

# A comparison of Gaussian, Student-t and Vine copulas for modelling geophysical measurements along a rock drill core

E Addo Jr<sup>1</sup>

AV Metcalfe<sup>2</sup>

EK Chanda<sup>3</sup>

Received 17 November 2017; revised 18 July 2018

## Abstract

This case study is based on measurements made approximately at 20cm lengths along a down-the-hole diamond drill core from a pyrite mine in South Australia. The measurements are the P-wave velocity, magnetic susceptibility and impedance. The trivariate distribution is modelled using Gaussian, Student-t and vine copulas and the results are compared in terms of goodness of fit and differences in extreme values from distributions obtained by simulation from the copulas. The vine copula provides the best fit for the variables. Trivariate linear spatial Gaussian, Student-t and vine copulas are used to predict magnetic susceptibility one step below the depth of the drill core. The vine copula allows for more detailed modelling of the error structure, and so

---

DOI:10.21914/anziamj.v59i0.12646, © Austral. Mathematical Soc. 2018. Published October 19, 2018, as part of the Proceedings of the 13th Biennial Engineering Mathematics and Applications Conference. ISSN 1445-8810. (Print two pages per sheet of paper.) Copies of this article must not be made otherwise available on the internet; instead link directly to the DOI for this article.

provides more accurate 90% prediction intervals. The 90% prediction interval for the vine copula is wider than that for the Student-t copula, and both are wider than the interval obtained with the Gaussian copula. In general, copulas provides a more realistic modelling of geological variables and hence allows for accurate assessment of risk and uncertainty.

*Subject class:* Mathematical Modelling;Prediction

*Keywords:* Copula; Estimation; Mining

# Contents

<b>1</b>	<b>Introduction</b>	<b>C217</b>
<b>2</b>	<b>Methods</b>	<b>C218</b>
<b>3</b>	<b>Results</b>	<b>C220</b>
3.1	Fitting marginal distributions . . . . .	C222
3.2	Fitting trivariate copulas . . . . .	C222
3.3	Goodness of fit test . . . . .	C224
3.4	Models for predicting at further depths . . . . .	C226
<b>4</b>	<b>Conclusions</b>	<b>C227</b>
<b>5</b>	<b>Acknowledgements</b>	<b>C228</b>

# 1 Introduction

Deep seated orebodies, low metal grades and fluctuating commodity prices have a high impact on the mining industry potentially reducing profit margins. It follows that accurate modelling of all geological variables is needed to reduce

the risk associated with mineral prospects. It is claimed that, the precise modelling of geological variables is the most significant factor in the success of mining projects [8]. In many geostatistical applications, kriging is used for predicting geological variables at unknown locations, however this method is only optimum if the distribution of the variables are multivariate Gaussian MVG. In reality, most geological variables exhibit a skewed distribution, which makes kriging inaccurate for predicting at unknown locations. Copula models are ideal in dealing with highly skewed and tail dependent distribution, these models encompasses all mutivariate distributions including the MVG [1, 6].

In this paper, the versatility and potential advantages of copula modelling of multivariate relationships in the context of a down-hole diamond drill core taken during prospecting has been demonstrated. The geological variables are impedance, magnetic susceptibility and P-wave velocity. These three geological variables are important path-finder variables when exploring for pyrite mineralisation.

The first objective is to describe the multivariate distribution of the three variables using copulas. The second objective is to compare Gaussian, Student-t and vine copula models for predicting magnetic susceptibility at further depths. The vine copula allows for detailed modelling of the error structure and therefore provided an accurate 90% prediction interval for magnetic susceptibility at further depth.

## 2 Methods

This section gives a brief overview of trivariate Gaussian and Student-t copulas [7], and a brief description of pair copula construction [5].

Suppose that  $Z$  is a continuous random variable with a cumulative distribution function (cdf)  $F(z)$ . Since  $Z$  is a random variable so too is  $F(Z)$ , and it follows from its definition that  $F(Z)$  has a uniform distribution on the interval

$[0,1]$ . This is known as the probability integral transform of  $Z$ . Copulas are multivariate uniform distributions. They encompass all multivariate distributions because the marginal cdfs are uniformly distributed. Moreover, copulas also yield multivariate distributions by expanding the uniform margins to any probability distributions which can all be different.

