



# The effect of serial correlation in environmental conditions on estimates of extreme events

Ed Mackay<sup>a,\*</sup>, Guillaume de Hauteclocque<sup>b</sup>, Erik Vanem<sup>c</sup>, Philip Jonathan<sup>d,e</sup>

<sup>a</sup> College of Engineering, Mathematics and Physical Sciences, University of Exeter, UK

<sup>b</sup> Bureau Veritas, Paris, France

<sup>c</sup> DNV, Høvik, Norway

<sup>d</sup> Department of Mathematics and Statistics, Lancaster University, UK

<sup>e</sup> Shell Research Ltd., London, UK

## ARTICLE INFO

### Keywords:

Long-term extreme  
Extreme response  
Environmental contour  
Short-term variability  
Extremal index

## ABSTRACT

In offshore engineering, it is common practice to estimate long-term extremes under the assumption that environmental conditions are independent. However, many environmental variables, such as winds and waves, exhibit correlation over several days. In this work, we consider the impact that this has on estimates of return values of metocean variables, environmental contours and long-term extreme responses. It is shown that methods which neglect serial correlation over-estimate the size of extreme events at a given return period. We introduce a new definition of a sub-asymptotic extremal index, and show how this can be used to quantify the effect of neglecting serial correlation. Simple examples are presented to illustrate why neglecting serial correlation leads to positive bias. We show how the size of the bias is related to the average shape of storm events and the shape of the tail of the distribution of storm peak values, with the latter having the dominant effect. Storm peak distributions with longer tails lead to larger biases when serial correlation is neglected. In the examples presented, neglecting serial correlation resulted in relative errors of over 50% in the 25-year extreme response estimates in some cases. The examples presented show that accounting for serial correlation in estimates of environmental contours and long-term extreme responses can reduce over-conservatism and result in more efficient designs.

## 1. Introduction

### 1.1. Motivation

Environmental variables such as winds, waves, currents and surges are serially correlated. Conditions can persist over time scales of hours to days. In the statistics literature, the effects of serial dependence on the modelling of extreme values has been discussed widely (see e.g. Leadbetter et al., 1983; Coles, 2001; Beirlant et al., 2004; Chavez-Demoulin and Davison, 2012). In offshore and coastal engineering, there are inconsistencies between the way in which serial dependence is treated in estimates of univariate extremes, multivariate extremes and long-term extreme responses. For univariate extremes, it is common practice to use methods such as annual maxima or peaks-over-threshold (Jonathan and Ewans, 2013), which remove the effects of serial correlation by considering only peak events that are sufficiently separated in time that they are effectively independent. Other methods, such as the average conditional exceedance rate (ACER) method (Naess and Gaidai, 2009), explicitly treat the serial correlation in a univariate

time series. In contrast, multivariate extremes or long-term extreme responses, are often calculated based on the distribution of all observations, under the tacit assumption that sequential observations are independent.

The distribution of all observations is different to distribution of peak events. When the interest is in assessing the probability of structural failure due to an extreme response, the pertinent problem is to estimate how often a storm event will occur in which a given response level is exceeded. If such an event occurs, it does not matter how many times a failure would occur within the storm event. In this case, it is the distribution of peak events that is of interest. However, if we are interested in the total number of times a level is exceeded (e.g. for low-cycle fatigue analysis), then it is the distribution of all observations that is of interest. In this paper, we focus on the problem of estimating the occurrence of extreme events for the purpose of assessing the risk of structural failure.

The main point that we wish to emphasise, is that neglecting serial correlation will result in over-estimates of the occurrence of extreme

\* Corresponding author.

E-mail address: [e.mackay@exeter.ac.uk](mailto:e.mackay@exeter.ac.uk) (E. Mackay).

events. To illustrate this point, consider the following simple example. Suppose we have a 100-year time series of significant wave height,  $H_s$ , in which there is one storm which exceeds  $H_s = 10$  m, and all other storms have  $H_s < 10$  m. Suppose that in the largest storm, there are  $r$  hours which exceed 10 m. If each hour is treated as an independent event, then the empirical estimate would be that  $H_s = 10$  m is exceeded  $r$  times in 100 years, on average. Whereas, in fact, there was only one event in 100 years which exceeds  $H_s = 10$  m. Of course, the actual return period of the largest storm in the record is not known exactly, but the example illustrates the difference between the two interpretations.

It is important to emphasise that neglecting serial correlation in the data does not necessarily lead to bias in the estimate of the distribution of all observations. Rather, the simple example above illustrates that the distribution of all observations is not what is required for estimating the occurrence of extreme events. The distribution of all observations tells us the proportion of time that a level is exceeded. It tells us nothing about how those exceedances are distributed in time, in particular it does not specify the rate of up-crossing of a particular level.

The problem of dependence is also important in the estimation of spatial extremes, and many of the methods to account for spatial dependence are similar to those for temporal (serial) dependence. However, the most common problem in offshore engineering is to estimate extreme events based on a time series of historic observations, so the focus here is on the effect of temporal rather than spatial dependence.

The arguments we present in this work will be familiar to some people in the offshore and coastal engineering communities. However, from various conversations with practitioners involved with estimating both extreme environmental conditions and extreme loads on marine structures, it was clear that the effect of neglecting serial correlation is not widely appreciated in the offshore community. Moreover, a number of widely-used methods for estimating extreme events do not account for serial correlation. For example, environmental contours are usually constructed from the distribution of all observations, without accounting for serial correlation (Haselsteiner et al., 2021). Similarly, estimates of long-term extreme response calculated by integrating the short-term distribution function over the long-term distribution of environmental conditions (e.g. Fogle et al., 2008; Sagrilo et al., 2011; Naess and Moan, 2013), also neglect serial correlation. The various methods for combining short-term response functions with long-term distributions of environmental conditions has been considered by a various authors in the past (e.g. Sagrilo et al., 2011; Forristall, 2008; Mackay and Johanning, 2018c). The current work extends these studies by considering the impact of serial dependence in a range of contexts, including univariate and multivariate problems.

The main aim of this paper is to illustrate why it is important to consider dependence in estimates of extreme events and provide examples which quantify the impact of neglecting serial correlation in practical cases. We also consider the theoretical treatment of serial dependence in the statistical literature, and show how the effect of neglecting serial correlation can be quantified in terms of the extremal index, defined in the following sub-section. The examples are intended to provide a more intuitive understanding of why neglecting serial correlation causes bias, and explain how various factors influence the size of bias.

## 1.2. Statistical treatment of dependence in time series extremes

In the statistics literature, extremes of time-series have been considered by a number of authors over many years. Chavez-Demoulin and Davison (2012) provide a relatively recent review. Informally, the analysis considers a time-series  $X_1, \dots, X_n$  such that the marginal distribution of  $X_i$  is common to all choices of  $i \in [1, 2, \dots, n]$ . It is then assumed that the time-series exhibits serial dependence, with certain conditions. Importantly, the focus of the analysis is typically the estimation of the common marginal distribution  $F$ . As discussed above,

this is not typically the main concern of the ocean engineer interested in estimation of return values and similar quantities.

For the practical analysis of dependent sequences, some constraints are usually assumed, which limit the long-range dependence at extreme levels. A widely used constraint is the so-called  $D(u_n)$  condition (see e.g. Leadbetter et al., 1983; Coles, 2001; Beirlant et al., 2004). Informally, for sequences which meet this condition, the events  $X_i > u$  and  $X_j > u$  are approximately independent, provided that  $u$  is high enough and times  $i$  and  $j$  are sufficiently separated.

The key metric used to quantify temporal dependence in extreme values is the extremal index,  $\theta$ . The extremal index can be defined in a number of equivalent ways. In serially correlated time series, extreme events occur in clusters. Asymptotically, extremes of a stationary sequence can be shown to occur in clusters with mean size  $1/\theta$  (Hsing, 1987a; Hsing et al., 1988). An equivalent way to define the extremal index is in terms of the distribution of the maximum values of a sample. Suppose that  $X_1, \dots, X_n$  are a stationary process with some level of serial correlation, and with marginal distribution function  $F$ . Assuming that  $F$  is in the domain of attraction of an extreme value distribution (see e.g. Coles, 2001), the distribution of the maximum of a sample will tend to a limit  $\Pr(\max\{X_1, \dots, X_n\} \leq x) \rightarrow G(x)$  as  $n \rightarrow \infty$ , where  $G$  is a member of the generalised extreme value (GEV) family. Now consider a sequence of independent variables,  $\tilde{X}_1, \dots, \tilde{X}_n$ , also with common distribution function  $F$ . It can be shown that  $\Pr(\max\{\tilde{X}_1, \dots, \tilde{X}_n\} \leq x) \rightarrow \tilde{G}(x)$  as  $n \rightarrow \infty$ , where  $\tilde{G}$  is also a member of the generalised extreme value (GEV) family, and  $G(x) = [\tilde{G}(x)]^\theta$ , where  $\theta \in [0, 1]$  is the extremal index (see e.g. Coles, 2001, Theorem 5.2). In the case that  $\theta < 1$ , we have  $G(x) = [\tilde{G}(x)]^\theta > \tilde{G}(x)$ , so if the serially correlated data are treated as independent, this will result in an over-estimate of the probability of extreme events.

For a series of independent observations we have  $\theta = 1$ , but there are also dependent series for which  $\theta = 1$ . For example, Chavez-Demoulin and Davison (2012) note that linear Gaussian autoregressive-moving average models have  $\theta = 1$ , corresponding to independent extremes at very high levels, despite clustering at lower levels. This highlights a key feature of the extremal index. Coles (2001, p97) notes that “a series for which  $\theta = 1$  means that dependence is negligible at asymptotically high levels, but not necessarily so at extreme levels that are relevant for any particular application”. Tawn (1990) proposed a penultimate approximation for the extremal index,  $\theta_x$ , which depends on the level,  $x$  and noted that in general  $\theta_x \leq \theta$ . Ledford and Tawn (2003) define the sub-asymptotic extremal index as  $\theta_x(n) = \Pr(X_2, X_3, \dots, X_n \leq x | X_1 > x)$ . A review of methods for estimating the extremal index was presented by Ancona-Navarrete and Tawn (2000), who noted that most methods for estimating  $\theta$  actually estimate  $\theta_x$ . Some authors have proposed incorporating estimates of  $\theta_x$  explicitly into inferences of extremes from serially correlated data (e.g. Eastoe and Tawn, 2012). However, most studies concerned with the effect of serial correlation, only consider the asymptotic extremal index,  $\theta$ . In the examples presented in Sections 3–5, we show that understanding how  $\theta_x$  varies with  $x$  (or equivalently with return period), is important for various problems in offshore design. In Section 2, we introduce a new definition of a sub-asymptotic extremal index and show how this is related to biases in return values and return periods when serial correlation is neglected.

A multivariate extension to the extremal index has also been proposed (Nandagopalan, 1994). Whereas in the univariate case, the asymptotic dependence is summarised by a single number, in the multivariate case the extremal index is a function, describing the dependence as a function of angle (Beirlant et al., 2004). In the present work, the examples considered can all be reduced to univariate problems, so our focus will be on the univariate extremal index.

Another important metric for dependence in time series extremes is the limiting conditional probability  $\lim_{u \rightarrow \infty} \Pr(X_{i+\tau} > u | X_i > u)$ , which estimates the extent of asymptotic dependence in the time series at lag  $\tau$ . This corresponds to the so-called  $\chi$  statistic used in

spatial extremes (Coles et al., 1999), and has been referred to as the extremogram (Davis and Mikosch, 2009). For stationary series, the quantity  $\Pr(X_{t+\tau} > u | X_t > u) - \Pr(X_t > u)$  will tend to zero when  $X_t$  and  $X_{t+\tau}$  are independent. This quantity can be plotted as a function of  $\tau$  for various thresholds  $u$ , to select de-correlation timescales used in declustering routines (Mackay and Johanning, 2018c). The corresponding  $\bar{\chi}$  statistic, characterising asymptotic independence can also be estimated (Coles et al., 1999).

For univariate environmental data it is more common to adopt either the annual maxima method or peaks-over-threshold method. In the latter, a suitable declustering scheme is used to select events that are approximately independent, thus ensuring that  $\theta \approx 1$ . It is important to note that, in the statistics literature, the distribution of quantities such as cluster maxima are considered mainly as a pragmatic approach to obtain the marginal distribution of all observations. In ocean engineering, however, it is often appropriate to focus on the distributions of cluster maxima or storm peak events.

### 1.3. The effect of seasonal and long-term variation

Throughout this work we will be mainly concerned with the effect of serial correlation in environmental conditions over time-scales of hours to several days. Seasonal variation in environmental conditions introduces a longer range dependence into conditions. However, this has a slightly different effect to short-range serial dependence. Provided that storm peaks are sufficiently separated in time, storm peak values of variables such as wind speeds or wave heights, can be considered as independent realisations from a distribution that is dependent on the season. That is, the storm peak value does not depend on the size of the previous storm peak, but only the time of year when it occurs. Mackay and Johanning (2018c) showed that when the seasonal signal is removed from time series of significant wave height,  $H_s$ , values of storm peak  $H_s$  separated by 4–5 days were effectively independent.

Seasonality can be modelled as a covariate effect. There is some debate about whether it is necessary to model seasonality and other covariate effects in order to accurately characterise the annual (or omni-covariate) distribution (Jonathan et al., 2008; Mackay et al., 2010). Using a non-stationary model can improve the accuracy in some cases, but this is dependent on the particular covariate model used and the strength of covariate influence (Mackay and Jonathan, 2020a). The naive use of non-stationary models, where the data is subdivided into seasonal bins and an independent model is fitted in each bin, will lead to larger biases than using stationary models (Mackay et al., 2010; Mackay and Jonathan, 2020a).

Alternatively, the effect of seasonality can be modelled by first removing the seasonal signal (see e.g. Vanem, 2018). However, this generally requires stronger assumptions than covariate approaches, for example that there is no change in the shape of the distribution over the year and that seasonality only affects the mean and variance.

Longer-term climate patterns such as the El Niño Southern Oscillation (ENSO) or the North Atlantic Oscillation (NAO) can also introduce long-range dependence into environmental conditions. However, as with seasonality, these can also be modelled as covariate effects. While neglecting short-range dependence will always result in a positive bias in estimates of return values, the effect of neglecting covariate effects such as seasonality or long-term climatic variations is case dependent and can result in either positive or negative biases.

### 1.4. Definition of return periods for serially correlated data

Suppose, as above, that  $X$  is any observation of a serially correlated sequence in time, with common marginal distribution function  $F$ . An equivalent independent sequence,  $\tilde{X}$ , also with marginal distribution function  $F$ , can be defined by randomising the order of observations in the dependent sequence. Alternatively, inferences from  $\tilde{X}$  can be interpreted as those that would be obtained if the serial correlation in

the dependent sequence,  $X$ , is ignored and observations are treated as independent.

We denote cluster maxima in the dependent sequence as  $X_c$ , where it is assumed that clusters are defined such that  $X_c$  are sufficiently separated in time that they are effectively independent. The marginal distribution of the cluster maxima is denoted  $F_c$ . It is assumed there are a fixed number,  $n$ , observations of  $X$  (or  $\tilde{X}$ ) per year and  $m$  cluster maxima on average per year.

The return period of level  $x$  for the dependent sequence,  $X$ , is defined as

$$T(x) = \frac{1}{m(1 - F_c(x))}. \quad (1)$$

Note that if annual maxima of  $X$  are considered, rather than cluster maxima, then definition (1) still applies, under the interpretation that  $F_c$  is the distribution of annual maxima and  $m = 1$ . In this case, return values can only be defined for  $T > 1$ . As it is often useful to consider the 1-year return value, it is useful to define return periods in terms of cluster maxima rather than annual maxima.

