Extreme value methods for the estimation of offshore structure failure probability

Matthew Speers, B.Sc.(Hons.), M.Res

Submitted for the degree of Doctor of Philosophy at Lancaster University.

February 2026

STOR-i
excellence with impact

Abstract

Estimation of the failure probability of offshore structures exposed to extreme ocean environments is critical to their safe design and operation. This estimation must consider the joint extremal behaviour of the metocean environment surrounding the structure, and the physical interaction between the environment and the structure. This thesis uses multivariate conditional extreme value models for the joint metocean environment, coupled with simulation from physically-based models for short-term wave variability and wave-structure interaction.

Using this forward approach, we can capture the uncertainty and short-term variability in each stage of estimation. Existing approaches for offshore structure design, such as environmental contour methods, do not account for this short-term variability in the ocean environment. They make simplifying assumptions about the wave-structure interaction to avoid the need for simulation from physically-based models, since this simulation can be computational infeasible in practice. We demonstrate that these assumptions are not valid for all structural scenarios, making structural risk assessment with these contour methods unreliable.

We develop two alternative methods for the estimation of structural failure probability when full forward simulation is infeasible. These consist of an efficient sampling technique that combines importance sampling and parallel tempering Markov-chain Monte-Carlo, and a Gaussian emulator for the physical model with associated utility function for active learning. Each of these approaches outperform a standard Monte-

Carlo sampling method for representative structural scenarios.

When using the conditional extreme value model, it is required to select a threshold in one 'conditioning variable', such that any data with conditioning variable above the threshold is used for inference. Care must be taken when choosing this threshold to best balance the associated bias-variance trade-off. Our method for automatic threshold selection allows for optimised use of the observed data. We demonstrate the benefits of our method when estimating joint extreme probabilities and model parameters over existing methods.

Acknowledgements

I would first like to thank my supervisors Jon Tawn, Phil Jonathan, and David Randell.

I've been very lucky to have been supervised by Jon throughout my PhD. Jon is approachable, friendly, funny, and the hardest worker I have ever met. I've learned so much from him over the past few years, and his discipline and work ethic is inspiring. This PhD has given me so much that I wouldn't have otherwise, and I've got Jon to thank for all of that.

Phil was always there to offer advice on work, careers, and people. Our days spent writing papers and working on code were so important in getting this thesis completed, and I'll always enjoy a trip to Ruthin for fish and chips.

Thanks to David too for giving me my proper introduction to Python coding, and for all the help and patience in explaining random wave simulation. Also thank you to David and everybody else at Shell for being so welcoming during my visits to London and Amsterdam.

The support offered by members of the TIDE research hub at the University of Western Australia has been invaluable. Thank you to Paul, Jana, Ed, Lachy, Hugh, Kath, and everybody else who I've met these past few years. You all made me feel very welcome out in Perth.

The STOR-i CDT has been a fantastic place to work, and I'd like to thank everybody involved in the organisation of the department for their work and support. Thank you to each of my friends in STOR-i, particularly Carla, Robert, Wanchen, Eleanor, Connie,

Rui, Kristina, Callum and Thomas for all of your support inside and out of work. The last few years have been made so much better by you all.

Thank you to my parents and grandparents for all of your work which has lead me to this stage. I would never have been able to get here without your help. The sacrifices that you made to give me my start in life have made all of this possible, and in this way you've contributed to this PhD as much as anyone.

最后，我想要感谢 Qianqian。我每天都在想我有多幸运，因为没有你的支持，我不可能取得这些成就。我为所有我们一起做过的事情感到自豪，并且十分期待我们的未来。

Declaration

I declare that the work in this thesis has been done by myself and has not been submitted elsewhere for the award of any other degree.

Chapter 3 has been published as Speers, M., Randell, D., Tawn, J., & Jonathan, P. (2024). Estimating metocean environments associated with extreme structural response to demonstrate the dangers of environmental contour methods. Ocean Engineering, 311, 118754.

Chapter 4 has been published as Speers, M., Tawn, J., & Jonathan, P. (2026). Sequential design for the efficient estimation of offshore structure failure probability. Ocean Engineering, 349, 123990.

Chapter 6 is a result of the Lancaster and Maynooth University contribution for a data challenge competition as part of the 2023 Extreme Value Analysis conference at the University of Bocconi, Italy. This has been published as André, L. M., Campbell, R., D' Arcy, E., Farrell, A., Healy, D., Kakampakou, L., Murphy, C., Murphy-Barltrop, C. J. R., and Speers, M. (2025). Extreme value methods for estimating rare events in Utopia: EVA (2023) conference data challenge: team Lancopula Utopiversity. Extremes, 28:23–45. https://doi.org/10.1007/s10687-024-00498-w. My primary contributions are in Section 6.5.

The word count for this thesis is approximately 49800.

Matthew Speers

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

Extreme value theory is the study of the statistical behaviour of extremes of random phenomena. Development of univariate extreme value theory began with the work of Fréchet (1927), Fisher and Tippett (1928), Von Mises (1936) and Gnedenko (1943), who together defined the extreme value theorem, the equivalent of the central limit theorem for block maxima. This theory was then extended to the bivariate setting by the likes of Geffroy (1958), Tiago de Oliveira (1959), and Sibuya (1960), and then to the multivariate setting by e.g., de Haan and Resnick (1977), Resnick (2008) and Pickands (1981). Extreme value theory remains an active area of research, with works such as those of Heffernan and Tawn (2004), Keef et al. (2013), Nolde (2014) and Nolde and Wadsworth (2022) seeing continuing development.

Models derived from extreme value theory provide a theoretically justified framework for modelling the tails of distributions of observed processes, allowing for extrapolation beyond the range of historical observations. These models have found application in a variety of areas that require understanding of rare events, including sports, finance and environmental sciences. In environmental sciences in particular, the use of princi-

pled statistical models that provide good quality inference for rare events is becoming increasingly important due to the rising occurrence of extreme weather and climate events. This increased rate of extreme events leads to higher risk to human life, as well as to residential, commercial and energy infrastructure. Extreme value theory has already seen extensive application in environmental modelling and risk mitigation. For instance, at the High Court, extreme value theory was utilised to determine the sequence of events that led to the sinking of the HMS Derbyshire (Heffernan and Tawn, 2003). In addition, extreme value theory has seen application into other areas, each requiring varying levels of extrapolation. At the most extreme end, risk assessment for nuclear power plants requires the estimation of events with annual exceedance probabilities as low as 1-in-10,000 due to strict safety regulation; extreme value theory provides methods which enable this estimation.

In this thesis the focus is on the use of extreme value theory to model storm events at offshore locations important to energy production (such as the North Sea), providing inference for the extremes of metocean variables, such as wave height, wave period or steepness, wind speed, and still water level, and the associated values of wave direction. There are currently 1,500 oil and gas offshore platforms sited around the world, and each are subject to exposure to the hazards from the surrounding waves, wind and other environmental conditions. In extreme storms, this can result in partial or complete failure of the structure, incurring monetary and human cost. A recent example occurred on the 21st October 2023, when 89 personnel required airlift evacuation from a UK North Sea drilling platform due to damage caused by Storm Babet. Such evacuations are expensive and dangerous; therefore, it is essential to design structures at the initial stage to prevent, or at least reduce, the risk of these events sufficiently. The level of risk to a structure can be quantified in terms of a design life, or period P, defined as the number of years the structure can be expected to survive in a location before failing. The design period P can be deduced from the annual failure probability of the

structure p via $P = 1/p$, assuming the process is identically distributed between years. Often, resulting design periods are in excess of 1000 years, far beyond the range of most observational or hindcast data, corresponding to failure events with exceedance probabilities much smaller than those of historical observations. Current literature focuses on both the marginal and joint modelling of such data (Jonathan et al., 2013, 2014b; Tendijck et al., 2023a), since structures are often more susceptible to damage during joint extreme events, i.e., simultaneously large wave height and wave steepness may be more likely to induce structural failure than marginal extreme events of the same, or even larger, magnitudes.

What is considered structural failure varies with structure scenario, with definitions including deformation of the structure beyond a given tolerance, the occurrence of a wave in deck event, or complete structure collapse (see e.g., Santo et al. 2020 for a discussion on types of loading and experimental results on wave in deck events). Estimation of the probability of failure thus requires both modelling of the joint environment (e.g., significant wave height H_s and peak wave steepness S_e) and knowledge of the loading L and response R induced on the structure as a result of impact of individual waves in the time period when (H_s, S_e) is observed. In particular, for individuals waves with a surface elevation η, velocity u and acceleration $\dot{u}$ that are observed within a joint environment with underlying conditions (H_s, S_e).

Figure 1.1.1 lays out the relationship between these variables. Each storm event has a storm peak sea state, denoted $\mathbf{X}^{sp}$, which corresponds to the sea state (of length e.g., 3 hours) within the storm with the most severe joint environment conditions, and contains the values of H_s^{sp} and S_e^{sp} within this peak 3-hour period. The sea state variables $\mathbf{X}_1, \mathbf{X}_2, \ldots$ describe the temporal evolution of the storm around this peak sea state, in the form of temporally dependent copies $\mathbf{X}_1, \mathbf{X}_2, \ldots$ of $\mathbf{X} = (H_s, S_e)$ corresponding to each of the 3-hour periods surrounding the storm peak. The variables $\mathbf{X}_1, \mathbf{X}_2, \ldots$ can be modelled conditional on the value of storm peak sea state $\mathbf{X}^{sp}$ (see, e.g., Tendijck et al.

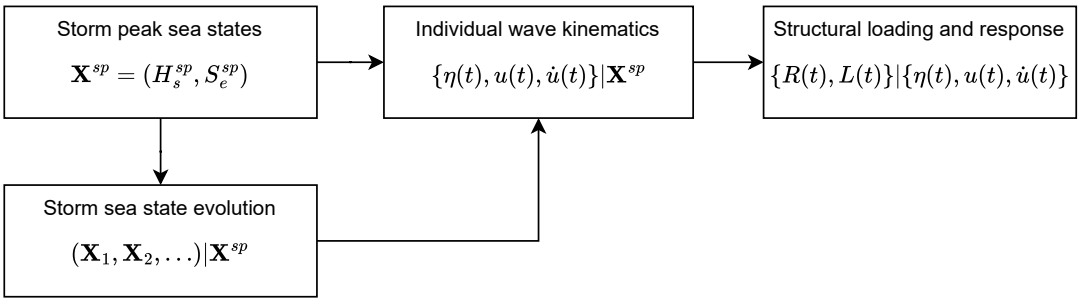


Figure 1.1.1: Illustration of key stages in full probabilistic analysis of structure failure probability. The flowchart describes the relationships between storm peak sea state variable $\mathbf{X}^{sp} = (H_s^{sp}, S_e^{sp})$, general sea state variable $\mathbf{X} = (H_s, S_e)$ with value $\mathbf{X}_t = (H_{s,t}, S_{e,t})$ at time $t = 1, 2, \ldots$, sea surface elevation $\eta(t)$, water velocity $u(t)$, water acceleration $\dot{u}(t)$, fluid loading $L(t)$ and structural response $R(t)$, for time t.

2019), and so we write $(\mathbf{X}_1, \mathbf{X}_2, \ldots)|\mathbf{X}^{sp}$. To model the physical interaction between the joint environment and the structure within a given storm, the assumption is often made that the structural responses induced within the storm peak sea state $\mathbf{X}^{sp}$ dominate those induced by sea states $(\mathbf{X}_1, \mathbf{X}_2, \ldots)|\mathbf{X}^{sp}$. Under this assumption, it is enough to obtain the individual wave kinematics for only the 3-hour period corresponding to $\mathbf{X}^{sp}$; for example sea surface elevation $\eta(t)$, water particle velocity $u(t)$ and water particle acceleration $\dot{u}(t)$ at times $t \in \mathbb{R}$ within the 3-hour sea state corresponding to $\mathbf{X}^{sp}$, and to study their behaviour conditional on the values of H_s^{sp} and S_e^{sp}. By obtaining water velocities and accelerations for the storm peak sea state $\mathbf{X}^{sp}$, and propagating these properties through models for the structural loading $L(t)$ and response $R(t)$, we can understand the effect on the structure induced by a given storm.

Our approach fully captures the long term stochasticity in the effect of the environment on the structure, by combining both aspects to propagate uncertainty through each stage of the analysis shown in Figure 1.1.1, providing a full probabilistic analysis of structure reliability. In contrast, previous statistical methods for structural risk assessment, e.g., Coles and Tawn (1994), only consider the joint modelling of H_s^{sp} and S_e^{sp}, neglecting the modelling of individual wave variability and making simplifying assumptions about the wave-structure interaction, leading to inferences that ignore the

stochastic nature of the ocean surface. On the other hand, oceanographers and engineers have developed models for the short term variability of the ocean surface and its interaction with the structure (e.g., Morison et al. 1950, Taylor et al. 1997, Taylor et al. 2024 and Orszaghova et al. 2025) but do not utilise statistical models for the long term extreme environment.

1.1.1 Joint modelling of storm peak sea states

Estimation of the joint distribution of the storm peak sea state variable $\mathbf{X}^{sp}$ (i.e., the stage of the analysis shown in the top left of Figure 1.1.1) is carried out using extreme value methods. Marginal extreme values of sea state variables H_s^{sp} and S_e^{sp} can be modelled using the peaks over threshold method of Davison and Smith (1990). This involves modelling the marginal variable values below some threshold $u \in \mathbb{R}$ empirically from historical observations, and modelling the excess values above u using the generalised Pareto distribution (GPD), with scale parameter $\sigma > 0$ and shape parameter $\xi \in \mathbb{R}$.

When modelling joint extreme events, even when the marginal distributions of H_s^{sp} and S_e^{sp} are well modelled, there is substantial sensitivity to inference from the strength of dependence between H_s and S_e estimated. Thus, it is essential to correctly characterise the extreme environment through suitable multivariate extreme value models. Two environmental variables are said to be asymptotically dependent if, in the limit as one becomes large, the probability that the other is large is 1. On the other hand, two variables are said to be asymptotically independent if this limiting conditional probability is 0; see e.g., Coles et al. (1999) for further discussion of extremal dependence types. Incorrectly modelling asymptotically dependent variables as asymptotically independent will therefore result in underestimation of the probability of failure for a structure that is more susceptible to joint extremes, leading to under design and increased risk to personnel and equipment. Conversely, incorrectly modelling asymptotically independent variables as asymptotically dependent will lead to overestimation of

joint probabilities and consequently may result in over design, incurring unnecessary cost.

The issue of selecting whether to model bivariate data as being asymptotically dependent or asymptotically independent is exacerbated in cases of limited or low quality data at the location of interest, which necessitates methods to work with a small number of historic events to estimate events corresponding to return (or equivalently design) periods (i.e., the expected time to observe one occurrence of the event) much longer than the observational period. Furthermore, when working beyond the bivariate setting, modelling extremal dependence structures becomes much more complicated, as different subsets of variables can exhibit asymptotic dependence and asymptotic independence, see the discussion of Section 2.2.4. Additionally, many environmental phenomena exhibit seasonal variability, or experience long term trends induced by climate change, so it is important that any methodology used to model extreme environments is able to capture these variations.

Multivariate extreme value models are already employed in the joint modelling of met-ocean variables, such as wave height and wind speed, with previous work in this area leading to the reinforcement of billions of dollars of offshore facilities. Of particular note is the conditional multivariate extreme value model of Heffernan and Tawn (2004), which has been applied to this context, for example by Jonathan et al. (2010), Tendijck et al. (2019), Towe et al. (2019) and Shooter et al. (2021b). This model provides a flexible framework for modelling the joint behaviour of multiple variables, given that at least one of the variables is large, and it is capable of capturing both asymptotic dependence and asymptotic independence, and thereby prevents issues that arise from mis-characterisation of dependence in joint extreme environments and the resulting potential for underestimation or overestimation of joint exceedance probabilities. The conditional extreme value model of Heffernan and Tawn (2004) has also been successfully applied in non-stationary contexts (Jonathan et al., 2013), allowing inferences to

be made regarding structural failure probability that correctly reflect underlying trends in the environment.

As of yet, however, there is no approach that allows the conditional multivariate extreme value model to automatically make best use of the available data, through optimally balancing the bias and variance trade-off. When performing inference using an observed sample, the user must specify which subset of the sample can be considered extreme by choosing an exceedance threshold for one variable, with observations where this variable is below the threshold being discarded from the conditional inference. Incorrectly retaining too many data (selecting too low a threshold) can lead to biased inference, as non-extreme values will be included in model parameter estimation and violate underlying theoretical assumptions. Conversely, if too many data are discarded (by selecting too high a threshold) the uncertainty in model parameters and resulting estimates of joint environment and structural failure probabilities will be unnecessarily increased. We introduce a novel technique to optimally choose the exceedance threshold for use in the conditional extreme value model. Our approach allows us to retain the maximum amount of data in our modelling, without mistakenly including non-extreme data, which introduces bias in the inferences. This ensures our method is optimal in reducing uncertainty.

1.1.2 Physical simulation

Estimates of structure failure probability should consider the specifics of the wave-structure interaction for the given scenario, via the simulation of individual wave kinematics (i.e., top middle of Figure 1.1.1) and associated fluid loading and response (i.e., top right of Figure 1.1.1). For example, an oil platform may be resistant to all waves up to its deck height but susceptible to wave in deck events, whereas a monopile (or 'stick-type') structure will have no 'critical' wave height in this sense but will instead experience deformation of the jacket support structure proportional to the amount of fluid

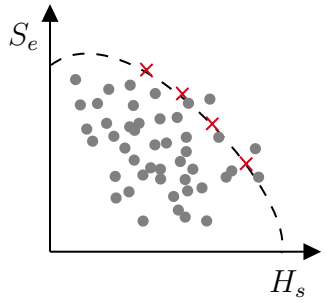


Figure 1.1.2: Toy example of a P-year environmental contour show as the dashed line, constructed using the sample of (H_s, S_e) shown in grey. The points marked by red crosses indicate possible design conditions, chosen as points along the contour boundary. If a structure withstands simulated wave kinematics and associated structural response corresponding to these design conditions, it can be said to possess the design period P.

in a wave impact of any height. When modelling the interaction between an extreme environmental event and a particular structure, it is therefore crucial to consider the specific type of interaction present via the design process described in Figure 1.1.1, by using physically motivated models for individual wave surface elevation, wave kinematics, and the wave spectrum, as well as models for the load or response on the structure resulting from wave impact. Models for these properties are well-developed; however, they are often computationally costly to simulate from for a large number of events, making estimation of structure failure probability infeasibly expensive.

1.1.3 Environmental contours

One way to circumvent this computational cost is through the use of environmental contour methods (e.g., Chai and Leira 2018, Ross et al. 2020, Haselsteiner et al. 2021, Mackay and Haselsteiner 2021, Hafver et al. 2022 and Mackay and de Hauteclocque 2023). Environmental contour methods are statistical approaches that provide a contour in the environment space which acts a set of design conditions (i.e., values of $\mathbf{X}$) associated with a given design period P. A structure is then said to possess the design period P if it can survive all the design conditions, determined by simulation from wave kinematic and structural response models at a subset of the values for $\mathbf{X}$ along

the contour. See Figure 1.1.2 for an illustration of this idea, showing an example P-year contour and corresponding design conditions.

Environmental contour approaches have seen much application, with a key example being the IFORM contour approach of Hafver et al. (2022), which has been recommended by both NORSOK N-003 (2017) and DNVGL-RP-C205 (2017) standards. This method obtains a design contour by transforming observations to the joint Gaussian space using the Rosenblatt transform (Rosenblatt, 1952), constructing a contour of constant density that is associated with the desired return period P in the joint Gaussian space, then mapping this contour back to the original space.

As a trade-off to their computational savings, environmental contours make simplifying assumptions about the linearity of failure boundaries in the environment space, typically assuming that any structural response increases monotonically with increases in each separate marginal environmental variables. This is not a valid assumption for all structural types; for instance floating structures can experience amplifying resonance affects at intermediate wave heights at their resonant frequencies (e.g., see Wang et al. 2024), resulting in more severe structural responses than those seen at larger wave heights at non-resonant frequencies. Additionally, the approximation error present in these methods is unknown and so cannot be properly accounted for in inference, thereby motivating the development of methods without such an approximation, such as those proposed in this thesis.

1.1.4 Efficient sampling methods

As alternatives to environmental contours, methods have been developed that make optimal use of a limited number of simulations from wave kinematic and structural interaction models when estimating structural failure probability. These include approaches for efficient sampling using importance sampling and Markov chain - Monte Carlo (MCMC), see e.g., Xiao et al. (2020) and Lystad et al. (2023), which target

regions in the space of environmental variables that are most likely to induce large structural response and also most likely to occur. Alternatively, emulators based on statistical methods such as Gaussian Processes, see e.g., Gramstad et al. (2020) and Castellon et al. (2023), can be trained on a small number of physical model outputs and then used as surrogates for further evaluation. These emulators can then be updated with the output from subsequent physical model simulations at environmental conditions chosen according to utility functions, designed to optimally reduce uncertainty in failure probability estimates.

In a case study based on observed North Sea environmental conditions, presented in Chapter 3, we demonstrate the drawbacks of environmental contours resulting from the limiting assumptions they make regarding types of wave structure interactions by combining the conditional multivariate extreme value model for the joint environment with simulations from models for wave kinematics and fluid loading. We propose two methods for the efficient estimation of structural failure probability: one utilising importance sampling coupled with parallel tempering MCMC (e.g., Vousden et al. 2015), and another utilising Gaussian Process emulators with an adaptive sampling algorithm. When combined with extreme value models for the environment, these methods present a complete procedure for the efficient estimation of structural failure probability.

1.2 Overview of thesis

This thesis is divided into a further six chapters. Chapter 2 provides a review of existing approaches for modelling univariate and multivariate extreme values. We discuss the advantages and drawbacks of each approach, finishing with a justification of our choice of model that is used in the subsequent chapters.

In Chapter 3, we apply multivariate extreme value methods alongside physical models for wave kinematics to evaluate the probability of structural failure for jacket-type

offshore structures. Extreme value analysis (EVA) uses data to estimate long-term extreme environmental conditions for variables such as significant wave height and period, for the design of marine structures. Together with models for the short-term evolution of the ocean environment and for wave-structure interaction, EVA provides a basis for full probabilistic design analysis. Alternatively, environmental contours provide an approximate approach to estimating structural integrity, without requiring structural knowledge. These contour methods also exploit statistical models, including EVA, but avoid the need for structural modelling by making what are believed to be conservative assumptions about the shape of the structural failure boundary in the environment space. These assumptions, however, may not always be appropriate, or may lead to unnecessary wasted resources from over design. We demonstrate a methodology for efficient fully probabilistic analysis of structural failure. From this, we estimate the joint conditional probability density of the environment (CDE), given the occurrence of an extreme structural response. We use CDE as a diagnostic to highlight the deficiencies of environmental contour methods for design; none of the IFORM environmental contours considered characterise CDE well for three example structures. This chapter has been published as Speers et al. (2024).

In Chapter 4, we explore methods for efficient sampling of environment variables for offshore structure design. Estimation of the failure probability of offshore structures exposed to extreme ocean environments is critical to their safe design and operation. The CDE quantifies regions of the space of long term environment responsible for extreme structural response. Moreover, the probability of structural failure is obtained by simply integrating the CDE over the environment space. In this work, two methodologies for estimation of the CDE and failure probability are considered. The first (IS-PT) combines parallel tempering MCMC (for CDE estimation) with important sampling (for eventual estimation of failure probability). The second (AGE) combines adaptive Gaussian emulation with Bayesian quadrature. Both approaches provide large reductions in

the number of function evaluations of the complex structural response to fluid loading, relative to naive brute-force calculations. We evaluate IS-PT and two variants of the AGE procedure in application to a simple artificial structure with multimodal CDE, and a monopile structure exhibiting non-linear resonant response. IS-PT provides reliable results for both applications. The AGE procedures require balancing exploration and exploitation of the environment space, using a typically-unknown weight parameter, λ. When λ is known, perhaps from prior engineering knowledge, AGE provides an order of magnitude reduction in computational cost relative to IS-PT. However, with λ unknown, IS-PT is more reliable. This chapter has been accepted for publication in Ocean Engineering, and is currently available as preprint Speers et al. (2026).

In Chapter 5, we propose a novel methodology for the automated selection of exceedance thresholds to be used in the conditional multivariate extreme value model. We develop a metric that captures the bias in a conditional extremes model fitted above a given threshold. Studying this metric for a range of candidate thresholds allows for informed, semi-automatic threshold selection. We also extend the metric by applying a weighted variance penalty. After selection of an appropriate weight value, this formulation allows for fully-automatic selection of the conditioning threshold through minimisation of the variance-weighted metric. We demonstrate the performance of our method on synthetic datasets that represent a range of extremal dependence structures, comparing to random threshold choices and the method of Wan and Davis (2019), who develop a threshold selection for a strong form of extremal dependence, i.e., asymptotic dependence. In contrast our method works well for both asymptotic dependent and asymptotic independent cases.

Chapter 6 introduces a variety of methods to capture the extremal behaviour of complex environmental phenomena in practice. These methods were used by the Lancopula Utopiversity team to tackle the EVA (2023) Conference Data Challenge. This data challenge was split into four challenges, labelled C1-C4. Challenges C1 and C2

comprise univariate problems, where the goal is to estimate extreme quantiles for a non-stationary time series exhibiting several complex features. For these, we propose a flexible modelling technique, based on generalised additive models, with diagnostics indicating generally good performance for the observed data. Challenges C3 and C4 concern multivariate problems where the focus is on estimating joint probabilities. For challenge C3, we propose an extension of available models in the multivariate literature and use this framework to estimate joint probabilities in the presence of non-stationary dependence. Finally, for challenge C4, which concerns a 50-dimensional random vector, we employ a clustering technique to achieve dimension reduction and use a conditional modelling approach to estimate extremal probabilities across independent groups of variables. This work has been published as André et al. (2025).

Chapter 2

Literature Review

2.1 Univariate extreme value theory

2.1.1 Modelling block maxima

Take a continuous univariate random variable $X \sim F$ with an unknown distribution function F on $\mathbb{R}$. Consider the random variable $M_n = \max\{X_1, \ldots, X_n\}$, the maximum of a sequence of $n < \infty$ independently identically distributed (IID) replicates of X. Classical extreme value theory (see e.g., Gumbel 1958) seeks the cumulative distribution function of the block maxima M_n. This can be found via

$$\begin{aligned}
\Pr(M_n \leq x) &= \Pr(X_1 < x, \ldots, X_n \leq x), \\
&= \Pr(X_1 \leq x) \times \ldots \times \Pr(X_n \leq x) \\
&= \Pr(X \leq x)^n = F^n(x),
\end{aligned}$$

since $X_1, \ldots, X_n$ are IID. However, if $x < x_+$ for upper endpoint $x_+ = \inf\{x \in: F(x) = 1\}$ of F, then

$$\Pr(M_n \leq x) \to 0 \text{ as } n \to \infty,$$

and if $x > x_+$, then for all n

$$\Pr(M_n \leq x) = 1.$$

This causes the distribution F^n to degenerate to a point mass on x_+ in the limit as $n \to \infty$, making this an unreliable basis for a statistical approach.

This degeneracy issue is resolved by transforming M_n using sequences $\{a_n > 0\}$ and $\{b_n \in \mathbb{R}\}$ such that the normalised block maximum

$$M_n^* = \frac{M_n - b_n}{a_n}$$

has non-degenerate limiting distribution G as $n \to \infty$. It is shown by Leadbetter et al. (2012) that the family of distributions G such that $M_n^* \sim G$ in the limit as $n \to \infty$ is the family of 'max-stable' distributions, given in Definition 2.1.1 as by Coles (2001).

Definition 2.1.1 (Max-Stability)**.** *A distribution G is max-stable if, for every $n = 2, 3, \ldots$, there exist constants $A_n > 0$ and $B_n \in \mathbb{R}$ such that, for all $z \in \mathbb{R}$*

$$G^n(A_n z + B_n) = G(z).$$

That is, if $X \sim G$ and $M_n = \max\{X_1, \ldots, X_n\} \sim G_{M_n}$, then G and G_{M_n} are identical up to a change of scale and location.

Leadbetter et al. (2012) provide further details of the max-stable distributions, showing that any distribution that satisfies max-stability belongs to the family of distributions specified by Definition 2.1.2: the Generalised Extreme Value (GEV) family. This distribution family has three distinct types: the Fréchet, the negative Weibull, and the Gumbel, determined by the value of the shape parameter $\xi \in \mathbb{R}$. For a random variable Z with GEV distribution G_{GEV}, we write $Z \sim G_{\text{GEV}}(\mu, \sigma, \xi)$ for location $\mu \in \mathbb{R}$, scale $\sigma > 0$ and shape $\xi \in \mathbb{R}$.

This equivalence of max-stable distributions G and members G_{GEV} of the GEV

family was originally formalised by Jenkinson (1955), who identified that the three distributions shown by Gnedenko (1943) to satisfy max-stability could be combined into the GEV family as in Theorem 2.1.3. Following this, we can say the normalised block maxima M_n^* is distributed $M_n^* \sim \text{GEV}(\mu, \sigma, \xi)$ with cumulative distribution function (2.1.1), in the limit as $n \to \infty$. Specific examples of this convergence in distribution are given by Coles (2001), for particular choices of distribution F and corresponding values of normalising sequences $\{a_n > 0\}$ and $\{b_n \in \mathbb{R}\}$.

Definition 2.1.2 (Generalised extreme value (GEV) family of distributions)**.** *A probability distribution whose cumulative distribution function has the following general form*

$$G_{GEV}(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \tag{2.1.1}$$

on $\{z \in \mathbb{R} : 1 + \xi(z-\mu)/\sigma > 0)\}$*, for location* $-\infty < \mu < \infty$*, scale* $\sigma > 0$ *and shape* $-\infty < \xi < \infty$*, is said to belong to the GEV family of probability distributions. If* $\xi > 0$*, we have the Fréchet family, if* $\xi < 0$ *we have the negative Weibull family, and in the limiting case of* $\xi \to 0$ *we have the Gumbel family.*

Theorem 2.1.3 (Extremal Types Theorem)**.** *If there exist sequences of constants* $\{a_n > 0\}$ *and* $\{b_n \in \mathbb{R}\}$ *such that*

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq z\right) \to G(z) \quad \text{as } n \to \infty, \tag{2.1.2}$$

where G *is a non-degenerate distribution function, then* G *is a member of the GEV family. Here* F*, the distribution of continuous random variable* X*, is said to be in the domain of attraction of* G*.*

For a random variable $Z \sim G_{\text{GEV}}$, a widely employed measure of extremal behaviour is the return level associated with a return period $P = 1/p$, for corresponding exceedance probability $p \in [0, 1]$. The return level z_p is defined implicitly via

$\Pr(M_n > z_p) = p$, meaning $G_{\text{GEV}}(z_p) = 1 - p$, so z_p is the $(1-p)$th quantile of Z. By inverting the GEV distribution given in expression (2.1.1), z_p can be written as

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - y_p^{-\xi}\right], & \text{for } \xi \neq 0, \\ \mu - \sigma \log y_p, & \text{for } \xi = 0, \end{cases}$$

for $\mu, \xi \in \mathbb{R}$, $\sigma > 0$, where $y_p = -\log(1-p)$, see e.g., Coles (2001). Return levels provide an intuitive interpretation of the GEV model. For instance, consider $X \sim F$ with F in the domain of attraction of G_{GEV}, such that $M_n = \max\{X_1, \ldots, X_n\} \sim G_{\text{GEV}}$. If $P = 10$ and M_n is the yearly block maxima, then z_p is the level the annual maxima is expected to exceed, on average, once every 10 years.

2.1.2 Inference for the generalised extreme value distribution

Consider a sequence of $m \times n$ observations of the random variable $X \sim F$, from which are obtained m observations $M_{n,1}, \ldots, M_{n,m}$ of the (unnormalised) block maxima M_n. When F is unknown, knowledge of the normalising sequences $\{a_n > 0\}$ and $\{b_n \in \mathbb{R}\}$ is not required to perform inference on observations of the block maxima M_n. As argued by e.g., Coles (2001), if, for large n,

$$\Pr(M_n^* \leq z) = \Pr\left\{\left(\frac{M_n - b_n}{a_n}\right) \leq z\right\} \approx G_{\text{GEV}}(z),$$

for all $z \in \mathbb{R}$ for GEV family distribution G_{GEV}, then

$$\Pr(M_n \leq z) \approx G_{\text{GEV}}\left(\frac{z - b_n}{a_n}\right) = G_{\text{GEV}}^*(z), \quad \text{for all } z \in \mathbb{R},$$

from max-stability, where G_{GEV}^* is also a member of the GEV family of distributions. That is, the distribution G_{GEV}^* of M_n is approximately GEV, with different location μ and scale σ (but identical shape ξ) to the approximate distribution G_{GEV} of M_n^*.

