19th International Congress on Modelling and Simulation, Perth, Australia, 12–16 December 2011
http://mssanz.org.au/modsim2011

A review of the application of copulas to improve modelling of non-bigaussian bivariate relationships (with an example using geological data)

R.C. Boardman[a] and J.E. Vann[a,b,c]

[a] *Quantitative Group PO Box 1304 Fremantle WA 6959*
[b] *Centre for Exploration Targeting, The University of Western Australia, Crawley WA 6009.*
[c] *School of Civil Environmental and Mining Engineering, The University of Adelaide, Adelaide SA 5000.*
Email: rb@qgroup.net.au

Correctly modelling bivariate relationships between geological variables is vital in mineral resource estimation. Often these relationships are complex and more simplistic methods of modelling such as Monte-Carlo Simulations (MCS), a bigaussian distribution or linear regression are not suitable. MCS where correlation coefficients are specified are inherently problematic because they can only reproduce the marginal distribution and a specified rank correlation coefficient, they cannot reproduce complex dependency structures. Bigaussian modelling is only appropriate if the data is indeed bigaussian (which is essentially never the case for grade variables). Elementary linear regression models can only model linear relationships and are often used in a deterministic manner. Copulas offer a framework to model and simulate multivariate relationships that go beyond correlation coefficients. They allow the strength of dependence to vary in different quantiles. Copulas have been used extensively in the recent years, but they have only made a limited appearance in the mining industry; for example, they have been used in spatial (geostatistical) simulations (Bardossy and Li, 2008). This paper focuses on the use of copulas to model bivariate relationships in a non-spatial sense. The intended audience for this paper is the non-statistician, such as a resource geologist, or geochemist. We introduce the concept of copulas including fitting, simulation and validation for the unfamiliar reader. A geologically relevant case study is provided where copulas are compared to other traditional methods of bivariate simulations.

Copulas offer a framework to model multivariate data structures because the marginal distributions are modelled separately from the dependency structure (which is contained in the copula itself). A copula is a multivariate distribution with uniform margins on the interval [0,1]. There are many types of copulas; each one has a specific dependency structure. For example a Clayton copula has high level of dependence in the lower tail and low dependence in the upper tail. Copulas can be used for simulation purposes; a Bivariate Distribution linked via a Copula (BDC). Figuratively speaking, to simulate from a BDC the marginal distributions are simulated separately from one another and they are joined together in a manner which is consistent with the dependency structure of the copula. The objective of copula fitting is to fit a known copula to the rescaled ranks of the observed data.

The case study uses alumina and iron data from an iron ore deposit (figure 1). Several copulas were fitted to the pseudo observations of the bivariate data and as comparison a MCS, linear regression and bigaussian model were also fitted to the raw data. The goodness of fit of the copulas was assessed by a range of statistics and graphical measures. The copulas that provided a reasonable fit were carried through to the fitting of a BDC, in this case the Clayton, Frank, Plackett and normal copulas. The Clayton BDC was chosen as the most suitable to model the data by a selection of statistics and graphical methods. The performance of the comparative (non-copula) modelling techniques was poor; none of them accurately modelled the high strength of dependence in the lower tail. The Clayton BDC modelled the unusual dependency structure better and captured the high level of dependence in the lower tail. Although the Clayton BDC was the most suitable the simulation could have been better. This suggests that a more bespoke copula may give more favourable results.

In this case study it was possible to visually discern that the Clayton model was better than the other methods. A quantitative technique for validating across all the methods of bivariate simulations is recommended in future studies. The main conclusions from this case study were that a Clayton copula captured and modelled the high level of dependence in the lower tail better than the comparative modelling techniques. This case study demonstrates that using copulas to model the complex relationships that often exist between geological phenomena could be a promising alternative to other simplistic methods of modelling.

