Proceedings of the ASME 2013 32nd International Conference on Ocean, Offshore and Arctic Engineering
OMAE2013
June 9-14, 2013, Nantes, France

OMAE2013-10154

THE EXTREMAL DEPENDENCE OF STORM SEVERITY, WIND SPEED AND SURFACE LEVEL PRESSURE IN THE NORTHERN NORTH SEA

Ross Towe
Mathematics and Statistics
Lancaster University
Lancaster, United Kingdom

Emma Eastoe
Mathematics and Statistics
Lancaster University
Lancaster, United Kingdom

Jonathan Tawn
Mathematics and Statistics
Lancaster University
Lancaster, United Kingdom

Yanyun Wu
Statistics and Chemometrics
Shell Technology Centre
Chester, United Kingdom

Philip Jonathan
Statistics and Chemometrics
Shell Technology Centre
Chester, United Kingdom

ABSTRACT

Characterising the joint distribution of extremes of significant wave height and wind speed is critical for reliable design and assessment of marine structures. The extremal dependence of pairs of oceanographic variables can be characterised using one of a number of summary statistics, which describe two different types of extremal dependence. Quantifying the type of extremal dependence is an essential pre-requisite to joint or spatial extreme value modelling, and ensures that appropriate model forms are employed for extreme value analysis.

We estimate extremal dependence between storm peak significant wave height and storm peak wind speed (H_s, WS) for locations in a region of the northern North Sea. The extremal dependence itself may vary with storm direction. As a result, we introduce new covariate-dependent forms of the extremal dependence measures that account for the direction of the storm.

We discuss the implications of all of the estimates for marine design, including specification of joint design criteria for extended spatial domains, and statistical downscaling to incorporate the effects of climate change on design specification.

Keywords: covariate effect, extremal dependence, extreme waves, North Sea, statistical model, statistical downscaling and WAM hindcast

1 INTRODUCTION

In standard analysis of environmental data the dependence between two variables can be determined by calculating their correlation, which gives a measure of the dependence of the variables for their entire distribution. However, when interest is in the extreme events, we are no longer interested in the entire joint distribution of a pair of variables. Instead we need to use an alternative dependence measure, which gives us a description of the extremal dependence structure of the two variables. For example in the case of storm severity, we would be interested in the relationship between extreme significant wave height and wind speed [1].

The occurrence of extreme storm peak significant wave height and the simultaneous occurrence of extreme wind speed has had devastating effects. For example, in 1982, the Ocean Ranger semi-submersible drilling rig sank due to the joint occurrence of wave heights reaching 20m combined with 100-knot winds. This incident led to loss of all of the crew members on the rig [2]. As a result, trying to understand the relationship between extreme (significant) wave height and wind speed is of utmost importance.

North Sea winter storm waves are generated by wind forcing, itself a result of differences in atmospheric pressure in the prevailing pressure field. In this paper we look at the relationship between extreme storm peak significant wave height (H_s) and negative surface level pressure (*NSLP*) as well as the relationship between extreme significant wave height and wind speed (*WS*). We consider a wave hindcast dataset for the northern North Sea with the relationship between significant wave height and surface level pressure shown in Figure 1(a) and the relationship between extreme significant wave height and wind speed shown in Figure 1(b). The extremal dependence structure of these data sets is hard to see on their original scale. To better visualise this dependence structure, the data set is transformed to have common marginal distributions, here Fréchet margins. The transformation of two variables *(S,T)* onto Fréchet margins *(X,Y)* is made by using the probability integral transform and is given as follows,

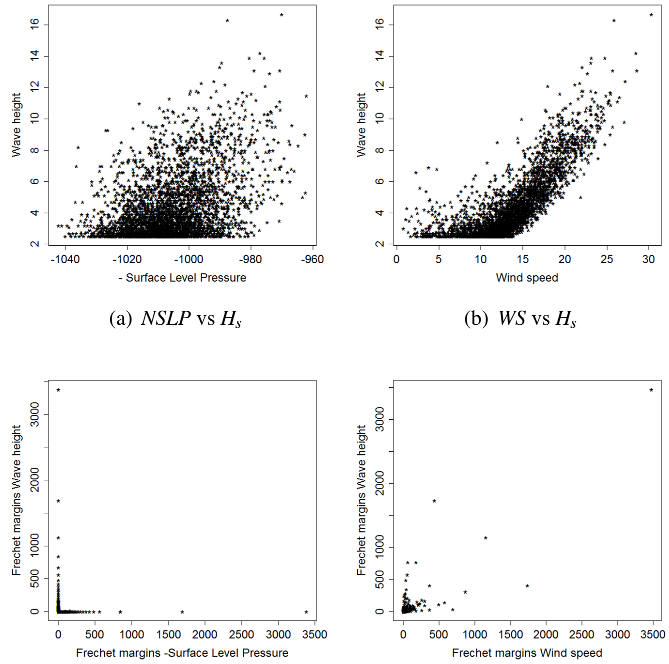
$$X = -\frac{1}{\log\left[F_S(S)\right]} \text{ and } Y = -\frac{1}{\log\left[F_T(T)\right]}, \qquad (1)$$

whereby $F_S(S)$ and $F_T(T)$ are the cumulative distribution functions of the variables S and T respectively. The choice of Fréchet marginal transformation enables us to solely investigate any dependence between *(X,Y)* as they both have common heavy tail marginal distributions [3].

