Simulating a multivariate sea storm using Archimedean copulas

Stefano Corbellaa,b, Derek D. Stretcha,*

a Centre for Research in Environmental, Coastal & Hydrological Engineering, School of Engineering, University of KwaZulu-Natal, Durban, 4041, South Africa
b eThekwini Municipality, Coastal, Stormwater and Catchment Management Department, P.O. Box 680, Durban, 4001, South Africa

ARTICLE INFO

Article history:
Received 6 March 2012
Received in revised form 15 January 2013
Accepted 18 January 2013

Keywords:
Sea storms
Waves
Multivariate statistical models
Copulas

ABSTRACT

In order to provide realistic storm simulations and to quantify coastal risks the dependencies between storm parameters such as wave height, wave period and storm duration need to be considered. Copulas provide a means to achieve this by enabling the development of multivariate statistical models of sea storms. Although there are many families of copulas, Archimedean copulas are appealing to engineers because of their mathematical tractability. The dependencies between wave height, wave period, storm duration, water level and storm inter-arrival time (or calm period) were investigated in a case study on the east coast of South Africa using Kendall’s tau correlation coefficient as a dependency metric. Three methods of creating multivariate copulas were applied and the results were compared using (1) Kendall’s measure; (2) empirical multivariate distributions; and (3) simulations. Only the wave height, wave period and storm duration were found to be significantly associated. Hierarchical copulas provided the best trivariate model for the case study data. The trivariate analysis extends previous bivariate analyses and thereby enables a more detailed description of sea storms to be incorporated in the statistical model. A significant limitation of the current model is that it fails to link wave parameter statistics to physical forcing and physical constraints. Ways of overcoming these and other limitations are discussed.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The quantification of coastal storm risks has traditionally been done by considering only the wave height \( H \) of an event (Isaacson and MacKenzie, 1981). This practice is flawed since the destructive potential of the wave event is also a function of the storm duration \( D \), water level \( L \), inter-arrival time or calm period \( I \), wave direction \( A \) and peak period \( T \). In turn these factors are interdependent and these dependencies have to be considered when modeling the sea state.

It is intuitive that the larger the wave height, the higher the damage potential, and that the longer a storm's duration, the more damage can be inflicted. The effects of the other variables are not as intuitive with regard to risks and require further discussion. A high water level alters the shoaling of waves allowing larger waves to occur closer to shore and consequently act further inland which heightens the risk to the shoreline. Woodroffe (2003) and Sorensen (2006) both state that the significant of storm surge coinciding with high tide or spring tide cannot be over emphasized.

Storm inter-arrival times are of particular importance to coastal erosion. Beaches act as a defense to coastal developments. They are eroded during storm events and will have less dry beach width to defend inland developments while they are recovering from those events. According to the equilibrium profile theory a given water level or wave condition will produce a specific equilibrium profile once those conditions have been maintained for a long enough time. Provided that the equilibrium profile has not been achieved during the initial storm a subsequent storm, of appropriate magnitude, may be thought of as an extension to the previous storm’s duration allowing it to reach the equilibrium profile. Erosion aside, if another storm event occurred during the recovery period the lack of a beach buffer may mean that the wave-runup reaches coastal developments. Therefore the smaller the inter-arrival time the greater the risk.

The wave direction is significant for sediment movement as well as for loadings on structures. For example the closer a wave is to striking a structure at right angles the larger the force it exerts (Goda, 2008). Wave direction is also important from a sheltering perspective as a beach may have a lower risk to the most probable extreme events. For example a beach sheltered by a headland from south easterly swell may erode less from a 12 m south easterly wave than a 6 m north easterly wave. Unfortunately wave direction has not been considered in this paper due to limitations in the case study’s data set.

An increase in wave period has been shown to result in an increase in erosion (van Gent et al., 2008; van Thiel de Vries et al., 2008). Having said that, in terms of fluid acceleration, all else being equal, the shorter the wave period the greater the fluid acceleration and consequentially the greater the bottom shear stress. Wave periods also have both deterministic and random links to wave height and storm duration: deterministic in terms of physical limitations such as maximum wave steepness, but random in so far as different combinations of wave height and period may be produced from similar meteorological forcing depending on the travel distance. The relationship between wave period, erosion and...
meteorological forcing is complicated but it is not a major concern for statistical models because the probabilistic relationship between wave height and period will include randomness while the consequential erosion effects will typically be modeled using a process-based morphological model. The consideration of all these factors allows the true risk of an event to be considered. Copulas provide a tool to model the dependence between all the aforementioned variables that characterize coastal storm events.

Wave climate simulation models and beach morphology models are becoming increasingly popular in coastal, port and marine engineering (e.g., SWAN, Delft3D, Mike3, SBench, Xbeach, Unibest). The application of these models often requires the simulation of the effects of sea storms. To model a series of sea storms without considering the dependencies between the storm parameters can degrade the realism of these simulations. For example extreme wave height may be unlikely to occur during short duration storms. Similarly the coincidence of extreme wave heights with an extremely long storm duration may also be improbable. Incorporating the statistical dependencies between extreme wave heights and storm durations should produce a more probable set of storm events. The proposed multivariate copula model can provide a simulation of a wave climate for a given probability level or be used to conditionally simulate storm events. For example Corbella and Stretch (2012a) illustrate an application of the methods described in this paper for a study of coastal erosion trends.

