https://doi.org/10.1007/s11053-016-9314-3

Natural Resources Research manuscript No.
(will be inserted by the editor)

Spatial pair-copula modelling of grade in ore bodies: a case study

G. Nishani Musafer · **M. Helen Thompson** ·
E. Kozan · **R. C. Wolff**

Received: date / Accepted: date

Abstract A real world mining application of pair-copulas is presented to model the spatial distribution of metal grade in an ore body. Inaccurate estimation of metal grade in an ore reserve can lead to failure of a mining project. Conventional kriged models are the most commonly used models for estimating grade, and other spatial variables. However, kriged models use the variogram or covariance function, which produces a single average value to represent the spatial dependence for a given distance. Kriged models also assume linear spatial dependence. In the application, spatial pair-copulas are used to appropriately model the non-linear spatial dependence present in the data. The spatial pair-copula model is adopted over other copula based spatial models since it is better able to capture complex spatial dependence structures. The performance of the pair-copula model is shown to be favourable compared to a conventional lognormal kriged model.

Keywords Geostatistical modelling · Pair-copula · Kriging · Non-linear spatial dependence · Mining

G. N. Musafer
School of Mathematical Sciences, Queensland University of Technology (QUT), Australia.
Cooperative Research Centre for Optimising Resource Extraction (CRC ORE), Australia.

M. H. Thompson
School of Mathematical Sciences, Queensland University of Technology (QUT), Australia.
Cooperative Research Centre for Optimising Resource Extraction (CRC ORE), Australia.
E-mail: helen.thompson@qut.edu.au

E. Kozan
School of Mathematical Sciences, Queensland University of Technology (QUT), Australia.
Cooperative Research Centre for Optimising Resource Extraction (CRC ORE), Australia.

R. C. Wolff
W.H. Bryan Mining and Geology Research Centre, University of Queensland (UQ), Australia.
Cooperative Research Centre for Optimising Resource Extraction (CRC ORE), Australia.

1 Introduction

This paper presents the first application of spatial pair-copulas to mining, with the purpose of illustrating the advantages of spatial pair-copula models over traditional kriged models in mining. This paper additionally aims to provide practitioners with a detailed guide to fitting spatial pair-copulas, which is lacking in the literature. This research forms part of a larger project with the Australian mining industry to reduce the uncertainty in estimates of economic risk in mining a potential ore reserve. The choice of the pair-copula model was motivated by the non-linear spatial dependence of multiple geological and geometallurgical variables apparent in the ore body and the need to estimate the variability in estimates of the spatial distribution of metal grade to facilitate a more informative risk evaluation. Accurate estimation of metal grade is one of the most important and influential factors for success in mining projects (Peattie and Dimitrakopoulos 2013).

Any method used to model a geological variable should be capable of accurately estimating the true spatial dependence (correlation). Spatial dependence describes the relationship between realisations of a geological variable sampled at different locations (Getis 2007). In reality, the *in-situ* spatial dependence structure may be non-linear, that is, it may vary over the distribution of the variable of interest (Journal and Alabert 2007). Although some kriged models, such as lognormal kriging and multi-Gaussain kriging, are able to model skewed geological variables, these models inherently assume linear spatial dependence through the use of the variogram (Diggle and Ribeiro 2007). Similarly, simulation methods that are based on conventional kriged models, such as conditional simulation (Khosrowshahi and Shaw 2001), also assume linear spatial dependence. The accuracy of local distributions from conditional simulations are also highly dependent on the number of simulations, and the method for finding the optimal number of simulations remains and open problem. Whilst multiple indicator kriging (MIK) is able to address spatial non-linearity, MIK can lead to higher estimates of recoverable material for higher cut-off grades due to the inconsistency of indicator models from one cut-off to the next as a result of the indicator variables being treated separately. This issue is known as the order relation problem (Vann and Guibal 2001). MIK also suffers from a loss in statistical power to detect the true relationship between variables due to binary transformation.

Bárdossy and Li (2008) introduced a copula based geostatistical model that uses bivariate copulas to model spatial dependence. Spatial copula models do not require a Gaussian assumption, are capable of modelling extreme measurements and also permit non-linear spatial dependence (Li 2010). However, most readily available copulas in the literature are unable to be extended to higher dimensions, which is required for spatial data, or do not provide good parameterisation for the dependence structure to appropriately reflect the spatial configuration of the data points (Bárdossy and Li 2008). Gaussian and Student t copulas fulfil both requirements but these copulas cannot be used to model asymmetric dependence structures. Whilst the non-central chi copula Bárdossy (2006) can model asymmetric dependence structures, this model is very computationally expensive for large scale data sets. For example, for n observations, 2^n calculations are needed to obtain estimates at unsampled locations. Additionally, the spatial copula model of Bárdossy and Li (2008) assumes the same copula family for each separation vector $\boldsymbol{h}$, and multivariate dependence, which is required in the interpolation process, is also modelled using the same family of higher dimensional copula. The spatial pair-copula model of Gräler and Pebesma (2011) not only possesses the desirable features of the Bárdossy and Li (2008) spatial copula model, but additionally permits the use of different types of copula families for different separating vectors and also for higher order dependencies. Thus, non-linear spatial dependence can be captured more accurately using a spatial pair-copula compared to more simple spatial copulas.

Although copula based modelling is a new avenue for geostatistics (Kazianka and Pilz 2010), it has been widely used in non-spatial applications in fields where it is essential to deal with non-linear dependence, such as in finance and actuarial sciences (Bárdossy 2006). In the literature, simple copula models have been used in only a few spatial applications, for example, to model hydrology properties (Bárdossy and Li 2008), soil properties (Marchant et al. 2011), air pollutants (Kazianka and Pilz 2011) and in mining (Musafer et al. 2013). The pair-copula model has been used in only a few spatial (Gräler and Pebesma 2011; Gräler 2014; Musafer and Thompson 2016*a*,*b*) and spatial-temporal (Erhardt et al. 2015*a*,*b*) applications. However, the pair-copula model has not yet been applied to mining applications.

The main objectives of this research are to fit a pair-copula model to estimate the metal grade of an ore reserve obtained from a real mine site, and to estimate the distribution of metal grade at unsampled locations, conditional on the local neighbourhood of sampled locations. Since the data are positively skewed, the pair-copula model is compared to a lognormal kriged model to facilitate comparison between a model that is, and a model that is not, able to capture non-linear spatial dependence.

This paper contains four sections. Section 2 describes copulas, pair-copulas and the pair-copula model for spatial data. In Section 3, the pair-copula model is applied to data on metal grade from a real mine and the corresponding results on model fit are given in the same section. Section 4 is devoted to conclusions driven by the results and discussion on the pair-copula model.