The difference in extremal dependence, which was difficult to see on the original margins in Figure 1(a) and 1(b), is now much clearer. In the case of Figure 1(c) there is weak dependence in the extremes, whereas in the case of Figure 1(d) it is clear that stronger dependence exists in the extremes of *WS* and H_s as the transformed data no longer lie close to the axes suggesting that the largest values of *WS* can occur simultaneously with the largest values of H_s. The two measures $\bar{\chi}$ and χ, which are introduced in Section 2, are a way of quantifying different levels of extremal dependence. In Section 5, it is found that the relevant measure to relate the extremal dependence between *NSLP* and H_s is $\bar{\chi}$, however in the case of *WS* and H_s the appropriate measure to consider is χ.

In previous analysis, it has been assumed that the extremal dependence structure for a pair of variables is stationary. Previous work by [4] showed that storm direction influenced the marginal extremes of H_s. As a result, we want to determine whether storm direction affects the extremal dependence structure of H_s and either *NSLP* or *WS*.

The relationship between pressure differences and H_s may be considered to be of more importance than the relationship between *SLP* and H_s. However, the latter relationship is of more interest to us, as we wish to use this methodology to analyse projections from a global climate model. For the global climate models currently available, the projections of *SLP* are currently much more accurate than the equivalent projections of pressure differences of *SLP*. Furthermore, the pressure differences are an approximation to wind speed and for this particular application the wind speed data are available.



(a) *NSLP* vs H_s (b) *WS* vs H_s (c) *NSLP* vs H_s (d) *WS* vs H_s

FIGURE 1. Storm peak significant wave Height (H_s), negative surface level pressure (*NSLP*) and storm peak significant wave Height (H_s) and wind speed (*WS*) on original (top) and Fréchet margins (bottom) at a typical location in the region

2 DEPENDENCE MEASURES FOR BIVARIATE EXTREMES

The following section derives the two measures of extremal dependence, $\bar{\chi}$ and χ. For the case of asymptotic independence when two extreme events are unlikely to occur simultaneously, we consider the measure $\bar{\chi}$. For the case of asymptotic dependence, when two extreme events are likely to occur simultaneously, we consider the measure χ. The measures $\bar{\chi}$ and χ are interlinked and allow us to determine whether a bivariate data set is asymptotically independent or asymptotically dependent.

Given a random vector (X,Y) with X and Y having common unit Fréchet margins and joint cumulative distribution function $F_{(X,Y)}$, our interest lies in the tail behaviour of $F_{(X,Y)}$. The following model for the joint survivor function was derived in Ledford and Tawn [5]. Interest lies in the joint tail where the components of the random vector (X,Y) exceed their respective $(1-p)^{th}$ quantiles which are $x_p = -1/\log(1-p)$. As $p \to 0$, this can be approximated by,

$$\mathbb{P}\left\{X > x_p \, , \, Y > x_p\right\} \sim \mathscr{L}\left(\tfrac{1}{p}\right) p^{1/\eta} \, , \tag{2}$$

where the function $\mathscr{L}$ is slowly varying at infinity [1]. We adopt the same approach as [5] and assume that the slowly varying function $\mathscr{L}(\frac{1}{p}) = c$, where c is a constant and for p sufficiently close to 0. The other parameter η is defined over the range $0 < \eta \leq 1$ and is termed the coefficient of tail dependence. The coefficient of tail dependence η is the key feature that determines the joint tail behaviour of the random vector (X,Y).

Extremal dependence can be classed as being either asymptotically independent or asymptotically dependent. The former means that extreme events of (X,Y) are very unlikely to occur simultaneously and results in a value of $\eta < 1$. If asymptotic dependence is observed it means that if X is extreme then it is also possible for Y to be simultaneously extreme and results in a value of $\eta = 1$. The following reparameterisation of η is used to characterise the level of asymptotic independence,

$$\bar{\chi} \; = \; 2\eta - 1. \tag{3}$$

Since $\eta \in (0,1]$, it follows that $-1 < \bar{\chi} \leq 1$. Different values of $\bar{\chi}$ determine the level of asymptotic independence,

$$\bar{\chi} = \begin{cases} 1 & \text{if asymptotically dependent;} \\ (0,1) & \text{if asymptotically independent with positive association;} \\ 0 & \text{if independent;} \\ (-1,0) & \text{if asymptotically independent with negative association;} \end{cases} \tag{4}$$

If the estimate of $\bar{\chi}$ is found to be not equal to one, then $\bar{\chi}$ is sufficient in describing the extremal dependence. However for the case where $\bar{\chi} \approx 1$, there is a second measure χ, which determines the strength of extremal dependence [5].