The return period of level  $x$  for the independent sequence,  $\tilde{X}$ , is defined as

$$\tilde{T}(x) = \frac{1}{n(1 - F(x))}. \quad (2)$$

Eq. (2) is sometimes used (incorrectly) to define return periods and return values for the dependent sequence. Our main concern in this paper will be to quantify the impact of using Eq. (2) to calculate return periods and return values when the data exhibit serial correlation.

For serially correlated data, arguably, it is clearer to define return values in terms of the distribution of the maximum value in the return period. If we denote the distribution of the maximum value of  $X$  in  $T$  years as  $F_T(x)$  then the  $T$ -year return value can be defined as the solution of

$$F_T(x) = \exp(-1). \quad (3)$$

If we suppose that occurrences of cluster maxima are independent and a Poisson point process with mean rate of occurrence  $m$  per year, then  $F_T$  can be written

$$F_T(x) = \sum_{k=0}^{\infty} f_p(k) [F_c(x)]^{kT}, \quad (4)$$

where  $f_p(k)$  is the Poisson probability mass function, given by  $f_p(k) = e^{-m} m^k / k!$ . Noting that  $\exp(x) = \sum_{k=0}^{\infty} x^k / k!$ , we can write:

$$F_T(x) = \exp(-mT(1 - F_c(x))). \quad (5)$$

Equating (3) and (5) shows that the two definitions of return periods for the dependent sequence, (1) and (3), are equivalent. The use of definition (1) requires the identification of independent cluster maxima, whereas definition (3) does not explicitly include the requirement to define independent events, only the requirement to establish the distribution of the maximum value in  $T$ -years. Of course, the problem of estimating  $F_T$  will require some method to account for the serial correlation in the data, for example through declustering to identify independent peaks, the use of annual maxima or through explicit treatment of the time series dependence structure. As will be seen in Section 2, definition (3) is useful when considering the theoretical aspects. However, for the practical examples considered in Sections 3–5, definition (1) is more useful.

### 1.5. Outline of the paper

In Section 2 we introduce a new definition of a sub-asymptotic extremal index,  $\theta_T$ , which is dependent on return period  $T$ . We show how the impact of neglecting serial correlation on estimates of extremes can be quantified in terms  $\theta_T$ . We then go on to consider three types of problem involving serial dependence. Firstly, in Section 3, we consider the estimation of univariate extremes of an observed variable, such as

an environmental state or measured response. Methods for dealing with serial correlation are well understood for this case. However, it is instructive to start by considering this simple case to illustrate the general principle. The examples demonstrate why neglecting serial correlation leads to biases, and how the size of the biases (or, equivalently, the value of  $\theta_T$ ) are related to the average shape of storm events and the shape parameter for the tail of the distribution of storm peak events. The theory and results developed in Sections 2 and 3 are then applied to more complex problems. In Section 4, we consider the effect of serial correlation in a multivariate context, where extremes are quantified in terms of environmental contours. In the third example, in Section 5, we go on to consider the effect of serial correlation on long-term extremes of a short-term response. In this case, the short-term response is not directly observed, but is specified in terms of a distribution conditional on the environmental state. If the short-term process is not directly observed, then the historic time series of environmental conditions combined with the short-term response function only gives a probabilistic description of historic response, so the methods required to account for serial correlation are slightly different to the case where we are interested in the extremes of an observed variable. Finally, the conclusions of the work are presented in Section 6.

## 2. Theoretical aspects

In this section, we demonstrate how the effect of neglecting serial correlation on return values and return periods, can be quantified in terms of the extremal index. In Section 2.1, we introduce a new definition of a sub-asymptotic extremal index, relevant for any return period of interest. In Section 2.2 we show how this new definition of the sub-asymptotic extremal index is consistent with the usual asymptotic definition of the extremal index, introduced in Section 1.2.

### 2.1. Relation between extremal index, return periods and return values at sub-asymptotic levels

As before, suppose that  $X$  is a serially-correlated random process, with marginal distribution function  $F$ . According to the results discussed in Section 1.2, there exists a value  $\theta_T \in [0, 1]$  such that the distribution of the maximum value in  $T$  years is given by  $F_T(x) = [F(x)]^{\theta_T n T}$ , where  $n$  is the (fixed) number of observations per year. Here,  $\theta_T$  is a sub-asymptotic extremal index, with  $\theta_T \leq \theta$ . Using definition (3), the return period of  $x$ , is the solution of

$$[F(x)]^{\theta_T n T} = \exp(-1). \tag{6}$$

The return period for the equivalent independent sequence is the solution of

$$[F(x)]^{n \bar{T}} = \exp(-1). \tag{7}$$

Equating (6) and (7) and taking logarithms, we have

$$\theta_T(x) = \frac{\bar{T}(x)}{T(x)}. \tag{8}$$

So, the sub-asymptotic extremal index,  $\theta_T$ , can be interpreted as the ratio of the return period of  $x$ , calculated under the assumption of independence, to the true return period of level  $x$ .

Whilst the effect of serial correlation on return periods of a given level is dependent only on  $\theta_T$ , the effect on return values depends on the shape of the distribution. To illustrate this, we assume that the distribution of the maximum value in  $T$  years for the dependent sequence is  $[F(x)]^{\theta_T n T} \approx G(x)$ , where  $G$  is a member of the GEV family, with cumulative distribution function (CDF)

$$G(x) = \begin{cases} \exp(-\exp(-s)), & \xi = 0, \\ \exp\left(-\left(1 + \xi s\right)_+^{-1/\xi}\right), & \xi \neq 0, \end{cases} \tag{9}$$

where  $z_+ = \max\{z, 0\}$ ,  $s = (x - \mu)/\sigma$ ,  $\mu \in \mathbb{R}$  is the location parameter,  $\sigma > 0$  is the scale parameter and  $\xi \in \mathbb{R}$  is the shape parameter. Under this assumption, it is straightforward to show that  $[F(x)]^{n T} \approx \tilde{G}(x)$  is also a GEV distribution, where  $\tilde{G}^{\theta_T}(x) = G(x)$  and the parameters of  $\tilde{G}$  are given by

$$\tilde{\mu} = \begin{cases} \mu - \sigma \log(\theta_T), & \xi = 0, \\ \mu - \sigma \frac{1 - \theta_T^{-\xi}}{\xi}, & \xi \neq 0, \end{cases} \tag{10a}$$

$$\tilde{\sigma} = \sigma \theta_T^{-\xi}, \tag{10b}$$

$$\tilde{\xi} = \xi. \tag{10c}$$

So serial correlation has no effect on the shape of the tail of the distribution, or on the upper end point in the case that  $\xi < 0$ , where the upper end point is given by  $x^F = \sup\{x : F(x) < 1\} = \mu - \sigma/\xi = \tilde{\mu} - \tilde{\sigma}/\tilde{\xi}$ . The size of the bias in the  $T$ -year return value, caused by neglecting serial correlation, can be calculated by comparing the quantiles of  $G$  and  $\tilde{G}$  at probability level  $p = \exp(-1)$ . The inverse distribution function for the GEV is given by:

$$G^{-1}(p) = \begin{cases} \mu - \sigma \log(-\log(p)), & \xi = 0, \\ \mu + \frac{\sigma}{\xi} \left( (-\log(p))^{-\xi} - 1 \right), & \xi \neq 0, \end{cases} \tag{11}$$

for  $p \in [0, 1]$ . Using definition (3) and substituting  $p = \exp(-1)$  into (11) gives

$$\frac{\tilde{x}_T - x_T}{\sigma} = \begin{cases} -\log(\theta_T), & \xi = 0, \\ \frac{\theta_T^{-\xi} - 1}{\xi}, & \xi \neq 0, \end{cases} \tag{12}$$

where  $x_T$  is the true return value and  $\tilde{x}_T$  is the return value for the equivalent independent sequence. The size of the bias is determined entirely by  $\sigma$ ,  $\theta_T$  and  $\xi$ . The bias is independent of the location parameter,  $\mu$ , and scales linearly with the scale parameter,  $\sigma$ . The bias in return value estimates is shown in Fig. 1 as a function of  $\xi$  for three values of  $\theta_T$ . For a given value of  $\theta_T$ , the bias increases with  $\xi$ .

In the discussion above,  $\theta_T$  is defined in terms of the distributions of the  $T$ -year maximum,  $F_T(x)$ . In practical cases of interest, it is not known a priori how  $\theta_T$  varies with  $T$ . In the simulations presented in Section 3, it will be shown that for a given level of serial correlation around the storm peak, the value of  $\theta_T$  is strongly dependent on the shape parameter of  $F_T$ . Therefore, quantifying bias in return values in terms of  $\xi$  and  $\theta_T$  alone does not tell the full story. The relationship between bias and  $\xi$  shown in Fig. 1, must be understood as the relationship for a fixed value of  $\theta_T$ . The analysis presented in Section 3 indicates that if the average shape of a storm is roughly constant, then  $\theta_T$  will decrease as  $\xi$  increases. In the case that  $\xi < 0$ , it will be shown that  $\theta_T \rightarrow 1$  as  $T \rightarrow \infty$ , and hence the bias caused by neglecting serial correlation also tends to zero. However, when  $\xi \geq 0$ , we can have  $\theta_T \rightarrow \theta < 1$  as  $T \rightarrow \infty$  and the bias tends to a constant value.

### 2.2. Asymptotic considerations

The discussion above has focused on the effect of serial correlation at sub-asymptotic levels. A somewhat surprising asymptotic result is that as the threshold tends toward the upper end point of the distribution, the conditional distribution of a randomly chosen threshold exceedance converges to the conditional distribution of cluster maxima (see e.g. Hsing, 1987b; Anderson, 1990; Leadbetter, 1991). To state this more precisely, suppose, as above, that  $X$  is a serially-correlated random process, with marginal distribution function  $F$ , and upper endpoint  $x^F \leq \infty$ . Under the assumption that  $F$  is in the domain of attraction of an extreme value distribution, we have

$$\lim_{u \rightarrow x^F} \Pr(X > x | X > u) = \bar{H}(x), \tag{13}$$

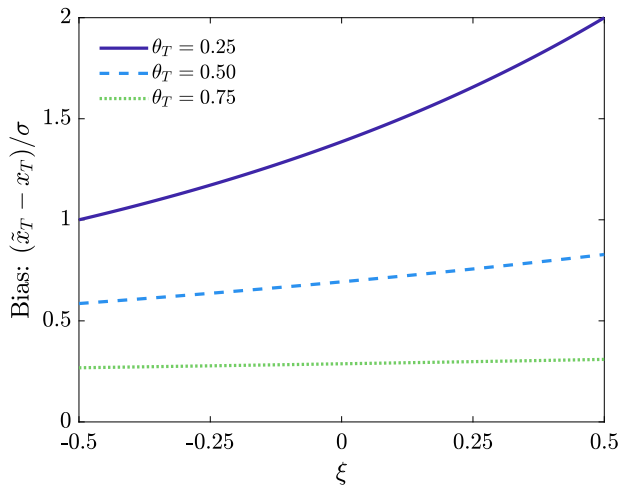


Fig. 1. Bias in the return value estimated under the assumption of independent observations,  $\bar{x}_T$ , relative to true return value,  $x_T$ , for various values of the sub-asymptotic extremal index,  $\theta_T$ . Calculations assume that the true distribution function of the  $T$ -year maximum is a GEV distribution with arbitrary location parameter, scale parameter,  $\sigma$ , and shape parameter,  $\xi$ .

where  $\bar{H} = 1 - H(x)$  and  $H(x)$  is a member of the generalised Pareto distribution family. Similarly, for cluster maxima,  $X_c$ , we have

$$\lim_{u \rightarrow x^F} \Pr(X_c > x | X_c > u) = \bar{H}(x). \tag{14}$$

To examine the effect that the asymptotic equivalence of these distributions has on estimates of return values, we need to consider the non-conditional probabilities  $\Pr(X > x)$  and  $\Pr(X_c > x)$ . Denote the threshold exceedance probabilities for all observations and cluster maxima as  $p_a = \Pr(X > u)$  and  $p_c = \Pr(X_c > u)$ . Then  $\Pr(X > x) = p_a \bar{H}(x)$  and  $\Pr(X_c > x) = p_c \bar{H}(x)$ . Using (1), the return period of level  $x > u$  for the dependent sequence is given by

$$T(x) = \frac{1}{mp_c \bar{H}(x)}, \quad u \rightarrow x^F. \tag{15}$$

Similarly, using (2), the return period of level  $x > u$  for the equivalent independent sequence is given by

$$\tilde{T}(x) = \frac{1}{np_a \bar{H}(x)}, \quad u \rightarrow x^F. \tag{16}$$

Note that  $mp_c$  is the expected number of clusters per year with peak value that exceeds  $u$ , and  $np_a$  is the expected number of observations per year that exceed  $u$ . Therefore, for  $x > u$ ,

$$\frac{\tilde{T}(x)}{T(x)} = \frac{mp_c}{np_a} = \theta_u = \theta, \quad u \rightarrow x^F, \tag{17}$$

where we have used the interpretation that  $\theta_u$  is the inverse of the mean cluster size at level  $u$ , noted in Section 1.2. As this is an asymptotic result,  $\theta_u$  is equal to the asymptotic extremal index  $\theta$  for  $u \rightarrow x^F$ . So, although the conditional distributions of cluster maxima and all observations are asymptotically equivalent, the difference in the expected number of clusters and observations exceeding a given level,  $u$ , results in different return periods for the dependent and independent sequences, consistent with the analysis in the previous section.

### 3. Effect of serial correlation on univariate extremes

In this section, we consider two examples to illustrate the effect of serial correlation on univariate extremes. In the first example, we assume that the time series consists of a series of independent ‘storm’ events, where the evolution of the variable over the storm is given by a simple parametric form. The motivation for using a parametric representation of the time series, is that it allows us to adjust the level

of serial correlation around the storm peaks. In the second example, we use more realistic time histories, resampled from measured time series. Throughout this section, our interest lies in the extremes of an observed variable (i.e. there is no unobserved short-term variability). The case where we have an unobserved short-term response is considered separately in Section 5.

#### 3.1. Examples with parametric time series

In this example, we consider time series of significant wave height,  $H_s$ , but the conclusions from the analysis are applicable to other variables as well. We assume that the time series of  $H_s$  is composed of discrete storm events, where the peak values of  $H_s$  in each storm, denoted  $H_s^{peak} = A$ , are assumed to be independent GEV variables, with CDF given in (9). In the present example, we fix the location parameter as  $\mu = 4$  and scale parameter as  $\sigma = 1$ . The duration of each storm,  $B$ , is assumed to be related to the storm peak, by the fixed relation  $B = 150A^{-0.25}$  hours (a similar relation was reported in Mackay and Johanning, 2018a). The time series of  $H_s$  within each storm is given by the ‘power storm’ model (Arena et al., 2014):

$$H_s(t) = A \left[ 1 - \left( \frac{2|t - t_0|}{B} \right)^\lambda \right], \quad -B/2 \leq t - t_0 \leq B/2, \quad \lambda > 0 \tag{18}$$

where  $t_0$  is the time corresponding to the storm peak. When  $\lambda = 1$ , the storms are triangular. When  $\lambda < 1$ , the storms have sharper peaks, and when  $\lambda > 1$ , the storms have more rounded peaks. The storm shape parameter,  $\lambda$ , determines the level of serial correlation around the storm peak. For  $\lambda \rightarrow 0$ , the peak becomes narrower, the storm tends to a delta function, with value  $A$  at the storm peak and  $H_s \rightarrow 0$  away from the storm peak. In this case, the serial correlation close to the peak tends to zero. Conversely, as  $\lambda \rightarrow \infty$ , the storm tends to a constant rectangular shape, with constant  $H_s = A$  throughout the storm. In this case, the serial correlation within the storm is equal to one.