2 Method

This section provides an explanation of the statistical theories, utilised by Gräler and Pebesma (2011), that underpin the construction of geostatistical models based on pair-copulas. Instructions for the application of pair-copula models to spatial data, as summarised from Gräler and Pebesma (2011) and Gräler (2014), then follows.

2.1 Copula

Copula theory, which was introduced by Sklar (1959), forms the basis for any copula based spatial model. A copula describes the dependence structure between random variables. A copula does not need any information about the marginal distribution of the random variables to describe the dependence structure. Moreover, a copula can be defined as a multivariate distribution function of uniformly distributed random variables. Conversely, the copula can be constructed using the multivariate distribution function. An introduction to copula theory can be found in Nelsen (2006) and Trivedi and Zimmer (2007). For an applied review of copulas, the reader is referred to Boardman and Vann (2011).

2.2 Pair-copula

The pair-copula model can be classified as a hierarchical model building concept. Aas et al. (2009) initially introduced this method to estimate the joint multivariate distribution of random variables using a set of bivariate copulas based on the work of Joe (1996), Bedford and

Cooke (2002), and Kurowicka and Cooke (2006). Aas et al. (2009) present a worked example for the construction of a multivariate distribution for four random variables. To provide a simple demonstration of Aas et al.'s (2009) method, a small example for three variables is given below.

Let the joint density function of X_1, X_2, X_3 be $f_{123}(x_1, x_2, x_3)$. This can be factorised as

$$f_{123}(x_1,x_2,x_3) = f_3(x_3) f_{2|3}(x_2|x_3) f_{1|23}(x_1|x_2,x_3). \quad (1)$$

From Sklar's (1959) theorem, any multivariate distribution function F with marginals $F_1(x_1), \ldots, F_n(x_n)$ can be written as

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)),$$

where C is an n dimensional copula. Hence the joint density function can be written as

$$f(x_1, \ldots, x_n) = c_{1,2,\ldots,n}(F_1(x_1), \ldots, F_n(x_n)) \cdot f_1(x_1) \cdot \ldots \cdot f_n(x_n), \quad (2)$$

where $c_{1,2,\ldots,n}$ is the copula density.

Using Eq. (2), the second term of Eq. (1) can be written as

$$\begin{aligned} f_{2|3}(x_2|x_3) &= \frac{f(x_2,x_3)}{f(x_3)} \\ &= \frac{c_{23}(F_2(x_2),F_3(x_3)) \cdot f_2(x_2) \cdot f_3(x_3)}{f_3(x_3)} \\ &= c_{23}(F_2(x_2),F_3(x_3)) \cdot f_2(x_2). \end{aligned} \quad (3)$$

Again, using Eq. (2), the third term of Eq. (1) can be written as

$$\begin{aligned} & f_{1|23}(x_1|x_2,x_3) = \\ & c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \cdot c_{12}(F_1(x_1),F_2(x_2)) \cdot f_1(_1). \end{aligned} \quad (4)$$

Substituting Eqs. (3) and (4) into Eq. (1) gives

$$\begin{aligned} f_{123}(x_1,x_2,x_3) = & f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot c_{12}(F_1(x_1),F_2(x_2)) \\ & \cdot c_{23}(F_2(x_2),F_3(x_3)) \cdot c_{13|2}(F_{1|2}(x_1|x_2),F_{3|2}(x_3|x_2)). \end{aligned}$$

This equation states that the density of the three dimensional copula can be decomposed into a set of three bivariate copulas. The copulas $c_{12}(F_1(x_1),F_2(x_2))$ and $c_{23}(F_2(x_2),F_3(x_3))$ are unconditional bivariate copulas (unconditional pair-copulas) and $c_{13|2}(F_{1|2}(x_1|x_2),F_{3|2}(x_3|x_2))$ is a conditional bivariate copula (conditional pair-copula). Here, three pair-copulas have been used for the decomposition. In general, to decompose an n dimensional density function, $n(n-1)/2$ pair-copulas are required.

Marginal conditional distributions are required when constructing the conditional pair-copula. Joe (1996) showed that

$$F(x|\boldsymbol{v}) = \frac{\partial C_{x,v_j|\boldsymbol{v}_{-j}}(F(x|\boldsymbol{v}_{-j}),F(v_j|\boldsymbol{v}_{-j}))}{\partial F(v_j|\boldsymbol{v}_{-j})}, \quad (5)$$

where $\boldsymbol{v}$ is a d dimensional vector, v_j is one arbitrarily selected variable and $\boldsymbol{v}_{-j}$ denotes the vector $\boldsymbol{v}$ excluding v_j. If $\boldsymbol{v}$ is univariate, such that $\boldsymbol{v} = v$, then

$$F(x|v) = \frac{\partial C_{x,v}(F(x),F(v))}{\partial F(v)}.$$

However, this pair-copula decomposition is not unique. For example, there are 240 different constructions for a five dimensional density. Each decomposition approximates the full copula density differently (Aas et al. 2009). A graphical model, called a regular vine model, was developed by Bedford and Cooke (2002) to organise the large number of pair-copula constructions. Canonical vines and D-vines are special cases of regular vines. Canonical vines can be used if one can identify the key variable that governs the interaction of the data set. If dependence between variables needs to be treated in a specific order, then D-vines can be used.

Figures 1 and 2, which are reproduced from Aas et al. (2009), represent the graphical model used to illustrate the D-vine and a canonical vine for five variables, respectively. Each figure consists of four trees $T_j, j = 1,2,3,4$. Tree T_j has $6-j$ nodes and $5\,j$ edges. Each edge represents the corresponding pair-copula and the label of the edge represents the subscript of the pair-copula. Nodes in the figure are only used for determining the labels of the edges.

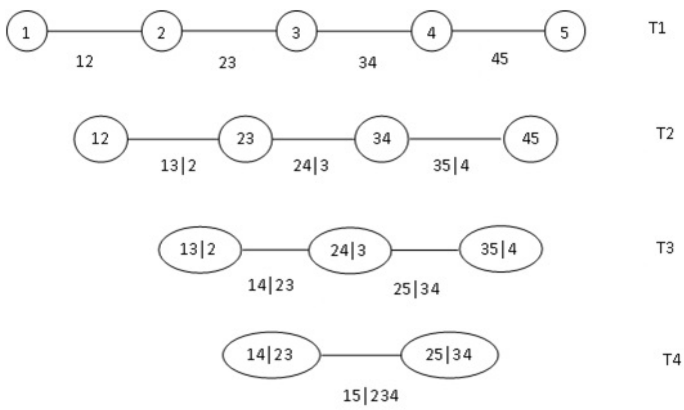


Fig. 1: D-vine for five variables.