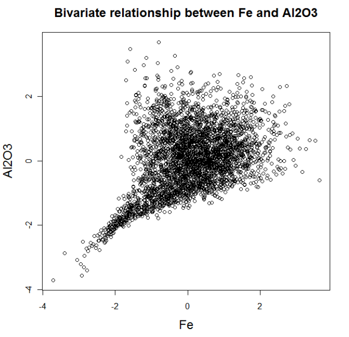


Figure 1: Scatter plot of observed data

Keywords: *Copulas, mining, dependency, correlation, bivariate modelling, Monte-Carlo Simulations*

1. INTRODUCTION

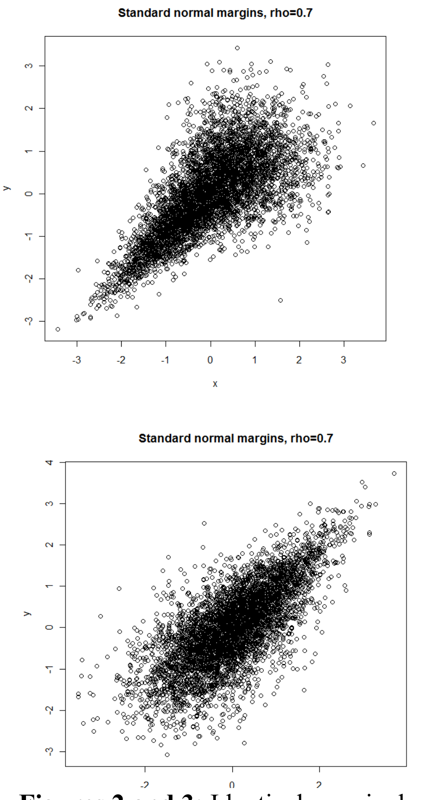
1.1 Problem definition

In geostatistics there has been some research using copulas for purposes of modelling multivariate dependencies for spatial conditional simulations (Bardossy and Li, 2008). This paper deals with non-spatial applications of copulas for geological data in the mining industry and will introduce the concept and methodology of using copulas for a non-statistical reader. See Nelson 2006 for an introductory text on copulas.

Currently in the mining industry Monte Carlo Simulation (MCS) and elementary regressions are used to model relationships between variables (Zou, 2007). In this text MCS refers to the use of the Iman-Conover (IC) algorithm (Iman and Conover, 1982) to simulate from a bivariate distribution. Software such as @RISK use the IC algorithm to provide a random sample where correlations are defined between variables. The IC algorithm is a distribution free method of simulating correlated variables and relies on the versatile Spearman's rank correlation coefficient. The IC algorithm involves defining numerous marginal distributions and a pair-wise Spearman's rank correlation coefficient. The weakness in this methodology is that specification of margins and a correlation coefficient does not completely describe the behaviour of the joint distribution. Simulation techniques that use the IC algorithm are thus not able to reproduce complex dependency structures; rather they just reproduce the rank correlation and the margins. Elementary linear regression ('best line fit') is also widely used in the mining industry but is very limited because it can only model linear relationships. Furthermore it is often used in a deterministic fashion, with more advanced methods such as a Bayesian approach, often not used in practice. In summary, methods that are currently used in the mining industry are unlikely to produce satisfactory results when complex dependency structures exist and there is significant uncertainty (randomness). This is problematic in the mining industry because many phenomena have complex relationships.

1.2 Correlation coefficients

Correlation is fairly misunderstood concept for non-statisticians. Perhaps correlation has enjoyed some undeserved attention over the years because of its ease of implementation and because it is (apparently) easy to understand. For most mining industry geologists, correlation coefficients and linear regressions comprise the entire tool kit for exploring and modelling bivariate relationships. However, correlation coefficients are just one available tool to describe dependencies between random variables. Pearson's linear correlation coefficient is only optimal when the multivariate structure is multivariate normal and a linear relationship actually exists. On the other hand, rank correlations (e.g. Kendall's tau (τ) (Kendall, 1938) and Spearman's rho (ρ) (Spearman, 1904)) make no such assumptions regarding the multivariate or univariate distributions nor do they assume any specific dependency structure. For this reason, rank correlations have become quite popular. A common misunderstanding for non-statisticians is that knowledge of the marginal distributions and the correlation defines the joint distribution. There is in-fact an infinite number of multivariate distributions which could fit such a description, unless of course the structure is of true multivariate normal form. Figures 2 and 3 illustrate this fallacy; the scatterplots show two identical marginal distributions (standard normal) and the same correlation coefficient (ρ=0.7) been used to create two very different bivariate distributions. These figures illustrate how knowledge of a correlation coefficient and margins does not necessarily completely inform knowledge of the bivariate structure.