We use numerical simulation to generate random storm peaks,  $A$ , with the values of  $H_s$  over the course of the storm, conditional on  $H_s^{peak}$ , given by (18). An example of random storms generated using the power storm model is shown in Fig. 2. In this example, the shape parameter for the GEV distribution of storm peak heights is  $\xi = 0$ , and the storms have been generated using exponents of  $\lambda = 0.5, 1$  and  $2$ . The distribution of storm peak heights is not affected by the storm shape parameter  $\lambda$  (it is determined solely by the GEV parameters), whereas the marginal distribution of  $H_s$  is directly affected by the value of  $\lambda$ .

Simulations were conducted for three values of the GEV shape parameter, with  $\xi = -0.2, 0$  and  $0.1$ . For each value of  $\xi$ ,  $10^8$  random storm peak heights were simulated and hourly values of  $H_s$  within each storm were generated for storm shape parameters  $\lambda = 0.5, 1$ , and  $2$ . The total lengths of the simulations were approximately  $2 \times 10^6$  years (this is related to the fixed relationship between  $A$  and  $B$ , noted above).

Fig. 3 shows return values estimated from the independent storm peaks and from all hourly values of  $H_s$ , under the assumption of independence. Return values are shown for return periods up to  $10^4$  years, where the sampling uncertainty is an acceptable level (see Mackay and Jonathan, 2021). It is clear that both the GEV shape parameter and the storm shape parameter influence the differences between the return values from the storm peak values and the return values calculated from hourly  $H_s$  under the assumption of independent observations. As expected, the effect of neglecting serial correlation has a greater impact when the correlation around the storm peak is higher, corresponding to the higher values of  $\lambda$ .

For the simulations with  $\xi = -0.2$  and  $\lambda = 0.5$ , neglecting serial correlation has negligible effect for return periods greater than 10 years. The reason for this is that when a storm occurs with a peak height exceeding the 10-year return value, typically there is only one hour in the storm when this occurs, so exceedances of this level are effectively independent. However, for higher values of the storm shape parameter there are typically multiple hours in a storm exceeding the

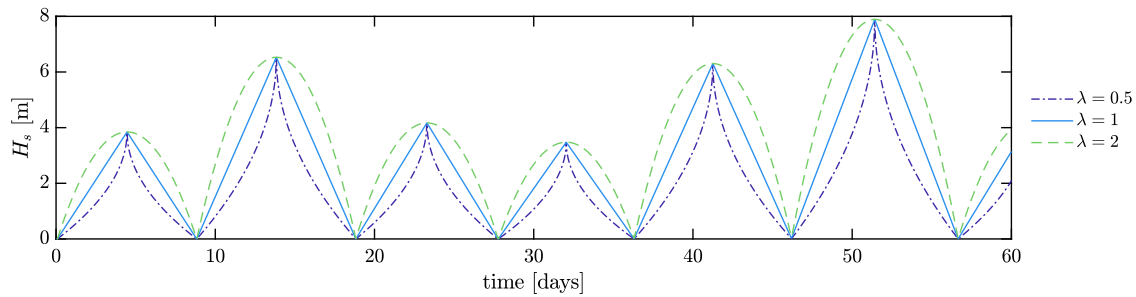


Fig. 2. Simulated time series using parametric storm shapes with random peak heights and various storm shape parameters,  $\lambda$ . The storm shape affects the marginal distribution of  $H_s$ , but not the distribution of peak heights.

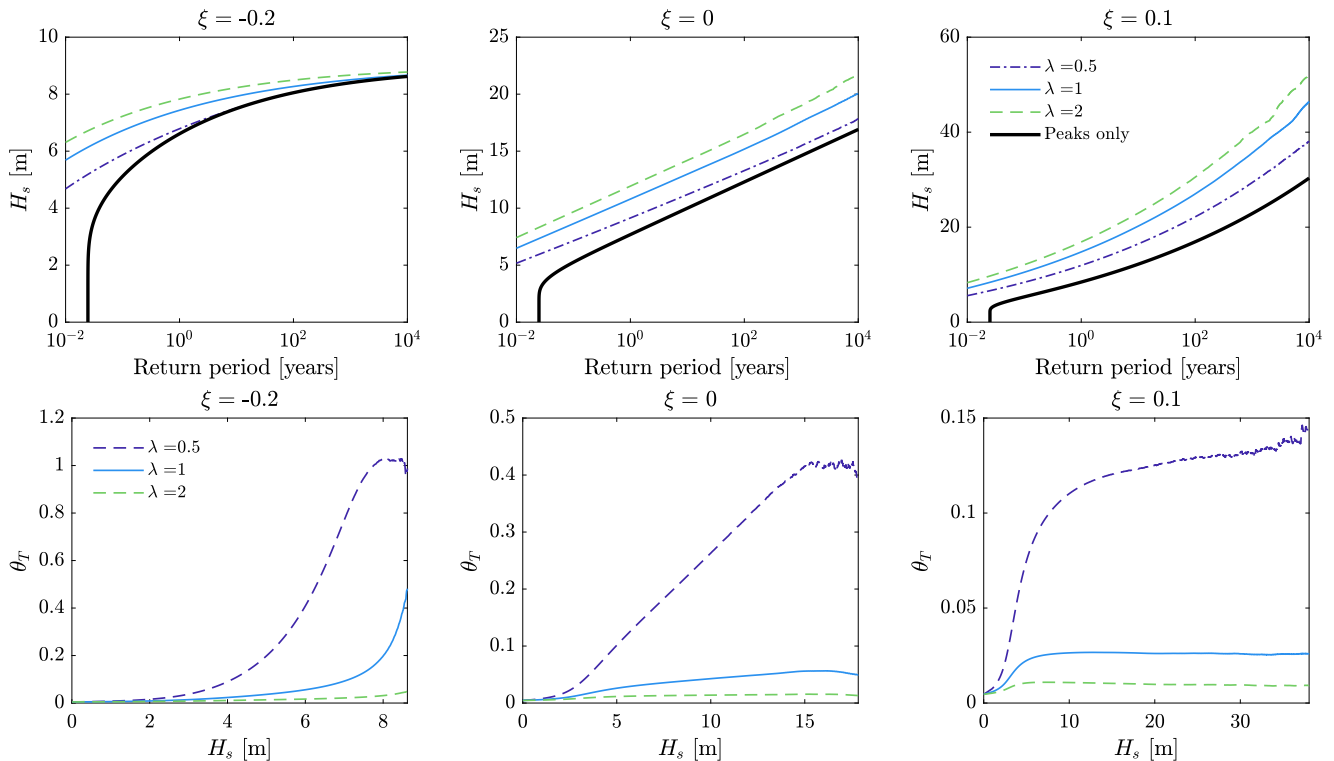


Fig. 3. Top row: Return values of  $H_s$  calculated using the time series of parametric storms, for various storm shape parameters,  $\lambda$ , and GEV shape parameter,  $\xi$ . Black lines show return values calculated from storm peak values only. Coloured lines indicate return values calculated from all values of  $H_s$  under the assumption of independent hourly observations. Bottom row: estimated sub-asymptotic extremal index,  $\theta_T$ . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

true 10-year return level, which are all treated as independent events if serial correlation is ignored. In the latter case, the more correlated observations that are treated as independent, the greater the bias will be.

The GEV shape parameter has a similar effect. In the case that the distribution of storm peaks has a longer tail (i.e. higher values of  $\xi$ ), the difference between e.g. the 10-year and 100-year return value is larger. So, when a storm occurs where the peak value exceeds the 100-year level, there will be many hourly values which also exceed the 10-year level. So if these are all treated as independent events then this will introduce bias. In contrast, when the tail of the distribution of storm peaks is shorter, the difference between the 10-year and 100-year level is smaller, so for a storm with peak value exceeding the 100-year level, there will be fewer hourly values exceeding the 10-year level.

Estimates of  $\theta_T$  are shown in Fig. 3. It is evident that  $\theta$  is dependent on both the storm shape parameter  $\lambda$  and the GEV shape parameter  $\xi$ . For the cases with  $\xi = -0.2$ ,  $\theta_T$  increases with level for all values of  $\lambda$ . For the more peaked storms with  $\lambda = 0.5$ , the extremal index tends towards 1 for  $H_s > 8$  m (note that finite sample size effects mean that

$\theta_T$  is not exactly equal to one). However, for  $\xi = 0$  and  $\xi = 0.1$ , the extremal index for  $\lambda = 0.5$  tends to a constant less than 1 (note that the upper limit of the x-axis has been set at the  $10^4$ -year return value, so that sampling variability is kept to a reasonable level). So, despite the level of serial correlation around the storm peaks being constant,  $\theta_T$  varies with the shape of the distribution of storm peak heights. Since  $\theta_T$  effectively quantifies the level of bias introduced by neglecting serial correlation, the reason for the dependence of  $\theta_T$  on  $\xi$  is clear, given the discussion above. When  $\xi < 0$ , the GEV distribution has a finite upper end point,  $x^F = \mu - \sigma/\xi$ . In this case, for storms which have a fixed shape with a single largest value, there will be only one point in a storm that exceeds a level  $x$  as  $x \rightarrow x^F$ . This means that exceedances of  $x$  are independent as  $x \rightarrow x^F$  and hence  $\theta_T \rightarrow 1$ . In contrast, for  $\xi \geq 0$  and fixed storm shape,  $\theta_T$  will tend to a finite value.

The examples presented in this section are relatively simplistic, but serve to illustrate how the storm shape and the distribution of the storm peak values both influence the bias related to neglecting serial correlation. In the next section, we use more realistic examples to illustrate the effect of serial correlation in real datasets.

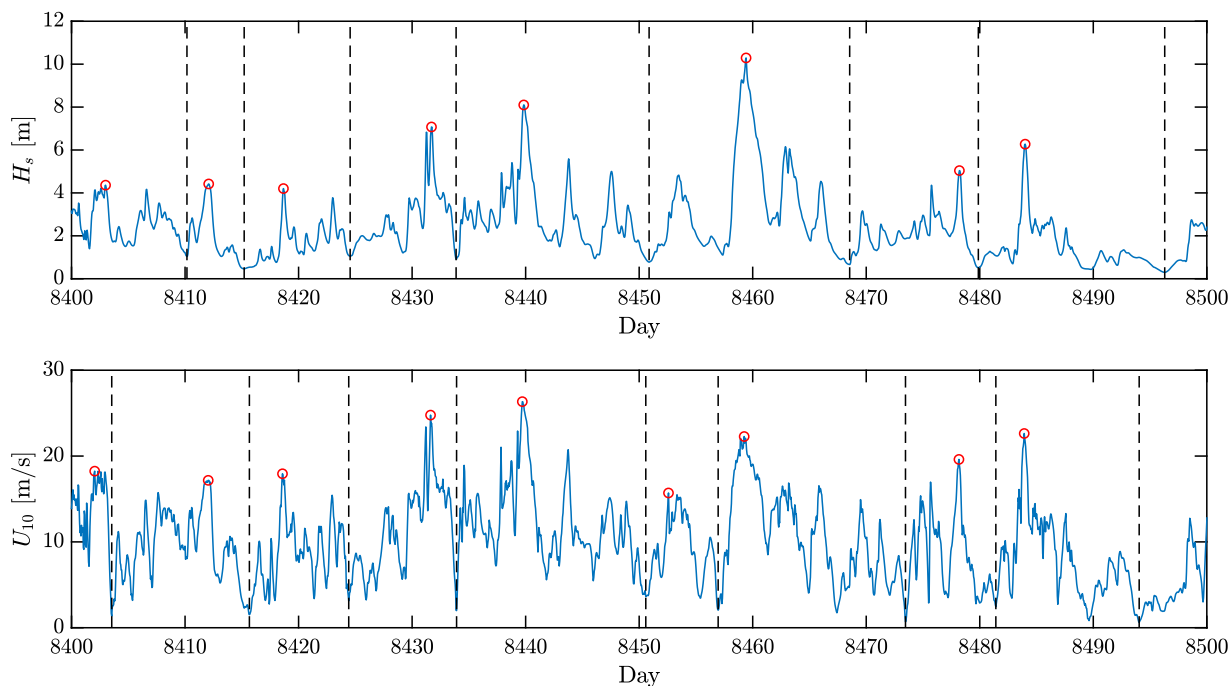


Fig. 4. Examples of time series of  $H_s$  and  $U_{10}$ , with declustered peaks circled in red and minima between adjacent peaks indicated by dashed lines.

### 3.2. Examples with resampled time series

In this section, we illustrate the effect of serial correlation in a dataset for a location in the central North Sea (dataset E from the recent benchmarking exercise for environmental contours, described in Haselsteiner et al., 2019). The dataset consists of a 50 year time series of hindcast values of  $H_s$  and wind speed at 10 m above water level,  $U_{10}$ , at hourly intervals. If we used the original 50-year time series to compare empirical estimates of return values from declustered peaks and all hourly observations, then the largest return values will coincide from both methods. However, this is an effect of finite sample size, which will happen for any sample, since the largest storm peak is, by definition, equal to the largest observation.

To circumvent this problem, we can resample the time series to create synthetic time series of arbitrary length. We use an approach similar to that outlined in Mackay and Jonathan (2020b), although here we consider each variable on its own, rather than modelling their joint distribution. The method is as follows. The time series of  $H_s$  and  $U_{10}$  are declustered to identify local peaks which are approximately independent, with peaks defined as local maxima within a moving window of length  $\pm 5$  days, to ensure peaks are separated by a minimum of 5 days. (Changing the minimum separation time used for the declustering mainly affects the isolation of lower peaks which have little influence on the extremes). Each time series is then partitioned into blocks where the peak values are assumed independent. The boundaries between the blocks are defined as the minimum value between adjacent peaks. This definition is somewhat arbitrary, but ensures that large values are not located close to block boundaries, so we do not separate one ‘event’ into separate blocks. Fig. 4 shows examples of how the time series of each variable are partitioned into blocks.

The storm-peak values are modelled with a GEV distribution, with CDF given by (9). It is more common to model the storm-peaks exceeding some threshold level with a generalised Pareto (GP) distribution. However, here we wish to fit a distribution to *all* storm-peak values, which can be used for simulating artificial time series. This could be achieved by using a two-part model with a parametric distribution for the body of the distribution and a GP model for the tail. The problem with this approach is that it is difficult to ensure continuity of the

density function on the boundary between body and tail. The GEV gives a simple, flexible model for all storm-peak values, and avoids the need to fit a two-part model. In practical cases of estimating univariate extremes, this would not be necessary, as only a model for the tail is required.

Fig. 5 shows the fitted GEV distributions and parameter estimates for the peak values of  $H_s$  and  $U_{10}$ . Whilst the GEV distribution is asymptotically justified as a model for maxima of blocks of constant size and sufficient length, the model diagnostics for the current data suggests that it is still a reasonable model for both variables in this example, despite the non-constant block sizes and finite block length. It should be emphasised that the aim here is not to fit an optimal model for the data or to advocate that data should be modelled in this way in general. The aim is to fit a simple parametric model for all storm peak values, which can be used to generate synthetic time series that are reasonably realistic. Fig. 5 indicates that the GEV distribution is a reasonable model for the peaks for this purpose.