The Gaussian copula is equivalent to a multivariate Gaussian distribution and the Student-t copula is equivalent to the multivariate  $t$  distribution. They can both be applied to multivariate data with a large number of components. There is also a very wide range of different forms of bivariate copulas. However, these bivariate copulas do not generally extend beyond bivariate data without restrictions, such as equal correlations between components, that limit their applicability. Vine copulas, also known as pair-copulas, provide a neat solution to this limitation.

The trivariate Gaussian copula belongs to the family of elliptical copulas and is equivalent to the standard MVG model. The cdf of the trivariate Gaussian copula is:

$$C(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3; \Sigma) = \Phi_{\Sigma} \left[ \Phi^{-1}(\mathbf{u}_1), \Phi^{-1}(\mathbf{u}_2), \Phi^{-1}(\mathbf{u}_3) \right], \quad (1)$$

where  $\Phi_{\Sigma}$  is the standardised MVG cdf and  $\Phi^{-1}$  is the inverse standard Gaussian cdf. The trivariate Student-t copula is defined, with respect to its multivariate Student-t distribution, in a similar fashion.

It follows from the multiplicative rule of probability that any multivariate distribution can be factorised in several ways using conditional distributions. In particular, a copula can be factorised as a product of the marginal distributions and the bivariate conditional copulas. Such factorisations are called pair-copula models. The pair-copula decomposition is however not unique, for example, a five dimensional density can have about 240 different forms of construction. Each decomposition expresses the full copula density differently. [5, 2] used a graphical model called the regular vines to arrange the large number of pair-copula constructions. This regular vine copula are made up of special cases of D-vines and canonical vines [3]. Canonical vines

are used when one can find a key variable that controls the relationships of the entire datasets. In this application, the canonical vine was selected. One of the variables used in this case study was found to control all the other variables. Moreover, that same variable is an important key path-finder for the exploration of pyrite mineralisation. Figure 1 shows the derivation of the canonical vine (C-vine) copula [5] for three variables, and has the density:

$$f_{123}(z_1, z_2, z_3) = f_1(z_1) \cdot f_2(z_2) \cdot f_3(z_3) \cdot c_{12}(F_1(z_1), F_2(z_2)), \\ \cdot c_{13}(F_1(z_1), F_3(z_3)) \cdot c_{23|1}(F_{2|1}(z_2|z_1), F_{3|1}(z_3|z_1))). \quad (2)$$

### 3 Results

The Brukunga mine site is located in the Southern Mount Lofty Ranges in South Australia. The town is located 5 km north-east of Nairne and 40 km east of Adelaide. The geology of the project area is primarily the Cambrian calc-silstones lying within the north-south trending Kanmantoo Trough, the youngest sequence in the southern part of the Adelaide Geosyncline. Iron-sulphide mineralisation occurs as three steeply east dipping conformable lenses that are separated by waste beds. Mineralization is pyrite and pyrrhotite with some minor sphalerite, chalcopyrite, galena and arsenopyrites.

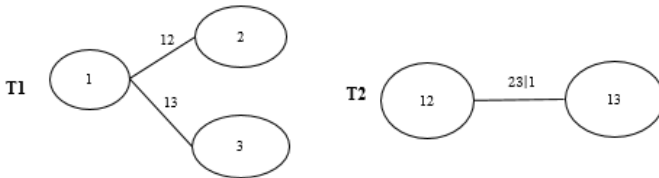


Figure 1: Canonical vine for three variables

A down-the-hole diamond drill core drilled to a total depth of 324.02 m, with an average sampling interval of 20 cm was sampled concurrently for the measurements of magnetic susceptibility, P-wave velocity and impedance. These three variables are the main path-finder variables for the exploration of pyrite mineralisation. Summary statistics of all three geological variables are given in Table 1. Histograms of the geological variables is shown in Figure 2, negative skewness is evident in impedance and P-wave velocity whilst positive skewness is evident in magnetic susceptibility. The Pearson correlations between the three geological variables are also shown in Figure 2. All the correlations are positive and the highest (0.95) is that between the impedance and P-wave velocity.