The GEV distribution is thereby an appropriate model for the observed block maxima $M_{n,1}, \ldots, M_{n,m}$.

Inference is typically carried out my maximising the log-likelihood $\ell(\mu, \sigma, \xi; \mathbf{z})$ of the observed maxima $\mathbf{z} = (z_1, \ldots, z_m)$, over the parameter space $\mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}$. When $\xi \neq 0$, the log-likelihood is written

$$\begin{aligned} \ell(\mu, \sigma, \xi; \mathbf{z}) = &- m \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{m} \log \left[1 + \xi \left(\frac{z_i - \mu}{\sigma}\right)\right] \\ &- \sum_{i=1}^{m} \left[1 + \xi \left(\frac{z_i - \mu}{\sigma}\right)\right]^{-1/\xi}, \end{aligned} \tag{2.1.3}$$

provided that

$$1 + \xi \left(\frac{z_i - \mu}{\sigma}\right) > 0,$$

for $i = 1, \ldots, m$. When $\xi = 0$, the likelihood is

$$\ell(\mu, \sigma; \mathbf{z}) = -m \log \sigma - \sum_{i=1}^{m} \left(\frac{z_i - \mu}{\sigma}\right) - \sum_{i=1}^{m} \exp\left\{-\left(\frac{z_i - \mu}{\sigma}\right)\right\}. \tag{2.1.4}$$

No analytic maximum of the likelihood exists, but maximisation of (2.1.3) and (2.1.4) with respect to the parameter vector (μ, σ, ξ) using standard numerical techniques yields parameter estimates

$$(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = \underset{\mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}}{\operatorname{argmax}}\, \ell(\mu, \sigma, \xi; \mathbf{z}).$$

Care must be taken when numerically optimising in the space where ξ is close to zero, typically solved by using (2.1.4) for values of ξ within some small interval around zero (Coles, 2001).

Selection of the block size n must also be considered carefully when performing inference. Since the number nm of observations of X is fixed, increasing the value of n leads to a smaller number m observations of M_n over which to maximise the likelihood, resulting in increased parameter estimate variability. Conversely, decreasing

the value of n can invalidate the finite approximation of the limiting behaviour (2.1.2), causing biased inference. In practice, the fit of the GEV to data can be studied for various block sizes using standard techniques (such as QQ-plots or AIC) to select the best block size, see e.g., Coles (2001). In the case of non-IID X observations, this issue carries additionally complexity, since inappropriate choices of n lead to non-IID M_n and making (2.1.3) and (2.1.4) invalid. In practice, the selection of block size for non-IID data is informed by the application; for instance, taking yearly maxima can avoid issues arising from serial dependence, intra-year trends and seasonality in many environment applications.

2.1.3 Modelling threshold exceedances

Use of the block maxima approach described in Section 2.1.1 results in discarding any observation that is not a block maximum. This is undesirable primarily due to the loss of data, but also for additional reasons. Consider an IID sequence with a large variance, which we divide into blocks. If in one block we observe a large number of extreme realisations, only the maximum of these will be used considered in the likelihood when fitting the GEV distribution. Conversely, another block may have a comparatively small maximum, meaning we have discarded data larger than some which have been kept.

The alternative approach of Davison and Smith (1990) uses any observations that are considered large, avoiding the need to only retain block maxima. Consider a sequence of IID realisations of a random variable $X \sim F$, where X is in the domain of attraction of a GEV distribution G_{GEV} with parameters $\mu \in \mathbb{R}$, $\sigma > 0$ and $\xi \in \mathbb{R}$, relying on normalising sequences $\{a_n > 0\}$ and $\{b_n \in \mathbb{R}\}$. Building on the work of Pickands (1975), Davison and Smith (1990) introduce a modelling scheme for the exceedances of X above a threshold u_n as $n \to \infty$ and u_n tends to the upper endpoint x_+ of F.

Consider the distribution of exceedances above a threshold $u_n = a_n u + b_n$ via

$$\Pr\left(X > a_n z + u_n | X > u_n\right) = \Pr(X > a_n(z+u) + b_n | X > a_n u + b_n),$$

where $z > 0$, where u and $u + z$ are in the domain of attraction of G_{GEV}. As F is in the domain of attraction of G_{GEV}, we have, as $n \to \infty$,

$$\begin{aligned} F^n(a_n z + b_n) &\to G_{\text{GEV}}(z) \qquad (2.1.5)\\ \implies n \log F(a_n z + b_n) &\to \log G_{\text{GEV}}(z)\\ \implies n\{1 - F(a_n z + b_n)\} &\to -\log G_{\text{GEV}}(z) \end{aligned}$$

by the Taylor expansion of $\log F$ for $F \to 1$. Following this, we have, as $n \to \infty$,

$$\begin{aligned} \Pr(X > a_n(z+u) + b_n | X > a_n u + b_n) &= \frac{n(1 - F(a_n(z+u) + b_n))}{n(1 - F(a_n u + b_n))} \to \frac{\log G_{\text{GEV}}(z+u)}{\log G_{\text{GEV}}}\\ = \frac{\left[1 + \xi\left(\frac{z+u-\mu}{\sigma}\right)\right]_+^{-\frac{1}{\xi}}}{\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]_+^{-\frac{1}{\xi}}} &= \left[1 + \xi\left(\frac{z}{\sigma + \xi(u-\mu)}\right)\right]_+^{-\frac{1}{\xi}} = \left(1 + \xi\frac{z}{\sigma_u}\right)_+^{-\frac{1}{\xi}}, \qquad (2.1.6) \end{aligned}$$

where the limit holds from (2.1.5), with $[y]_+ = \max(y, 0)$ for $y \in \mathbb{R}$, and $\sigma_u = \sigma + \xi(u - \mu)$. If we assume the above limit holds approximately for some finite fixed threshold $u_n \in \mathbb{R}$, then the right hand side of expression (2.1.6) is equal to the survivor function of the excesses of X above this threshold. That is, we can write

$$\Pr\{(X > z + u) | (X > u)\} = \left(1 + \xi\frac{z}{\sigma_u}\right)_+^{-\frac{1}{\xi}},$$

for $z > 0$, where σ_u absorbs a_n. Definition 2.1.4 describes the resulting distribution function of $Z_u = (X - u)|(X > u)$, the generalised Pareto distribution (GPD). We write $Z_u \sim G_{\text{GPD}}(\sigma_u, \xi)$ for scale $\sigma_u > 0$ and shape $\xi \in \mathbb{R}$. The shape ξ is the same as that of the corresponding GEV, whereas the scale σ_u is different as it is dependent on

the exceedance threshold u, defined as $\sigma_u = \sigma + \xi(u - \mu)$ for GEV parameters μ and σ.

Definition 2.1.4 (The generalised Pareto distribution (GPD))**.** *A random variable Z_u with cumulative density function*

$$G_{GPD}(z) = \Pr(Z_u \leq z) = 1 - \left(1 + \xi \frac{z}{\sigma_u}\right)_+^{-\frac{1}{\xi}}, \tag{2.1.7}$$

on $z > 0$, with scale parameter $\sigma_u > 0$ and shape parameter $-\infty < \xi < \infty$, is said to have a generalised Pareto distribution. The random variable Z_u is termed a generalised Pareto (GP) random variable. Occasionally, an alternative parametrisation with $k = -\xi$ is given, for example in Davison and Smith (1990).

Return levels can be defined for a random variable $Z_u \sim G_{\text{GPD}}$. Take $X \sim F$ such that $Z_u = (X - u)|(X > u) \sim G_{\text{GPD}}$ for some $u > 0$. For large m, the mth-observation return level of X, i.e., the value that X exceeds on average every m observations, is written

$$x_m = \begin{cases} u + \frac{\sigma_u}{\xi}[(m\zeta_u)^\xi - 1], & \text{for } \xi \neq 0, \\ u + \sigma_u \log(m\zeta_u), & \text{for } \xi = 0, \end{cases}$$

for $\sigma_u > 0$, $\xi \in \mathbb{R}$, where $\zeta_u = 1 - F$, and a_n and b_n are absorbed into ζ_u. These 'per-observation' return levels can be converted to GPD equivalents of the 'per-time period' return levels discussed for the GEV in Section 2.1.1, see e.g, Coles (2001) for details.

2.1.4 Inference for the generalised Pareto distribution

Inference using the GPD first requires the selection of a finite exceedance threshold $u \in \mathbb{R}$, before observations of $Z_u = (X - u)|(X > u)$ can be obtained for modelling purposes. As with block size n in the GEV case, the choice of u is non-trivial. A larger value of u will help to ensure suitability of the GPD model assumptions, however, it will result in fewer observations for likelihood inference and therefore higher parameter

uncertainty. Conversely, a lower threshold will yield more observations for inference, but will risk invalidating the arguments leading to limit distribution (2.1.6). The selection of exceedance threshold u is therefore an important consideration when performing inference for the GPD.

The simplest approach for threshold selection is to assess the stability of estimates for the parameters $\tilde{\sigma} = \sigma_u - \xi u$ and ξ, estimated using observations of $(X-u)|(X>u)$ above a range of values of u; if a choice of u is reasonable, then estimates of $\tilde{\sigma}$ and ξ for any thresholds above u will not vary significantly after accounting for sampling uncertainty. See Coles (2001) for further justification of this method, as well as example applications.

The above threshold stability approach is often criticised for its subjectivity, since it is not always possible to tell for certain at which threshold the parameters become stable from visual inspection. See, e.g., the review of Scarrott and MacDonald (2012) for discussion on this point. Alternatively, automated and semi-automated methods have been developed. For instance, the probability weighted moments approach of Hosking et al. (1985). They define the rth probability weighted moment of a random variable $Z_u \sim G_{\text{GPD}}$, written

$$\alpha_r = E\left[Z_u\{1 - G_{\text{GPD}}(Z_u)\}^r\right] = \frac{\sigma_u}{(r+1)(r+1-\xi)}, \tag{2.1.8}$$

for a particular threshold choice u. This theoretical value (2.1.8) can be compared with empirical probability weighted moment values. If theoretical and empirical equivalents are the same up to sampling error (also given by Hosking et al. 1985) then the GPD is approximately valid for Z_u, and the value of u being considered is deemed suitable. If they are not the same up to sampling error, then the GPD approximation is invalid as the considered u value is too low, and a higher threshold must be chosen.

Another method is the graphical approach of mean residual life plots, described by Davison and Smith (1990). Let $u_0 \in \mathbb{R}$ be the smallest threshold for which the GPD is

a valid model for threshold exceedances, i.e., the smallest value for which the arguments leading to (2.1.4) hold. We can write, for $\xi < 1$ and $u > u_0$,

$$\mathbb{E}\{(X-u)|(X>u)\} = \frac{\sigma_u}{1-\xi} = \frac{\sigma_{u_0} + \xi(u-u_0)}{1-\xi}, \tag{2.1.9}$$

see Yang (1978) and Hall and Wellner (2020) for further details. Following (2.1.9), the mean residual life $\mathbb{E}(X-u|X>u)$ is linear with respect to $u > u_0$. Therefore, empirical sample means of $(X-u)|(X>u)$ for $u > u_0$ should be linear up to sampling error. Inspection of empirical mean residual life plots can thus be used to select an optimal value of u by determining where this linearity begins.

We also point to the recent works of Wadsworth (2016), Northrop et al. (2016), Varty et al. (2021) and Murphy et al. (2025) as examples of more novel methodologies, which use less subjective methods. In particular, we note the method of Murphy et al. (2025), who design a metric that quantifies the quality of fit of the GPD model with respect to the choice of threshold, enabling automatic threshold selection via optimisation of this metric across threshold candidates.

Given selection of an appropriate exceedance threshold u, the GPD log-likelihood for (σ_u, ξ) can be written as

$$\ell(\sigma_u, \xi; \mathbf{z}) = \begin{cases} -m\log\sigma_u - \left(\frac{1}{\xi}+1\right)\sum_{i=1}^{m}\log\left(1+\frac{\xi z_i}{\sigma_u}\right), & \text{for } \xi \neq 0, \\ -m\log\sigma_u - \frac{1}{\sigma_u}\sum_{i=1}^{m} z_i, & \text{for } \xi = 0. \end{cases} \tag{2.1.10}$$

for observations $\mathbf{z} = (z_1, \ldots, z_m)$ of $Z_u = (X-u)|(X>u)$, with $1 + \xi z_i/\sigma_u > 0$ for $i = 1, \ldots, m$. As with the GEV case, the GPD log-likelihood has no analytic maximum (except when $\xi = 0$); however, parameter estimates can be obtained by finding the maxima of likelihood (2.1.10) over the parameter space $\mathbb{R}^+ \times \mathbb{R}$ using standard numerical techniques.

2.1.5 Temporally dependent extremes

Up to now, we have discussed models for sequences of IID continuous random variables, that is, those with identical marginal distributions across observations (identically distributed) and with no inter-observation dependence (independence). The assumption of identical distributions is reasonable in many applications, see e.g, Coles (2001), and so we do not relax this assumption here. However, when considering the (typically environmental) data to which extreme value models are applied, the assumption of independence is often unreasonable. We therefore relax the assumption of independence in observations, seeking models for the extremal behaviour of a sequence $X_1, \ldots, X_n$ of temporally dependent replicates of $X \sim F$, with identical marginal distribution F, i.e., a stationary process.

Consider the stationary process $\{X_t\}_{t=1}^{\infty}$. In order to model the extremal behaviour of this sequence, some constraints must be imposed on the type of dependence exhibited by the sequence. In the context of extreme value modelling, the effect of long-range dependence is typically assumed to be limited in comparison to short term dependence, allowing the development of models which solely focus on the short-term interaction between observations. This assumption is formalised by the 'long-range weak dependence' property $D(u_n)$, defined in Definition 2.1.5, as given by Leadbetter et al. (2012). Property $D(u_n)$ is equivalent to near-independence for observations of X observed sufficiently far apart in time.

Definition 2.1.5. *A sequence of random variables* $\{X_1, \ldots, X_n\}$ *satisfies the condition* $D(u_n)$ *if, for any integers,*

$$1 \leq i_1 < \cdots < i_p < j_1 < \cdots < j_p \leq n$$

for which $j_1 - i_p \geq l$*, we have*

$$|F_{1,\ldots,i_p,j_1,\ldots,j_p}(u_n) - F_{1,\ldots,i_p}(u_n)F_{j_1,\ldots,j_p}(u_n)| \leq d_{n,l_n},$$

where $F_{X_{k_1},\ldots,X_{k_n}}$ *denotes the joint distribution of* $\{X_{k_1},\ldots,X_{k_n}\}$ *and* $d_{n,l_n} \to 0$ *as* $n \to \infty$ *for some sequence* $l_n = o(n)$*. The property* $D(u_n)$ *is sometimes referred to as the long-range weak dependence property.*

Consider a stationary sequence of random variables $X_1,\ldots,X_n$. If the $D(u_n)$ condition holds for some sequence u_n, for instance $u_n = a_n z + b_n$ as in limit (2.1.3), then the GEV is a valid model for the block maxima $M_n = \max\{X_1,\ldots,X_n\}$, where the form of this GEV distribution is dependent on the extent of temporal dependence in the data. The form of the GEV must change because extreme observations of a dependent sequence are more likely to occur in sequentially observed 'clusters' than those of an IID sequence; how the GEV form changes due to this dependence is summarised in Theorem 2.1.6.

Theorem 2.1.6. *Let* $X_1, X_2, \ldots$ *be a stationary sequence of dependent variables and* $X_1^I, X_2^I, \ldots$ *a sequence of independent variables with the same marginal distribution as* X_1*. Define* $M_n = \max\{X_1,\ldots,X_n\}$ *and* $M_n^I = \max\{X_1^I,\ldots,X_n^I\}$*. Under the regularity conditions specified in Theorem 2.1.3, if*

$$\Pr\{(M_n^I - b_n/a_n) \leq z\} \to G_1(z),$$

for all $z \in \mathbb{R}$*, as* $n \to \infty$ *for normalising sequences* $\{a_n > 0\}$ *and* $\{b_n\}$*, where* G_1 *is a non-degenerate distribution function, and if the long-range weak dependence property* $D(u_n)$ *of Definition 2.1.5 is satisfied, then*

$$\Pr\{(M_n - b_n)/a_n \leq z\} \to G_2(z),$$

for all $z \in \mathbb{R}$*, where*

$$G_2(z) = G_1^{\theta}(z),$$

for a constant $0 < \theta \leq 1$.

The constant θ is termed the extremal index. The probability of observing a randomly selected threshold exceedance remains the same between the stationary and independent sequences, since the marginal distribution F is unchanged between them. When $\{X_t\}_{t=1}^{\infty}$ exhibits positive temporal dependence, there are a smaller number of independent threshold exceedances in the stationary sequence than in the independent sequence, meaning observed maxima tend to be smaller in the dependent sequence. Formally, the effect of temporal dependence on the GEV distribution, i.e., the extremal index θ, is determined by the mean cluster size over $u_n = a_n z + b_n$ in the limit as the threshold $u_n \to x_+$. Consider a sequence $X_1, \ldots, X_{p_n}$ of observations of the stationary process, for $p_n = o(n)$. Let $N_{p_n}(u_n)$ be the number of observations of that sequence to exceed the threshold u_n. A cluster is formally defined as the set $\{X_i > u_n : i = 1, \ldots, p_n\}$, provided that $N_{p_n}(u_n) \geq 1$. In this case, $N_{p_n}(u_n)$ denotes the cluster size, with limiting probability mass function

$$\pi(m) = \lim_{n \to \infty} \Pr\left(N_{p_n}(u_n) = m | N_{p_n}(u_n) \geq 1\right),$$

for $m \in \mathbb{N}$, see Hsing et al. (1988). From this, Leadbetter et al. (2012) show that the extremal index is $\theta = \{\sum_{m=1}^{\infty} m\pi(m)\}^{-1}$, under mild regularity conditions. Hence θ is the reciprocal of the limiting mean cluster size. As a result of Theorem 2.1.6, if a dependent sequence satisfies condition $D(u_n)$, then its block maxima will be distributed according to a GEV with location $\mu^* = \mu - \sigma(1 - \theta^{\xi})/\xi$, scale $\sigma^* = \sigma\theta^{\xi}$, and shape ξ, where μ, σ, and ξ are the respective location, scale and shape parameters of the GEV distribution for the block maxima of the equivalent independent sequence.

The method of identification of clusters to be used during inference is a non-trivial

problem, with solutions offered by e.g., Davison and Smith (1990), Smith and Weissman (1994) and Ferro and Segers (2003). Specifically, the approach of Smith and Weissman (1994), which defines a cluster as a sequence of observations separated by at most a specified number of consecutive observations below the threshold. Consider a stationary sequence $X_1, \ldots, X_n$, and define

$$W_i = \begin{cases} 1 & X_i > u, \\ 0 & X_i \leq u, \end{cases}$$

$i = 1, \ldots, n$, for indicating exceedances for threshold $u \in \mathbb{R}$. Also define $N_n = \sum_{i=1}^{n} W_i$, the number of exceedances of u observed in the sequence. The random variable

$$Z_n = \sum_{i=1}^{n-l} \left\{ W_i \prod_{j=1}^{l} l(1 - W_{i+j}) \right\}$$

is the number of times the last observation in a cluster is followed by $l \in \mathbb{N}$ consecutive non-exceedances. Intuitively, this is a measure of the number of observed clusters, as a 'run' of l non-exceedance observations indicates a gap between two adjacent clusters. The corresponding estimate of the extremal index based on threshold u and this cluster identification method is thereby given as

$$\hat{\theta}_u = \frac{Z_n}{N_n}. \tag{2.1.11}$$

The value l is referred to as the run length, and must be chosen prior to estimation of the extremal index via (2.1.11). In particular, Ferro and Segers (2003) give an automated selection method.

The dependence in the sequence means likelihood inference cannot be carried out using expression (2.1.10), since this relies on IID observations for the threshold exceedances, which is not the case for exceedances within the same cluster. To circum-

vent this issue, Davison and Smith (1990) suggest using the GPD to model the observed maxima of each cluster (as defined above) of exceedances above a threshold. Leadbetter et al. (2012) provide further justification for this approach, relating the point process behaviour of threshold exceedances of the independent sequence to the behaviour of cluster exceedances in the dependence sequence via the extremal index θ.

Fawcett and Walshaw (2007) note that the asymptotics used to justify the GPD model to cluster maxima may not be valid at finite values of threshold u, leading to bias in parameter estimates when the GPD is used for inference. In response, Eastoe and Tawn (2012) propose a sub-asymptotic extension of the GPD (2.1.7) for use in modelling the observed cluster maxima variable of threshold u, denoted Z_u^c. The model takes the form

$$\Pr\{Z_u^c > x + u | Z_u^c > u\} = 1 - \frac{\theta(x+u,l)}{\theta(u,l)} \left(1 + \frac{\xi x}{\sigma_u}\right)_+^{-1/\xi},$$

for $x > 0$, where

$$\theta(x,l) = \Pr\left\{\max\left(X_2, \ldots, X_l\right) < x \mid X_1 > x\right\},$$

is the sub-asymptotic extremal index (Ledford and Tawn, 2003) for a stationary sequence $X_1, \ldots, X_l$ of $X \sim F$. Here l is the run length used to detect temporally-dependent clusters of observations using the method of Ferro and Segers (2003).

2.1.6 Independent and not identically distributed extremes

All models discussed so far require each observation in the process that is being modelled to have identically marginal distribution functions. This is often not the case in environmental applications, for instance individual ocean storm intensities may vary seasonally (Jonathan et al., 2013), or annual maximum temperatures may show a positive trend due to climate change (Brown, 2020). These settings motivate the modelling

of random variables dependent on external covariates, e.g., seasons and time in the two examples above. In the general, non-extreme case, consider the following example. Take a process $X_t \sim F_{X_t}$ with associated p-dimensional covariate $\mathbf{Z}_t \sim F_{\mathbf{Z}_t}$ evolving deterministically, or with known distribution $F_{\mathbf{Z}_t}$ in time $t \in \mathcal{T}$. To fully understand the process X_t, we require knowledge of the distribution $F_{X_t|\mathbf{Z}_t}$ for $X_t|\mathbf{Z}_t$, given which we can write

$$F_{X_t}(x) = \int_{\mathcal{Z}} F_{X_t|\mathbf{Z}_t}(x|\mathbf{z}_t) f_{\mathbf{Z}_t}(\mathbf{z}_t) \mathrm{d}\mathbf{z}_t,$$

where $f_{\mathbf{Z}_t}$ is the density function of $\mathbf{Z}_t$ on $\mathcal{Z}$. In the case of extremes, we seek distributions for either block maxima or threshold exceedances, conditional on the values of appropriate covariates. Approaches for the both block maxima and threshold exceedance frameworks have been developed, however, the latter is more widely applied. We therefore focus on the threshold exceedance case in this review, referring to Coles (2001) and Parey et al. (2013) as examples of covariate dependent block maxima modelling.

Take a random variable X_t, for which we wish to model the exceedances $(X_t - u)|(X_t > u)$ via a GPD distribution with covariate $\mathbf{Z}_t = \mathbf{z}_t$. A typical approach is to model the parameters $\sigma_u > 0$ and $\xi \in \mathbb{R}$ of the GPD distribution according to the regression

$$\log \sigma_u(\mathbf{z}_t) = \mathbf{z}_t^{\mathrm{T}} \boldsymbol{\alpha} \text{ and } \xi(\mathbf{z}) = \mathbf{z}_t^{\mathrm{T}} \boldsymbol{\beta}, \tag{2.1.12}$$

for regression coefficients $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^p$, where the log transform is used to ensure positivity of the scale parameter $\sigma_u(\mathbf{z}_t)$. See e.g., Davison and Smith (1990) and Coles (2001) for examples. Eastoe and Tawn (2008) point out that care must be taken when applying the log transform to σ_u, since this invalidates the ideas behind threshold stability discussed in Section 2.1.3; instead they suggest $\sigma_u(\mathbf{z}_t) = \mathbf{z}_t^{\mathrm{T}} \boldsymbol{\alpha}$, with care taken in choosing the range of $\boldsymbol{\alpha}$ to ensure that $\sigma_u(\mathbf{z}_t) > 0$ for all possible $\mathbf{z}_t \in \mathcal{Z}$. In practice, the shape parameter ξ is often kept to be an unknown constant as it is difficult to estimate when

assumed to be dependent on covariates; note that when fixing the shape parameter in this way, care must first be taken to ensure it is a reasonable assumption, e.g., using pooled QQ plots (Heffernan and Tawn, 2003). It is also possible to model the exceedance rate of threshold u, i.e., $\lambda(u|\mathbf{z}_t) = \Pr(X_t > u|\mathbf{Z}_t = \mathbf{z}_t)$, as a function of covariates $\mathbf{z}_t$. For instance, we may write

$$\text{logit}\lambda(u|\mathbf{z}_t) = \mathbf{z}_t^{\mathrm{T}}\boldsymbol{\gamma},$$

for regression coefficient $\boldsymbol{\gamma} \in \mathbb{R}^p$, where the logit-link function is used to ensure exceedance probability $\lambda(u|\mathbf{z}_t)$ lies on the unit interval [0, 1] for all $\mathbf{z}_t \in \mathbb{R}^p$.

Variations on the covariate formulation (2.1.12) have been developed for particular applications. For instance, Coles and Tawn (1994) model covariates for in temperature data using harmonics to capture seasonal trends, whilst Jonathan et al. (2013) regress GPD parameters on Fourier series for periodic covariates related to extreme sea states. More recently, a generalisation of (2.1.12) using generalised additive models (GAMs) has become popular. GAMs extend the linear predictors of (2.1.12) to include sums of smooth functions over covariates, allowing for a wider class of covariate behaviour. The use of GAMs for extreme value covariate modelling was original proposed by Chavez-Demoulin and Davison (2005) and then popularised by Youngman (2019a) as the extreme value generalised additive model (EVGAM) framework. It has since been applied by the likes of Brown (2020) in modelling heat waves occurrences, and Richards et al. (2023) in modelling precipitation.

2.2 Multivariate extreme value theory

2.2.1 Defining extreme events

The methods of Section 2.1 can be extended to the multivariate setting when we need to understand the joint extremal behaviour of more than one variable. This is useful in

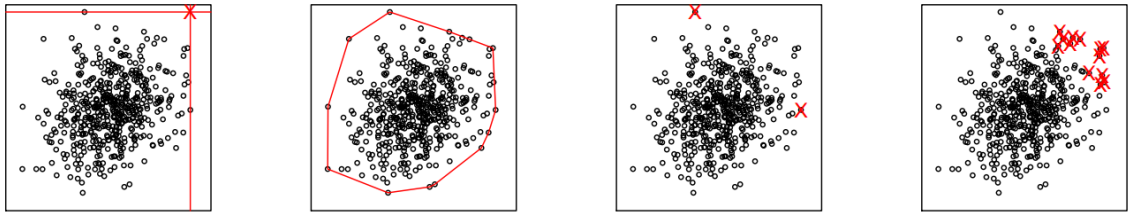


Figure 2.2.1: Illustration of various ways to order extremes (presented by Tawn, 2022).

certain applications; for instance, an offshore structure may be able to withstand severe wind and wave conditions separately, but joint extreme events in both variables may induce failure. It is therefore important to understand not only the extremal behaviour of the wind and wave conditions marginally, but also how these variables interact in their extremes.

The first point to consider is what should be referred to as a 'multivariate extreme event'. Unlike the univariate case, there is no single natural interpretation of the extreme of a set of multivariate observations. Barnett (1976) discusses this point in detail, providing four possible orderings for multivariate events, including: (a) the componentwise maxima of a set of observations; (b) a convex hull surrounding the set, with observations on the hull boundary taken as extreme; (c) observations which are large in a least one component (pictured in the third panel of Figure 2.2.1 as the points where one variable is its largest value); and (d) a subregion of a set that is extreme in the context of a particular application. Cases (a)-(d) are each illustrated in Figure 2.2.1 from left to right. Models for (a) and (c) are discussed in Section 2.2.2 and Section 2.2.5 respectively. Methods describing equivalents of (b) and (d) are covered in context for applications discussed in this thesis, see Chapters 3 and 4 for respective examples.

2.2.2 Models for multivariate block maxima

We begin by discussing models for componentwise maxima, the multivariate extension of the block maxima setting discussed in Section 2.1.1. As with block maxima in the

univariate case, initial development of multivariate extreme value theory was based on this ordering, seeking to extend the GEV to a multivariate distribution capable of modelling componentwise maxima. Much of this development was carried out by the likes of de Haan and Resnick (1977), Pickands (1981) and Resnick (2008), amongst others. Here, we give the multivariate extension of the GEV as in Coles and Tawn (1991), and then go on to describe parametric models such as those given by Gumbel (1960), Tawn (1988), Tawn (1990), Coles and Tawn (1991), Naveau et al. (2009) and Huser and Davison (2014).

Take a d-dimensional random vector $\mathbf{X} \sim F_d$ with unknown multivariate distribution function F_d. Consider a sequence of n independent observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ of $\mathbf{X}$, where $\mathbf{X}_i = (X_{1,i}, \ldots, X_{d,i})$ for $i = 1, \ldots, n$. Define the componentwise maximum $\mathbf{M}_n = \{M_{1,n}, \ldots, M_{d,n}\}$, where $M_{j,n} = \max(X_{j,1}, \ldots, X_{j,n})$ for $j = 1, \ldots, d$ (as illustrated in the first panel of Figure 2.2.1). Extending from the case of univariate block maxima discussed in Section 2.1.1, we assume the existence of normalising sequences $\{a_{j,n} > 0\}$ and $\{b_{j,n} \in \mathbb{R}\}$, $j = 1, \ldots, d$, where

$$\Pr\left(\frac{M_{1,n} - b_{1,n}}{a_{1,n}} \leq z_1, \ldots, \frac{M_{d,n} - b_{d,n}}{a_{d,n}} \leq z_d\right) \to G_d(z_1, \ldots, z_d), \tag{2.2.13}$$

as $n \to \infty$, for $(z_1, \ldots, z_d) \in \mathbb{R}^d$, such that G_d is non-degenerate in each margin. The distribution function G_d, termed the multivariate extreme value distribution, can be viewed as the multivariate extension of the GEV distribution (2.1.1) for univariate block maxima. As with the GEV, G_d must be max-stable (see e.g., Geffroy 1958) meaning, for every $n = 2, 3, \ldots$, there exist constants $\mathbf{A}_n = (A_{1,n}, \ldots, A_{d,n}) \in \mathbb{R}^d_+$ and $\mathbf{B}_n = (B_{1,n}, \ldots, B_{d,n}) \in \mathbb{R}^d$ such that, for all $\mathbf{z} = (z_1, \ldots, z_d) \in \mathbb{R}^d$,

$$G_d^n(\mathbf{A}_n \mathbf{z} + \mathbf{B}_n) = G_d(\mathbf{z}). \tag{2.2.14}$$

This is the multivariate extension of the univariate max-stability property given in Def-

inition 2.1.1, and implies that the margins of G_d are each themselves GEV distributions (see Tawn 1988 for further discussion). In the case of standard Fréchet margins for G_d, with marginal distribution function $\exp(-1/z)$ for $z > 0$, setting $\mathbf{A}_n = (n, \ldots, n)$ and $\mathbf{B}_n = (0, \ldots, 0)$ satisfies expression (2.2.14).

Given this knowledge of the marginal behaviour of G_d, we now require information on the dependence structure. de Haan and Resnick (1977) show that G_d may be written

$$G_d(\mathbf{z}) = \exp(-V(\mathbf{z})), \quad \text{for } \mathbf{z} \in \mathbb{R}^d_+, \tag{2.2.15}$$

when G_d has Fréchet margins as discussed above, for some function $V(\mathbf{z})$ on $\mathbb{R}^d_+ \mapsto \mathbb{R}$, which is homogeneous of order -1; the function $V(\mathbf{z})$ is termed the exponent measure by de Haan and Resnick (1977). It is also shown by Pickands (1981) that the exponent measure $V(\mathbf{z})$ can be written as

$$V(\mathbf{z}) = \int_{S_d} \max_{1 \le j \le d} \left(\frac{w_j}{z_j} \right) \mathrm{d}H(\mathbf{w}), \tag{2.2.16}$$

where $\mathbf{w} = (w_1, \ldots, w_d)$, for a positive finite measure H defined on the unit simplex

$$S_d = \left\{ (w_1, \ldots, w_d) : \sum_{j=1}^{d} w_j = 1, \ w_j \ge 0 \ j = 1, \ldots, d \right\},$$

with H satisfying the condition

$$\int_{S_d} w_j \mathrm{d}H(\mathbf{w}) = 1, \quad j = 1, \ldots, d.$$

Where H has mass on the unit simplex S_d (i.e., it's interior, boundaries and vertices) can be used to deduce the full extremal dependence structure of $\mathbf{X} = (X_j; j \in D = \{1, \ldots, d\})$. Denote by h_l the mass that H places on the subregion of S_d corresponding to a subset $(X_j; i \in l)$ for $l \subseteq D$ being large but $(X_j; j \in D \setminus l)$ all being small.