For the random vector (X,Y), the asymptotic dependence measure χ is defined to be,

$$\chi \; = \; \lim_{p \to 0} \mathbb{P}\left\{X > x_p \middle| Y > x_p\right\}.$$

[1] In order for a function ℓ to be slowly varying at infinity, the following convergence limit has to hold for $t > 0$

$$\frac{\ell(tr)}{\ell(r)} \; \to \; 1 \text{ as } r \to \infty.$$

By applying the Ledford & Tawn model in equation (2), it follows that,

$$\chi \; = \; \begin{cases} c \;\text{ if } \lim_{p\to 0} \mathscr{L}(\frac{1}{p}) = c > 0 \;\text{ and }\; \bar{\chi} = 1 \\ 0 \;\text{ if } \lim_{p\to 0} \mathscr{L}(\frac{1}{p}) = 0 \qquad \text{or} \quad \bar{\chi} < 1 \end{cases}, \tag{5}$$

where c is a constant and determines the strength of asymptotic dependence of the random vector (X,Y). If χ is equal to 0, then the variables are asymptotically independent and $\bar{\chi}$ is the relevant measure of dependence.

3 EXISTING METHODOLOGY TO ESTIMATE $\bar{\chi}$ AND χ

We follow the methods in [5] to estimate $\bar{\chi}$ and χ. Consider the variable Z, where $Z = \min(X,Y)$, otherwise known as the structure variable. By the application of the Ledford & Tawn model in equation (2), it follows for a high enough threshold u, that for $z > u$, the distribution of Z is approximately,

$$\begin{aligned} \mathbb{P}(Z > z) &= \mathbb{P}(\min(X,Y) > z) \\ &= \mathbb{P}(X > z \, , \, Y > z) \\ &\approx cz^{-1/\eta}. \end{aligned} \tag{6}$$

In order to make inferences about c and η using the distribution in equation (6), the censored likelihood approach of [5] is adopted. For a given data set $(x_1,y_1),\ldots,(x_n,y_n)$, we calculate the structure variables $z_1,\ldots,z_n$ and let $z_{(1)},\ldots,z_{(n_u)}$ be the n_u values that are defined above the threshold u. The set of variables $z_1,\ldots,z_{(n_u)}$ have cumulative distribution function F_z and corresponding probability density function f_z. For a large enough threshold u, the likelihood function is defined in terms of the constant c and coefficient of tail dependence η, and is as follows,

$$\begin{aligned} L(z_1,\ldots,z_n;\eta,c) &= \prod_{i=n_u+1}^{n} F_z(u) \prod_{i=1}^{n_u} f_z(z_{(i)};\eta,c) \\ &= F_z(u)^{n-n_u} \prod_{i=1}^{n_u} f_z(z_{(i)};\eta,c) \\ &= \left[1 - \frac{c}{u^{\frac{1}{\eta}}}\right]^{n-n_u} \prod_{i=1}^{n_u} \left[\frac{c}{\eta} z_{(i)}^{-\left(\frac{1}{\eta}+1\right)}\right]. \end{aligned} \tag{7}$$

By maximising equation (7), the maximum likelihood estimate of $\bar{\chi}$ is obtained by substituting the maximum likelihood estimate, $\hat{\eta}$ into equation (3) to give,

$$\hat{\bar{\chi}} = \frac{2}{n_u}\left(\sum_{i=1}^{n_u}\log\left[\frac{z_{(i)}}{u}\right]\right) - 1.$$

The estimate $\hat{\bar{\chi}}$ has the following asymptotic variance,

$$\text{var}(\hat{\bar{\chi}}) = \frac{(\hat{\bar{\chi}}+1)^2}{n_u}.$$

Of course due to sampling variability, $\bar{\chi}$ is unlikely to be exactly equal to one under asymptotic dependence, so to determine whether $\bar{\chi}$ is significantly less than one [6], we can perform a generalised likelihood ratio test $\bar{\chi} = 1$ v $\bar{\chi} < 1$ based on equation (3). In the case when $\bar{\chi} = 1$ is not rejected then the data are deemed to be asymptotically dependent and so χ is the appropriate measure. To estimate χ the censored likelihood given in equation (7), assuming that $\eta = 1$ is used. This produces a likelihood for the constant $c > 0$,

$$\begin{aligned} L(z_1,\ldots,z_n;c) &= \prod_{i=n_u+1}^{n} F_z(u)\prod_{i=1}^{n_u} f_z(z_{(i)};c) \\ &= \prod_{i=n_u+1}^{n}\left[1-\frac{c}{u}\right]\prod_{i=1}^{n_u}\left[\frac{c}{z_{(i)}^2}\right]. \end{aligned} \quad (8)$$

Through standard maximisation techniques and using equation (5), the maximum likelihood estimate for χ is,

$$\hat{\chi} = \frac{un_u}{n}.$$

This estimate has asymptotic variance,

$$\text{var}(\hat{\chi}) = \frac{u^2 n_u(n-n_u)}{n^3}.$$

These maximum likelihood estimates of $\bar{\chi}$ and χ and their respective asymptotic variances are threshold dependent. Therefore, careful consideration has to be given to the choice of threshold [7].

4 ESTIMATION OF $\bar{\chi}$ AND χ WITH COVARIATES

In the previous sections it was assumed that the asymptotic dependence between a single pair of concurrent observations (X_i, Y_i) was the same for all i. However for our data the level of extremal dependence is expected to depend on storm direction θ. We now discuss how to model $\bar{\chi}$ and χ in this case by assuming that the covariate dependent estimates of $\bar{\chi}$ and χ vary as a smooth function of direction.