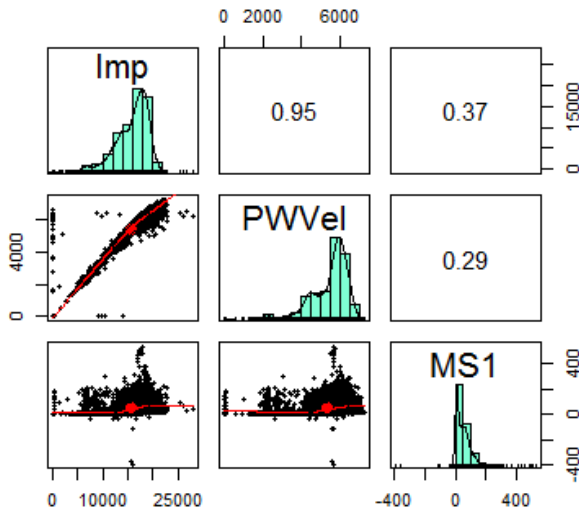


Figure 2: Pearson correlation matrix between all three variables

Table 1: Summary statistics of trivariate dataset

Statistics	Impedance (ohm)	P-wave Velocity (km/s)	Magnetic Susceptibility (1)
Number of Samples	11079	11079	11079
Minimum Value	0	0	-389.80
First Quartile	13799	4804	8.89
Median	16625	5742	41.21
Mean	15709.72	5370.29	52.10
Standard Deviation	3540.16	1099.14	49.96
Kurtosis	2.14	3.10	7.69
Skewness	-1.25	-1.57	1.62
Third Quartile	18268	6116	77.02
Maximum Value	27975	7270	532.67

### 3.1 Fitting marginal distributions

The first step in copula modelling is to estimate the marginal distribution functions of the variables. Therefore, the cdfs can be calculated from all ( $N = 11079$ ) observations. Without any prior information from all three distributions, a non-parametric estimation was chosen for the marginal distributions [9]. A kernel margin approximated by the empirical distribution functions (ecdf) were fitted for all three variables. For example

$$F_{X,T}(x) = \frac{1}{T+1} \sum_{i=1}^T I(X_i \leq x) = \frac{\text{rank}(X_i)}{T+1}, \quad (3)$$

where  $I()$  is 1 if its argument is true and 0 otherwise.

### 3.2 Fitting trivariate copulas

The ecdfs are used to transform the trivariate dataset to uniform scale. The trivariate Gaussian copula was fitted by maximum likelihood using the

function *fitCopula* in the package *Vine Copula* [10] in R. Figure 3 (upper row) shows contour plots from the trivariate Gaussian copula. These plots represent the marginal bivariate copula density functions of the fitted copula for all the pairs from the triples. The trivariate Student-t copula was fitted

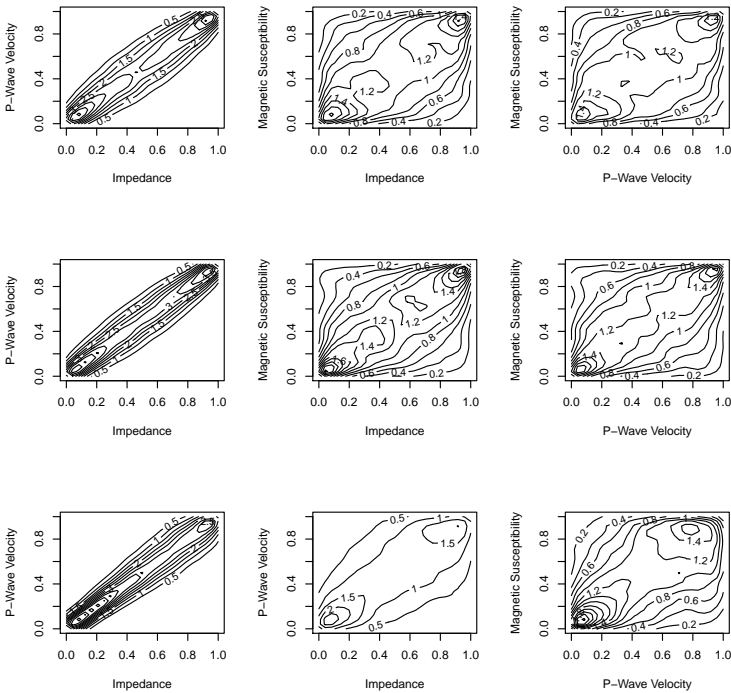


Figure 3: Fitted theoretical contour plots for pairs of variables using Gaussian (upper row), Student-t (middle row) and Vine copulas (lower row)

to the datasets using the function *fitCopula* in the package *Vine Copula* [10] in R. The degrees of freedom for the fitted t distribution was  $\nu$  equal to 6. Figure 3 (middle row) shows the contour plots from the fitted Student-t copula. The C-vine structure was fitted by using the decomposition expressed in the Equation (2), the joint density function of the datasets was estimated.