To simulate synthetic time series, we simulate random storm peaks and resample and rescale the observed blocks to match the simulated peak heights. However, the correlation structure of the observed time series is dependent on the block-peak value. To illustrate this, we can consider the mean storm shape as a function of peak level. If hourly values are treated as independent, then the order in which they occur within each block does not matter. So we can re-order each block into a descending sequence, to examine how many hours there are where  $H_s$  or  $U_{10}$  is close to the peak value. Fig. 6 shows the average shape of the ordered blocks (or ‘storms’), when storms are binned by the peak value. The average storm shape is similar for both  $H_s$  and  $U_{10}$ . In both cases, the larger storms decay more rapidly away from the peak. To preserve this effect, for each simulated storm peak, measured blocks are resampled at random from the five blocks with the closest peak values. This does not provide a way to extrapolate the shape of the storm outside the range of observations, but provided the extrapolation is not too large, the results should be representative. It should be noted that although there may be covariate effects on the distribution of storm peak  $H_s$  or  $U_{10}$  in the original data (e.g. dependence on season or direction), by design, there are no covariate effects in the simulated time series.

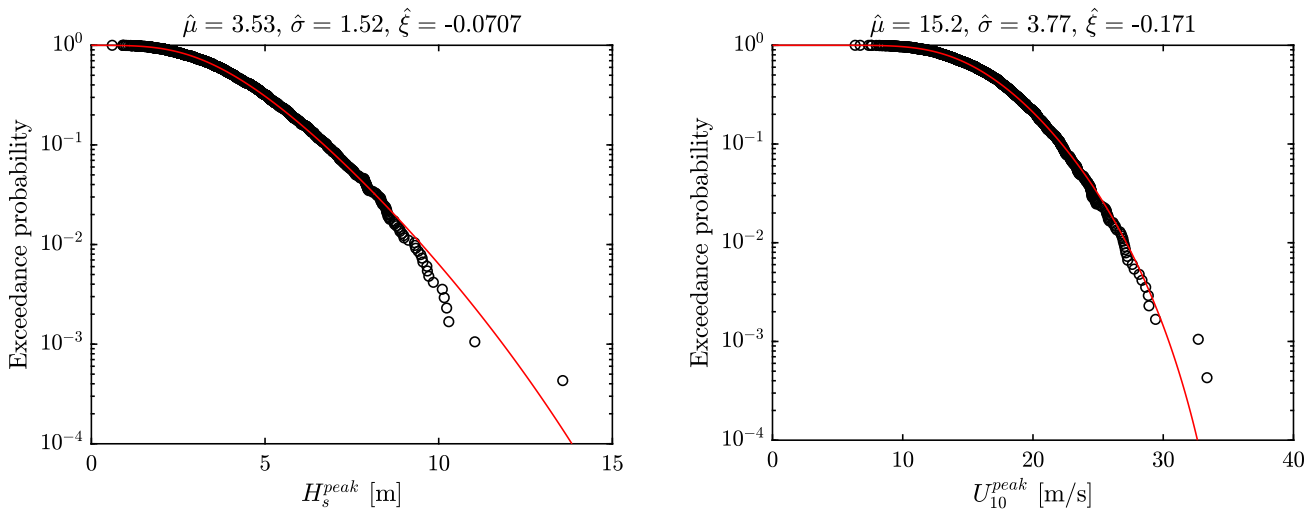


Fig. 5. GEV fit to declustered storm peaks for North Sea dataset. GEV parameter estimates are shown above each plot.

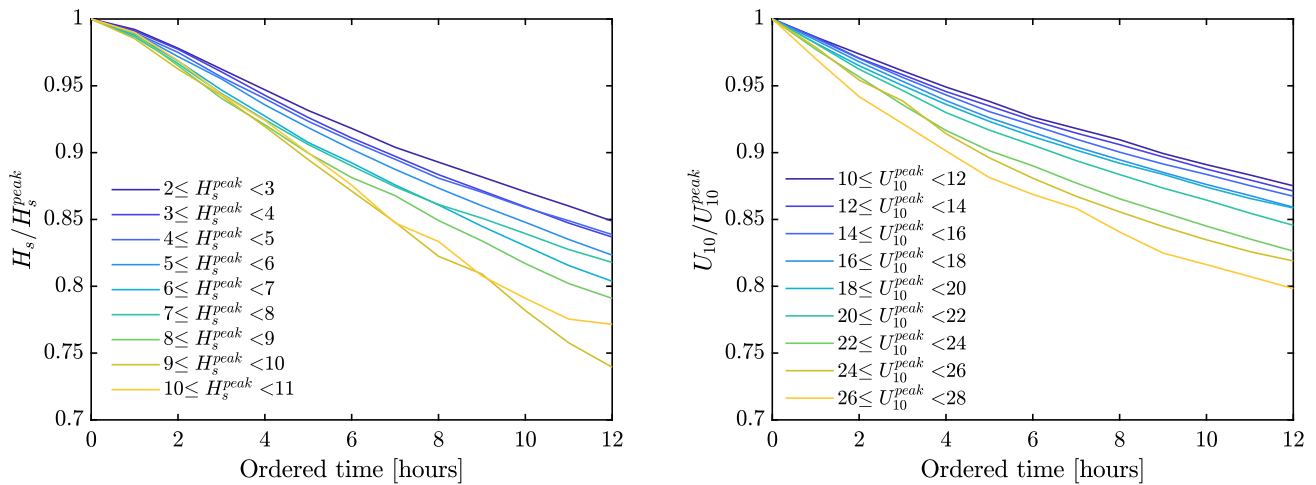


Fig. 6. Average shapes of ordered storms for North Sea dataset.

A total of  $10^6$  random storm peaks were simulated from the fitted GEV distributions for  $U_{10}$  and  $H_s$ . This gave a total length of approximately 30,000 years of hourly records. Fig. 7 shows the return values of  $H_s$  and  $U_{10}$  from the peak values and the hourly values from the resampled time series. Return values are shown for return periods between 0.1 and 100 years, for which the sampling uncertainties in the return values are very low from the 30,000-year time series (see Mackay and Jonathan, 2021). For  $H_s$ , there is a difference of more than 2 m at the 1-year level and around 1 m at the 100-year level. For  $U_{10}$  the difference at the 1-year level is around 1.7 m/s, but at the 100-year level, the difference is relatively small (around 0.1 m/s). Fig. 6 indicated that the level of serial correlation around the storm peaks was similar for winds and waves at this location. The difference between the effect of neglecting serial correlation is therefore related to the shape of the tail of the distribution of storm peaks. The tail of the distribution of  $U_{10}^{peak}$  is shorter than the tail for  $H_s^{peak}$  (i.e. the GEV shape parameter was more negative, as shown in Fig. 5); as discussed in the previous section, this results in exceedances of higher levels being effectively independent. It is important to note that these conclusions about neglecting serial correlation are particular to this example. In other locations, where the shape of the distribution of storm peaks is different or where the average ordered storm shape is different, then the effect of serial correlation would change accordingly, as indicated by the simple examples in the previous section.

#### 4. Effect of serial correlation on environmental contours

Environmental contours (EC) are widely used as an approximate method of estimating the long-term extreme response of a marine structure, based on response calculations for a small number of design conditions. In the EC method, the probabilistic description of the environment is separated from structural design, by making certain assumptions about the region of the environmental variable space where the structure fails (the failure region). The most commonly used EC approach is the inverse first-order reliability method (IFORM) (Winterstein et al., 1993). In this method, it is assumed that the failure surface (the boundary of the failure region) can be approximated by a straight line in two dimensions (or hyperplane in higher dimensions) at the design point (the point on the failure surface with the highest probability of occurrence). The IFORM approximation to the failure surface is illustrated in Fig. 8. The method can be applied in higher dimensions, but we will restrict our attention to the two-dimensional case.

Typically, short-term variability in the response is not considered directly in the EC method. Instead, the short-term response is evaluated at a fixed quantile. Often the median value of the distribution of the maximum response in each environmental state is used, but some design standards recommend using a higher quantile to compensate for neglecting the effect of short-term variability — see e.g. Winterstein



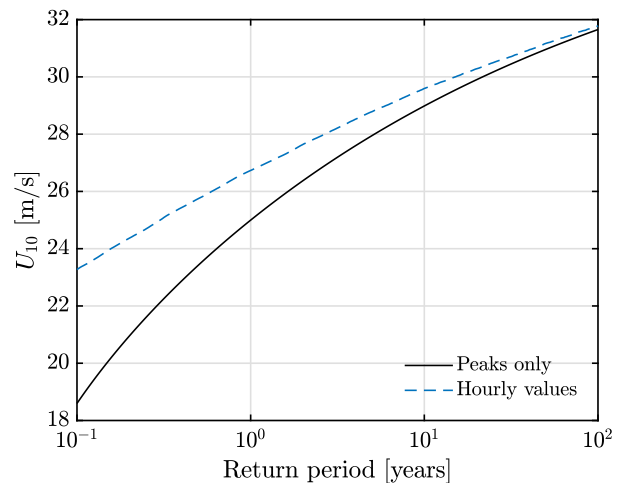
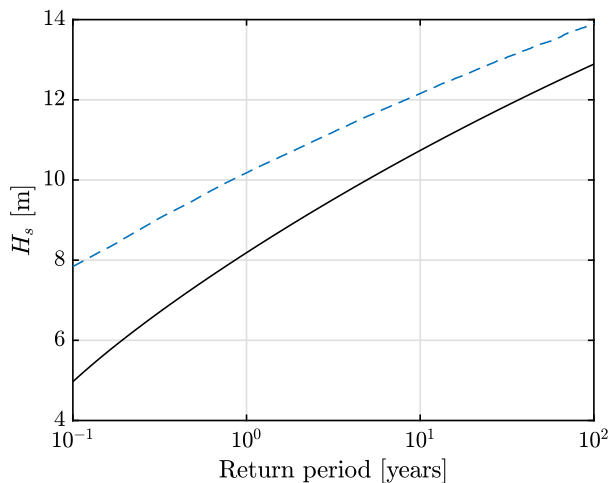


Fig. 7. Return values of  $H_s$  and  $U_{10}$  from resampled North Sea data.

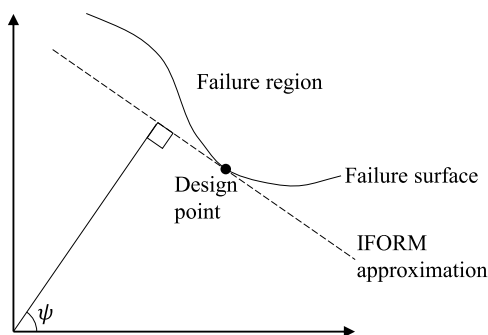


Fig. 8. Illustration of IFORM approximation to failure surface.

et al. (1993), Winterstein and Engebretsen (1998) and Derbanne et al. (2017). This is equivalent to assuming that the short-term response function is deterministic, with the value given by the fixed quantile of the distribution of the maximum response in each environmental state.

Under these assumptions, the problem of estimating extreme structural responses is reduced to estimating the probability that environmental conditions fall within the half-plane region defined by the linearised failure surface. Since the location of the failure region is not known, the  $\alpha$ -exceedance environmental contour is defined by the intersection of all half-plane regions containing a probability  $1 - \alpha$ . The definition of these regions is not unique. In the original IFORM method (Winterstein et al., 1993), the regions are calculated by applying the Rosenblatt transformation to the joint density function, which maps the original variables to independent standard Gaussian variables. In the Gaussian space, the half-plane areas containing a probability  $1 - \alpha$  are at a constant radius of  $\Phi^{-1}(1 - \alpha)$  from the origin, where  $\Phi$  is the CDF of the standard Gaussian distribution. Huseby et al. (2013, 2015) noted that it was not necessary to apply the Rosenblatt transformation to calculate these exceedance regions and proposed that the boundary of the half-plane regions could be calculated in the original parameter space.

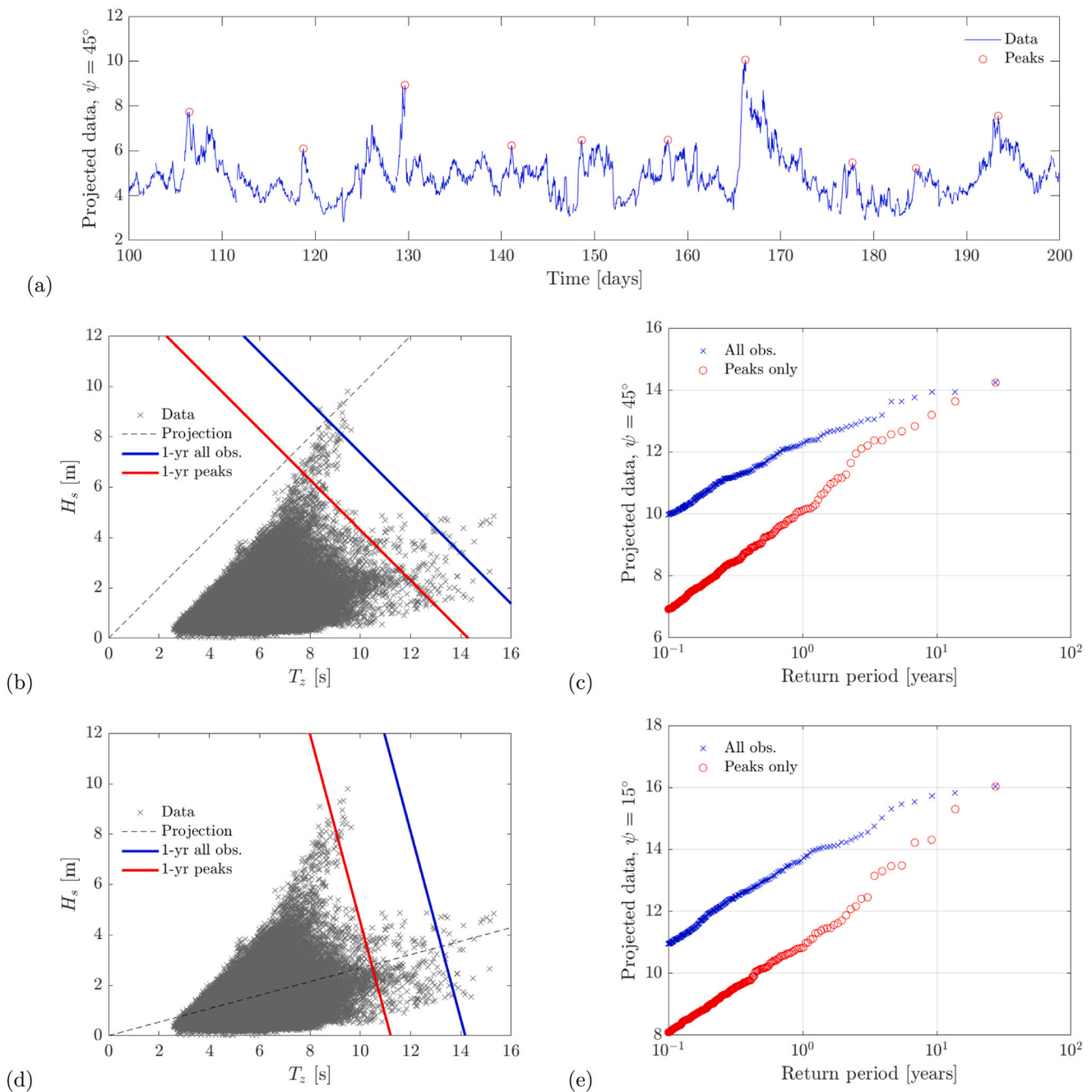
Both the original and modified IFORM methods can be calculated by projecting the variables onto lines at various angles,  $\psi$ , to the origin (either in Gaussian space or the original parameter space) and calculating the  $1 - \alpha$  quantile for each projection angle. The problem of calculating multivariate exceedance regions is therefore reduced to a series of univariate analyses of deterministic variables. (Note that this is different to the cases considered in Section 5, since short-term variability is not considered directly in the EC method). In both approaches, it is only the proportion of observations in a region that is

counted, without accounting for serial correlation in the data. As such, the positive biases resulting from neglecting serial correlation, that were discussed in the previous section, will also affect environmental contours calculated in this way. However, since IFORM contours are defined in terms of univariate exceedance regions, the techniques of univariate analysis can be applied here as well. Derbanne and de Hauteclocque (2019) proposed using a POT method on projections of the data onto lines at given angles, in the original variables space. In their method, the projected data are first declustered to identify independent peaks, and a generalised Pareto (GP) model is fitted to the declustered peak values at each angle.

