Estimation of direction of increase of gold mineralisation using pair-copulas

Emmanuel Addo [a]**, Emmanuel K. Chanda** [a]**, Andrew V. Metcalfe** [b]

[a] *School of Civil, Environmental and Mining Engineering, University of Adelaide*

[b] *School of Mathematical Sciences, University of Adelaide*

Email: emmanuel.addo@adelaide.edu.au

Abstract: The case study is based on a gold deposit in western Ghana. One thousand five hundred surface soils samples are available on an area of 7810 hectares. The distribution of the grade appears to be a realization of a non-stationary anisotropic random process. The objectives of the analysis are to model the gold grade and to extrapolate into the near surrounding area to determine regions of highest gold value for a future drilling campaign. In the analysis we compare possible transformations of the data to reduce the influence of outliers, and in the case of copulas achieving a marginal uniform distribution. The anisotropy of the gold grades is described with empirical copula density plots for each distance class and for two orthogonal (135° and 45°) directions. The non-stationarity is modelled by regression methods including periodic variation if appropriate. The residuals from the regression are modelled with spatial pair copulas. An investigation of the possible benefits of increasing the number of nearby locations modeled in the spatial pair copula construction is utilized. Predictions are done for unknown sample locations to the North-East (NE) and South-West (SW) of the main field. The different approaches of increasing the nearby points are compared in terms of minimum, maximum and average predicted grade at all unknown sampling locations outside the main field.

Keywords: *Copula, geostatistical modelling, pair-copula, kriging*

1. INTRODUCTION

In many geostatistical applications the objective is to predict a spatial variable at points, for which there are no direct measurements, from a neighboring set of measurements. This requires modelling of spatial dependence between points. Regression methods can be used to model spatial trends, and the errors in the regression model can then be taken to represent a stationary field variable. The regression model can be fitted to the field data and the residuals are treated as data from a stationary field. Kriging is the standard method for estimation of grade at unknown points. Kriging depends on the variogram which represents expected values of the squared differences in grade at two points as a function of distance between them. Kriging is optimum if the distribution of the grade is multivariate Gaussian (MVG). However, gold grades are not always well

Figure 1. Schematic diagram showing how a weighted correlogram is defined. P_0P_1 is the direction of maximum range

modelled as MVG, even after transformation, as extreme outliers are common. Indicator kriging, which models grade as above or below some threshold, is a modification that is less sensitive to outliers (Journal and Alabert, 1990 and Goovaerts et al. 2005). The indicator approach is highly empirical and is not based on any stochastic model, so it requires large data sets for adequate precisions. In particular, indicators are fitted for each threshold separately which can lead to problems with monotonicity of the estimates.

Copulas are multivariate uniform distributions, provide a versatile model for the spatial distribution of grade variables, including multivariate Gaussian as special case, and offer an alternative to indicator kriging. Examples of spatial applications of copulas include: Bárdossy and Li (2008) used spatial copulas to model groundwater parameters, Gräler (2014) used pair-copula to model particulate matter concentrations (PM_{10}); and Musafer et al. (2016) applied copulas to mining applications.

This paper describes the modelling of anisotropic surface soil sample gold grades using spatial pair-copulas in order to extrapolate into the near surrounding area to determine regions of highest gold value for a future drilling campaign. The anisotropy of the gold grades is described with empirical copula contour plots for a direction of maximum range and the orthogonal direction. The non-stationarity is modelled by

Table 1. Summary statistics of the surface soil samples grades

Statistics	Value
Number of Samples	1500
Minimum Value (ppm)	0.05
First Quartile (ppm)	1
Median (ppm)	8
Mean (ppm)	20.81
Standard Deviation (ppm)	114.66
Kurtosis	891.82
Skewness	27.19
Third Quartile (ppm)	14
Maximum Value (ppm)	3906
MAE (ppm)	23.91

regression and the residuals from the regression are modelled with spatial pair-copulas. The spatial pair-copula model is used to predict gold grade at 20 exterior locations (NE corner and SW corner).