4.1 Non stationary estimation of $\bar{\chi}$

To incorporate covariates into c and η, we use a Fourier based approach. The covariate forms of c and η given in equation (9) are modelled as a sum of the first d terms in a Fourier series of direction,

$$\begin{aligned} c(\theta) &= \exp\left(c_0 + \sum_{i=1}^{d} c_{1i}\cos(c_{2i}+i\theta)\right), \\ \eta_+(\theta) &= \exp\left(\alpha_0 + \sum_{i=1}^{d}\alpha_{1i}\cos(\alpha_{2i}+i\theta)\right), \\ \eta(\theta) &= \min\{\eta_+(\theta),1\}, \end{aligned} \quad (9)$$

with $0 \le c(\theta)$ due to the using of the link function, $\theta \in [0, 360°)$ and $0 \le \eta(\theta) \le 1$.

The censored likelihood in equation (7) is extended to incorporate the covariates defined in equation (9) and the new covariate dependent likelihood is given below,

$$\begin{aligned} & L(z_1,\ldots,z_n,\theta_1,\ldots,\theta_n;\eta,c) \\ &= \prod_{i=n_u+1}^{n} F_z(u;\eta(\theta_{(i)}),c(\theta_{(i)}))\prod_{i=1}^{n_u} f_z(z_{(i)};\eta(\theta_{(i)}),c(\theta_{(i)})) \\ &= \prod_{i=n_u+1}^{n}\left[1-\frac{c(\theta_{(i)})}{u^{\frac{1}{\eta(\theta_{(i)})}}}\right]\prod_{i=1}^{n_u}\left[\frac{c(\theta_{(i)})}{\eta(\theta_{(i)})} z_{(i)}^{-\left(\frac{1}{\eta(\theta_{(i)})}+1\right)}\right], \end{aligned} \quad (10)$$

where $\theta_{(i)}$ is the direction of $z_{(i)}$. The number of Fourier terms in equation (9) is determined by using the AIC (Akaike's Information Criterion) [8].

4.2 Non stationary estimation of χ

In order to incorporate covariates into the measure of asymptotic dependence χ, the likelihood given in equation (10) is adopted under the assumption that $\eta(\theta_{(i)}) = 1$ for all values of θ. However we no longer fit the functional form of $c(\theta)$ given in equation (9) but instead a logistic link function to relate $c(\theta)$ to the linear predictor,

$$c(\theta) = \frac{\exp\left(\beta_0 + \sum_{i=1}^{d}\beta_{1i}\cos[\beta_{2i}+i\theta]\right)}{1+\exp\left(\beta_0+\sum_{i=1}^{d}\beta_{1i}\cos[\beta_{2i}+i\theta]\right)},$$

where $\theta \in [0, 360^\circ)$. Using a logistic link function provides an alternative way to ensure that the estimates of $c(\theta)$ lie between 0 and 1. Furthermore the chance of estimating $c(\theta) = 1$ is very unlikely as in reality the probability of observing perfect dependence between two variables is very unlikely. If no directional effect is present $c(\theta) = \exp(\beta_0)/[1+\exp(\beta_0)]$ over all values of θ. This result gives a constant level of asymptotic dependence for all pairs of (X_i, Y_i). Using the likelihood given in equation (10) and assuming that $\eta(\theta) = 1$, the likelihood for χ given in terms of c is as follows,

$$
\begin{aligned}
L(z_1,\ldots,z_n,\theta_1,\ldots,\theta_n;c) &= \prod_{i=n_u+1}^{n} F_z(u) \prod_{i=1}^{n_u} f_z(z_{(i)}; c(\theta_{(i)})) \\
&= \prod_{i=n_u+1}^{n} \left[1 - \frac{c(\theta_{(i)})}{u}\right] \prod_{i=1}^{n_u} \left[\frac{c(\theta_{(i)})}{z_{(i)}^2}\right].
\end{aligned}
\quad (11)
$$

Then the estimate of $\chi(\theta)$ is equal to $\hat{c}(\theta)$ for a sufficiently large threshold u. To determine the number of Fourier terms needed in the model, the AIC was used [8].

4.3 Uncertainty intervals for $\bar{\chi}$ and χ

Uncertainty in the estimates for $\bar{\chi}$ and χ can be determined by using a non-parametric bootstrap. This non-parametric bootstrap means that we first sample with replacement from our original data set. When sampling, vectors (X_i, Y_i, θ_i) are sampled together, to produce a bootstrap sample with the same dependence structure as the original data set. From this sample, we then estimate the parameters given in either equation (10) or equation (11). Then from these estimates, we calculate the estimate of either $\bar{\chi}(\theta)$ or $\chi(\theta)$ for all the possible values of θ.

We repeat this procedure m times to give us m estimates of either $\bar{\chi}(\theta)$ or $\chi(\theta)$. For each value of θ we determine the 2.5% and 97.5% quantiles of the m estimates to give the 95% pointwise confidence intervals.

5 APPLICATION

5.1 Data set

Data were obtained from the WAM-Hindcast data set [9], with three hourly observations in the North Sea from 1957-2009, over an area of 175 by 350km. The oceanographic variables available are, significant wave height (H_s) [m] mean surface level pressure [hPa (Pascal)] wind speed in 10m above sea level (WS) [m/s] wind direction [degrees] and mean wave direction [degrees] with $\theta = 0^\circ$ corresponding to the North. The majority of severe storms will come from the Atlantic Ocean due to the longer fetches, we will focus on four sites with interesting local characteristics. A brief description of these four sites is as follows along with their exact location given in Figure 2.

