Estimation of associated values from conditional extreme value models

Ross Towe[a], David Randell[b], Jennifer Kensler[c], Graham Feld[d], Philip Jonathan[a,e,*]

[a]*Shell Research Limited, London SE1 7NA, United Kingdom.*
[b]*Shell Global Solutions International BV, 1031 HW Amsterdam, The Netherlands.*
[c]*Shell International Exploration and Production, Houston TX 77082-3101, USA.*
[d]*Shell UK Ltd., Aberdeen AB12 3FY, United Kingdom.*
[e]*Department of Mathematics and Statistics, Lancaster University LA1 4YF, United Kingdom.*

Accepted for publication, January 2023

Abstract

The design and reanalysis of offshore and coastal structures usually requires the estimation of return values for dominant metocean variables (such as significant wave height) and associated values for other variables (such as peak spectral period or wind speed) from a finite sample of data; these are typically estimated using extreme value analysis. Yet the parameters of extreme value models can only be estimated with error from finite data. Different choices available to summarise uncertain information about the characteristics of the tail of a multivariate distribution in a small number of summary statistics (such as return values and associated values) complicates their estimation, especially for small sample sizes: choices regarding the ordering of mathematical operations lead to estimators of return values and associated values with different finite sample bias and variance characteristics. The current work extends a previous study (Jonathan et al. 2021) into the performance of estimators for marginal return values in the presence of sampling uncertainty, to estimators of associated values based on the bivariate conditional extremes model (Heffernan and Tawn 2004) and competitors. Using a large designed simulation experiment, we explore the performance of combinations of 12 different estimators and three bivariate model candidates. The rich set of results from the simulation experiment are reported and explained in detail. Briefly: (a) calculation of associated values is only always feasible from small samples using two of the 12 estimators, which should be preferred; (b) estimators exploiting the median rather than the mean to summarise a distribution are more robust, and should also be preferred, especially for small sample sizes; (c) extreme value models incorporating appropriate descriptions of marginal and dependence provide better estimation of associated values for larger sample size; and (d) summarising the joint tail of metocean variables (in terms of return values and associated values) should be avoided where possible, in favour of probabilistic risk analysis of structural failure incorporating full uncertainty propagation.

Keywords: extreme; return value; associated value; multivariate; metocean; conditioning

1. Introduction

Estimating summary statistics to characterise extremes of multivariate distributions is important in environmental risk analysis. Return values are widely used to quantify the severity of a univariate environmental process, and their estimation is required by design standards such as NORSOK N-006 (2015) and ISO19901-1 (2015). For a bivariate process, it is common practice (e.g. Haver 1985, Tromans and Vanderschuren 1995) to estimate a return value corresponding to some return period of N years for the dominant variable $X > 0$, and then estimate an *associated value* for the other variable $Y > 0$, given that the dominant variable is equal to its return value.

Return value estimation typically requires the adoption of an extreme value model for large values (such as peaks over threshold, or block maxima) of X. The quality of inference is influenced by sample size and other effects, including the action of covariates (such as direction, season and water depth). For very large, representative samples, return values can be estimated with low bias and uncertainty. However, as shown by Jonathan et al. (2021) for small samples, return value estimation is problematic; care should be taken in deciding the most appropriate approach to characterising the distribution of the tail.

Estimating associated values requires the adoption of a model for the joint structure of large X and associated Y. Historically, the pragmatic engineering approach was to fit a straight line through data for the joint tail of X and Y, possibly after marginal transformations to appropriate scale. The statistics literature provides a number of theoretically well-founded approaches to tackle this problem, and in particular to quantify the quality of models and estimates.

*Corresponding author `philip.jonathan@shell.com`

Given that access to data and statistical tools is now straightforward, statistical approaches should be preferred in general because of their careful mathematical underpinning. Ewans and Jonathan (2014) gives an introduction for the ocean engineering community. Typically, as outlined in Section 5, a statistical model for the joint distribution of X and Y is expressed as the combination of models for the marginal distributions of X and of Y, and a dependence or copula model for the dependence between the variates on standard marginal scale. There are many types of dependence model, which impose different characteristics on the joint distribution of X and Y when one or both are large (e.g. Gudendorf and Segers 2010). Coles et al. (1999) showed that different classes of extremal dependence exist, and Heffernan (2000) gives a useful directory of tail dependence forms for different bivariate cases. Under *asymptotic dependence*, the conditional probability $\Pr(Y_S > x | X_S > x)$ converges to a positive constant as x increases to infinity, where $X_S \in \mathbb{R}$ and $Y_S \in \mathbb{R}$ represent X and Y transformed independently to a common standard marginal scale; that is, large values of X_S and Y_S continue to occur together. The so-called logistic or Gumbel model (see Section 6.1) provides one example. In contrast, under *asymptotic independence*, the conditional probability $\Pr(Y_S > x | X_S > x)$ converges to zero with increasing x; now, on standard marginal scale, the probability of seeing large X_S and Y_S tends to zero. The bivariate Gaussian distribution with standard margins, and correlation $\rho \in (0, 1)$ (again see Section 6.1) is a typical example. Estimating associated values in the presence of uncertainty (e.g. due to a small sample for inference) is clearly more challenging therefore than estimation of a marginal return value, since it involves inference for many more models and parameters, and additional model choice.

There is a growing literature on joint and conditional extremes modelling. In the statistics literature, Tawn (1988a), Tawn (1988b) and Coles and Tawn (1991) provide early guidance. The dependence models used in the current study are based on the conditional extremes model of Heffernan and Tawn (2004) describing asymptotic independence (and asymptotic dependence at a corner of the parameter space), on standard copulas (e.g. Joe 2014) and on linear regression. There is a close correspondence between the underpinning theory of joint and conditional extremes and that of spatial extremes (e.g. Davison et al. 2012) and temporal extremes (e.g. Chavez-Demoulin and Davison 2012). A number of extensions of the conditional extremes model of Heffernan and Tawn (2004) relevant to the current study have been proposed, including Tendijck et al. (2021). In the engineering literature, the work of Haver (1987), Bitner-Gregersen and Haver (1989) and Mathisen and Bitner-Gregersen (1990) provide motivation. More recent papers using the conditional extremes model for metocean design include Jonathan et al. (2010) and Jonathan et al. (2012), and Bitner-Gregersen (2015). The estimation of associated values has features in common with environmental design contour estimation (e.g. Montes-Iturrizaga and Heredia-Zavoni 2017, Chai and Leira 2018, Ross et al. 2020, Haselsteiner et al. 2021 and references within).

Motivation, objective and outline of paper

The ocean environment can be viewed as a non-stationary multivariate spatio-temporal process. Characterising (multivariate) extremes of that environment therefore requires adequate characterisation of (a) the effects of time, space and multiple covariates (e.g. direction, fetch, water depth; Davison and Smith 1990, Chavez-Demoulin and Davison 2005, Zanini et al. 2020) on the marginal properties of the process, as well as (b) the extremal temporal and spatial dependence present (e.g. Chavez-Demoulin and Davison 2012, Davison et al. 2012, Tendijck et al. 2019, Shooter et al. 2022), and (c) the cross-dependence between different components (e.g. wind, wave, current; Heffernan and Tawn 2004, Gudendorf and Segers 2010, Rootzen et al. 2018, Tendijck et al. 2021) of the process. Different model estimation schemes are available (as discussed in Jonathan et al. 2021, including e.g. maximum likelihood, Smith and Naylor 1987; moment methods, Furrer and Naveau 2007; empirical Bayes, Zhang 2010; maximum entropy, Petrov et al. 2013, Chen et al. 2021), some of which have been demonstrated to be advantageous in certain settings. There is a large and growing literature on these topics in statistics and ocean engineering journals. In real-world applications, all of these issues need to be considered in a careful analysis. However, the purpose of this article is different: here, we seek to quantify bias (and uncertainty) in estimates of associated values using typical extreme value models (including the conditional extremes model) and standard likelihood inference, as a function of the size and characteristics of the sample, and various specific choices of estimator under uncertainty, from samples with known distributional characteristics. Of course, other characterisations of multivariate extremes (including e.g. environmental contours) are likely to suffer similar bias effects to those reported here for associated values.

The objective of the current work is to (a) provide a precise description of some estimators for associated values in the presence of uncertainty; (b) assess the performance of estimators in combination with different statistical models for joint and conditional extremes, in a simulation study using samples generated from plausible combinations of marginal and dependence models and parameter values; and (c) provide the practitioner with quantitative guidance regarding appropriate model and estimator choice for metocen design purposes.

The layout of the article is as follows. To motivate subsequent development, Section 2 provides examples of bivariate samples for variables (X, Y), typical of metocean applications, for which estimates of return values (for X) and associate values (for $Y|X$) might be required in practice. Section 3 summarises possible estimators for the marginal return value

of the conditioning variate X, drawing in part on Jonathan et al. (2021). Building on these definitions, Section 4 then introduces the conditional extremes model, and Section 5 introduces a number of plausible estimators for the associated value. Next, Section 6 describes a large simulation experiment, designed to explore the performance of different combinations of estimators and statistical models for associated values over a range of conditions of practical metocean interest. Results of the simulation study are reported in Section 6.3 (for return values) and Section 7 (for associated values). Section 8 provides discussion and conclusions. The appendix outlines estimation of the distribution of annual maxima, from the distribution of individual events. The appendix also provides supporting evidence for estimator performance, referenced from Section 7. Supplementary plots and software for the simulation study are given at Jonathan (2022).

2. Motivating application

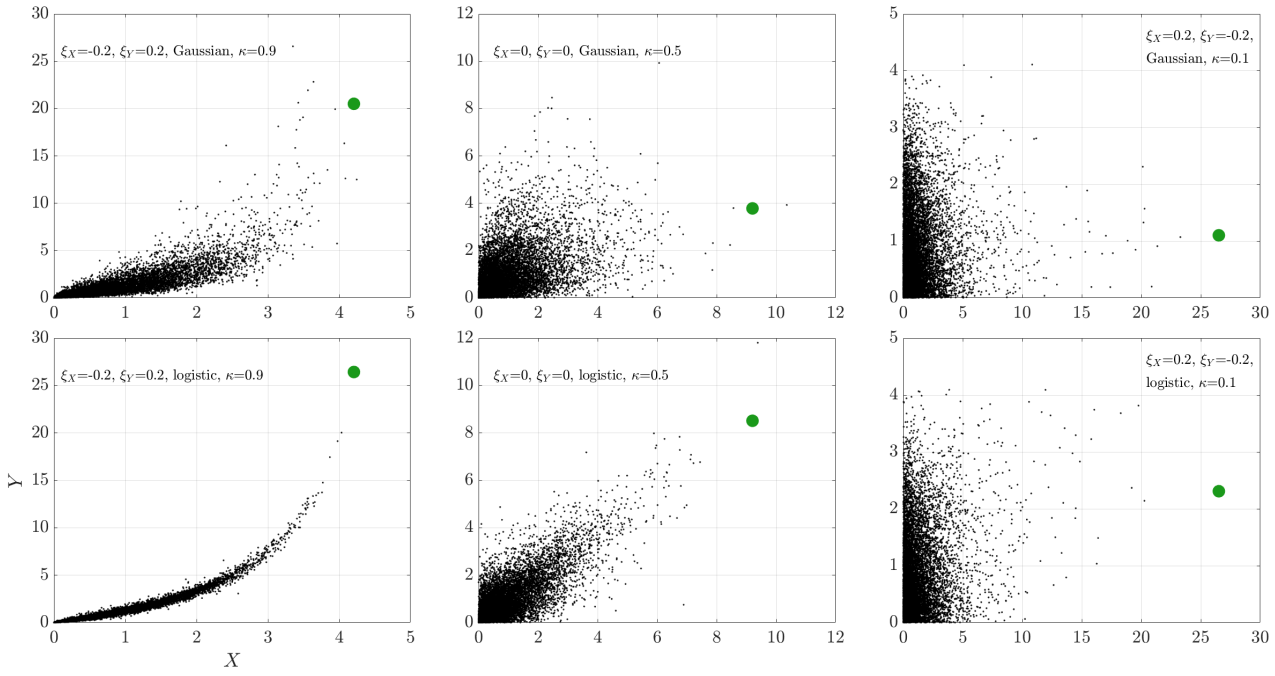

Figure 1: Illustrative samples of size 10,000, simulated under bivariate models with marginal and dependence characteristics as labelled (for notation and simulation procedure, see Section 6.1). Green discs indicate the location of the return value (in X, for the same period as the sample) and the associated value of Y. The marginal distributions of X and Y are assumed to be generalised Pareto, with zero threshold, unit scale and shapes ξ_X and ξ_Y. The dependence between X and Y follows either a Gaussian or logistic copula, with dependence strength $\kappa \in [0, 1]$, with large κ indicating high dependence.

Figure 1 gives plots for pairs of variables, with different marginal and dependence characteristics illustrative of the kinds of dependence that might be expected in metocean applications. We are interested in estimating the location of the green disc in each case, corresponding to a marginal return value x of X (with return period equal to the period of the sample) and the corresponding *associated value* of Y, namely its conditional mean $\mathbb{E}[Y|X = x]$. Estimating pairs of return values and associated values in this way is challenging, because it involves establishing an appropriate model for the bivariate cloud of points, and in particular for the cloud's characteristics when at least X is large.