2. METHODS

2.1. Theory of copulas

A copula is a multivariate uniform distribution with all margins uniform over the interval ($U[0,1]$). This is very versatile as any marginal probability distribution can be transformed to $U[0,1]$. Specifically, if Z is a random variable with cumulative distribution (cdf) $F(\ )$ then $F(Z) \sim U[0,1]$. Two examples of bivariate copula cdf are the Gaussian, which is equivalent to fitting a bivariate Gaussian distribution after transforming margins to normality, and Archimedean copulas which have cdf $C(u_1, u_2, .., u_p) = \psi\left(\psi^{-1}(u_1) + \psi^{-1}(u_2) + \cdots + \psi^{-1}(u_p)\right)$, where $\psi: [0,1] \to [0, \infty]$ is a continuous, strictly decreasing generator function with $\psi(1) = 0$. Sklar's Theorem (1959) states that any multivariate distribution can be expressed as a copula, and an expansion of a copula with any marginal distribution, which can all be different, is a valid multivariate distribution. A key result for pair-copula construction is that the probability density function (pdf) of a bivariate distribution $f(z_1, z_2)$ can be expressed as $f(z_1, z_2) = c(z_1, z_2)f(z_1)f(z_2)$ where $c(,)$ is the copula pdf. Most bivariate copula have a single parameter, which can be expressed as some function of Kendall's tau (Musafer et al., 2016), and which controls the dependence.
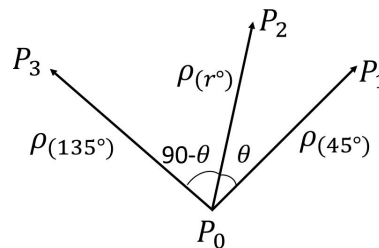
2.2 Pair copulas

In this application we compare the results obtained with the multivariate distribution of grade at four, six and eleven points. The canonical vine pair-copula construction is a factorization of the multivariate pdf into bivariate copulas (Aas et al. 2009). For three variables the factorization can be expressed as shown in Eq. (1)

$$f_{123}(z_1, z_2, z_3) = f_1(z_1).f_2(z_2).f_3(z_3).c_{12}\big(F_1(z_1), F_2(z_2)\big).$$

$$c_{13}\big(F_2(z_1), F_3(z_3)\big).c_{23|1}\Big(F_{2|1}(z_2|z_1), F_{3|1}(z_3|z_1)\Big) \tag{1}$$

For four variables the factorisation can be expressed as shown in Eq. (2)

$$f_{1234}(z_1, z_3, z_3, z_4) = f_1(z_1).f_2(z_2).f_3(z_3).f_4(z_4).c_{12}\big(F_1(z_1), F_2(z_2)\big)$$

$$.c_{13}\big(F_1(z_1), F_3(z_3)\big).c_{14}\big(F_1(z_1), F_4(z_4)\big).c_{23|1}\Big(F_{2|1}(z_2|z_1), F_{3|1}(z_3|z_1)\Big)$$

$$.c_{24|1}\Big(F_{2|1}(z_2|z_1), F_{4|1}(z_4|z_1)\Big).c_{34|12}\Big(F_{3|12}(z_3|z_1, z_2), F_{4|12}(z_4|z_1, z_2)\Big) \tag{2}$$

The pattern continues. The steps for fitting and estimation of the bivariate copulas follow.

Step 1: Empirical Bivariate Copula Contour or Densities for Orthogonal Directions

The initial step of copula based geostatistical modelling is to estimate the marginal univariate distribution function of the variable of interest, (Z). Under the stationarity assumptions the distribution for every location of the variable are identical. $F(z)$ can be calculated from all the observations $(z_1, \ldots \ldots, z_N)$ where N is the total number of samples in the domain. The estimated distribution function is used to transform all the observations to a unit interval $[0, 1]$.

Figure 2. Location of points with soil samples (•) and exterior points (▪)