Site A is the furthest south of all of the four sites, and is most likely to be influenced by storms coming from the southern North Sea.

Site B is the closest site to the Shetland Islands and as a result this land shadow will reduce the severity of storms. The most severe storms will then be expected to come from the southern North Sea.

Site C is situated in the north east of the region and is expected to be strongly influenced by Atlantic storms. These Atlantic storms dominate the region as they have much longer fetches and produce the largest waves observed in the northern North Sea.

Site D is in the north west of the region and will also be affected by the Atlantic storms but should also see the effects of storms coming south the Norwegian Sea.

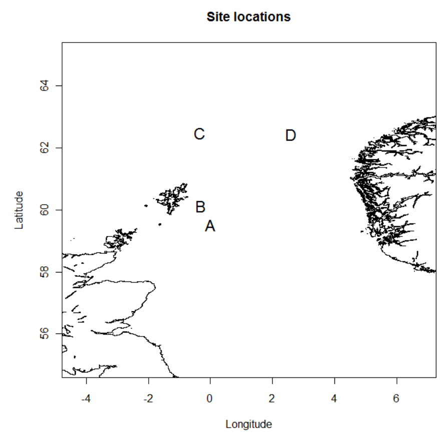


FIGURE 2. Locations of Sites A-D in the North Sea

5.2 Storm peak algorithm

In order to ensure that each storm peak event is independent in time, the following algorithm is deployed.

For each site in the data set, the time series of significant wave height $H_{s(t)}$, wind speed $WS_{(t)}$, mean surface level pressure $SLP_{(t)}$ and wave direction $WvD(t)$ are selected, where $t = 1,\ldots,m$ and m is the number of observations.

For the time series $H_{s(t)}$ a suitable threshold v is determined. For comparison across sites, it is sensible to have a quantile threshold rather than a specific value. In this particular case a 50% quantile threshold for v was chosen. Once a threshold v has been defined, points are identified whether they are $H_{s(t)} > v$ or $H_{s(t)} < v$. An up-crossing is defined to be the time at which H_s goes above v. Similarly a down-crossing is defined to be the time at which H_s goes below v. The time period between consecutive up- and down-crossings is known as the storm period, the process of defining the up-crossing and down-crossings is repeated for the entire time series and produces y storm periods. In order to ensure independence between storm peaks, a time window of length $\tau = 16$ (2 days) is defined. If two storm peaks y_a y_b are within time τ of one another, the two storm periods are combined with the largest storm peak selected.

Once independent storm periods have been determined, the maximum value of H_s in each storm period is recorded, as well as the associated values of the covariates of interest. If multiple occurrences of the maximum value of H_s are observed, the average of the covariates at these time points are taken.

Henceforth NSLP, WS and H_s will refer to observations of storm peak NSLP, storm peak WS and storm peak H_s selected using the storm peak selection algorithm.

5.3 Extreme significant wave height and surface level pressure

Given that the lowest values of surface level pressure correspond to the highest values of significant wave height, we consider *NSLP* and use the marginal transformation given in equation (1). When carrying out the marginal transformation, the marginal cumulative distribution $F_S(S)$ and $F_T(T)$ are replaced with their empirical estimates $\hat{F}_S(S)$ and $\hat{F}_T(T)$ respectively.

No direct relationship exists between surface level pressure and extreme significant wave height, as *NSLP* generates wind speed and changes in wind speed generate the largest waves. Due to no clear relationship existing, we would expect them to be asymptotically independent.

5.3.1 Directionally independent estimates As an initial analysis, for each of sites A-D given in Section 5.1, the measure $\bar{\chi}$ is given in Table 1. Estimates were found by using a 80% quantile threshold u in likelihood (10). From Table 1, it is clear that the estimates of c are consistent across sites, however slight variations can be seen in the estimates of $\bar{\chi}$. The values of $\bar{\chi}$ are of particular interest as they are significantly different from either -1 or 1, thus we conclude that the data are asymptotically independent and in this case with positive association. This conclusion is consistent with physical intuition. The next question to determine is whether or not the level of asymptotic independence varies with storm direction.

Site	$\hat{c}$ (95% *CI*)	$\hat{\bar{\chi}}$ (95% *CI*)
A	0.55 (0.50,0.60)	0.61 (0.50,0.72)
B	0.54 (0.49,0.59)	0.61 (0.50,0.72)
C	0.57 (0.51,0.63)	0.59 (0.47,0.70)
D	0.53 (0.48,0.58)	0.65 (0.53,0.77)

TABLE 1. Estimates of c and $\bar{\chi}$ for the 4 sites with the threshold set at the respective 80% quantile estimation threshold