3. Estimators of marginal return values

Suppose that random variable $A_X > 0$ represents the maximum value of some physical quantity X (such as storm peak H_S) per annum, and that $A_{X,N} > 0$ represents for corresponding N-year maximum for $N \in \mathbb{Z}_{\geq 1}$. Suppose that the distributions of A_X and $A_{X,N}$ are known conditional on uncertain real-valued extreme value model parameters $\boldsymbol{Z}$, the values of which are estimated by fitting to a sample of data, for example. If the conditional distribution of threshold exceedances of X is described by a generalised Pareto (GP) distribution, then $\boldsymbol{Z}$ might be the set of threshold, shape and scale parameters for that distribution. Given $\boldsymbol{Z}$, the distributions of X, A_X and $A_{X,N}$ can be written in closed

form (if, for example, we assume for A_X and $A_{X,N}$ that occurrences of events X are Poisson-distributed with known annual rate). Then as described in Jonathan et al. (2021), we can consider (at least) four different estimators of the N-year return value, namely

$$\begin{aligned} q_1 &\coloneqq F^{-1}_{A_X|\boldsymbol{Z}}(1-1/N \mid \mathbb{E}_{\boldsymbol{Z}}[\boldsymbol{Z}]) \\ q_2 &\coloneqq \mathbb{E}_{\boldsymbol{Z}}[F^{-1}_{A_X|\boldsymbol{Z}}(1-1/N \mid \boldsymbol{Z})] \\ q_3 &\coloneqq \tilde{F}^{-1}_{A_X}(1-1/N) \\ q_4 &\coloneqq \tilde{F}^{-1}_{A_{X,N}}(\exp(-1)) \end{aligned} \tag{1}$$

where $F_{A_X|\boldsymbol{Z}}$ is the distribution function of A_X given $\boldsymbol{Z}$ and $F^{-1}_{A_X|\boldsymbol{Z}}$ the corresponding quantile function. $\tilde{F}_{A_X}$ and $\tilde{F}_{A_{X,N}}$ are predictive distribution functions given by

$$\begin{aligned} \tilde{F}_{A_X}(x) &\coloneqq \mathbb{E}_{\boldsymbol{Z}}[F_{A_X|\boldsymbol{Z}}(x|\boldsymbol{Z})] \\ \tilde{F}_{A_{X,N}}(x) &\coloneqq \mathbb{E}_{\boldsymbol{Z}}[F_{A_{X,N}|\boldsymbol{Z}}(x|\boldsymbol{Z})] \end{aligned} \tag{2}$$

with quantile functions $\tilde{F}^{-1}_{A_X}$ and $\tilde{F}^{-1}_{A_{X,N}}$. An outline of the calculation of $\tilde{F}_{A_X}$ is given in the appendix. Differences between estimators q_1, q_2, q_3 and q_4 are due entirely to the location of the expectation operator in the definitions (1) above; taking expectations over uncertain model parameters at different points in the return value calculation, results in different estimators for the return value in the presence of uncertainty. For example, for q_1 we take expectations over uncertain $\boldsymbol{Z}$ directly, using the mean parameters in the closed form expression for the distribution of A_X, and then find its $1-1/N$ quantile. For q_2, expectations are taken over return values, each of which is estimated using a given uncertain $\boldsymbol{Z}$. For q_3 (and q_4), expectations are taken over distributions of the annual (and N-year) maximum, each of which is estimated using a given uncertain $\boldsymbol{Z}$. In the absence of parameter uncertainty, the quantities q_1, q_2 and q_3 coincide; q_4 also converges to the others with increasing N.

In Section 5, we define estimators of associated values under uncertainty. As will be discussed there, when the uncertainty in model parameters $\boldsymbol{Z}$ is large (e.g. when model parameters are estimated from fitting using small samples), it is reasonable to expect that the performance of estimators for associated values might be poor, especially for those estimators based on expectations as opposed to more robust summary operators such as the median. Hence in the current work, for comparison, we also consider an estimator based on the median over uncertain $\boldsymbol{Z}$

$$q_5 \coloneqq \mathrm{med}_{\boldsymbol{Z}}[F^{-1}_{A_X|\boldsymbol{Z}}(1-1/N \mid \boldsymbol{Z})]. \tag{3}$$

Estimator q_5 is similar to q_2, except that we take a median as opposed to a mean over uncertain $\boldsymbol{Z}$. We expect that q_5 provides a more robust estimate for the return value under uncertainty.

4. The conditional extremes model

We seek to characterise the central features of the conditional distribution of an associated variable Y given extremes of variable X, using $\mathbb{E}_Y[Y|X=x]$ for $x > u > 0$, for some high threshold u. Typically the value of x would be set to a marginal return value of interest for X, estimated using one of the estimators in Section 3. Given knowledge of the joint distribution of (X,Y), this task is relatively straightforward. Otherwise, the joint distribution of (X,Y) for $X > u$ needs to be estimated from data in some way. In this work, we adopt the conditional extremes methodology of Heffernan and Tawn (2004) for variates (X_L, Y_L) on standard Laplace marginal scale as the main route to estimation of associated values.

4.1. Model form

For Laplace-scale variables $X_L \in \mathbb{R}$, $Y_L \in \mathbb{R}$, and high conditioning Laplace-scale threshold $u_L > 0$, the conditional extremes model then assumes

$$Y_L|(X_L = x_L) \;=\; \alpha x_L + x_L^{\beta}(\mu + \zeta W) \text{ for } x_L > u_L \tag{4}$$

with parameters $\alpha \in [-1,1]$, $\beta \in (-\infty, 1]$, $\mu \in \mathbb{R}$ and $\zeta > 0$. Standardised residuals W follow an unknown distribution, estimated using an empirical sample of standardised residuals from a model fit to data. Marginal transformation between the physical-scale variables (X,Y) and their Laplace-scale counterparts (X_L, Y_L) is achieved using functions

(g_{LX}, g_{LY}) such that $X_L = g_{LX}(X)$ and $Y_L = g_{LY}(Y)$. That is, $g_{LX}(\cdot) = F_L^{-1}F_X(\cdot)$ and $g_{LY}(\cdot) = F_L^{-1}F_Y(\cdot)$, where F_L is the cumulative distribution function of the standard Laplace distribution, given by

$$F_L(x) = \begin{cases} \exp(x)/2 & x \leq 0 \\ 1 - \exp(-x)/2 & x > 0. \end{cases}$$

F_X and F_Y are the marginal distributions of X and Y on physical measurement scale, the right-hand tails of which are assumed to take GP form. In practice, the pairs of functions (F_X, F_Y) and hence (g_{LX}, g_{LY}) are unknown an must also be estimated from data.

4.2. Sources of sampling uncertainty

Inferences under the conditional extremes model, including associated values, will therefore be subject to sampling uncertainty from multiple sources, including (a) the estimation of the marginal models for (F_X, F_Y), (b) estimation of marginal conditioning return values x of X (using estimated F_X and the rate of occurrence of events), and (c) estimation of the conditional extremes model. We are interested in estimators of associated values with reasonable bias and variance characteristics given these sources of uncertainty.

In practice, marginal transformation between the physical-scale variables (X, Y) and their Laplace-scale counterparts $(X_L|\boldsymbol{Z}, Y_L|\boldsymbol{Z})$ is achieved using functions $(g_{LX|\boldsymbol{Z}}, g_{LY|\boldsymbol{Z}})$ estimated from data; hence the dependence of both the functions and their Laplace-scale outputs on uncertain $\boldsymbol{Z}$. The marginal distributions $F_{X|\boldsymbol{Z}}$ and $F_{Y|\boldsymbol{Z}}$ are estimated empirically below some thresholds u_X, u_Y, and by fitting a GP distribution above the thresholds. We choose to repeat the conditional extremes model form as

$$Y_L|(X_L = x_L, \boldsymbol{Z}) \;=\; \alpha(\boldsymbol{Z})(x_L|\boldsymbol{Z}) + (x_L|\boldsymbol{Z})^{\beta(\boldsymbol{Z})}(\mu(\boldsymbol{Z}) + \zeta(\boldsymbol{Z})W) \text{ for } (x_L|\boldsymbol{Z}) > u_L \tag{5}$$

to emphasise the dependencies on $\boldsymbol{Z}$.

5. Estimators of associated values

The construction of the conditional extremes model above is similar to that of (bivariate) copula models (e.g. Joe 2014, and Section 6.1), in that the dependence between X and Y is stated in terms of transformed variables X_S and Y_S with a standard marginal distribution. The full joint distribution of X and Y on physical scale then requires specification of marginal distributions (for each of X and Y, allowing transformation to X_S and Y_S) as well as the dependence or copula distribution itself. The effects of estimation uncertainty in conditional extremes and copula models are also similar. Usually the marginal models are estimated first (with error), and then following (conditional) transformation to standard marginal scale (with error), the (conditional) dependence model is estimated (with error). For all these models therefore, it is appropriate to consider estimators of associated values that reflect this construction.

We now define estimators $q_{kk'}$ $k = 1, 2, ..., 6$, $k' = 1, 2$ providing a useful summary of the location of the tail in X, and the conditional "body" in Y, corresponding to a return period of N years. For example, for any of the five choices q_k, $k = 1, 2, ..., 5$ from Equations (1) and (3), we define a related estimator for the location of conditional values of Y to specify $q_{kk'}$

$$\begin{aligned} q_{k1} &\coloneqq \mathbb{E}_{\boldsymbol{Z}}\left[\mathbb{E}_{Y|\boldsymbol{Z}}[g^{-1}_{SY|\boldsymbol{Z}}((Y_S|X_S = g_{SX|\boldsymbol{Z}}(q_k|\boldsymbol{Z}), \boldsymbol{Z})|\boldsymbol{Z})]\right] \\ q_{k2} &\coloneqq \text{med}_{\boldsymbol{Z}}\left[\mathbb{E}_{Y|\boldsymbol{Z}}[g^{-1}_{SY|\boldsymbol{Z}}((Y_S|X_S = g_{SX|\boldsymbol{Z}}(q_k|\boldsymbol{Z}), \boldsymbol{Z})|\boldsymbol{Z})]\right]. \end{aligned} \tag{6}$$

The first of these estimators first identifies the conditional distribution of $(Y_S|X_S = g_{SX|\boldsymbol{Z}}(q_k|\boldsymbol{Z}), \boldsymbol{Z})$, where $g_{SX|\boldsymbol{Z}}(q_k|\boldsymbol{Z})$ is an estimate of the marginal return value on appropriate "standard scale" (e.g. standard Laplace for conditional extremes estimation), uncertain since it is defined using an estimated marginal model for X. (We adopt the notation X_S, Y_S here, rather than specific Laplace-scale X_L, Y_L to emphasise that the definition is applicable to other copula estimation schemes, as well as the conditional extremes model). This conditional distribution is then transformed to physical scale using the inverse of function $g_{SY|\boldsymbol{Z}}$, which is itself dependent on $\boldsymbol{Z}$ through uncertain marginal parameters for Y. Next we take expectation over Y, and finally over uncertain parameters $\boldsymbol{Z}$. The second estimator is similar to the first, with the final expectation over $\boldsymbol{Z}$ replaced by the median for robustness.

We also consider estimators q_{61} and q_{62} defined by

$$\begin{aligned} q_{61} &\coloneqq \mathbb{E}_{\boldsymbol{Z}}\left[\mathbb{E}_{Y|\boldsymbol{Z}}[g^{-1}_{SY|\boldsymbol{Z}}(Y_S|X_S = F^{-1}_{A_S}(1 - 1/N), \boldsymbol{Z})|\boldsymbol{Z})]\right] \\ q_{62} &\coloneqq \text{med}_{\boldsymbol{Z}}\left[\mathbb{E}_{Y|\boldsymbol{Z}}[g^{-1}_{SY|\boldsymbol{Z}}(Y_S|X_S = F^{-1}_{A_S}(1 - 1/N), \boldsymbol{Z})|\boldsymbol{Z})]\right]. \end{aligned} \tag{7}$$

For the estimators in Equation (7), the conditioning value of standard-scale X_S is set to $F_{A_S}^{-1}(1-1/N)$, where F_{A_S} is the cumulative distribution function of the annual maximum on standard scale. That is, the value of X_S is set to yield the desired marginal tail probability $F_{A_X|\boldsymbol{Z}}(1-1/N \mid \boldsymbol{Z})$ (see the appendix) such that $F_{A_X|\boldsymbol{Z}}(1-1/N \mid \boldsymbol{Z}) = F_{A_S}^{-1}(1-1/N)$ by definition for the conditioning variate on Laplace scale, but the conditioning value of X_S does not depend on the physical-scale marginal characteristics of X nor on $\boldsymbol{Z}$. For brevity, we therefore choose to describe estimators q_{61}, q_{62} as being *conditioned on probability*, and all other estimators as *conditioned on value*. However, the estimated conditional extremes model for $Y_S|X_S, \boldsymbol{Z}$ does depend on the marginal X fit, through the function $g_{SX|\boldsymbol{Z}}$ used for transformation from X to X_S. We note the close definitional correspondence between estimators $q_{2\cdot}$, $q_{5\cdot}$ and $q_{6\cdot}$, all of which exploit the conditioning return value $F_{A_X|\boldsymbol{Z}}(1-1/N \mid \boldsymbol{Z})$ for given $\boldsymbol{Z}$.

Perhaps Equations 6 and 7 appear unwieldy in terms of the number of times that $\boldsymbol{Z}$ appears. Nevertheless, these serve to illustrate that dependence on uncertain $\boldsymbol{Z}$ enters the estimation of associated values at multiple points, and that the ordering of operations in the expressions should be expected to impact the bias and variance characteristics of the resulting estimators.

6. Numerical study of estimator performance

In Jonathan et al. (2021), a combination of theoretical arguments and numerical simulation were used to quantify the relative bias characteristics of return value estimators q_k, $k = 1, 2, 3, 4$. Unfortunately, there is no easy theoretical route to explore the relative characteristics of estimators for associated values, and we must rely on simulation. Therefore, in this section we describe bivariate models used for sample simulation (Section 6.1) and estimation (Section 6.2). We also provide a summary of the characteristics of marginal return value estimator q_5 relative to its competitors q_k, $k = 1, 2, ..., 4$, because this comparison has not been reported in the past. Discussion of the performance of combinations of estimators and estimation schemes for associated values then follows in Section 7.