Distances between every pair of observations $z_i(x_i, y_i)$ and $z_j(x_j, y_j); i \neq j, \forall i, j = 1, 2, \ldots, N$ are calculated as $h = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$. Each paired datum, $\big\{F(z(x_i)), F\big(z(x_j)\big)\big\}$, is calculated and placed in its respective distance class $[(0, h_1), (h_1, h_2), \ldots \ldots, (h_{l-1}, h_l)]$, where h_l is define as the distance beyond which there is no significant dependence between the observations. The coordinates for each pair $\big\{F(z(x_i)), F\big(z(x_j)\big)\big\}$, that is in a two-dimensional (2D) case; $pt1 = (x_i, y_i) \ and \ pt2 = (x_j, y_j)$ are rotated using the rotation matrix in Eq. (3). Given a defined tolerance angle (ε) and an angle (θ), all pairs of points that lies within an angle $\pm\varepsilon$ are selected as the new pairs of points for that directional angle (θ). This step is repeated for all pairs of data in all distance class. The mean distance between pairs of each distance class is computed and that represents the distance for any particular class interval. The principal axis is taken as the direction with the longest range.
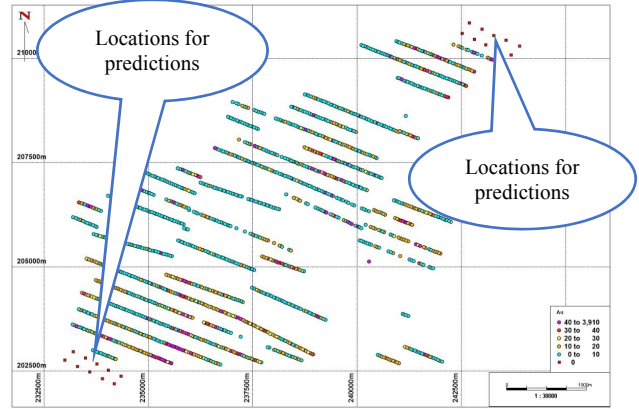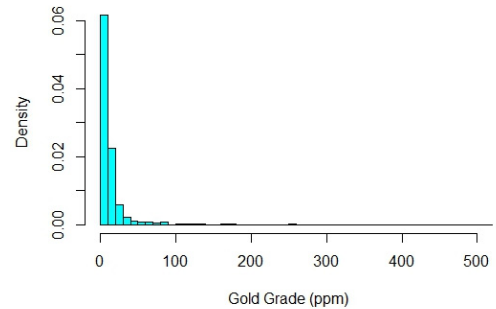

Figure 3. Histogram of the surface soil samples grades

$$\begin{pmatrix} x_i' \\ y_i' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix} \tag{3}$$

same rotation is used to transform (x_j, y_j) to (x_j', y_j'). Points are considered to lie in the direction class θ if

$$slope = \left| \frac{(y_j' - y_i')}{(x_j' - x_i')} \right| < \varepsilon \tag{4}$$

The weighted correlogram, Figure 1 is defined as

$$\rho_{(\theta)} = \rho_{(135°)} \left(\frac{\theta}{90} \right) + \rho_{(45°)} \left(\frac{90 - \theta}{90} \right) \tag{5}$$

The number of pairs for each distance class is chosen to be large enough for the kernel smoothing density function (KernSmooth) in R to be used to generate contour plots.

Step 2: Theoretical Bivariate Copula Densities and Spatial Copula Construction

For each distance class we have a number of grade pairs. The grades are transformed to uniform distribution, using the empirical cdf of all grades, and various types of copula are fitted by maximum likelihood. The copula with the highest value of the likelihood is chosen. Furthermore, Kendall's tau, which is a non-parametric measure of correlation based on ranks, is also calculated for each distance class. All selected spatial copulas for all distance class are used to obtain the distance-direction-dependent convex combinations of copula given in Eq. (6). The copula $C._{,h}$ used in the convex combination in this instance depends on the distance and direction. This can be described by the distance-direction dependent copula $C_h(u, v)$. This spatial copula must approach the Fréchet–Hoeffding bound Nelsen (1999) $M(u,v) = min(u,v)$ (which shows perfect positive dependence) when the distance approaches zero and the product copula $\prod(u,v) = uv$ (explains independence) when the distance approaches the range where the data are spatially correlated. Instead of restricting the spatial copula to a single family, the convex combination of copulas for several distances-directions are used. That is $h_1, \ldots, h_l$, M denotes zero separation and Π denotes the maximum range h_l. Then the spatial copula is given with $\lambda_i := (h_i - h)/(h_i - h_{i-1})$ by

Figure 4. Histogram of residuals from regression model and superimposed Gaussian pdf.

$$C_h(u,v) = \begin{cases} \lambda_1 M(u,v) + (1 - \lambda_1) C_{1,h}(u,v), & 0 \le h \le h_1 \\ \lambda_i C_{i-1,h}(u,v) + (1 - \lambda_i) C_{i,h}(u,v), & h_{i-1} \le h \le h_i \\ \lambda_l C_{l-1,h}(u,v) + (1 - \lambda_l) C_{l,h}(u,v), & h_{l-1} \le h \le h_l \end{cases} \tag{6}$$