Unconditional survival Joe-Frank (BB8) and survival Joe-Gumbel (BB6) were fitted by maximum likelihood in the first tree using the functions *BiCopSelect* and *RVineMatrix* in the R package *Vine Copula* [10]. The density of the full pair-copula is a product of all the bivariate copula densities following the decomposition in Tree 1, a rotated 90° Joe-Frank (BB8) copula was fitted in the last tree. A table of all competing copulas for each tree fitting is given in Appendix A. The canonical vine structure explicitly includes the single variable marginal distribution (in this application, the impedance) as the root of the decomposition. Impedance was found to be the key variable that controls the relationship of the dataset based on the Pearson's correlations with all other variables. In addition, impedance is the main path-finder variable for the prospecting of pyrite orebodies. Figure 3 (lower row) shows the bivariate contour plots from the fitted C-vine structure.

### 3.3 Goodness of fit test

The cdfs of the three fitted copulas obtained with a Monte-Carlo procedure were plotted against the empirical copula  $\hat{C}_N$  in Equation 4 and Figure 4.

$$\hat{C}_N(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \frac{1}{N} \sum_{i=1}^N \mathbb{I} \left( \frac{D_i}{N+1} \leq \mathbf{u}, \frac{P_i}{N+1} \leq \mathbf{v}, \frac{M_i}{N+1} \leq \mathbf{w} \right), \quad (4)$$

where  $D_i$ ,  $P_i$  and  $M_i$  are the ranks of the impedance, P-wave velocity and magnetic susceptibility respectively. The plotted points corresponds to  $\mathbf{u} = \mathbf{v} = \mathbf{w}$  from 0.01 to 0.99 in steps of 0.01. The measure of fit is based on how close the points are to the diagonal line  $\mathbf{y} = \mathbf{x}$ . The vine copula provides a better fit compared with the Gaussian and Student-t copulas, because the points appear closer to the  $\mathbf{y} = \mathbf{x}$  diagonal line. This was quantified by calculating the mean of the squared vertical distances between the empirical copula and the fitted copula just before (when the empirical copula is below the fitted copula), or just after (when the empirical copula is above the fitted copula), each of the 11079 data. This gave values of 0.00138, 0.00117 and

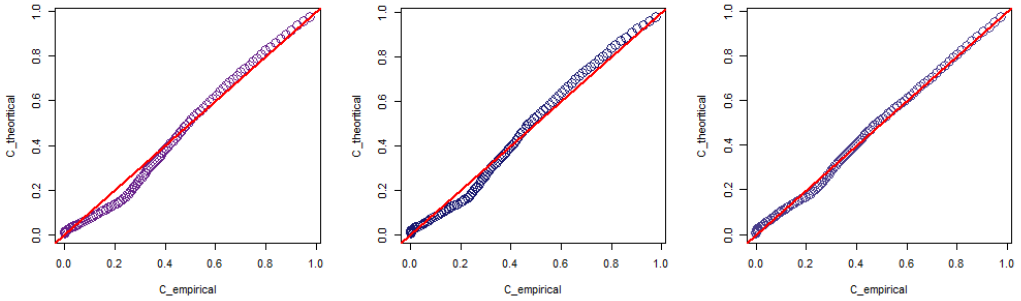


Figure 4: Fitted Gaussian against empirical copulas (left), fitted Student-t against empirical copulas (centre) and fitted vine against empirical copulas (right).

0.00023 for the Gaussian, Student-t and vine copulas respectively. The best fit, by this measure, is given by the vine copula. We formally compare the fit of the vine copula with the Kolmogorov-Smirnov test. The maximum deviation between the copula cdf and empirical density function is 0.0404. The probability of a value as large or large than this when sampling 11079 random deviates from the vine copula is found by simulation. The differences in the tail proportions of the three copulas and comparison of the field data was investigated by generating  $(10^5)$  random deviates from each copula. This was achieved with the R function `mvdc` in the package `copula` [4]. The proportion of triples with all three components below 0.05, and above 0.95, quantiles are shown in Table 2. The difference between the goodness of fit test for the Gaussian and Student-t copula is not apparent from Figure 4 (left) and 4 (centre). However, the proportion below and above 0.05 and 0.95 quantiles respectively shows that the Student-t copula draws large number of trivariate points to its upper and lower tails which is a consequence of its tail dependence. The vine copula allows for tail dependence when it is apparent but can also allow for lower levels of correlation in the tails than the Gaussian copula. In this application, the vine copula gives a considerably better fit in