The creation of mathematically consistent multivariate copulas can be difficult and methods of combining three 2-copulas into a 3-copula are still debated (Salvadori and Michele, 2006). An appreciation of the difficulty associated with constructing n-copulas is given by Nelsen (2006). There have been numerous methods proposed for the construction of multivariate copulas. For Archimedean copulas some examples include: hierarchical (Nelsen, 2006), conditional mixtures (De Michele et al., 2007) and the structure proposed by Chakak and Koehler (1995). Recently multi-parameter multivariate extreme value (MEV) copulas have been proposed by Salvadori and De Michele (2010), Salvadori et al. (2011). This paper focuses on multivariate Archimedean copulas and since only the Gumbel copula is both MEV and Archimedean, we preclude multi-parameter extreme value copulas. Multivariate statistical models can be challenging for routine engineering applications because they tend to require higher levels of statistical and mathematical fluencies. Archimedean copulas may not be as flexible as MEV or vine copulas (Brechmann and Schepsmeier, 2011), but they are designed to be more tractable mathematically and to provide a significantly better model of dependency than traditional techniques. Archimedean copulas therefore improve on classical univariate techniques while being simpler to apply than other copula classes. They also have the advantage of being able to create multivariate copulas using the simple hierarchical technique which often fails with other copula classes.

This paper investigates which sea storm parameters are independent and tests which multivariate construction technique is appropriate to model and simulate sea storms.

We initially provide an overview of Archimedean copulas and the available multivariate construction techniques. The selection of an appropriate copula is then described followed by the associated dependencies. We then use hierarchical, condition mixtures and the Chakak and Koehler (1995) structure to construct multivariate Archimedean copulas for wave data from a case study on the east coast of South Africa. We identify the best model through a simulation study.

2. Theoretical background and methods

2.1. Archimedean copulas

Copulas are mathematical functions that join or couple multivariate probability distribution functions \( F(x_1, \ldots, x_n) \) to their one-dimensional marginal distribution functions \( F_1(x_1), \ldots, F_n(x_n) \). For a detailed introduction to copulas refer to Nelsen (2006), Salvadori and De Michele (2010), De Michele et al. (2007), Joe (1997).

Archimedean copulas are a special class of copulas. They have been used in a wide range of applications because of properties that make them easy to construct (Nelsen, 2006). An Archimedean copula is a solution to the functional equation

\[
\phi(C(u,v)) = \phi(u) + \phi(v)
\]

where \( u = F(x) \) and \( v = F(y) \) are marginal distribution functions \((u,v)\) and \( \phi \) is the generator function. Simply put a copula disentangles the marginal distributions and the dependence structure of a multivariate distribution. An example of an Archimedean copula is the Clayton copula with generator function

\[
\phi(t) = \frac{1}{\theta} (t^\theta - 1)
\]

where \( \theta \) is the dependence parameter and \( t \) is a number from 0 to 1. Nelsen (2006) provides an extensive list of other Archimedean copulas.

Kendall’s tau \( \tau_X \) is the sample version of the measure of association defined in terms of concordance, namely

\[
\tau_X = \frac{c - d}{c + d}
\]

where \( c \) is the number of concordant pairs and \( d \) is the number of discordant pairs. A valuable property of Archimedean copulas is that \( \tau_X \) can be expressed as a function of the generator:

\[
\tau_X = 1 + 4\int_0^1 \frac{\phi'(t)}{\phi(t)} \, dt.
\]

The dependence parameter \( \theta \) can be found from Eq. (4) using Kendall’s tau.

2.2. Constructing multivariate Archimedean copulas

An n-copula cannot simply be used to “couple” another \((n-1)\)-copula with a variate by setting them as its marginal distributions. For example attempting to solve

\[
C^n(u_1, u_2, \ldots, u_n) = C(C^{n-1}(u_1, u_2, \ldots, u_{n-1}), u_n)
\]

often fails and is referred to as the compatibility problem (Nelsen, 2006). The literature does however offer various techniques to create multivariate distributions from copulas. Some of these techniques are presented below. The first two techniques we describe are based on conditional laws. A conditional distribution describes the probability of observing a variate given that we know the probability of occurrence of an associated variate. In physical terms, we may want to know, given that there will be a wave height of 8.5 m, what storm durations can occur. So \( F_{X|Y}(y) \) would be the probability of observing \( X \) for a known value of \( Y \). Mathematical details of the following methods are provided in Appendix A1.

Chakak and Koehler (1995) present one of the simplest ways of forming a 3-copula that uses bivariate conditional distributions to create a trivariate copula (Eq. (A.1)). The method is appealing because of its simple form and the fact that all the dependence parameters are available from the corresponding 2-copulas. However the resulting 3-copula is not uniquely determined and is dependent on the order in which the copulas are combined.