6.1. Study set-up

The simulation study used to assess the relative performance of the estimators for return values and associated values introduced in Sections 3 and 5 is as follows. We simulate samples of size $n = 200$, 1,000 and 10,000 from bivariate distributions constructed with GP margins and either Gaussian or logistic dependence D, representing asymptotic independence and asymptotic dependence respectively. That is, we simulate pairs of values of X, Y with distribution function

$$\mathbb{P}(X \leq x, Y \leq y; \xi_X, \xi_Y, \kappa) = F_{XY}(x, y; \xi_X, \xi_Y, \kappa) = F_X(x; \xi_X) F_Y(y; \xi_Y) C(F_X(x; \xi_X), F_Y(y; \xi_Y); \kappa) \tag{8}$$

where cumulative distribution functions $F_X(\cdot; \xi_X)$ and $F_Y(\cdot; \xi_Y)$ represent marginal GP distributions for X and Y with shape parameters ξ_X and ξ_Y, common scale parameter $\varsigma = 1$ and threshold $\eta = 0$, with functional form

$$F(x; \xi, \varsigma, \eta) = 1 - \left(1 + \frac{\xi}{\varsigma}(x - \eta)\right)_+^{-1/\xi}. \tag{9}$$

The study design considers all combinations of cases $\xi_X, \xi_Y \in \{-0.4, -0.2, 0, 0.2\}$ representing a range of tail heaviness for X and Y common in environmental applications. $C(\cdot, \cdot; \kappa)$ is a bivariate copula parameterised in terms of a scalar $\kappa \in [0, 1]$. For the logistic copula ("Lgs"), $1 - \kappa$ represents the logistic parameter (usually denoted α) $\in [0, 1]$, and for the Gaussian copula ("Gss"), κ represents the correlation $\in [0.1]$, such that

$$C(u_X, u_Y|\kappa) = \begin{cases} C(u_X, u_Y|\kappa) = \exp\left(-\left((-\log u_X)^{-1/(1-\kappa)} + (-\log u_Y)^{-1/(1-\kappa)}\right)^{1-\kappa}\right) & \text{Logistic} \\ C(u_X, u_Y|\kappa) = \Phi(\Phi^{-1}(u_X), \Phi^{-1}(u_Y); \boldsymbol{\Sigma}) & \text{Gaussian} \end{cases} \tag{10}$$

for $\kappa \in (0, 1)$, with $\kappa = 0$ and $= 1$ corresponding to perfect independence and perfect dependence respectively. We choose to express the logistic dependence in terms of $1 - \kappa$ so that the strength of dependence increases with κ for both Gaussian and logistic copulas. $\Phi(\cdot)$ is the cumulative distribution function of the standard Gaussian distribution, and $\Phi(\cdot, \cdot; \boldsymbol{\Sigma})$ is the cumulative distribution function of a bivariate Gaussian distribution with zero mean and covariance $\boldsymbol{\Sigma}$, a 2×2 matrix with unit diagonal elements and κ on the off-diagonal. For the study, we consider the cases $\kappa \in \{0.1, 0.5, 0.9\}$ representing low, medium and high strength of dependence respectively. Figure 1 gives an impression of the effect of ξ_X, ξ_Y, D and κ on sample characteristics.

A total of $m = 1000$ sample realisations was generated for each combination of ξ_X, ξ_Y, dependence copula type D ("Gss" or "Lgs") and strength κ, for subsequent estimation of sets $(q_k, q_{kk'})$ of return and associate values. This corresponds to a total of 96 ($4 \times 4 \times 2 \times 3$) combinations (or "cases") of design factors for sample generation.

6.2. Estimation of return and associated values

For each of m sample realisations, we estimate pairs of return value and associated value $(q_k, q_{kk'})$, $k = 1, 2, ..., 6$, $k' = 1, 2$ as defined in Section 5 using each of three models types for the joint distribution of (X, Y). Of primary interest in the conditional extremes model (Section 4.1); for comparison, we also include a Gaussian or logistic copula model (i.e. the dependence model under which a sample was generated), and a naive model based on simple linear regression motivated by historical based practice (e.g. ISO19901-1 2015). These inference schemes are described in more detail below.

Once these models have been estimated, we calculate summary statistics of interest, including associated values using Equations 6, by simulation, numerical integration or equivalent under the fitted model. Estimation uncertainty is quantified using bootstrapping: that is, empirical estimates for $\mathbb{E}_{\boldsymbol{Z}}$ and $\text{med}_{\boldsymbol{Z}}$ (in Equations 6 and 7) are calculated as sample means and medians over $n_B = 100$ bootstrap resamples for each sample realisation.

6.2.1. Estimation schemes

Estimation of the three models is now described.

The conditional extremes scheme ("CntExt" for brevity) requires (a) estimation of marginal extreme value models for X and Y, (b) marginal transformation to standard Laplace scale, and (c) estimation of the conditional extremes model. For marginal modelling (a), we assume we know that the full sample follows a GP distribution, so that an extreme value threshold $= 0$ is always appropriate; however, in general, we assume that marginal parameters must be estimated. Since the conditional extremes model (c) is motivated asymptotically, we estimate it for a sequence of Laplace scale thresholds with non-exceedance probabilities $\tau \in \{0.5, 0.8, 0.9\}$. We expect to see a bias-variance trade-off as a result: as the value of τ increases, estimated parameter bias might be expected to reduce because the asymptotic model form is more appropriate, but estimated parameter variance increases due to reduced sample size.

The copula scheme ("CplEst" for brevity) involves (a) marginal extreme value estimation, followed by (b) marginal transformation to the uniform scale (see Equation 10), and (c) estimation of copula parameter κ. In this work, for each sample realisation, we assume that the copula type D is known, and hence we do not consider copula misspecification. Further, we assume we know that the copula model can be applied to the full sample, which is appropriate since the whole sample was generated under the copula.

The "naive" simple linear estimation scheme ("RgrEst" for brevity) estimates the dependence between X and Y by fitting a linear regression (with intercept and slope terms) to all data exceeding a marginal threshold in X with given non-exceedance probability τ. The marginal X return value is estimated by fitting a GP model to the full X sample, as for other schemes. This estimation scheme is included as a pragmatic benchmark approach for comparison only.

As a result, for each sample realisation, there are 8 $(3 \times 3 - 1)$ combinations of model type and dependence threshold to be considered. In total therefore, the design consists of $n_C = 768$ $(= 96 \times 8)$ different inferences cases, and $n_B = 100$ bootstrap resamples per case to evaluate return values and associated values for a given sample realisation. We consider $m = 1000$ sample realisations per case to estimate the sampling distributions of return values and associated values.

6.2.2. Estimation of fractional bias

The performance of the estimators $(q_k, q_{kk'})$, $k = 1, 2, ..., 6$, $k' = 1, 2$ is quantified by comparison with the known underlying values of marginal return values and associated values, in terms of fractional bias. For quantity θ with estimated value $\widehat{\theta}_{ij}$ for sample realisation i, $i = 1, 2, ..., m$ and case j, $j = 1, 2, ..., n_C$ from the experimental design, fractional bias is given by $(\widehat{\theta}_{ij}/\theta) - 1$. We then use the sample $\{(\widehat{\theta}_{ij}/\theta) - 1\}_{i=1}^{m}$ to quantify the performance of the return value or associated value estimators for case j.

Given either known underlying parameter values, or estimates for these, return values can be estimated in closed form. Associated values for Gaussian dependence are estimated either by simulation or numerical integration; numerical integration is used in the case of logistic dependence. Carefully constructed algorithms are necessary to ensure that associated values are estimated reliably even for values of dependence κ near the limits of its $[0, 1]$ domain. Similarly, transformation between different marginal scales needs to be performed reliably regardless of the estimated values of GP shape and scale parameters ξ, ς. MATLAB code for the simulation is available at Jonathan (2022).

6.3. Return values

The main focus of the current work is the estimation of associated values, but it is interesting also to consider the performance of the new estimator q_5 for return value, relative to competitors q_k, $k = 1, 2, ..., 4$ examined previously in Jonathan et al. (2021). Figure 2 shows the fractional bias of all return value estimators as a function of GP shape parameter ξ_X for different sample sizes. A similar plot is shown in Jonathan et al. (2021) (Figure 3), except that the new estimator q_5 has been added. Interestingly, the new estimator q_5, the bootstrap median of the return

value estimated per bootstrap resample, performs similarly to estimator q_1, namely the return value calculated using bootstrap mean parameter estimates. We note further that estimator q_3, namely the $1-1/N$ quantile of the predictive distribution of the annual maximum, has a relatively large positive bias: it overestimates the return value, especially for $\xi_X > -0.1$. As sample size increases, fractional bias reduces in magnitude for all estimators, highlighting that differences in estimators are of most concern for small samples.

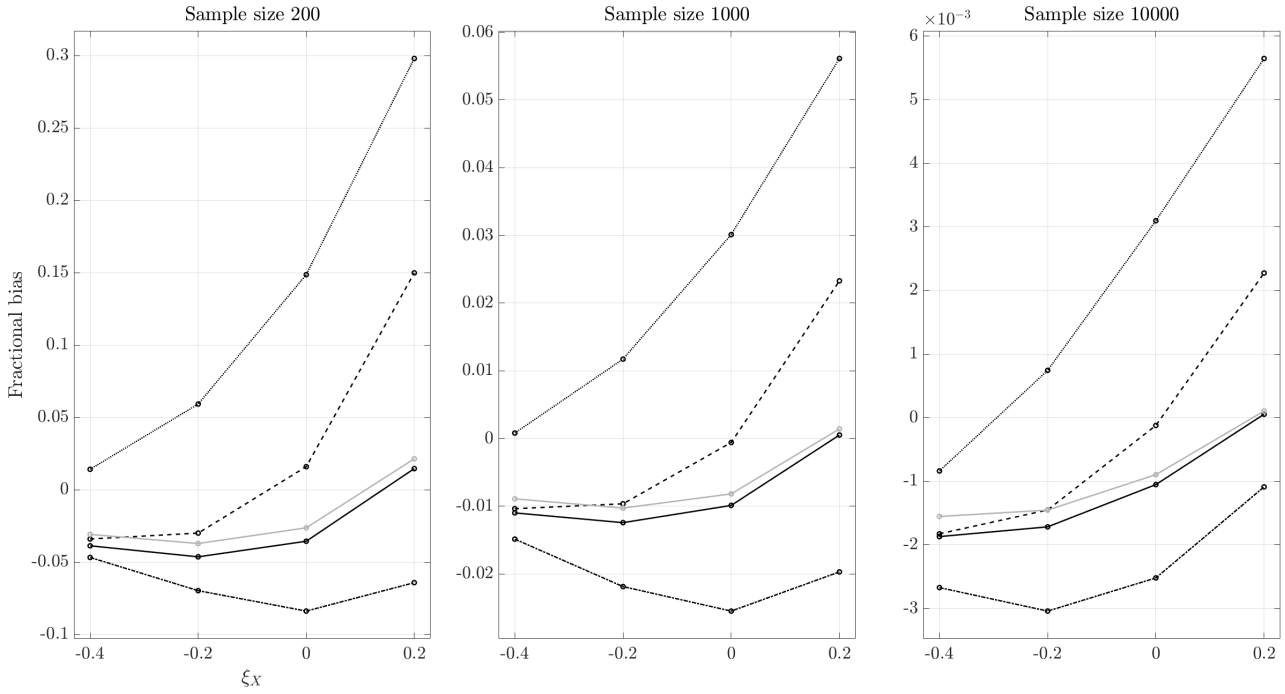

Figure 2: Fractional bias of return value estimators q_k, $k = 1, 2, ..., 5$ as a function of GP shape parameter ξ_X for sample sizes $n = 200, 1,000$ and $10,000$. Estimators are defined in Section 3, and have line styles black solid (q_1), black dashed (q_2), black dotted (q_3), black dash-dotted (q_4) and grey solid (q_5).

7. Performance of models and estimators for associated values over the design

In this section, we perform an analysis of the fractional bias results for associated values from the computer experiment. Our main objective is to determine whether any of the estimators for associated values is to be preferred in general for metocean design. Estimators of associated value are more complex to calculate that those of return values, and in particular require the estimation of a larger number of parameters. We would expect therefore that, for a given sample size, the uncertainty of associated values might be greater than that of return values. In this section, we start (in Section 7.1) by considering the feasibility of calculation of the estimators $q_{kk'}$, $k = 1, 2, ..., 6$, $k' = 1, 2$ over the design. We find that for estimators *conditioned on value* (namely $q_{kk'}$, $k = 1, 2, ..., 5$, $k' = 1, 2$) it is not always possible to perform the associated value calculation; this issue is particularly problematic for small sample size n and corresponding estimators q_k for return values which are biased high. Indeed, for $n < 10,000$, we show and explain that only estimators $q_{6\cdot}$ *conditioned on probability* can be guaranteed to provide feasible computations under bootstrap resampling. For estimators conditioned on value, we are forced to reject a proportion of bootstrap resamples, since these do not yield feasible calculations of associated values.

In Section 7.2, we illustrate the distributions of fractional bias for different design cases over sample realisations to give an initial impression of fractional bias data. Given that there are $n_C = 768$ different design cases to be considered, visual assessment alone of fractional bias is not realistic. We also estimate the Spearman rank correlation between estimates from all pairs of estimators $q_{kk'}$ for different sample sizes n. We also report plots of the median and inter-quartile range (IQR) for all estimators over the full design. Further, in Section 7.3 and Jonathan (2022), we seek to explore the characteristics of fractional bias for estimators as a function of design variables, and further quantify this using a non-homogeneous Gaussian regression (e.g. Williams et al. 2014). This regression model helps us understand how underlying sample characteristics affect the fractional bias of associated values.