For instance, when $l = D$, h_l represents how the mass is placed in the interior of S_d, whereas when $l = \{j\}$ for a single $j \in D$, h_l is the mass placed at the vertex of S_d where $w_j = 1$. When H places a point mass at the centre of S_d, identical extreme values of *all* components of $\mathbf{X}$ can occur simultaneously, and if all the mass in S_d is at this centre then the componentwise maxima are jointly identical, i.e., perfectly dependent. Conversely, if $h_l = 0$ for all $\{l \subseteq D; |l| > 1\}$, i.e., when H places point masses at each of the d vertices of S_d, but not at the edges or the interior of S_d, the components of $\mathbf{X}$ cannot be large together, and this corresponds to the componentwise maxima variables being independent. In the intermediate case, if H places mass on an edge of S_d corresponding to some strict subset $l \subset D$, i.e., $h_l > 0$ where $|l| < d$, then it is possible for extremes of elements of l to occur together whilst elements of $(X_j; j \in D \setminus l)$ are not extreme.

Any distribution function G_d that has unit Fréchet margins and also satisfies expression (2.2.13) is of the form (2.2.15), see e.g., Pickands (1981). Additionally, any multivariate extreme value distribution with non-Fréchet GEV margins can be aligned with (2.2.15) by transformation of the margins to unit Fréchet, using the probability integral transform componentwise.

For inference purposes, there has also been development of non-parametric representations for the dependence structure of the multivariate extreme value distribution estimating either V or H, see e.g., Pickands (1981), Smith et al. (1990), Capéraà et al. (1997), Hall and Tajvidi (2000), Guillotte et al. (2011) and Cormier et al. (2014). Smith et al. (1990) compare the choice of parametric vs non-parametric models for the dependence structure of G_d, concluding that parametric forms are preferable for small d. For this reason, we opt to discuss only parametric distribution functions which are specific cases of (2.2.15), such as those of Gumbel (1960), Tawn (1988), Joe (1990) and Coles and Tawn (1991). We discuss two examples from Gumbel (1960) and Tawn (1990) below.

The logistic dependence structure for the multivariate extreme value distribution

function, originally defined by Gumbel (1960), is given in Definition 2.2.1, for d-dimensional random variables. This is the earliest form of parametric model for the multivariate extreme value distribution, and relies on symmetry in the dependence between components of $\mathbf{X}$, i.e., the probability of observing X_j and X_k large together must be equal to the probability of observing $X_{j'}$ and $X_{k'}$ large together, for $j \neq j'$ and $k \neq k$. This model has all the mass of H in the interior of S_d. Tawn (1990) extended the logistic form to the asymmetric logistic distribution, see Definition 2.2.2. In its most general form it has mass on the interior of S_d and on each of the edges and vertices of S_d. The asymmetric logistic distribution reduces to the logistic distribution when $\alpha_{c_d} = \alpha$ for all $c_d \in C_d$, $\theta_{i,c_d} = 1$ for $\{c_d : |c_d| = d\}$, $i = 1, \ldots, d$ and $\theta_{i,c_p} = 0$ for $\{c_d : |c_d| \neq d\}$, $i = 1, \ldots, d$.

Definition 2.2.1 (Logistic Distribution)**.** *The general form of the d-dimensional logistic distribution function with unit Fréchet margins, as given in Tawn (1990), is*

$$G_{Log}(\mathbf{z}) = \exp\left\{-\left(\sum_{i=1}^{d} z_i^{-1/\alpha}\right)^{\alpha}\right\},$$

for $\mathbf{z} = (z_1, \ldots, z_d) \in \mathbb{R}_+^d$, *with* $0 < \alpha \leq 1$. *This is a multivariate extreme value distribution function, as defined by expression* (2.2.14). *A random variable* $\mathbf{Z}$ *with distribution function* G_{Log} *is said to be a multivariate logistic random variable. If* $\alpha = 1$ *we have complete independence of the components of* $\mathbf{Z}$, *whereas, in the limit as* $\alpha \to 0$ *we have perfect dependence of components of* $\mathbf{Z}$.

Definition 2.2.2 (Asymmetric Logistic Distribution)**.** *The general form of the d-dimensional asymmetric logistic distribution function with unit Fréchet margins, as defined by Coles and Tawn (1991), is*

$$G_{AsmLog}(\mathbf{z}) = \exp\left[-\sum_{c \in C}\left\{\sum_{i \in c}\left(\frac{\theta_{i,c}}{z_i}\right)^{1/\alpha_c}\right\}^{\alpha_c}\right],$$

on $\mathbb{R}_+^d$*, where* C_d *is the set of all non-empty subsets of* $\{1, \ldots, d\}$*,* $0 < \alpha_{c_d} \leq 1$ *for* $c_d \in C_d$*,* $\theta_{i,c} = 0$ *if* $i \in c_d$*,* $\theta_{i,c_d} \geq 0$*,* $i = 1, \ldots, d$ *and* $\sum_{c_d \in C_d} \theta_{i,c_d} = 1$*. This is a multivariate extreme value distribution, as defined by expression* (2.2.14)*. A random variable* $\mathbf{Z}$ *with distribution function* G_{AsmLog} *is said to be an multivariate asymmetric logistic random variable.*

Consider $m \in \mathbb{N}$ observations of componentwise maxima $\mathbf{M}_n$ of blocks of size $n \in \mathbb{N}$, analogous to the case of univariate block maxima in Section 2.1.1. Inference under the above models typically involves: fitting of d marginal GEV distributions to the observed componentwise block maxima; followed by transformation of the block maxima to the unit Fréchet marginal scale using these marginal fits; and finishing with the fitting of a single joint parametric form for the multivariate extreme value distribution, e.g., the logistic or asymmetric logistic, to the unit Fréchet marginal componentwise block maxima. The choice of block size n has similar implications to those in the univariate case, where choosing n too small can violate the assumption of the limiting behaviour in (2.2.13) holding at finite levels and thus induce bias, whereas choosing block size n too large will result in too few observations and thus increase the variance of parameter estimates. As in the univariate case, n is often chosen according to the application.

Given a value for the block size n, the d GEV marginal distributions can be fitted by maximising the likelihood (2.1.3), replacing evaluations with those of likelihood (2.1.4) when ξ is within some small interval around zero. The resulting maximum likelihood estimates for GEV location $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)$, scale $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_d)$ and shape $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_d)$ parameters can then be used in the transformation

$$Z_j(M_{n,j}) = \left[1 + \xi \left(\frac{M_{n,j} - \mu_j}{\sigma_j}\right)^{-1/\xi_j}\right]^{-1},$$

for $j = 1, \ldots, d$, yielding IID vector random variables $\mathbf{Z}_i = (Z_{1,i}, \ldots, Z_{d,i})$ for $i = 1, \ldots, m$, with dependence structure equal to that of $\mathbf{M}_n = (M_{n,1}, \ldots, M_{n,d})$ and with

IID unit Fréchet marginals. Finally, parameter values for the chosen form for the extremal dependence structure, such as those given in Definitions 2.2.1 and 2.2.2, can be estimated using maximum likelihood. For example, let $\boldsymbol{\theta}$ be the model parameters for parametric forms of distribution function $G_d(\mathbf{z};\boldsymbol{\theta})$ defined by expression (2.2.15), such as α in the case of the logistic dependence structure given Definition 2.2.1. The parameter $\boldsymbol{\theta}$ can estimated by maximisation of the appropriate log-likelihood, written here in the general form as

$$\ell(\boldsymbol{\theta};\{\mathbf{z}_i\}_{i=1}^m) = \sum_{i=1}^m \log \frac{\partial^d G_d(\mathbf{z}_i;\boldsymbol{\theta})}{\partial z_1 \ldots \partial z_d} = \sum_{i=1}^m \log \left\{ \frac{\partial^d}{\partial z_1 \ldots \partial z_d} \exp(-V(\mathbf{z}_i)) \right\}, \qquad (2.2.17)$$

for exponent measure V and observed componentwise block maxima $\mathbf{z}_1, \ldots, \mathbf{z}_m$, transformed to unit Fréchet margins by transformation (2.2.17).

2.2.3 Point process model

The modelling and inference of Section 2.2.2 relies on the first ordering of Barnett (1976), componentwise block maxima. This can result in a loss of useful data, analogous to that in the case of univariate block maxima, whilst additionally relying on the subjective choice of componentwise block maxima for multivariate ordering. Additionally, these models often cannot be used to explain the joint behaviour of actual observations, as these are unlikely to correspond to the componentwise maxima since individual variable maxima may not occur concurrently. Instead, Coles and Tawn (1991) provide an alternative model for all events where at least one component of a multivariate random variable is suitably large, analogous to the threshold model of Section 2.1.3.

Formally, Coles and Tawn (1991) consider the probability of observing realisations of the d-dimensional random variable $\mathbf{X} \sim F_d$, with unit Fréchet margins. Consider the sequence $\{n^{-1}\mathbf{X}_i\}_{i=1}^n$ of $n \in \mathbb{N}$ normalised IID observations of $\mathbf{X}$. It is shown by

de Haan (1985) that, as $n \to \infty$, the process of points $\{n^{-1}\mathbf{X}_i\}_{i=1}^n$ converges to a non-homogeneous Poisson process, on any set $A \subset \mathbb{R}_+^d \setminus \{\mathbf{0}_d\}$, with intensity measure

$$\mu_I(\mathrm{d}r \times \mathrm{d}\mathbf{w}) = \frac{\mathrm{d}r}{r^2}\mathrm{d}H(\mathbf{w}), \tag{2.2.18}$$

for angular-radial coordinates

$$w_j = \frac{x_j}{\sum_{i=1}^d x_i} \quad \text{and} \quad r = \sum_{i=1}^d x_i, \quad j = 1, \ldots, d, \tag{2.2.19}$$

for $(x_1, \ldots, x_d) \in \mathbb{R}_+^d$, and positive finite measure H given in (2.2.16). Coles and Tawn (1991) leverage this property by arguing that, for large enough n, the points in $\{n^{-1}\mathbf{X}_i\}_{i=1}^n$ which lie in the set

$$A_u = \left\{ \mathbf{x} \in \mathbb{R}_+^d; \sum_{i=1}^d x_i > u \right\}, \tag{2.2.20}$$

for finite $u > 0$ and $\mathbf{x} = (x_1, \ldots, x_d)$, behave approximately as a Poisson process P_{μ_I} with intensity μ_I of expression (2.2.18). The summation in (2.2.20) defines the set A_u as the region in $\mathbb{R}_+^d$ where the radial coordinate in (2.2.19) exceeds the threshold u. The distance of A_u from $\mathbf{0}_d$ (i.e., the value of u) is important for this argument to be valid, and the required minimum distance is determined by the rate of convergence of $\{n^{-1}\mathbf{X}_i\}_{i=1}^n$ to P_{μ_I}, see Coles and Tawn (1991) for details. Selection of the threshold u is considered by Wan and Davis (2019), who develop an automated test for the limiting Poisson process behaviour for a given candidate u. See Section 2.2.6 for a summary of their approach. In Chapter 5, we develop a novel threshold selection method for the multivariate extreme value model discussed in Section 2.2.5, drawing inspiration from their work.

Notably, expression (2.2.18) shows that, as $n \to \infty$, the behaviour of the $(d-1)$-dimensional angular coordinate variable $\mathbf{W}$ is independent of the radial coordinate R

(which has a known Pareto distribution), and is described entirely by the measure H. Since the angle $\mathbf{W}$ determines the relative strength of components of $\mathbf{X}$ between one another, the measure H completely describes the dependence between the extreme components of the random variable $\mathbf{X}$ as $n \to \infty$, as discussed in Section 2.2.2.

Expression (2.2.18) also reveals how to estimate the probability of variable $\mathbf{X}$ lying in a set beyond the range of observed data. Consider the set $B_u \subseteq nA_u$ for A_u with sufficiently large u, and define the scaled set tB_u for $t \geq 1$. We wish to find the relationship between the probabilities $\Pr(\mathbf{X} \in tB_u)$ of $\mathbf{X}$ and $\Pr(\mathbf{X} \in B_u)$, for $t \geq 1$. These probabilities can be related by studying the integrated intensity function Λ of the Poisson process P_{μ_I} over both of these sets, i.e., $\Lambda(tB_u)$ and $\Lambda(B_u)$. The integrated intensity function Λ for P_{μ_I} over tB_u can be written as

$$\begin{aligned}\Lambda(tB_u) &= \int_{tB_u} \frac{\mathrm{d}r}{r^2}\mathrm{d}H(\mathbf{w}) = \int_{B_u} \frac{t\mathrm{d}s}{(ts)^2}\mathrm{d}H(\mathbf{w}) \\ &= \frac{1}{t}\int_{B_u} \frac{\mathrm{d}s}{s^2}\mathrm{d}H(\mathbf{w}) = \frac{1}{t}\Lambda(B_u),\end{aligned}$$

for $t \geq 1$, where the substitution $r = st$ is made. We can thus write

$$\Pr(\mathbf{X} \in tB_u) = \frac{1}{t}\Pr(\mathbf{X} \in B_u), \tag{2.2.21}$$

for all ≥ 1. Property (2.2.21) is useful when estimating extremal set probabilities. For instance, if we have observed values of $\mathbf{X}$ lying in and around B_u and have subsequently estimated $\Pr(\mathbf{X} \in B_u)$, say empirically, we can then use probability extrapolation (2.2.21) to provide an estimate for $\Pr(\mathbf{X} \in tB_u)$ for a set tB_u without any observations in this set, provided that u is large enough for the limiting point process assumption to be valid. Note that the decay probability of sets along the rays mapping B_u to tB_u, in Fréchet margins, is only affected by a change in the radial variable r (by multiplication by t), with no change in $\mathbf{w}$. This is as we would expect since $\mathbf{w}$ is fixed along these rays for all elements of B_u.

The point process representation of Coles and Tawn (1991) can be reconciled with the componentwise maxima formulation and consequent multivariate extreme value distributions of Section 2.2.2. That is, we can recover expression (2.2.16) for the form of multivariate extreme value distribution G_d. Consider the normalised componentwise block maxima $n^{-1}\mathbf{M}_n$ of $\{n^{-1}\mathbf{X}_i\}_{i=1}^n$, and the set $A_M = \mathbb{R}_+^d \setminus \{(0, z_1) \times \ldots \times (0, z_d)\}$ for $\mathbf{z} = (z_1, \ldots, z_d) \in \mathbb{R}_+^d$. Coles and Tawn (1991) show that, as $n \to \infty$,

$$\Pr\left(n^{-1}\mathbf{M}_n \leq \mathbf{z}\right) = \Pr\left(n^{-1}\mathbf{X}_i \notin A_M, \forall i = 1, \ldots, n\right) \to \exp(-\Lambda(A_M)),$$

for

$$\Lambda(A_M) = \int_{A_M} \frac{\mathrm{d}r}{r^2}\mathrm{d}H(\mathbf{w}) = \int_{S_d}\int_{r=\min_j(z_j/w_j)}^{\infty} \frac{\mathrm{d}r}{r^2}\mathrm{d}H(\mathbf{w}) = \int_{S_d} \max_{1\leq j\leq d}\left(\frac{w_j}{z_j}\right)\mathrm{d}H(\mathbf{w}).$$

We can then write $G_d(\mathbf{z}) = \lim_{n\to\infty}\Pr\left(n^{-1}\mathbf{M}_n \leq \mathbf{z}\right) = \exp(-V(\mathbf{z}))$, yielding expression (2.2.16) as required, as $V(\mathbf{z}) = \Lambda(A_M)$.

Inference using the point process approach to fit models of the form (2.2.15), such as the logistic and asymmetric logistic distributions, is typically carried out in a two-stage process. Consider a random variable $\mathbf{Y} = (Y_1, \ldots, Y_d) \in \mathbb{R}^d$ with unknown marginal distributions F_{Y_j} such that $Y_j \sim F_{Y_j}$, $j = 1, \ldots, d$, of which we have observed values $\{\mathbf{y}_i\}_{i=1}^n$. Firstly, observations of the random variable $\mathbf{Y}$ must be transformed to observations $\{\mathbf{x}_i\}_{i=1}^n$ of the Fréchet marginal random variable $\mathbf{X} = (X_1, \ldots, X_d) \in \mathbb{R}^d$ with the same dependence structure as $\mathbf{Y}$. This can be achieved using the probability integral transform along with estimates $\hat{F}_{Y_j}$ of the marginal distributions F_{Y_j}, $j = 1, \ldots, d$. See, for instance, Coles and Tawn (1991) who use the generalised Pareto distribution (GPD) of Section 2.1.3 to model exceedances of Y_j above some threshold u_j, $j = 1, \ldots, d$, and

the empirical distribution below u_j, leading to the probability integral transform

$$X_j(Y_j) = \begin{cases} -\left(\log\left[1-\zeta_{u,j}\{1+\xi_j(Y_j-u_j)/\sigma_{u,j}\}^{-1/\xi_j}\right]\right)^{-1} & \text{for } Y_j > u_j, \\ -\log[\{R(Y_j)/(n+1)\}]^{-1} & \text{for } Y_j \leq u_j. \end{cases} \tag{2.2.22}$$

where the probability $\zeta_{u,j} = \Pr(Y_i > u_j)$ is estimated empirically, $\sigma_{u,j}$ and ξ_j are parameters for the GPD model of the excesses of Y_j above threshold u_j, and $R(Y_j)$ denotes the rank of Y_j in $(Y_{j,1}, \ldots, Y_{j,n})$. In other words, the marginal distribution F_{Y_j} of Y_j is modelled by a GPD above a threshold u_j and by the empirical distribution below u_j, for $j = 1, \ldots, d$. Parameters for the GPD models can be estimated as discussed in Section 2.1.3, using maximum likelihood estimation for $\sigma_{u,j}$ and ξ_j, and threshold selection techniques for u_j.

Applying transformation (2.2.22) to the marginal components of the observations $\{\mathbf{y}_i\}_{i=1}^n$ yields the required Fréchet marginal observations $\{\mathbf{x}_i\}_{i=1}^n$. The second step involves maximum likelihood estimation of the dependence model parameter $\boldsymbol{\theta}$ (i.e., α in the logistic dependence structure). Coles and Tawn (1991) show the likelihood over the set A_M can be written as

$$L(\boldsymbol{\theta}, \boldsymbol{\sigma}_u, \boldsymbol{\xi}; \{\mathbf{y}_i\}_{i=1}^n) = \exp\{-V(\mathbf{v})\} \prod_{i=1}^{n_{A_M}} \Bigg(h\left(\mathbf{w}_i; \boldsymbol{\theta}\right) (nr_i)^{-(d+1)} \\ \times \prod_{\substack{j=1,\ldots,d: \\ x_{i,j} > nv_j}} \left[\sigma_j^{-1} \zeta_{u,j}^{1/\xi_j} x_{i,j}^2 \exp\left(1/x_{i,j}\right) \left\{1 - \exp\left(-1/x_{i,j}\right)\right\}^{1-1/\xi_j}\right] \Bigg),$$

for angular coordinates $\mathbf{w}_1, \ldots, \mathbf{w}_{n_{A_M}}$ and radial coordinates $r_1, \ldots, r_{n_{A_M}}$ of $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,d})$ for $i = 1, \ldots, n_{A_M}$, scale parameter vector $\boldsymbol{\sigma}_u = (\sigma_{u,1}, \ldots, \sigma_{u,d})$, and shape parameter vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_d)$, where $\mathbf{v} = (v_1, \ldots, v_d)$ with $v_j = X_j(u_j)$ are the thresholds u_j, $j = 1, \ldots, d$, on the Fréchet scale, and n_{A_M} is the number of observations in A_M. The parametric form h for the measure density of H in intensity (2.2.18) is unique to the type of dependence structure being fitted (e.g., logistic or asymmetric logistic). See

Coles and Tawn (1991) for examples.

2.2.4 Extremal dependence types

The discussion of the measure H above alludes to differing forms of extremal dependence, quantified in this case by the placement of mass by H on the unit simplex. We now formalise this concept with the following definition of two extremal dependence types; asymptotic dependence and asymptotic independence.

Definition 2.2.3 (Asymptotic Dependence and Asymptotic Independence)**.** *Consider a d-dimensional random variable* $\mathbf{X} = (X_j; j \in D = \{1, \ldots, d\})$*, where each component has the same marginal distribution* F *with upper endpoint* x_+*. We say that a subset* $(X_j; j \in l)$ *for* $l \subseteq D$ *of the components are asymptotically dependent if*

$$\chi_l = \lim_{x \to x_+} \frac{\Pr\left(X_j > x; j \in l\right)}{1 - F(x)} > 0. \tag{2.2.23}$$

Otherwise, we say the variables in $(X_j; j \in l)$ *are asymptotically independent as a collective.*

Expression (2.2.23) is equivalent to

$$\chi_l = \lim_{x \to x_+} \Pr\left(X_j > x; j \in l \text{ when } k \in l | X_k > x\right), \tag{2.2.24}$$

that is, the probability of all variables in l being greater than x together, given than a particular X_k is greater than x, $k \in l$. For $d = 2$, $\chi_{\{1,2\}} > 0$ indicates asymptotic dependence, whereas $\chi_{\{1,2\}} = 0$ indicates asymptotic independence. For instance, the bivariate logistic distribution has $\chi_{\{1,2\}} = 2 - 2^{\alpha}$ for logistic dependence parameter α; when $\alpha = 1$ there is complete independence between the variables with $\chi_{\{1,2\}} = 0$, whereas in the limit as $\alpha \to 0$ there is perfect dependence between all variables with $\chi_{\{1,2\}} = 1$.

Care must be taken when interpreting the value of χ_l when $d > 2$. Notably, it is possible for a subset $(X_j; j \in l)$ to be asymptotically independent with $\chi_l = 0$, but a second subset $(X_j; j \in l')$ for $l' \subset l$ to be asymptotically dependent with $\chi_{l'} > 0$. For instance, consider the trivariate case of (X_1, X_2, X_3). It could be that $\chi_{\{1,2,3\}} = 0$, indicating that the trio of variables cannot appear extreme together, however, this does not exclude the possibility that $\chi_{\{1,2\}} > 0$, meaning X_1 and X_2 could be asymptotically dependent. The value of $\chi_{\{1,2\}}$ would need to be evaluated separately to determine if this is the case. More generally, we have that $\chi_l = 0 \implies \chi_{l'} = 0$ for all $l' \subset l$, but that $\chi_l > 0 \implies \chi_{l'} > 0$ for all $l' \subset l$.

The measure χ_l can be related to the placement of mass by H on the unit simplex S_d discussed in Section 2.2.2, however, care must be taken when doing so in the case of $d > 2$. For instance, $\chi_l > 0$ does not imply that the measure H places mass on the corresponding subspace of the unit simplex. It instead tells us that, for some superset $l^* \supseteq l$, we have H placing mass on the corresponding subspace. Again consider the trivariate example of Section 2.2.2 with $d = 3$. If $\chi_{\{1,2\}} > 0$, then we could either have mass placed on the subspace of S_3 where only $\{W_1, W_2\}$ are non-zero, or where all of $\{W_1, W_2, W_3\}$ are all non-zero, or both. Specifically, we do not know for certain that mass is placed on the subspace for which only $\{W_1, W_2\}$ are non-zero, as it may be a requirement that the third variable X_3 is also large when both X_1 and X_2 are large together. Similarly, $\chi_l > 0$ doesn't imply that H places mass on the subspace of S_d associated with the subset l. For instance, for a trivariate case of the logistic distribution (2.2.1), $\chi_l > 0$ for all subsets l of $\{X_1, X_2, X_3\}$; however, mass is only placed on the subset for which all of $\{W_1, W_2, W_3\}$ are non-zero. These subtleties to the interpretation of χ_l when $d > 2$ motivate alternative descriptions of extremal dependence between variables, such as that given by Simpson et al. (2020).

Models discussed so far, of the form (2.2.15), are unable to account for these complex dependence structures. For instance, even in the bivariate case $d = 2$, any extreme value

distribution of form (2.2.15) can capture asymptotic dependence between variables, but it is only capable of capturing asymptotic independence in the form of complete independence. That is, if the variables are determined to exhibit dependence (using standard statistical techniques), then they are assumed to be asymptotically dependent with $\chi_{\{1,2\}} > 0$. These limitations can lead to biased inference when applied to sub-asymptotically dependent, but asymptotically independent $\mathbf{X}$, since estimates of joint survivor probabilities will assume asymptotic dependence between components of $\mathbf{X}$ and will thus overestimate the true joint survivor probabilities.

To understand why sub-asymptotic dependence matters, we consider the works of Ledford and Tawn (1996), Ledford and Tawn (1997) and Eastoe and Tawn (2012) who build on that of Coles and Tawn (1991), offering an equivalent of extrapolation (2.2.21) in the asymptotically-independent case. Eastoe and Tawn (2012) consider unit Fréchet marginal random variables $\mathbf{X} = (X_j; j \in D)$, showing that the joint survivor function of a subset $l \subseteq D$ can be written as

$$\Pr\left(X_j > x; j \in l\right) \sim \begin{cases} x^{-1} & \text{for perfect positive dependence,} \\ \chi_l x^{-1} & \text{for asymptotic dependence } 0 < \chi_l < 1, \\ x^{-|l|} & \text{for exact independence,} \end{cases} \tag{2.2.25}$$

as $x \to \infty$, using straightforward Taylor expansions of joint Frećhet survivor probabilities. They then link the edge cases in expression (2.2.25) with the expression

$$\Pr\left(X_j > x; j \in l\right) \sim \mathcal{L}_l(x) x^{-\frac{1}{\eta_l}}, \tag{2.2.26}$$

as $x \to \infty$, for a constant $1/|l| \leq \eta_l \leq 1$ and slowly varying function $\mathcal{L}_l(x) > 0$. Using

expression (2.2.26), the conditional probability in (2.2.24) can be written as

$$\Pr\left(X_j > x; \forall j \in l \text{ and } k \in l | X_k > x\right) = \frac{\Pr\left(X_j > x; j \in l\right)}{\Pr\left(X_k > x\right)} \sim \frac{\mathcal{L}_l(x) x^{-\frac{1}{\eta_l}}}{1 - \exp\left(-\frac{1}{x}\right)} = \mathcal{L}_l^*(x) x^{1-\frac{1}{\eta_l}},$$

as $x \to \infty$, for some slowly varying function $\mathcal{L}_l^*(x)$. The value of η_l can thus be related the extremal dependence structure of $\mathbf{X}$, since it defines the value of the conditional probability seen in the limit (2.2.23). When all components of l are asymptotically dependent, $\eta_l = 1$ and $\mathcal{L}_l^*(x) \to \chi_l > 0$ as $x \to \infty$. When components of l exhibit either near or complete independence, $\eta_l = 1/|l|$. In the intermediate case of sub-asymptotic dependence (with $\chi_l = 0$), for *bivariate* l with $|l| = 2$, the constant has range $1/2 < \eta_l < 1$. This final point cannot be easily extended to $|l| > 2$, since the constant η_l will also lie in this range if some subset $l' \subset l$ exhibits asymptotic dependence with $\chi_{l'} > 0$ and $\eta_{l'} = 1$, but the remaining elements in $l \backslash l'$ are completely independent. In this case, knowledge of the constant $\eta_{l'}$ for all possible subsets $l' \subset l$ is required to determine the complete extremal dependence structure of the set indexed by l, see e.g., Eastoe and Tawn (2012). For example, say that X_1 and X_2 are asymptotically dependent, but X_3 is completely independent from X_1 and X_2. In this case we have $\eta_{\{1,2,3\}} = 2/3$ with $1/3 < \eta_{\{1,2,3\}} < 1$, suggesting all three variables as a collective fall into the class of asymptotic independence with sub-asymptotic dependence. Instead, evaluation of $\eta_{\{1,2\}} = 1$ and $\eta_{\{1,3\}} = \eta_{\{2,3\}} = 1/2$ is needed to reveal the full extremal dependence structure.

Ledford and Tawn (1997) develop an extension of the point process approach of Coles and Tawn (1991) which is applicable for bivariate $\mathbf{X}$ with Fréchet margins to the intermediate case of $1/2 < \eta_l < 1$. They consider the process of points $\{n^{-\eta_l}\mathbf{X}_i\}_{i=1}^n$ and show that it converges to a Poisson process with intensity function

$$\mu_{I,\eta_l}(\mathrm{d}r \times \mathrm{d}w) = \frac{\mathrm{d}r}{r^{1+1/\eta_l}} \mathrm{d}H_{\eta_l}(w),$$

for r and w in (2.2.19), where H_{η_l} is a positive, finite measure on $w \in (0,1)$ when $\eta_{\{1,2\}} < 1$ and $r > 0$. Consider the set $B_u \subseteq A_{u,\epsilon}$, where $A_{u,\epsilon} = \{(w,r) : r > u, \epsilon < w < 1-\epsilon\}$ for any $0 < \epsilon < 1/2$. Applying similar arguments to those used to obtain the extrapolation (2.2.21), we can write the integrated intensity function of this Poisson process over tB_u for $t \geq 1$ as

$$\begin{aligned}\Lambda(tB_u) &= \int_{tB_u} \frac{\mathrm{d}r}{r^{1+1/\eta_l}} \mathrm{d}H_{\eta_l}(w) = \int_{B_u} \frac{t\mathrm{d}s}{(ts)^{1+1/\eta_l}} \mathrm{d}H_{\eta_l}(w) \quad (\text{where } r = st) \\ &= \frac{1}{t^{1/\eta_l}} \int_{B_u} \frac{\mathrm{d}s}{s^2} \mathrm{d}H_{\eta_l}(w) = \frac{1}{t^{1/\eta_l}} \Lambda(B_u),\end{aligned}$$

and so obtain

$$\Pr\left(\mathbf{X} \in tB_u\right) = \frac{1}{t^{1/\eta_l}} \Pr\left(\mathbf{X} \in B_u\right),$$

for $t \geq 1$, the equivalent extrapolation for a random variable $\mathbf{X}$ with constant η_l. Setting $\eta_l = 1$ as in the asymptotically dependence case yields the original extrapolation in (2.2.21), whereas for $\eta_l < 1$ the probability of the extrapolated event decays faster than when $\eta_l = 1$. This development was followed by the likes of Bortot and Tawn (1998) and Bortot et al. (2000) who propose models for the asymptotically independent, but sub-asymptotically dependent case; for a review of these models and examples of associated extremal dependence types see Coles et al. (1999).

The values χ_l and η_l reveal information on the extremal dependence structure of the joint behaviour of $\{X_j; j \in l\}$ in the limit, meaning they cannot be used when working with a finite set of observations. Instead, Coles et al. (1999) provide a measure for extremal dependence useful at finite levels, for bivariate $\mathbf{X} = (X_1, X_2)$, with $X_1, \sim F_1$ and $X_2 \sim F_2$ for known distributions F_1 and F_2. Take $(X_{U,1}, X_{U,2}) = \{F_1(X_1), F_2(X_2)\}$ on uniform margins, and rewrite the non-limit element of expression (2.2.23) as

$$\chi_{\{1,2\}}(u) = \Pr\left(X_{U,2} > u | X_{U,1} > u\right), \tag{2.2.27}$$

for $0 < u < 1$, providing a measure of the dependence at the quantile u, and a measure of the asymptotic dependence in its limit as $u \to 1$. The measure $\chi_{\{1,2\}}(u)$ can thus be used for large finite u to determine if a pair (X_1, X_2) exhibit asymptotic dependence. For asymptotic dependent variables, $\chi_{\{1,2\}}(u) > 0$ for all u. For perfectly dependent variables, $\chi_{\{1,2\}}(u) = 1$ for all u, whereas for completely independent variables $\chi_{\{1,2\}}(u) = 1 - u$ for all u. However, in the case of asymptotically independent variables, the convergence of $\chi_{\{1,2\}}(u)$ to zero as $u \to 1$ can be slow, resulting in the measure misleadingly appearing to be positive or greater than $1-u$ when tested at finite levels. Additionally, $\chi_{\{1,2\}}(u)$ provides no detail of the dependence of asymptotically independent variables as $\chi_{\{1,2\}}(u) \to 0$ as $u \to 1$, only indicating that they fall into this class. To avoid these issues, Coles et al. (1999) define a second dependence measure in the bivariate case as

$$\bar{\chi}_{\{1,2\}}(u) = \frac{2 \log \Pr\left(X_{U,1} > u\right)}{\log \Pr\left(X_{U,1} > u, X_{U,2} > u\right)} - 1 \quad \text{for} \quad 0 \leq u \leq 1,$$

where $-1 < \bar{\chi}_{\{1,2\}}(u) \leq 1$ for all $0 \leq u \leq 1$ and $\bar{\chi}_{\{1,2\}}(u) \to \bar{\chi}_{\{1,2\}}$ as $u \to 1$, for

$$\bar{\chi}_{\{1,2\}} = 2\eta_l - 1.$$

For asymptotically dependent variables, $\bar{\chi}_{\{1,2\}} = 1$, and conversely $\bar{\chi}_{\{1,2\}} \leq 1$ for asymptotically independent variables. In the latter case, $\bar{\chi}_{\{1,2\}}(u)$ provides a useful measure of the strength of dependence, in a sense determining how close to asymptotic dependence the variables are. Also, convergence of $\bar{\chi}_{\{1,2\}}(u)$ to $\bar{\chi}_{\{1,2\}}$ as $u \to 1$ is easier to assess than the convergence of $\chi_{\{1,2\}}(u)$ to zero as $u \to 1$, facilitating more accurate interpretation at finite levels. See Coles (2001) for illustrations of the relative convergence rate of the two measures.