By using the decompositions shown in Figure 1, the joint density function of five random variables can be approximated using a D-vine as follows (Aas et al. 2009).

$$
\begin{aligned}
&f_{12345}(x_1,x_2,x_3,x_4,x_5) = \\
&\quad f_1(x_1)\cdot f_2(x_2)\cdot f_3(x_3)\cdot f_4(x_4)\cdot f_5(x_5)\cdot \\
&\quad c_{12}(F_1(x_1),F_2(x_2))\cdot c_{23}(F_2(x_2),F_3(x_3))\cdot c_{34}(F_3(x_3),F_4(x_4))\cdot \\
&\quad c_{45}(F_4(x_4),F_5(x_5))\cdot c_{13|2}(F_{1|2}(x_1|x_2),F_{3|2}(x_3|x_2))\cdot \\
&\quad c_{24|3}(F_{2|3}(x_2|x_3),F_{4|3}(x_4|x_3))\cdot c_{35|4}(F_{3|4}(x_3|x_4),F_{5|4}(x_5|x_4))\cdot \\
&\quad c_{14|23}(F_{1|23}(x_1|x_2,x_3),F_{4|23}(x_4|x_2,x_3))\cdot \\
&\quad c_{25|34}(F_{2|34}(x_2|x_3,x_4),F_{5|34}(x_5|x_3,x_4))\cdot \\
&\quad c_{15|234}(F_{1|234}(x_1|x_2,x_3,x_4),F_{5|234}(x_5|x_2,x_3,x_4)).
\end{aligned}
$$

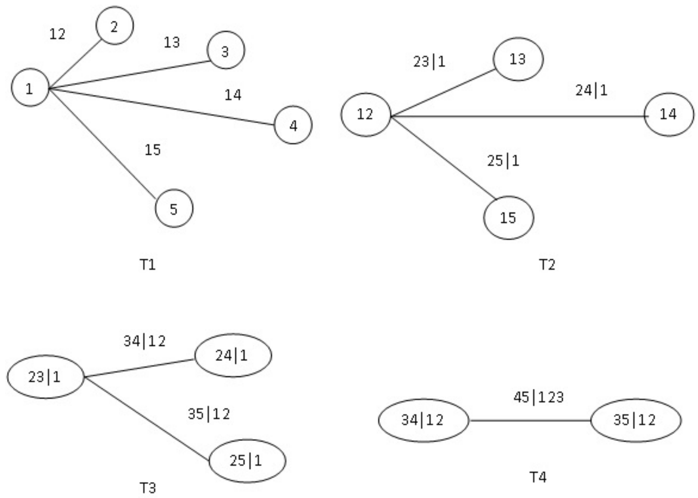


Fig. 2: Canonical vine for five variables.

From Figure 2, approximation of the joint density function for five random variables can be written using a canonical vine as follows (Aas et al. 2009).

$$
\begin{aligned}
&f_{12345}(x_1,x_2,x_3,x_4,x_5) = \\
&\quad f_1(x_1)\cdot f_2(x_2)\cdot f_3(x_3)\cdot f_4(x_4)\cdot f_5(x_5)\cdot \\
&\quad c_{12}(F_1(x_1),F_2(x_2))\cdot c_{13}(F_1(x_1),F_3(x_3))\cdot c_{14}(F_1(x_1),F_4(x_4))\cdot \\
&\quad c_{15}(F_1(x_1),F_5(x_5))\cdot c_{23|1}(F_{2|1}(x_2|x_1),F_{3|1}(x_3|x_1))\cdot \\
&\quad c_{24|1}(F_{2|1}(x_2|x_1),F_{4|1}(x_4|x_1))\cdot c_{25|1}(F_{2|1}(x_2|x_1),F_{5|1}(x_5|x_1))\cdot \\
&\quad c_{34|12}(F_{3|12}(x_3|x_1,x_2),F_{4|12}(x_4|x_1,x_2))\cdot \\
&\quad c_{35|12}(F_{3|12}(x_3|x_1,x_2),F_{5|12}(x_5|x_1,x_2))\cdot \\
&\quad c_{45|123}(F_{4|123}(x_4|x_1,x_2,x_3),F_{5|123}(x_5|x_1,x_2,x_3)).
\end{aligned}
$$

2.3 Pair-copula construction for spatial data

Gräler and Pebesma (2011) introduced pair-copula construction to the spatial framework. Spatial pair-copulas allow modelling of complex spatial dependence in a fully flexible way. A canonical vine structure is used to construct a pair-copula for spatial data, since this structure benefits spatial interpolation by giving higher priority to the interaction between the unobserved locations and nearby locations, if unobserved locations are selected as the root element.

2.3.1 Assumptions of copula based geostatistical models

As with conventional geostatistical models, copula based models assume that the set of measured values of the variable of interest are realisations of a random field (Bárdossy and Li 2008). However, when fitting copula based models, a stationary random field (see the definition in Gaetan and Guyon (2010)) is assumed over the domain of interest. This assumption is stronger than the conventional linear geostatistical assumption of a second-order stationary random field over the domain of interest because the copula based model requires that all the moments of the data generating process be unaffected by a change of spatial distance. However, copula based modelling has more advantages when compared to conventional geostatistical modelling, even though it requires a more limiting assumption, such as the ability to obtain the full conditional distribution, ability to remove the influences of marginal distributions when modelling the dependence structure and the ability to model non-linear spatial dependence (Haslauer et al. 2010). Based on this strong stationarity assumption, the marginal distributions of the variable of interest for each location in the domain are identical, that is, $F_i(z_i) = F(z_i)$. The empirical bivariate copula can be used to explore the spatial variability. As with the variogram, it is assumed that the bivariate spatial copula C_s at any two locations only depends on the separation vector $\boldsymbol{h}$ and is independent of the locations x (Bárdossy 2006; Bárdossy and Li 2008), that is,

$$\begin{aligned} C_s(\boldsymbol{h},u,v) &= Pr\left(F(Z(x)) \leq u, F(Z(x+\boldsymbol{h})) \leq v\right) \\ &= C\left(F(Z(x)), F(Z(x+\boldsymbol{h}))\right). \end{aligned}$$

All of the above mentioned assumptions are also applicable to pair-copula modelling of spatial data. To simplify application of the pair-copula model, spatial dependence is restricted to the isotropic case here. In isotropic situations, spatial dependence is assumed to vary only with distance and not with direction. In this case, the vector $\boldsymbol{h}$ is simply distance h.

2.3.2 Spatial interpolation using pair-copulas

The steps for carrying out spatial interpolation using pair-copulas, based on Gräler and Pebesma (2011) and Gräler (2014), is described as follows.

Step 1: Empirical bivariate copula densities.

Since the marginal univariate distributions of the variable of interest for each location are identical (based on the stationarity assumption), the empirical marginal distribution function $F(z)$ can be estimated using all the observations $z(x_1),\dots,z(x_N)$, where N is the total number of sample locations. A unit interval transformation is then applied to the observations using the estimated distribution function.

