = \begin{cases} \hat{c}\left(\dfrac{5}{2}\dfrac{\| h_{ij} \|}{a} - \dfrac{5}{2}\dfrac{\| h_{ij} \|^{3}}{a^{3}} + \dfrac{\| h_{ij} \|^{5}}{a^{5}}\right) \text{ pour } 0 \leqslant \| h_{ij} \| \leqslant a \\ \hat{c} \qquad\qquad\qquad \text{pour } \| h_{ij} \| \geqslant a \end{cases}$	
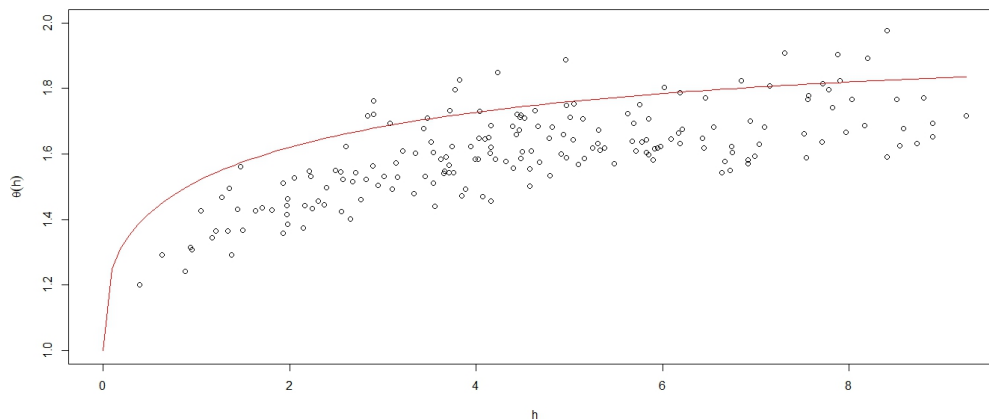
Figure 3: Graph of the extremal coefficient

4. Conclusion

The results of the study provides important characterizations of the variogram, the correlogram in a copula framework. Especially, they show on one hand that these tools are limited when data includes extremes values, in an other hand, that they have a copulawise extension, allowing the copula to model the data in the spatial context. Moreover, the study provides tools to analyze data and perform a comparative study with existing tools.

References

[1] Barro, D. (2009). Conditional Dependence of Trivariate Generalized Pareto Distributions, Asian Journal of Mathematics & Statistics Vol. 2 No.2, 20-32. DOI: ajms.2009.20.32

[2] Barro, D. (2012). Geostatistical Analysis with Conditional Extremal Copulas, International Journal of Statistics and Probability; Vol. 1 No.2, 2012. ISSN 1927-7032 E-ISSN 1927-7040

[3] Beirlant J., Goegebeur Y, Segers J. and Teugels,J. (2005). Statistics of Extremes: theory and application, Wiley, Chichester, England.

[4] Bondar, I., McLaughlin, K., & Israelsson, H. (2005). Improved Event Location Uncertainty Estimates, 27th Seismic Research Review, 299-307

[5] Bordossy, A. (2006). Copula based geostatistical models for groundwater quality parameters. Water Resources Research 42.

[6] Cooley D. Poncet P. and P. Naveau (2006). Variograms for max-stable random .elds. In Dependence in 8 Probability and Statistics. Lecture Notes in Statistics 187 373.390. Springer, New York

[7] Dossou-Gbete S., B., Somé and Barro, D. (2009). Modelling the Dependence of Parametric Bivariate Extreme Value Copulas, Asian Journal of Mathematics & Statistics, Vol 2, Issue3, 41-54 DOI: .3923/ajms.2009.41.54

[8] Kazianka, H.(2009). Spatial modeling and interpolation using copulas. PhD thesis, University of Klagenfurt. Ribatet (2011)- Statistical Modelling of Spatial Extremes A. C. Davison, S. A. Padoan and M. Ribatet October 3, 2011

[9] Nelsen, R.B. (1999). An Introduction to copulas. Lectures notes in Statistics 139, Springer-Verlag Joe H., 1997, Multivariate Models and Dependence Concepts. Monographs on Statistics and Applied Probabilty 73, Chapman and Hall, London. ISBN 978-0-412-7331

[10] Pickands, J. (1981). Multivariate extreme values distributions. in Bulletin of international Statistical Institute (Buenos Aires), Procedings of the 43 rd Session, pp 859-878.

[11] Sklar, A.(1973). Random variables, joint distribution functions, and copulas.Kybernetika, 9(6),449-460.9

[12] Schmitz, V. (2003). Copulas and Stochastic Processes, Aachen University, PhD dissertation

[13] M.Frechet Sur la loi de probabilitÃ© de l'Ã©cart maximum. Annales de la sociÃ©tÃ© Polonaise de MathÃ©matiques,6.(1927).93-116.