Figures 2 and 3: Identical marginal dist. and correlation used to create different bivariate structures

1.3 Mathematical definitions for bivariate copulas

Copulas are in themselves a *p* dimensional distribution whose margins are uniform on the interval [0,1]. Since copulas are distributions; density, quantile and probability functions exist for them. Copulas have a parameter vector **θ**. A multivariate distribution of two variables u and v, C(u,v) is considered a copula if and only if:

1. U and V ~ Uniform[0,1]
2. For u and v, C(u,v)=a when u or v is equal to one and the other is equal to a
3. C(u,v) is isotonic; i.e., C(**a**)≤C(**b**) for all a,b in $[0,1]^2$, a≤b

4. C(u,v) is increasing
5. C(u,v) =0 if u or v is equal to 0

Sklar's theorem (Sklar, 1959) is fundamental to copula analysis. Consider two random variables X and Y and their corresponding density functions F(x) and G(y). If H is the joint density and C is a copula then:

$$H(X,Y) = C\{F(x), G(y)\} \qquad X, Y \in \mathbb{R} \tag{1}$$

The theorem states that, for a given joint distribution, a copula exists that can model the multivariate structure by linking the marginal distributions. If this joint density H(X,Y) is explicitly known then the copula C, and the density functions of the two variables F(x) and G(y) can be known with absolute certainty. In practice, knowledge of H(X,Y) is often very difficult to derive so the copula C usually needs to be estimated. This is made easier if the data is transformed into the domain as the copula, that is the two marginal variables are each transformed to be uniform over the interval [0,1], i.e., $[0,1]^2$.

1.4 Copulas and simulation of bivariate distributions

A copula is a multivariate distribution which exhibits a specific dependency structure. There are many types of copulas each one has a different dependency structure; an appropriate copula choice depends on the dependency structure evident in the observed data. As an illustrative example, figure 4 shows the distribution of a Gumbel copula (notice it is only defined on the interval $[0,1]^2$). Figures 5 and 6 illustrate the same copula used to link together various univariate distributions, note the simulation is now defined in the space of the univariate distributions. The Gumbel copula has a strong level of dependency in the upper tail and a weak dependence in the lower tail – this characteristic is evident in figures 4, 5 and 6.