Distances between every pair $x_i - x_j = h; i \neq j, \forall i,j = 1,2,\dots,N$ are then calculated and, thereafter, each pair $\{F(z(x_i)), F(z(x_j))\}$ is placed into a relevant distance class from the following classes $[0,h_1),[h_1,h_2),\dots,[h_{l-1},h_l)$, where h_l is the maximum distance at which significant dependence is observed. The mean distance is taken as the representative value for each class.

The empirical bivariate copula densities can be calculated using kernel density smoothing if the number of pairs per distance class is large enough, otherwise the empirical bivariate copula can be calculated by defining a regular grid on the unit square and calculating the cumulative frequency of values for each grid. The next step is to fit the theoretical copula model to the empirical copula densities. This is similar to fitting a theoretical model to the experimental variogram.

Step 2: Theoretical bivariate copula densities and spatial copula construction.

Maximum likelihood can be used to estimate the bivariate copula densities. In the spatial setting, several copula families must be estimated for each distance class in order to fit the most suitable spatial copula. For example, if there are 10 distance classes and nine copula families are to be compared for each distance class, altogether, ninety bivariate copulas need to be estimated in the first step of pair-copula construction. This may be computationally demanding and time consuming. It is simpler and faster to calculate the inverse of Kendall's tau (or Spearman's rho) for a distance class and convert this value to an estimate of the dependence parameter using the functional relationship between Kendall's tau and the dependence parameter of the copula family (Genest and Rivest 1993). Following this, the copula that produces the maximum likelihood, amongst the copulas for a given distance class, is selected as the spatial copula for the corresponding class and is assigned to the mean distance of the distance class. The set of selected spatial copulas is then used to obtain distance dependent convex combinations of copulas as follows.

$$C_h(u_1,u_2) = \begin{cases} \lambda_1 \cdot M(u_1,u_2) + (1-\lambda_1)\cdot C_{1,h}(u_1,u_2), & 0 \le h < h_1 \\ \vdots & \\ \lambda_i \cdot C_{i-1,h}(u_1,u_2) + (1-\lambda_i)\cdot C_{i,h}(u_1,u_2), & h_{i-1} \le h < h_i \\ \vdots & \\ \lambda_k \cdot C_{k-1,h}(u_1,u_2) + (1-\lambda_k)\cdot L(u_1,u_2), & h_{k-1} \le h < h_k \\ L(u_1,u_2), & h_k \le h < h_l \end{cases} \tag{6}$$

where $h_1,\ldots,h_l$ are the boundaries of the distance classes, h_l is the maximum distance at which significant dependence is observed, $L(u_1,u_2) = u_1 \cdot u_2$ (independence for far away locations), $M(u_1,u_2) = \min(u_1,u_2)$ (perfect dependence for very close locations), $\lambda_i = \frac{h_i - h}{h_i - h_{i-1}}$, and u_1 and u_2 are the calculated cumulative values for the two locations of interest. This convex combination ensures consistency between distance classes.

Step 3: Pair-copula construction and spatial interpolation.

Copula based methodology permits estimation of the full conditional distribution of $Z(x)$:

$$F(x,z) = Pr(Z(x) \le z | Z(x_1) = z_1, \ldots, Z(x_N) = z_N),$$

where N is the total number of observations. The full conditional distribution can be written using the corresponding conditional copula $C_{x,N}$:

$$F(x,z) = C_{x,N}(F(z)|u_1 = F(z_1), \ldots, u_N = F(z_N)).$$

It may be computationally intensive to use all N observations in calculating the full conditional distribution due to the large number of conditional pair-copulas that must be semi-parametrically fitted to the data. However, the full conditional distribution can be approximated based on a sufficient number of local neighbouring locations (Bárdossy and Li 2008). The number of locations used in the approximation is determined by randomly selecting a few locations, and estimating and plotting the density functions for different numbers of nearby locations. The smallest number of nearby locations n that produces nearly identical density functions for almost all considered locations is used for the approximation. The approximate full conditional distribution is given by

$$F(x,z) = C_{x,n}(F(z)|u_1 = F(z_1),\ldots,u_n = F(z_n)),$$

where $F(z_i) = F(z(x_i))$ for $i = 1,\ldots,n$ and the points x_i are observations in the neighbourhood of x.

The conditional density function can be derived as

$$\begin{aligned} f(z|z_1,\ldots,z_n) &= \frac{\partial F(x,z)}{\partial z} \\ &= \frac{\partial C(F(z)|u_1 = F(z_1),\ldots,u_n = F(z_n))}{\partial z} \\ &= \frac{\partial C(u|u_1 = F(z_1),\ldots,u_n = F(z_n))}{\partial u} \cdot \frac{\partial F(z)}{\partial z}, \end{aligned}$$

that is,

$$f(z|z_1,\ldots,z_n) = c(u|u_1 = F(z_1),\ldots,u_n = F(z_n)) \cdot f(z),$$

where $f(z)$ is the marginal density and $F(z)$ is its distribution function.

The procedure for constructing the conditional copula density $c(u|u_1 = F(z_1),\ldots,u_n = F(z_n))$ using a pair-copula construction is described using an example as follows.

Let the number of nearby locations be four. Figure 3, which is reproduced from Gräler and Pebesma (2011), depicts the pair-copula decomposition, based on a canonical vine structure, for obtaining the full five dimensional pair-copula density. In Figure 3, edges represent the bivariate copula and the two nodes connected to each edge represent the two arguments of a corresponding bivariate copula. The unobserved location is x_0, and x_1, x_2, x_3 and x_4 are nearby locations.

The estimation process of the copulas in the first tree, T_1, has already been discussed in step 2. By using these copulas, the marginal conditional distributions $F_{i|0}, i = 1,2,3,4$, can be calculated using Eq. (5). The conditional pair-copula in the second tree, T_2, can then be estimated. The same procedure is repeated to estimate the conditional copulas in other trees.

These conditional copulas are influenced not only by their conditional distribution function arguments but also by the value of the conditioning variable. For example, $c_{12|0}$ is influenced by its arguments $\big(F_{1|0}(z(x_1)|z(x_0)), F_{2|0}(z(x_2)|z(x_0))\big)$ and the value of $Z(x_0)$. However, in pair-copula construction, estimation of a conditional pair-copula is simplified by ignoring the influence from the value of the conditioning variable to keep the construction process more practicable (Haff et al. 2009). Haff et al. (2009) showed that, even though this simplified version has some limitations, it is a good approximation for the actual model.