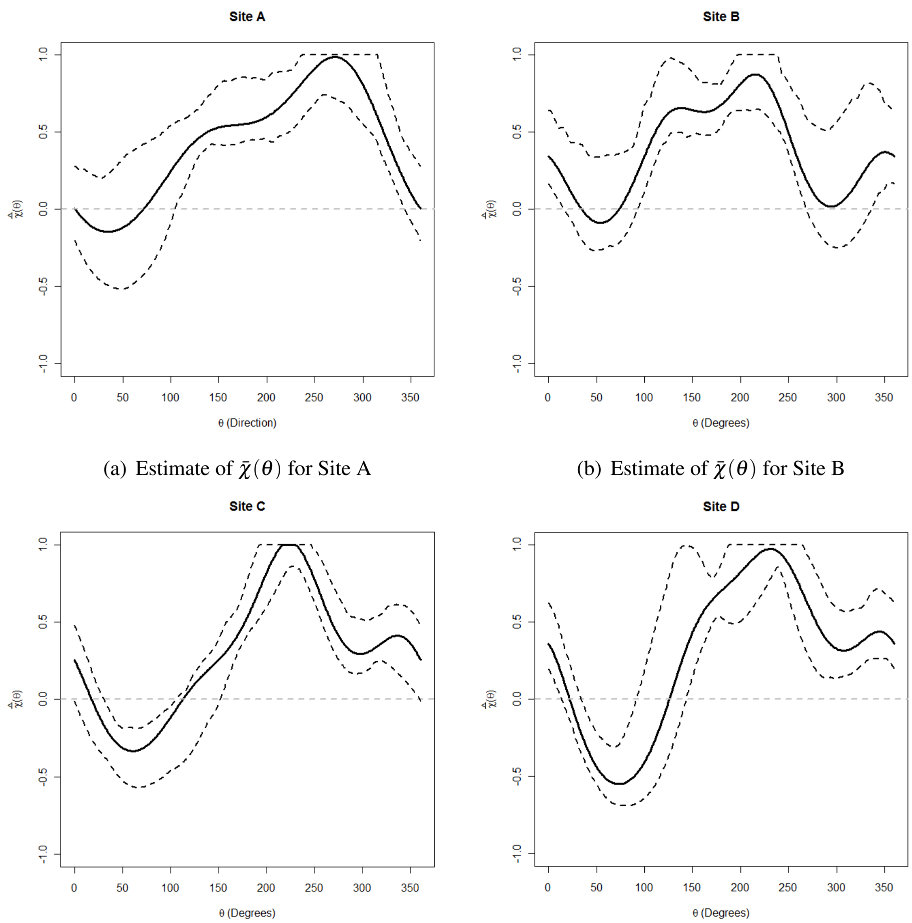
5.3.2 Directionally dependent estimates The model given in equation (10) was extended to include further Fourier terms. The estimates are plotted in Figures 3(a) to 3(d), along with their respective 95% pointwise confidence intervals.

From Figure 3(a) to 3(d), it can be seen that a different number of Fourier terms are needed for the different sites. In the case of sites A and B, two Fourier terms are sufficient, and in both cases the most uncertainty in the estimates lies between $(0, 180^\circ)$ and $(300, 360^\circ)$. For sites A and B this would correspond respectively to storms coming from the Arctic Ocean and the dampening of the Shetland land shadow on severe storms. All of the estimates of $\bar{\chi}(\theta)$ are significantly different from one apart from the sector corresponding to severe storms coming from the Atlantic Ocean. This is expected as severe storms that arise in the Atlantic dominate the region and pressure fields stretch over a large spatial area thus supporting the similarities between the estimates of $\bar{\chi}(\theta)$ at the sites. For the rest of the directional sectors in Figure 3(a) to 3(d), the estimate of $\bar{\chi}(\theta)$ ranges from asymptotic independence with positive association, independence and asymptotic independence with negative association. The asymptotic independence with positive association supports the physical processes as the lowest values of *NSLP* correspond to the highest values of H_s, and furthermore the asymptotic independence is expected as no direct relationship exists between *NSLP* and H_s. In the cases where *NSLP* and H_s are found to be asymptotically independent with positive association this corresponds to severe storms that have arisen from the Arctic ocean.

5.4 Extreme significant wave height and wind speed

The previous combination of oceanographic variables illustrated the idea of asymptotic independence. A more prominent relationship of interest to metocean engineers, is the relationship between extreme significant wave height and wind speed. Large wind speeds are known to be a key driver of storm severity, a consequence of these are extreme wave heights.

5.4.1 Directionally independent estimates Given that such a strong relationship exists between extreme significant wave height and the corresponding values of wind speed, we expect them to be asymptotically dependent. However, to check this assumption, estimates of $\bar{\chi}$ are first calculated for the four sites and are given in Table 2. From Table 2, the estimates of $\bar{\chi}$ are very close to 1. This suggests there is evidence that extreme significant wave height and wind speed are asymptotically dependent. Consequently estimates of χ are also taken for each site, these are also given in Table 2. It is clear that the maximum likelihood estimates of χ are significantly different from zero, suggesting that extreme significant wave height and wind speed are asymptotically dependent and at a strong level, which emphasises how closely related they are in determining storm severity. Interestingly, the estimates are consistent across the



(a) Estimate of $\bar{\chi}(\theta)$ for Site A (b) Estimate of $\bar{\chi}(\theta)$ for Site B

(c) Estimate of $\bar{\chi}(\theta)$ for Site C (d) Estimate of $\bar{\chi}(\theta)$ for Site D

FIGURE 3. Estimates of $\bar{\chi}(\theta)$ with Fourier term and bootstrapped pointwise confidence intervals for each of the 4 sites plotted against direction (θ). The maximum likelihood estimates for each of the sites is given by the black solid line and the 95% bootstrapped pointwise confidence intervals are given by the dashed black lines.

northern North Sea, this supports physical insight as wind speed is a key indicator of storm severity and wind speed is determined by changes in pressure fields that stretch over a large spatial area rather than a single location.