Step 3: Pair-Copula Construction and Spatial Interpolation

Finally, the random variable Z at unknown location x_0 also follows the distribution $F(z|z_1, \ldots, z_k)$ which is conditioned on the known values of grade at the nearest neighbours $x_1, \ldots \ldots, x_k$. The full multivariate distribution of grade $z_1, \ldots, z_k$ is given by

$$F(x, z_1, \ldots, z_k) = P(Z < z, Z_1 < z_1 \ldots, Z_k < z_k) \tag{7}$$

where k is the total number of observations. The conditional distribution is then expressed in terms of the conditional pair-copula $C_{x,k}$

$$F(z|z_1, \ldots z_k) = c(u|u_1, \ldots u_k) f(z) \tag{8}$$

Point estimates of grade at unknown locations x_0 can then be obtained by calculating the mean or the median (Bardossy and Li 2008)
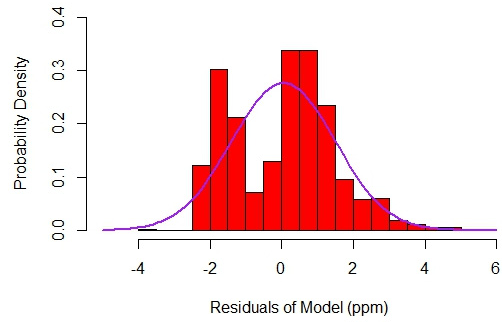
$$\hat{Z}_{mean}(x_0) = \int_0^1 F^{-1}(u)c(u|u_1, \ldots \ldots, u_n)du \tag{9}$$

$$\hat{Z}_{median}(x_0) = F_{\propto}^{-1}\left(u = C^{-1}(0.5|u_1, \ldots., u_n)\right) \tag{10}$$

The pair-copula modeling approach provides the full conditional distribution of the realisations at unknown locations, this full distribution can be used to generate prediction intervals.

3. APPLICATION

3.1. Overview of Project Area

Surface soil samples data from an operating mine in western Ghana have been used to demonstrate the copula based approach for geostatistical modelling. A total of 1500 surface soil samples (2-dimensional data) were taken over an area of 7810 hectares. The spatial correlation of the gold grade in surface soil samples is relatively low due to the mineralization style and the distances between the samples. In the local grid system of the mine, the project extends from an easting of 243258.6E to 233183.8E and a northing of 210425.9N to 202674.5N Figure 2 shows a 2D plot of the project area and locations of the surface soil samples. Also shown are the locations at which the grade is to be predicted. All the surface soil grades are in parts per million (ppm). All statistical analyses and the fitting of spatial pair-copula models were conducted using R (R Core Team 2016).

3.2. Summary Statistics and Fitting/Predicting from Quadratic Surface

Summary statistics of the surface soil sample grades are given in Table 1. The histogram of the grades in Figure 3 shows extremely high positive skewness. The non-stationarity of the surface soil grades was modelled by fitting a quadratic surface by regression. The natural logarithm (w) of the grade value was regressed on the x (X coordinate), y (Y coordinate), x^2, y^2 and the cross product xy. Equation 11 defines the relationship between the logarithm of grade and the coordinates. Where ε is the random error, which is expected to be spatially correlated with mean 0. The spatial correlation of the errors is estimated as a first step in fitting copulas or kriging (section 3.3). It does not affect the estimation of the coefficients but it does affect their standard errors. The estimated standard deviation of the errors is 1.421 on 1494 degrees of freedom, which is slightly smaller than the standard deviation of the logarithm of grade (1.480). The coefficient of multiple determination (R^2) is 0.082.

Figure 5. Contour of the fitted regression surface

$$w = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy + \varepsilon \tag{11}$$

Whilst this is quite low it is highly statistically significant with a sample of this size, even after allowing for the spatial correlation of the errors. We assume the residuals of the regression are a realisation of a stationary spatial process. The fitted regression surface can be used to make predictions within or slightly beyond the study area. Figure 5 shows the contour plot of the fitted quadratic surface. In total, twenty (20) sampling locations outside the project area (10 in the NE corner and 10 in the SW corner) were identified based on the fitted regression contour as shown in Figure 5 to be the best areas for further exploration. Ten sampling locations to the NE and SW respectively were identified. The fitted regression was used to predict the grade for all exterior points. Copula was used to estimate the deviation of logarithm of grade from the regression as shown in table 2.
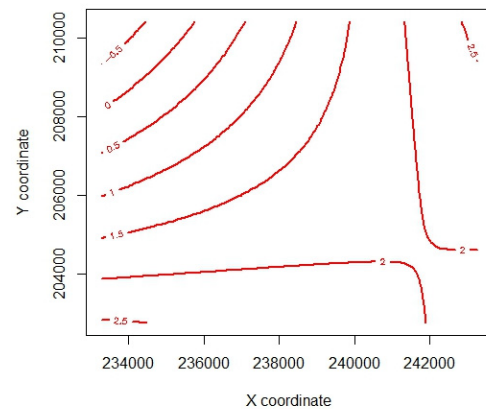
Table 2. Estimates of deviation of logarithm of grade from regression surface using copula *(Results using a randomly selected 95% of the data set are bracketed in italic)*