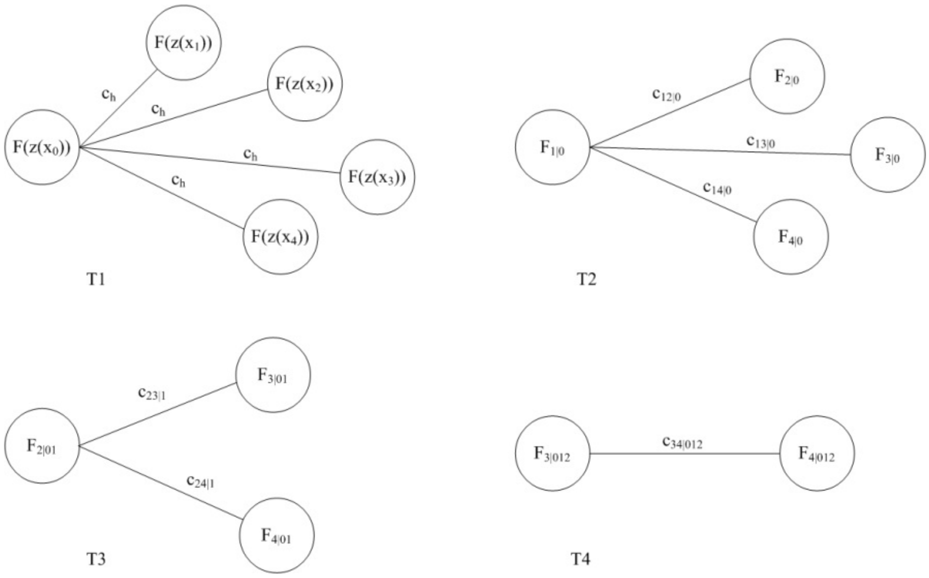


Fig. 3: Five dimensional spatial vine.

Finally, using the decomposition shown in Figure 3, the full five dimensional copula density can be written as

$$\begin{aligned}
&c(u_0,u_1,\dots,u_4) = \\
&c_h(F(z(x_0)),F(z(x_1)))\cdot c_h(F(z(x_0)),F(z(x_2)))\cdot c_h(F(z(x_0)),F(z(x_3)))\cdot \\
&c_h(F(z(x_0)),F(z(x_4)))\cdot c_{12|0}(F_{1|0}(z(x_1)|z(x_0)),F_{2|0}(z(x_2)|z(x_0)))\cdot \\
&c_{13|0}(F_{1|0}(z(x_1)|z(x_0)),F_{3|0}(z(x_3)|z(x_0)))\cdot \\
&c_{14|0}(F_{1|0}(z(x_1)|z(x_0)),F_{4|0}(z(x_4)|z(x_0)))\cdot \\
&c_{23|01}(F_{2|01}(z(x_2)|z(x_0),z(x_1)),F_{3|01}(z(x_3)|z(x_0),z(x_1)))\cdot \\
&c_{24|01}(F_{2|01}(z(x_2)|z(x_0),z(x_1)),F_{4|01}(z(x_4)|z(x_0),z(x_1)))\cdot \\
&c_{34|012}(F_{3|012}(z(x_3)|z(x_0),z(x_1),z(x_2)),F_{4|012}(z(x_4)|z(x_0),z(x_1),z(x_2))).
\end{aligned}$$

The conditional copula density of the variable of interest at the unsampled location can then be obtained as follows

$$c(u_0|u_1,\dots,u_4) = \frac{c(u_0,u_1,\dots,u_4)}{\int_0^1 c(v,u_1,\dots,u_4)dv}.$$

Finally, point estimates at unobserved location x_0 can be obtained. The mean and median are (Bárdossy and Li 2008)

$$\begin{aligned}
\hat{Z}_{mean}(x_0) &= \int_0^1 F^{-1}(u)c(u|u_1,\dots,u_n)du, \\
\hat{Z}_{median}(x_0) &= F_\alpha^{-1}(u = C^{-1}(0.5|u_1,\dots,u_n)).
\end{aligned}$$

Since the pair-copula method provides the full conditional distribution at an unsampled location, it is easy to obtain a more complete estimation of uncertainty, such as confidence intervals, when compared to the kriged model. Here complete is used to emphasise that the copula based model is fully capable of producing uncertainty estimation dependent on both the observations configuration and values. This feature is important for additional drilling campaigns, where a reduction in uncertainty is expected based on the influence of additional measurements (e.g., Musafer and Thompson (2016*a*,*b*)).

3 Case Study

Confidential data on one particular metal from a real mine site are available, in which there are nearly 80,000 measurements from over 2,000 drill holes. A small scale example is presented here based on a random subset of the spatial observations. The subset of 2,086 measurements of metal grade $z(x_i)$ at three dimensional locations $x_i = (x_{1i}, x_{2i}, x_{3i})$, $i = 1, \ldots, 2086$ are displayed in Figure 4. R software (R Core Team 2016) and the R package 'spcopula' (Gräler and Appel 2015) were used to fit the pair-copula models.

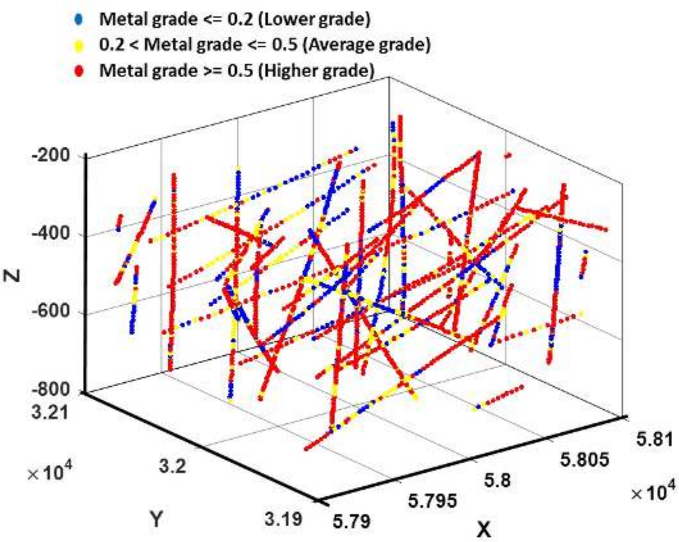


Fig. 4: Three dimensional spatial plot of metal grade.

Summary statistics for metal grade are given in Table 1. Non-parametric edge-weighted kernel density estimation was used to estimate the marginal distribution of metal grade with higher weights given to measurements that are close to zero. A histogram of the metal grades can be seen in Figure 5, from which positive skewness is apparent. The curve is the fitted weighted kernel density. Using the estimated marginal distribution, observed measurements were then transformed to the unit interval in order to construct the empirical copula densities to explore the spatial dependence structure.

Five metre by five metre classes were constructed. Selecting this width for the classes ensures high flexibility in the pair-copula model. Additionally, for this class width, each class contains more than 100 pairs.

Statistic	Value
n	2086
Mean	0.815
Standard deviation	0.723
Coefficient of variation	0.886
Min	0.005
First quartile Q1	0.243
Median	0.675
Third quartile Q3	1.158
Max	4.961

Table 1: Summary statistics for metal grade.