The conditional mixtures approach was used by Salvadori et al. (2007), Joe (1997) and De Michele et al. (2007) to combine two 2-copulas to form a 3-copula. This method is conceptually similar to that of Chakak and Koehler (1995). In this case the three dimensional distribution is obtained from the conditional bivariate distributions.
Other methods considered have mainly been used in financial modeling. They are the fully nested and partially nested methods suggested in Joe (1997) and presented by Savu and Trede (2006) as a joint distribution of asset returns. The method was also applied by Grimaldi and Serinaldi (2006), Nelsen (2006), Whelan (2004) and Embrechts et al. (2001). The term used is “hierarchical Archimedean copula” which is a copula that joins two or more bivariate or higher order copulas by another Archimedean copula.

Hierarchical Archimedean copulas can either be created by the fully nested method or the partially nested method. The partially nested method is more flexible than the fully nested (Savu and Trede, 2006), but it is however not possible to create a 3-copula since the lowest dimension is n = 4.

The major modeling limitation of hierarchical Archimedean copulas is that not all combinations of joint distribution are modeled uniquely. This may not be general enough for modeling a sea state and needs to be considered if the method is used for that application. The fully nested method provides an improved model in this regard.

The partially nested method has the benefit of producing a 4-copula from only two 2-copulas but is unable to produce a 3-copula. Three 2-copulas are required to create a 3-copula via the conditional mixtures approach (Eq. (A.3)) and by the Chakal and Koehler (1995) method (Eq. (A.1)). Eq. (A.3) is likely to provide the most complete model as it allows all the dependencies to be modeled by different copulas.

3. Case study

The east coast of South Africa has 18 years of reliable wave data from a wave recording buoy near the city of Durban. Corbella and Stretch (2012b) provide details of the data set. The data was aggregated into storm events following the methods of De Michele et al. (2007) and the three multivariate copula construction techniques were applied to the wave data to create a storm sea state model.

3.1. Univariate analysis

A storm event was defined as a wave event that exceeded a threshold significant wave height $H$ of 3.5 m (Fig. 1). Experience has shown that wave heights exceeding 3.5 m are typically associated with significant erosion. Each storm was assumed to commence when the wave height exceeded 3.5 m and to end when the wave height fell below 3.5 m and stayed there for at least two weeks. The two week delay was selected so that consecutive storm events were approximately statistically independent based on the decay of the Spearman autocorrelation of the wave heights which becomes small ($\ll 0.1$) for lags greater than two weeks. The storm data set was then manually screened to ensure that each storm event represented one meteorological system. Finally we selected, on average, the three storms with the largest significant wave heights per year. This method of selecting independent storms is similar to those previously used by Mendez et al. (2008), Salvadori et al. (2007), Callaghan et al. (2008), U.S. Army Corps of Engineers (2006).

The storm duration $D$ is the time in hours between the start and end of a storm. The storm inter-arrival time or calm period $I$ was defined as the time in hours between the end of one storm and the start of the next storm. The maximum peak period $T$ that coincided with the storm event was also considered along with simulated water levels $L$. Wave direction was not considered due to insufficient data.

The relative goodness of fit of various statistical models was assessed using the Akaike information criterion, namely

$$AIC = 2K - 2\ln(L)$$

where $K$ is the number of parameters in the fitted probability distribution and $L$ is the maximized value of the likelihood function for the parameter estimation. The variates $H, D, T$ and $L$ were best fitted with the generalized extreme value (GEV) distribution. The GEV distribution has been previously used to model wave heights by Guedes Soares and Scotto (2004), Chini et al. (2010), Mendez et al. (2008), Minguez et al. (2010) and Ruggiero et al. (2010) and $D$ by De Michele et al. (2007) and is defined by

$$y = \alpha^{-1} e^{-(\frac{x-\mu}{\alpha})} e^{-e^{-(\frac{x-\mu}{\alpha})}} : -\infty < x < \infty$$

where $\mu$ is a location parameter, $\sigma$ is a scale parameter, $k$ is a shape parameter. Although only $H$ is truly an extreme value the GEV provided superior Akaike information criterion for the other parameters when compared to several different distributions. There is no theoretical motivation to use a specific distribution (Goda, 2008) and because copulas allow the dependence structure to be treated separately from the marginal distributions the GEV distributions can be replaced with any distribution representative of the given data.

The storm frequency $1/(I + D)$ was modeled with the Poisson distribution

$$y = \frac{\lambda^x e^{-\lambda}}{x!},$$

where $\lambda$ is the mean and variance. Seasonality of storm frequency can be included by allowing $\lambda$ to vary in time. Seasonality has no relevance in this paper and will not be discussed further.

To explore the tail dependence of $H$ and $D$ and of $H$ and $T$ we can consider the pairs as single values that describe their energetics. De Michele et al. (2007) used a storm magnitude based on an equivalent triangular storm model (see Fig. 1), following Bocciotti (2000). Let the storm magnitude be

$$M = (H - \eta)D/2,$$

where $\eta$ is the wave height threshold taken here as 3.5 m. The storm magnitude provides a single quantity to measure the magnitude of a storm produced by the interdependence of $H$ and $D$. We extend this concept by also considering the wave power

$$P = \left(\frac{1}{4h}\right) E$$

where $g$ is gravitational acceleration, $\rho$ is the density of salt water and the average wave energy $E$ is given by

$$E = \frac{1}{g} \rho g H^2.$$
3.2. Bivariate analysis

The dependence between the variates is required prior to constructing the multivariate copulas. Kendall’s tau is a nonparametric measure of correlation and was used, along with the corresponding p value, to determine the degree of dependence between the pairs HD, HT, HI, HL, DT, DI, DL, TI, TL, and IL. Kendall’s τb considers the number of tied pairs and is given by

\[ \tau_b = \frac{(n_1-n_2)}{\sqrt{(n_1+n_2)(n-n_1-n_2)}} \]  