Nos of Neighbouring Points	NE Points (Mean Prediction)			SW Points (Mean Prediction)		
	Min Grade	Max Grade	Mean Grade	Min Grade	Max Grade	Mean Grade
3	0.59 *(0.54)*	2.29 *(2.18)*	1.59 *(1.51)*	0.67 *(0.65)*	2.37 *(2.39)*	1.79 *(1.79)*
5	0.80 *(0.74)*	2.33 *(2.19)*	1.77 *(1.68)*	0.77 *(0.69)*	2.50 *(2.57)*	1.85 *(1.82)*
10	0.85 *(0.83)*	2.48 *(2.49)*	1.80 *(1.74)*	1.79 *(1.48)*	2.64 *(2.77)*	2.23 *(2.24)*

The expected value of logarithm of grade at a point with coordinates (x_p, y_p) is given by the equation $\hat{w} = \hat{\beta}_o + \hat{\beta}_1 x_p + \cdots + \hat{\beta}_5 x_p y_p$ where $\hat{\beta}_i$ the estimated coefficients in equation (11). The means of the predicted logarithms of grade at the 10 SW exterior points and the 10 NE exterior points are both 2.61. However, these predictions ignore the information available from the residuals which are assumed to be correlated with grade at neighbouring points.

3.3. Constructing Empirical Copula Contours and Spatial Copula Construction

The residuals from the regression were considered as a realisation of a stationary process. A normal marginal distribution was fitted to the residuals as shown in Figure 4, and the residuals appear to be positively skewed. Therefore a rank transformation $r_{i:n} \rightarrow \frac{i}{n+1}$, where $r_{i:n}$ is the ith smallest residual was used to transformed the residuals to the uniform interval [0, 1]. This enabled the construction of empirical copula contours to explore the spatial dependence structure at all lag distances and directions. Using the kernel smoothing density package in R to generate the contours. Figure 6 shows the empirical copula contours for directions 135° and 45°, where 45° corresponds to the longest range. It can be observed from Figure 6 that at distance of (0-300m) and (0-600m) the spatial structure for direction 135° and 45° shows different empirical contour plots. The difference in the contour plots provides evidence for anisotropy. The Kendall's tau correlogram for the directions 135° and 45° were estimated with ranges around 1000 and 1100m respectively. Weighted correlogram between 135° and 45° was estimated using eqaution (5). In all, ten bivariate copulas competed at each distance class, their dependence parameters were estimated by maximum likelihood. The bivariate copulas with the highest likelihood value at each distance was chosen as the best fitting copula for that class.
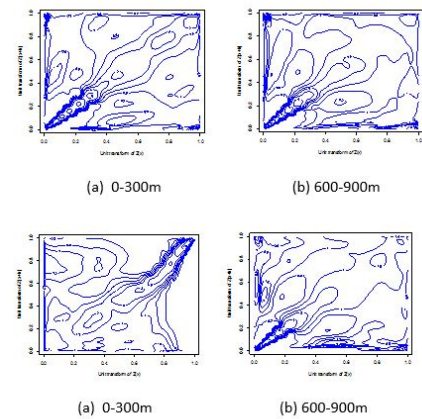

(a) 0-300m   (b) 600-900m

(a) 0-300m   (b) 600-900m

Figure 6. Empirical copula contours of residuals for direction 135° (upper) and 45° (lower)

The estimate of expected grade at the exterior points is given by $exp(\hat{w}_p + \hat{r}_p + s^2/2)$ where s^2 is the estimated variance of the errors in equation (11). The adjustment of $exp(\hat{w}_p + \hat{r}_p)$ by the factor of $exp(s^2/2)$ is to allow for the difference between the median and mean of an assumed log-normal distribution. The results using 3, 5 and 10 points for the copula are given in Tables 3 and 4. An approximate standard error is taken as one quarter of the width of the 95% prediction interval for the mean grade based on the regression surface, which is $\pm 4\%$. This is an approximation because: it does not allow for the correlation between the 10 neighbouring points and so undersestimates the error; but it does not allow for the reduction in uncertainty due to use of the copula. The mean estimator for the pair-copula model with normal margin was used. The data were back transformed by adding the predicted residual value to the fitted values of the regression and taking exponential.