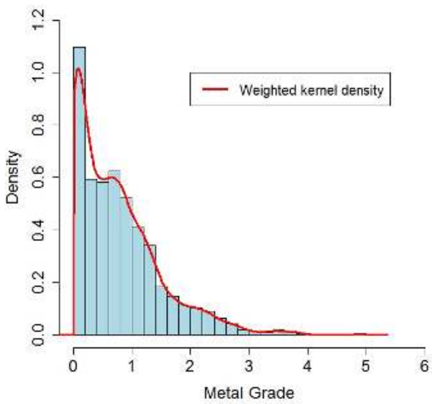


Fig. 5: Histogram of metal grade.

Figure 6 is a plot of the Kendall tau values against the mean of the distance classes. From this plot, spatial correlation of metal grade is estimated to decrease to zero for locations separated by more than 65 metres. Autocorrelation for the mean of each distance class was estimated using a polynomial fit to the Kendall tau values.

Figure 7 shows the empirical copula densities obtained for four of the 20 distance classes. If the spatial dependence is linear, then the empirical copula density plots should demonstrate a similar structure to that shown in Figure 8, which is a plot of a Gaussian copula density. Even though the distance class $[0,5)$ metres appears to have a linear spatial structure (Figure 7(a)), the other distance classes have more complex spatial structures. The empirical plot in Figure 7(d) confirms spatial independence between locations that are more than 65 metres apart.

Inversion of Kendall's tau was used to estimate the dependence parameter for a spatial copula. The copula with the highest log-likelihood value, amongst the Gaussian, Student t, Frank, Clayton, Gumbel, Joe, and survival version of the latter three, copulas was fitted to each distance class. Table 2 gives the best fitting spatial copula for each distance class.

The anisotropy of the data was evaluated in several directions. The variograms show fairly similar dependence structures for all directions. Hence, isotropic spatial dependence was assumed.

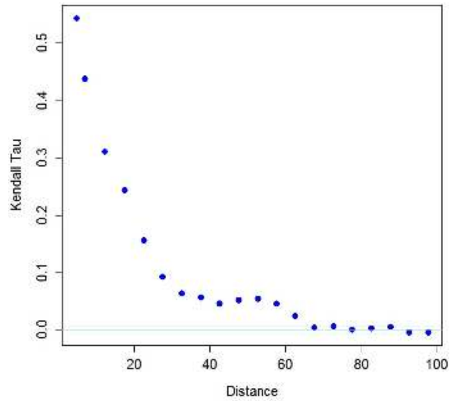


Fig. 6: Kendall tau values against the mean of the distance classes.

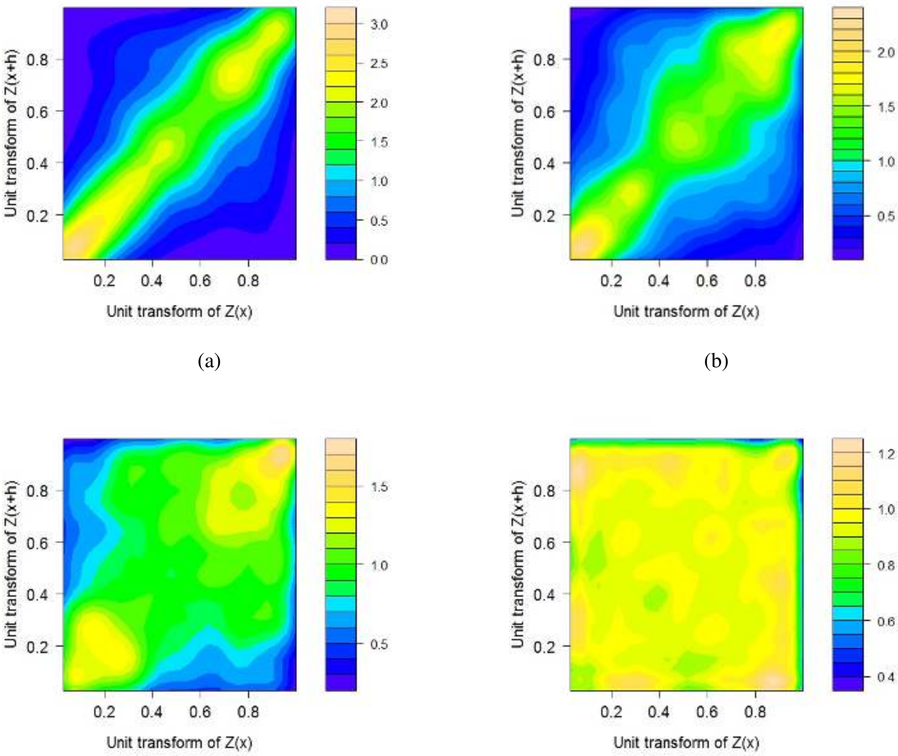
Class	Copula	Dependence Parameter	Degrees of freedom
0-5	Student *t*	0.709	4
5-10	Student *t*	0.646	4
10-15	Gumbel	1.504	-
15-20	Gumbel	1.327	-
20-25	Frank	1.622	-
25-30	Frank	1.109	-
30-35	Gumbel	1.088	-
35-40	Survival Gumbel	1.055	-
40-45	Survival Gumbel	1.036	-
45-50	Survival Gumbel	1.022	-
50-55	Survival Gumbel	1.017	-
55-60	Survival Gumbel	1.017	-
60-65	Independent	-	-

Table 2: Best fit copulas for each distance class.

Cross-validation was carried out to compare the performance of the pair-copula model with lognormal kriging. Figure 9 shows the experimental variogram that was used for lognormal kriging, where the exponential model was used to model spatial dependence. The estimated nugget, sill and range of the exponential model are 0.301, 2.570 and 60.450, respectively. The same distance classes as the pair-copula model were used in constructing the variogram model. Leave-one-out cross-validation was used, with 20 nearby locations in the interpolation process. Two estimators, the mean and median, were estimated from the pair-copula model.

The performance of the models was evaluated by calculating the accumulated error between observed and estimated values for all sampling points using two criteria: mean absolute error (MAE) and mean squared error (MSE). MAE and MSE are used to assess bias in prediction and model accuracy, respectively. Table 3 summarises these statistics.

Both the mean and median estimators from the pair-copula model performed well in terms of bias and accuracy of predictions in this application. From Table 3, the pair-copula models have smaller MAEs and MSEs than the lognormal kriged model with the mean



(a) (b) (c) (d)

Fig. 7: Empirical copula density of metal grade for (a) 0-5 m, (b) 10-15 m (c) 20-25 m and (d) 65-70 m distance classes.