2.2.5 Conditional multivariate extreme value models

We now discuss the model of Heffernan and Tawn (2004) which utilises the third ordering discussed in Section 2.2.1, modelling multivariate events where at least one component is large. As motivation, consider the metric $\chi_{\{1,2\}}(u)$ defined in expression (2.2.27). The value of this metric measures the strength of extremal dependence between uniform random variables $X_{U,1}$ and $X_{U,2}$ in terms of the probability that $X_{U,2}$ is greater than a value u, given that $X_{U,1}$ is known to be greater than u. In other words, it describes the behaviour of the 'unconditioned' variable $X_{U,2}$, given that the 'conditioned' variable $X_{U,1}$ exceeds some threshold, but only in the cases where $X_{U,2}$ is as extreme as $X_{U,1}$. The method of Heffernan and Tawn (2004) provides a framework to model this conditional behaviour, thereby capturing the extremal dependence structure between the two variables, whether the values of $X_{U,2}$ associated with $X_{U,1} > u$ grow at the same rate as $X_{U,1}$, or not. Unlike the models of Section 2.2.2 and 2.2.3, which are defined for Fréchet margins, Heffernan and Tawn (2004) work with Gumbel margins as they find these to be more suitable to studying extremal dependence under this ordering as it induces linear relationships when there is positive dependence. The later work of Keef et al. (2013), which builds on that of Heffernan and Tawn (2004), employs Laplace margins as this allows for more concise definition of model parameter support when negative dependence is feasible. For this reason, we present the method of Heffernan and Tawn (2004) as is done by Keef et al. (2013), using Laplace margins throughout.

Consider a multivariate random variable $\mathbf{X} = (X_1, \ldots, X_d) \in \mathbb{R}^d$ with standard Laplace univariate marginal distributions. Heffernan and Tawn (2004) define a model for the behaviour of the random variable $\mathbf{X}$ given that component X_j exceeds some threshold $v_j > 0$, by describing the linear asymptotic relationship between X_j and the remaining components of $\mathbf{X}$, when $X_j > v_j$, via unique location $\mathbf{a}_{|j}(x) : \mathbb{R} \mapsto \mathbb{R}_+^{d-1}$ and scale $\mathbf{b}_{|j}(x) : \mathbb{R} \mapsto \mathbb{R}_+^{d-1}$ normalising functions. In practice, these normalising functions

can be found to be of the form of a linear function and a power function respectively for a wide range of parametric copula examples, with these forming parsimonious parametric families for $\mathbf{a}_{|j}(x)$ and $\mathbf{b}_{|j}(x)$. Hence we focus on these parametric functions, with the gradients of the linear function being unknown parameters $\boldsymbol{\alpha}_{|j} \in [-1,1]^{d-1}$ and the powers of the power function being unknown parameters $\boldsymbol{\beta}_{|j} \in (-\infty,1]^{d-1}$. Then, Heffernan and Tawn (2004) assume, for $\mathbf{z}_{|j} \in \mathbb{R}^{d-1}$ and $x > 0$, that

$$\mathbb{P}\left(\frac{\mathbf{X}_{-j} - \boldsymbol{\alpha}_{|j} X_j}{X_j^{\boldsymbol{\beta}_{|j}}} < \mathbf{z}_{|j}, X_j - v_j > x | X_j > v_j\right) \to e^{-x} G_{|j}(\mathbf{z}_{|j}), \tag{2.2.28}$$

as $v_j \to \infty$, for distribution function $G_{|j} : \mathbb{R}^{d-1} \mapsto \mathbb{R}$ with non-degenerate marginals, where componentwise operations are assumed. Here, $\mathbf{X}_{-j}$ is the vector of all components X_k of $\mathbf{X}$, for $k \neq j$. The distribution $G_{|j}$ is termed the $(d-1)$-dimensional multivariate residual distribution, for residual

$$\mathbf{Z}_{|j} = \lim_{v_j \to \infty} \left\{ \frac{\mathbf{X}_{-j} - \boldsymbol{\alpha}_{|j} X_j}{X_j^{\boldsymbol{\beta}_{|j}}} : X_j > v_j \right\}.$$

The value of $\alpha_{k'|j}$ for a particular $k' \in D_j$, $D_j = \{1,\ldots,d\} \setminus j$, where $\boldsymbol{\alpha}_{|j} = (\alpha_{k|j}; k \in D_j)$, is the most important feature of the representation in terms of describing the extremal dependence structure of $(X_j, X_{k'})$, whilst the value of $\beta_{k'|j}$, where $\boldsymbol{\beta}_{|j} = (\beta_{k|j}; k \in D_j)$ describes features of the dependence structure which explain more subtle details of the relationship between the variables. When variables $(X_j, X_{k'})$ are asymptotically dependent, $(\alpha_{k'|j}, \beta_{k'|j}) = (1, 0)$ and vice versa. Also, when $(X_j, X_{k'})$ are asymptotically independent, $(\alpha_{k'|j}, \beta_{k'|j}) \in (-1,1) \times (-\infty, 1)$ and vice versa, subject to additional joint constraints on the parameter space that are discussed in Chapters 3 and 4. For any $0 < \alpha_{k'|j} < 1$, the variables exhibit positive dependence, and for any $-1 < \alpha_{k'|j} < 0$, they exhibit negative dependence. When the variables exhibit exact independence, $\alpha_{k'|j} = \beta_{k'|j} = 0$. When $\beta_{k'|j} > 0$ ($\beta_{k'|j} < 0$), there is greater (less) variability in $X_k|(X_j > v_j)$ respectively, than when $\beta_{k'|j} = 0$, where the relationship is

linear. The model of Heffernan and Tawn (2004) is therefore capable of capturing both forms of extremal dependence, for any combination of pairwise asymptotic dependence and asymptotic independence between elements of $\mathbf{X}$.

The product in the limit of (2.2.28) corresponds to the excesses of X_j above a threshold v_j being independent of residual $\mathbf{Z}_{|j}$. In practice, inference under this model is carried out by assuming this limiting behaviour holds for some finite large v_j. The selection of this value v_j is the focus of Chapter 5 of this thesis, where we build from the work of Wan and Davis (2019) who introduce a visual heuristic to aid the choice of finite radial threshold for which the limiting intensity measure (2.2.18) is approximately valid. Given a choice of v_j, for $j = 1, \ldots, d$, inference typically proceeds as follows. Consider a set of observations $\{\mathbf{x}_i\}_{i=1}^n$ of $\mathbf{X}$. Assuming the limit (2.2.28) holds for the chosen value of v_j, we can write

$$(X_j - v_j)|(X_j > v_j) \sim \text{Exp}(1) \quad \perp \quad \mathbf{X}_{-j}|\{X_j = x\} = \boldsymbol{\alpha}_{|j}x + x^{\boldsymbol{\beta}_{|j}}\mathbf{Z}_{|j}, \tag{2.2.29}$$

for $x \geq v_j$, with $(\boldsymbol{\alpha}_{|j}, \boldsymbol{\beta}_{|j})$ and residual $\mathbf{Z}_{|j}$ estimated from the sample $\{\mathbf{x}_i\}_{i=1}^n$. A parametric form for the joint distribution of $\mathbf{Z}_{|j}$ must be assumed when carrying out inference for $(\boldsymbol{\alpha}_{|j}, \boldsymbol{\beta}_{|j})$, which is typically taken to be independent Gaussians with unknown locations $\boldsymbol{\mu}_{|j} = (\mu_{k|j}; k \in D_j) \in \mathbb{R}^{d-1}$ and scales $\boldsymbol{\sigma}_{|j} = (\sigma_{k|j}; k \in D_j) \in \mathbb{R}_+^{d-1}$ as suggested by Heffernan and Tawn (2004). Under this assumption, parameter vectors $\boldsymbol{\alpha}_{|j}$ and $\boldsymbol{\beta}_{|j}$ can be estimated by maximisation of the log-likelihood

$$\ell_j(\boldsymbol{\alpha}_{|j}, \boldsymbol{\beta}_{|j}, \boldsymbol{\mu}_{|j}, \boldsymbol{\sigma}_{|j}; \{\mathbf{x}_i\}_{i=1}^{n_j}) = \sum_{i=1}^{n_j} \left\{ \sum_{k \in D_j} \ell_{j,k}(\alpha_{k|j}, \beta_{k|j}, \sigma_{k|j}, \mu_{k|j}; \{x_{-j,i}^{(k)}, x_{j,i}\}) \right\},$$

where the index i iterates over the n_j observations of $\mathbf{X}$ where $X_j > v_j$, and

$$\ell_{j,k}(\alpha_{k|j}, \beta_{k|j}, \sigma_{k|j}, \mu_{k|j}; \{x^{(k)}_{-j,i}, x_{j,i}\}) = -\log(x_{j,i}^{\beta_{k|j}}\sigma_{k|j}) - \frac{1}{2}\left[\frac{\left(x^{(k)}_{-j,i} - \alpha_{k|j}x_{j,i} - x_{j,i}^{\beta_{k|j}}\mu_{k|j}\right)^2}{x_{j,i}^{2\beta_{k|j}}\sigma^2_{k|j}}\right],$$

up to a constant, where here $x_{j,i} > v_j$, for $i = 1, \ldots, n_j$, for observed values $\mathbf{x}_{-j,1}, \ldots, \mathbf{x}_{-j,n_j}$ of $\mathbf{X}_{-j}$ and $x_{j,1}, \ldots, x_{j,n_j}$ of X_j, with $\mathbf{x}_{-j,i} = (x^{(k)}_{-j,i}; k \in D_j)$. This inference procedure can be repeated by conditioning on each component of $\mathbf{X}$ in turn, obtaining estimates for parameters $\boldsymbol{\alpha}_{|j}$ and $\boldsymbol{\beta}_{|j}$ for $j = 1, \ldots, d$.

Importantly, the Gaussian form assumed for maximum likelihood estimation is only used as a working assumption, and the residual distribution is not modelled in this way. Instead, it is typically modelled as the empirical distribution using the observed residuals $\{\mathbf{z}^{(i)}_{|j}\}_{i=1}^{n_j}$ (as in Heffernan and Tawn 2004), where

$$\mathbf{z}^{(i)}_{|j} = \frac{\mathbf{x}_{-j,i} - \hat{\boldsymbol{\alpha}}_{|j}x_{j,i}}{x_{j,i}^{\hat{\boldsymbol{\beta}}_{|j}}}$$

for $i = 1, \ldots, n_j$, for maximum likelihood estimates $\hat{\boldsymbol{\alpha}}_{|j}$ and $\hat{\boldsymbol{\beta}}_{|j}$ of $\boldsymbol{\alpha}_{|j}$ and $\boldsymbol{\beta}_{|j}$. Alternatively, the residual distribution can be modelled as the Gaussian kernel smoothed density estimate of the observed residuals $\{\mathbf{z}^{(i)}_{|j}\}_{i=1}^{n_j}$ (as in Winter and Tawn 2017). There are several works which consider more sophisticated models for the residuals $\mathbf{Z}_{|j}$, $j = 1, \ldots, d$, aiming to provide parametric or semi-parametric form for the dependence structure between $Z_{i|j}$ and $Z_{k|j}$ for $i \neq j \neq k$. For instance, Towe et al. (2019) place a Gaussian copula on the joint distribution of residuals, with marginal distributions given by the smoothed kernel density estimates of Winter and Tawn (2017). Alternatively, Wadsworth and Tawn (2022) assume a Gaussian copula with generalised-Gaussian margins. More recent work includes that of Farrell et al. (2025), who use a Gaussian graphical modelling approach for the copula of $\mathbf{Z}_{|j}$ and allow for asymmetry in the marginal residual distributions.

Given estimates for the parameter vectors $\boldsymbol{\alpha}_{|j}$ and $\boldsymbol{\beta}_{|j}$, and an estimate for the residual distribution function $G_{|j}$, for $j = 1, \ldots, d$, the probability $\Pr(\mathbf{X} \in A)$ for $A \subseteq \{\mathbf{X} \in \mathbb{R}^d : \sum_{j=1}^{d} I(X_j > v_j) > 0\}$, for identify function I, can be estimated under the conditional extremes model by considering the decomposition

$$\Pr(\mathbf{X} \in A) = \sum_{j=1}^{d} \Pr(\mathbf{X} \in A_j) = \sum_{j=1}^{d} \Pr(\mathbf{X} \in A_j | X_j > u_j) \Pr(X_j > u_j),$$

for $A_j = A \cap \{\mathbf{X} \in \mathbb{R}^d : F_L(X_j) > F_L(X_k) : k \in 1, \ldots, d, k \neq j\}$, so A_j is the part of A for which element X_j is the largest, for $j = 1, \ldots, d$. Here, F_L is the standard Laplace marginal distribution function, and $u_j = \inf_{\mathbf{x} \in A_j}(x_j)$ is the smallest value of x_j over all elements in A_j, for all $j = 1, \ldots, d$. Since the random variable X_j is Laplace distributed, the probability $\Pr(X_j > u_j) = 1 - F_L(u_j)$ can be obtained simply from the standard Laplace distribution function F_L. The conditional probability $\Pr(\mathbf{X} \in A_j | X_j > u_j)$, for $j \in 1, \ldots, d$, can be estimated under the conditional extremes model by drawing additional samples of $\mathbf{X}$ within the region where $X_j > u_j$, provided that $u_j \geq v_j$ for chosen v_j in expression (2.2.29). These samples are obtained by first drawing $(X_j - u_j)|(X_j > u_j) \sim \text{Exp}(1)$, then using the fitted model for $G_{|j}$ to sample the residual components $\mathbf{Z}_{|j}$. The sampled values of X_j and $\mathbf{Z}_{|j}$, along with the estimates for $\boldsymbol{\alpha}_{|j}$ and $\boldsymbol{\beta}_{|j}$, can then be used to obtain sampled values of $\mathbf{X}_{-j}$ via expression (2.2.29). This process yields samples of the full random vector $\mathbf{X} = (\mathbf{X}_{-j}, X_j)$ in the region $\{\mathbf{X} \in \mathbb{R}^d : X_j > u_j\}$, which can then be used to estimate the probability $\Pr(\mathbf{X} \in A_j | X_j > u_j)$.

This procedure can be applied for any arbitrary random vector $\mathbf{Y}$ on $\mathbb{R}^d$ with unknown marginals, by first mapping the observed data to Laplace margins, and then mapping the simulated data back to the original margins, via a transformation analogous to the transformation (2.2.22) applied for the point process model of Coles and Tawn (1991), see e.g., Section 3.3.2 of Chapter 3.

2.2.6 Multivariate threshold selection

In Section 2.2.3, when discussing the point process representation of multivariate extremes, we allude to the choice of a threshold $u > 0$ required to define the set A_u in expression (2.2.20), on which the approximate Poisson process behaviour is assumed to hold. Wan and Davis (2019) propose a method, for the selection of this threshold u in the bivariate case.

Take a bivariate random variable $(X_1, X_2) \in \mathbb{R}^2_+$ with unit Fréchet marginals, and consider the radial-angular variables

$$R = X_1 + X_2 \quad \text{and} \quad W = \frac{X_1}{X_1 + X_2},$$

as in expression (2.2.19). Multivariate extreme value models of the form (2.2.15) can be used to model the joint extremal dependence structure of (X_1, X_2) in the region $\{\mathbb{R}^2_+ : R > u\}$, for a finite radial threshold $u > 0$. This region is chosen such that it can be assumed that the independence between radius R and angle W given by the limiting expression (2.2.18) holds. Wan and Davis (2019) say that a candidate value for the threshold u can be considered valid when this independence can be observed empirically, and therefore develop an approach which relies on the assessment of the empirical dependence between R and W on the set $\{\mathbb{R}^2_+ : R > u\}$.

Consider candidate threshold u_n, which is to be assessed using some n observations of (X_1, X_2). Wan and Davis (2019) measure the dependence between observed values of angle $W|(R > u_n)$ and radius $R|(R > u_n)$ using the empirical distance covariance of the angle W and the normalised radius R/u_n, defined as

$$T_n = \int_{\mathbb{R}^2_+} |\hat{\phi}_{(R/u_n, W)|u_n}(s,t) - \hat{\phi}_{R/u_n|u_n}(s)\hat{\phi}_{W|u_n}(t)|^2 \mu(\mathrm{d}s, \mathrm{d}t),$$

for suitable measure μ on $\mathbb{R}^2_+$, and empirical characteristic functions $\hat{\phi}_{(R/u_n, W)|u_n}$, $\hat{\phi}_{R/u_n|u_n}$

and $\hat{\phi}_{W|u_n}$ of $(R/u_n, W)$, R/u_n and W respectively, all conditional on $R > u_n$. Wan and Davis (2019) show that, under certain conditions, for independent R/u_n and W,

$$n_u T_n \to \int_{\mathbb{R}^2_+} |Q(s,t)|^2 \mu(ds, dt) \tag{2.2.30}$$

as $n_u \to \infty$, where n_u is the number of observations such that $R > u_n$, and Q is a centred Gaussian process with particular covariance function specified in Wan and Davis (2019). They then use convergence (2.2.30) to establish the hypothesis testing framework

$$H_0 : R/u_n \text{ and } W \text{ are independent given } R > u_n,$$
$$H_1 : R/u_n \text{ and } W \text{ are not independent given } R > u_n,$$

with p-value for testing H_0 against H_1 given by

$$p = \Pr\left(\int_{\mathbb{R}^2_+} |Q(s,t)|^2 \mu(ds, dt) > u \right) \Bigg|_{u = n_u T_n}.$$

In practice, the test statistic is calculated using a bootstrapping procedure, since its integral is intractable.

This hypothesis test can be carried out for a range of values of candidate threshold u_n. Under H_0, p should be $U(0,1)$, and under H_1, p should be close to zero. Wan and Davis (2019) devise an algorithm to select the optimal value of u^* for a given set of observations of (X_1, X_2), firstly bootstrapping the original sample of (X_1, X_2) to obtain replicate data, and then studying the resulting bootstrapped sampled values of p corresponding to each candidate threshold u_n. Following from the design of the hypothesis test, if uniform values of p are observed for all $u_n > u^*$ for some u^*, then the threshold u^* is chosen as the best candidate value. Wan and Davis (2019) use the spline fitting approach of Fryzlewicz (2014) to automatically select the value of u^* above which the associated p-values become a sample from a joint distribution with uniform

marginal distributions; see Wan and Davis (2019) for further details of this approach.

2.2.7 Geometric extremes

Consider the bivariate random variable (X_1, X_2) with unit exponential marginals, so that $X_1, X_2 \sim \text{Exp}(1)$. Eddy and Gale (1981), Brozius and de Haan (1987) and Davis et al. (1987), amongst others, study the shape of the sample cloud

$$C_n = \{(X_{1,i}, X_{2,i}) : i = 1, \ldots, n\},$$

as $n \to \infty$, for n independent samples $(X_{1,1}, X_{2,1}), \ldots, (X_{1,n}, X_{2,n})$ from the joint distribution of (X_1, X_2), relating the shape of the boundary of this sample cloud to the extremal dependence structure of (X_1, X_2). Simpson and Tawn (2024) build on such work and model the extremal dependence structure of the two variables by studying the shape of the normalised sample cloud

$$C_n^* = \{(X_{1,i}, X_{2,i}) \,/\, \log n : i = 1, \ldots, n\},$$

as $n \to \infty$. In simpler cases (where mass is placed on either the centre or edge subspaces of the unit simplex, but not both), the limiting convex hull of C_n^* provides a geometric interpretation of the dependence structure of (X_1, X_2), see e.g., Figures 2.2.2a and 2.2.2b. In cases with more complex extremal dependence structure (where mass is placed on both the centre or edge subspaces of the unit simplex), it is the boundary of a sub-region lying within the limiting convex hull that describes the extremal dependence structure, see e.g., Figures 2.2.2c. In either scenario, this limiting shape can be described using the gauge function $g(x_1, x_2)$, defined by

$$-\log f\left(tx_1, tx_2\right) \sim tg\left(x_1, x_2\right), \quad t \to \infty, \quad x_1, x_2 \geq 0. \tag{2.2.31}$$

see e.g., Nolde (2014) and Nolde and Wadsworth (2022). In the limit as $n \to \infty$, the scaled sample cloud C_n^* converges to the limit set

$$G^* = \{(x_1, x_2) : g(x_1, x_2) \leq 1\} \subseteq [0, 1]^2,$$

which is assumed to exist. The boundary set $G = \{(x_1, x_2) : g(x_1, x_2) = 1\} \subset [0, 1]^2$ determines the limiting shape of the scaled sample cloud C_n^* as $n \to \infty$, thereby providing inference for the dependence structure of (X_1, X_2) since different structures will provide varying boundary sets G.

Nolde (2014) and Nolde and Wadsworth (2022) quantify the relationship between the shape of G and the extremal dependence structure of (X_1, X_2) by linking G to the dependence metrics and model parameters of Sections 2.2.4 and 2.2.5. Nolde (2014) show that the coefficient of extremal dependence $1/2 \leq \eta \leq 1$, defined by expression (2.2.26), can be linked to G via

$$\eta = \min\{s \in (0, 1] : [s, \infty]^2 \cap G = \emptyset\}.$$

Intuitively, the value of η can therefore be found by moving the set $[1, \infty]^2$ along the line $X_2 = X_1$ in the direction of the origin until it last intersects G. See the supplementary material to Simpson and Tawn (2024) for an illustration. Similarly, Nolde and Wadsworth (2022) relate the parameters $(\alpha_{2|1}, \alpha_{1|2})$ from the conditional extremes model of Heffernan and Tawn (2004) to the shape of G via

$$\alpha_{2|1} = \max\{\tilde{\alpha}_{2|1} \in [0, 1] : g(1, \tilde{\alpha}_{2|1}) = 1\},$$
$$\alpha_{1|2} = \max\{\tilde{\alpha}_{1|2} \in [0, 1] : g(\tilde{\alpha}_{1|2}, 1) = 1\}.$$

for gauge function g defined by expression (2.2.31). That is, the values of $(\alpha_{2|1}, \alpha_{1|2})$ are determined by where the boundary set G comes into contact with the edges of the

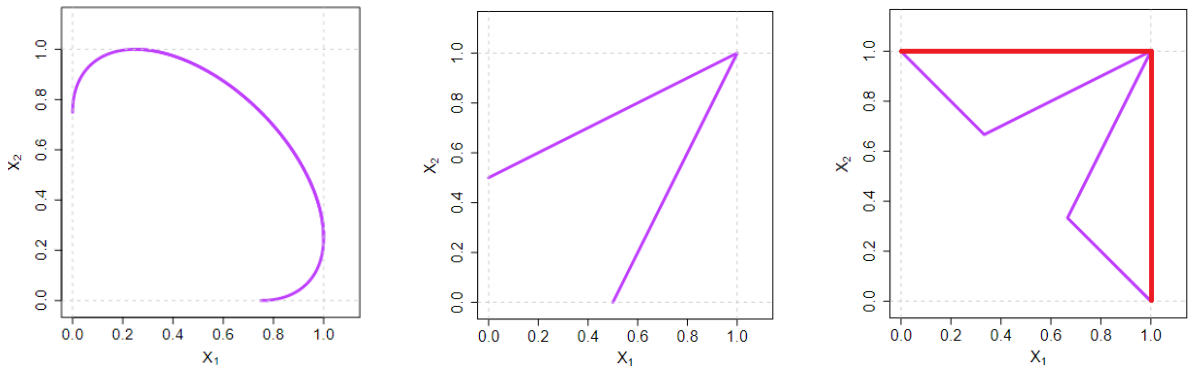
bounding box $[0,1]^2$. Notably, it is possible for G to come into contact with $[0,1]^2$ in more than two places; in this case, the contact points on each edge closest to the upper corner $(1,1)$ define the values of $(\alpha_{2|1}, \alpha_{1|2})$. Nolde and Wadsworth (2022) also relate G to the conditional model parameters $(\beta_{2|1}, \beta_{1|2})$, showing that

$$g(1, \alpha_{2|1} + u) = 1 + O\left\{u^{1/(1-\beta_{2|1})}\right\},$$

as $u \downarrow 0$, and similarly for $\beta_{1|2}$.

Figure 2.2.2 shows three examples of the boundary set G from Simpson and Tawn (2024), plotted for three example dependence structures for exponential margins. Figure 2.2.2a shows the boundary set g for a Gaussian copula with dependence parameter $\rho = 0.5$. Here, g does not come into contact with the upper corner $(1,1)$ of the bounding box $[0,1]^2$, indicating that $\eta < 1$ and thus corresponding to the categorisation of asymptotic independence provided by Ledford and Tawn (1996). Also, the contact points of g to the edges of $[0,1]^2$ at $(0.25, 0.25)$ correctly indicate that $(\alpha_{2|1}, \alpha_{1|2}) = (\rho^2, \rho^2)$, as discussed by Simpson and Tawn (2024). Similarly, Figures 2.2.2b and 2.2.2c both indicate that $\eta = \alpha_{2|1} = \alpha_{1|2} = 1$ since the boundary set touches the point $(1,1)$, correctly identifying the asymptotic dependence of the logistic and asymmetric logistic copulas. The additional structure see in Figure 2.2.2c reflects the fact that, under the asymmetric logistic copula, the variables can be extreme both together and in isolation, as is discussed in Section 2.2.2.

Simpson and Tawn (2024) provide methods of inference for this boundary set G for a bivariate random variable (X_1, X_2) with exponential margins, and hence for estimating the extremal dependence structure of (X_1, X_2). This has been followed by additional work on the estimation of the boundary set G. For instance, Wadsworth and Campbell (2024), Murphy-Barltrop et al. (2024) and Majumder et al. (2025) each develop semi-parametric inference approaches for estimation of the gauge function g. Most recently, Murphy-Barltrop et al. (2025) propose an inference approach for g in the non-stationary



(a) Gaussian dependence structure with $\rho = 0.5$.

(b) Logistic dependence structure with $\alpha = 0.5$.

(c) Asymmetric logistic structure with $\alpha = 0.5$.

Figure 2.2.2: Examples of sample cloud boundary set G for three representative bivariate dependence structures, as presented by Simpson and Tawn (2024). In third panel, the limiting convex hull is shown in red.

setting.

2.3 Choice of multivariate extreme value model

In the works presented in Chapters 3, 4 and 5 of this thesis, we choose to employ the conditional multivariate extreme value model of Section 2.2.5 over the other multivariate extreme value approaches discussed here, for the following reasons. As discussed in Section 2.2.4, the componentwise maxima and point process models of Sections 2.2.2 and 2.2.3 assume asymptotic dependence between any sets of variables which do not exhibit complete independence. This assumption leads to biased estimates of joint extremal probabilities when modelling the joint behaviour of asymptotically independent variables, and so these models are not well suited to the environmental applications in this thesis, as variables within environmental contexts will often exhibit asymptotic independence, or at least we are not certain that they exhibit asymptotic dependence. Similarly, models based on the work of Ledford and Tawn (1996), discussed briefly in Section 2.2.4, only allow for estimation of the probability of joint sets along the 'diagonal' of $\mathbb{R}^d$, $d > 1$, in the sense that all variables within the set must grow at the same rate. This is too limiting a restriction in our application, where we wish to estimate probabilities within any arbitrary region of $\mathbb{R}^d$, including those lying close to the axis. The geometric approach described in Section 2.2.7 overcomes the limitations of the other methods, however, it is currently not well-defined in higher dimensions. Additionally, it lacks the flexibility to easily introduce non-stationarity in the extremal dependence structure modelling, unlike the conditional extreme value framework for which this can be achieved through the addition of covariates onto the dependence parameters, see e.g., Jonathan et al. (2013). For these reasons, we believe the conditional extreme value framework presents the ideal balance of a flexible but well-established methodology, and therefore use it as our chosen multivariate extreme value model throughout this thesis.

2.4 Linear wave theory

2.4.1 Overview

Wave theory provides a framework for theoretical and numerical modelling of the ocean environment and ocean-structure interactions, including the behaviour of random variables relevant to quantifying the reliability of offshore structures. The statistical methods introduced in Sections 2.1-2.2 can be used to characterise joint extreme sea-state conditions (e.g., of significant wave height H_s and wave steepness S_e at some location over a given interval of time). In order to assess the reliability of an offshore structure to these sea-state conditions, we need to quantify the extreme loads on and responses of that structure, as outlined in Figure 1.1.1 and its discussion in Section 1.1. We achieve this using a combination of wave theory and fluid loading methodologies, outlined here.

Linear (or Airy) wave theory provides a linearised description of the propagation of gravity-driven waves on the surface of a homogenous fluid (which is a reasonable assumption for localised areas of the ocean), and associated wave kinematics (that is, velocities and accelerations). The theory is based on the solution of Newton's second law and the equation of conservation of mass for a fluid with zero viscosity and constant density, exhibiting irrotational flow. Wave amplitudes, velocities and accelerations are assumed to be small in size, so that second and higher order terms can be neglected from equations, thereby linearising their solutions. Velocities and accelerations from linear wave theory can be used e.g., as input to the empirically derived Morison equation (Morison et al., 1950) for the combined drag and inertial loads on an idealised structure in the fluid. Wave surface elevations can be used as input to other models, e.g., for inertial loading and response of monopile structures (Taylor et al., 2024).

This section is provides a brief introduction and justification of linear wave theory as a basis for the wave kinematic models utilised in Chapters 3 and 4. Section 2.4.2 introduces the concept of the material derivative of a fluid, a derivative operator following

fluid flow, essential to manipulation of equations of fluid motion and mass conservation. Section 2.4.3 uses arguments based on the conservation of mass to derive the Laplace field equation, demonstrating that linearised fluid flow can be considered an application of potential theory, and quantified in terms of a velocity potential. Section 2.4.4 uses arguments based on Newton's second law to derive the Bernoulli equation (e.g., Dean and Dalrymple 1991, relating the fluid velocity potential to pressure and gravitational forcing, and providing a so called dynamic surface boundary condition for the fluid motion. Section 2.4.5 constructs additional so called kinematic boundary conditions at the free surface and ocean floor, essentially ensuring 'no flow' out of the fluid, as explained by e.g., Holthuijsen (2010). Section 2.4.6 presents an analytical solution to the Laplace field equation subject to dynamic and kinematic boundary conditions, in terms of functional forms for the velocity potential, velocity vector and acceleration vector (all in 3-D space and time), and the free surface elevation (in 2-D space and time). All physical quantities are given in SI units throughout.

2.4.2 The material derivative

Consider a volume of fluid with a given velocity field $\mathbf{u} = (u_x, u_y, u_z) \in \mathbb{R}^3$, such that $\mathbf{u}(\mathbf{x}, t)$ is the velocity $\mathbf{u}$ at location $\mathbf{x} = (x, y, z) \in \mathbb{R}^3$ at time $t \in \mathbb{R}$. Suppose that this fluid has some general scalar property $\phi \in \mathbb{R}$ and vector property $\mathbf{b} \in \mathbb{R}^3$, e.g., density and momentum respectively. The following definitions of the 'material derivative' of scalar property ϕ (e.g., density ρ) and vector property $\mathbf{b}$ (e.g., velocity $\mathbf{u}$) provide Lagrangian derivatives which 'follow' the fluid in the direction of $\mathbf{u}$, and are widely used in calculations of fluid motion, see e.g., Vallis (2017). For scalar quantity ϕ, the material derivative is given as

$$\frac{\mathrm{D}\phi}{\mathrm{D}t} = \frac{\partial \phi}{\partial t} + \mathbf{u} \cdot \boldsymbol{\nabla} \phi, \tag{2.4.32}$$

for gradient operator $\boldsymbol{\nabla}$ such that $\boldsymbol{\nabla}\phi$ is the vector

$$\boldsymbol{\nabla}\phi = \left(\frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y}, \frac{\partial\phi}{\partial z}\right).$$

Similarly, for vector quantity $\mathbf{b}$, the material derivative is defined as

$$\frac{\mathrm{D}\mathbf{b}}{\mathrm{D}t} = \frac{\partial\mathbf{b}}{\partial t} + (\mathbf{u}\cdot\boldsymbol{\nabla})\mathbf{b}, \tag{2.4.33}$$

where

$$\mathbf{u}\cdot\boldsymbol{\nabla} = u_x\frac{\partial}{\partial x} + u_y\frac{\partial}{\partial y} + u_z\frac{\partial}{\partial z},$$

making expression (2.4.33) the componentwise equivalent of expression (2.4.32). The notation $\mathrm{D}/\mathrm{D}t$ is used to denote the material derivates (2.4.32) and (2.4.33) throughout this section.