To illustrate the effect of serial correlation on construction of environmental contours, we will use another dataset from the recent benchmarking exercise (Haselsteiner et al., 2019). The dataset is a 30-year time series of hourly measurements of significant wave height,  $H_s$ , and zero-upcrossing period,  $T_z$ , taken from NDBC buoy 41009, located off the coast of Florida, US. We consider the construction of IFORM contours in the original parameter space, following the approach of Huseby et al. (2013). Fig. 9 shows some steps involved in the construction of the contour. An extract of the time series when the data are projected onto a line at  $\psi = 45^\circ$  to the origin is shown (where the projected variable is defined by  $X = T_z \cos \psi + H_s \sin \psi$ ), together with peaks defined as local maxima within a moving window of length  $\pm 5$  days. Rather than fitting a GP model to the projected data, we consider the 1-year empirical return values, for which the sampling variability is reasonable for a 30-year dataset (see Mackay and Jonathan, 2021). The empirical return values from the peaks and all observations are shown in panel (c), and the boundaries to the half plane regions defined by the return values of the projected data are shown in panel (b). Another example for projections at  $\psi = 15^\circ$  is shown in panels (d) and (e).

For these projection angles, the tail of the distribution of storm peaks is approximately exponential (evident from the approximately linear relationship between return values and return periods, when plotted on a semi-log scale). As shown in the previous section, this results in a large difference between the 1-year return values from the peak values and all observations. By definition, the empirical estimates of return values coincide for the largest storm peak and largest observation in the 30-year dataset. However, as noted in the previous section, this is a finite sample size effect that will occur for any sample.

Fig. 10 shows the 1-year empirical IFORM contours for the Florida dataset, constructed from all observations and from peak values only. The contours are constructed as the intersection of the non-exceedance regions at projection angles  $\psi = 10^\circ, 20^\circ, \dots, 360^\circ$ . The effect of sampling uncertainty is evident in some of the ‘loops’ at various points on the contours (see Huseby et al., 2013 for a discussion of this effect and Vanem, 2019 for a discussion of improved sampling methods from



**Fig. 9.** Illustration of construction of IFORM contours by projecting data onto lines at various angles  $\psi$  to the origin. (a) Time series of  $(T_z, H_s)$  data projected onto line at  $45^\circ$  to the origin, together with local peaks within a 5-day moving window. (b) Scatter plot of  $H_s$  against  $T_z$ , together with empirical 1-year return values for projected data at  $\psi = 45^\circ$ . (c) Empirical return values for projected data at  $\psi = 45^\circ$ , calculated using all observations (blue) and peaks only (red). (d) and (e) same as (b) and (c), but for  $\psi = 15^\circ$ . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

fitted models). However, the differences between the two contours is clear. There is a large difference for the highest values of  $H_s$  and  $T_z$ , with nearly a 3 m difference in the maximum value of  $H_s$  and nearly 3 s difference in the maximum value of  $T_z$ . This is related to the long tails of these distributions, discussed in the previous section. In contrast, the differences between the two contours on the low period side is much lower, since the tail of the distribution of the projected data is shorter for these angles, due to the physical limit imposed by steep waves breaking.

As discussed in the previous section, the effect of neglecting serial correlation may decrease for higher return periods (e.g. when  $\xi < 0$ ). However, the size of the effect and how it varies with return period will depend on the particular joint distribution of interest. Further

discussion of the effect of neglecting serial correlation on response estimates from environmental contours is provided in [de Hauteclocque et al. \(2021\)](#).

Finally, it is important to note that neglecting serial correlation also affects contours defined in terms of ‘total exceedance’ probability, i.e. the probability than an observation falls anywhere outside the contour. These include highest density region contours ([Haselsteiner et al., 2017](#)) and inverse second-order reliability method (ISORM) contours ([Chai and Leira, 2018](#)). As with standard IFORM contours, highest density and ISORM contours are calculated in terms of the proportion of data outside a region, which tells us nothing about how these occurrences are distributed in time. As far as the present authors are aware, currently, no method has been proposed to account for serial

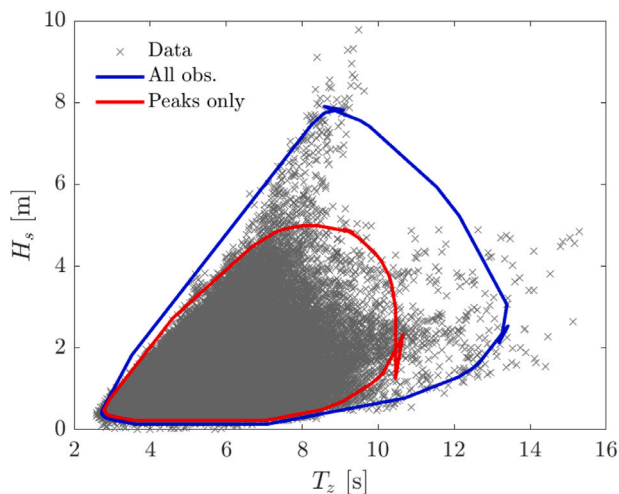


Fig. 10. 1-year IFORM contours for Florida dataset based on declustered data and all observations.

correlation in highest density or ISORM contours. Given that ‘total exceedance’ contours are significantly more conservative than contours defined in terms of marginal exceedance probabilities (i.e. the various types of IFORM contour) (Mackay and Haselsteiner, 2021), the effect of neglecting serial correlation will introduce yet further conservatism.

## 5. Effect of serial correlation on long-term extreme responses

In this section we consider the problem of how to calculate the long-term extremes of a short-term response that is conditional on the environmental state. The short-term response can represent a range of processes, such as sea surface elevation, the motion response of an offshore structure, or the load on a particular part of a structure. The environmental conditions are assumed to vary on a much slower time scale than the short-term response, such that the environmental condition can be considered stationary for a period of the order of a few hours, during which there are many cycles of the short-term response.

The key difference between this problem and the univariate cases discussed in Section 3, is that, typically, the short-term response is not directly observed, but is described in terms of a previously-established distribution, conditional on the environmental state. In this case, a record of historic environmental conditions only enables a probabilistic description of the historic response.

To illustrate why this problem is different to the cases considered in Section 3, consider the problem of estimating the long-term extreme individual wave height. Fig. 11 shows a time series of  $H_s$  over the course of a storm, together with 1-hour maximum wave heights (simulated assuming a Rayleigh distribution of wave heights and  $3600/T_z$  waves per hour). In this example, the maximum individual wave height in the storm does not occur at the time of the storm peak  $H_s$ . In the examples in Section 3, we were interested in the distribution of the peak values of an observed variable, and we only needed to consider the peak value of the variable in each storm. In the present example, it is clear that if we only consider the peak  $H_s$  in a storm, then we risk missing the maximum individual wave height. So in the case where we do not directly observe the short-term response, we need to account for all environmental conditions over a storm.

There is a common misconception about the need to account for serial correlation in the estimation of long-term extreme responses, related to the difference in the correlation time scales for the short-term response process and the environmental variables. In a stationary environmental state, the correlation time scales for most response processes are relatively short, and it is reasonable to expect that peak responses

separated by 1 h would be independent. However, since environmental conditions are non-stationary and correlated over time scales of hours to days, this introduces correlation into the values of peak responses in adjacent 1-hour periods. This effect is apparent in the lower panel of Fig. 11, where the serial correlation in 1-hour maximum wave heights over the storm is clearly visible. Fig. 12 shows a one year time series of  $H_s$  and 1-hour maximum wave height. In this figure, the dependence of the 1-hour maximum wave height on  $H_s$  is very clear, illustrating how the serial correlation in the environmental state introduces serial correlation into the short-term maximum responses.

We start by presenting a brief review of methods for calculating long-term extreme response in Section 5.1. Some examples of the effect of neglecting serial correlation on estimates of long-term extreme responses are then presented in Section 5.2.

### 5.1. Methods for calculating long-term extreme response

Methods for calculating the long-term extreme response are usually based on calculating the distribution of the maximum response in a random event, where the event is either a storm, a single short-term condition, or a single peak. The long-term distribution is then formed under the assumption that each event is independent. We shall refer to the methods as the ‘all-peak’, ‘short-term maxima’ and ‘storm-based’ methods. Comparisons of all-peak and short-term maxima methods were presented in Sagrilo et al. (2011). Further comparisons with storm-based methods were presented in Forristall (2008) and Mackay and Johannig (2018c). Here, we present a brief overview of the methods. As storm-based approaches are less widely used, we present more detail on these methods. In the following, we denote the vector of parameters representing the environmental state as  $\mathbf{S}$  and the short-term response peaks as  $R$ . Here we use the term ‘peaks’ to distinguish between local peaks (e.g. crest heights) and the continuous short-term response (e.g. surface elevation). The distribution of response peaks conditional on environmental state is denoted  $F_{R|\mathbf{S}}(r|\mathbf{s})$  and the joint density function of environmental conditions is denoted  $f_{\mathbf{S}}(\mathbf{s})$ .

#### 5.1.1. All-peaks methods

Methods for calculating the long-term distribution of all short-term response peaks have been proposed in Nordenström (1969), Battjes (1970) and Tucker (1989). In this work, the long-term distribution of all short-term peaks is calculated using the Battjes method (Battjes, 1970). In this method, the probability that a randomly chosen response peak does not exceed a level,  $r$ , is given by

$$F_R(r) = \bar{N}^{-1} \int_{\mathbf{S}} N(\mathbf{s}) F_{R|\mathbf{S}}(r|\mathbf{s}) f_{\mathbf{S}}(\mathbf{s}) d\mathbf{s}, \quad (19)$$

where  $N(\mathbf{s})$  is the average number of response peaks in condition  $\mathbf{s}$  and  $\bar{N}$  is the average number of response peaks in any environmental state chosen at random, given by:

$$\bar{N} = \int_{\mathbf{S}} N(\mathbf{s}) f_{\mathbf{S}}(\mathbf{s}) d\mathbf{s}. \quad (20)$$

The distribution of the maximum response in a  $T$ -year period is calculated under the assumption that all response peaks are independent:

$$F_{R_{max},T}(r) = [F_R(r)]^{\bar{N}nT}, \quad (21)$$

where  $n = 365.25 \times 24/d$  is the number of environmental conditions per year, and  $d$  is the duration of each condition in hours.

#### 5.1.2. Short-term maxima methods

Two methods are commonly used for calculating the long-term distribution of the maximum response in each environmental condition. One is based on ergodic mean (Naess, 1984; Krogstad, 1985) and the other based on an average over the probability of occurrence of environmental states. The two methods give almost identical results for return periods of interest for offshore design (see Sagrilo et al.,

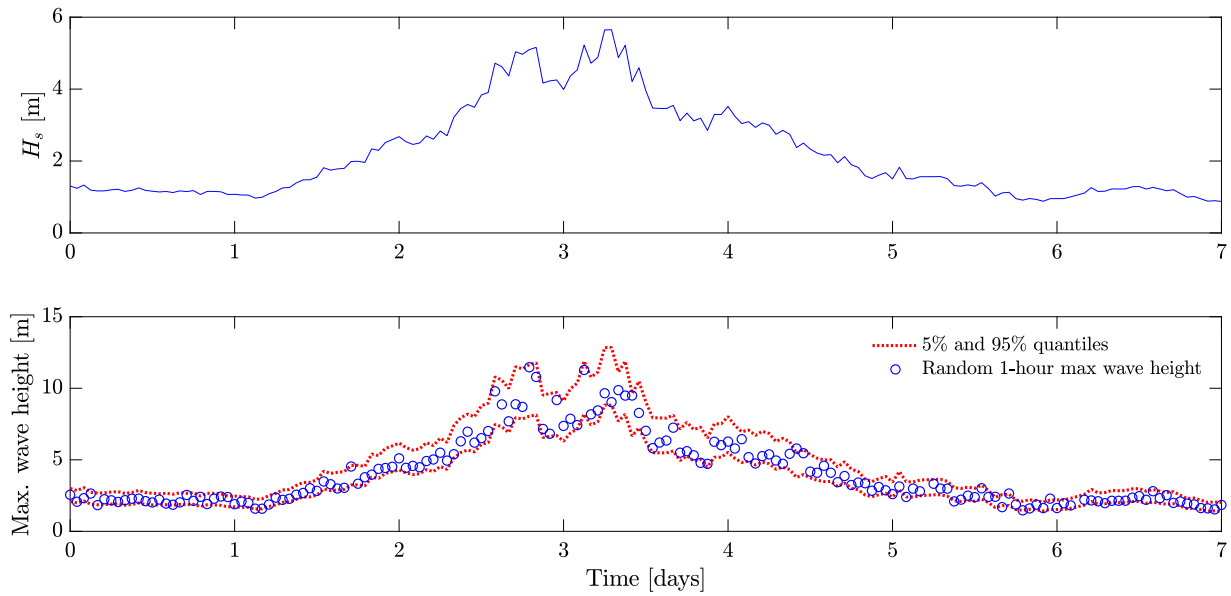


Fig. 11. Example of  $H_s$  and random 1-hour maximum wave heights over a storm. The 5% and 95% quantiles of the distribution of the 1-hour maximum wave height are also shown, assuming that individual wave heights follow a Rayleigh distribution. In this example the maximum individual wave height does not occur at the storm peak  $H_s$ .

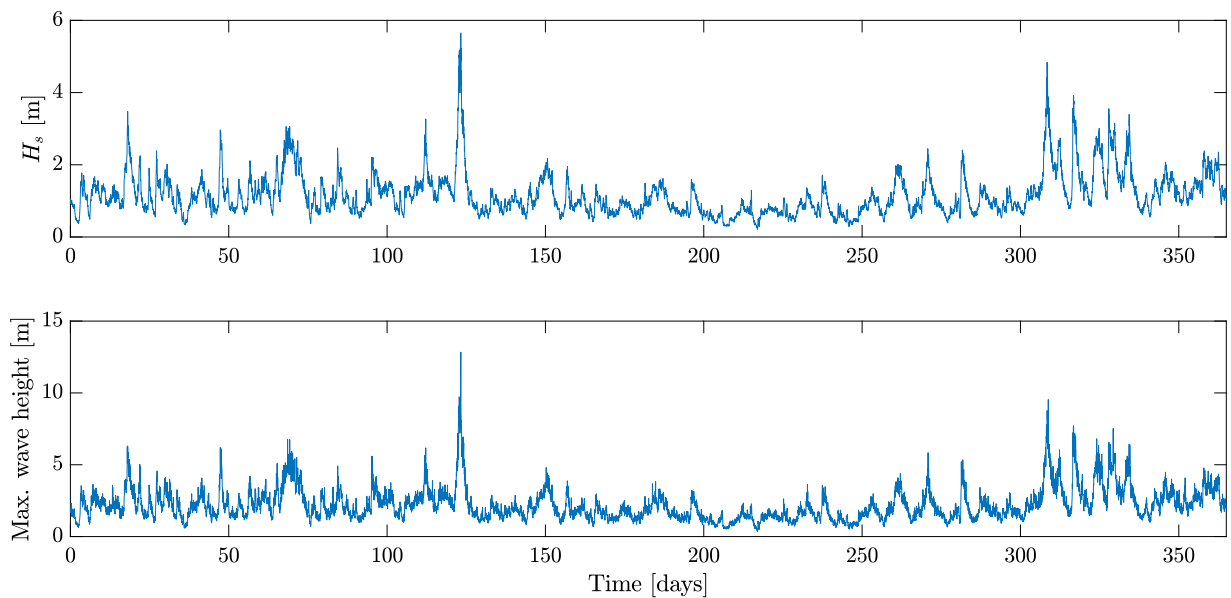


Fig. 12. Example time series of  $H_s$  and random 1-hour maximum wave heights over a year. The serial correlation in 1-hour maximum wave heights is evident.