7.1. Infeasible calculations

Inspection of the expressions (Equation 6) for associated values $q_{kk'}$, $k = 1, 2, ..., 5$, $k' = 1, 2$ conditioned on value indicates that conditioning is made on Laplace scale per bootstrap resample, using an estimate q_k of the N-year return value obtained independently beforehand on physical scale. Therefore, the actual Laplace-scale conditioning value $g_{LX|\boldsymbol{Z}}(q_k|\boldsymbol{Z})$ is uncertain, due to the transformation function $g_{LX|\boldsymbol{Z}}$ estimated (with error) from the sample. When the estimated GP shape parameter ξ_X is negative, the marginal X distribution has a finite upper end point. When the marginal fit (and hence the estimate of $g_{LX|\boldsymbol{Z}}$) is particularly poor (e.g. when the sample size n is small), then it is possible that the value of q_k (estimated beforehand) lies beyond the upper end point of the fitted marginal distribution, so that $g_{LX|\boldsymbol{Z}}(q_k|\boldsymbol{Z})$ is undefined; this outcome is also possible for a particularly poor estimate of q_k. Figure 3 explores this effect for all estimators, for different "corner" choices of pairs ξ_X, ξ_Y from the design, and for different sample sizes n, over all combinations of dependence design factors (i.e. dependence type D, strength κ and threshold non-exceedance probability τ), for conditional extremes estimation. The figure summarises the distribution of the proportion of rejected bootstrap resamples in the form of box-whisker plots. For example, estimators $q_{3\cdot}$ incur a rejection rate of around 70% on average for sample size $n = 200$ with $\xi_X = -0.4$ and $\xi_Y = -0.4$: that is, 70% of bootstrap resamples yield infeasible calculations of associated value. As might be expected, as sample size increases (green to orange to grey), the proportion of rejected bootstrap resamples reduces to zero; there are no rejections for any estimators with $n = 10,000$. Further the rejection rate is highest per estimator for strongly negative ξ_X, as a direct consequence of the marginal return value estimator q_3 (the quantile of the predictive distribution of the annual maximum with non-exceedance probability $1 - 1/N$) exhibiting more positive bias than other return value estimators. Plots for fitting under copula models exhibit analogous features.

Estimators $q_{6\cdot}$ (conditioned on probability, Equation 7) never incur rejections by construction. For these estimators, for any bootstrap resample, we condition on the N-year return value estimated using exactly the same resample for marginal fitting, with a fixed probability $1 - 1/N$ for annual exceedance. This value can be estimated without error provided that the rate of occurrence of events is assumed known; even if the rate of occurrence was only known approximately, this uncertainty could never lead to a Laplace-scale conditioning value exceeding the upper end point of the distribution, by definition. We conclude that estimators conditioned on probability are inherently preferable under a bootstrap procedure for uncertainty quantification. The source of the infeasibility issue is the fact that an

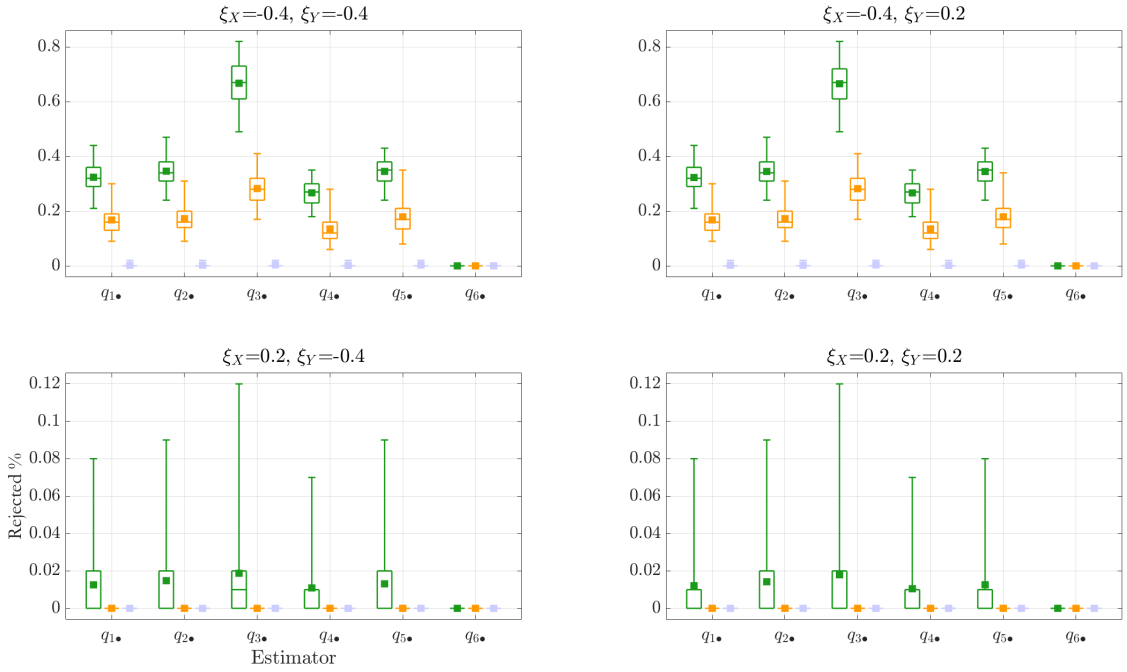

Figure 3: Box-whisker plots showing the proportion of bootstrap estimates rejected due to the conditioning value exceeding the upper end-point of the estimated marginal GP distribution for the tail of X, for the conditional extremes estimation scheme, encompassing all choices of dependence type D, strength κ and threshold with non-exceedance probability τ. Panels correspond to different "corner" combinations of GP shapes ξ_X and ξ_Y from the design, and colours indicate sample size $n = 200$ (green), 1,000 (orange) and 10,000 (grey). The abscissa variable indicates the estimator for associated values. For the final abscissa value $q_{6\cdot}$, conditioning is performed per bootstrap resample based on the appropriate Laplace-scale probability.

external estimate of the conditioning return value is used, rather than an estimate using the bootstrap resampling scheme adopted for uncertainty quantification. In passing we note that same difficulty would arise under a Bayesian inference, unless suitable prior distributional assumptions were imposed on marginal X tail models to prevent $g_{LX|\boldsymbol{Z}}$ being undefined at the conditioning value q_k. Of course, modification of estimation schemes and estimators is possible to mitigate infeasible calculations; but in this situation, the performance characteristics of estimators of associated values would then be dependent on the details of the mitigation used. In the following sections, we adopt a naive mitigation strategy: estimates for $q_{kk'}$, $k = 1, 2, ..., 5$, $k' = 1, 2$ conditioned on value are based on the subset of bootstrap resamples corresponding to feasible calculations.

7.2. Visual summaries of estimator performance

Figure 4 illustrates the empirical density of fractional bias for estimator q_{62} under the conditional extremes scheme, over all choices of dependence type D, strength κ and threshold (with non-exceedance probability τ), for the same "corner" combinations of ξ_X and ξ_Y as in Figure 3. In each panel, colour again indicates sample size. We observe that distributional width reduces with increasing sample size as would be expected. Further, the general size of bias increases dramatically with increasing ξ_Y. For $\xi_Y = 0.2$ and $n = 200$, the empirical density shows a long right-hand tail (positive skewness), but skewness reduces with increasing n. The mean fractional bias is also apparently influenced by sample size, as can be more easily seen for the approximately symmetric densities for $\xi_Y = -0.4$. For sample size $n = 200$, we should not be surprised to encounter fractional biases of -100% to $+500\%$ occasionally, depending on sample characteristics. For $n = 10,000$, fractional biases of $\pm 20\%$ are still possible for large ξ_Y. The characteristics of plots analogous to Figure 4 for other estimators are similar, notwithstanding rejection of infeasible bootstraps where necessary.

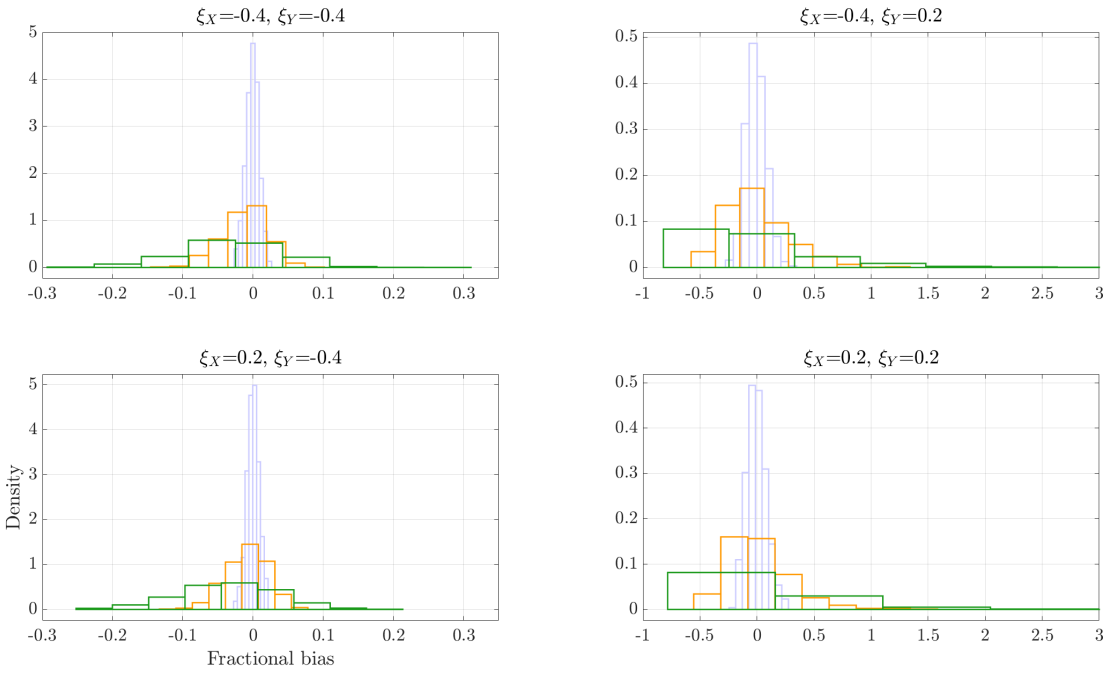

Figure 4: Empirical densities of fractional bias for q_{62} estimated using the conditional extremes model. Sample realisations exhibit Gaussian dependence and $\kappa = 0.9$. Panels correspond to different "corner" combinations of marginal GP shape parameters ξ_X and ξ_Y. Empirical densities in each panel correspond to $n = 200$ (green), $n = 1,000$ (orange) and $n = 10,000$ (grey). For panels corresponding to $\xi_Y = 0.2$, the abscissa is truncated at 3.0 for clarity. The maximum values observed for $n = 200$ are 4.36 (for $\xi_X = -0.4$) and 7.70 (for $\xi_X = 0.2$).

Figure 5 provides further evidence for the likely spread of fractional bias for different sample characteristics. Each panel shows the estimated inter-quartile range (IQR) of the distribution of fractional bias across all estimators $q_{kk'}$, $k = 1, 2, ..., 6$, $k' = 1, 2$, for a given combination of sample size n and dependence threshold with non-exceedance probability τ. Colours represent different estimation schemes: solid green lines correspond to conditional extremes estimation, including marginal estimation; dashed green lines correspond to conditional extremes estimation with the marginal distributions assumed known. Orange lines represent copula estimation, assuming that the correct dependence type D is known, with estimated margins (solid) and assumed known margins (dashed). Comparison of

solid and dashed curves therefore quantifies the extent to which uncertainty in associated values can be attributed to marginal uncertainty. The grey line corresponds to simple linear regression estimation. The value of IQR does not

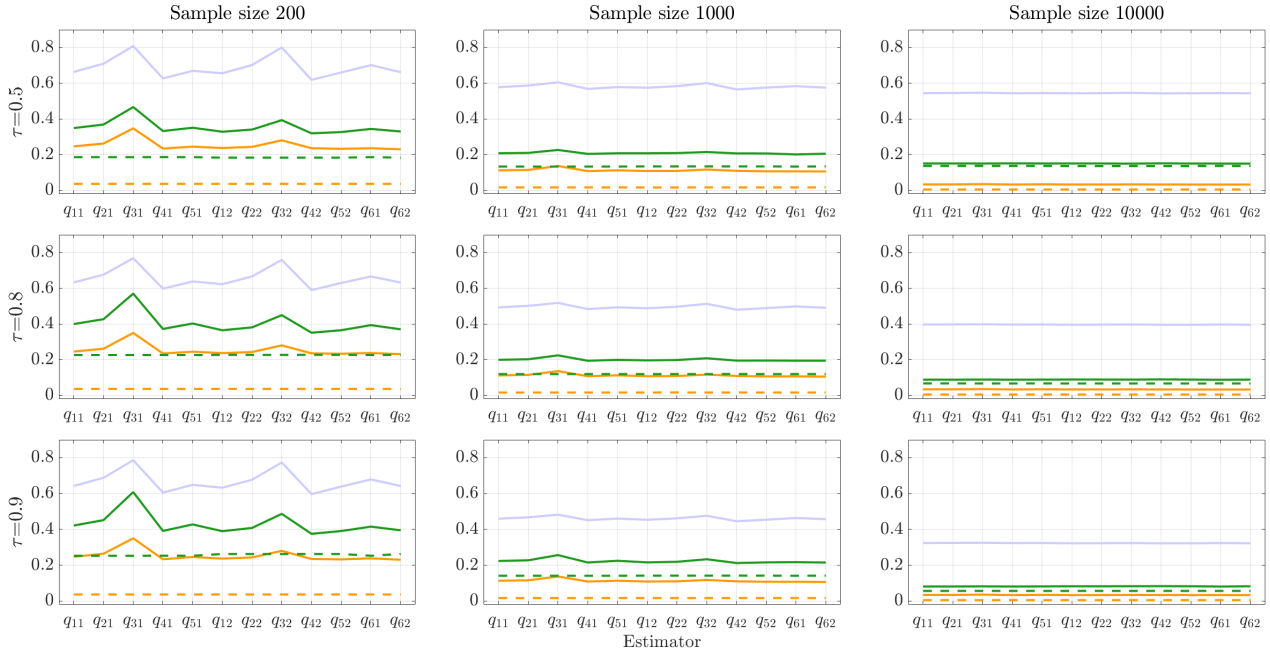

Figure 5: Inter-quartile range of fractional bias for all estimators, over all design combinations, for different sample sizes (n, columns) and dependence modelling thresholds (with non-exceedance probabilities τ, rows): simple regression estimation (RgrEst, grey) and conditional extremes estimation (CndExt, green) with unknown (solid) and known margins (dashed). Also shown as a benchmark, are corresponding estimates under the correct dependence copula D form (but unknown parameter κ, orange) with unknown (solid) and known margins (dashed), using the full sample for κ estimation. Note that for estimators $q_{kk'}$, $k = 1, 2, ..., 5$, $k' = 1, 2$, only feasible bootstrap resample calculations are used.

vary substantially across estimators, which is surprising because of the large bootstrap rejection rate for estimators conditioned on value; but this finding in turn may suggest that the naive mitigation strategy adopted (i.e. rejection of infeasible calculations) has its merits. There is evidence of increased IQR for q_{31} and q_{32}, reflecting the relative positive bias of the corresponding marginal estimator q_3. Knowledge of marginal characteristics reduces IQR, dramatically so for $n = 200$; in contrast for $n = 10,000$, the sample is large enough to provide good marginal estimation, so that there is little benefit of knowing marginal characteristics. Knowledge of the true copula type reduces IQR always; conditional extremes estimation provides higher IQR given τ and n, because it requires estimation of a larger parameter set (encoding multiple forms and extents of extremal dependence). Further, IQR reduces with increasing sample size in all cases. Figure 6 is similar to Figure 5, but provides median fractional bias. The simple linear regression estimation scheme yields negative bias over the design; otherwise all estimators have relatively low median fraction bias. Again, q_{31} and q_{32} perform differently to other estimators because of the marginal bias of q_3. Median fractional bias reduces in size with increasing n and τ.