Margin	Approach	MAE	MSE
Weighted kernel density	Pair-copula - mean	0.418	0.364
	Pair-copula - median	0.409	0.368
	Lognormal kriging	0.466	0.426

Table 3: Results of cross-validation.

estimator from the pair-copula model having the smallest MSE and the median estimator having the smallest MAE.

Figure 10, shows bias against the true metal grade. For all models, the bias of individual observations are generally larger as metal grade increases. However, both estimators from the pair-copula model appear to show less bias than the estimator from the kriged model over the distribution of metal grade.

The quantile plot in Figure 11 indicates that the pair-copula model reproduces the distribution of the data better than lognormal kriging.

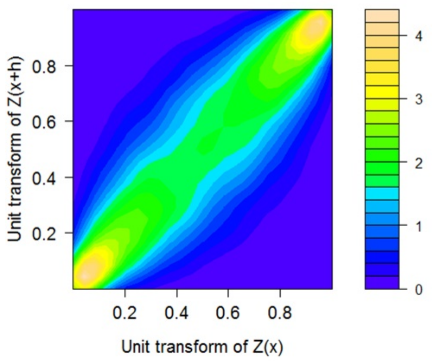


Fig. 8: Gaussian copula density.

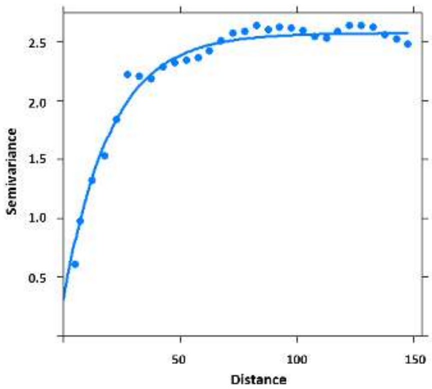


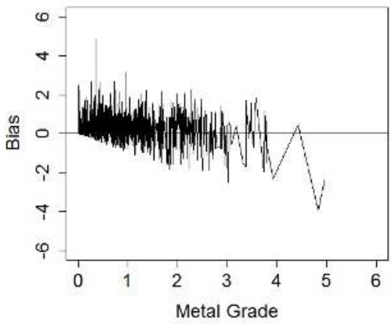
Fig. 9: Empirical variogram overlaid with fitted exponential model.

4 Discussion and Conclusion

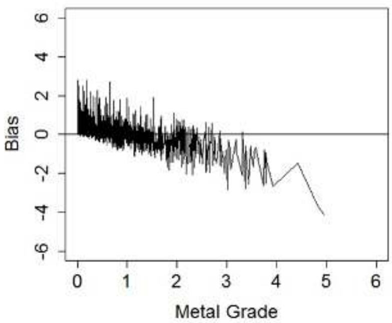
A spatial pair-copula model was fitted to model the distribution of metal grade from the ore body of a real mine site. Unlike conventional kriging, the pair-copula model is able to take account of non-linear spatial dependence and is, generally, more accurate than other, less flexible, copula based spatial models. The full conditional distribution of metal grade is available from the pair-copula model. Here, the mean and median estimators of metal grade were obtained. In the application, the pair-copula model outperformed lognormal kriging in terms of bias and accuracy of predictions.

It should be noted that, in mining applications, the mean estimator is expected to perform well because it has the ability to produce unbiased estimates for total metal content. This was shown to be the case for the mean estimator of the pair-copula model in the application.

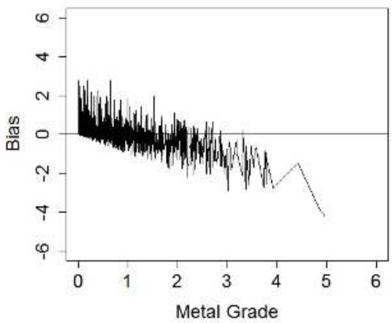
Figure 10 indicates the existence of conditional bias (lower values are overestimated and higher values are underestimated) in both the kriged and pair-copula models. The main



(a)



(b)



(c)

Fig. 10: Bias against true metal grade for (a) lognormal kriging, and (b) mean and (c) median estimate from pair-copula model..

reason for the conditional bias in kriging and indicator kriging is the smoothing effect of the variance of the estimator. Conditional bias arising from smoothing is well-documented and understood in the literature (Seo 2013; Mclennan and Deutsch 2004). Although the smoothing effect does not directly apply to the pair-copula model, this model uses several approximations and numerical integrations throughout the estimation process. It can be conjectured that this might be the reason for the existence of conditional bias in the estimators of the pair-copula model.

The pair-copula model has the potential to become a popular geostatistical model because of the ability to remove the influences of marginal distributions when modelling the dependence structure and the ability to model non-linear spatial dependence and tail dependence. As a result, the copula based model is fully capable of producing uncertainty estimation dependent on both the observations configuration and values. Hence more complete uncertainty estimation can be used to obtain more precise optimal designs than optimal designs obtained using a kriged model for additional drillings.

A major disadvantage of spatial pair-copula modelling is the rapid increase in computational time required to fit conditional distributions at unsampled locations with an increase

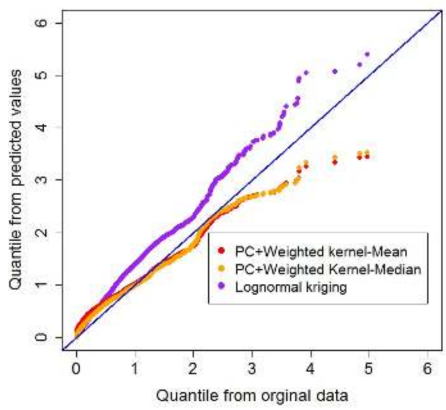


Fig. 11: Quantiles of the data against quantiles of the predicted values for the different estimators.

in the number of nearby locations. As the number of nearby locations increases, the number of conditional bivariate copulas in the pair-copula model that requires estimation increases rapidly. Consequently, a limited number of nearby locations are used in fitting pair-copula models with the assumption that the full conditional distribution is reasonably approximated, which may be difficult to verify. Additionally, an invalid multivariate distribution may be fitted by a pair-copula model when different types of copulas and different parameters are used to fit the conditional bivariate copulas.

When fitting the marginal distribution, the pair-copula model assumes observations are independent of each other. Hence, an inappropriate marginal distribution may be fitted to the data in situations where, for example, observations are clustered because they come from the same borehole or sampling has been carried out in areas where high grades are expected.

In the application, an isotropic dependence structure was assumed. Anisotropy should be evaluated in different directions. Inspection of variograms for different directions is insufficient to evaluate anisotropy when fitting a spatial pair-copula model. Instead, the empirical copula density of each distance class for different directions should be compared. This will be addressed in future research.

Further improvements in the pair-copula model are expected to be gained through, for example, development of an efficient method for defining the lag distance classes, use of advanced search strategies, e.g., quadrant search, to remove the obvious cluster effects, and use of more families of copulas. These improvements are the focus of current research.