(11)

where \( n_0 = n(n-1)/2 \), \( n_1 = \sum t_i(t_i-1)/2 \), \( n_2 = \sum j_i(j_i-1)/2 \), \( t_i \) is the number of tied values in the ith group of ties for the first quantity, \( j_i \) is the number of tied values in the jth group of ties for the second quantity, \( n_t \) is the number of discordant pairs, and \( n_d \) is the number of discordant pairs. A p value of less than 0.05 was considered acceptable to reject the null hypothesis that \( \tau_b = 0 \). Kendall’s tau was also calculated for the pairs as a function of storm magnitude and peak wave power.

3.3. Selecting 2-copulas

The following Archimedean copulas were considered: Clayton, Gumbel, Ali–Mikhail–Haq, Frank and copula 12 in Nelsen (2006, Table 4.1). These copula functions, their generator functions \( \phi(\cdot) \) and the relationship between their dependence parameter \( \theta \) and Kendall’s tau are shown in Table 1.

We found the best-fitting Archimedean copula using a nonparametric estimation procedure proposed by Genest and Rivest (1993) and which has been successfully applied by numerous other authors (e.g. Accioly and Chiyoshi, 2004; Dowd, 2008; Zhang and Singh, 2007). Let \( Z = H(X, Y) \) have a distribution function \( K(z) = P(H(X,Y) \leq z) \) in the interval (0,1). An empirical distribution of \( K(z) \) can be determined as \( \hat{K}(z) = \text{proportion of } Z_{ik} \leq z \), where \( Z_{ik} \) is the number of \( (X_j, Y_j) \) such that \( X_j < X_i \) and \( Y_j < Y_i \).

Genest and Rivest (1993) showed that a parametric estimator of \( K \) could be found using the generators of the Archimedean copulas. The generator estimate of \( K \), denoted by \( \hat{K}_p(z) \), is given by

\[ \hat{K}_p(z) = z - \frac{\phi(z)}{\phi'(z)} \]  

(12)

The best fitting copula was found by plotting \( \hat{K}_p(z) \) against \( \hat{K}(z) \) (known as a Q-Q plot) and performing a Chi-squared and Kolmogorov–Smirnov goodness of fit test. All p-values were calculated for a 95% confidence level. The smallest Chi-squared value, calculated from

\[ \chi^2 = \sum_{i=1}^n \frac{(\hat{K}(z_i) - \hat{K}_p(z_i))^2}{\hat{K}_p(z_i)} \]  

(13)

was used to select the best fitting copula.

3.4. Empirical non-exceedance probabilities

Various authors have proposed empirical non-exceedance probabilities as plotting position formulae and those appropriate to wave parameters are discussed by the U.S. Army Corps of Engineers (2006). We used the plotting position formula proposed by Gringorten (1963) since Zhang and Singh (2007) successfully applied it to a trivariate probability distribution. The empirical trivariate probability distribution can be expressed as

\[ F(u, v, w) = P(U \leq u, V \leq v, W \leq w) = \sum_{i=1}^N \sum_{m=1}^N \sum_{p=1}^N n_{i,j,p} \cdot 0.44 \]  

(14)

where \( N \) is the sample size and \( n_{i,j,p} \) is the number of occurrences of the combinations of \( u_i \) and \( v_n \) and \( w_p \).

4. Results

4.1. Dependence between variables H, D, T and I

Kendall’s tau and the corresponding p values were calculated for the following pairs: HD, HT, HI, HL, DT, DI, DL, TI, TL, and IL. Only HD, HT and DT showed a statistically significant correlation (Table 2). This is not unexpected as it shows that the three variates \( H, D \) and \( T \) are dependent while \( I \) and \( L \) are likely to be independent. This weakness in association allows us to simplify our model as \( I \) and \( L \) may be simulated independently from the other variates using only their marginal distributions.

The results in Table 2 show the chosen variate pairs to be positively dependent. This implies that the larger the wave heights and periods the longer the storm duration. That is the longer the storm duration the greater the wave power. Similarly the greater the storm magnitude the greater the wave period.

Fig. 2 shows the correlation between HD pairs for storm magnitudes exceeding specified threshold values. Similarly the correlation between HT pairs with peak wave powers above specified thresholds are also shown in Fig. 2. Both HD and HT correlations are positive for low thresholds, but become negative as the thresholds increase. The HD correlation becomes positive again at the high storm magnitudes. The HT correlation behaves differently for high peak wave power where after an initial increase it subsequently changes from positive to (perfect) negative correlation. This is thought to be attributable to the very few available data pairs in that extreme range. It is expected that with more data pairs the increase to positive correlations would continue into the upper range of peak wave power.