2011; Mackay and Johanning, 2018c). For consistency with the all-peaks and storm-based methods, we shall use the definition in terms of the probabilistic average. Denote the distribution of the maximum response in environmental state  $s$  as  $F_{R_{max}|s}(r|s)$ . Then the distribution of the maximum response in any environmental condition selected at random, is given by

$$F_{R_{max}}(r) = \int_{\mathbf{S}} F_{R_{max}|s}(r|s) f_{\mathbf{S}}(s) ds. \quad (22)$$

The distribution of the maximum response in a  $T$ -year period is calculated under the assumption that the short-term maximum responses are independent:

$$F_{R_{max},T}(r) = [F_{R_{max}}(r)]^{nT}. \quad (23)$$

As discussed above, short-term maximum responses exhibit serial correlation, so (23) will overestimate the probability of occurrence of

large response values. We note that (22) is a valid way of calculating this distribution of all short-term maximum responses (given the caveats about the differences between the ergodic mean and probabilistic mean, mentioned above). However, since short-term maximum response are serially-correlated, this distribution is not what is required for calculating long-term extreme responses. Calculating long-term extreme responses from  $F_{R_{max}}(r)$ , is analogous to calculating long-term extremes of an environmental variable,  $X$ , based on the distribution of all observations,  $F_X(x)$ , without accounting for serial correlation. As discussed in Sections 1–3, this leads to a positive bias in estimated return values.

### 5.1.3. Storm-based methods

Storm-based methods are derived in a similar way to the all-peaks or short-term maxima methods. We start by calculating the distribution of the maximum response in any storm selected at random,  $F_{R_{SP}}$ ,

where SP denotes storm peak. The long-term response distribution is then calculated under the assumption that storm peak responses are independent:

$$F_{R_{max},T}(r) = [F_{R_{SP}}(r)]^{mT}, \quad (24)$$

where  $m$  is the expected number of storms per year. Storm-based methods can be classed as either “equivalent storm” methods or Monte Carlo methods. In equivalent storm methods, the distributions of the maximum response in measured storms is parameterised in terms of a statistically equivalent storm and  $F_{R_{SP}}$  is calculated in an analogous way to the distributions of all peaks or all short-term maxima. In Monte Carlo methods,  $F_{R_{SP}}$  is estimated directly from the data, removing the need to define equivalent storms.

#### Equivalent storm methods:

Although sequential environmental states are correlated, the extreme responses in adjacent environmental states are dependent only on the environmental conditions and not on the extreme response in adjacent environmental states. That is, the sequence of values of  $R_{max,i}$  are independent realisations from  $F_{R_{max}|S_i}$ , but the sequence  $S_i$  is serially correlated. The correlation in  $S_i$  introduces correlation into sequence of values of  $R_{max,i}$  (see Fig. 11).

Given that  $F_{R_{max}|S_i}$  are conditionally independent, the distribution of the extreme response over a measured storm (MS) can be calculated as

$$F_{R_{SP}|MS}(r) = \prod_{i \in [i_0, i_1]} F_{R_{max}|S_i}(r|S_i), \quad (25)$$

where  $i_0$  and  $i_1$  are the indices of the start and end of the storm.

There is a potential area for misunderstanding here. In (25), the distribution of the maximum response in the measured storm is calculated as the product of the conditionally-independent distributions of maximum responses in each short-term condition. We could extend this argument to say that the distribution of the maximum response over a given year is the product of the distributions of maximum responses in each observed short-term condition in the given year. Since this distribution is calculated as a product, the order of the terms in the sequence does not matter. Based on this, some authors argue (incorrectly) that serial correlation therefore cannot affect long-term extreme responses, since the product could be rearranged into any sequence without affecting the result. However, this argument only applies to an observed time series. As discussed in Sections 1–3, if a year of observations was simulated from the distribution of all observations, without accounting for serial correlation, then this would overestimate the frequency of occurrence of large values. The method applied in (25), for calculating the distribution of the maximum response in an observed, serially-correlated time series of environmental variables, therefore cannot be extended to uncorrelated sequences of environmental variables simulated from  $f_S(s)$ .

Two types of method have been proposed for parameterising the distribution of the maximum response in a storm. One approach to parameterising  $F_{R_{SP}|MS}$  is to model the temporal evolution of environmental states in a storm using a simplified geometric form, such as a triangle (Boccotti, 1986, 2000; Arena and Pavone, 2006), power law (Fedele and Arena, 2010; Arena et al., 2014) or exponential law (Lafate and Arena, 2016). In this approach, the temporal evolution of only one variable is modelled (such as significant wave height or wind speed) and all other variables are modelled as taking a single value, conditional on the dominant variable. If storm peaks are sufficiently separated in time, the order in which environmental states occur within a storm does not matter. Each measured storm can therefore be rearranged into a monotonic sequence, for which a simple parametric form is a reasonable approximation for the largest values. Temporal evolution methods were derived for calculating return periods of wave and crest heights, for which  $H_s$  has the dominant influence on the short-term distribution. For responses such as loads or motions of floating

structures, which are dependent on multiple variables, modelling only the temporal evolution of a single variable may not provide an adequate model for the distribution of the maximum response in a storm.

The alternative is to define an equivalent storm as a parametric statistical distribution, fitted to  $F_{R_{SP}|MS}$  for each measured storm. Tromans and Vanderschuren (1995) proposed to model the distribution of the square of the maximum response in a storm as a Gumbel distribution. A generalisation of this method was proposed in Mackay (2017) and Mackay and Johanning (2018a), where  $F_{R_{SP}|MS}$  is modelled using the GEV distribution.

For both the temporal evolution methods and distribution modelling methods, the distribution of the maximum response in any storm selected at random is calculated as

$$F_{R_{SP}}(r) = \int_{\Omega} F_{R_{SP}|\Omega}(r|\omega) f_{\Omega}(\omega) d\omega, \quad (26)$$

where  $\Omega$  is the vector of equivalent storm parameters and  $f_{\Omega}(\omega)$  is the long-term joint density function of equivalent storm parameters. Often, the integral over the joint density of the equivalent storm parameters can be simplified by taking the mean value of some equivalent storm parameters conditional on the others, for example, taking the mean GEV shape parameter and modelling the scale parameter as conditional on the location parameter. Details are provided in the works cited above.

#### Monte Carlo methods:

Broadly speaking, two types of Monte Carlo methods have been proposed for combining long-term and short-term distributions. One type uses only the observed environmental conditions; we shall refer to this as MC1. The other uses some method to generate random time series of environmental conditions, which we will refer to as MC2.

Mackay and Johanning (2018b) proposed an MC1 method where a random realisation of the maximum response in each environmental state is generated. A POT analysis is then conducted, with the generalised Pareto (GP) distribution fitted to the random storm peak responses. This process is repeated a large number of times and the results are averaged over all trials. It should be noted that results will differ, depending on whether the GP parameters are averaged and return values are calculated, or the return values are calculated for each trial and then averaged. The difference is due to the nonlinear relationship between the GP parameters and return values. Methods used for estimating GP parameters and quantiles vary in the bias and variance (see e.g. de Zea Bermudez and Kotz, 2010; Mackay et al., 2011). Whether it is appropriate to average GP parameters or return values depends on the parameter estimation method. Jonathan et al. (2021) showed that if maximum likelihood was used to estimate GP parameters, then averaging return values over each trial results in a lower bias than calculating return values from the average GP parameters. However, if the empirical Bayesian method (EBM) (Zhang, 2010) is used to estimate GP parameters, then using the average GP parameters to calculate return values results in a lower bias.

Various types of MC2 methods have been proposed. The methods generally involve three steps: (1) a method to partition the observed time series into a sequence of non-overlapping storm events; (2) a joint model for the storm peak variables; and (3) a model for the distribution of variables conditional on the storm peak values. For multivariate problems, the peak values of each variable within a storm may not occur simultaneously. There is therefore some choice in how to choose the values used for the joint model of storm peak variables. Some practitioners opt to identify one dominant variable, such as significant wave height or wind speed, and select the values of the other variables at the time of the peak value of the dominant variable (e.g. Ewans and Jonathan, 2008). In other works, ‘characteristic’ values of variables over the course of a storm are defined (Hansen et al., 2020). Alternatively, the peak values of each variable within a storm can be modelled, where the peak values are not required to occur simultaneously (Mackay and Jonathan, 2020b). To obtain a model for

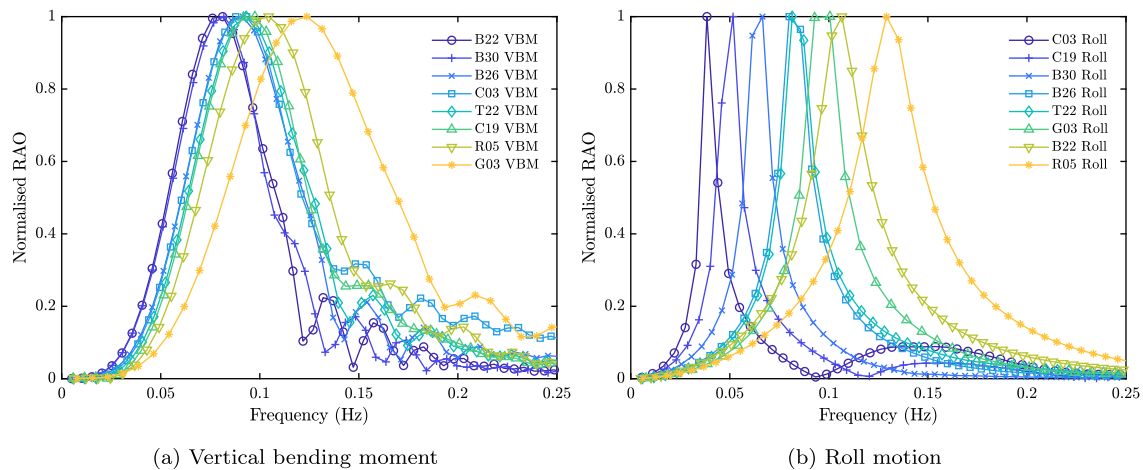


Fig. 13. Normalised RAOs considered in the study.

the joint distribution of variables, conditional on the storm peak values, measured storms can be resampled and rescaled (Ross et al., 2017; Hansen et al., 2020; Mackay and Jonathan, 2020b). Alternatively, a model for the time series evolution around the peak values can be estimated (Tendjick et al., 2019).

The MC1 method is simpler to implement than MC2, as it only requires the short-term response function and a standard univariate POT analysis. The disadvantage of MC1 is that a separate analysis is required for each response of interest. It also assumes that response function does not change significantly outside the range of observations. Some responses, such as mooring line snatch loads or wave-in-deck loads may exhibit a sharp change at some level. In comparison, MC2 is more complex to implement, but has the advantage that the extreme value analysis of environmental conditions can be separated from the extreme response analysis, and no assumptions are required about how the response function behaves outside the range of observations.

## 5.2. Examples using resampled time series

### 5.2.1. Short-term response model

We use response functions for the vertical bending moment (VBM) and roll motion of eight ships, considered in de Hauteclocque et al. (2021). The normalised VBM and roll response amplitude operators (RAOs) are shown in Fig. 13.<sup>1</sup> The natural periods for the roll responses are between 7.5 s and 25 s. Although, only two types of ship response are considered, roll is typical of many resonance driven responses, with a sharp-peaked narrow-band response. In contrast, VBM has a broader bandwidth. Thus the roll and VBM responses used here can be considered as representative of a wider range of responses of offshore structures.

The response spectrum is given by

$$S_{R|S}(\omega) = [\text{RAO}(\omega)]^2 S_{\eta|S}(\omega), \quad (27)$$

where  $S_{\eta|S}(\omega)$  is the surface elevation spectrum in sea state  $S$ . In the examples in this section it is assumed that each sea state has a JONSWAP spectrum with peak enhancement factor  $\gamma = 1.5$ . The significant response is defined as  $R_s = 2\sqrt{m_0}$ , where  $m_j$  is the  $j$ th moment of the response spectrum

$$m_j(s) = \int_0^\infty \omega^j S_{R|S}(\omega) d\omega. \quad (28)$$

<sup>1</sup> The RAOs used in this study are available at: <https://github.com/ec-benchmark-organizers/ec-benchmark/tree/master/publications/quantitative-assessment-of-contours/raos>.

Under the assumption of a linear, narrow-band response, peak responses follow a Rayleigh distribution (Naess and Moan, 2013):

$$F_{R|S}(r|s) = 1 - \exp\left(-2\left(\frac{r}{R_s}\right)^2\right). \quad (29)$$

If we make the simplifying assumption that all short-term response peaks in a sea state are independent, then

$$F_{R_{max}|S}(r|s) = [F_{R|S}(r|s)]^N, \quad (30)$$

where  $N = \nu d \times 3600$ ,  $\nu = \sqrt{(m_2(s)/m_0(s))/2\pi}$  is the mean number of peaks per unit time, and  $d$  is the duration of the sea state in hours. In reality, consecutive response peaks are correlated due to the grouping of waves. In practice, more advanced methods could be used to estimate  $F_{R_{max}|S}(r|s)$ , which do account for short-term correlation in response peaks. However, this simple model will suffice for the purpose of illustrating the effect of serial correlation in environmental conditions.

### 5.2.2. Time series model

To simulate arbitrarily long time series of  $H_s$  and  $T_z$ , we use the storm resampling approach described in Mackay and Jonathan (2020b) (in the notation of Section 5.1.3, this is an MC2-type model). The method for dividing the time series into discrete storm events is very similar to that described in Section 3.2. The details of the joint model for storm peaks are not repeated here, but is described in Mackay and Jonathan (2020b). The method was applied to three datasets of  $H_s$  and  $T_z$ , provided in the recent environmental contour benchmarking exercise (Haselsteiner et al., 2021). Comparisons of the observations and empirical isodensity contours for the resampled data are shown in Fig. 14 and further comparisons are presented in de Hauteclocque et al. (2021). Although there are some differences between the observed and simulated distributions, the simulated data are sufficiently representative of real datasets to illustrate the effect of neglecting serial correlation in the ‘all-peaks’ and ‘short-term maxima’ methods.