Next, consider the volume V of the fluid with velocity field $\mathbf{u}$. Over time, as the fluid is compressed and deformed by its own motion, the volume that this fluid occupies will vary. Therefore, the total amount of scalar property ϕ in this volume will also change. This change can be quantified by the material derivative of the volume integral of ϕ, defined (Vallis, 2017) as

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\int_V \phi \mathrm{d}V\right) = \int_V \left(\frac{\mathrm{D}\phi}{\mathrm{D}t} + \phi\boldsymbol{\nabla}\cdot\mathbf{u}\right)\mathrm{d}V, \tag{2.4.34}$$

where $\boldsymbol{\nabla}\cdot$ is the vector divergence operator defined by

$$\boldsymbol{\nabla}\cdot\mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z},$$

for $\mathbf{F} = (F_x, F_y, F_z)$. Equations (2.4.32) and (2.4.34) are the fundamental definitions needed to make calculations involving the material derivative. In addition, by definition (2.4.32) of the material derivative of a scalar applied to ϕ, equation (2.4.34) can be

written as

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\int_V \phi \mathrm{d}V\right) = \int_V \left(\frac{\partial \phi}{\partial t} + \boldsymbol{\nabla} \cdot (\phi \mathbf{u})\right) \mathrm{d}V. \tag{2.4.35}$$

An equivalent argument can be constructed for the vector property **b**, yielding the corresponding integral

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(\int_V \mathbf{b} \mathrm{d}V\right) = \int_V \left(\frac{\partial \phi}{\partial t} + \boldsymbol{\nabla} \cdot (\mathbf{b}\mathbf{u})\right) \mathrm{d}V.$$

2.4.3 Conservation of mass and the Laplace equation

In fluid dynamics, there are essentially two equivalent ways of analysing fluid motions. The first approach deals with problems such as evaluating the velocity and density of a fluid at fixed points (or over fixed regions and volumes) in space as time passes. This is known as the field or Eulerian viewpoint. The second approach involves tracking quantities such as the position and momentum of a fixed set of a fluid parcels, each uniquely identified by some label, in time. This is known as the Lagrangian viewpoint. Simply put, from an Eulerian perspective, we consider the properties of a fluid in time within a control volume fixed in space (but the density of the fluid within the volume can change in time). From a Lagrangian perspective, we consider the properties of a fixed number of fluid parcels in time (the mass of which cannot change, but whose volume can change in time). The same equation for conservation of mass for a fluid can be derived using either of these approaches.

Here we derive the equation for conservation of mass of a fluid using the Eulerian viewpoint, for a fixed control volume V. Fluid will pass through the volume V due to its motion, potentially changing the mass of fluid in the volume. Net accumulation or dispersion of the fluid in volume V can also occur, increasing or decreasing its mass (and equivalently density) respectively. These effects must balance, quantified in terms

of the material derivative by the equation

$$\frac{\mathrm{D}}{\mathrm{D}t}\int_V \rho \,\mathrm{d}V = 0,$$

which, by equation (2.4.35) for the material derivative of a volume integral applied with $\phi = \rho$, can be written

$$\frac{\mathrm{D}}{\mathrm{D}t}\int_V \rho \,\mathrm{d}V = \int_V \left(\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla}\cdot(\rho\mathbf{u})\right) \mathrm{d}V = 0, \tag{2.4.36}$$

for density ρ, velocity $\mathbf{u}$ and time t, see e.g., Vallis (2017). Intuitively, equation (2.4.36) balances the change in density $\partial\rho/\partial t$ of the fluid in all three dimensions (x, y, z) with the flow of fluid $\boldsymbol{\nabla}\cdot(\rho\mathbf{u})$ through the volume. Equation (2.4.36) applies for an arbitrary volume element, and therefore applies anywhere in the fluid. We can therefore write

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla}\cdot(\rho\mathbf{u}) = 0. \tag{2.4.37}$$

Next, assuming that water density ρ is constant over time and space, i.e., $\partial\rho/\partial t = 0$, and so

$$\boldsymbol{\nabla}\cdot\mathbf{u} = 0. \tag{2.4.38}$$

The velocity $\mathbf{u}$ is a vector, and so can be written as

$$\mathbf{u} = \boldsymbol{\nabla}\phi + \boldsymbol{\nabla}\cdot\boldsymbol{\psi}, \tag{2.4.39}$$

for *velocity potential* ϕ and rotation $\boldsymbol{\psi}$, with gradient and divergence operators $\boldsymbol{\nabla}$ and $\boldsymbol{\nabla}\cdot$ respectively; any vector can be deconstructed in this way, see e.g., Vallis (2017). Taking the fluid to have irrotational flow (i.e., $\boldsymbol{\psi} = \mathbf{0}$), we may write $\mathbf{u} = \boldsymbol{\nabla}\phi$, and so

by substituting equation (2.4.39) into equation (2.4.38), and letting $\boldsymbol{\psi} = 0$, we have

$$\boldsymbol{\nabla} \cdot \mathbf{u} = \boldsymbol{\nabla}^2 \phi = \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} = 0. \tag{2.4.40}$$

Expression (2.4.40) is the Laplace equation (e.g., Dean and Dalrymple 1991) for irrotational flow of a fluid with constant density.

2.4.4 Conservation of momentum and the Bernoulli equation

Next, we will appeal to conservation of momentum in the fluid to derive the Bernoulli equation. Let $\mathbf{m}(x, y, z)$ be the momentum per unit volume of the fluid, where $\mathbf{m} = \rho\mathbf{u}$ (density times velocity). By Newton's second law, the rate of change in momentum in a volume V is equal to the force $\mathbf{F}$ acting on it, so we can write

$$\frac{\mathrm{D}}{\mathrm{D}t} \int_V \rho\mathbf{u} \, \mathrm{d}V = \int_V \mathbf{F} \, \mathrm{d}V, \tag{2.4.41}$$

for material derivative $\mathrm{D}/\mathrm{D}t$, where $\mathbf{F}$ is the force per unit volume. Applying the equation (2.4.34) for the material derivative of a volume integral to the left hand side,

$$\begin{aligned}
\frac{\mathrm{D}}{\mathrm{D}t} \int_V \rho\mathbf{u} \, \mathrm{d}V &= \int_V \left(\frac{\mathrm{D}\left(\rho\mathbf{u}\right)}{\mathrm{D}t} + \rho\mathbf{u}(\boldsymbol{\nabla} \cdot \mathbf{u}) \right) \mathrm{d}V, \quad \text{by (2.4.34)}, \\
&= \int_V \left(\rho \frac{\mathrm{D}\mathbf{u}}{\mathrm{D}t} + \mathbf{u} \frac{\mathrm{D}\rho}{\mathrm{D}t} + \rho\mathbf{u}(\boldsymbol{\nabla} \cdot \mathbf{u}) \right) \mathrm{d}V \\
&= \int_V \left[\rho \frac{\mathrm{D}\mathbf{u}}{\mathrm{D}t} + \mathbf{u} \left(\frac{\mathrm{D}\rho}{\mathrm{D}t} + \rho \boldsymbol{\nabla} \cdot \mathbf{u} \right) \right] \mathrm{d}V \\
&= \int_V \left[\rho \frac{\mathrm{D}\mathbf{u}}{\mathrm{D}t} + \mathbf{u} \left(\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot (\rho\mathbf{u}) \right) \right] \mathrm{d}V, \quad \text{by (2.4.33)}, \\
&= \int_V \left[\rho \frac{\mathrm{D}\mathbf{u}}{\mathrm{D}t} + \mathbf{u} \left(0 \right) \right] \mathrm{d}V, \quad \text{by (2.4.37)},
\end{aligned}$$

giving that

$$\frac{\mathrm{D}}{\mathrm{D}t} \int_V \rho\mathbf{u} \, \mathrm{d}V = \int_V \rho \frac{\mathrm{D}\mathbf{u}}{\mathrm{D}t} \mathrm{d}V.$$

So, equation (2.4.41) can be written

$$\int_V \left(\rho \frac{\mathrm{D}\mathbf{u}}{\mathrm{D}t} - \mathbf{F} \right) \mathrm{d}V = 0.$$

Since the equation applies for an arbitrary volume V, it applies anywhere in the fluid, so that

$$\frac{\mathrm{D}\mathbf{u}}{\mathrm{D}t} = \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \boldsymbol{\nabla})\mathbf{u} = \frac{\mathbf{F}}{\rho}. \tag{2.4.42}$$

Now, the applied force $\mathbf{F}$ per unit volume can be written as the sum of a pressure gradient and gravity (e.g., Vallis 2017), that is, $\mathbf{F} = -\boldsymbol{\nabla} p - \rho g \mathbf{e}_z$, for pressure p, acceleration due to gravity g and unit vector $\mathbf{e}_z = (0, 0, 1)$ in the vertical direction. Combining this with expression (2.4.42), we obtain

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \boldsymbol{\nabla})\mathbf{u} = -\frac{\boldsymbol{\nabla} p}{\rho} - g\mathbf{e}_z.$$

Further, by neglecting small perturbations in $\mathbf{u}$ and thus such terms in $\mathbf{u}^2$ and higher, we can neglect the $(\mathbf{u} \cdot \boldsymbol{\nabla})\mathbf{u}$ term and write

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\boldsymbol{\nabla} p}{\rho} + g\mathbf{e}_z = 0. \tag{2.4.43}$$

Finally, expression (2.4.43) can be used to obtain the Bernoulli equation by again considering the decomposition (2.4.39) of $\mathbf{u}$, and resulting expression $\mathbf{u} = \boldsymbol{\nabla}\phi$ for velocity potential ϕ. Substituting $\mathbf{u} = \boldsymbol{\nabla}\phi$ into expression (2.4.43), we have

$$\frac{\partial \boldsymbol{\nabla}\phi}{\partial t} + \frac{\boldsymbol{\nabla} p}{\rho} + g\mathbf{e}_z = 0,$$

or, assuming we can reverse the order of spatial and temporal differentiation of ϕ,

$$\boldsymbol{\nabla} \left(\frac{\partial \phi}{\partial t} + \frac{p}{\rho} + gz \right) = 0,$$

for vertical spatial coordinate z, yielding

$$\frac{\partial \phi}{\partial t} + \frac{p}{\rho} + gz = C,$$

for constant C, which can be taken without loss of generality to be $C = 0$. This is the Bernoulli equation, see e.g., Holthuijsen (2010).

The Bernoulli equation can be used to derive the dynamic boundary condition for the wave kinematic model. At the free surface of the fluid η, the pressure can be taken to be zero, yielding

$$\frac{\partial \phi}{\partial t} + g\eta = 0, \tag{2.4.44}$$

at $z = 0$, the mean sea level in the sea-state.

2.4.5 Kinematic boundary conditions

Here we discuss boundary conditions that the fluid kinematics must satisfy at the free surface ($z = \eta$ for free surface elevation η) and the ocean floor ($z = -d$). Firstly, a geometric argument (see e.g., Holthuijsen 2010) shows that, at the free surface

$$u_z = \frac{\partial \eta}{\partial t} + u_x \frac{\partial \eta}{\partial x} + u_y \frac{\partial \eta}{\partial y}.$$

Neglecting second and higher order times in η and $\mathbf{u}$, this equation linearises to

$$u_z = \frac{\partial \eta}{\partial t},$$

at $z = 0$, or equivalently

$$\frac{\partial \phi}{\partial z} = \frac{\partial \eta}{\partial t}, \tag{2.4.45}$$

since $\mathbf{u} = \boldsymbol{\nabla}\phi$ and so $\partial\phi/\partial z = u_z$.

Finally, we can also construct boundary conditions for the kinematics at the ocean

floor ($z = -d$). As argued by e.g., Holthuijsen (2010), there can be no flow into the ocean floor, meaning

$$u_z = \frac{\partial \phi}{\partial t} = 0, \tag{2.4.46}$$

at $z = -d$.

2.4.6 Free surface elevation and kinematics models

We can now present a solution (Holthuijsen, 2010) to the Laplace field equation (2.4.40) subject to the dynamic (2.4.44) and kinematic (2.4.45) and (2.4.46) boundary conditions, in the form of a long-crested harmonic wave propagating in the positive x-direction. The progressive wave is defined by its surface elevation $\eta(t; x)$ at time $t \in R$ and location $x \in \mathbb{R}$, given by

$$\eta(t; x) = \sum_{n=1}^{N} \left\{ A_n \cos\left(2\pi f_n t - k_n x\right) + B_n \sin\left(2\pi f_n t - k_n x\right) \right\}, \tag{2.4.47}$$

with contributing frequencies $f_n > 0$, coefficients $A_n, B_n > 0$ and wave numbers k_n, for $n = 1, \ldots N$, where N is the number of contributing frequencies. The coefficients A_n and B_n are randomly distributed as

$$A_n|\{\mathbf{X} = \mathbf{x}\}, B_n|\{\mathbf{X} = \mathbf{x}\} \sim N(0, \Delta_f S(f_n; \mathbf{x})),$$

for $n = 1, \ldots, N$, where $\mathbf{x}$ is the chosen value of sea state $\mathbf{X} = (H_s, S_e)$, $S(\cdot; \mathbf{x})$ is the wave spectrum, and Δ_f is the width of the frequency band (or bin width) in the discretised form of the wave spectrum. The wave spectrum $S(\cdot; \mathbf{x})$ can be specified by a known function, e.g., the JONSWAP spectrum (Hasselmann et al., 1973), with parameters dependent on the desired values of the sea state parameters H_s and S_e. Equation (2.4.47) can therefore be used to model individual wave elevation within a particular sea state, see Chapter 3 Section 3.7 for further details.

Equation (2.4.47) for the wave surface elevation has corresponding velocity potential function

$$\phi(t;x,z)=\sum_{n=1}^{N}\left\{\frac{2\pi f_n\cosh[k_n(d+z)]}{k_n\sinh(k_nd)}\left[-A_n\sin(2\pi f_nt-k_nx)+B_n\cos(2\pi f_nt-k_nx)\right]\right\}, \tag{2.4.48}$$

at time $t\in\mathbb{R}$ and location $(x,z)\in\mathbb{R}^2$ (Holthuijsen, 2010). As discussed in Section 2.4.3, the velocity of the wave is given by $\mathbf{u}=\boldsymbol{\nabla}\phi$, meaning differentiation of the velocity potential (2.4.48) with respect to x and z yields equations

$$u_x(t;x,z)=\sum_{n=1}^{N}\left[\left\{A_n\cos(2\pi f_nt-k_nx)+B_n\sin(2\pi f_nt-k_nx)\right\}2\pi f_n\frac{\cosh\left[k_n(d+z)\right]}{\sinh(k_nd)}\right],$$

$$u_z(t;x,z)=\sum_{n=1}^{N}\left[\left\{-A_n\sin(2\pi f_nt-k_nx)+B_n\cos(2\pi f_nt-k_nx)\right\}2\pi f_n\frac{\sinh\left[k_n(d+z)\right]}{\sinh(k_nd)}\right],$$

for horizontal (in the x-direction) and vertical (in the z-direction) wave velocities u_x and u_z at time $t\in\mathbb{R}$ and location $(x,z)\in\mathbb{R}^2$. Furthermore, similar differentiation of the above velocity equations yields,

$$\dot{u}_x(t;x,z)=\sum_{n=1}^{N}\left[\left\{-A_n\sin(2\pi f_nt-k_nx)+B_n\cos(2\pi f_nt-k_nx)\right\}2\pi f_n^2\frac{\cosh\left[k_n(d+z)\right]}{\sinh(k_nd)}\right],$$

$$\dot{u}_z(t;x,z)=\sum_{n=1}^{N}\left[\left\{-A_n\cos(2\pi f_nt-k_nx)-B_n\sin(2\pi f_nt-k_nx)\right\}2\pi f_n^2\frac{\cosh\left[k_n(d+z)\right]}{\sinh(k_nd)}\right],$$

for horizontal (in the x-direction) and vertical (in the z-direction) wave accelerations $\dot{u}_x$ and $\dot{u}_z$ at time $t\in\mathbb{R}$ and location (x,z).

Finally, by substituting equation (2.4.47) for the surface elevation and equation

(2.4.48) for the velocity potential into dynamic boundary condition (2.4.44), it can be shown (Holthuijsen, 2010) that the wave number k_n and frequency f_n, $n = 1, \ldots, N$, are related by the expression

$$(2\pi f_n)^2 = g k_n \tanh(k_n d), \tag{2.4.49}$$

for water depth d and acceleration due to gravity g. The relationship (2.4.49) is known as the dispersion relation (e.g., Holthuijsen 2010). By solving equation (2.4.49) implicitly for contributing frequencies $f_1, \ldots, f_N$, each with coefficients $A_1, \ldots, A_N, B_1, \ldots, B_N$ corresponding to the specified values of (H_s, S_e), we can use the above equations to obtain surface elevation η and kinematics $\mathbf{u} = (u_x, u_z)$ and $\dot{\mathbf{u}} = (\dot{u}_x, \dot{u}_z)$ of waves within a sea state with given sea state parameters (H_s, S_e).

Chapter 3

Estimating Met-Ocean Extremes Associated with Extreme Structural Response

3.1 Introduction

3.1.1 Background

Ocean engineers use different approaches to quantify extreme conditions for design and reassessment of offshore and coastal structures. The natural full probabilistic approach (henceforth, the "forward" approach) is to construct a sequence of statistical models to characterise the extreme multivariate ocean environment, as well as the interaction between that environment and the structure (e.g., Towe et al. 2021). This approach considers the response of the structure to be a stochastic function of the environment, summarised by underlying sea state statistics, such as significant wave height and period, contrary to previous work (e.g., Coles and Tawn 1994) where a deterministic relationship is assumed. The forward approach thus seeks a multivariate distribution $F_{\mathbf{X}}$ for environmental variables $\mathbf{X}$, as well as a distribution $F_{R_L|\mathbf{X}}$, which characterises

the maximum stochastic response R_L induced on the structure by the environment $\mathbf{X}$ over the period of a sea state of duration L (e.g., 3) hours. A key property of this method is that uncertainty from the estimation of each distribution can be naturally quantified and propagated through the sequence of models. Structural risk assessment centres on the estimation of the probability of structural failure using the distribution F_{R_S} for the maximum response R_S in a random storm, or alternatively the distribution F_{R_A} for the maximum response R_A per annum, evaluated by marginalisation of the distribution $F_{R_L|\mathbf{X}}$ over the environment space. In addition, $F_{R_L|\mathbf{X}}$ may be used to find the probability of the event of structural failure $\mathcal{F}_L$ within a sea state (of duration L) with environmental conditions $\mathbf{X}$.

A combination of better models for the extreme ocean environment, techniques to reduce the computational complexity of the forward approach, and improved computational resources, have made forward estimation of structural failure probability more routinely achievable. These include the development of practically-useful statistical models for non-stationary, or covariate effected, margins (e.g., Chavez-Demoulin and Davison 2005; Randell et al. 2016; Youngman 2019b) as well as the conditional multivariate extremes model developed by Heffernan and Tawn (2004), conditional simulation of extreme time series (of waves and wave kinematics) proposed by Taylor et al. (1997), and efficient importance sampling from distributions (e.g., Gelman et al. 2013). In this work we demonstrate an efficient forward approach to estimate the tail of the distributions $F_{R_L|\mathbf{X}}$ and F_{R_S}, corresponding to return periods P of the order of 10^3 years. Specifically, we estimate the distribution of extreme base shear, approximated using the loading equation of Morison et al. (1950) on simple structures, following the procedure of Tromans and Vanderschuren (1995). The environmental and structural models utilised here are sufficiently complex to illustrate key methodological steps, whilst being simple enough to avoid unnecessary complexity. We will also estimate the joint conditional density of environmental variables (henceforth referred to as CDE_P)

given the occurrence of an extreme P-year Morison load on example structures. CDE_P is introduced as a diagnostic tool *emerging from* fully probabilistic analysis, highlighting regions of environmental space associated with extreme structural responses. We use CDE_P to assess the relevance of other approaches that aim to identify regions of the environment space important to design.

Historically, forward estimation of the probability of failure has proved computationally intractable or prohibitively expensive. Instead, metocean design has tended to focus on a dominant variable (such as significant wave height) at a location, placing less emphasis on other associated environmental variables (e.g., Feld et al. 2015); the metocean engineer's challenge is then to estimate marginal return values of the dominant variable, and perhaps follow an engineering recipe to specify design values for associated variables. The development of methods of structural reliability, associated with Madsen et al. (1986), made it clearer that good models for the joint distribution of environmental variables were necessary. For this reason, approaches to structural reliability that make use of more than one environmental variable became popular, such as environmental design contours derived from parametric hierarchical models of the environmental variables $\mathbf{X}$ (e.g., Haver 1987). A recent review of statistical methodologies for metocean design is provided by Vanem et al. (2022).

In one respect, contour methods are advantageous over the forward approach in that they characterise the environment only and so just require estimation of the joint environmental distribution $F_{\mathbf{X}}$. Therefore, combined with appropriate assumptions for the nature of the ocean-structure interaction, environmental contours can be used in principle to assess any structure in that environment. The decoupling of environment and structure is achieved by making what are thought to be conservative assumptions about the environment-structure interaction, leading to what are believed to be conservative estimates of structural reliability. Many different methods exist to estimate environmental design contours (e.g., Ross et al. 2020; Haselsteiner et al. 2021; Mackay

and Haselsteiner 2021; Hafver et al. 2022; Mackay and de Hauteclocque 2023). These methods, whilst each attempting to construct the P-year level contour, can produce quite different estimates to one another. In addition, the environmental contour produced by a particular contour estimation method will vary with the estimate of $F_{\mathbf{X}}$ used to obtain it. In this work, we consider the IFORM environmental contour method (e.g., Haver and Winterstein 2009), recommended by both NORSOK N-003 (2017) and DNVGL-RP-C205 (2017) standards. We focus on IFORM due to its popularity in the ocean engineering community. We will assess the relative performance of different hierarchical models for $F_{\mathbf{X}}$ and their respective IFORM environmental contours, by quantifying their capacity to enclose CDE_P for a given structure. We will show that, regardless of the hierarchical model form for $F_{\mathbf{X}}$, environmental contours do not provide reliable coverage of CDE_P across a set of simple structure examples. Therefore, evaluation of response along these IFORM contour boundaries will not provide reliable realisations of the desired P-year response.

3.1.2 Objectives and Layout

The objective of the chapter is to promote the use of fully probabilistic design in favour of environmental contour-based methods. To achieve this, we provide the following analysis. (a) We demonstrate that fully probabilistic design (using the "forward approach") can be achieved in a computationally efficient manner. (b) Using the forward approach, we estimate the conditional density of the environment (CDE) across different example structures. We view CDE as a design diagnostic which identifies regions of the space of environmental variables contributing to extreme structural responses. We observe that this region changes from structure to structure. (c) We highlight the deficiencies of environmental contour methods for design, by assessing their ability to characterise the CDE. Specifically, we show that (c-i) a specified approach to environmental contour estimation is not suitable to characterise CDEs corresponding to

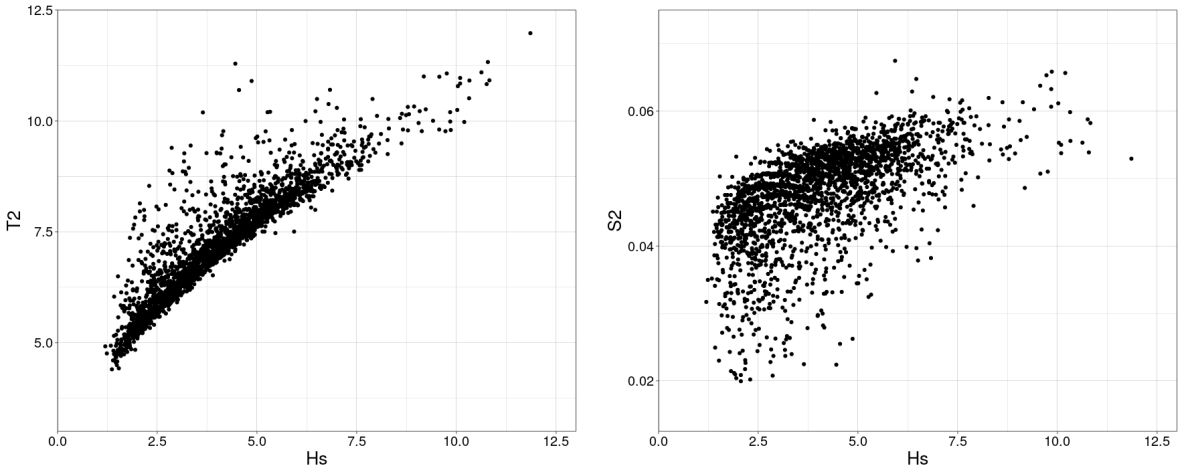
different structures (because the environmental contour is structure-independent), and (c-ii) for reasonably-sized samples, the characteristics of the estimated environmental contour are sensitive to the modelling choices underpinning contour estimation, and that making these choices well is challenging. Our findings from (c) further motivate rejection of contour-based design in favour of the fully probabilistic alternative. Further, if contour-based methods are to be used, they should be calibrated for the specific structural archetype under consideration. To establish this calibration, the full forward model would nevertheless need first to be evaluated for the archetype.

The layout of the chapter is as follows. In Section 3.2, we seek to motivate our analysis using a sample of data for storm peak significant wave height and second spectral moment wave period from a location in the central North Sea. Section 3.3 describes the methodologies combined to achieve the efficient forward approach for estimation of the distributions F_{R_S} and F_{R_A}. The approach used to estimate environmental design contours, and the various parametric forms considered for the hierarchical estimation of $F_{\mathbf{X}}$ are given in Section 3.4. In Section 3.5, we present estimates of CDE_P for three variants of a simple stick structure, and use these to quantify the performance of different IFORM environmental contours, demonstrating that none of the IFORM contours estimates performs well for all example structures. In Section 3.5.4, we discuss the implications of our results, and make recommendations for structural design practice. We provide a Supplementary Material (SM) in Section 3.7 with a fuller description of aspects of the procedures above.

3.2 Motivating metocean dataset

We motivate the analysis using hindcast data for sea state significant wave height and second spectral moment wave period for a location in the central North Sea. The data consist of 124671 observations for the period January 1979 to September 2013,

calculated for consecutive 3-hour sea states. Intervals corresponding to storm events are isolated from the hindcast data, using the approach of Ewans and Jonathan (2008), resulting in a total of 2462 values for storm peak significant wave height (H_S) and corresponding wave period (T_2), for an average of 73 storm events per annum. Figure 3.2.1a shows the storm peak data (H_S, T_2). Despite the fact that we expect these variables not to be identically distributed due to environmental covariates (e.g., direction and season), for the purposes of the current work we assume these to be independently and identically distributed. We choose to use storm peak wave steepness S_2 in favour of



(a) Storm peak variables (H_S, T_2). (b) Storm peak variables (H_S, S_2).

Figure 3.2.1: Data of storm peak sea state variables from a location in the central North Sea. Storm peaks are extracted from hindcast data of consecutive 3-hour sea states, using the method of Ewans and Jonathan (2008), in terms of significant wave height H_S [m], second spectral moment wave period T_2 [s] and wave steepness S_2.

wave period T_2 in the analysis below because extreme events generally correspond to large values of H_S and S_2. This simplifies joint modelling of extreme environments when compared to joint modelling of H_S, and T_2, because it is generally small values of T_2 which correspond to extreme events, and multivariate extreme value models are not typically designed to model simultaneously small and large variables. Values of S_2 are calculated via $S_2 = 2\pi H_S/(gT_2^2)$, where g is the acceleration due to gravity. Figure 3.2.1b shows the resulting storm peak data $\mathbf{X} = (H_S, S_2)$. Note throughout the

chapter, that we restrict the notation H_S, S_2 and T_2 to refer to storm peak quantities, and further that all numerical values are quoted in SI units. Additionally, all physical properties are taken to be real-valued unless stated otherwise.

The choice of (H_S, S_2) over (H_S, T_2) is motivated by the fact that extreme value models are generally developed for the joint upper tail of variables. Extreme environmental loads are typically generated by large values of H_S, large values of S_2 but non-extreme values of T_2; that is, joint extremes of (H_S, S_2) in the upper tail induce the most extreme environments and structural responses. Therefore, it is appealing to structure the analysis in terms of (H_S, S_2) (e.g., Myrhaug 2018).

3.3 Environment and response modelling

3.3.1 Outline of the forward approach and CDE_P

Storm peak characteristics and intra-storm evolution

We illustrate the methodologies combined to form the forward approach for direct estimation of the distribution of extreme structural response (specifically, base shear), in order to estimate the conditional distribution of the environment CDE_P. For generality, and alignment with the work of others (e.g., Towe et al. 2021), initially we present a form of the forward approach that incorporates full intra-storm evolution and the effects of covariates. We subsequently restrict the model, and focus our analysis on storm peak variables as introduced in Section 3.2, so as to emphasise the key methodological steps only.

We assume access to metocean data for storm peak variables $\mathbf{X}^{\text{sp}}$ (e.g., H_S and S_2 from Section 3.2). Since the characteristics of storm peak events generally vary with respect to covariates (e.g., Randell et al. 2015) we also assume access to storm peak covariates $\boldsymbol{\Theta}^{\text{sp}}$, which encode this information. The response of an offshore structure to environmental loading occurs continuously, and in particular for the full duration of

the storm event, consisting of a series of sea states (of duration L hours) indexed by $s \in \mathcal{S}_T = \{1, 2, \ldots, T\}$, for unknown storm length T. To estimate the distribution of the maximum structural response R_S in a storm, we need to consider (a) the variability of the duration T of a random storm, (b) the full evolution of sea state variables $\{\mathbf{X}_s\}_{s\in\mathcal{S}_T}$ over a given storm (themselves not identically distributed given the sea state covariates $\{\boldsymbol{\Theta}_s\}_{s\in\mathcal{S}_T}$) and (c) the resulting maximum response $R_S|\{(\mathbf{X}_s, \boldsymbol{\Theta}_s)\}_{s\in S_T}$ induced over sea states within the storm.

As sea states are dependent over time, $R_{L,s}$ and $R_{L,s'}$, for any $s, s' \in \mathcal{S}_T$ where $s \neq s'$, will also be dependent, where $R_{L,s}$ is the maximum response in a sea state of length L at time index s. However, this dependence is purely due to the sea state characteristics $\{(\mathbf{X}_s, \boldsymbol{\Theta}_s)\}_{s\in\mathcal{S}_T}$ evolving over time. This is because a load $R_{L,s}$ from the model of Morison et al. (1950) is produced by an individual random wave event, and the statistical properties of wave events within a sea state at time index s are determined by the characteristics $(\mathbf{X}_s, \boldsymbol{\Theta}_s)$, yet the interval of time over which consecutive waves are correlated is considerably shorter than the length L of a sea state. Therefore, the conditional dependence between the variables $R_{L,s}$ and $R_{L,s'}$, given $(\mathbf{X}_s, \boldsymbol{\Theta}_s)$ and $(\mathbf{X}_{s'}, \boldsymbol{\Theta}_{s'})$, is negligible. We exploit this reasoning to facilitate step (c), where we assume that the random variables $\{R_{L,s}\}_{s\in\mathcal{S}_T}$ are conditionally independent given sea states $\{(\mathbf{X}_s, \boldsymbol{\Theta}_s)\}_{s\in\mathcal{S}_T}$, which leads to the following simplification, for response $r > 0$

$$F_{R_S|\{(\mathbf{X}_s,\boldsymbol{\Theta}_s)\}_{s\in S_T}}(r|\{(\mathbf{x}_s, \boldsymbol{\theta}_s)\}_{s\in S_\tau}) = \prod_{s\in\mathcal{S}_T} F_{R_L|(\mathbf{X}_s,\boldsymbol{\Theta}_s)}(r|\mathbf{x}_s, \boldsymbol{\theta}_s), \tag{3.3.1}$$

where we omit the subscript s in writing $F_{R_L|(\mathbf{X}_s,\boldsymbol{\Theta}_s)}$ since the dependence of R_L on s is contained through the vales of $(\mathbf{X}_s, \boldsymbol{\Theta}_s)$.