Note that estimates of fractional bias mean and standard deviation (given in Figures 12 and 13 of the Appendix), sensitive to a small number of *exceptional values* of fractional bias, are considerably less well behaved that their more robust analogues visualised in Figures 5 and Figure 6. Here, we define an exceptional value as an estimate of fractional bias which exceeds 10. Exceptional values occur because of relatively poor estimation of marginal and dependence models, unavoidable for small sample sizes (unless constraints on values of parameter estimates are imposed, or different estimation schemes more suitable for small samples are adopted). Over the complete design, 0.16% of the total number of sample realisations resulted in estimates of fractional bias exceeding 10 for $n = 200$, reducing to 3.4×10^{-3}% for $n = 1,000$. No exceptional values were encountered at all for $n = 10,000$. The presence of exceptional values of fractional bias is to be expected, especially when the marginal Y tail is estimated to be long, and estimated $\xi_Y \geq 0$; this in turn is to be expected when the sample size is small. The proportions of exceptional values by design variables is illustrated in Figure 7, where the green (orange) histograms shows proportions of exceptional values across all sample sizes n (and $n = 1,000$, 10,000 only) by design variable. Over all sample sizes (green), exceptional values predominantly correspond to $n = 200$ and $\xi_Y = 0.2$. For $n > 200$, exceptional values occur exclusively at $n = 1,000$ and ξ_Y=0.2.

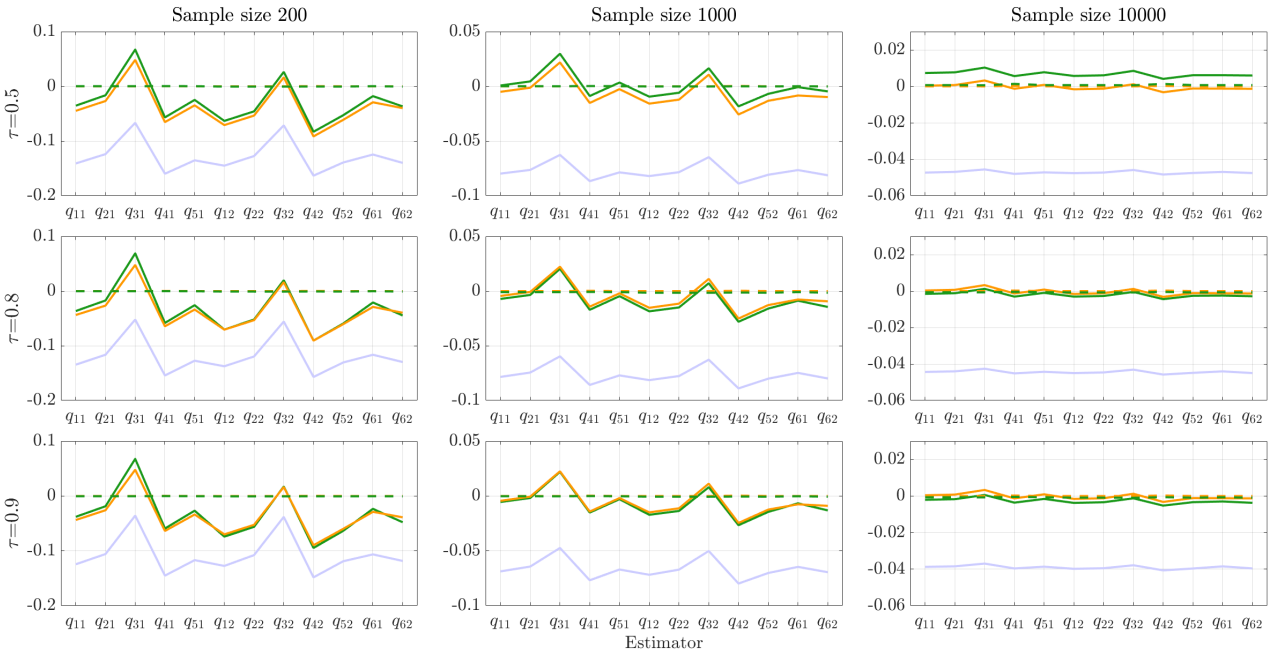

Figure 6: Median fractional bias for all estimators, over all design combinations, for different sample sizes (n, columns) and dependence modelling thresholds (ψ, with non-exceedance probability τ, rows). For details, see caption of Figure 5. Note that for estimators $q_{kk'}$, $k = 1, 2, ..., 5$, $k' = 1, 2$, only feasible bootstrap resample calculations are used.

It is also noteworthy that occurrences of exceptional values for estimators q_{61}, q_{62} (conditioned on probability) and q_{k2}, $k = 1, 2, ..., 5$ (conditioned on value, but using the median rather than mean to summarise over estimates for different bootstrap resamples, see Equation 6) are rare, even for $n = 200$. This suggests that these estimators are more appropriate for general use, especially for small sample sizes.

Figure 8 shows Spearman rank correlations for all pairs of estimates as a function of sample size, over all other design variables. Again, for estimators $q_{kk'}$, $k = 1, 2, ..., 5$, $k' = 1, 2$, only those bootstrap resamples admitting feasible calculation of associated values are considered. Estimated values for all estimators are highly correlated for all sample sizes, and the general features of the correlation maps reflect the different constructions of return and associated values. For $n = 10,000$ all rank correlations are close to unity. For $n = 200$, we observe that estimators q_{k1}, $k = 1, 2, ..., 5$ are generally more highly inter-correlated, as are q_{k2}, $k = 1, 2, ..., 5$. However, estimators q_{31} and to a lesser extent q_{32} are exceptions to this trend, due to the fact that return value estimator q_3 tends to be positively biased. Estimators q_{61} and q_{62} are in general more correlated with q_{k2}, $k = 1, 2, ..., 5$ than q_{k1}, $k = 1, 2, ..., 5$.

7.3. Quantifying sources of fractional bias

Results in Section 7.2 suggest that the fractional bias of estimators of associated values varies systematically over the design. In this sub-section, we build regression models to explore this effect further, with the intention of understanding whether fractional bias from some estimators and estimation schemes is more sensitive to underlying sample characteristics: we would prefer that the performance of estimates of associated values was consistently good across the design. As a result, associated values could be estimated from samples with unknown characteristics with a level of confidence.

Figure 9 shows fractional bias "main effects" for estimator q_{62} on all design variables, as box-whisker plots, for sample size $n = 200$. The first (second) row of panels illustrates results for conditional extremes estimation, given a true Gaussian (logistic) dependence. The third and fourth rows follow a similar pattern, for the simple linear regression estimation scheme. Figure 10 shows the analogous plots for $n = 10,000$. The range of fractional bias is very large for $n = 200$, but reduces considerably for $n = 10,000$ under the conditional extremes estimation scheme only. For $n = 200$ and conditional extremes estimation, the distribution of fractional bias for different values of ξ_X appears relatively stable, with a wider spread and positive skewness under Gaussian dependence. The spread of fractional bias increases with increasing ξ_Y; mean fractional bias increases with ξ_Y under Gaussian dependence, but reduces with ξ_Y under logistic dependence. The mean and median fractional bias both tend to zero with increasing strength κ of dependence

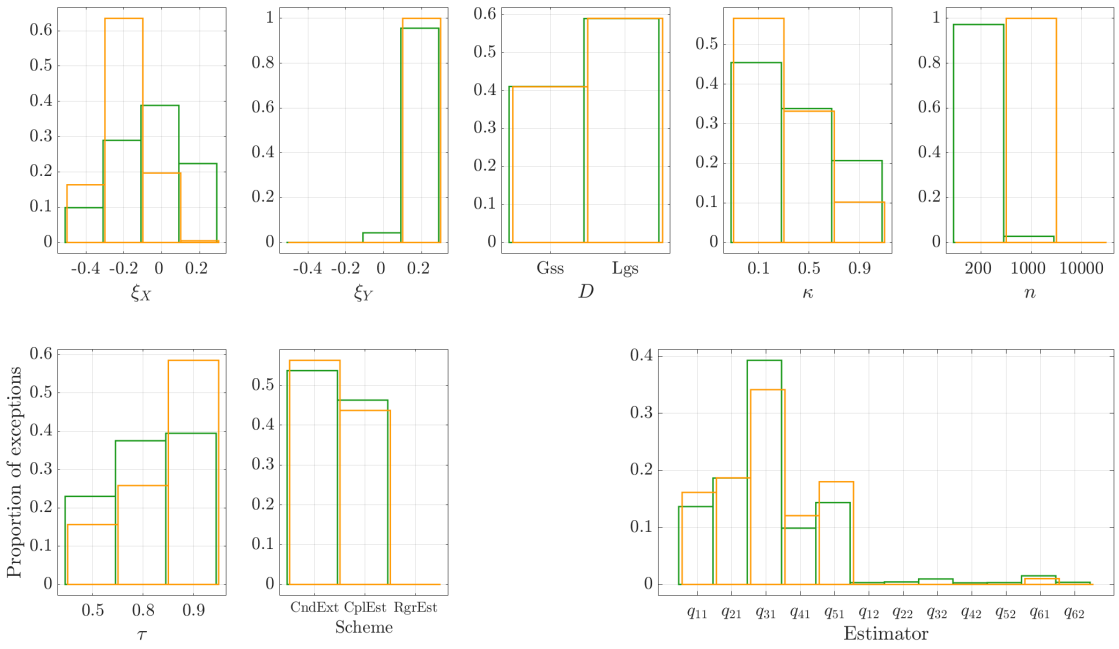

Figure 7: Occurrences of exceptional associate values with magnitudes > 10. Proportions (green) with respect to the design variables ξ_X, ξ_Y, dependence type D, strength κ, threshold ψ, sample size n and estimation scheme, and associated value estimator. Proportions in orange correspond to sample sizes $n = 1000$ and 10,000 only. Note that for estimators $q_{kk'}$, $k = 1, 2, ..., 5$, $k' = 1, 2$, only feasible bootstrap resample calculations are used.

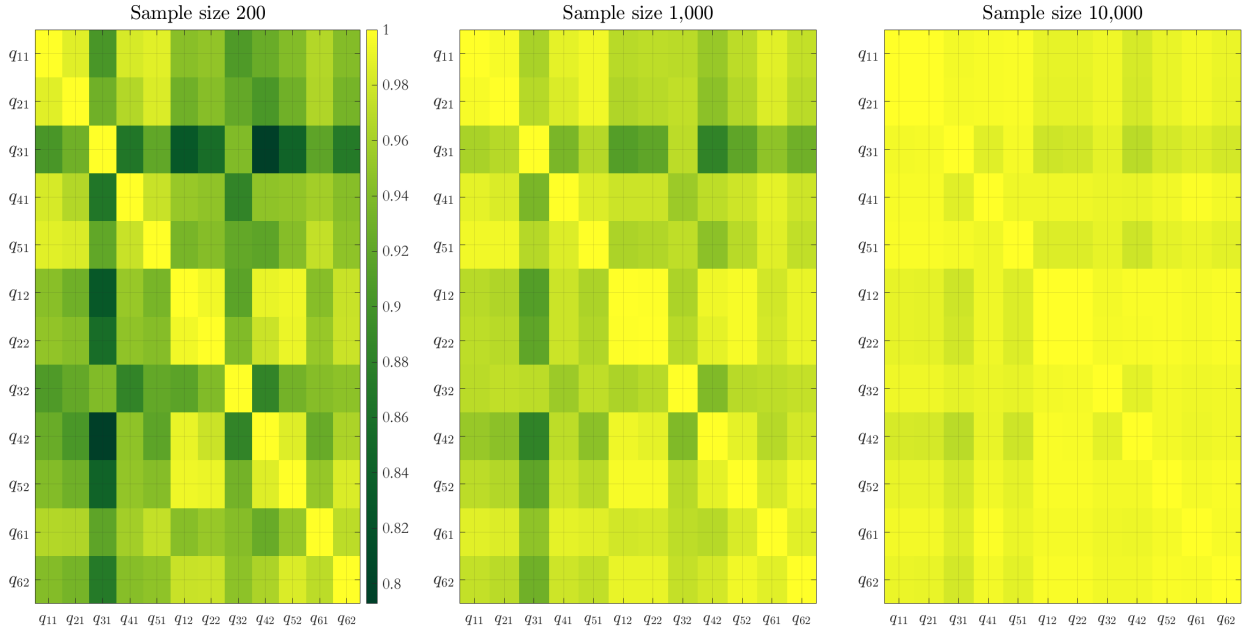

Figure 8: Spearman rank correlation coefficients between all pairs of estimators, for different sample sizes. Colour legend is given on the right-hand side of the left panel. Note that for estimators $q_{kk'}$, $k = 1, 2, ..., 5$, $k' = 1, 2$, only feasible bootstrap resample calculations are used.