The trend of the storm magnitude can be explained as follows: small magnitudes are associated with positive HD correlations because as relatively small wave events increase so do the event durations. The middle range becomes negative because relatively small waves coincide with long durations or large waves coincide with short durations. The trend becomes positive again because extreme storm magnitudes are formed by large wave events having long durations and the larger the waves the longer the duration. A similar explanation can be applied to the peak wave power. These results can be significantly influenced by the storm event definition and De Michele et al. (2007), following Boccotti.

### Table 1

<table>
<thead>
<tr>
<th>Family</th>
<th>Function, C(u,v)</th>
<th>Generator</th>
<th>Kendall’s tau</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton</td>
<td>([\max(u^{-1} + v^{-1} - 1, 0)]^{-0.9})</td>
<td>((z - 0.9)^{10})</td>
<td>(\theta (\theta + 2))</td>
</tr>
<tr>
<td>Gumbel–Hougaard</td>
<td>(\exp(-((1 - (u^-1) + (v^-1))^{0.9}))</td>
<td>((-1 - v)^{10})</td>
<td>(\theta (\theta + 2))</td>
</tr>
<tr>
<td>Ali–Mikhail–Haq</td>
<td>(uv/(1 - \theta (1 - u) (1 - v)))</td>
<td>(\phi (\phi + 2))</td>
<td>(\theta (\theta + 2))</td>
</tr>
<tr>
<td>Frank</td>
<td>(-\frac{1}{2} \ln (1 + (\frac{u}{v} - 1) \frac{(v - u)}{(v + u - 2)}))</td>
<td>(-\ln (\frac{u - v}{2}))</td>
<td>(\theta (\theta + 2))</td>
</tr>
<tr>
<td>Nelsen (2006)</td>
<td>((1 + {u^{-1} - 1}^{0} + (v^{-1} - 1)^{0} (1 + 1)^{0})^{-1})</td>
<td>((z - 1)^{10})</td>
<td>(\theta (\theta + 2))</td>
</tr>
</tbody>
</table>
4.3. Fully nested hierarchical copulas

In an attempt to explore the compatibility problem the trivariate copulas were constructed with different 2-copulas as well as with different copulas that did and did not satisfy the requirement \( \theta < \theta_{12} \) (refer to Appendix A1). Table 4 shows the fitting results of the trivariate copulas, Kendall’s tau, dependence parameters and goodness of fit between the empirical trivariate distribution and the copula based trivariate distribution. Let the marginal distribution functions of wave height, wave period and storm duration be \( F(H) = h, F(T) = t \) and \( F(D) = d \) respectively. Table 4 shows that \( C(h,d) \) and \( t \) nested with the Clayton copula has the best fit to the empirical distribution. However, \( C(h,t) \) and \( d \) do not satisfy the condition \( \theta < \theta_{12} \). Fig. 4 shows that the empirical cumulative distributions of both trivariate copulas are similar. The nesting should be performed with \( C(h,d) \) and \( t \) which does satisfy the condition \( \theta < \theta_{12} \).

4.4. Conditional mixtures and the Chakak–Koehler approach

The conditional mixtures and the Chakak and Koehler (1995) approach can be used to construct multivariate copulas from only the 2-copulas. The results yielded the smallest Chi-squared value of 1.04 while the conditional mixture approach was 4.43.

4.5. Simulation comparison

We initially simulate \( F(H), F(T) \) and \( F(D) \) from the bivariate copulas \( C(h,t) \) and \( C(h,d) \). The bivariate simulations yield a scatter plot very similar to the measured data (Fig. 5).

The bivariate Clayton copulas provide an appropriate model for the association between \( HD \) and \( HT \). We now consider which construction technique yielded the best trivariate model. The construction of the trivariate copula \( C(h,d,t) \) following both Chakak and Koehler (1995) and the conditional mixtures approaches are performed by coupling three 2-copulas into a 3-copula conditionally. The similarities in the techniques suggest that the simulations should be similar. The results are shown in Fig. 6 where it is evident that both methods do indeed produce similar results. Fig. 7 shows that the simulation results of the hierarchical copula \( C(C(h,d), t) \) are also similar to those of the Chakak and Koehler (1995) and the conditional mixtures methods. A visual comparison with Fig. 5 indicates that the hierarchical model has the most similar simulation results to the bivariate copula simulations. The hierarchical copula is expected to produce the best simulation results as it had the best fit to the empirical data.

5. Discussion

5.1. Copula construction methods

The p-values in Table 5 show that the conditional mixture construction technique did not provide an appropriate fit to the empirical trivariate distribution. The hierarchical trivariate Clayton copula has the best Chi-squared statistic while the Chakak and Koehler (1995) construction technique had the best Kolmogorov–Smirnov statistic. Ultimately both techniques provide acceptable results. The appropriateness of the trivariate Clayton copula is dependent on

---

**Table 3**
The best fitting copulas, dependence parameters and the Chi-squared (\( \chi^2 \)) and Kolmogorov–Smirnov (ks) statistics for the pairs HD, HT and DT. The p-value at a 95% level of confidence is also shown.