The performance of the copula is compared with kriging by calculating the mean absolute error (MAE) at interior points for 10 neighbouring points, as shown in Table 3.

Table 3. Comparison of pair copula and kriging

Model	MAE
Pair copula (Gaussian Margin)	17.7
Lognormal kriging	18.3

4. DISCUSSION AND CONCLUSIONS

Gold grades from a mine in Ghana have been analysed. The 1500 gold grades surface soil samples within the field were analysed and show slight, but statistically significant, non-stationarity. The non-stationarity was accounted for by fitting a quadratic regression surface. The residuals from the regression were considered as a realisations of a stationary process. All possible pairs of residuals from this regression were grouped by their distances of separation, and the angle between them. The Kendall tau τ rank correlation was calculated from the corresponding pairs of grades. There was evidence of spatial correlation which depends on the direction of the vector joining the points. The correlation had a longer range in the 45° direction than in the 135° direction. This anisotropy was apparent in the contour plots of the empirical copulas as shown in (Figure 6). Copulas were fitted by maximum likelihood and were used to predict the distribution of the grades of residuals outside the main surveyed field (Total of 20 sampling points, 10 apiece at NE and SW respectively). The minimum, maximum and mean predicted grades for all sampling points were estimated for (3, 5 and 10) neighboring points. Table 2 shows the predicted grades based on the contouring of the grade deviation with the regression surface. The minimum, maximum and mean predicted grades for the NE and SW exterior points are highlighted in the table 4. The estimates change with the number of neighbouring points which suggest that at least 10 points should be used in this application, particularly in the SW corner. We recommend that the company should concentrate further exploration in the SW corner of the main field.

Table 5. Estimate of the expected grade at the exterior points *(Results using a randomly selected 95% of the data set are bracketed in italic)*

Nos of Neighbouring Points	NE Points (Mean Prediction)			SW Points (Mean Prediction)		
	Min Grade	Max Grade	Mean Grade	Min Grade	Max Grade	Mean Grade
3	59.86 *(55.26)*	409.82 *(348.76)*	231.29 *(201.43)*	65.16 *(63.20)*	457.13 *(480.31)*	263.62 *(266.99)*
5	78.46 *(71.45)*	434.98 *(351.25)*	258.45 *(220.17)*	76.48 *(71.11)*	522.03 *(577.14)*	276.59 *(275.15)*
10	71.49 *(73.39)*	463.56 *(447.74)*	270.17 *(242.39)*	208.12 *(148.47)*	581.39 *(627.15)*	358.64 *(380.21)*

REFERENCES

Aas, K, Czado, C, Frigessi, A & Bakken, H (2009). 'Pair-copula constructions of multiple dependence', *Insurance: Mathematics and economics*, vol. 44, no. 2, pp. 182-198.

Bárdossy, A & Li, J (2008). 'Geostatistical interpolation using copulas', *Water Resources Research*, vol. 44, no. 7.

Goovaerts, P, AvRuskin, G, Meliker, J, Slotnick, M, Jacquez, G & Nriagu, J (2005). 'Geostatistical modeling of the spatial variability of arsenic in groundwater of southeast Michigan', *Water Resources Research*, vol. 41, no. 7.

Gräler, B (2014). 'Developing Spatio-temporal Copulas'.

Journel, AG & Alabert, FG (1990). 'New method for reservoir mapping', *Journal of Petroleum technology*, vol. 42, no. 02, pp. 212-218.

Musafer, GN & Thompson, MH (2016). 'Non-linear optimal multivariate spatial design using spatial vine copulas', *Stochastic Environmental Research and Risk Assessment*, pp. 1-20.

Nelsen, RB (1999). 'Introduction', *An Introduction to Copulas*, Springer New York, New York, NY, pp. 1-4.

R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/)

Sklar, A (1959)., 'Fonctions Deépartion à n Dimensions et Leurs Marges', *Publications de l'Institut de Statistique de l'Université de Paris*. l'Université de Paris, Paris.