under conditional extremes estimation, and there is evidence under Gaussian dependence that the spread of fractional bias also reduces with increasing κ. Further for $n = 200$ and conditional extremes estimation, there is little evidence

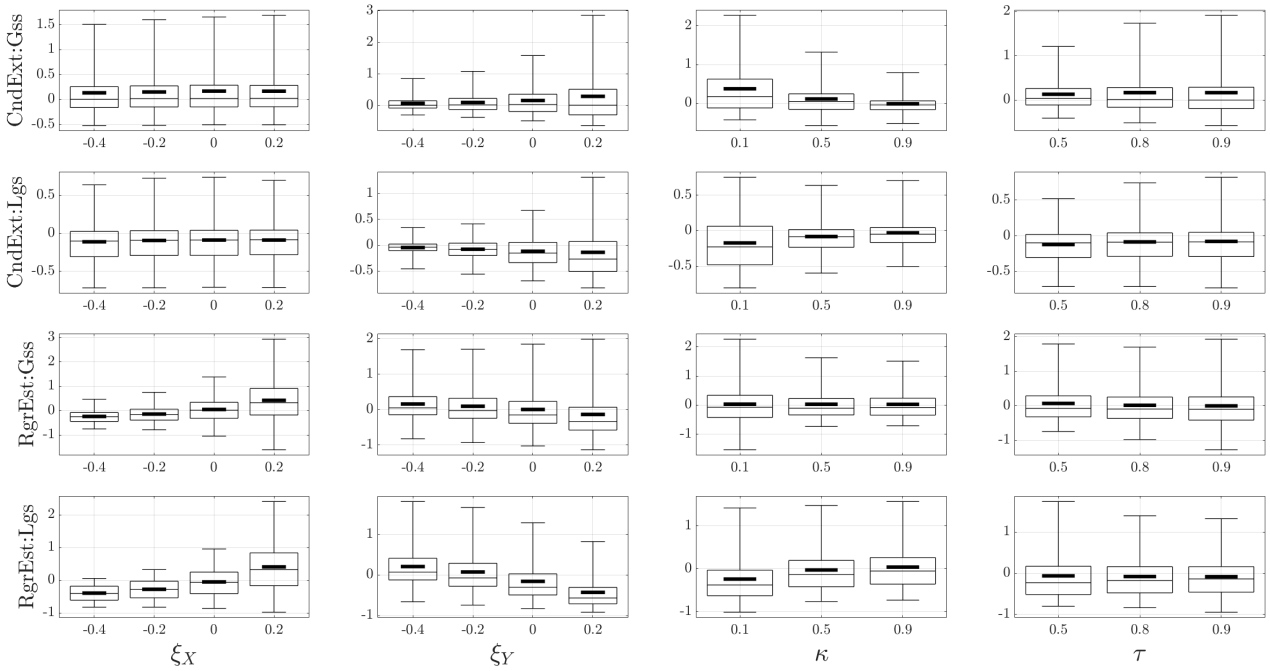

Figure 9: Box-whisker plots of fractional bias in associate value estimator q_{62} as a function of sample covariates for conditional extremes (CndExt) and simple linear regression (RgrEst) estimation schemes, and Gaussian (Gss) and logistic (Lgs) dependence. Sample size $n = 200$. In the box-whisker structure, thin horizontal lines in the box represent sample 25%, 50% and 75% quantiles, and whiskers 2.5% and 97.5% quantiles. The thick horizontal line corresponds to the sample mean.

that the mean and median fractional bias trend with threshold non-exceedance probability τ. In contrast, the simple regression estimation scheme shows sensitivity to both ξ_X and ξ_Y. At $n = 10,000$, Figure 10 indicates that many of the trends already mentioned persist. For simple regression estimation, sensitivity of fractional bias to both ξ_X and ξ_Y is clear, as is the considerably larger general spread of fractional bias. Additionally there is evidence that the mean and median fraction bias under conditional extremes estimation reduce towards zero with increasing τ.

Plots analogous to Figures 9 and 10 for all combinations of estimators and sample sizes are given at Jonathan (2022). Again, we emphasise that for associated value estimators ($q_{kk'}$, $k = 1, 2, ..., 5$, $k' = 1, 2$) conditioned on value, only estimates from feasible calculations can possibly be considered. Trends present in the plots are consistent with the discussion here and in Sections 7.2.

The trends with design variables in Figures 9 and 10 and at Jonathan (2022) suggest we consider estimating a statistical model for the distribution of fractional bias as a function of design variables. The figures indicate that the spread of fractional bias also varies systematically over the design. For this reason, we adopt a non-homogeneous Gaussian regression form

$$(B|\boldsymbol{C} = \boldsymbol{c}) = \boldsymbol{\varphi}'_\mu \boldsymbol{c} + \boldsymbol{\varphi}'_\sigma \boldsymbol{c} \times N(0, 1) \tag{11}$$

for response B and a vector $\boldsymbol{c}$ of covariates, where $N(0, 1)$ indicates a standard Gaussian-distributed random variable. Vectors of mean and error parameters $\boldsymbol{\varphi}_\mu$ and $\boldsymbol{\varphi}_\sigma$ are to be estimated jointly. In our case, B is the fractional bias for each of the combinations of estimators and estimation schemes under consideration, in turn, and $\boldsymbol{C}$ is a vector containing the full "response surface" in design variables (that is, the intercept, linear terms and all squared and cross-terms in design variables) as listed in the x-axis label of Figure 11. The regression was estimated using non-linear optimisation.

We emphasise that the purpose of the regression is not to produce a perfect predictive model for fractional bias to be used in practical application, since in practice we will rarely if ever know the values of design variables for an application; rather, we estimate the regression model to identify whether combinations of associated value estimators and estimation schemes exist which are relatively insensitive to the design, and hence useful across a range of applications. Further, since the sample size for the regression ($> 350,000$ observations per dependence type) is huge, there is little purpose to considering parameter estimate "significance" in the usual sense. Figure 11 gives parameter estimates for vectors $\boldsymbol{\varphi}_\mu$ (odd rows) and $\boldsymbol{\varphi}_\sigma$ (even rows) by dependence type (columns) for conditional extremes estimation (first two rows)

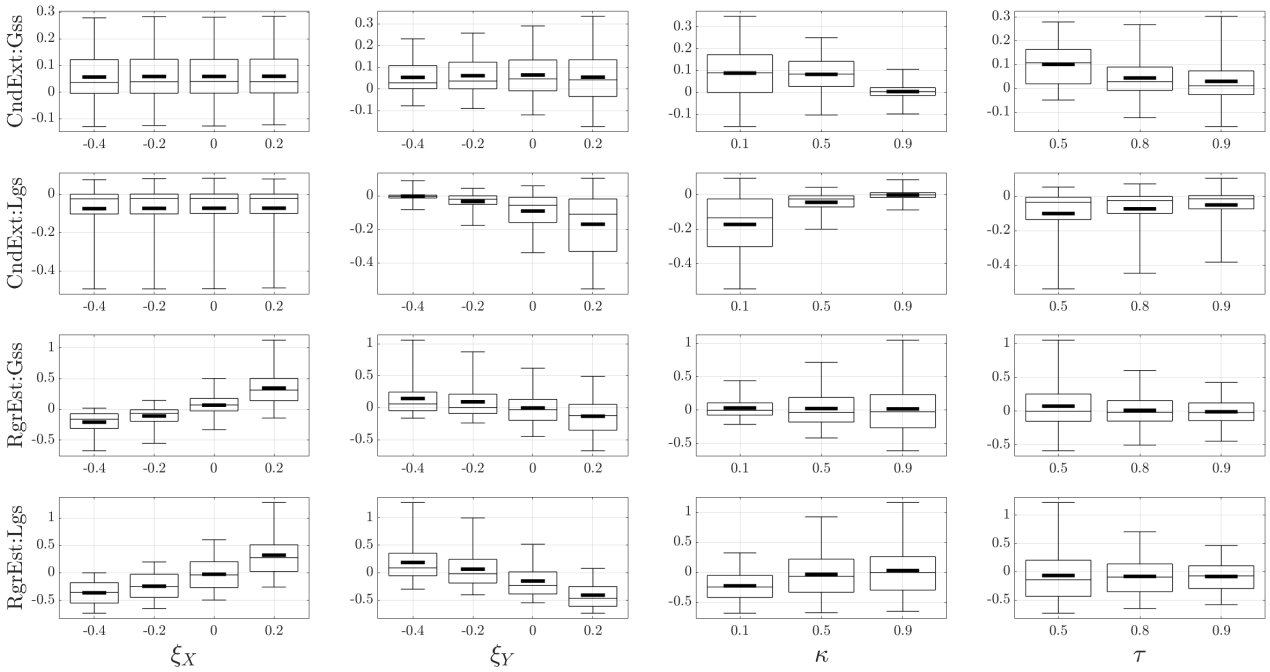

Figure 10: Box-whisker plots of fractional bias in associate value estimator q_{62} as a function of sample covariates for conditional extremes (CndExt) and simple linear regression (RgrEst) estimation schemes, and Gaussian (Gss) and logistic (Lgs) dependence. Sample size $n = 10,000$.

and simple linear regression (third and fourth rows). For sample size $n = 200$ $(1,000)$, green (orange) lines indicate parameter vectors for q_{61} (solid) and q_{62} (dashed). For $n = 200$ in particular, regression analysis for these and other estimators is problematic because the empirical distribution of fractional bias has a very long right-hand tail due to the presence of exceptional values, particularly when employing the mean (for estimators $q_{\cdot 1}$) rather than the median (for estimators $q_{\cdot 2}$). For this reason, parameter estimates for $q_{\cdot 1}$ with $n = 200$ are not shown, since they are not reliably described using the non-homogeneous Gaussian regression. For $n = 10,000$, grey lines indicate parameter vectors for all estimators $q_{kk'}$, $k = 1, 2, ..., 6$, $k' = 1, 2$; as might be inferred from the rank correlation matrix for $n = 10,000$ in Figure 8, parameter estimates for different estimators are very similar in this case.

The regression model is estimated using covariates standardised to mean zero and unit standard deviation over the design, so that the magnitudes of parameter estimates indicate their relative contribution in explaining the variation of fractional bias. Comparing the first and third rows of the figure, for conditional extremes (CntExt) and regression estimation (RgrEst) respectively for both $n = 1,000$ (orange) and $n = 10,000$ (grey), we note that the values of the largest parameter estimates for μ are much smaller in magnitude for conditional extremes estimation; that is, mean fractional bias estimated under the conditional extremes model is less sensitive to covariate variation over the design. Parameter estimates for μ using simple linear regression (RgrEst) for dependence modelling are relatively insensitive to sample size, but the strong positive effect of increasing ξ_X and decreasing ξ_Y on μ are clear. For samples with Gaussian dependence estimated using conditional extremes, fractional bias increases in mean μ and spread σ with increasing ξ_Y, but decreases with increasing dependence strength κ and threshold (with non-exceedance probability τ). These effects are present for all sample sizes. For samples with logistic dependence estimated using conditional extremes, the trends in σ with ξ_Y, κ and τ are the same as for Gaussian dependence, but trends in μ tend to be reversed. All the fitted models illustrated provide descriptions of the variation of fractional bias over the design which are materially (and "significantly") better than a constant model, because of the huge sample size involved. For $n = 10,000$, there is little difference between regression fit for the different estimators (all shown in grey).

8. Discussion and conclusions

We present a computer experiment to assess the relative performance of different estimators and estimation schemes for associated values. The experimental design is chosen to cover combinations of marginal and dependence characteristics representative of typical metocean samples. The focus is on an appropriate choice of estimator to use in

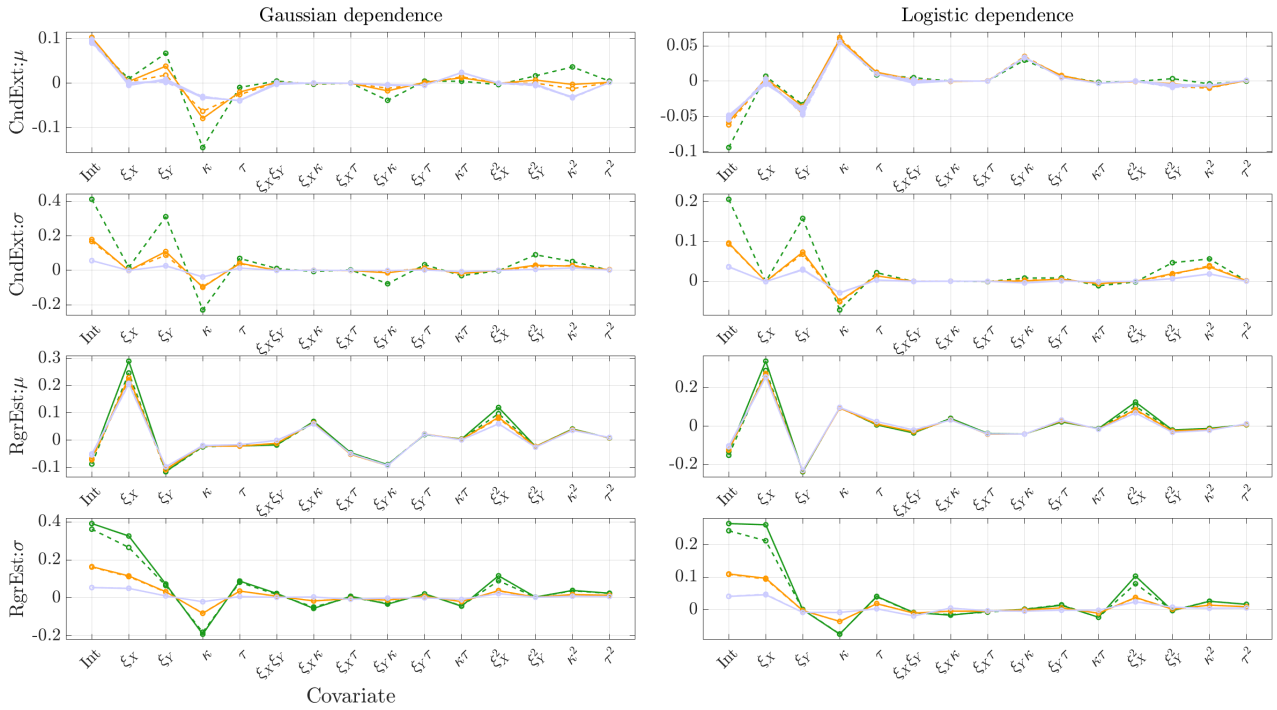

Figure 11: Parameter estimates (φ_μ, φ_σ, assuming standardised covariates) for non-homogeneous Gaussian regression mean μ and error standard deviation σ, for conditional extremes (CndExt) and simple linear regression (RgrEst) estimation schemes, and Gaussian and logistic dependence. Lines in green (orange) correspond to estimators q_{61} (solid) and q_{62} (dashed) for $n = 200$ ($n = 1000$). Lines in grey correspond to *all* estimators for $n = 10,000$. Parameter estimates for estimator q_{61} with $n = 200$ are not shown, as explained in the main text.

combination with the conditional extremes model. The effect of uncertainty in model parameters, estimated from finite samples, on estimates of associated values, is quantified in terms of fractional bias, using a large number of sample replications.