Distribution of maximum response per storm and per annum

The forward approach estimates the cumulative distribution function F_{R_S} of the maximum structural response R_S in a storm, and subsequently the distribution of maximum response R_A per annum. Estimation of F_{R_S} requires (a) modelling of multivariate storm peak variables $\mathbf{X}^{\text{sp}}$ given storm peak covariates $\boldsymbol{\Theta}^{\text{sp}}$, (b) characterisation of the conditional time-varying within-storm evolution of sea state characteristics $\{(\mathbf{X}_s, \boldsymbol{\Theta}_s)\}_{s\in\mathcal{S}_T}$ given storm peak characteristics $\mathbf{X}^{\text{sp}}, \boldsymbol{\Theta}^{\text{sp}}$, and (c) the estimation of the maximum response $R_{L,s}$ given sea state characteristics $(\mathbf{X}_s, \boldsymbol{\Theta}_s)$ for an L-hour sea state. We must also consider the variability in the duration T of a storm event. See Towe et al. (2021) for previous discussions of similar models. Given knowledge of the above, and exploiting assumption (3.3.1) made in Section 3.3.1, the distribution of R_S can be written as

$$\begin{aligned} F_{R_S}(r) = \int_{\boldsymbol{\theta}^{\text{sp}}} \int_{\mathbf{x}^{\text{sp}}} \int_{(\{(\mathbf{x}_s,\boldsymbol{\theta}_s)\}_{s\in\mathcal{S}_\tau},\tau)} \prod_{s\in\mathcal{S}_\tau} F_{R_L|(\mathbf{X}_s,\boldsymbol{\Theta}_s)}(r|\mathbf{x}_s,\boldsymbol{\theta}_s) \\ \times f_{(\{(\mathbf{X}_s,\boldsymbol{\Theta}_s)\}_{s\in\mathcal{S}_T},T)|\mathbf{X}^{\text{sp}},\boldsymbol{\Theta}^{\text{sp}}}\left(\{(\mathbf{x}_s,\boldsymbol{\theta}_s)\}_{s\in\mathcal{S}_\tau}, \tau \mid \mathbf{x}^{\text{sp}}, \boldsymbol{\theta}^{\text{sp}}\right) \\ \times f_{\mathbf{X}^{\text{sp}}|\boldsymbol{\Theta}^{\text{sp}}}(\mathbf{x}^{\text{sp}}|\boldsymbol{\theta}^{\text{sp}}) \times f_{^{\text{sp}}}(\boldsymbol{\theta}^{\text{sp}})\, \mathrm{d}(\{(\mathbf{x}_s,\boldsymbol{\theta}_s)\}_{s\in\mathcal{S}_\tau},\tau)\, \mathrm{d}\mathbf{x}^{\text{sp}}\, \mathrm{d}\boldsymbol{\theta}^{\text{sp}}, \end{aligned} \tag{3.3.2}$$

for $r > 0$, where $f_{\boldsymbol{\Theta}^{\text{sp}}}$ is the joint probability density of storm peak covariates, $f_{\mathbf{X}^{\text{sp}}|\boldsymbol{\Theta}^{\text{sp}}}$ is the joint probability density of storm peak variables given storm peak covariates, and $f_{(\{(\mathbf{X}_s,\boldsymbol{\Theta}_s)\}_{s\in\mathcal{S}_T},T)|\mathbf{X}^{\text{sp}},\boldsymbol{\Theta}^{\text{sp}}}$ is the joint probability density for the full time-series evolution of within-storm sea state characteristics (and storm duration) given storm peak characteristics.

If we assume that the number of occurrences of storm events in a year is Poisson-distributed with expectation λ per annum, we can use F_{R_S} to estimate the corresponding cumulative distribution function F_{R_A} of the maximum response R_A in a year, i.e.,

$$F_{R_A}(r) = \sum_{m=0}^{\infty} [F_{R_S}(r)]^m \frac{\lambda^m e^{-\lambda}}{m!} = \exp\left[-\lambda\left(1 - F_{R_S}(r)\right)\right],$$

for $r > 0$. From this expression, we may define return values for maximum response corresponding to a return period of P years as $r_P = F_{R_A}^{-1}(1 - 1/P)$.

Reduced forward approach

The objective of the current work is to compare estimates for CDE_P made using the forward approach, summarised by (3.3.2), with environmental contours at the P-year level. This comparison is useful even in the absence of covariate effects, and whilst neglecting intra-storm evolution. Therefore, to minimise computational complexity, we now assume that the effects of covariates $\boldsymbol{\Theta}^{\text{sp}}$ and $\boldsymbol{\Theta}_s$ can be ignored, and that the maximum response in a storm always occurs in the storm peak sea state, so that intra-storm evolution can also be ignored. As a result, integral (3.3.2) for F_{R_S} reduces to

$$F_{R_S}(r) = \int_{\mathbf{x}^{\text{sp}}} F_{R_L|\mathbf{X}^{\text{sp}}}(r|\mathbf{x}^{\text{sp}}) f_{\mathbf{X}^{\text{sp}}}(\mathbf{x}^{\text{sp}}) \mathrm{d}\mathbf{x}^{\text{sp}},$$

for $r > 0$, where $F_{R_L|\mathbf{X}^{\text{sp}}}$ is the cumulative distribution function of the maximum response over an L-hour sea state given storm peak variables $\mathbf{X}^{\text{sp}}$. For brevity, we henceforth omit the storm peak superscript and write

$$F_{R_S}(r) = \int_{\mathbf{x}} F_{R_L|\mathbf{X}}(r|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \mathrm{d}\mathbf{x}, \tag{3.3.3}$$

for $r > 0$, where $f_{\mathbf{X}}$ is now the joint density of storm peak variables $\mathbf{X} = (X_1, \ldots, X_p)$, and $F_{R_L|\mathbf{X}}$ is the distribution of maximum response over an L-hour storm peak sea state with variables $\mathbf{X}$, where the value of L is determined by the duration of sea states in the data being used for inference. This brings us back to the setting of Section 3.2, where we introduced observations of storm peak sea state data $\mathbf{X}$ with $p = 2$. The key steps in evaluating the reduced forward approach in (3.3.3) therefore become the estimation of $f_{\mathbf{X}}$ and $F_{R_L|\mathbf{X}}$. The first of these inferences is achieved using the conditional extremes model of Heffernan and Tawn (2004), as described in Section 3.3.2. The second

inference involves conditional simulation of environmental time-series following Taylor et al. (1997), and importance sampling from the tail of $F_{R_L|\mathbf{X}}$ following Towe et al. (2021), as described in Section 3.3.3.

Probability of structural failure

When designing offshore structures, it is often desirable to determine the probability of the event of structural failure $\mathcal{F}_L$ for a given L-hour sea state with variables $\mathbf{X} = \mathbf{x}$. Given an estimate for the distribution of $R_L|(\mathbf{X} = \mathbf{x})$, we can evaluate this probability using the expression

$$\Pr\left(\mathcal{F}_L|\mathbf{X} = \mathbf{x}\right) = \int_{r_L} \Pr\left(\mathcal{F}_L|R_L = r_L\right) f_{R_L|\mathbf{X}}(r_L|\mathbf{x})\,\mathrm{d}r_L,$$

where the failure probability $\Pr\left(\mathcal{F}_L|R_L = r_L\right)$ depends on the nature of the structure, as illustrated in Figure 3.5.6 in Section 3.5.

Conditional density of the environment

The conditional density of the environment $f_{\mathbf{X}|R_L}(\cdot|r_P)$ describes the joint density of the environmental variables $\mathbf{X}$, conditional on the appearance of a P-year maximum response r_P within a sea state of length L hours. Given estimates for $f_{\mathbf{X}}$, $F_{R_L|\mathbf{X}}$, F_{R_L} and r_P, CDE_P can therefore be evaluated using Bayes' rule

$$f_{\mathbf{X}|R_L}(\mathbf{x}|r_P) = \frac{f_{R_L|\mathbf{X}}(r_P|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{f_{R_L}(r_P)},$$

for $\mathbf{x} \in \mathbb{R}^p$, where $f_{R_L|\mathbf{X}}$ and f_{R_L} are the densities corresponding to the distributions $F_{R_L|\mathbf{X}}$ and F_{R_L} respectively. Examples of CDE_P for the central North Sea application are given in Section 3.5.3.

3.3.2 Joint modelling of storm peak conditions

Outline of the conditional extremes model

The upper extremes of the marginal and joint distributions of the environmental variables $\mathbf{X} = (X_1, \ldots, X_p)$, corresponding to the storm peak sea state, are described using the conditional extremes model of Heffernan and Tawn (2004). For our illustrative example, in Section 3.2, $p = 2$, but we present the methodology here and for environmental contours for dimension p, to cover more general cases. This asymptotically justified flexible framework allows for the characterisation of joint tail behaviour from a sample of independent identically distributed observations of $\mathbf{X}$, without the need for making a subjective choice for a particular form of extremal dependence model (copula) between variables. This method has been applied extensively to the modelling of oceanographic data (e.g., Jonathan et al. 2014b; Towe et al. 2019; Shooter et al. 2021b; Tendijck et al. 2023a).

The conditional extremes method uses univariate extreme value techniques to characterise the distribution of each variable individually, with the joint structure specified for variables on standard (typically Laplace) marginal scales (e.g., Keef et al. 2013). Estimation of the conditional extremes model is thus performed in two stages: (a) marginal extreme value modelling of each variable X_j $(j = 1, \ldots, p)$ in turn, followed by the marginal transformation $X_j \mapsto Y_j$ of each variable to standard Laplace scale Y_j and (b) estimation of the conditional extremes model for the set of Laplace-scale variables $\mathbf{Y} = (Y_1, \ldots, Y_p)$. Subsequently, we estimate $f_{\mathbf{X}}$ within an extreme joint tail region using the fitted conditional extremes model. The above steps are discussed below.

Marginal modelling and marginal transformation to Laplace scale

We adopt the approach of Davison and Smith (1990) for marginal modelling of storm peak variable X_j for $j = 1, \ldots, p$. We fit a generalised Pareto distribution (GPD) to exceedances of high threshold u_j, and model threshold non-exceedances empirically.

Our marginal model F_{X_j} for the cumulative distribution function of X_j can thus be written

$$F_{X_j}(x) = \begin{cases} \tilde{F}_{X_j}(x) & x \leq u_j \\ \tilde{F}_{X_j}(u_j) + \{1 - \tilde{F}_{X_j}(u_j)\} F_{\text{GPD},j}(x; u_j, \sigma_j, \xi_j) & x > u_j, \end{cases} \tag{3.3.4}$$

where $\tilde{F}_{X_j}$ is the empirical distribution of X_j and

$$F_{\text{GPD},j}(x; u_j, \sigma_j, \xi_j) = 1 - \left(1 + \frac{\xi_j(x - u_j)}{\sigma_j}\right)_+^{-1/\xi_j},$$

for $x > u_j$, with scale and shape parameters $\sigma_j > 0$ and $\xi_j \in \mathbb{R}$ and where $y_+ = \max(y, 0)$ for $y \in \mathbb{R}$. The values of u_j $(j = 1, \ldots, p)$ are selected using the univariate extreme threshold selection methods summarised by Coles (2001), see SM. Parameters σ_j and ξ_j are jointly estimated using standard maximum likelihood techniques. The probability integral transform

$$Y_j = \begin{cases} \log\left\{2F_{X_j}(X_j)\right\} & \text{for } X_j < F_{X_j}^{-1}(0.5) \\ -\log\left\{2\left[1 - F_{X_j}(X_j)\right]\right\} & \text{for } X_j > F_{X_j}^{-1}(0.5), \end{cases}$$

is then applied to each variable in turn, obtaining the standard Laplace-scaled equivalents $\mathbf{Y}$ of $\mathbf{X}$, such that, for $j = 1, \ldots, p$, the distribution of Y_j is

$$F_{Y_j}(y) = \begin{cases} \frac{1}{2}\exp(y) & y \leq 0 \\ 1 - \frac{1}{2}\exp(-y) & y > 0. \end{cases}$$

Joint dependence modelling

Having transformed environmental variables $\mathbf{X}$ to standard Laplace-scale equivalents $\mathbf{Y}$, we now apply the model of Heffernan and Tawn (2004) to estimate the joint distribution

$F_{\mathbf{Y}}$ in the upper tail region. It is shown by Heffernan and Tawn (2004) and Keef et al. (2013) that, for $j = 1, \ldots, p$, there exist unique values of parameters $\boldsymbol{\alpha}_{|j} \in [-1, 1]^{p-1}$, $\boldsymbol{\beta}_{|j} \in (-\infty, 1]^{p-1}$, satisfying the constraints of Keef et al. (2013), and $\mathbf{z}_{|j} \in \mathbb{R}^{p-1}$, $y > 0$, such that

$$\lim_{v_j \to \infty} \Pr\left(\frac{\mathbf{Y}_{-j} - \boldsymbol{\alpha}_{|j} Y_j}{Y_j^{\boldsymbol{\beta}_{|j}}} < \mathbf{z}_{|j}, Y_j - v_j > y | Y_j > v_j \right) = e^{-y} G_{|j}(\mathbf{z}_{|j}), \tag{3.3.5}$$

where the $(p-1)$-vector $\mathbf{Y}_{-j}$ denotes the p-vector $\mathbf{Y}$ with jth element Y_j removed, $G_{|j}$ is a $(p-1)$-dimensional distribution function with non-degenerate marginals, and componentwise operations are assumed. Property (3.3.5) can be leveraged by assuming a non-linear regression of $\mathbf{Y}_{-j}$ onto Y_j holds for all values of $\mathbf{Y}$ within the region $\{\mathbf{Y} \in \mathbb{R}^p : Y_j > v_j\}$, for some suitably large finite threshold $v_j > 0$. For conditioning variable Y_j $(j = 1, \ldots, p)$, the form of this regression is

$$(\mathbf{Y}_{-j} | \{Y_j = y\}) = \boldsymbol{\alpha}_{|j} y + y^{\boldsymbol{\beta}_{|j}} \mathbf{Z}_{|j}, \quad y > v_j, \tag{3.3.6}$$

where $\mathbf{Z}_{|j} \sim G_{|j}$ is a $(p-1)$-dimensional residual random variable that is independent of Y_j given $Y_j > v_j$. We estimate parameter vectors $\boldsymbol{\alpha}_{|j}$ and $\boldsymbol{\beta}_{|j}$ using standard maximum likelihood techniques, assuming for model fitting only that $G_{|j}$ corresponds to independent Gaussian distributions with unknown means and variances. The distribution $G_{|j}$ is modelled via the kernel density estimate of the observed values of the $(p-1)$-dimensional residual

$$\mathbf{Z}_{|j} = \frac{\mathbf{Y}_{-j} - \boldsymbol{\alpha}_{|j} Y_j}{Y_j^{\boldsymbol{\beta}_{|j}}}, \quad \text{for } Y_j > v_j,$$

as in Winter and Tawn (2017).

Simulation under the conditional model and estimation of the environment joint density

Inferences in $\mathbb{R}^p$ using the fitted conditional extremes model are typically made by careful combination of Laplace-scale simulations in each of the upper tail regions $\{\mathbf{Y} \in \mathbb{R}^p : Y_j > v_j\}$, for $j = 1, \ldots, p$, together with empirical estimation in the remaining region $\{\mathbf{Y} \in \mathbb{R}^p : Y_j \leq v_j \, \forall j\}$, as described in Heffernan and Tawn (2004), to give a set of size N_{sim} realisations from the estimate of the joint density $f_{\mathbf{Y}}$. The p fitted marginal models (3.3.4) can then be used componentwise to transform this sample of $\mathbf{Y}$ with Laplace-scale marginals to a sample of $\mathbf{X}$ on the original physical scale.

We can use the simulated sample to estimate the probability of $\mathbf{X}$ being in subregions of $\mathbb{R}^p$. Specifically, if D is the set of feasible $\mathbf{X}$ values such that $\Pr(\mathbf{X} \in \mathbb{R}^p \setminus D) = 0$, then we consider a partition $(D_1, \ldots, D_M)$ of D. Then, if $N_{\text{sim},i}$ is the number of realisations in set D_i, we can estimate $\Pr(\mathbf{X} \in D_i)$, for any $i = 1, \ldots, M$, as $\Pr(\mathbf{X} \in D_i) = N_{\text{sim},i}/N_{\text{sim}}$. To obtain an estimate $f_{\mathbf{X}}(\mathbf{x})$ for any $\mathbf{x} \in D$ we exploit the property that, if $\mathbf{x} \in D_i$ and $|D_i|$ is sufficiently small and $f_{\mathbf{X}}$ is reasonably constant for all $\mathbf{x} \in D_i$, then

$$\Pr(\mathbf{X} \in D_i) = \int_{\mathbf{x}' \in D_i} f_{\mathbf{X}}(\mathbf{x}') \mathrm{d}\mathbf{x}' \approx |D_i| f_{\mathbf{X}}(\mathbf{x}),$$

yielding the estimate $\hat{f}_{\mathbf{X}}(\mathbf{x}) = N_{\text{sim, i}}/(N_{\text{sim}}|D_i|)$ for all $\mathbf{x} \in D_i$. We can achieve the required conditions for the approximation to be reliable by taking M to be sufficiently large and selecting all D_i such that $|D_i| \propto M^{-1}$.

3.3.3 Estimation of maximum response in a storm peak sea state given storm peak variables

Outline of estimation of $F_{R_L|\mathbf{X}}$

To derive properties of R_L we first need to model the behaviour of the maximum response R_I to an individual wave in sea state $\mathbf{X}$. However, to estimate the distribution $F_{R_I|\mathbf{X}}$ of the maximum response due to the action of an individual wave in sea state $\mathbf{X}$, we first evaluate the distribution $F_{R_I|(\mathbf{X},C)}$ of the maximum response R_I to an individual wave in the sea state $\mathbf{X}$ with crest elevation C, so that we can efficiently obtain distributions of the response given the full range of crest elevation C, for a particular sea state $\mathbf{X}$. This is achieved by simulation of wave fields under sea state conditions $\mathbf{X}$ with known crest elevation C, followed by propagation of the resulting stochastic wave fields through to the structural response model; see later in this Section for details. Then, integrating out C, we have

$$F_{R_I|\mathbf{X}}(r|\mathbf{x}) = \int_{\mathbb{R}^+} F_{R_I|\mathbf{X},C}(r|\mathbf{x},c) f_{C|\mathbf{X}}(c|\mathbf{x}) \mathrm{d}c, \tag{3.3.7}$$

for $r > 0$, where $f_{C|\mathbf{X}}$ is the density of crest elevation in the sea state $\mathbf{X}$, where we assume that crests are Rayleigh-distributed, with density

$$f_{C|\mathbf{X}}(c|\mathbf{x}) = \frac{16c}{h(\mathbf{x})^2} \exp\left(-8\frac{c^2}{h(\mathbf{x})^2}\right), \tag{3.3.8}$$

for $c > 0$, with sea state $\mathbf{X} = \mathbf{x}$ with significant wave height $h(\mathbf{x})$. Computationally efficient estimation of $F_{R_I|\mathbf{X}}$ following (3.3.7) is achieved using importance sampling; see later in this Section. Finally, we obtain $F_{R_L|\mathbf{X}}$ from $F_{R_I|\mathbf{X}}$ by assuming that (a) there are a fixed number of waves per L-hour sea state $\mathbf{x}$, given by $Q_L(\mathbf{x}) = 60^2 L/t_2(\mathbf{x})$ where $t_2(\mathbf{x})$ is the second spectral moment wave period for the sea state, and (b) individual-wave maximum responses (i.e., the R_I) in a given sea state are independent of each

other. Assumption (a) approximates the stochastic number of waves per sea state with an 'average' value, and (b) holds since individual base shears calculated via numerical simulation are observed for fractions of a second, significantly less than the typical wave period; therefore, there is no correlation between responses induced by different waves for a known sea state. Combining these assumptions gives

$$F_{R_L|\mathbf{X}}(r|\mathbf{x}) = \{F_{R_I|\mathbf{X}}(r|\mathbf{x})\}^{Q_L(\mathbf{x})}, \tag{3.3.9}$$

for $r > 0$, following the definition for the distribution of the maximum of independent random variables.

Simulation of maximum response to the action of an individual wave, given sea state variables and crest elevation

We estimate the distribution of $R_I|(\mathbf{X}, C)$ in two stages: (a) simulation of realisations of wave fields under sea state conditions $\mathbf{X}$ with known crest elevation C, followed by (b) propagation of the resulting wave fields through a suitable structural response model. The details of each stage are outlined below.

The model of Taylor et al. (1997) allows for conditional simulation of a wave field given the occurrence of a turning point of surface elevation in time, with specified crest elevation $C = c > 0$ at the structure location at time $t = 0$, for a given sea state $\mathbf{X} = \mathbf{x}$ with wave spectrum $S(\cdot; \mathbf{x})$, using linear wave theory. The JONSWAP spectrum of Hasselmann et al. (1973) is chosen as the form of $S(\cdot; \mathbf{x})$ due to its applicability to the North Sea wave conditions (Holthuijsen, 2010), where offshore structures of the kind we consider are commonly located, see SM for further details. Taylor et al. (1997) provides expressions for linear crest elevation $E(t; \mathbf{x}, c)$, horizontal velocity $U(t; z, \mathbf{x}, c)$, and horizontal acceleration $\dot{U}(t; z, \mathbf{x}, c)$, at time $t \in \mathbb{R}$ and vertical position $z \in \mathbb{R}$, relative to the mean water level, each conditioned on the wave process (a) attaining a turning point of E at time $t = 0$, with (b) $E(t = 0; \mathbf{x}, c) = c$, both at the location of

the structure. The forms of $E(t;\mathbf{x},c)$ and $U(t;z,\mathbf{x},c)$ are

$$E(t;\mathbf{x},c)=\sum_{n=1}^{N}\left\{(A_n+\mathcal{Q}\sigma_n^2)\cos(\omega_n t)+(B_n+\mathcal{R}\sigma_n^2\omega_n)\sin(\omega_n t)\right\}, \quad (3.3.10)$$

and

$$U(t;z,\mathbf{x},c)=\sum_{n=1}^{N}\omega_n\frac{\cosh(k_n(d+z))}{\sinh(k_n d)}\left\{(A_n+\mathcal{Q}\sigma_n^2)\cos(\omega_n t)+(B_n+\mathcal{R}\sigma_n^2\omega_n)\sin(\omega_n t)\right\}, \quad (3.3.11)$$

for $z<E(t;\mathbf{x},c)$ and zero otherwise, for a regular grid of angular frequencies $\omega_1,\ldots,\omega_N>0$ with spacing $\delta\omega>0$ and $N\in\mathbb{N}$ specified below, where $A_n,B_n\in\mathbb{R}$ $(n=1,\ldots,N)$ and $\mathcal{Q},\mathcal{R}\in\mathbb{R}$ are random coefficients, d is water depth and k_n $(n=1,\ldots,N)$ are wave numbers given implicitly by $\omega_n^2=gk_n\tanh(k_n d)$. The equation for $\dot{U}(t;z,\mathbf{x},c)$ can be found by differentiation of $U(t;z,\mathbf{x},c)$ with respect to t. Coefficients A_n,B_n $(n=1,\ldots,N)$ are a series of independently and identically distributed $N(0,\sigma_n^2)$ random variables with variance $\sigma_n^2=S(\omega_n;\mathbf{x})\delta\omega$, the integrated spectral density in the frequency band $(\omega_n-\delta\omega/2,\omega_n+\delta\omega/2)$ of the discretised wave spectrum. The random coefficients $\mathcal{Q}$ and $\mathcal{R}$ are defined as

$$\mathcal{Q}=\frac{1}{\sum_n\sigma_n^2}\left(c-\sum_{n=1}^{N}A_n\right)\text{ and }\mathcal{R}=\frac{1}{\sum_n\omega_n^2\sigma_n^2}\left(-\sum_{n=1}^{N}\omega_n B_n\right).$$

Next, we estimate the total base shear response of the structure to the simulated conditional wave field. We assume the wave-structure interaction to be quantified by the equation of Morison et al. (1950), which estimates drag and inertial loads applied by the ocean environment on a stick structure. These loads are calculated from the wave velocity and acceleration fields respectively. Under the assumptions of linear wave theory, these fields can be derived entirely from knowledge of the wave spectrum. For the applications described in this work, we assume that the values of sea state (H_S,S_2) are sufficient to define the wave spectrum; hence, for a given structure, the 2-dimensional

storm peak representation $\mathbf{X} = (X_1, X_2) = (H_S, S_2)$ from Section 3.2 is sufficient to describe the extreme ocean environment and its associated structural response.

Under our simplifying assumptions, waves are assumed to be unidirectional, propagating in a single direction towards the vertical cylindrical structure with nominal small diameter. Waves are assumed to pass through the structure, whilst also exerting force, without being obstructed, and the effects of current and wind are ignored. This wave field model provides a basis to approximate the induced load on a jacket structure. The Morison loading equation estimates the base shear $M(t; z, \mathbf{x}, c)$ induced on a cylinder by the wave at time t and vertical position z, and is given by

$$M(t; z, \mathbf{x}, c) = \rho c_m(z) V \dot{U}(t; z, \mathbf{x}, c) + \frac{1}{2}\rho c_d(z) A U(t; z, \mathbf{x}, c)|U(t; z, \mathbf{x}, c)|, \qquad (3.3.12)$$

where $c_m(z), c_d(z) > 0$ are inertia and drag coefficients, $\rho = 1024$ (recall that SI units are used throughout) is the density of water, V is the volume of the body and A is the area of the structure perpendicular to the wave propagation. We assume a cylindrical structure with diameter 1 and height of 150 situated within water of depth $d = 100$. Since the probability of a crest elevation greater than 50 is near zero for all relevant sea states, this structure scenario amounts to a cylinder of infinite height. In order to approximate models of different structure types, $c_m(z)$ and $c_d(z)$ can be made to vary with z, as discussed in Section 3.5. To evaluate the total base shear $B_S(t; \mathbf{x}, c)$ on the structure at time t, we integrate $M(t; z, \mathbf{x}, c)$ to give

$$B_S(t; \mathbf{x}, c) = \int_{-d}^{E(t;\mathbf{x},c)} M(t; z, \mathbf{x}, c)\mathrm{d}z, \qquad (3.3.13)$$

the total Morison load induced up the water column at the structure location.

The response $R_I|(\mathbf{X} = \mathbf{x}, C = c)$ may be obtained by considering the portion of the time series $\{B_S(t; \mathbf{x}, c)\}_{t\in\mathbb{R}}$ that corresponds to the central wave conditioned to attain $E(t = 0; \mathbf{x}, c) = c$; that is, the period of time $t \in \mathcal{T}_0 \subset \mathbb{R}$, with $0 \in \mathcal{T}_0$, for which

the wave surrounding the conditioning crest of elevation c at time $t = 0$ acts on the structure. We define

$$R_I|(\mathbf{X} = \mathbf{x}, C = c) = \max_{t \in \mathcal{T}_0} B_S(t; \mathbf{x}, c). \tag{3.3.14}$$

We obtain realisations of $R_I|(\mathbf{X} = \mathbf{x}, C = c)$ from a time series of the base shear response (3.3.13) evaluated using Morison loads (3.3.12), in turn calculated from wave fields simulated according to expressions (3.3.10) and (3.3.11). The interval of time over which wave fields are simulated corresponds to a period of 120 seconds, sufficiently large to ensure reliable performance of the FFT algorithm (Cooley and Tukey, 1965), meaning here $\mathcal{T}_0 \subset [-60, 60]$. Realisations of conditional crest elevation and wave kinematics are simulated for a regular grid $\{(t_i, z_j)\}_{i=1,j=1}^{n_t,n_z}$ of values $t \in [-60, 60]$ and $z \in [-100, 150]$. We set $N = n_t$ in expressions (3.3.10) and (3.3.11), which is necessary to evaluate the wave field equations using the FFT algorithm; see SM for details. The values of $n_t = 480$ and $n_z = 50$ are chosen, sufficient to ensure reasonable response approximation. The simulated kinematics are then propagated through the Morison equation, providing a realisation of $\{M(t_i; z_j, \mathbf{x}, c)\}_{i=1,j=1}^{n_t,n_z}$. Numerical evaluation of integral (3.3.13) with respect to z yields a realisation of the time series $\{B_S(t_i; \mathbf{x}, c)\}_{i=1}^{n_t}$. A realisation of the maximum individual wave response $R_I|(\mathbf{X} = \mathbf{x}, C = c)$ is then obtained by applying (3.3.14), using the set $t \in \{t_i\}_{i=1}^{n_t} \cap \mathcal{T}_0$ as an approximation to $\mathcal{T}_0$. Given conditioning crests $\{c_i\}_{i=1}^{k}$, the above procedure can be used to map $c_i \mapsto r_i$, for $i = 1, \ldots, k$, obtaining a set of maximum responses $\{r_i\}_{i=1}^{k}$, for a given sea state $\mathbf{x}$.

Importance sampling of simulated maximum responses

The procedure discussed in Section 3.3.3 is used to obtain realisations $\{r_i\}_{i=1}^{k}$ of $R_I|(\mathbf{X}, C)$, for a set of k conditioning crests $\{c_i\}_{i=1}^{k}$ and specified values of storm peak variables $\mathbf{X}$. These are then used in integral (3.3.7) to estimate the distribution $F_{R_I|\mathbf{X}}$ of the maximum response to an individual wave in sea state $\mathbf{X}$, assuming a Rayleigh distribu-

tion (3.3.8) for $C|\mathbf{X}$. However, evaluation of integral (3.3.7) via Monte Carlo methods sampling from the Rayleigh density is inefficient in targeting the tail of the response distribution $F_{R_I|\mathbf{X}}$. Given our interest in extreme structural responses, we therefore employ the importance sampling approach described by Towe et al. (2021), writing integral (3.3.7) as

$$F_{R_I|\mathbf{X}}(r|\mathbf{x}) = \int_r F_{R_I|\mathbf{X},C}(r|\mathbf{x},c)\frac{f_{C|\mathbf{X}}(c|\mathbf{x})}{g^{(\epsilon)}_{C|\mathbf{X}}(c|\mathbf{x})}g^{(\epsilon)}_{C|\mathbf{X}}(c|\mathbf{x})\mathrm{d}c, \tag{3.3.15}$$

for $r > 0$, where $g^{(\epsilon)}_{C|(\mathbf{X}=\mathbf{x})}$ is the density of the Uniform$[0, \epsilon h(\mathbf{x})]$ distribution, for significant wave height $h(\mathbf{x})$ and some $\epsilon > 0$. For a fixed number of conditional wave simulations, sampling of the conditioning crest c from $g^{(\epsilon)}_{C|\mathbf{X}}$ ensures greater coverage of the feasible range of large crest elevations and of the induced maximum response than is achieved when sampling c from $f_{C|\mathbf{X}}$. Therefore, the upper tail of the distribution $F_{R_I|\mathbf{X}}$ is more efficiently estimated using the sampling distribution $g^{(\epsilon)}_{C|\mathbf{X}}$. The value of ϵ in (3.3.15) is selected so that $g^{(\epsilon)}_{C|\mathbf{X}}$ provides adequate coverage of the domain of $f_{C|\mathbf{X}}$, i.e., the exceedance probability

$$\Pr\{C > \epsilon h(\mathbf{x})|\mathbf{X} = \mathbf{x}\} = \int_{c>\epsilon h(\mathbf{x})} f_{C|\mathbf{X}}(c|\mathbf{x})\mathrm{d}c = \exp\left(-8\epsilon^2\right), \tag{3.3.16}$$

is sufficiently close to zero. We set $\epsilon = 2$, which gives a value of probability (3.3.16) in the order of 10^{-14}.