Table 2: Number (parts per million) of triples below and above 0.05 quantiles and 0.95 quantiles for all three variables

Data	Number of Triples	Number below 0.05	Number above 0.95
Field observations	11079	3520	632
Gaussian copula	$10^5$	8190	8140
Student-t copula	$10^5$	12990	13010
Vine copula	$10^5$	11480	1390

the upper tails, which are more important in mining applications. There is less variation in the fits for the lower tails, but it can be seen that the vine copula is not constrained to be symmetrical between the two tails. Comparing all three fitted theoretical copula to the empirical field observations it seems that the vine copula provides a better fit to the trivariate datasets.

### 3.4 Models for predicting at further depths

The trivariate Gaussian, trivariate Student-t and vine copulas were used to predict magnetic susceptibility which is the primary path-finder variable for pyrite mineralisation, one step below the depth of drill core. The principle is to fit a trivariate copula to contiguous triples in the linear spatial series using maximum likelihood. The trivariate Gaussian copula fitted to magnetic susceptibility had a correlation of 0.97 at a lag of 1 and 0.93 at a lag of 2. The fitted trivariate Student-t copula to the magnetic susceptibility had a correlation of 0.99 at a lag of 1, and 0.97 at a lag of 2, with 3 degrees of freedom. The correlation for the fitted magnetic susceptibility for the first tree (Tree 1) in the C-vine were 0.99 and 0.97, the last tree (Tree 2) had a correlation of 0.82.

The predictions one step below the drill core for all three copulas, after back transforming are given in Table 3 below. The vine copula had the lowest predicted value compared with the Student-t and Gaussian copulas but the

differences are slight by comparison with the prediction interval. The 90% prediction intervals are narrowest for the Gaussian copula and widest for the vine copula. Both the Student-t copula and vine copula can allow for tail dependence if it is appropriate, and the vine copula allows for more flexibility in the modelling of the error structure than the Student-t copula. The wider prediction intervals are a consequence of the more detailed modelling and the 90% prediction interval from the vine copula should be the most accurate.

Table 3: Gaussian, Student-t and Vine copulas predictions and 90% prediction intervals

Model	Prediction	90% Prediction Interval
Gaussian copula	5.76	(4.93, 6.95)
Student-t copula	6.08	(5.32, 7.32)
Vine copula	5.71	(4.26, 8.99)

## 4 Conclusions

Two applications of copulas have been considered, and for each application three copula types have been compared. The first application was the modelling of trivariate distribution of impedance, P-wave velocity and magnetic susceptibility using copulas. The magnetic susceptibility is highly associated with pyrite mineralization which is the primary economic variable. However, the impedance and P-wave velocity provide further confirmation about the response properties of the rocks, such as the grindability and fragmentation, that affect the cost of processing. Estimating the relationship between these variables is important when deciding whether to exploit a prospect. The vine copula is more flexible than the Student-t copula or Gaussian copula and provided better fit in terms of closeness to the fitted empirical copula and the correspondence between its upper tails and those of the empirical copula.

The other application was for the prediction of magnetic susceptibility one step beyond the end of the drill core. Such predictions are valuable for establishing points for further exploration. The vine copula provided smaller prediction and far wider 90% prediction interval one step further depth. This finding is a consequence of the very large number of outlying values in the distribution of magnetic susceptibility (kurtosis of 7.69). In general, copulas provide a means for dealing with outliers as they only contribute to the fitting of the copula through their ranks. This is particularly valuable for geological data where outlying values are common and should not be dismissed as erroneous. In addition, copulas provides more accurate model, and also the uncertainty associated with predictions are more reliable.

## 5 Acknowledgements

This research is supported by Australian Government Research Training Program Scholarship awarded to Mr. Emmanuel Addo Jr. The authors are thankful to the mining company for providing the down-the-hole diamond drill core datasets used in this case-study.

The authors are also grateful to the reviewers for their comments and suggestions, which have improved the practical application of this manuscript.