<table>
<thead>
<tr>
<th>Pairs</th>
<th>Copula</th>
<th>Dependence parameter (( \theta ))</th>
<th>Goodness of fit</th>
<th>( \chi^2 )</th>
<th>ks</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD</td>
<td>Clayton</td>
<td>2.08</td>
<td></td>
<td>0.219</td>
<td>0.100</td>
<td>0.911</td>
</tr>
<tr>
<td>HT</td>
<td>Clayton</td>
<td>0.543</td>
<td></td>
<td>0.210</td>
<td>0.100</td>
<td>0.911</td>
</tr>
<tr>
<td>DT</td>
<td>Clayton</td>
<td>0.662</td>
<td></td>
<td>0.187</td>
<td>0.067</td>
<td>0.999</td>
</tr>
</tbody>
</table>

---

**Table 2**
Kendall’s tau and p values for the pairs HD, HT and DT.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Kendall’s tau</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD</td>
<td>0.594</td>
<td>1.22 \times 10^{-11}</td>
</tr>
<tr>
<td>HT</td>
<td>0.214</td>
<td>0.0159</td>
</tr>
<tr>
<td>DT</td>
<td>0.249</td>
<td>0.00402</td>
</tr>
</tbody>
</table>

---

**Fig. 2.** Kendall’s tau for HD pairs as a function of storm magnitude threshold (solid line) and HT pairs as a function of peak wave power threshold (dotted line).
the bivariate relationships. All our bivariate models are best modeled by the Clayton copula and \( C(h,t) \) and \( C(d,t) \) have similar dependence parameters (Table 3). This means that in our case study the limitations of hierarchical copulas are not realized. It is expected that one of the other two construction techniques would produce a better model if the bivariate distributions were not all described by the Clayton copula. So although the fully nested hierarchical copula provided the best fitting model it is unlikely to do so in all other circumstances as it does not uniquely model all the dependencies between variates.

The Chakak and Koehler (1995) construction technique is the simplest and gave satisfactory model results (see Section 4.4 and Fig. 6). The fully nested hierarchical technique yielded the smallest Chi-squared statistic and is relatively simple to apply. The conditional mixtures approach is the most difficult construction technique of the three. The integral (Eq. (A.3)) cannot be solved explicitly and a numerical solution can be demanding. The Chakak and Koehler (1995) approach is the most appealing from a practical point of view because it is so easy to apply.

The Clayton copula best represents the case study data set and satisfies the storm model described by De Michele et al. (2007) (following Boccotti, 2000) where it is assumed that the lower tail has a positive correlation and that the association converges towards the limit of \(-1\) in the upper tail. However, we have hypothesized that there is a return to positive correlations for the most extreme events. As we approach a physical limit to wave height and storm duration the only way the storm magnitude can increase is by extreme waves coinciding with extreme storm durations. This will result in a positive correlation in the upper tail. The Clayton copula is not appropriate for such a model but copula 12 from Nelsen (2006) can model both upper and lower tail dependencies, but has the worst fit to our case study data. Despite this, Fig. 8 shows that the simulation results for copula 12 appear to have an appropriate scatter when compared to the empirical data. Fig. 9 illustrates the changes in the simulations using copula 12 for cases where the dependence parameter increases (i.e., the correlation approaches 1). Fig 9b shows that even with a strong correlation copula 12 still models a weaker association about the centre.

It is likely that most data sets will be modeled best by the Clayton copula. This is because most data sets would not have enough upper tail data to describe the positive correlation and so we suggest that the Clayton copula would only be modeling the lower tail and the central bulk of the distribution. We suspect that a more substantial data set including numerous extreme events could be more appropriately modeled by copula 12.

5.2. The model limitations

This paper aims to identify appropriate techniques for creating multivariate statistical models. In its current form the model has numerous limitations, some of which can be improved. Firstly the model has no link to physical meteorological forcing processes. On the east coast of South Africa storm waves are generally produced by either tropical cyclones, cut-off lows or cold fronts. Tropical cyclones are most frequent in February and cut-off lows in March. These events behave similarly in terms of wave production with the exception of wave direction. Tropical cyclones generally produce north easterly swell while cut-off lows produce south easterly swell. Although 93% of all the South African east coasts’ offshore wave heights exceeding 3.5 m fall between south and east, inclusion of the wave direction would improve the model. Between 1962 and 2005 only seven cyclones affected the eastern parts of South Africa (Krugler et al., 2010). Assuming an average of 3 wave events per year means that cyclones only account for 5% of the wave events. Similarly, only 4% of the case study’s wave data is produced by cyclones of which the wave characteristics are very similar to those produced by cut-off lows. The statistical model is therefore modeling a mixture of cold fronts and cut-off lows. The cut-off lows form further out to sea than the cold fronts. The cut-off lows are associated with the large wave heights and wave periods while the cold fronts produce the smaller wave heights and shorter wave periods.

Seasonality affects the frequency and intensity of events and the exclusion of these effects is a major limitation of the model as discussed here. However as explained in Section 1 this can be included by using non-stationary probability distributions — an example of this approach is given by Corbella and Stretch (2012a).

The model also has no physical constraints such as maximum wave steepness or a maximum water depth at wave breaking. The model is purely statistical and could therefore produce a wave height that could not exist at the given water depth. This however is not a major concern as the simulation results may be conditioned to restrict physically impossible events.