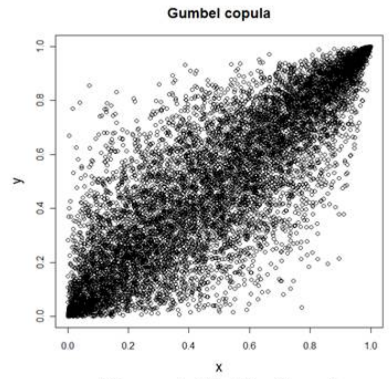


Figure 4: Distribution of a Gumbel copula

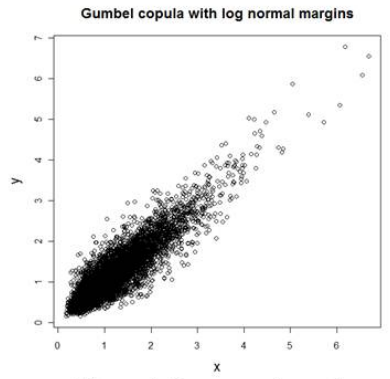


Figure 5: Log normal margins and a Gumbel copula

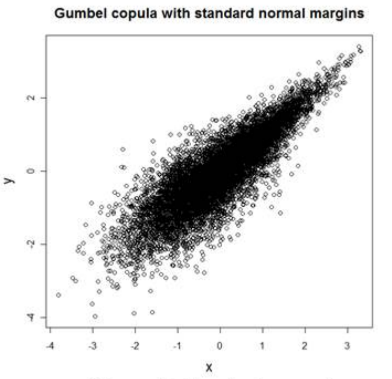


Figure 6: Standard normal margins and a Gumbel copula

1.5 Fitting copulas to data

The Probability Integral Transform (PIT) transforms any continuous univariate distribution into a standard uniform distribution. It is a monotonic increasing transform. The PIT is used to derive a cdf from a pdf:

$$F_x(a) = \int_{-\infty}^{a} f_x(a).\,dx \tag{2}$$

After applying the PIT, the domain of the cdf $F_x(x)$ is [0,1] – this fits in nicely to the domain of a copula! Equation (2) applies the pdf; $f_x(x)$, but if the pdf is not known and we are working with sample data we can use an empirical counterpart to derive the cdf (edf):

$$\hat{F}_x(x_i) = R_i/n \tag{3}$$

Where R_i is the rank of the i^{th} observation. In copula analysis this transformation is slightly altered by putting (n+1) on the denominator to avoid problems at the boundary of $[0,1]^2$ (Kojadinovic and Yan, 2010). Suppose we are working with a bivariate dataset (X,Y) where R_i and S_i denotes the ranks of X and Y respectively;

$$U_i = \frac{R_i}{n+1} \qquad and \qquad V_i = \frac{S_i}{n+1} \tag{4}$$

This transformation produces what are referred to as the 'pseudo observations'. It is worth noting U and V are uniformly distributed over the discrete set {1,..,n}/(n+1), wrt to $\hat{F}_x(x)$ and $\hat{G}_y(y)$. Importantly (U, V) contain the information about the dependency structure of (X,Y) (Genest and Favre, 2007). The pseudo observations (U, V) are sometimes referred to as the 'empirical copula' (Deheuvels, 1979). The objective of copula analysis is to define a known copula which is as close to the empirical copula as possible.

A copula is invariant under monotonic increasing transformations. Let φ denote a monotonic increasing transformation, for example the pseudo transformation defined in equation (4). Let U= φ(X) and V=φ(Y) denote two pairs of random variables; X and Y can be considered the original data and U and V can be considered the pseudo observations (transformed to uniform $[01]^2$). Genest and Favre (2007) showed that the copula defined for (X,Y) is the same as the copula defined for (U, V). This justifies fitting the copula to the transformed pseudo observations. Genest and Favre (2007) and Kojadinovic and Yan (2010) therefore advocate fitting the copula to the pseudo observations rather than cdf derived from the PIT. This is advantageous because it does not require the estimation of the marginal distribution to perform the PIT (equation 1), leaving one less place to make an error in the analysis and introduce bias.

As well as facilitating copula fitting; pseudo observations can also be a very useful visual tool for an exploratory data analysis viewpoint. A plot of the pseudo observations can speak volumes about the dependency structure evident in the data. This is because the marginal behaviour (e.g., skewness, dispersion, etc.) cannot distort the dependency structure. Referring again to figure 4, 5 and 6; 5 and 6 are both described by the copula shown in 4. This would have been very difficult to detect by looking at the raw data because it is distorted by the characteristics of the marginal distributions.