Site	$\hat{\bar{\chi}}$ (95% *CI* $\hat{\bar{\chi}}$)	$\hat{\chi}$ (95% *CI* χ)
A	0.95(0.81,1.00)	0.73(0.69,0.79)
B	0.86(0.73,0.98)	0.74(0.69,0.78)
C	0.85(0.71,0.99)	0.73(0.68,0.78)
D	0.84(0.71,0.97)	0.72(0.67,0.77)

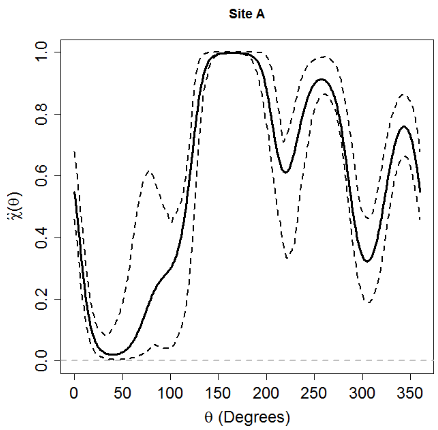
TABLE 2. Estimates of ($\bar{\chi}$ and χ) for the four sites with the threshold set at the respective 80% quantile estimation threshold

5.4.2 Directionally dependent estimates

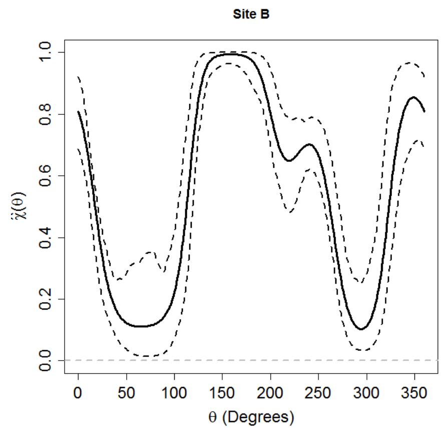
The estimates of χ given in Table 2 are consistent across sites, suggesting that one summary measure would be sufficient to describe the level of asymptotic dependence across sites. However, it is known that each of the four sites have very different local characteristics particularly in terms of where the most severe storms come from. These different environmental characteristics should be more evident once direction has been introduced as a covariate.

The model given in equation (5) was extended to include Fourier terms. It was found that for each of the four sites, three Fourier terms were sufficient to describe the directional structure of $\chi(\theta)$. The estimates are plotted in Figures 4(a) to 4(d), along with their respective 95% pointwise confidence intervals.

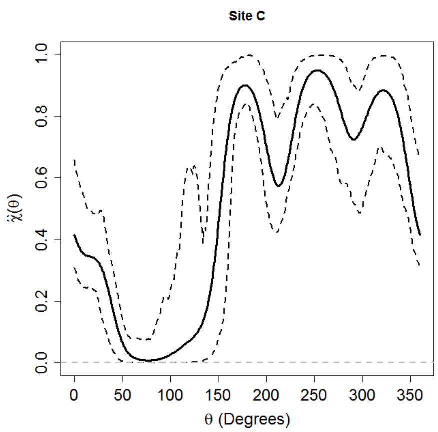
Through the addition of the extra Fourier terms it is apparent that each of the sites have very different spatial characteristics.



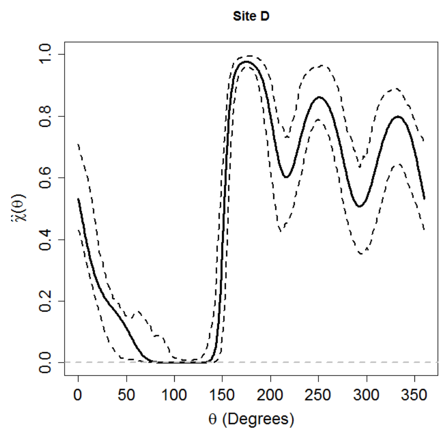
(a) Estimate of $\chi(\theta)$ for Site A



(b) Estimate of $\chi(\theta)$ for Site B



(c) Estimate of $\chi(\theta)$ for Site C



(d) Estimate of $\chi(\theta)$ for Site D

FIGURE 4. Estimates of $\chi(\theta)$ with 3 Fourier terms and bootstrapped pointwise confidence intervals for each of the four sites plotted against direction (θ). The maximum likelihood estimates for each of the sites is given by the black solid line and the 95% bootstrapped pointwise confidence intervals are given by the dashed black lines.