The statistical modeling discussed herein has been restricted to Archimedean copulas that are not as flexible as some other classes of copulas but are relatively easier to apply. As demonstrated in this paper relatively simple techniques can be used to produce multivariate

---

### Table 3

<table>
<thead>
<tr>
<th>Nesting</th>
<th>Copula</th>
<th>Kendall’s tau</th>
<th>Dependence parameter ((\rho))</th>
<th>Goodness of fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C(h,t)) &amp; (d)</td>
<td>Gumbel-Hougaard</td>
<td>0.489</td>
<td>1.90</td>
<td>1.02</td>
</tr>
<tr>
<td>(C(h,d)) &amp; (t)</td>
<td>Clayton</td>
<td>0.244</td>
<td>0.646</td>
<td>0.742</td>
</tr>
</tbody>
</table>

---

---

### Fig. 3

Q-Q plots of the best fitting copulas to the pairs (a) HD; (b) HT; (c) DT.
distributions from hierarchical Archimedean copulas. These techniques often fail for other copula classes and more complicated techniques are required to combine 3 or more copulas. In addition a parametric estimate of $K$ can easily be found for Archimedean copulas (Genest and Rivest, 1993) while other copula classes may require more complicated techniques (see e.g., Genest and Rémillard, 2008; Kojadinovic and Yan, 2011). In summary, although Archimedean copulas have limitations they do provide a relatively simple means of producing multivariate models that are superior to the classical techniques used in engineering.

6. Conclusions

In this paper we have explored the construction of multivariate statistical models for describing sea storms. The inter-dependencies of sea storm parameters have been analyzed. Only the parameters $H$, $T$ and $D$ displayed a significant interdependence all of which were positively correlated. The dependence between $H$ and $D$ was the strongest, followed by $H$ and $T$. Three construction techniques were investigated for the creation of a trivariate copula. From a theoretical perspective the conditional mixtures approach provides the most complete model with no limitations regarding dependencies. The Chakak and Koehler (1995) method is the most appealing from a practical point of view as it provides similar results to the conditional mixtures, only requires 2-copulas to create a 3-copula, and avoids solving the complicated integral produced by the conditional mixtures. The fully nested method of creating hierarchal copulas provided the best results for our case study. This method does not model all the dependencies uniquely and the reason why it provided the best results is largely a function of the case study data set. The results highlight the importance of doing a thorough analysis of the bivariate data prior to creating a multivariate model because a simple construction technique may be appropriate and allow one to avoid the difficulties associated with the more complete conditional mixtures.

We have extended De Michele et al. (2007)’s bivariate analysis of sea storm data. De Michele et al. (2007) considered the dependencies of wave height and duration as a function of storm magnitude, while we have included an analysis of wave height and period as a function of peak wave power. From these bivariate analyses we have proposed an extension to the sea storm model suggested by Boccotti (2000). We propose that there are three levels of storm magnitude and peak wave power.
power: the lowest levels display positive associations, the middle levels negative associations, and as storms approach their physical limits there is another positive correlation. This concept can be modeled well by copula 12 from Nelsen (2006), which can mimic both the lower and upper tail dependencies. The Boccotti (2000) model is well modeled by the Clayton copula.

Fig. 6. Simulations of wave height and storm duration as marginals $F(H)$ and $F(D)$ (top plot) and physical parameters $H$ and $D$ (bottom plot) from a trivariate copula (a) constructed by Chakak and Koehler (1995) and (b) constructed by the conditional mixtures. The simulated data is shown by the gray dots and the empirical data is shown by the black dots.

Fig. 7. Simulations of wave height and storm duration as marginals $F(H)$ and $F(D)$ (top plot) and physical parameters $H$ and $D$ (bottom plot) from a trivariate hierarchical Clayton. The simulated data is shown by the gray dots and the empirical data is shown by the black dots.
In order to fully exploit the realism of numerical models the storm event inputs must be equally realistic. Statistical models and copula methods can provide appropriate tools to simulate realistic events for forcing numerical models of coastal and marine processes.

Appendices

The appendices provide the mathematical expressions for the three construction techniques described in Section 2.2 and the simulation algorithm used in this study.

Appendix A. Multivariate copula construction techniques

The Chakak and Koehler (1995) method uses the bivariate conditional distribution of \((X,Y)\), given that \(Z = w\), to create a trivariate copula

\[
C_{XZ}(u,v,w) = wC_{XY}(u,v,w) + C_{YZ}(v,w) - C_{XY}(u,v) - C_{XZ}(u,w) + C_{XY}(u,v) + C_{XZ}(u,w) - C_{XZ}(w,v) - C_{XZ}(u,v) + C_{XZ}(u,w) + C_{XZ}(w,v)
\]

where \(u,v,w \in [0,1]\). The method has the compatibility problem and there is no guarantee that

\[
wC_{XY}(u,v,w) + C_{YZ}(v,w) = vC_{XY}(u,v) + C_{XZ}(v,w).\]

The conditional mixtures approach is conceptually similar to that of Chakak and Koehler (1995) and was used by Salvadori et al. (2007), Joe (1997) and De Michele et al. (2007) to combine two 2-copulas to form a 3-copula. The three dimensional distribution can be obtained from the conditional distributions by

\[
F_{XZ}(x,y,z) = \int_{-\infty}^{x} C_{XZ}(k(t), F_{X|Y}(k(t))) F_{Y}(dt)
\]

where \(F_{X|Y}, F_{Y|Z}, F_{Z|X}\) are the conditional distributions and can be found from the fitted two copulas

\[
F_{(X|Y)(Y|Z)} = \int F_{(X|Y)}(y|x)F_{(Y|Z)}(z|y)\,dy
\]

A similar expression exists for \(F_{Z|X}\). The notation can be rather confusing but the conditional distributions are simply the partial derivatives of the relevant 2-copulas. Note that \(C_{YZ}\) is basically a measure of conditional dependence between \(X\) and \(Z\), given the behavior of \(Y\). Generally an analytic solution of the integrals cannot be found and a numerical method has to be employed.