Integral (3.3.15) is then estimated as follows. The k conditional crests $\{c_i\}_{i=1}^k$ are sampled from the uniform proposal density $g^{(\epsilon)}_{C|\mathbf{X}}$. Corresponding realisations $\{r_i\}_{i=1}^k$ of single-wave maximum responses are obtained using the procedure described in Section 3.3.3. The distribution $F_{R_I|\mathbf{X}}$ is then estimated as

$$\hat{F}_{R_I|\mathbf{X}}(r|\mathbf{x}) = \frac{\sum_{i=1}^k \mathbb{1}_{\{r_i \le r\}}\iota^{(\epsilon)}(c_i|\mathbf{x})}{\sum_{i=1}^k \iota^{(\epsilon)}(c_i|\mathbf{x})}, \tag{3.3.17}$$

for $r > 0$, where $\iota^{(\epsilon)}(c|\mathbf{x}) = f_{C|\mathbf{X}}(c|\mathbf{x})/g^{(\epsilon)}_{C|\mathbf{X}}(c|\mathbf{x})$ is the importance sampling ratio and $\mathbb{1}_{\{r_i \leq r\}} = 1$ if $r_i \leq r$, zero otherwise, for $i = 1, \ldots, k$. Estimate (3.3.17) is an empirical cumulative distribution function of the simulated responses, weighted to remove bias introduced from sampling crests from $g^{(\epsilon)}_{C|\mathbf{X}}$ rather than $f_{C|\mathbf{X}}$. We use estimate (3.3.17) to evaluate the distribution of maximum response per L-hour sea state using relation (3.3.9). Given an estimate for $f_{\mathbf{X}}$ obtained as in Section 3.3.2, we may then calculate the marginal maximum response distribution using integral (3.3.3).

3.4 Environmental contours

3.4.1 Overview of environmental contours

Environmental contours provide a method of determining extremal conditions that are in some way related to an extreme structural response. These contours often make assumptions about the interaction between environment and response, usually regarding the shape of some failure boundary in the environment space such that environmental conditions beyond the boundary will result in structural failure. For instance, IFORM (e.g., Winterstein et al. 1993) contours assume a convex form for this boundary, whereas ISORM (e.g., Chai and Leira 2018) assumes it to be concave. These assumptions may or may not be valid depending on the specific features of the structure type in question. Here, we outline the methodology of the IFORM contour (Section 3.4.2) and the fitting approach we employ to estimate it for our example dataset (Section 3.4.3).

3.4.2 IFORM design contours

Section 3.3.1 details how, given an estimate for the distribution of $R_L|\mathbf{X}$ from Section 3.3, in principle we can evaluate the probability of structural failure $\Pr(\mathcal{F}_L|\mathbf{X} = \mathbf{x})$ for a sea state (of duration L) with variables $\mathbf{X} = \mathbf{x}$. IFORM offers an approach to structural design that avoids direct calculation of $\Pr(\mathcal{F}_L|\mathbf{X} = \mathbf{x})$, by attempting to

make conservative assumptions. For an ocean environment represented by a set of random variables $\mathbf{X}$ transformed to independent standard Gaussian random variables $\mathbf{U}$, IFORM assumes that $\Pr(\mathcal{F}_L|\mathbf{U}=\mathbf{u})$ is deterministic for all $\mathbf{u}$, taking values $\{0,1\}$, contrary to the failure probability discussed in Section 3.3.1 which takes any value in $[0,1]$. Writing the region of environmental space corresponding to failure $\mathcal{F}_L$ as $\mathcal{F}$, IFORM assumes that the boundary $\partial\mathcal{F}$ of $\mathcal{F}$ is linear, and lies tangential to a contour of constant transformed environmental density $f_{\mathbf{U}}$, making the assumed location of $\partial\mathcal{F}$ dependent on the joint Gaussian distribution $f_{\mathbf{U}}$. The assumption of a failure boundary of this type is typically conservative, in that estimates for $\Pr(\mathbf{U}\in\mathcal{F})$ using it have positive bias.

The transformation of $\mathbf{X}\mapsto\mathbf{U}$ is achieved via the method of Rosenblatt (1952), which proceeds as follows. For storm peak variables $\mathbf{X}=(X_1,\ldots,X_p)$, suppose we can estimate the nested conditional distributions $F_{X_1}, F_{X_2|X_1}, F_{X_3|(X_1,X_2)},\ldots,F_{X_p|(X_1,\ldots,X_{p-1})}$. Estimation of these distributions is non-trivial as it involves estimating a sequence of conditional dependence models with results dependent on the sequencing of the p environmental variables; see Section 3.4.3 for an example approach. The Rosenblatt transformation maps a realisation $\mathbf{x}=(x_1,\ldots,x_p)$ of $\mathbf{X}$ to the realisation $\mathbf{u}=(u_1,\ldots,u_p)$ of $\mathbf{U}$ via $u_1=\Phi^{-1}\{F_{X_1}(x_1)\}$ and $u_j=\Phi^{-1}\left\{F_{X_j|(X_1,\ldots,X_{j-1})}(x_j|x_1,\ldots,x_{j-1})\right\}$, for $j=2,\ldots,p$, where Φ is the standard Gaussian cumulative distribution function.

In $\mathbf{U}$-space, contours of constant probability joint density correspond to boundaries of hyperspheres centred at the origin. In particular, the P-year IFORM contour in $\mathbf{U}$-space is the boundary of a hypersphere with radius

$$\beta_I=\Phi^{-1}\left(1-\frac{1}{N_{\text{an}}P}\right), \tag{3.4.18}$$

where N_{an} is the average number of independent storm peak observations per annum. That is, the probability of a point lying outside the set enclosed by the P-year IFORM contour is $1/(N_{\text{an}}P)$. Consider failure region $\mathcal{F}_{\beta_I}$ with boundary $\delta\mathcal{F}_{\beta_I}$ tangential to the

hypersphere centred at the origin with radius β_I. For any angle $\boldsymbol{\psi}$, in spherical polar coordinates for a point on the hypersphere, we have

$$\Pr\left(\mathbf{U} \in \mathcal{F}_{\beta_I}\right) = \Pr\left(\sum_{i=1}^{p} w_i(\boldsymbol{\psi}) U_i > \beta_I\right),$$

where $\sum_{i=1}^{p} w_i^2(\boldsymbol{\psi}) = 1$, due to the linearity of $\delta\mathcal{F}_{\beta_I}$ and it being tangential to the hypersphere at radius β_I. So, as $W = \sum_{i=1}^{p} w_i(\boldsymbol{\psi}) U_i \sim N(0,1)$ given the independence of $(U_1, \ldots, U_p)$, it follows that $\Pr\left(\mathbf{U} \in \mathcal{F}_{\beta_I}\right) = \Pr\left(W > \beta_I\right)$, which directly gives expression (3.4.18) for β_I.

Once the environmental contour has been estimated in **U**-space, it can be represented in the original **X**-space via the inverse Rosenblatt transformation $x_1 = F_{X_1}^{-1}\{\Phi(u_1)\}$ and $x_j = F_{X_j|(X_1,\ldots,X_{j-1})}^{-1}\{\Phi(u_j)|x_1, \ldots, x_{j-1}\}$, for $j = 2, \ldots, p$. Unlike the hypersphere-shaped contours in the **U**-space, these contours are not guaranteed to be convex (see Section 3.5.2). The procedure for construction of a P-year IFORM contour in terms of environmental variable $\mathbf{X} = (X_1, \ldots, X_p)$ is summarised in Algorithm 1.

Algorithm 1 IFORM contour calculation for $\mathbf{X} = (X_1, \ldots, X_p)$

Input Return period P; Average number of independent storm peaks per annum N_{an}; Estimates of distributions $F_{X_1}, F_{X_2|X_1}, F_{X_3|(X_1,X_2)}, \ldots, F_{X_p|(X_1,\ldots,X_{p-1})}$.

Output P-year IFORM contour.

1: Define $\beta_I = \Phi^{-1}\left(1 - \frac{1}{N_{\text{an}}P}\right)$ where Φ is the standard normal cdf.

2: Obtain a set of k equally spaced points $\{(u_{1(j)}, \ldots, u_{p(j)})\}_{j=1}^{k}$ on the hypersphere given by $u_1^2 + \ldots + u_p^2 = \beta_I^2$, using a regular grid of values of $\boldsymbol{\psi}$ in spherical polar coordinates. j in $1, \ldots, k$

3: Compute $x_{1(j)} = F_{X_1}^{-1}\{\Phi(u_{1(j)})\}$.

4: Compute $x_{2(j)} = F_{X_2|X_1}^{-1}\{\Phi(u_{2(j)})|x_{1(j)}\}$.

5: $\vdots$

6: Compute $x_{p(j)} = F_{X_p|(X_1,\ldots,X_{p-1})}^{-1}\{\Phi(u_{p(j)})|(x_{1(j)}, \ldots, x_{p-1(j)})\}$. $\{(x_{1(j)}, \ldots, x_{p(j)})\}_{j=1}^{k}$ points along the p-dimensional IFORM contour.

3.4.3 Joint parametric models for storm peak variables

Construction of the IFORM contour for environmental variables $\mathbf{X} = (X_1, X_2) = (H_S, S_2)$ via Algorithm 1 requires estimates for the marginal distribution F_{H_S} and conditional distribution $F_{S_2|H_S}$ (while F_{S_2} and $F_{H_S|S_2}$ could equally be used, we use the former to reflect past approaches). Mirroring the hierarchical approach of Winterstein et al. (1993), we select the GPD tail model (3.3.4) for marginal H_S, and evaluate a range of parametric forms for the distribution of $S_2|H_S$, selecting the most appropriate based on an assessment of predictive performance. We estimate the model $S_2|H_S$ as follows.

We allow for two sources of flexibility: (a) the conditional distributional form for $S_2|H_S$, and (b) the nature of the parametric form for how the parameters of the distribution vary as a function of H_S. The distributions we consider in modelling step (a) are the Lognormal(μ_L, σ_L), as in Winterstein et al. (1993), Gamma(α, β), Weibull(λ, k) and the

Generalised Extreme Value distribution GEV(μ_G, σ_G, ξ); see SM. We also consider conditional distributions fitted to the transformed negative steepness $S_2^- = \min(S_2) - S_2$, enabling the right-hand tail of the distribution to be fitted to small S_2; see SM for a summary of all model combinations. In step (b), we impose linear, quadratic and exponential forms for the S_2 distribution parameters as functions of H_S. To assess the performance of candidate models for $S_2|H_S$ and $S_2^-|H_S$, we use a cross validation approach in which we evaluate the predictive likelihood of the models, focusing on the performance in the tail region (for large H_S). This proceeds as follows.

The sample $\{\mathbf{x}_i\}_{i=1}^{N_{\text{all}}}$ (see Section 3.2) is partitioned into a 'body' where $H_S \leq v$ and a 'tail' where $H_S > v$ for $v > 0$, denoted $\{\mathbf{x}_i^B\}_{i=1}^{N_B}$ and $\{\mathbf{x}_i^T\}_{i=1}^{N_T}$ respectively, with $N_B + N_T = N_{\text{all}}$. The tail portion $\{\mathbf{x}_i^T\}_{i=1}^{N_T}$ is itself partitioned into K subsets $\{\mathcal{S}_j\}_{j=1}^{K}$ each of sizes $\lfloor N_T/K \rfloor$ or $\lceil N_T/K \rceil$. A K-fold cross validation is performed using training set $\{\mathbf{x}_i\}_{i=1}^{N} \setminus \mathcal{S}_j$ and test set $\mathcal{S}_j$ at fold $j \in \{1, \ldots, K\}$. That is, we always utilise the entirety of the body of the data within the training set alongside all but a single fold of the tail. Excluding a subset of the tail from the training data in this way is appropriate for estimation of the conditional distribution for $S_2|H_S$, but leads to biased estimation of the marginal distribution of H_S, so it is important to note estimation of the marginal model (3.3.4) (see Section 3.3.2) is carried out using the entire dataset. The predictive likelihood is then only calculated on extreme data points, and so measures the fit of each model to the extremes of the data.

We repeat the above process for $K = 5, 10$ over values of $v = 0, 0.8, 0.9$ to determine the sensitivity of model performance to the choice of extreme threshold. Setting $v = 0$ recovers a standard cross validation approach for assessing fit to all of data, which we also include to ensure the best performing models fit the body of the data well. In addition, we also evaluate the model fit using AIC. The AIC and cross validation scores are each standardised by dividing by the number of observations for which we evaluate the optimised likelihood (negated when considering AIC, which is standardised over

both terms). This standardisation results in a set of loosely comparable scores for each model across different threshold choices, and these scores for each model are averaged over values of both K and v to obtain a single score, referred to as the aggregate score (AS); see SM in Section 3.7 for details. The model with the largest AS is deemed to be the best predictive model for the data.

3.5 Results

3.5.1 Estimating the joint density of storm peaks

We employ the forward methodology of Sections 3.3.2 and 3.3.3 to estimate the environmental density $f_{\mathbf{X}}$ and response distribution $F_{R_L|\mathbf{X}}$ for our motivating dataset of $\mathbf{X} = (H_S, S_2)$ introduced in Section 3.2. In turn, these are used to evaluate the distributions F_{R_S} and F_{R_A} (as in Sections 3.3.1 and 3.3.1) and subsequently CDE_P (as in Section 3.3.1). The CDE_P is then compared with various IFORM contour estimates using the methods of Section 3.4. The purpose of the comparison of CDE_P and IFORM contours is to demonstrate that none of the hierarchical environmental model forms underpinning the IFORM contour provide an adequate description of CDE_P for all example structures considered. That is, the IFORM contours estimated in general do not identify the correct region of environmental space responsible for extreme structural response. Therefore, evaluation of response along these IFORM contour boundaries will not provide reliable realisations of the desired P-year response.

We model the joint environment $\mathbf{X}$ using the method of Heffernan and Tawn (2004) discussed in Section 3.3.2, fitted to data in the region $A_\nu = \{\mathbf{X} \in \mathbb{R}^2 : H_S > \nu\}$ for conditioning threshold ν. We take $\nu = \tilde{F}^{-1}_{H_S}(0.95)$. Inspection of plots for the variability of conditional extremes model parameters with respect to threshold indicated this choice of threshold to be within the interval for which these parameters are invariant. Using this fitted model, the density $f_{\mathbf{X}}$ is estimated in A_ν as described in Section 3.3.2. The

density in the complement $\mathbb{R}^2 \setminus A_\nu$ is modelled empirically. Since our interest lies in environments with large H_S and associated structural responses, we are not concerned with (a) smooth estimation of this lower portion of the density and (b) the density in the region corresponding to large S_2 but small H_S.

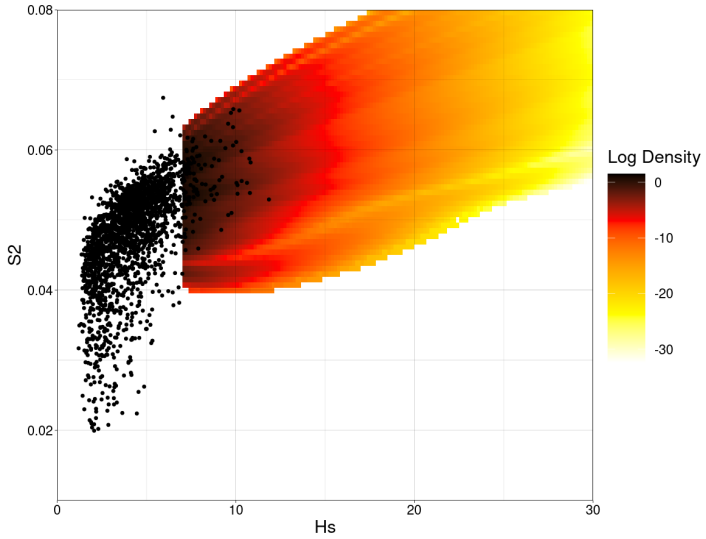


Figure 3.5.2: Storm peak density for H_S [m] and S_2 in A_ν on log scale, estimated from the fitted conditional extremes model. Observations of storm peak $\mathbf{X} = (H_S, S_2)$ are shown as black dots.

In the 2-dimensional case $\mathbf{X} = (H_S, S_2)$, the parameters of model (3.3.6) reduce to $\alpha \in [-1, 1]$ and $\beta \in (-\infty, 1]$ when conditioning on H_S. We find corresponding estimates $\hat{\alpha} = 0.378$ and $\hat{\beta} = 0.533$, as well as an estimate $\hat{\xi} = -0.02$ for the H_S marginal shape parameter ξ in (3.3.4), the latter indicating a near exponential upper tail for H_S. The effect of $\hat{\alpha}$ is seen in Figure 3.5.2, as a positive trend in the resulting density estimate. This density also appears to agree with the shape of the data in the extreme region $\mathcal{A}_\nu$. Figure 3.5.2 also shows the presence of rays in the estimated density, caused by sampling from the kernel density of observed residual values (see Winter and Tawn 2017 or discussion in Section 3.3.2). The strength of these rays is determined by the value of the kernel smoothing parameter. Higher values provide stronger smoothing and therefore less prominent rays, however, too large a value will result in over-smoothing and thus less precise density estimates. When in doubt, we favour under-smoothing.

3.5.2 Selection of model form for the conditional distribution

For IFORM, we consider four distributional forms for each of $S_2|H_S$ and $S_2^-|H_S$ (summarised in Section 3.4.3) each with two distribution parameters modelled as functions of H_S. We do not model the GEV shape ξ as a function of H_S as its estimation using maximum likelihood is difficult for finite samples; instead, we assume it to be an unknown constant. The variation of each of these eight distribution parameters with H_S is represented by one of three parametric forms (linear, quadratic and exponential), giving a total of 72 combined candidate models for $S_2|H_S$ and $S_2^-|H_S$. These models are ranked using the AS, introduced in Section 3.4.3, yielding results given in full in the SM. Table 3.5.1 summarises these results, showing the optimal model for each of $S_2|H_s$ and $S_2^-|H_S$, with unique $a, b, c \in \mathbb{R}$ for each distributional parameter, constrained such that their respective domains are not violated. Standard errors are found as the sample standard deviation of the AS, evaluated over thirty replicates of each cross validation setting.

Label	Distribution (with optimal functional form of parameters)			AS
$\mathcal{C}_P^1$	$S_2^-\|\{H_S = h\} \sim \text{GEV}(\mu_G, \sigma_G, \xi)$	$\mu_G = a + b\exp(ch)$	$\sigma_G = a + b\exp(ch)$	**3.999 (0.002)**
$\mathcal{C}_P^2$	$S_2\|\{H_S = h\} \sim \text{Weibull}(k, \lambda)$	$k = a + bc$	$\lambda = a(h+b)^2 + c$	**3.983 (0.003)**
$\mathcal{C}_P^3$	$S_2^-\|\{H_S = h\} \sim \text{Lognormal}(\mu_L, \sigma_L)$	$\mu_L = a + bh$	$\sigma_L = a + bh$	**3.963 (0.001)**
$\mathcal{C}_P^4$	$S_2^-\|\{H_S = h\} \sim \text{Gamma}(\alpha, \beta)$	$\alpha = a(h+b)^2 + c$	$\beta = a + bh$	3.963 (0.006)
$\mathcal{C}_P^5$	$S_2^-\|\{H_S = h\} \sim \text{Weibull}(k, \lambda)$	$k = a + b\exp(ch)$	$\lambda = a + b\exp(ch)$	3.897 (0.002)
$\mathcal{C}_P^6$	$S_2\|\{H_S = h\} \sim \text{Gamma}(\alpha, \beta)$	$\alpha = a + bh$	$\beta = a + bh$	3.873 (0.002)
$\mathcal{C}_P^7$	$S_2\|\{H_S = h\} \sim \text{Lognormal}(\mu_L, \sigma_L)$	$\mu_L = a(h+b)^2 + c$	$\sigma_L = a + b\exp(ch)$	3.824 (0.005)
$\mathcal{C}_P^8$	$S_2\|\{H_S = h\} \sim \text{GEV}(\mu_G, \sigma_G, \xi)$	$\mu_G = a + b\exp(ch)$	$\sigma_G = a(h+b)^2 + c$	3.533 (0.000)

Table 3.5.1: AS (with standard error) for the optimal forms of each distribution considered for $S_2|H_S$ and $S_2^-|H_S$. Large values of AS indicate good performance. The three best performing models' scores are indicated in bold. Complete results are given in the SM.

The models in Table 3.5.1 are used with Algorithm 1 to construct the IFORM contours in Figure 3.5.3. Each contour corresponds to a return period of $P = 1000$ years. Contours are labelled $\mathcal{C}_P^1$ to $\mathcal{C}_P^8$ and ordered according to their AS, with the best fitting models having the lowest labelling. All of the contour estimates provide plausible descriptions of the shape of the sample, but from an engineering design perspective, we

note clear differences in the shape and position of the contours for larger H_S. Even so, the three highest ranking models generate contours that agree to a reasonable degree in all regions. These three contours also appear visually to be the best descriptions of the shape of the data. In comparison, the other contours do not agree in the region of large H_S, and fail to capture the shape of the main body of the data. We therefore select the highest three ranking contours as the best representations of IFORM to compare to CDE_P in Section 3.5.3.

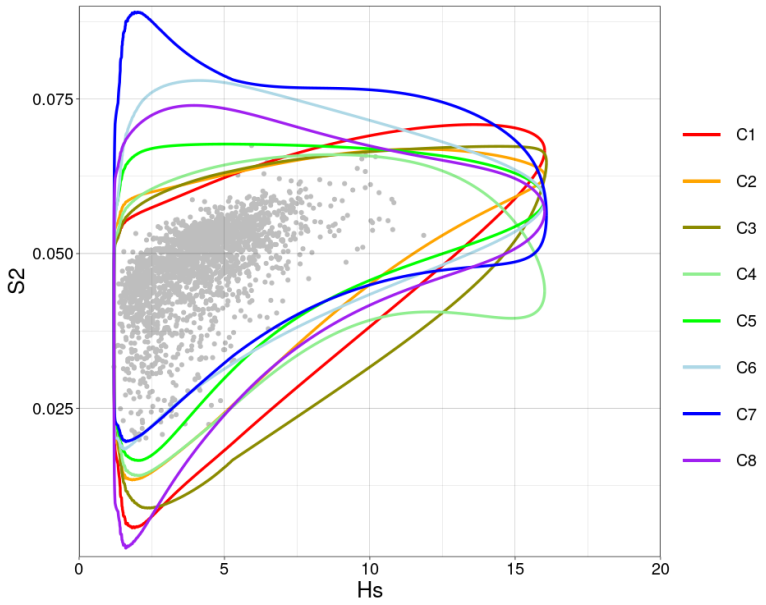


Figure 3.5.3: IFORM contours in H_S [m] and S_2 constructed from the models in Table 3.5.1, corresponding to an exceedence probability of $p = 10^{-3}/73$, or a return period of $P = 1000$ years for data with $N_{\text{an}} = 73$ observations per annum. Contours $\mathcal{C}_P^1$ to $\mathcal{C}_P^8$ are listed and coloured in order of decreasing performance AS, from red to purple, with contour $\mathcal{C}_P^i$ labelled Ci for $i = 1 \ldots, 8$.

3.5.3 Estimating the conditional density of associated environmental variables

Estimation of CDE_P for example structure models

We evaluate CDE_P for three examples of the stick structure model (Section 3.3.3), denoted A, B & C. These structures assume different values for drag and inertia co-

efficients $c_d(z)$ and $c_m(z)$ along their height z, as shown in Table 3.5.2. Structure A represents the simplest stick structure, with homogeneous drag and inertia coefficients along its entire height, i.e., $c_d(z) = c_m(z) = 1$ for all $z \in [-100, 150]$. We mimic wave-in-deck loads for structure B, with $c_d(z) = c_m(z) = 100$ increased for a portion of the structure above sea level. A portion of structure C near the sea bed incurs increased load.

	Value of $c_d(z) = c_m(z)$			
Structure		$5 < z \leq 15$	$-95 < z \leq -85$	Elsewhere
	A	1	1	1
	B	100	1	1
	C	1	100	1

Table 3.5.2: The drag and inertia coefficients of (3.3.12) with $c_d(z) = c_m(z)$ for all z, for varying z associated with stick structures A, B & C.

Figure 3.5.4 shows the corresponding estimates for CDE_P for $P = 1000$ years. The shape and position of the conditional density varies between structures, due to their differing loading characteristics. For structure C in particular, the conditional density extends to larger H_S, and over a wider interval of S_2; we comment further on this feature in the discussion of environmental contours in Figure 3.5.6. That is, Figure 3.5.4 demonstrates that more than one region of the (H_S, S_2) domain contributes to the distribution of the P-year response, particularly for structure C.

Comparison between CDE_P and IFORM contours

Provided we have accurate models for the series of p nested conditional distributions $F_{X_1}, \ldots, F_{X_p|(X_1,\ldots,X_{p-1})}$, and provided that the assumptions underlying IFORM are valid, the P-year IFORM contour gives design points at which evaluation of a response model will provide conservative estimates of the P-year response, as indicated by Winterstein et al. (1993). That is, it aims to provide environmental conditions at least as extreme as those that induce the P-year response for specific structures. It is natural therefore to consider assessing IFORM contour performance using CDE_P,

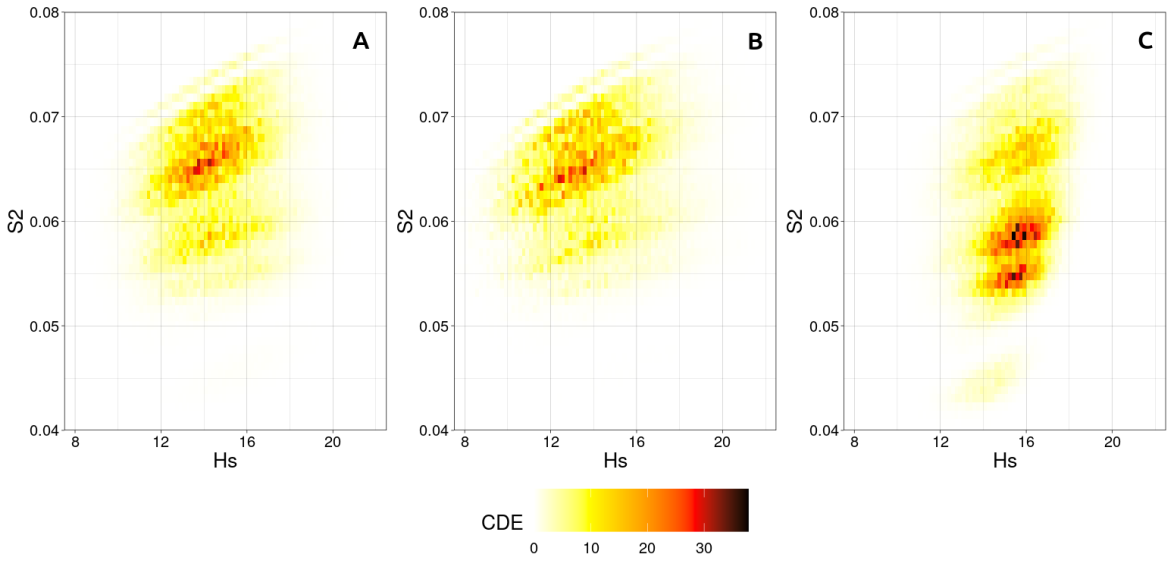


Figure 3.5.4: Estimated CDE_P for the three structure models given in Table 3.5.2 (left to right A, B & C), for a period of $P = 1000$ years. These are evaluated for the example dataset given in Section 3.2, using the methods of Section 3.3.

since the latter provides asymptotically-justified estimates of the environmental conditions corresponding to the P-year response, obtained from application from the forward approach of Section 3.3.

Assuming that CDE_P provides a valid estimate of the environment density conditional on the P-year response, we reason that a well-estimated IFORM contour $\mathcal{C}_P$ should intersect with CDE_P, and that the discrepancy between $\mathcal{C}_P$ and CDE_P points to inadequacy of the IFORM methodology, potentially due to a mis-specified parametric environmental model or invalid IFORM assumptions regarding the characteristics of the wave-structure interaction for the application at hand. For instance, see Figure 3.5.5, which shows the estimated CDE_P for structures A, B and C, for $P = 1000$. Contour $\mathcal{C}_P^2$ in the left panel overlaps with darker points of CDE_P so provides estimates of the P-year response that agree with those obtained from the forward approach, meaning it performs well in this region of the contour. The centre panel of Figure 3.5.5 shows $\mathcal{C}_P^1$ taking (H_S, S_2) values more extreme than the darker points of CDE_P, indicating that the contour will provide estimates for the P-year response more extreme than those ob-

tained from the forward approach, i.e., the conservative outcome intended. Conversely, the right panel shows $\mathcal{C}_P^2$ taking (H_S, S_2) values less extreme than the darker points of CDE_P, suggesting estimates for the P-year response obtained from points along this contour will be smaller than those obtained from the forward approach and hence they fail to give conservative estimates, which is contrary to their claimed properties. More generally, the majority of the presented IFORM contours exhibit over-conservatism in regions of small S_2, i.e., they lie in regions of the environment space with more severe H_S than in the regions with non-zero CDE_P, and this could lead to substantial over-design if this region of the environmental space was important.

We observe from the above analysis of Figure 3.5.5 that the overlap between the region of non-zero CDE_P and the region $\mathcal{A}_P$ bounded below by the IFORM contour $\mathcal{C}_P$ relates to the level of conservatism, and therefore is a good scalar metric for measuring performance of IFORM contours relative to the more strongly justified forward approach. If this overlap is zero, then the contour lies in a region less extreme than the region of non-zero CDE_P, indicating non-conservatism. Conversely, if the overlap includes all of the non-zero region of CDE_P, the contour appears to be conservative. To formalise this observation, we define the following metric, which quantifies the level of this overlap. Consider

$$\zeta(P, \mathcal{A}_P) = 2 \int_{\mathbf{x}} \mathbb{1}_{\mathcal{A}_P}(\mathbf{x}) f_{\mathbf{X}|R_L}(\mathbf{x}|r_P) \mathrm{d}\mathbf{x} - 1,$$

for $f_{\mathbf{X}|R_L}$ as in Section 3.3.1, with $\mathbb{1}_{\mathcal{A}_P}(\mathbf{x}) = 1$ if $\mathbf{x} \in \mathcal{A}_P$ and $\mathbb{1}_{\mathcal{A}_P}(\mathbf{x}) = 0$ otherwise, and where r_P is the P-year response of the structure. The metric $\zeta(P, \mathcal{C}_P)$ takes values in [-1, 1]. Here, $\zeta(P, \mathcal{C}_P) > 0$ indicates conservatism of $\mathcal{C}_P$ due to overestimation of the P-year response, $\zeta(P, \mathcal{C}_P) < 0$ indicates underestimation of the P-year response and so non-conservatism, and $\zeta(P, \mathcal{C}_P) \approx 0$ indicates accurate estimates for the P-year response and so optimal overall structural design given by IFORM.

To justify our findings regarding $\zeta(P, \mathcal{C}_P)$ in a general case, we apply the following

heuristic argument. First, we observe from Figure 3.5.4 that CDE_P exhibits approximate marginal symmetry with respect to both H_S and S_2 (i.e., little marginal skew). We also assume $\mathcal{A}_{P_1} \subset \mathcal{A}_{P_2}$ for any $P_2 > P_1 \geq 1$, and that a contour $\mathcal{C}_P$ varies smoothly with H_S and S_2 for a given P. Then, (a) if $\zeta(P, \mathcal{C}_P) \approx 0$, the integral over CDE_P within $\mathcal{A}_P$ is approximately equal to the integral over CDE_P within $\mathcal{A}_P^c$. We interpret this as indicating that points on $\mathcal{C}_P$ coincide with high values of CDE_P, as in the cases seen in Figure 3.5.5, hence structural responses corresponding to points on $\mathcal{C}_P$ will be of similar magnitudes to those from the high density regions of CDE_P. However, (b) if $\zeta(P, \mathcal{C}_P) > 0$, $\mathcal{A}_P$ contains the environmental region where CDE_P is non-zero, hence points on $\mathcal{C}_P$ produce structural responses beyond the P-year level, resulting in conservative design using IFORM contour $\mathcal{C}_P$. Further, (c) if $\zeta(P, \mathcal{C}_P) < 0$, the intersection between $\mathcal{A}_P$ and the non-zero CDE_P region of the environment space is negligible. Given our assumptions, this arises only when $\mathcal{A}_P$ occupies a region of the environmental space less extreme than that with non-zero CDE_P. Under these circumstances, points on $\mathcal{C}_P$ will produce structural responses corresponding to return periods less than P, resulting in a lack of conservatism in design.

Estimates for $\zeta(P, \mathcal{C}_P)$ in Table 3.5.3 support this interpretation, relative to our estimates in Figure 3.5.5. For structure A, only contour $\mathcal{C}_P^2$ appears to pass through the highest density region of CDE_P. The other contours lie beyond the highest density region of CDE_P, corresponding to $\zeta(P, \mathcal{C}_P) > 0$. Observations for structure B are similar, since CDE_P does not vary considerably between structures A and B. For structure C, relative to A and B, the highest density regions in CDE_P occur at higher H_S but lower S_2. Hence, despite no change in the locations of contour $\mathcal{C}_P^j$ ($j = 1, 2, 3$), now only contour $\mathcal{C}_P^3$ is conservative. Contours $\mathcal{C}_P^1$ and $\mathcal{C}_P^2$ are clearly non-conservative, as confirmed in Table 3.5.3.