### 5.2.3. Results

For each dataset,  $10^4$  years of synthetic hourly time series of  $H_s$  and  $T_z$  were simulated. For each hour in the records, a random maximum response was simulated for each of the 16 response functions described in Section 5.2.1. Return periods were calculated from the empirical distribution function (EDF) for the simulated storm peak values (the maximum response in each resampled block) and for hourly values under the assumption of independence. Examples of return values of VBM for ship R05 are shown in Fig. 15. For dataset A, the effect of neglecting serial correlation is negligible for return periods above 10 years. In contrast, for datasets B and C, there are relatively large

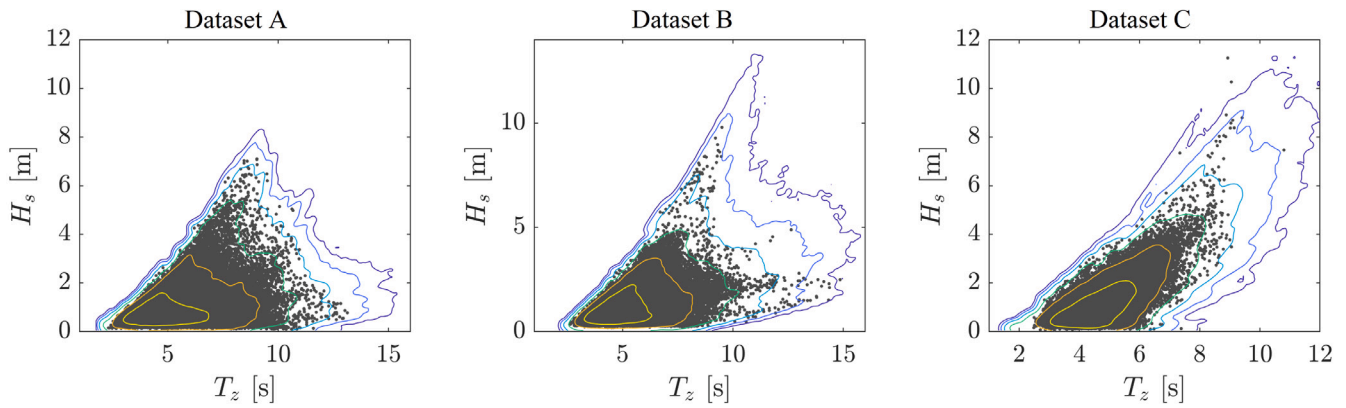


Fig. 14. Scatter plot of observed  $H_s$  and  $T_z$  (dots), overlaid with empirical isodensity contours for the resampled data. Contours shown at probability densities of  $10^{-n}$  for  $n = 1, \dots, 6$ .

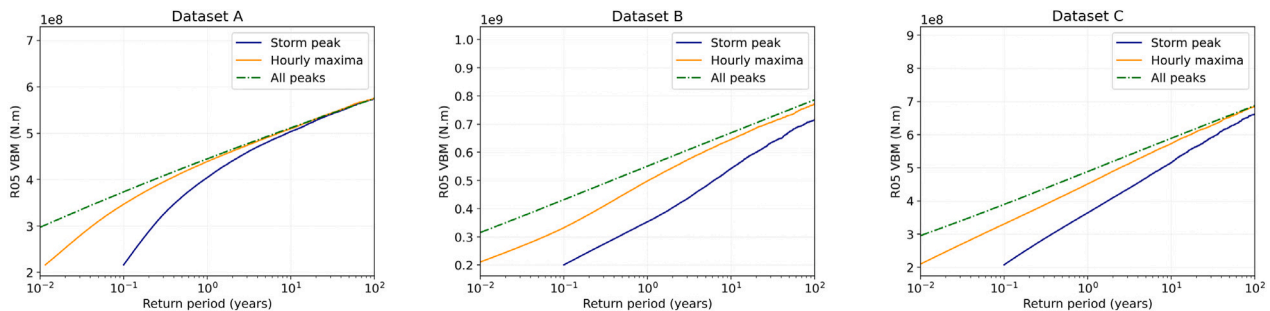


Fig. 15. Examples of return values of vertical bending moment for ship R05 for the three datasets, calculated using various methods.

differences at the 10-year level between the ‘true’ return values and those which neglect serial correlation.

Return values of VBM calculated using the all-peaks method are also shown in Fig. 15. Instead of generating time series of all peaks within the 10,000 year datasets, a discrete analogue of (19) was applied to the hourly time series to calculate the distribution of a randomly chosen response peak:

$$F_R(r) = \frac{1}{M\bar{N}} \sum_{i=1}^M N(s_i) F_{R|S_i}(r|s_i), \quad (31)$$

where  $M = 10^4 \times 8760$  is the number of hourly sea states in the 10,000-year time series. Return values were then calculated as described in Section 5.1.1. For dataset A, there is negligible difference in the return values for the all-peaks and short-term maxima methods above the 1-year level, whereas for datasets B and C there are still some differences at the 10-year level. Mackay and Johannning (2018c) showed that if the distribution of the maximum short-term response is formed under the assumption that all peaks are independent, then the return values from the all-peak method will converge to those from the short-term maxima method as  $T \rightarrow \infty$ . This is because both methods make the assumption of no serial correlation between individual peaks or sea states.

It should be noted that after the random maximum response in each hour is simulated, we now have a univariate time series of observed variables, so the situation is identical to that discussed in Section 3. Therefore, for a given response, the effect of neglecting serial correlation will depend on the shape of the tail of the distribution of storm peak response and the average shape of the ‘storms’, where in this case ‘storms’ refer to clusters of large responses. Moreover, note that for a given dataset both the shape of the tail of storm peaks and the mean storm shape will depend on the particular response function.

To summarise the results over all datasets, Fig. 16 shows the sub-asymptotic extremal index  $\theta_T$  for each of the response functions and each of the datasets. In this plot,  $\theta_T$  is calculated as the ratio of the return period from the short-term maxima method to that from the

storm peak method. There are similar trends for each response function. For dataset A,  $\theta_T$  increases to close to 1 at  $T = 100$  years, whereas for datasets B and C, there is more scatter, but lower values on average.

To assess the influence of the shape of the tail of storm peak response  $F_{R_{SP}}$ , Fig. 17(a) shows a scatter plot of  $\theta_T$  for  $T = 25$  years against the ratio  $R_{10}/R_{100}$ , where  $R_T$  is the  $T$ -year return value of the response. The ratio  $R_{10}/R_{100}$  is a proxy for the shape of the tail of the distribution of  $F_{R_{max}|SP}$  around the 25-year return level, with lower values of the ratio corresponding to higher values of  $\xi$ .<sup>2</sup> The relatively tight banding of the results over all 16 responses evaluated for the three datasets, indicates that the mean storm shape is fairly constant for each case. The colour of the points in Fig. 17 indicates the period of the peak response. It is apparent that the longer period responses tend to have distributions of storm peak response with longer tails, and hence small values of  $R_{10}/R_{100}$ . This is related to the shape of the joint distribution of  $H_s$  and  $T_z$ , discussed in Section 4. For the lower period responses, the tail of the distribution of storm peak response will be shorter, since the minimum period at a given  $H_s$  is limited by wave breaking. Fig. 17(b) shows the relative error in the 25-year return value for short-term maxima methods compared to storm-based methods (defined as  $(\bar{R}_{25} - R_{25})/R_{25}$ , where  $\bar{R}_{25}$  is the estimate from the short-term maxima method and  $R_{25}$  is storm-based estimate). When the tail of the distribution of  $F_{R_{max}|SP}$  is short and  $R_{10}/R_{100} \geq 0.6$ , the relative error is negligible. However, for smaller values of  $R_{10}/R_{100}$ , the biases can be very large — just under 100% in the largest case. For this case, a structure designed using response estimates from a short-term maxima method would be extremely conservative.

Finally, we consider the differences in the effect of serial correlation with and without short-term variability. Fig. 16 shows the values of

<sup>2</sup> The correspondence between the ratio and the tail shape can be seen by considering the Hill estimator (Hill, 1975) for tail index. If we denote an ordered sample  $x_{(1)} \leq \dots \leq x_{(n)}$  and denote the ratio  $v_{ik} = x_{(n-i+1)}/x_{(n-k)}$ , then the Hill estimator is  $\hat{\xi}_k = \frac{1}{k} \sum_{i=1}^k \log(v_{ik})$ .

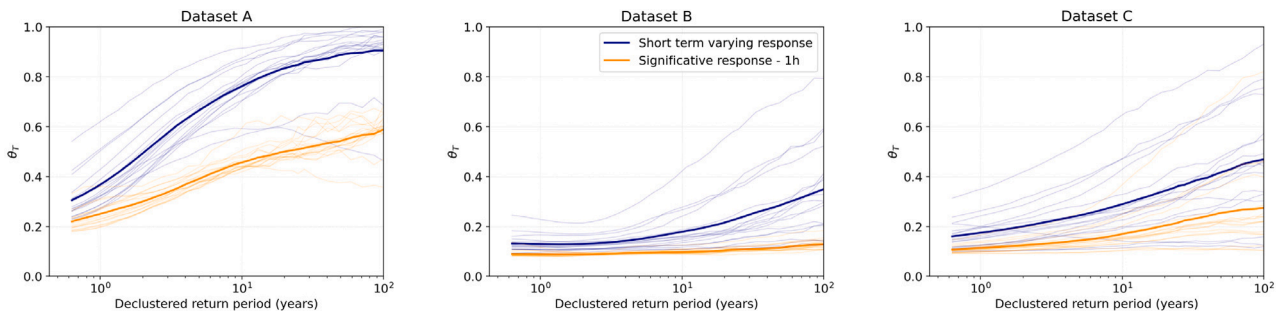


Fig. 16. Sub-asymptotic extremal index for three dataset and various ship responses. Thin lines are values for each response, bold lines are mean over all responses.

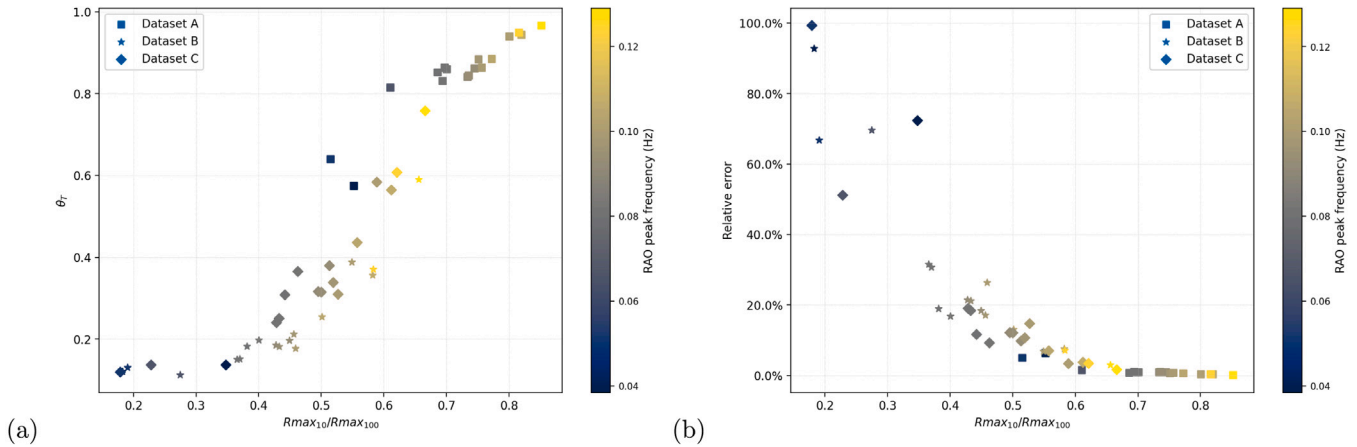


Fig. 17. (a) Sub-asymptotic extremal index and relative error at  $T = 25$  years, for each of the 16 responses and three datasets.  $R_{10}/R_{100}$  is the ratio between the 10-year and 100-year response, used as a proxy to characterise tail shape. (b) Relative error in 25-year return value for short-term maxima methods compared to storm-based methods. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$\theta_T$  for the significant response,  $R_s$ , for each of the response functions (also presented in de Hauteclocque et al., 2021). It is apparent that including short-term variability increases the value of  $\theta_T$ . As discussed before, a change in  $\theta_T$  could be due to either a change in the shape of the tail of the distribution of storm peak response, or a change in the mean storm shape. Quantile–quantile plots of the random storm peak response against the storm peak significant response indicated a linear relationship (not shown). So for the present examples, it can be inferred that short-term variability influences the mean storm shape (where the storm shape is defined in the same way as in Section 3.2, as the ordered sequence of response values over the storm, averaged over storms with similar peak response values). However, for other response functions, short-term variability could also influence shape the distribution of storm peak response.

When the short-term distribution is Rayleigh, given by (29), the distribution of the maximum response in  $N$  cycles tends to a Gumbel distribution as  $N \rightarrow \infty$ , (a GEV distribution with CDF given in (9) and  $\xi = 0$ ) with parameters  $\mu = R_s \sqrt{\log(N)}/2$  and  $\sigma = R_s/2\sqrt{2\log(N)}$ . As a measure of the variability of  $R_{max}$  we define a coefficient of variability of the short-term maximum response as  $V_{ST} = \text{STD}(R_{max}(s))/\text{mode}(R_{max}(s)) = \pi/(2\sqrt{6}\log(N))$ . Here we have normalised by the most probable maximum value,  $\mu$ , rather than the expected maximum value to simplify the expression. In the case that  $V_{ST}$  is large relative to the variation of  $R_s$  within storms, the presence of short-term variability dominates any serial correlation in significant response, making time series of  $R_{max}(s)$  effectively independent. In the present examples,  $V_{ST}(s)$  is relatively constant at around 10% for all response functions and the ranges of sea states considered. So clusters of extreme values in the time series of  $R_s$  and  $R_{max}$  will occur at approximately the same times (similar to the case illustrated in Fig. 11).

To investigate the effect of additional short-term variability on the mean storm shape, consider the following simple example. We assume

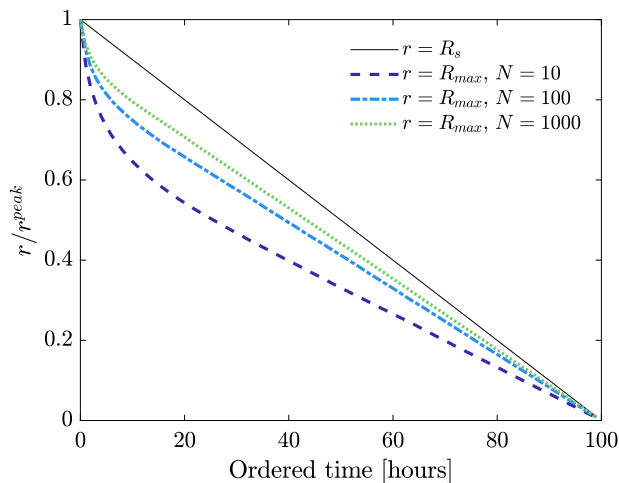
that over the course of a storm,  $R_s$  increases linearly from zero to  $R_s^{peak}$  over 100 h, and that the distribution of the maximum response in each hour follows a Gumbel distribution with parameters given above. We simulate random values of  $R_{max}$  in each hour of the storm for  $N = 10, 100$ , and 1000. For each simulation, the values of  $R_{max}$  over the storm are ordered, and the process is repeated 1000 times to calculate the mean shape of the ordered values of  $R_{max}$ . The results are shown in Fig. 18. It is evident that the mean shape of the ordered values of  $R_{max}$  is more peaked than ordered values of  $R_s$ , and that the peakedness increases with  $V_{ST}(s)$ , as expected. This implies that the reason that  $\theta_T$  is reduced when considering random responses rather than significant responses is that the mean storm shapes are more peaked.

## 6. Conclusions

Sections 1 and 2 present theoretical arguments suggesting how a rational extreme value analysis from serially-correlated data should be conducted in the ocean engineering context. We show that basing analysis on storm peaks or cluster maxima of variables leads to intuitive interpretations of return values and return periods. We have presented theoretical arguments and simple examples to show how neglecting serial correlation leads to positive bias in the estimation of extreme events.

A new definition of a sub-asymptotic extremal index,  $\theta_T$ , was introduced. It was shown  $\theta_T(x) = \tilde{T}(x)/T(x)$ , where  $T$  is the true return period of  $x$ , and  $\tilde{T}(x)$  is the return period when serial correlation is neglected. It was also shown that the bias in return values introduced by neglecting serial correlation can be quantified in terms of  $\theta_T$ . The sub-asymptotic extremal index is more relevant to offshore engineering than the standard extremal index, which describes the effect of serial correlation at asymptotically high levels, corresponding to return periods outside the usual range of interest. We show that the value of





**Fig. 18.** Effect of short-term variability on mean storm shape. Solid line shows ordered values of  $R_s$  over a hypothetical storm. Other lines show mean of ordered values of 1-hour maximum response,  $R_{max}$ , over the same storm, where there are  $N$  peaks per hour, and the response follows a Rayleigh distribution.