The construction of a multi-level hierarchical Archimedean copula is conceptually simple but computationally and notationally challenging (Savu and Trede, 2006). To make the notation slightly less confusing the operation \(\circ\) is used to indicate the composition of functions such that \(\varphi_{n-1}\circ\varphi_{n-2}(y) = \varphi_{n-1}(\varphi_{n-2}(y))\). The \(n\)-dimensional copula for the fully nested case requires \((n-1)\) generators, \(\varphi_{1}, \ldots, \varphi_{n-1}\),

\[
C(u_1, \ldots, u_n) = \varphi_{n-1}^{-1}(\varphi_{n-1}^{-1}(\ldots(\varphi_{2}^{-1}(\varphi_{1}^{-1}(u_1) + u_1)) + \varphi_{2}^{-1}(u_2)) \ldots + \varphi_{n-1}^{-1}(u_{n-1}) + u_{n-1})).
\]

To make the method clearer consider the 3-dimensional copula \(C(u,v,w)\). It can be written as a fully nested hierarchical copula \(C(C(u,v),w)\), assuming the same dependence parameter. Writing the copulas in terms of their generating functions \(\varphi\) (Eq. (1)) would produce an equation similar to Eq. (A.5).

Although the partially nested method is more flexible than the fully nested method, it is not possible to create a 3-copula since the lowest dimension is \(n=4\), whence

\[
C(u_1, \ldots, u_4) = \varphi^{-1}_{1}(\varphi^{-1}_{2}(\varphi^{-1}_{3}(\varphi^{-1}_{4}(u_1) + \varphi_{3}(u_1) + \varphi_{4}(u_1))))
\]

with three generators \(\varphi_1, \varphi_2, \varphi_3\). The conditions required for nesting are satisfied if \(\theta < \theta_{13}\) and \(\theta < \theta_{24}\) (Grimaldi and Serinaldi, 2006; Hofert, 2008, 2011; Savu and Trede, 2006). Only the random variates \(U_1\) and \(U_2\) along with \(U_3\) and \(U_4\) are interchangeable. The major modeling limitation is that the joint distribution of \((U_1, U_2, U_3, U_4)\) is equal to the joint distributions of \((U_2, U_3, U_4)\) and \((U_1, U_2, U_3, U_4)\). The fully nested method provides a better model in this regard since for \(n=4\) dimensions, only \((U_1, U_2)\), \((U_1, U_3)\) and \((U_1, U_4)\) will have the same joint distribution. Generally there are \(n(n-1)/2\) (the number of bivariate marginals) ways to couple \(n\) variables and since there are only \(n-1\) generators only part of all possible mutual dependences will be uniquely modeled.

![Fig. 8](image-url)

**Fig. 8.** Simulations of (a) the wave height marginals \(F(H)\) and the storm duration marginal \(F(D)\) and (b) the wave height and storm duration from the bivariate copula 12 from Nelsen (2006). The simulated data is shown by the gray dots and the empirical data is shown by the black dots.
Appendix B. Simulation

The aim of a simulation is to generate a vector \((U_1, \ldots, U_d)\) whose variables are interdependent and lie in the interval \([0,1]\). Let their joint distribution be a copula \(C(U_1, \ldots, U_d)\). A sample from \(C\) can be simulated by the conditional inversion method (De Michele et al., 2007; Nelsen, 2006; Savu and Trede, 2006, 2010). In general a sample \(u_n\) can be simulated from \(U_n\) based on the conditional law of \(U_n\) given the values \(U_1, \ldots, U_{n-1}\).

\[
C_n(u_n|u_1, \ldots, u_{n-1}) = P(U_n < u_n|U_1 = u_1, \ldots, U_{n-1} = u_{n-1}) = \frac{\partial u_{n-1} \cdots \partial u_1}{\partial u_n} C(u_1, \ldots, u_n, 1, \ldots, 1) \tag{B.1}
\]

for \(n = 2, \ldots, d\). The conditional law \(C_n\) can then be used in the following algorithm to generate \(u_n^*\):

- Simulate \(d\) independent random variables \(t_1, \ldots, t_d\) on \(I\).
- Set \(t_i = u_i\).
- For \(n = 2, \ldots, d\), evaluate the inverse of the conditional distribution function to generate \(u_n = C_n^{-1}(t_n|u_1, \ldots, u_{n-1})\).

Evaluation of the inverse conditional distribution becomes increasingly complicated as more variates are included in the model and can usually only be solved numerically. Other simulation algorithms have been proposed and examples of these can be found in Chebana and Ouarda (2011), Chakak and Koehler (1995), Whelan (2004) and Embrechts et al. (2001).

References