1.6 Parameter estimation

The copula parameter(s) are closely related to the strength of dependence between the variables. In this study we only consider copulas with one parameter. The correlation coefficient used for copula analysis must be invariant under the monotonic increasing transformations; i.e.:

$$\omega(X,Y) = \omega(U,V) \tag{5}$$

Where ω is some correlation measure, X and Y is the raw data and U and V are the transformed data. It has been shown that rank correlation measures such as τ and ρ are invariant under monotonic transforms (Al-Harthy et. al, 2007). On the other hand, Pearson's correlation is not invariant under monotonic transforms and is therefore not appropriate for copula analysis. The three most common methods of estimating copula parameter(s) are τ, ρ and Maximum Pseudo Likelihood (MPL). Rank measures of correlation can be related to the copula density and hence the copula parameter θ. Parameter estimation using rank correlation is very popular because closed form solutions exist for many commonly used copulas. For example the τ can be related to the Gumbel copula density by:

$$\tau = 1 - \theta^{-1} \tag{6}$$

ρ can also be related to copula densities in a similar fashion. MPL is analogous the Maximum Likelihood Estimation (MLE) but has been adapted to handle pseudo observations, the method was formalised by Genest et al. (1995). Deriving the MPL estimate requires maximising the following function by changing the parameter θ:

$$l(\theta) = \sum_{i=1}^{n} log\left\{c_\theta\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right)\right\} \tag{7}$$

Where c_θ is the copula density. MPL is often considered a superior method for copula parameter estimation (Genest and Favre, 2007). The `copula` package (Kojadinovic and Yan, 2010 and Yan, 2006) in the `R` software (R Development Core Team, 2011) provides functions to carry out copula parameter estimation using MPL, τ and ρ method.

1.7 Assessing the goodness of fit of a copula

Given that the pseudo observations are derived by non-parametric means (U,V) they are the most reasonable bench mark for validating how well a copula fits the data (Genest et al., 2009). The Cramér-von Mises statistic (CVM, see Anderson, 1962) is a non-parametric method for assessing the goodness of fit of a known probability distribution to the empirical distribution. It is also used in copula analysis to test the hypotheses that the empirical copula (pseudo observations) is compatible with the estimated copula. This test can be implemented for the Clayton, Gumbel, Frank, normal and Plackett copulas in the `copula` package. The p-value for this test can be determined a bootstrap method and the faster multiplier method (Kojadinovic and Yan, 2009). The CVM test statistic S_n is (a small value of S_n indicates a better fit):

$$S_n = \int_{[0,1]^2} n\left\{C_n(\boldsymbol{u}) - C_{\theta_n}(\boldsymbol{u})\right\}^2 dC_n(\boldsymbol{u}) \tag{8}$$

$$= \sum_{i=1}^{n}\left\{C_n(u_i, v_i) - C_{\theta_n}(u_i, v_i)\right\}^2 \tag{9}$$

The log likelihood associated with the MPL parameter estimate can also be used as a measure of the goodness of fit. The larger the value of the log likelihood, the better the fit is considered to be. A plot of the pseudo observations can be compared to a large value of random numbers generated from the copula to see how well they agree. If the copula mimics the structure of the pseudo observations then the copula may be an appropriate choice to model the data.

2. CASE STUDY FROM IRON ORE DATA

2.1 Background to data

Drill hole data was sourced from an iron mining company. The data provided consisted of the main chemical assays of the ore (iron (Fe), phosphorous (P), silica (SiO_2), alumina (Al_2O_3) and LOI or 'loss on ignition'). The bivariate relationships of the various combinations of these assays are very important to both resource modelling and day to day quality management of the iron ore business. They are usually modelled in industry using simple linear regression models, Pearson's correlation or (less commonly) MCS. Our analysis was carried out on the bivariate distribution of Fe and Al_2O_3 because this relationship looked the most interesting and complex (figure 1). The data were transformed into standard normal prior to commencement of copula analysis. The relationship between Fe and Al_2O_3 was originally negative. This negative relationship is usual for these two variables because the rock is largely comprised of iron minerals, thus an increase in silicates (and thus silica SiO_2) generally leaves less space for iron minerals and corresponds to a decrease in Fe. For copula analysis it was necessary to manipulate this relationship into a positive one, Al_2O_3 values were multiplied by -1. This step is mechanical and was taken because most of the copulas available in the `copula` package only model positive relationships; τ for this pair was 0.240. The parameters of the margins were estimated by Maximum Likelihood Estimation (MLE).