Acknowledgements This research was funded by the Australian Government's Cooperative Research Centre for Optimising Resource Extraction grant P3C-030. The authors thank the reviewers for their comments and guidance, which greatly improved discussion and practical aspects of the application.

References

Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2), 182–198.

Bárdossy, A. (2006). Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research*, 42(11), W11416.

Bárdossy, A. and Li, J. (2008). Geostatistical interpolation using copulas. *Water Resources Research*, 44(7), W07412.

Bedford, T. and Cooke, R. M. (2002). Vines: a new graphical model for dependent random variables. *The Annals of Statistics*, 30(4), 1031–1068.

Boardman, R. C. and Vann, J. E. (2011). A review of the application of copulas to improve modelling of non-bigaussian bivariate relationships (with an example using geological data). In F. Chan, D. Marinova and R. S. Anderssen (Eds.), *Proceedings of the 19th International Congress on Modelling and Simulation (MODSIM2011)*. Modelling and Simulation Society of Australia and New Zealand (MSSANZ). Perth, Australia. pp. 627–633. http://www.mssanz.org.au/modsim2011/A9/boardman.pdf

Diggle, P. J. and Ribeiro, P. J. (2007). Classical parameter estimation. In *Model-based Geostatistics*. New York: Springer. pp. 99–133.

Erhardt, T. M., Czado, C. and Schepsmeier, U. (2015*a*). R-vine models for spatial time series with an application to daily mean temperature. *Biometrics*, 71(2), 323–332.

Erhardt, T. M., Czado, C. and Schepsmeier, U. (2015*b*). Spatial composite likelihood inference using local C-vines. *Journal of Multivariate Analysis*, 138(C), 74–88.

Gaetan, C. and Guyon, X. (2010). Second-order spatial models and geostatistics. In *Spatial Statistics and Modeling*. New York: Springer. pp. 1–52.

Genest, C. and Rivest, L. (1993). Statistical inference procedures for bivariate Archimedian copulas. *Journal of the American Statistical Association*, 88(423), 1034–1043.

Getis, A. (2007). Reflections on spatial autocorrelation. *Regional Science and Urban Economics*, 37(4), 491–496.

Gräler, B. (2014). Modelling skewed spatial random fields through the spatial vine copula. *Spatial Statistics*, 10, 87–102.

Gräler, B. and Appel, M. (2015). 'spcopula'. http://r-forge.r-project.org/projects/spcopula/

Gräler, B. and Pebesma, E. (2011). The pair-copula construction for spatial data: a new approach to model spatial dependency. *Procedia Environmental Sciences*, 7, 206–211.

Haff, I. H., Aas, K. and Frigessi, A. (2009). On the simplified pair-copula construction – simply useful or too simplistic?. *Journal of Multivariate Analysis*, 101(5), 1296–1310.

Haslauer, C. P., Li, J. and Bárdossy, A. (2010). Application of copulas in geostatistics. In P. M. Atkinson and C. D. Lloyd (Eds.), *geoENV VII – Geostatistics for Environmental Applications*. Netherlands: Springer. pp. 395–404.

Joe, H. (1996). Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependencee parameters. In *Distributions with Fixed Marginal and Related Topics*. Vol. 28 of *Lecture Notes-Monograph Series*. Institute of Mathematical Statistics. pp. 120–141.

Journal, A. G. and Alabert, F. (2007). Non-Gaussian data expansion in the Earth Sciences. *Terra Nova*, 1(2), 123–134.

Kazianka, H. and Pilz, J. (2010). Copula-based geostatistical modeling of continuous and discrete data including covariates. *Stochastic Environmental Research and Risk Assessment*, 24(5), 661–673.

Kazianka, H. and Pilz, J. (2011). Bayesian spatial modeling and interpolation using copulas. *Computers & Geosciences*, 37(3), 310–319.

Khosrowshahi, S. and Shaw, W. (2001). Conditional simulation for resource characterization and grade control. In A. C. Edwards (Ed.), *Mineral Resource and Ore Reserve Estimation - The AusIMM Guide to Good Practice*. Australasian Institute of Mining and Metallurgy (AusIMM). pp. 285–292.

Kurowicka, D. and Cooke, R. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley.

Li, J. (2010). Application of copulas as a new geostatistical tool. PhD dissertation. Institute for Water and Environmental System Modeling. University of Stuttgart, Germany. http://dx.doi.org/10.18419/opus-332

Marchant, B. P., Saby, N. P. A., Jolivet, C. C., Arrouays, D. and Lark, R. M. (2011). Spatial prediction of soil properties with copulas. *Geoderma*, 162(3–4), 327–334.

Mclennan, J. A. and Deutsch, C. V. (2004). Conditional non-bias of geostatistical simulation for estimation of recoverable reserves. *CIM Bulletin*, 97(1080), 68–72.

Musafer, G. N. and Thompson, M. H. (2016*a*). Non-linear optimal multivariate spatial design using pair-copulas. *Stochastic Environmental Research and Risk Assessment*, pp. 1–20. doi:10.1007/s00477-016-1307-6.

Musafer, G. N. and Thompson, M. H. (2016*b*). Optimal adaptive sequential spatial sampling of soil using pair-copulas. *Geoderma*, 271, 124–133.

Musafer, G. N., Thompson, M. H., Kozan, E. and Wolff, R. C. (2013). Copula-based spatial modelling of geometallurgical variables. In S. Dominy (Ed.), *Proceedings of The Second AusIMM International Geometallurgy Conference (GeoMet2013)*. Australasian Institute of Mining and Metallurgy (AusIMM). Brisbane, Australia. pp. 239–246.

Nelsen, R. B. (2006). *An Introduction to Copulas*. New York: Springer.

Peattie, R. and Dimitrakopoulos, R. (2013). Forecasting recoverable ore reserves and their uncertainty at Morila Gold Deposit, Mali: an efficient simulation approach and future grade control drilling. *Mathematical Geosciences*, 45(8), 1005–1020.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.r-project.org/

Seo, D. J. (2013). Conditional bias-penalized kriging (CBPK). *Stochastic Environmental Research and Risk Assessment*, 27(1), 43–58.

Sklar, A. (1959). Fonctions de répartition á n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris*, 8, 229–231.

Trivedi, P. K. and Zimmer, D. M. (2007). *Copula Modeling: An Introductoin for Practitioners*. Boston: Now.

Vann, J. and Guibal, D. (2001). Beyond ordinary kriging - an overview of non-linear estimation. In A. C. Edwards (Ed.), *Mineral Resource and Ore Reserve Estimation - The AusIMM Guide to Good Practice*. Australasian Institute of Mining and Metallurgy (AusIMM). pp. 249—256.