$\theta_T(x)$  is dependent on the average shape of storm events and the shape parameter of storm peak values,  $\xi$ . It was shown that larger values of  $\xi$  lead to larger biases at a given return period. For the examples considered here, the effect of neglecting serial correlation is reduced at higher return periods. The rate at which  $\theta_T$  increases with  $T$  is also dependent on  $\xi$ .

The examples presented show how neglecting serial correlation can lead to biased, overly-conservative estimates of environmental contours and long-term extreme responses. For the roll motion and bending moment response functions considered in Section 5, it was shown that when serial correlation was neglected, the 25-year return values of the long-term extreme response could have relative errors of nearly 100% in some cases, although in other cases the relative errors were much smaller. Accounting appropriately for the effect of serial correlation can therefore reduce over-conservatism in offshore design, leading to more rational design decisions.

#### CRediT authorship contribution statement

**Ed Mackay:** Conceptualization, Methodology, Formal analysis, Writing – original draft. **Guillaume de Hauteclocque:** Methodology, Formal analysis, Writing – original draft. **Erik Vanem:** Writing – review & editing. **Philip Jonathan:** Formal analysis, Writing – original draft.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

EM was funded by the EPSRC Supergen Offshore Renewable Energy Hub, United Kingdom [grant no: EP/S000747/1] Flexible Fund project “Improved Models for Multivariate Metocean Extremes (IMEX)”. EM also acknowledges funding from the Tidal Stream Industry Energiser Project (TIGER), France, a European Union INTERREG V A France (Channel) England Research and Innovation Programme, which is co-financed by the European Regional Development Fund (ERDF), United Kingdom.

#### References

- Ancona-Navarrete, J., Tawn, M.A., 2000. A comparison of methods for estimating the extremal index. *Extremes* 3, 5–38. <http://dx.doi.org/10.1023/A:1009993419559>.
- Anderson, C.W., 1990. Discussion of “models for exceedances over high thresholds” by A.C. Davison and R.L. Smith. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 52, 393–442.
- Arena, F., Malara, G., Romolo, A., 2014. On long-term statistics of high waves via the equivalent power storm model. *Probab. Eng. Mech.* 38, 103–110.
- Arena, F., Pavone, D., 2006. The return period of non-linear high wave crests. *J. Geophys. Res.* 111, C08004.
- Battjes, J., 1970. Long-Term Wave Height Distribution at Seven Stations Around the British Isles. Report No. A 44, National Institute of Oceanography, Godalming.
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. *Statistics of Extremes: Theory and Applications*. Wiley, Chichester, UK.
- Boccotti, P., 1986. On coastal and offshore structure risk analysis. In: *Excerpta of the Italian Contribution to the Field of Hydraulic Engng.*, Vol. 1. pp. 19–36.
- Boccotti, P., 2000. *Wave Mechanics for Ocean Engineering*. Elsevier, New York.
- Chai, W., Leira, B.J., 2018. Environmental contours based on inverse SORM. *Mar. Struct.* 60, 34–51.
- Chavez-Demoulin, V., Davison, A., 2012. Modelling time series extremes. *Revstat* 10, 109–133.
- Coles, S., 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer, <http://dx.doi.org/10.1198/tech.2002.s73>.
- Coles, S., Heffernan, J., Tawn, J., 1999. Dependence measures for extreme value analyses. *Extremes* 2, 339–365.
- Davis, R., Mikosch, T., 2009. The extremogram: A correlogram for extreme events. *Bernoulli* 15, 977–1009.
- de Hauteclocque, G., Mackay, E., Vanem, E., 2021. Quantitative assessment of environmental contour approaches. <http://dx.doi.org/10.13140/RG.2.2.10068.12161>, Submitted to this special issue.
- de Zea Bermudez, P., Kotz, S., 2010. Parameter estimation of the generalized Pareto distribution-Parts I&II. *J. Statist. Plann. Inference* 140 (6), 1353–1388. <http://dx.doi.org/10.1016/j.jspi.2008.11.019>.
- Derbanne, Q., de Hauteclocque, G., Dumont, M., 2017. How to account for short-term and long-term variability in the prediction of the 100 years response?. *International Conference on Offshore Mechanics and Arctic Engineering*, Vol. Volume 3A: Structures, Safety and Reliability, <http://dx.doi.org/10.1115/OMAE2017-61701>.
- Derbanne, Q., de Hauteclocque, G., 2019. A new approach for environmental contour and multivariate de-clustering. In: *38th International Conference on Ocean, Offshore and Arctic Engineering*. Glasgow. p. OMAE2019/95993. <http://dx.doi.org/10.1115/OMAE2019-95993>.
- Eastoe, E.F., Tawn, J.A., 2012. Modelling the distribution of the cluster maxima of exceedances of subsymptotic thresholds. *Biometrika* 99 (1), 43–55. <http://dx.doi.org/10.1093/biomet/asr078>.
- Evans, K., Jonathan, P., 2008. The effect of directionality on Northern North Sea Extreme wave design criteria. *J. Offshore Mech. Arct. Eng.* 130 (4), <http://dx.doi.org/10.1115/1.2960859>.
- Fedele, F., Arena, F., 2010. Long-term statistics and extreme waves of sea storms. *J. Phys. Oceanogr.* 40, 1106–1117.
- Fogle, J., Agarwal, P., Manuel, L., 2008. Towards an improved understanding of statistical extrapolation for wind turbine extreme loads. *Wind Energy* 11 (6), 613–635. <http://dx.doi.org/10.1002/we.303>.
- Forristall, G., 2008. How should we combine long and short term wave height distributions? In: *27th International Conference on Offshore Mechanics and Arctic Engineering*. Estoril, Portugal. pp. OMAE2008-58012.
- Hansen, H.F., Randell, D., Zeeberg, A.R., Jonathan, P., 2020. Directional-seasonal extreme value analysis of North Sea storm conditions. *Ocean Eng.* 195, 106665. <http://dx.doi.org/10.1016/j.oceaneng.2019.106665>.
- Haselsteiner, A., Coe, R.G., Manuel, L., Chai, W., Leira, B., Clarindo, G., Guedes Soares, C., Hannesdóttir, Á., Dimitrov, N., Sander, A., Ohlendorf, J.H., Thoben, K.D., de Hauteclocque, G., Mackay, E., Jonathan, P., Qiao, C., Myers, A., Rode, A., Hildebrandt, A., Schmidt, B., Vanem, E., Huseby, A.B., 2021. A benchmarking exercise for environmental contours. *Ocean Eng.* 236, 109504. <http://dx.doi.org/10.1016/J.OCEANENG.2021.109504>.
- Haselsteiner, A.F., Coe, R.G., Manuel, L., Nguyen, P.T.T., Martin, N., Eckert-Gallup, A., 2019. A benchmarking exercise on estimating extreme environmental conditions: Methodology & baseline results. In: *38th International Conference on Ocean, Offshore and Arctic Engineering*. Glasgow, UK. pp. OMAE2019-96523.
- Haselsteiner, A.F., Ohlendorf, J.-H., Wosniok, W., Thoben, K.-D., 2017. Deriving environmental contours from highest density regions. *Coast. Eng.* 123, 42–51.
- Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. *Ann. Statist.* 3 (5), 1163–1174. <http://dx.doi.org/10.1214/aos/1176343247>.
- Hsing, T., 1987a. On the characterisation of certain point processes. *Stochastic Process. Appl.* 26, 297–316.
- Hsing, T., 1987b. On the characterization of certain point processes. *Stochastic Process. Appl.* 26 (C), 297–316. [http://dx.doi.org/10.1016/0304-4149\(87\)90183-9](http://dx.doi.org/10.1016/0304-4149(87)90183-9).
- Hsing, T., Husler, J., Leadbetter, M.R., 1988. On the exceedance point process for a stationary sequence. *Probab. Theory Related Fields* 78, 97–112.
- Huseby, A.B., Vanem, E., Natvig, B., 2013. A new approach to environmental contours for ocean engineering applications based on direct Monte Carlo simulations. *Ocean Eng.* 60, 125–135.

- Huseby, A., Vanem, E., Natvig, B., 2015. Alternative environmental contours for structural reliability analysis. *Struct. Saf.* 54, 32–45.
- Jonathan, P., Ewans, K.C., 2013. Statistical modelling of extreme ocean environments with implications for marine design: a review. *Ocean Eng.* 62, 91–109.
- Jonathan, P., Ewans, K., Forristall, G., 2008. Statistical estimation of extreme ocean environments: The requirement for modelling directionality and other covariate effects. *Ocean Eng.* 35 (11–12), 1211–1225. <http://dx.doi.org/10.1016/j.oceaneng.2008.04.002>.
- Jonathan, P., Randell, D., Wadsworth, J., Tawn, J., 2021. Uncertainties in return values from extreme value analysis of peaks over threshold using the generalised Pareto distribution. *Ocean Eng.* 107725. <http://dx.doi.org/10.1016/j.oceaneng.2020.107725>.
- Krogstad, H.E., 1985. Height and period distributions of extreme waves. *Appl. Ocean Res.* 7 (3), 158–165. [http://dx.doi.org/10.1016/0141-1187\(85\)90008-2](http://dx.doi.org/10.1016/0141-1187(85)90008-2).
- Lafage, V., Arena, F., 2016. A new equivalent exponential storm model for long-term statistics of ocean waves. *Coast. Eng.* 116, 133–151. <http://dx.doi.org/10.1016/j.coastaleng.2016.06.011>.
- Leadbetter, M.R., 1991. On a basis for ‘peaks over threshold’ modeling. *Statist. Probab. Lett.* 12, 357–362.
- Leadbetter, M.R., Lindgren, G., Rootzen, H., 1983. *Extremes and Related Properties of Random Sequences and Series*. Springer, New York, USA.
- Ledford, A.W., Tawn, J.A., 2003. Diagnostics for dependence within time series extremes. *J. R. Stat. Soc. B* 65, 521–543.
- Mackay, E., 2017. Return periods of extreme loads on wave energy converters. In: 12th European Wave and Tidal Energy Conference. Cork, Ireland.
- Mackay, E., Challenor, P.G., Bahaj, A.B.S., 2010. On the use of discrete seasonal and directional models for the estimation of extreme wave conditions. *Ocean Eng.* 37, 425–442. <http://dx.doi.org/10.1016/j.oceaneng.2010.01.017>.
- Mackay, E., Challenor, P.G., Bahaj, A.S., 2011. A comparison of estimators for the generalised Pareto distribution. *Ocean Eng.* 38 (11–12), 1338–1346. <http://dx.doi.org/10.1016/j.oceaneng.2011.06.005>.
- Mackay, E., Haselsteiner, A.F., 2021. Marginal and total exceedance probabilities of environmental contours. *Mar. Struct.* 75, 102863. <http://dx.doi.org/10.1016/j.marstruc.2020.102863>.
- Mackay, E., Johanning, L., 2018a. A generalised equivalent storm model for long-term statistics of ocean waves. *Coast. Eng.* 140, 411–428.
- Mackay, E., Johanning, L., 2018b. A simple and robust method for calculating return periods of ocean waves. In: Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE, Vol. 11B. American Society of Mechanical Engineers (ASME), pp. OMAE2018–78729. <http://dx.doi.org/10.1115/omae2018-78729>.
- Mackay, E., Johanning, L., 2018c. Long-term distributions of individual wave and crest heights. *Ocean Eng.* 165, 164–183. <http://dx.doi.org/10.1016/j.oceaneng.2018.07.047>.
- Mackay, E., Jonathan, P., 2020a. Assessment of return value estimates from stationary and non-stationary extreme value models. *Ocean Eng.* 207, 107406. <http://dx.doi.org/10.1016/j.oceaneng.2020.107406>.
- Mackay, E., Jonathan, P., 2020b. Estimation of environmental contours using a block resampling method. In: 39th International Conference on Ocean, Offshore and Arctic Engineering. pp. OMAE2020–18308. <http://dx.doi.org/10.1115/OMAE2020-18308>.
- Mackay, E., Jonathan, P., 2021. Sampling properties and empirical estimates of extreme events. *Ocean Eng.* 239, 109791. <http://dx.doi.org/10.1016/j.oceaneng.2021.109791>.
- Naess, A., 1984. On the long-term statistics of extremes. *Appl. Ocean Res.* 6 (4), 227–228. [http://dx.doi.org/10.1016/0141-1187\(84\)90061-0](http://dx.doi.org/10.1016/0141-1187(84)90061-0).
- Naess, A., Gaidai, O., 2009. Estimation of extreme values from sampled time series. *Struct. Saf.* 31, 325–334.
- Naess, A., Moan, T., 2013. *Stochastic Dynamics of Marine Structures*. Cambridge University Press, pp. 1–410. <http://dx.doi.org/10.1017/CBO9781139021364>.
- Nandagopalan, S., 1994. On the multivariate extremal index. *J. Res. Natl. Inst. Stand. Technol.* 99 (4), 543. <http://dx.doi.org/10.6028/jres.099.052>.
- Nordenström, N., 1969. Methods for Predicting Long Term Distributions of Wave Loads and Probability of Failure for Ships. DNV Report No. 69-21-S, Det Norske Veritas (Høvik).
- Ross, E., Randell, D., Ewans, K., Feld, G., Jonathan, P., 2017. Efficient estimation of return value distributions from non-stationary marginal extreme value models using Bayesian inference. *Ocean Eng.* 142, 315–328. <http://dx.doi.org/10.1016/j.oceaneng.2017.06.059>.
- Sagrilo, L.V.S., Naess, A., Doria, A.S., 2011. On the long-term response of marine structures. *Appl. Ocean Res.* 33 (3), 208–214. <http://dx.doi.org/10.1016/j.apor.2011.02.005>.
- Tawn, J.A., 1990. Discussion of “models for exceedances over high thresholds” by A.C. Davison and R.L. Smith. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 52, 393–442.
- Tendjick, S., Ross, E., Randell, D., Jonathan, P., 2019. A non-stationary statistical model for the evolution of extreme storm events. *Environmetrics* 30, e2541.
- Tromans, P.S., Vanderschuren, L., 1995. Response based design conditions in the north sea: application of a new method. In: Offshore Technology Conference, Houston (OTC-7683).
- Tucker, M.J., 1989. An improved ‘Battjes’ method for predicting the probability of extreme waves. *Appl. Ocean Res.* 11 (4), 212–218. [http://dx.doi.org/10.1016/0141-1187\(89\)90020-5](http://dx.doi.org/10.1016/0141-1187(89)90020-5).
- Vanem, E., 2018. A simple approach to account for seasonality in the description of extreme ocean environments. *Mar. Syst. Ocean Technol.* 13, 63–73. <http://dx.doi.org/10.1007/s40868-018-0046-6>.
- Vanem, E., 2019. 3-dimensional environmental contours based on a direct sampling method for structural reliability analysis of ships and offshore structures. *Ships Offshore Struct.* 14 (1), 74–85. <http://dx.doi.org/10.1080/17445302.2018.1478377>.
- Winterstein, S.R., Engebretsen, K., 1998. Reliability-based prediction of design loads and responses for floating ocean structures. In: OMAE 1998: 17th International Conference on Offshore Mechanics and Arctic Engineering.
- Winterstein, S.R., Ude, T.C., Cornell, C.A., Bjerager, P., Haver, S., 1993. Environmental parameters for extreme response: Inverse FORM with omission factors. In: 6th International Conference on Structural Safety & Reliability (ICOSSAR)..
- Zhang, J., 2010. Improving on estimation for the generalized Pareto distribution. *Technometrics* 52 (3), 335–339. <http://dx.doi.org/10.1198/TECH.2010.09206>.