[14] D. Barro Contribution Ã la modÃ©lisation statistique des valeurs extrÃªmes multivariÃ©es. ThÃ¨se de l'UniversitÃ© de Ouagadougou. DÃ©cembre (2010)

[15] P.J.Diggle & P.J.Ribeiro Jr. Model-based Geostatistics. Springer Series in Statistics, Springer Science and Business Media, LLC, Springer. New York. (2007).

[16] C. Dombry ThÃ©orie spatiale des extrÃªmes et propriÃ©tÃ©s des processus max-stables. Document de synthÃ¨se en vue de l'habilitation Ã diriger des recherches, UniversitÃ© de Poitiers, UFR Sciences Fondamentales et AppliquÃ©es. Novembre (2012).

[17] R.A. Fisher & L.H.C. Tippett. Limiting Forms of the Frequency of the Largest or Smallest Member of a Sample. Proceedings of the Cambridge Philosophical Society, 24.(1928). 180-190.

[18] C. Fonseca, L. Pereira, H. Ferreira & A.P. Martins. Generalized Madogram and Pairwise Dependence of Maxima over two Regions of a Random Fiel. arXiv: 1104.2637v2 [math.ST] 24 Jan 2012. (2012).

[19] A.L. FougÃ¨res & P. Soulier. Limit Conditional Distributions for Bivariate Vectors with Polar Representation. Stochastic Models, 26(1). (2010). 54-77.

[20] M. FrÃ©chet. Sur la loi de probabilitÃ© de l'Ã©cart maximum. Annales de la sociÃ©tÃ© Polonaise de MathÃ©matiques, 6. (1927). 93-116. C. Gaet

[21] Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2005). *Statistics of Extremes: theory and application* -Wiley, Chichester, England.

[22] Coles, S. (2001). *An introduction to statistical modeling of extreme values*- Springer-Verlag (London), 2001.

[23] Degen M. (2006). *On Multivariate Generalised Pareto Distributions and High Risk Scenarios* - thesis, Department of Mathematics, ETH Zürich.

[24] D. Barro (2009) *Conditional Dependence of Trivariate Generalized Pareto Distributions.* Asian Journal of Mathematics & Statistics Year: 2009 | Volume: 2 | Issue: 2 | Page No.: 20-32. DOI: 10.3923/itj.2012.76.84

[25] Diakarya Barro, Moumouni Diallo, and Remi Guillaume Bagrã©, "*Spatial Tail Dependence and Survival Stability in a Class of Archimedean Copulas*," Inter. J. of Mathematics and Mathematical Sciences, vol. 2016, Article ID 8927248, 8 pages, 2016. doi:10.1155/2016/8927248

[26] Ferreira, H. and Ferreira,M. (2012) *Fragility Index of block tailed vectors*. ScienceDirect. ELsevier vol. 142 (7), 1837–1848 http ://dx.doi.org/10.1016/j.bbr.2011.03.031

[27] Fougã¨res, A., L (1996). *Estimation non paramã©trique de densitã©s unimodales et de fonctions de dã©pendance-* thã¨ses doctorat de l'Universitã© Paul Sabatier de Toulouse.

[28] Husler, J., Reiss, R.-D.(1989). *Extreme value theory - Proceedings of a conference held in Oberwolfach*, Dec. 6-12,1987. Springer, Berlin etc. Lenz, H.

[29] Joe, H. (1997). *Multivariate Models and Dependence Concepts - Monographs on Statistics and Applied Probabilty* 73, Chapman and Hall, London.

[30] Kotz, S., Nadarajah, S. (2000). Extreme Value Distributions, Theory and Applications - Imperial College Press - [48] S. Lang , Linear Algebra

[31] Michel, R.(2006). *Simulation and Estimation in Multivariate Generalized Pareto Distributions-* Dissertation, Fakultät für Mathematik, Universität Würzburg, Würzburg.

[32] Nelsen, R.B. (1999). *An Introduction to copulas-* Lectures notes in Statistics 139, Springer-Verlag

[33] Resnick, S.I. (1987). *Extreme Values, Regular Variation and Point Processes-* Springer-Verlag.

[34] Schmitz, V. (2003). *Copulas and Stochastic Processes*, Aachen University, PhD dissertation

[35] Tajvidi, N. (1996a)- *Confidence Intervals and Accuracy Estimation for Heavy-tailed Generalized Pareto Distribution-* Thesis article, Chalmers University of Technology. http://www.maths.lth.se/matstat/staff/~nader/.