2.2 Fitting methodology for copulas and Bivariate Distribution linked via a Copula (BDC)

The copulas fitted to the data were: Clayton, Gumbel, Frank, Normal, Plackett, Galambo, Husler Reiss (HR) and Tawn. The following methodology was carried out for *each* copula.

Each copula parameter was estimated by the MPL method detailed in section 1.6. For each copula fitted to the pseudo observations the parameter estimate, standard error and MPL were recorded. Only the copulas which were deemed a close fit to the pseudo observations were considered for use in modelling the BDC. The goodness of fit was assessed by the value of the pseudo log likelihood and the appropriateness of the copula dependency structure. The latter was visually gauged by how well 10,000 random numbers from each copula fitted the observed pseudo observations. For many of the copulas it was evident that the dependency structure was not appropriate. Additionally, the Cramér-von Mises test statistic was considered as a measure of goodness of fit.

For the copulas deemed appropriate, a bivariate distribution linked via a copula (BDC) was initially created using the copula parameter estimated by MPL. The margins were specified as normal with the MLE for the mean and variance. This initial BDC served as a 'guesstimate' or starting point for more rigorous fitting and optimisation methods. These methods involved maximising the log likelihood of the BDC by changing the marginal parameters and the copula parameter. The BDC was fitted to the raw observed data to produce a final estimate of the copula parameter and the parameters of the marginal distributions. From this final BDC, 10,000 random numbers were generated and the log-likelihood was recorded.

2.3 Methodology for comparative modelling techniques

A bigaussian distribution was also simulated using the sample variance matrix and sample means as the parameters. A MCS was also completed in the @RISK software. The margins were specified as normal and using the ML estimates for the mean and variance.

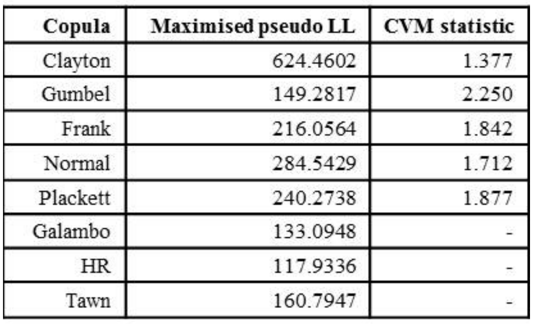
2.4 Validation procedure

To validate the adequacy of each simulation method (linear regression, bigaussian, MCS and copulas) a range of statistics and graphical diagnostics were used. Firstly each simulation was checked to see that the marginal distribution was preserved. This was carried out via a Kolmogorov-Smirnov (KS) test and Q-Q plots. To assess how well each simulation method reproduced the dependency structure the scatter plots of the final simulations were also visually compared to the observed data.

3. RESULTS AND CONCLUSIONS

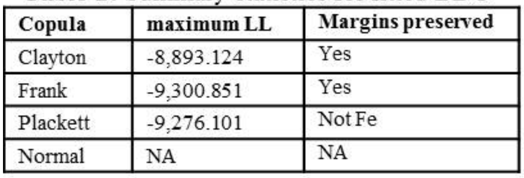
The results of the copula fitting are summarised in table 1. According to the maximised pseudo log-likelihood the Clayton copula offers the best fit. Additionally the CVM test statistic was the smallest for the Clayton copula, which is confirmatory of this result. The Frank, Normal and Plackett copulas all performed very similarly to each other with regards to the MPL and CVM statistic. Based on these statistics and inspection of 10,000 random numbers generated from each copula the Clayton, Frank, Plackett and Normal copula were carried through to further analysis of fitting a BDC. The summary statistics for the Clayton, Frank, Plackett and Normal BDC are summarised in table 2.