Estimators can be categorised according the nature of the conditioning performed. One group of estimators, referred to here as being *conditioned on value* exploit a previously independently-estimated return value (on physical scale); see Equation 6. The second group, referred to as being *conditioned on probability* impose return value conditioning using the appropriate extreme marginal quantile (with fixed annual non-exceedance probability) per bootstrap resample; see Equation 7. Calculation of the first group of estimators from a sample is sometimes infeasible, especially when the sample size is small. The second group of estimators can always be calculated. Of course, the procedure for calculation of estimators conditioned on value can be modified to mitigate infeasible calculations (e.g. Towe et al. 2017); in this case however, the properties of resulting estimators will also depend on the specifics of the mitigation strategy. In this work, bootstrap resamples resulting in infeasible calculations have simply been ignored, and associated values calculated using the output of feasible calculations only; results of the simulation study suggest that this approach has some merit.

Estimators defined in Equations 6 and 7 can also be categorised according to the choice of operator used to estimate the associated value over uncertain model parameters. One group of estimators (q_{k1}, $k = 1, 2, ..., 6$) uses the mean, and the second (q_{k2}, $k = 1, 2, ..., 6$) the median. For small sample sizes, the distribution of estimators $q_{\cdot 2}$ is less skewed by the occurrence of exceptional values of fractional bias. Of course, the procedure for calculation of estimators can also be modified to mitigate the occurrence of exceptional values, with the consequence that properties of resulting estimators will again depend on the specifics of the mitigation strategy.

The conditional extremes (CndExt) and copula (CplEst) estimation schemes require the fitting of both marginal and dependence models. For small sample sizes, uncertainty in marginal estimation of return values contributes to approximately half the inter-quartile range of the fractional bias of the estimated associated value. For large sample size, uncertainty in the fractional bias of estimated associated values is dominated by the quality of dependence model fit. In the experiments considered, it is assumed for copula modelling that the true underlying dependence type is known. As a result, CplEst out-performs CndExt estimation. For small sample size, the estimated conditional extremes model sometimes yields exceptionally large associated values; this is not surprising, given that the number of observations for conditional extremes model fitting, $n(1 - \tau)$ is between 20 and 100 in this case. In comparison,

the simple regression model (RgrEst) for dependence provides estimates for fractional bias which are biased (away from zero) over the design, and more variable than those estimated using conditional extremes or copula estimation. Fractional bias varies systematically over the design, for all combinations of estimators for associated values and estimation schemes. The size of the effect reduces with increasing sample size. For $n = 10,000$, all estimators are approximately equally sensitive to design variables for a given estimation scheme.

For small sample sizes, specific estimation schemes, including the empirical Bayes approach of Zhang and Stephens (2009), Zhang (2010), the method of moments and probability-weighted moments (see Jonathan et al. 2021) have been developed for improved estimation of marginal GP parameters. Jonathan et al. (2021) demonstrates that the empirical Bayes estimator in particular performs well in estimating the marginal return value, in simulations for a range of GP shape parameters. It is not inconceivable that specific estimators could be developed for improved estimation of associated values. However, one major restriction with estimators such as these, not based on generic maximum likelihood inference, is that the incorporation of important model features (such as covariate effects) within inference is not straightforward.

The ocean engineering literature contains many examples of bivariate (and higher-dimensional) data sets with characteristics broadly similar to those considered in this work, illustrated in Figure 1, produced using simulation models described in Section 6. For example, Figures 1-4 of Jonathan et al. (2010) illustrate bivariate samples of significant wave height H_S and spectral peak period T_P for locations in the northern North Sea, the Gulf of Mexico and on the north-west shelf of Australia; sample sizes are 145, 505, 620, 827. It is clear that the analysis of the smallest sample here would be highly sensitive to arbitrary modelling decisions, including the choice of estimator of associated value. Jonathan et al. (2014a) considers the construction of environmental design contours using the same date; these too would be influenced by similar arbitrary choices of estimator for contour location. There is a large literature on the estimation of engineering design contours; the recent benchmark study of Haselsteiner et al. (2019) and the review of Ross et al. (2020) provide an introduction. Figure 2 of Jonathan et al. (2013) illustrates a bivariate sample of 1163 pairs of H_S and T_P from a northern North Sea location, with a directional covariate. It appears from the figure that the characteristics of the dependence between H_S and T_P indeed varies with direction. The sample of storm peak H_S and associated T_P values available for extreme value analysis in this case is very small. Within specific directional sectors, the figure shows that the effective sample size for analysis is of the order of 100 or less. Again, it is clear that the analysis here is highly sensitive to the choice of estimator for associated value. Jonathan et al. (2014b) presents a similar analysis; Figure 2 therein is illustrative of sample characteristics. The practical engineering significance of the current work in the context of these applications is therefore clear: decisions regarding the selection of data, the inference procedure for marginal and dependence modelling, and in particular the choice of estimator for return value and associated value (from a set of plausible estimators) will change the values of design criteria developed by the metocean engineer. This can have considerable implications for the design and re-analysis of offshore and coastal structures.

The set of estimators of associated values considered here could be extended easily to include other variants. Examples include

$$\begin{aligned} &\mathbb{E}_{\boldsymbol{Z}}\left[g^{-1}_{SY|\boldsymbol{Z}}\left(\mathbb{E}_{Y_S|\boldsymbol{Z}}\left[(Y_S|X_S = g_{SX|\boldsymbol{Z}}(q_k|\boldsymbol{Z}), \boldsymbol{Z})|\boldsymbol{Z}\right]\right)\right] \\ &\text{med}_{\boldsymbol{Z}}\left[g^{-1}_{SY|\boldsymbol{Z}}\left(\mathbb{E}_{Y_S|\boldsymbol{Z}}\left[(Y_S|X_S = g_{SX|\boldsymbol{Z}}(q_k|\boldsymbol{Z}), \boldsymbol{Z})|\boldsymbol{Z}\right]\right)\right] \end{aligned} \tag{12}$$

for which expectation over Y is taken on Laplace scale (as $\mathbb{E}_{Y_S|\boldsymbol{Z}}$), prior to transformation to physical scale.

In this work, a frequentist (bootstrap resampling) approach is used to quantify the effects of epistemic uncertainty on estimates of associated values. Adopting a Bayesian approach to inference would produce similar general findings, since the effects of uncertain models on quantile inference are similar in both cases.

The main recommendations from the experiment can be summarised as follows

1. Estimators of associated values *conditioned on probability* should be preferred, because their calculation is always feasible and their characteristics are at least as good as those of competitors. (In the notation of Section 5, therefore, estimators $q_{6\cdot}$ are preferred.)

2. Estimators of associated values using the median (as opposed to the mean) to summarise the distribution over uncertain model parameters are more robust, especially for smaller sample sizes. (Therefore, estimator q_{62} is preferred to q_{61}.)

3. Sample estimates of associated values reflect the uncertainties of estimates of marginal return values required. Over-estimation of marginal return values (e.g. using estimators based on predictive distributions for small samples) leads to over-estimation of associated values when the variables concerned exhibit positive dependence.

4. The conditional extremes estimation scheme performs reasonably across the design. However, estimation of associated values using sample sizes of the order of 200 is generally problematic, and should be avoided; both marginal and dependence inference results in parameter estimates with large uncertainty, propagating into uncertain estimates of associated values. For sample size $10,000$, uncertainty in estimated associated values is dominated by that of dependence model fit.

5. Estimating the tail dependence between two variables using simple linear regression on physical scale yields estimates of associated values, the fractional bias of which varies systematically with underlying sample marginal and dependence characteristics. The spread of the fractional bias of associated values is relatively large (see Figure 5) even for large sample sizes. This approach should therefore be avoided in general. The simple linear regression model is competitive only for small sample size, for which none of the estimation schemes performs particularly well.

More generally, the findings of this study reinforce and extend those of Jonathan et al. (2021). We therefore enhance one conclusion of that article as follows. Summarising multivariate distributions for metocean variables in terms of return values and associated values has obvious advantages in terms of conciseness of description of an extreme ocean environment, for communication between different parties involved in offshore structural design. However, in reality, the accurate estimation of system risk or probability of structural failure should be the clear focus of analysis. From a predictive perspective, the effects of all sources of modelling uncertainty should be propagated carefully through the entire sequence of design calculations, expressed probabilistically, so that (a) the estimation of failure probability reflects these uncertainties as fully and fairly as possible, and (b) resources can hence be devoted to reducing the largest sources of uncertainty on failure probability in a rational and systematic manner. In this light, the use of summary statistics such as metocean return values and associated values at intermediate design stages in place of estimates for full distributions of metocean variables should be avoided (see also e.g. Serinaldi 2015).

Acknowledgement

The authors gratefully acknowledge useful discussions with Emma Ross at Shell in Amsterdam. We thank two anonymous reviewers for their comments on the work.

Appendix

Distribution of annual maximum

According to asymptotic statistical theory, under the assumption that the random variable X belongs to the maximum domain of attraction of a non-degenerate distribution, the distribution of X exceeding threshold η converges (Pickands 1975) to the GP distribution as the threshold increases. We therefore typically assume, for sufficiently large threshold η that the conditional distribution function $F_{X|X>\eta,\boldsymbol{Z}}$ for parameters $\boldsymbol{Z}$ follows GP form. Usually, the value of threshold is set prior to estimation of ξ and ς from the sample of values for X (e.g. by examining a mean residual life plot, Coles 2001), although this is not always the case (Scarrott and MacDonald 2012). Given $F_{X|X>\eta,\boldsymbol{Z}}$, the distribution $F_{A|\boldsymbol{Z}}$ of annual maxima can be derived using $F_{A|\boldsymbol{Z}}(x|\boldsymbol{Z}) = \Pr(A \leq x|\boldsymbol{Z}) = \sum_{k=0}^{\infty} f_C(k) F_{X|\boldsymbol{Z}}^k(x|\boldsymbol{Z})$ where C is the number of occurrences of X per annum, with probability mass function f_C and $F_{X|\boldsymbol{Z}}(x|\boldsymbol{Z}) = \tau + (1-\tau)F_{X|X>\eta,\boldsymbol{Z}}(x|\boldsymbol{Z})$ where $\tau = \Pr(X < \eta)$. In practice, f_C is unknown and must also be estimated from data. Density f_C is often described by a Poisson distribution such that $f_C(k) = \exp[-\lambda]\lambda^k/k!$, $k = 0, 1, 2, ...$, for annual rate $\lambda > 0$ to be estimated. Assuming for simplicity that λ is known, the expression for $F_{A|\boldsymbol{Z}}$ simplifies (e.g. Ross et al. 2017) to

$$F_{A|\boldsymbol{Z}}(x|\boldsymbol{Z}) = \exp[-\lambda\left(1 - F_{X|\boldsymbol{Z}}(x|\boldsymbol{Z})\right)].$$

By setting $F_{A|\boldsymbol{Z}} = 1 - 1/N$ in the above, we estimate return value estimators given $F_{X|X>\eta,\boldsymbol{Z}}$ (with parameters ξ and ς conditional on $\boldsymbol{Z}$) for any return period N.

Estimated mean and standard deviation of fractional bias over design

Building on Figures 5 and 6, Figures 12 and 13 give empirical means and standard deviations of fractional bias for all estimators, for different sample sizes and dependence threshold non-exceedance probabilities τ. Means and standard deviations for estimators $q_{kk'}$, $k = 1, 2, ..., 6$, $k' = 1, 2$ are only reported over bootstrap resamples for which the associated value calculation is feasible. The y-axis intervals shown have been restricted for visibility, resulting in summary statistics for some combinations of estimators and estimation schemes not appearing in the figures. This issue is particularly problematic for $n = 200$; estimates of the mean and standard deviation are influenced by a small

number of exceptionally high values of fractional bias, especially for the conditional extremes estimation. This is not surprising, given that the number of observations for conditional extremes model fitting, $n(1-\tau)$ is between 20 and 100 in this case. The performance of estimators q_{61}, q_{62} is relatively good for $n = 1,000$ and 10,000, but there are again some issues with marginal fitting (indicated by missing green crosses in figures, and closer inspection of detailed results) for $n = 200$. The simple regression estimation scheme yields considerably larger standard deviations of fractional bias than other estimation schemes when $n = 10,000$, reflecting findings for IQR in Figure 5.