## References

- [1] E Addo, EK Chanda, and AV Metcalfe. “Spatial Pair-Copula Model of Grade for an Anisotropic Gold Deposit”. In: *Mathematical Geosciences* (2018), pp. 1–26. DOI: [10.1007/s11004-018-9757-7](https://doi.org/10.1007/s11004-018-9757-7) (cit. on p. C218).

- [2] Tim Bedford and Roger M Cooke. “Vines: A new graphical model for dependent random variables”. In: *Annals of Statistics* (2002), pp. 1031–1068. URL: <http://www.jstor.org/stable/1558694> (cit. on p. C219).
- [3] Claudia Czado. “Pair-copula constructions of multivariate copulas”. In: *Copula theory and its applications*. Springer, 2010, pp. 93–109. DOI: [10.1007/978-3-642-12465-5\\_4](https://doi.org/10.1007/978-3-642-12465-5_4) (cit. on p. C219).
- [4] Marius Hofert et al. “copula: Multivariate dependence with copulas”. In: *R package version 0.999-9, URL http://CRAN.R-project.org/package=copula* (2014). URL: <http://copula.r-forge.r-project.org/> (cit. on p. C225).
- [5] A. Frigessi H. Bakken K. Aas C. Czado. “Pair-copula constructions of multiple dependence”. In: *Mathematics and Economics* 44.2 (2009). DOI: <https://doi.org/10.1016/j.insmatheco.2007.02.001> (cit. on pp. C218, C219, C220).
- [6] G Nishani Musafar and M Helen Thompson. “Non-linear optimal multivariate spatial design using spatial vine copulas”. In: *Stochastic environmental research and risk assessment* 31.2 (2017), pp. 551–570. DOI: [10.1007/s00477-016-1307-6](https://doi.org/10.1007/s00477-016-1307-6) (cit. on p. C218).
- [7] R. B. Nelsen. “An Introduction to Copulas”. In: *Springer, New York, 2nd Edition* (2006) (cit. on p. C218).
- [8] Richard Peattie and Roussos Dimitrakopoulos. “Forecasting Recoverable Ore Reserves and Their Uncertainty at Morila Gold Deposit, Mali: An Efficient Simulation Approach and Future Grade Control Drilling”. In: *Mathematical Geosciences* 45 (8 November 2013), pp. 1005–1020. DOI: [10.1007/s11004-013-9478-x](https://doi.org/10.1007/s11004-013-9478-x) (cit. on p. C218).
- [9] Olivier Scaillet, Arthur Charpentier, and Jean-David Fermanian. “The estimation of copulas: Theory and practice”. In: (2007) (cit. on p. C222).

- [10] Ulf Schepsmeier et al. *VineCopula: Statistical Inference of Vine Copulas*. R package version 2.0.5. 2016. URL: <https://github.com/tnagler/VineCopula> (cit. on pp. C223, C224).

## Author addresses

1. **E Addo Jr**, School of Civil, Environmental and Mining Engineering, University of Adelaide , South Australia 5005, AUSTRALIA, ORCID:<https://orcid.org/0000-0002-5046-8934>  
<mailto:emmanuel.addojunior@adelaide.edu.au>
2. **AV Metcalfe**, School of Mathematical Sciences, University of Adelaide, South Australia 5005, AUSTRALIA.  
<mailto:andrew.metcalfe@adelaide.edu.au>
3. **EK Chanda**, School of Civil, Environmental and Mining Engineering, University of Adelaide, South Australia 5005, AUSTRALIA.  
<mailto:emmanuel.chanda@adelaide.edu.au>

Table 4: Appendix A: Competing bivariate copulas for C-vine fitting

<b>Copula Models</b>			
Independence	Gaussian	Student-t	Clayton
Frank	Gumbel	Joe	BB1
BB7	BB8	BB6	Survival Clayton
Survival BB1	Survival Gumbel	Survival Joe	Survival BB6
Survival BB7	Rotated Clayton (90°)	Survival BB8	Rotated Gumbel (90°)
Rotated BB6 (90°)	Rotated Joe (90°)	Rotated BB1 (90°)	Rotated BB7 (90°)
Rotated Gumbel (270°)	Rotated BB8 (90°)	Rotated Clayton (270°)	Rotated Joe (270°)
Rotated BB7 (270°)	Rotated BB1 (270°)	Rotated BB6 (270°)	Rotated BB8 (270°)