Table 1:Summary of fitted copulas



Copula	Maximised pseudo LL	CVM statistic
Clayton	624.4602	1.377
Gumbel	149.2817	2.250
Frank	216.0564	1.842
Normal	284.5429	1.712
Plackett	240.2738	1.877
Galambo	133.0948	-
HR	117.9336	-
Tawn	160.7947	-

Table 2: Summary statistics for fitted BDC



Copula	maximum LL	Margins preserved
Clayton	-8,893.124	Yes
Frank	-9,300.851	Yes
Plackett	-9,276.101	Not Fe
Normal	NA	NA

The log likelihood of the Normal BDC was undefined thus it was unable to be fitted to the data. The Clayton log likelihood is larger than the Frank and Plackett copula indicating a better fit for the Clayton copula. Additionally the Clayton and Frank preserved the margins Fe and Al_2O_3 whereas the Plackett copula did not preserve the marginal distribution of Fe.

Figure 7 shows the scatterplots from the three fitted BDC. The grey dots are the simulated values and the black points are the observed values. The Clayton BDC has captured the high strength of dependence in the lower tail and the weak dependence in the upper tails. The Frank and Plackett copula did not capture the dependency in the lower tail at all. These plots and the statistics in table 2 are taken as evidence that the Clayton copula has modelled the data better than the Frank or Plackett copula. However it is evident that the Clayton copula has not been able to model the dependency entirely satisfactorily, although there is a presence of a higher dependency in the lower tail it may not be as strong as we would like.

Figure 8 shows the scatter plots for the other simulation methods namely MCS and the bigaussian model. Neither simulation captured the high level of dependency in the lower tail; however, both methods did preserve the marginal distribution of Fe and Al_2O_3. In this analysis the Clayton copula modelled the dependency structure of the bivariate distribution of Fe and Al_2O_3 better than the other copulas and comparative methods.

The main limitation in this study was that there was no quantitative method to compare how each simulation method reproduced the dependency structure. Visual inspections of the scatter plots were the only means to assess reproduction of dependency structure. This qualitative approach was not considered a major limitation in this instance, because the Clayton copula was the better performer and the other simulation methods did not model the data accurately at all. The use of scatter plots to compare copulas and other simulation techniques would not be appropriate if it were not visually obvious how well each simulation performed. Moreover this strategy becomes impossible when the dimensionality is greater than 3.

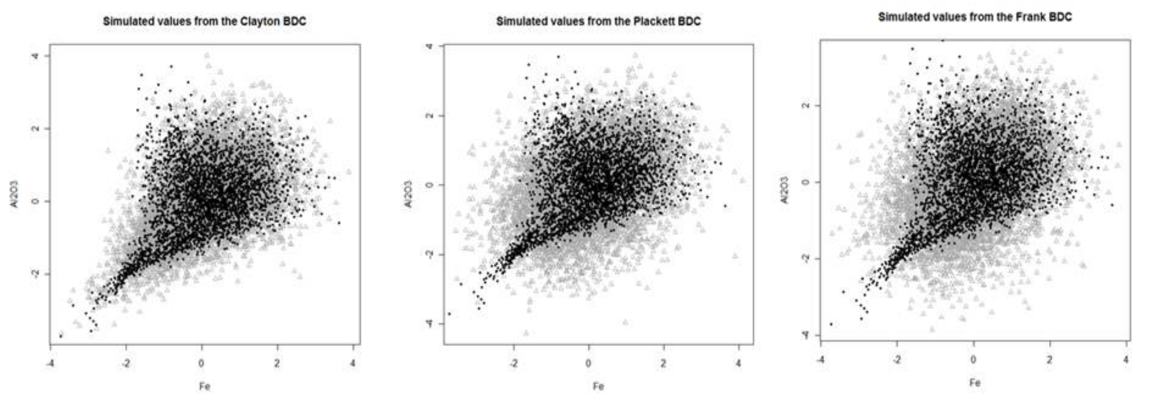


Figure 7: Scatter plots of fitted BDC simulations

4. CONCLUSIONS AND RECCOMENDATIONS

This study has demonstrated that although copulas are a useful tool for modelling data with complex dependency structures they are not a magic bullet. In this instance the Clayton copula outperformed the other simulation technics and the other copulas. However the Clayton copula did not produce a completely satisfactory result and improvements could be made. This indicates that this particular bivariate relationship could not be explained by the ‘off the shelf’ copulas and a more bespoke approach may provide better results. Also there may be a complex interaction between all five variables and this could explain the dramatic change in behaviour at (-3/2, -3/2) for Fe and Al_2O_3. Modelling the entire dataset with all five variables with a five dimensional copula is recommended