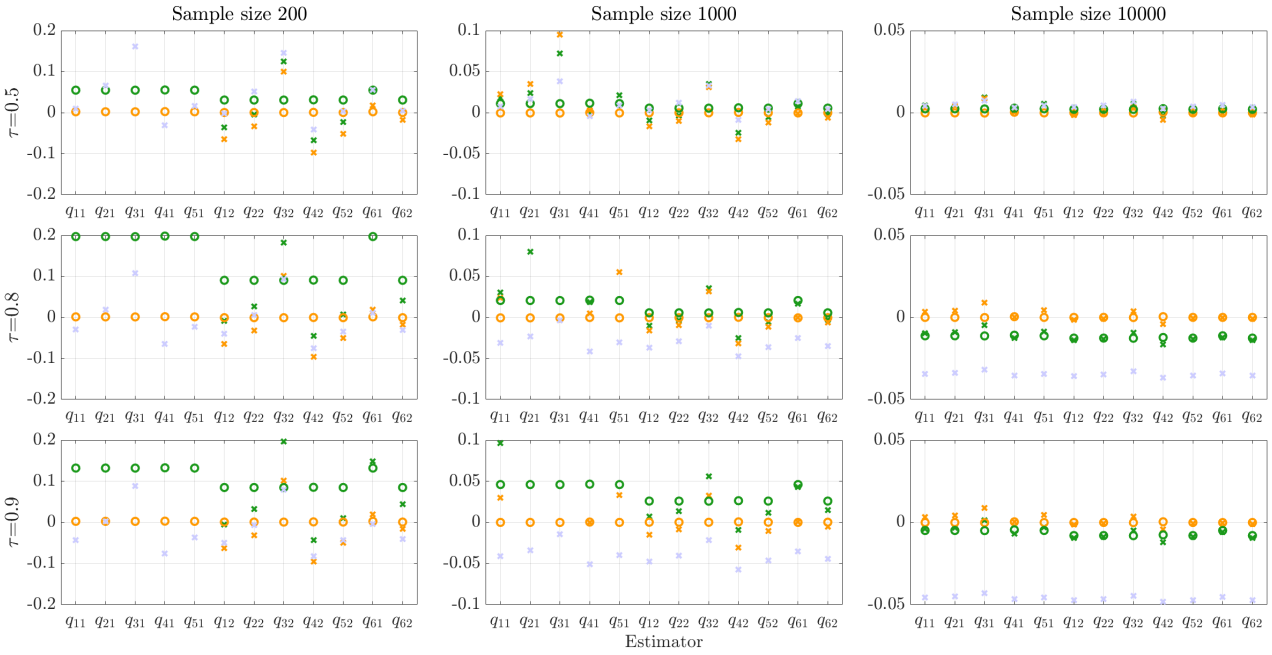

Figure 12: Mean fractional bias for all estimators, over all design combinations, for different sample sizes (n, columns) and dependence modelling thresholds (with non-exceedance probability τ, rows). Colours and symbols indicate simple regression estimation (RgrEst, grey crosses) and conditional extremes estimation (CndExt, green) with unknown (crosses) and known margins (circles). Also shown as a benchmark, are corresponding estimates under the correct dependence copula D form (but unknown parameter κ, orange) with unknown (crosses) and known margins (circles), using the full sample for κ estimation. For details, see caption of Figure 5. Note further that for estimators $q_{kk'}$, $k = 1, 2, ..., 5$, $k' = 1, 2$, only accepted bootstrap resamples are used (see Section 7.1), and that y-axis limits have been limited.

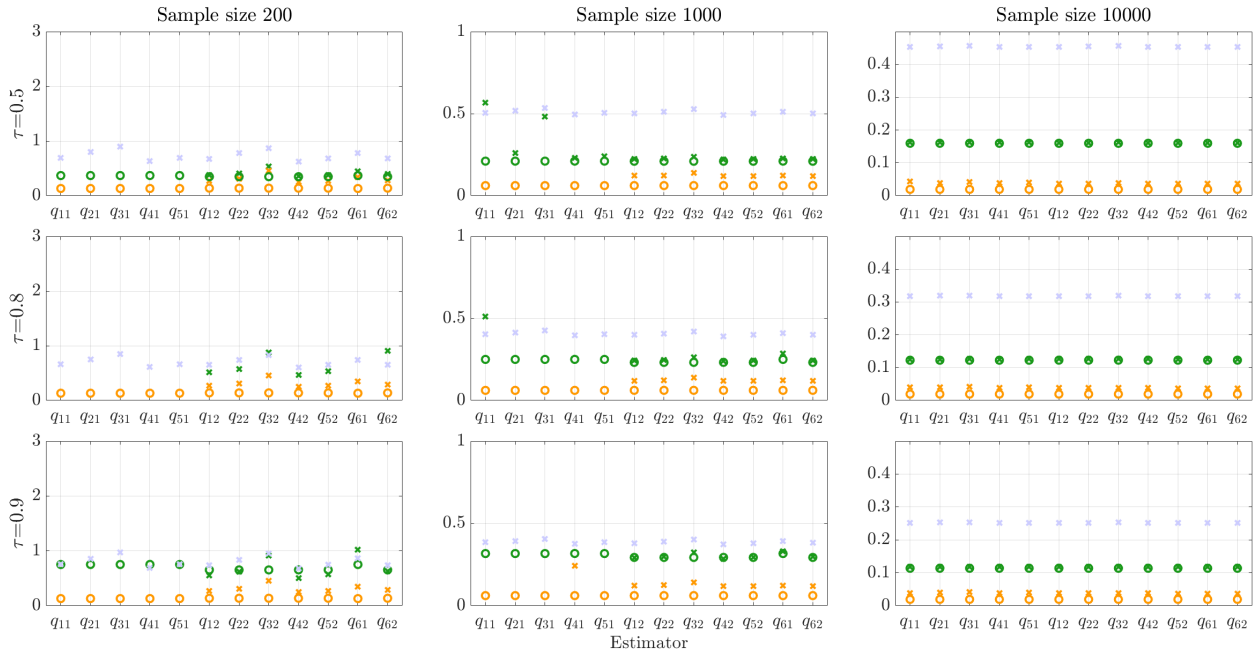

Figure 13: Standard deviation of fractional bias for all estimators, over all design combinations, for different sample sizes (n, columns) and dependence modelling thresholds (with non-exceedance probabilities τ, rows). For details, see caption of Figure 12. Note that for estimators $q_{kk'}$, $k = 1, 2, ..., 5$, $k' = 1, 2$, only accepted bootstrap resamples are used, and that y-axis limits have been limited.

References

Bitner-Gregersen, E.M., 2015. Joint met-ocean description for design and operations of marine structures. Appl. Ocean Res. 51, 279–292.

Bitner-Gregersen, E.M., Haver, S., 1989. Joint long term description of environmental parameters for structural response calculations., in: Proc. 2nd International workshop on wave hindcasting and forecasting.

Chai, W., Leira, B.J., 2018. Environmental contours based on inverse SORM. Mar. Struct. 60, 34–51.

Chavez-Demoulin, V., Davison, A., 2005. Generalized additive modelling of sample extremes. J. Roy. Statist. Soc. Series C: Appl. Stat. 54, 207–222.

Chavez-Demoulin, V., Davison, A., 2012. Modelling time series extremes. Revstat 10, 109–133.

Chen, B.Y., Kou, Y., Zhao, D., Wu, F., Wang, L.P., Liu, G.L., 2021. Maximum entropy distribution function and uncertainty evaluation criteria. China Ocean Eng. 35, 238–249.

Coles, S., 2001. An introduction to statistical modelling of extreme values. Springer.

Coles, S., Heffernan, J., Tawn, J., 1999. Dependence measures for extreme value analyses. Extremes 2, 339–365.

Coles, S.G., Tawn, J.A., 1991. Modelling multivariate extreme events. J. Roy. Statist. Soc. B 53, 377–92.

Davison, A., Smith, R.L., 1990. Models for exceedances over high thresholds. J. R. Statist. Soc. B 52, 393.

Davison, A.C., Padoan, S.A., Ribatet, M., 2012. Statistical modelling of spatial extremes. Statist. Sci. 27, 161–186.

Ewans, K.C., Jonathan, P., 2014. Evaluating environmental joint extremes for the offshore industry. J. Mar. Syst. 130, 124–130.

Furrer, R., Naveau, P., 2007. Probability weighted moments properties for small samples. Stat. Probab. Letters 70, 190?195.

Gudendorf, G., Segers, J., 2010. Extreme-value copulas, in: Copula theory and its applications. Lecture Notes in Statistics 198 (Edts. Jaworski P., Durante F., Hardle W., Rychlik T.), Springer, Berlin, Heidelberg. pp. 127–145.

Haselsteiner, A.F., 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 Engineering 236, 109504.

Haselsteiner, A.F., Lehmkuhl, J., Pape, T., Windmeier, K.L., Thoben, K.D., 2019. A software to compute multivariate extremes using the environmental contour method. SoftwareX 9, 95–101.

Haver, S., 1985. Wave climate off northern Norway. Appl. Ocean Res. 7, 85–92.

Haver, S., 1987. On the joint distribution of heights and periods of sea waves. Ocean Eng. 14, 359–376.

Heffernan, J., 2000. A directory of coefficients of tail dependence. Extremes 3, 279–290.

Heffernan, J.E., Tawn, J.A., 2004. A conditional approach for multivariate extreme values. J. R. Statist. Soc. B 66, 497–546.

ISO19901-1, 2015. Petroleum and natural gas industries. Specific requirements for offshore structures. Part 1: Metocean design and operating considerations. First edition. International Standards Organisation.

Joe, H., 2014. Dependence modelling with copulas. CRC Press.

Jonathan, P., 2022. Estimation of associated values from conditional extreme value models (MATLAB software and supplementary figures). `https://github.com/ygraigarw/AssValCntExt`.

Jonathan, P., Ewans, K., Flynn, J., 2014a. On the estimation of ocean engineering design contours. ASME J. Offshore Mech. Arct. Eng. 136:041101.

Jonathan, P., Ewans, K.C., Flynn, J., 2012. Joint modelling of vertical profiles of large ocean currents. Ocean Eng. 42, 195–204.

Jonathan, P., Ewans, K.C., Randell, D., 2013. Joint modelling of environmental parameters for extreme sea states incorporating covariate effects. Coastal Eng. 79, 22–31.

Jonathan, P., Ewans, K.C., Randell, D., 2014b. Non-stationary conditional extremes of northern North Sea storm characteristics. Environmetrics 25, 172–188.

Jonathan, P., Flynn, J., Ewans, K.C., 2010. Joint modelling of wave spectral parameters for extreme sea states. Ocean Eng. 37, 1070–1080.

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. 220, 107725.

Mathisen, J., Bitner-Gregersen, E., 1990. Joint distributions for significant wave height and wave zero-up-crossing period. Appl. Ocean Res. 12, 93–103.

Montes-Iturrizaga, R., Heredia-Zavoni, E., 2017. Assessment of uncertainty in environmental contours due to parametric uncertainty in models of the dependence structure between metocean variables. Appl. Ocean Res. 64, 86 – 104.

NORSOK N-006, 2015. NORSOK Standard N-006:2015: Assessment of structural integrity for existing offshore load-bearing structures. NORSOK, Norway.

Petrov, V., Soares, C.G., Gotovac, H., 2013. Prediction of extreme significant wave heights using maximum entropy. Coastal Eng. 74, 1–10.

Pickands, J., 1975. Statistical inference using extreme order statistics. Ann. Statist. 3, 119–131.

Rootzen, H., Segers, J., Wadsworth, J.L., 2018. Multivariate generalized pareto distributions: parametrizations, representations, and properties. J. Multivariate Anal. 165, 117–131.

Ross, E., Astrup, O.C., Bitner-Gregersen, E., Bunn, N., Feld, G., Gouldby, B., Huseby, A., Liu, Y., Randell, D., Vanem, E., Jonathan, P., 2020. On environmental contours for marine and coastal design. Ocean Eng. 195, 106194.

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.

Scarrott, C., MacDonald, A., 2012. A review of extreme value threshold estimation and uncertainty quantification. Revstat 10, 33–60.

Serinaldi, F., 2015. Dismissing return periods! Stoch. Env. Res. Risk A. 29, 1179–1189.

Shooter, R., Ross, E., Ribal, A., Young, I.R., Jonathan, P., 2022. Multivariate spatial conditional extremes for extreme ocean environments. Ocean Eng. 247, 110647.

Smith, R.L., Naylor, J.C., 1987. A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. J. Roy. Statist. Soc. C 36, 358–369.

Tawn, J.A., 1988a. Bivariate extreme value theory: models and estimation. Biometrika 75, 397–415.

Tawn, J.A., 1988b. Modelling multivariate extreme value distributions. Biometrika 77, 245–253.

Tendijck, S., Eastoe, E., Tawn, J., Randell, D., Jonathan, P., 2021. Modeling the extremes of bivariate mixture distributions with application to oceanographic data. J. Am. Statist. Soc. doi:10.1080/01621459.2021.1996379.

Tendijck, S., Ross, E., Randell, D., Jonathan, P., 2019. A non-stationary statistical model for the evolution of extreme storm events. Environmetrics 30, e2541.

Towe, R., Eastoe, E., Tawn, J., Jonathan, P., 2017. Statistical downscaling for future extreme wave heights in the North Sea. Ann. Appl. Stat. 11, 2375–2403.

Tromans, P.S., Vanderschuren, L., 1995. Variable based design conditions in the North Sea: application of a new method. Offshore Technology Confernence, Houston (OTC–7683) .

Williams, R., Ferro, C., Kwasniok, F., 2014. A comparison of ensemble post-processing methods for extreme events. Quart. J. R. Met. Soc. 140, 1112–1120.

Zanini, E., Eastoe, E., Jones, M., Randell, D., Jonathan, P., 2020. Covariate representations for non-stationary extremes. Environmetrics 31, e2624.

Zhang, J., 2010. Improving on estimation for the generalized Pareto distribution. Technometrics 52, 335–339.

Zhang, J., Stephens, M.A., 2009. A new and efficient estimation method for the generalized Pareto distribution. Technometrics 51, 316–325.