For each of the figures, the effects of the directional sectors are detected and the effects vary depending on the local characteristics of the site. For all sites, apart from site B, the minimum point is observed from 45° to 90°. These directions correspond to severe storms coming from Norway, which are rarely observed at site B and instead is dominated by severe storms from the south. The sites' proximity to the Shetland land shadow is also apparent in Figure 4(b) due to the estimate of χ only being about 0.1. The small width of the tolerance interval either side of the direction of $\theta = 180^\circ$ reflects the fact that the site is dominated by storms from the southern North Sea. For sites A, B and D, the main directional sectors have been clearly identified to be storms coming from the southern North Sea, Atlantic Ocean and the Arctic Ocean with clear differences arising dependent on the location of the specific site. At all four sites there is found to be a strong asymptotic dependence between extreme significant wave height and wind speed that significantly depends upon direction. For site B and D it was found there was evidence for perfect dependence between extreme significant wave height and wind speed that correspond to storms coming from a direction of $\theta = 180^\circ$.

6 CONCLUDING REMARKS

The asymptotic extremal dependence measures $\bar{\chi}$ and χ are vital summary statistics when modelling bivariate extremes. We have extended these measures to include a third variable direction to see whether the extremal dependence measures change with direction. We have used $\bar{\chi}$ and χ to provide us with information about the asymptotic relationship between oceanographic variables such as storm peak significant wave height, wind speed and surface level pressure that are known to be key influences of storm severity. It was found that extreme storm peak significant wave height and the corresponding values of surface level pressure are asymptotically independent. In contrast, extreme storm peak significant wave height and the corresponding wind speed were found to be asymptotically dependent. These conclusions are consistent with our physical understanding; pressure differences cause winds which cause waves. We might then not expect a clear relationship to exist between significant wave height and surface level pressure, whereas the relationship between wind and waves should be clearer. High wind speeds cause the largest waves. Therefore, we would expect them to be asymptotically dependent.

As previous authors [4] had found storm direction is an important covariate in the marginal behaviour of H_s, we thought that storm direction may also affect the extremal dependence structures. As a result, the extremal dependence measures $\bar{\chi}$ and χ were extended to include direction as a covariate. For both $\bar{\chi}$ and χ, the covariate was introduced in the form of a Fourier series, which assumes a smooth function for the estimates of $\bar{\chi}(\theta)$ and $\chi(\theta)$. The modelling techniques could easily be extended to include other covariates that are influential to $\bar{\chi}$ or χ. For both measures, it was found that direction was a highly important covariate and the harmonic terms correctly identified the different directional storm sectors in the North Sea. This modelling approach of introducing the covariate direction resulted in very different estimates of the extremal dependence measures. For example, in some cases suggesting that for certain directions the measure changes from being asymptotically dependent to asymptotically independent. These findings motivate the adoption of the conditional extremes approach of [10] for reliable estimation of design criteria, since it is able to handle both asymptotic independence and dependence, and covariate effects [11].

Further research is currently being undertaken which prewhitens a data set before any dependence modelling is performed. The pre-whitening of data set involves removing any marginal effects present in a data set before the joint modelling takes place. This further analysis will help to determine whether the dependence observed in the extremal dependence measures is from marginal effects that have been unaccounted for or whether the measures are genuinely capturing the extremal dependence between a pair of variables [11].

ACKNOWLEDGMENT

This research would have not been possible without funding from Lancaster Doctoral Training centre and Shell. We would like to thank Graham Feld (Shell) in particular for advice concerning the WAM hindcast, and further thank Simon Brown (UK MetOffice), Kevin Ewans (Shell), Nicolas Fournier (Shell) and Jan Kysely (Institute of Atmospheric Physics, Czech Academy of Sciences) for providing us with data and for many illuminating discussions.

REFERENCES

[1] B. Kinsman, *WIND WAVES*. Prentice-Hall, Inc, 1965. their generation and propagation on the ocean surface.

[2] C. Heising and W. Grenzebach, “The ocean ranger oil rig disaster: A risk analysis,” *Risk Analysis*, vol. 9, pp. 55–62, 1989.

[3] S. Coles and J. Tawn, “Modelling extreme multivariate events,” *J. Roy. Statist. Soc. Ser. B*, vol. 53, no. 2, pp. 377–392, 1991.

[4] P. Jonathan and K. Ewans, “The effect of directionality on extreme wave design criteria,” *Ocean Engineering*, vol. 34, pp. 1977–1994, 2007.

[5] A. Ledford and J. Tawn, “Statistics for near independence in multivariate extreme values,” *Biometrika*, vol. 83, no. 1, pp. 169–187, 1996.

[6] J. Heffernan, S. Coles, and J. Tawn, “Dependence measures for extreme value analyses,” *Extremes*, vol. 2, pp. 339–365, 1999.

[7] S. Coles, *An introduction to statistical modeling of extreme values*. Springer Series in Statistics, London: Springer-Verlag London Ltd., 2001.

[8] H. Akaike, “A new look at the statistical model identification,” *IEEE Trans. Automatic Control*, vol. AC-19, pp. 716–723, 1974. System identification and time-series analysis.

[9] WAMDIGroup, *The WAM Model*. Guideline document, 1988. A Third Generation Ocean Wave Prediction Model.

[10] J. Heffernan and J. Tawn, “A conditional approach for multivariate extreme values,” *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 66, no. 3, pp. 497–546, 2004. With discussions and reply by the authors.

[11] K. Ewans and P. Jonathan, “Evaluating environmental joint extremes for the offshore industry,” *Journal of Marine Systems (submitted October 2012)*, 2013.