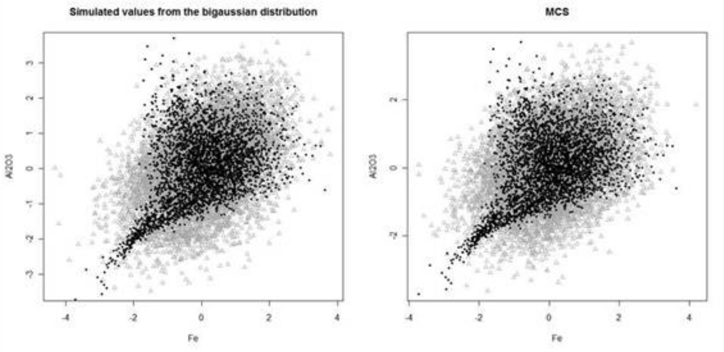


Figure 8: Scatter plots for comparative simulations

This case study has demonstrated that copulas can be a viable alternative to other methods of bivariate modelling when complex dependency structures exist. In this study the Clayton BDC clearly modelled the data more accurately than other BDCs or comparative techniques. Copulas are recommended over other more simplistic approaches because they better handle complex dependency structures, however more bespoke implementation may sometimes be required. Additionally they are easily implemented and are freely accessible in the `R` package `'copula'`.

REFERENCES

Al-Harthy, M., S. Begg, et al. (2007). Copulas: A new technique to model dependence in petroleum decision making, *Journal of Petroleum Science and Engineering* 57: 195-208.

Anderson, T. (1962), "On the distribution of the two-sample Cramer-von Mises Criterion." *Annals of Mathematical Statistics*, 33(3), 1148-1159.

Bardossy, A. and J. Li (2008). "Geostatistical interpolation using copulas." Water Resources Research 44.

Deheuvels, P. (1979), "La fonction de dépendance empirique et ses propriétés: Un test non paramétrique d’indépendance." *Bull. Cl. Sci,* 65(6), 274-292.

Genest, C., K. Ghoudi, et al. (1995), A semiparametric estimation procedure of dependence pararmeters in multivariate families of distributions. *Biometrika*, 82(3), 543-552.

Genest, C. and A. Favre (2007), Everything you always wanted to know about copula modelling but were afraid to ask. *Journal of Hydrologic Engineering* 12(4): 347-368.

Genest, C., B. Remillard, et al. (2009), Goodness-of-fit tests for copulas: A review and power study. *Insurance: Matheamatics and Economics* 44(1), 199-213.

Iman, R. and L. Conover (1982), A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics- Simulation and Computation* 11(3), 311-334.

Kendall, M. (1938), A New Measure of Rank Correlation. *Biometrika*, 30(1), 81-89.

Kojadinovic, I. and J. Yan (2010), Modelling multivariate distributions with continuous margins using the `copula R` package. *Journal of Statistical Software* 34(9), 1-20.

Kojadinovic, I. and J. Yan (2009). A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistical Computing*, 21(1), 17-30.

Nelson, R. (2006). *An Introduction to Copulas*. New York, Springer.

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

Spearman, C. (1904), The proof and measurement of association between two things. *The American Journal of Psychology* 15(1): 72-101.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Yan, J. (2006), Enjoy the Joy of Copulas: With a Package `copula`. *Journal of Statistical Software* 21(4): 1-21.

Zou, H (2007). *Quantitative Geochemistry*. Imperial College Press.