Contour	$\zeta(P, \mathcal{C}_P)$		
	Structure A	Structure B	Structure C
$\mathcal{C}_P^1$	0.353	0.475	-0.259
$\mathcal{C}_P^2$	-0.125	-0.016	-0.539
$\mathcal{C}_P^3$	0.176	0.200	0.042

Table 3.5.3: Estimates of $\zeta(P, \mathcal{C}_P)$ from (3.5.3) for IFORM contours corresponding to the three best fitting models for H_S, S_2 from Table 3.5.1, for each of structures A, B and C from Table 3.5.2.

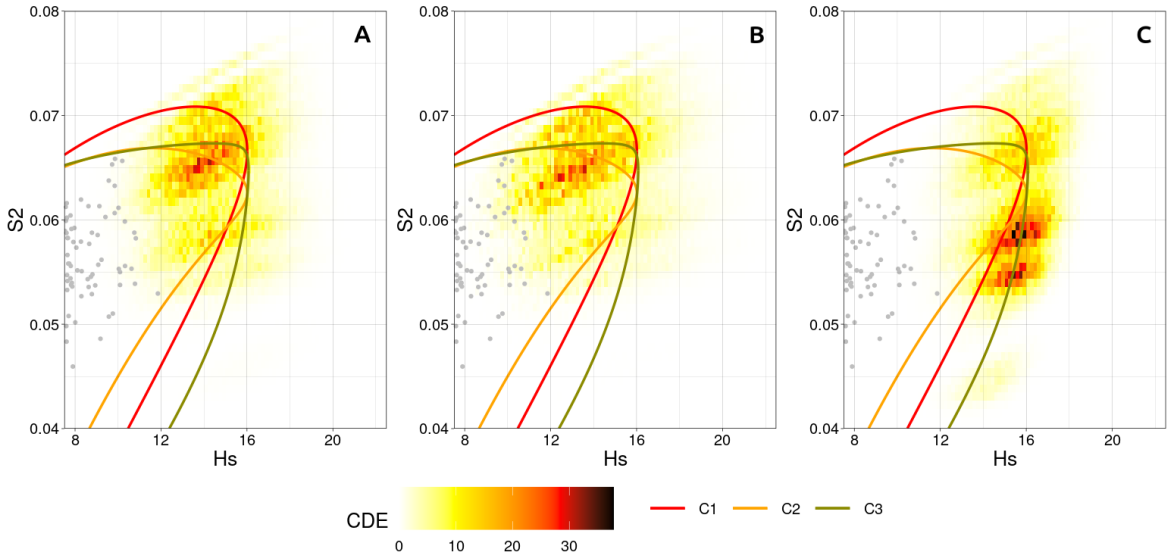


Figure 3.5.5: Density of storm peak sea state parameters (H_S, S_2) conditioned on observing the 1000-year marginal response on stick-type structures. Overlaid are the three highest scoring IFORM fits by AS, with contour $\mathcal{C}_P^i$ labelled Ci for $i = 1 \ldots, 3$. The colouring of the contours (red, orange, grey) indicates the order of ranking, in terms of decreasing predictive performance.

Figure 3.5.6 shows the estimated log probability of exceeding the P-year response 3-hour sea state, given by $\log\{1 - F_{R_L|\mathbf{X}}(r_p|\mathbf{x})\}$ where $F_{R_L|\mathbf{X}}$ is estimated as in Section 3.3, for $P = 1000$ and for all $\mathbf{X} = (H_S, S_2)$ within a subset $[0, 25] \times [0, 0.08]$ of the environment space $\mathbb{R}^2$. IFORM contours using the highest ranking environmental model are overlaid. The panels show that the probability of exceeding P-year response for given environmental conditions varies with structure type, resulting in differing locations of

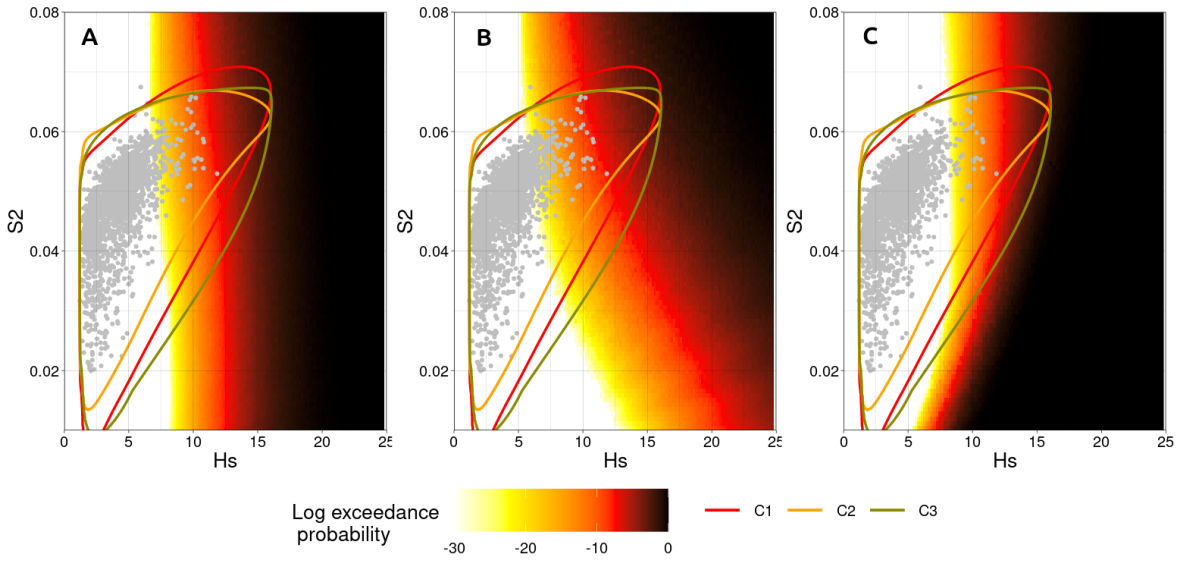


Figure 3.5.6: Estimated log probability of exceeding the P-year response for $P = 1000$ within a 3-hour sea state as a function of H_S and S_2, for structures A, B and C (left to right), obtained using the foward method of Section 3.3. Overlaid are IFORM contours for the three best-fitting environmental models, with contour $\mathcal{C}_P^i$ labelled Ci for $i = 1 \ldots, 3$. The original sample of storm peak (H_S, S_2) is shown as grey dots.

a 'frontier' where this probability becomes non-zero (i.e., where the log probability exceeds roughly -30), corresponding to the shaded regions on the plots. For structure A, the frontier lies along constant H_S, indicating that H_S alone affects the base shear induced on the structure. For structures B and C, we see convex and concave curvature of the frontier respectively, indicating that these structures are more susceptible to high- and low-steepness conditions respectively. These differences in the underlying wave-structure interaction are reflected in the positions of CDE_P in Figures 3.5.4 and 3.5.5, but not in the locations of the IFORM contours which remain the same across all structures. This change in frontier location is therefore the cause of different contour performances across structures, seen in Table 3.5.3 and Figure 3.5.5. For instance, contours $\mathcal{C}_P^1$ and $\mathcal{C}_P^2$ give non-conservative estimates for the P-year response on structure C because they fail to account for the concave curvature of the frontier seen in the third panel of Figure 3.5.6, and so not react to the increased severity of response seen at lower values of S_2. In summary, none of the three contours $\mathcal{C}_P^1$, $\mathcal{C}_P^2$, $\mathcal{C}_P^3$ provides

the same level of conservatism for all three example structures. Indeed some contours, although fitting the sample data well, are non-conservative for design.

3.5.4 Discussion

Offshore structures are subjected to extremes of environmental conditions (in the current context, significant wave height H_S and wave steepness S_2), making structural risk assessment a critical step in the design process. Ideal methods combine models for environmental extremes with direct estimation of the environment-structure interaction from fluid loading. We demonstrate an efficient approach to estimating the distribution of extreme base shear on three stick structures, in a central North Sea environment. Environmental modelling is carried out using a combination of asymptotically justified models for univariate (Davison and Smith, 1990) and conditional extremes (Heffernan and Tawn, 2004). Simulation of short term conditional wave fields then exploits the conditional method of Taylor et al. (1997) with efficient sampling (Towe et al., 2023) to estimate the distribution of Morison base shear (see Morison et al. 1950).

Using the full probabilistic analysis, we estimate the conditional density of the environment variables (CDE_P) given occurrence of the P-year response. We adopt CDE_P as a diagnostic tool with which to evaluate the usefulness of any approach that claims to identify regions of the environmental space associated with extreme structural response.

Due to the perceived computational complexity of the forward approach, metocean engineers often use environmental contour methods as computationally simpler alternatives to approximate a full probabilistic analysis. These characterise the joint environment only, and then rely on assumptions about the nature of the environment-structure interaction to be useful for design (see Ross et al. 2020; Haselsteiner et al. 2021 for a recent review). The usual argument in favour of adopting an environmental contour approach is that the contour can be estimated without knowledge of the struc-

ture; this is correct. However, once the contour is applied to assess the reliability of an actual structure, an assumption is made that the contour boundary is informative for environmental conditions likely to generate extreme structural responses. In this work, we demonstrate that this assumption is not correct in general.

Results comparing CDE_P with IFORM contour estimates highlight two deficiencies of the IFORM method, and hence of design contours in general. First, although design contours are intended to be conservative by construction, they are unable to reflect the type of structure under consideration and so may perform well for one structure but not for another. Second, identification of a good environmental model underpinning the contour is challenging, and there is considerable variability in the location of the contour due to the choice of environmental model; current practice tends to ignore this source of uncertainty. More generally, IFORM represents just one of a number of possible variants of environmental design contours (see Ross et al. 2020; Haselsteiner et al. 2021; Mackay and Haselsteiner 2021; Mackay and de Hauteclocque 2023); it is often not clear which contour is most appropriate in a given application. (CDE_P allows us to identify the most appropriate contour approach for a given application, but at the cost of undertaking the full probabilistic analysis; in this instance, we would always choose the full probabilistic analysis in favour of the contour approach.)

Within the forward model, we adopt the conditional extremes model of Heffernan and Tawn (2004) to estimate the joint distribution of extremes only of environmental variables. Using the forward model, we estimate the conditional density of the environment (CDE_P), given an extreme response corresponding to a given return period P. We then use CDE_P to compare with different environmental contours corresponding to the same return period. We use the IFORM procedure to estimate the environmental contour, which requires a statistical model for the full joint distribution of the environmental variables. We explore a range of hierarchical models for this purpose, reflecting the model forms typically used in the ocean engineering literature; these typically do

not include the conditional extremes model. In particular, we establish that fitting of hierarchical models for H_S and $S_2|H_S$ is a large source of uncertainty in environmental contour estimation (see also e.g. de Hauteclocque et al. 2022). We acknowledge that in future work it would also be interesting to explore the conditional extremes model for contour estimation further (following the work of e.g. Jonathan et al. 2010, 2012, 2014a; Ross et al. 2020 or Towe et al. 2024). In fact, Jonathan et al. (2010) already provides a direct comparison of environmental contours estimated using the H_S, $T_P|H_S$ hierarchical model with that estimated using the conditional extremes model, where T_P is the spectral peak period. See also Tendijck et al. (2023b) for a discussion on extremal characteristics of hierarchical models.

We employ techniques for selecting the extreme thresholds in the models of Davison and Smith (1990) and Heffernan and Tawn (2004) that can lead to subjective choices for each. Recent work by Murphy et al. (2025) provides an automated approach to threshold selection that eliminates this subjectivity in the univariate case. Additional methods for handling the choice of conditioning threshold for the model of Heffernan and Tawn (2004) exist, such as testing for the independence between exceedance and residual values (discussed by Jonathan et al. 2012) or bootstrap sampling to quantify the uncertainty in parameter estimates due to threshold choice (see e.g., Jonathan et al. 2010). Future analysis could therefore be improved by developing methods to automate the choice of conditional model threshold which incorporate the aforementioned techniques, to be used alongside the univariate method of Murphy et al. (2025).

In this work, simple stick structures are considered, with a response dependent on only two environmental variables (H_S and S_2). We believe that this framework provides sufficiently realistic wave surface and kinematic models for our structure types. Despite this, future analysis might benefit from the inclusion of more complex structure models and wave-structure interactions. In reality, there are factors, such as wind and current, alongside additional directional and seasonal covariate effects, not accommodated in

this work. Structure models that include effects such as local loading and wave breaking could also be utilised, alongside improved models for the environment itself. For example, linear wave theory may be extended by transforming linear wave characteristics to their respective non-linear equivalents following the approach outlined in Swan (2020) and Gibson (2020). Use of more complex structure and environment models, however, incurs higher computational cost, and so an approach for efficient estimation of CDE_P that avoids the need for numerical simulation from fluid loading models may be desirable. Moreover, adoption of methods of full probabilistic structural design must be undertaken with care, to ensure rational evidence-based evolution of design procedures. For example, Standard Norge (2022) identifies that a number of the features of the methodology of the LOADS joint industry project (Swan, 2020; Gibson, 2020) are either not compatible with the NORSOK standard, or not yet sufficiently well stress-tested for adoption within the standard. From this perspective, contours retain the advantages of being less computationally costly to employ, whilst also only requiring knowledge of the joint environment.

One approximate approach to reduce the bias in extreme response and associated risk statistics estimated from an environmental contour is to calibrate contour characteristics for a specific structural archetype. The necessary calibration would be estimated by applying the full forward probabilistic analysis for the structural archetype, then adjusting contour characteristics and/or associated calculations to reduce the observed bias in extreme response, or any other statistic of interest, estimated using the contour; indeed, the CDE might prove a useful basis for contour calibration. It might be possible to apply different calibration corrections to the contour in different parts of the environmental space in a systematic manner, with reduced need for user judgement, so that the contour better mimics CDE. The appropriate calibration could then be used with an environmental contour applied to that specific archetype. Note that this adjustment of contour characteristics to accommodate structural characteris-

tics, is quite different from contour adjustment for short-term environmental variability (recommended by some standards; e.g., NORSOK N-003 2017), and to adjustment of contour estimates made from serially-correlated data (e.g., de Hauteclocque et al. 2022).

During the review process, one referee queried (a) the usefulness of CDE_P as a diagnostic for environmental contours, and (b) whether the arguments given in the current work are an adequate reflection of the dangers (or otherwise) of environmental contour methods for design. We believe that the novelty of CDE_P as a diagnostic for environmental contours stems from its ability to capture the specific regions of the environment space responsible for extreme structural response, in a clear and systematic fashion with little need for user judgement. In contrast, current evaluation (and calibration) of environmental contours focuses only on comparison of responses calculated at user-selected points on the contour frontier with the relevant percentiles of the marginal short-term response distribution (e.g., Ross et al. 2020). The additional information provided by CDE_P, in our opinion, provides improved qualitative understanding together with a basis for systematic quantification of the drawbacks of (IFORM) environmental contours when applied to different structural archetypes. The current analysis illustrates the benefits of full probabilistic structural analysis relative to approximate analysis using environmental contours. Wherever possible, we recommend the application of full probabilistic structural design, or alternatively of contour methods carefully calibrated for specific structural archetypes using full probabilistic analysis, potentially using the conditional density of the environment as a basis for calibration.

3.6 Acknowledgements

The work was completed while Matthew Speers was part of the EPSRC funded STOR-i centre for doctoral training (grant no. EP/S022252/1), with part-funding from the ARC TIDE Industrial Transformational Research Hub at the University of Western

Australia. The authors wish to acknowledge the support of colleagues at Lancaster University and Shell.

3.7 Supplementary material

3.7.1 Overview

Here we present further details on aspects of the methodology covered in the main text. Section 3.7.2 describes the diagnostic techniques used to select the marginal exceedance thresholds for the generalised Pareto distribution (GPD) tail model (3.3.4) of Section 3.3.4. In Section 3.7.3, we provide further details on the origin of the wave kinematics equations given in Section 3.3.3, as well as information on the efficient simulation of wave fields from this model. Section 3.7.4 gives the specific parametric form of the models considered for $S_2|H_S$ and $S_2^-|H_S$ in Section 3.4.3, as well as the complete results summarised by Table 1 in Section 3.5.2.

3.7.2 Univariate threshold selection for significant wave height modelling

The suitable threshold for the modelling of H_S and S_2 via (4) in Section 3.2.2 is selected using standard extreme value diagnostics, such as stability plots and mean-residual-life plots (see Coles 2001 for examples). These methods are summarised below.

First, we consider the choice of threshold u_{H_S} for H_S. Figure 3.7.7 shows the values of the GPD standardised scale $\sigma^* = \sigma_u - \xi u$ (for σ_u estimated with threshold $u > 0$) and shape parameter ξ for choices of threshold non-exceedance probability $p_{u_{H_S}} = \tilde{F}_{H_S}(u_{H_S})$, alongside their respective 95% confidence intervals, obtained using block bootstrapping. The estimates of each parameter appear stable beyond the 0.8^{th} percentile. We thus select $u_{H_S} = \tilde{F}_{H_S}^{-1}(0.8)$ as the exceedance threshold for the marginal

modelling of H_S. We select threshold u_{S_2} for marginal modelling of S_2 using the same approach. Figure 3.7.8 shows the equivalent plots obtained when fitting to S_2. Again, we see stability of parameter estimates for thresholds with exceedance probability past 0.8, and so we select $u_{S_2} = \tilde{F}_{S_2}^{-1}(0.8)$ as the exceedance threshold for the marginal modelling of S_2.

We verify these threshold choices using mean residual life plots (see Coles 2001 for details). The form of the GPD implies that a linear trend in mean excesses with respect to threshold will occur above thresholds satisfying the model conditions. Figure 3.7.9 shows that there is a linear trend in mean excess point estimates when conditioning on thresholds above the 0.8^{th} percentile for both H_S and S_2.

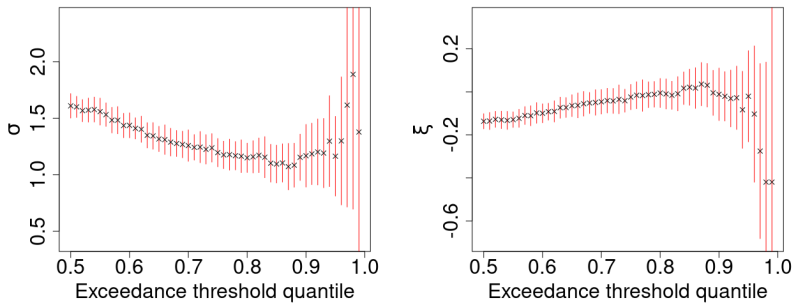


Figure 3.7.7: Stability plots for estimated values of the GPD scale σ and shape ξ parameters when fitting to H_S peak data, with respect to varying levels of the conditioning threshold percentile. Point estimates obtained via maximum likelihood estimation are marked in black. Corresponding 95% confidence intervals are shown in red.

3.7.3 Wave field simulation

Airy wave theory

Section 3.3.2 introduces the wave model of Taylor et al. (1997) which is itself derived from the work of Airy (1845), commonly referred to as linear wave theory, which provides physically based models for the surface elevation of the wave surface and associated kinematics. Airy (1845) models the stochastic surface elevation E at time $t \in \mathbb{R}$ and

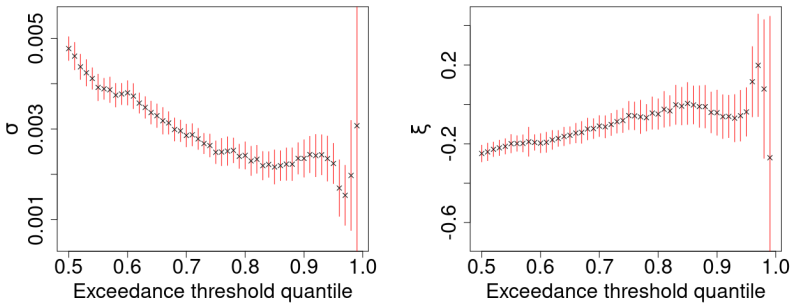


Figure 3.7.8: Stability plots for estimated values of the GPD scale σ and shape ξ parameters when fitting to S_S peak data, with respect to varying levels of the conditioning threshold percentile. Point estimates obtained via maximum likelihood estimation are marked in black. Corresponding 95% confidence intervals are shown in red.

location $x \in \mathbb{R}$ as

$$E(t;x) = \sum_{n=1}^{N} \left\{ A_n \cos\left(\omega_n t - k_n x\right) + B_n \sin\left(\omega_n t - k_n x\right) \right\},$$

with contributing angular frequencies $\omega_n > 0$, coefficients $A_n, B_n > 0$ and wave numbers k_n as defined in Section 3.3.3, for $n = 1, \ldots, N$. This is the non-conditioned form of expression (10), for an arbitrary location $x \in \mathbb{R}$. This model ensures conservation of mass, momentum satisfaction of necessary boundary conditions in a simple setting; see Holthuijsen (2010) for details.

Airy (1845) also introduces the 'velocity potential' equation

$$\phi(t;x,z) = \sum_{n=1}^{N} \left\{ \frac{\omega_n \cosh[k_n(d+z)]}{k_n \sinh(k_n d)} \left[-A_n \sin(\omega_n t - k_n x) + B_n \cos(\omega_n t - k_n x) \right] \right\}, \tag{3.7.19}$$

for time $t \in \mathbb{R}$, horizontal position $x \in \mathbb{R}$ and vertical position $z \in \mathbb{R}$ relative to mean surface level, with water depth d. As in the main text, all physical quantities are given in SI units. Expression (3.7.19) possesses the property

$$U(t;x,z) = \frac{\partial \phi}{\partial x} \quad \text{and} \quad V(t;x,z) = \frac{\partial \phi}{\partial z},$$

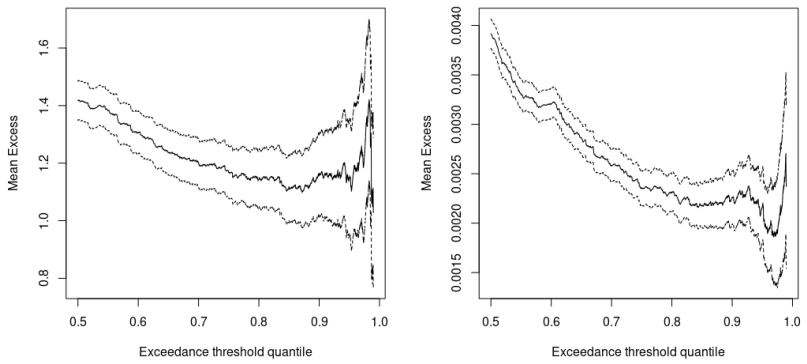


Figure 3.7.9: Mean residual life (mean excess) plots for excesses of H_S and S_2. Point estimates and 95% confidence intervals of mean excess with given exceedance probability are shown. The GPD implies a linear trend in mean excesses with respect to threshold. Thus, exceedance probabilities for which this linear trend is observed provide valid exceedance thresholds.

where $U(t;x,z)$ and $V(t;x,z)$ are horizontal and vertical velocities respectively. Hence, differentiation of (3.7.19) yields

$$U(t;x,z) = \sum_{n=1}^{N} \left[\{A_n \cos(\omega_n t - k_n x) + B_n \sin(\omega_n t - k_n x)\} \, \omega_n \frac{\cosh\left[k_n(d+z)\right]}{\sinh(k_n d)} \right],$$

$$V(t;x,z) = \sum_{n=1}^{N} \left[\{-A_n \sin(\omega_n t - k_n x) + B_n \cos(\omega_n t - k_n x)\} \, \omega_n \frac{\sinh\left[k_n(d+z)\right]}{\sinh(k_n d)} \right],$$

$$\dot{U}(t;x,z) = \sum_{n=1}^{N} \left[\{-A_n \sin(\omega_n t - k_n x) + B_n \cos(\omega_n t - k_n x)\} \, \omega_n^2 \frac{\cosh\left[k_n(d+z)\right]}{\sinh(k_n d)} \right],$$

$$\dot{V}(t;x,z) = \sum_{n=1}^{N} \left[\{-A_n \cos(\omega_n t - k_n x) - B_n \sin(\omega_n t - k_n x)\} \, \omega_n^2 \frac{\cosh\left[k_n(d+z)\right]}{\sinh(k_n d)} \right],$$

for horizontal velocity U, horizontal acceleration $\dot{U}$, vertical velocity V and vertical acceleration $\dot{V}$. Note that in the main text we only utilise models derived from the above expressions for the horizontal velocity U and acceleration $\dot{U}$ with $x = 0$; the expressions for V and $\dot{V}$ are included here for completeness.

JONSWAP wave spectrum

In Section 3.3.3, we assume the wave spectrum $S(\omega; \mathbf{X})$ has a JONSWAP parametric form, which is shown by Hasselmann et al. (1973) to be suitably flexible to capture measure offshore spectral behaviour. The JONSWAP spectrum has spectral density function

$$S(\omega; \mathbf{X} = \mathbf{x}) = \alpha\omega^{-r} \exp\left\{-\frac{r}{4}\left(\frac{|\omega|}{\omega_p(\mathbf{x})}\right)^{-4}\right\} \gamma\delta(\omega; \mathbf{x}),$$

for $\omega > 0$, where $\mathbf{X} = (H_s, S_2)$ and $\omega_p(\mathbf{x}) = 2\pi/t(\mathbf{x})$, where $t(\mathbf{x})$ is the observed value of the second spectral moment wave period $T_2 = \sqrt{(2\pi H_S)/(gS_2)}$ in sea state $\mathbf{X} = \mathbf{x}$, with

$$\delta(\omega; \mathbf{x}) = \exp\left\{-\frac{1}{2\left(0.07 + 0.02 \cdot I\{\omega_p(\mathbf{x}) > |\omega|\}\right)^2}\left(\frac{|\omega|}{\omega_p(\mathbf{x})} - 1\right)^2\right\},$$

and constants $r, \alpha, \gamma > 0$. The normalising constant α is chosen so that

$$4 \cdot \left\{\int_{-\infty}^{\infty} S(\omega; \mathbf{x}) \mathrm{d}\omega\right\}^{\frac{1}{2}} = h(\mathbf{x}),$$

where $h(\mathbf{x})$ is the observed value of significant wave height H_S in sea state $\mathbf{X} = \mathbf{x}$.

Efficient wave simulation using the fast Fourier transform

We apply the fast Fourier transform (FFT) algorithm (Nussbaumer, 1982) to efficiently compute the linear wave behaviour given by expressions (3.3.10) and (3.3.11) of Section 3.3.3, by formulating them as the discrete Fourier transforms of appropriately chosen 'link functions'. Specifically, for

$$A'_n = A_n + \mathcal{Q}\sigma_n^2 \quad \text{and} \quad B'_n = B_n + \mathcal{R}\sigma_n^2\omega_n,$$

we write

$$E(t_j) = \sum_{n=1}^{N} g_n^{(\eta)} \exp\left\{\frac{-2\pi i}{N} jn\right\}, \quad g_n^{(\eta)} = A'_n + B'_n i,$$
$$U(t_j, z) = \sum_{n=1}^{N} g_n^{(u)} \exp\left\{\frac{-2\pi i}{N} jn\right\}, \quad g_n^{(u)} = (A'_n + B'_n i) \cdot \omega_n \cdot \frac{\cosh\{k_n(z+d)\}}{\sinh(k_n d)},$$
$$\dot{U}(t_j, z) = \sum_{n=1}^{N} g_n^{(\dot{u})} \exp\left\{\frac{-2\pi i}{N} jn\right\}, \quad g_n^{(\dot{u})} = (B'_n - A'_n i) \cdot \omega_n^2 \cdot \frac{\cosh\{k_n(z+d)\}}{\sinh(k_n d)},$$

where $i^2 = -1$ and $j = 1, \ldots, N$. This allows us to utilise the fast Fourier transform (Nussbaumer, 1982) to efficiently compute the linear wave behaviour given by expressions (3.3.10) and (3.3.11) for $z \in \mathbb{R}$ at times $\{t_j\}_{j=1}^N$. Due to the intractability of the above kinematics at values of $z > 0$, we employ simple linear kinematic stretching and evaluate the above equations at

$$z' = \begin{cases} z & \text{for } z \leq 0, \\ 0 & \text{for } z > 0, \end{cases}$$

rather than z.

3.7.4 Storm peak hierarchical model selection

Distributional forms for steepness modelling

Here we give the parametric forms of the four distributions considered for the modelling of $S_2|H_S$ and $S_2^-|H_S$ in Section 3.4.3, described by the most appropriate of their distribution and density functions. First, is the Generalised Extreme Value (GEV) distribution, which has distribution function

$$F_{\text{GEV}}(x) = \exp\left\{-\left[1 + \xi\left(\frac{x - \mu_G}{\sigma_G}\right)\right]^{-1/\xi}\right\},$$

defined on the set $\{x : 1 + \xi(x - \mu)/\sigma > 0\}$, with location $\mu_G \in \mathbb{R}$, scale $\sigma_G > 0$ and shape $\xi \in \mathbb{R}$. Second, the Weibull distribution, which has distribution function

$$F_{\text{Weibull}}(x) = 1 - e^{-(x/\lambda)^k},$$

for $x \geq 0$, with scale $\lambda > 0$ and shape $k > 0$. Third, the Lognormal distribution, with density function

$$f_{\text{Lognormal}}(x) = \frac{1}{x\sigma_L\sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu_L)^2}{2\sigma_L^2}\right),$$

for $x > 0$, with location μ_L and log-scale $\sigma_L > 0$. Fourth, the gamma distribution, with density function

$$f_{\text{Gamma}}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x},$$

where Γ is the gamma function, for $x \geq 0$, with shape $\alpha > 0$ and rate $\beta > 0$.

Detailed cross validation results

Here we present the full details of the results summarised by Table 1 in Section 3.5.2, summarised in Table 3.7.4 which shows the AIC, cross validation scores and AS for all 72 candidate models for $S_2|H_S$ and $S_2^-|H_S$. The distributional form and function of H_S imposed on distribution parameters are given, as well as individual scores for each of the assessment criteria. Columns titled 'p_vCVK' contain K-fold cross validation scores for an extreme threshold $v = \tilde{F}_{S_2}(p_v)$. The AS score is obtained by averaging of AIC and cross validation scores. Each cross validation scenario is repeated for 30 replicates for each data set, allowing the calculation of standard errors. The highest ranking model per distribution per scoring method is marked in bold. In the event of a tie, we opt for the model with the highest AIC.

Distribution	Parameter Forms	AIC	0.0CV5	0.0CV10	0.8CV5	0.8CV10	0.9CV10	0.9CV10	Aggregate Score (AS)
$S_2^-\|H_S \sim$ GEV	Exp μ_G, Exp σ_G	**3.931**	3.931 (0.002)	3.931 (0.001)	4.032 (0.014)	4.037 (0.004)	4.024 (0.002)	**4.11 (0.002)**	**3.999 (0.002)**
$S_2^-\|H_S \sim$ GEV	Exp μ_G, Lin σ_G	3.930	**3.937 (0.004)**	**3.936 (0.003)**	3.945 (0.027)	3.959 (0.014)	3.998 (0.004)	4.084 (0.005)	3.97 (0.005)
$S_2^-\|H_S \sim$ GEV	Exp μ_G, Qua σ_G	3.927	3.93 (0.001)	3.93 (0.001)	**4.034 (0.007)**	4.041 (0.003)	**4.025 (0.005)**	4.106 (0.004)	**3.999 (0.002)**
$S_2^-\|H_S \sim$ GEV	Lin μ_G, Exp σ_G	3.918	3.919 (0.001)	3.919 (0.001)	4.012 (0.006)	4.014 (0.005)	3.975 (0.001)	4.061 (0.001)	3.974 (0.004)
$S_2^-\|H_S \sim$ GEV	Lin μ_G, Lin σ_G	3.913	3.926 (0.004)	3.925 (0.003)	3.953 (0.017)	3.951 (0.01)	3.975 (0.005)	4.062 (0.001)	3.958 (0.003)
$S_2^-\|H_S \sim$ GEV	Lin μ_G, Qua σ_G	3.915	3.917 (0.001)	3.916 (0.001)	4.015 (0.007)	4.016 (0.003)	3.98 (0.004)	4.06 (0.003)	3.974 (0.001)
$S_2^-\|H_S \sim$ GEV	Qua μ_G, Exp σ_G	3.929	3.929 (0.004)	3.93 (0.001)	4.026 (0.012)	4.029 (0.01)	4.011 (0.013)	4.101 (0.01)	3.994 (0.003)
$S_2^-\|H_S \sim$ GEV	Qua μ_G, Lin σ_G	3.927	3.934 (0.004)	3.933 (0.004)	3.955 (0.016)	3.97 (0.016)	3.992 (0.012)	4.086 (0.008)	3.971 (0.004)
$S_2^-\|H_S \sim$ GEV	Qua μ_G, Qua σ_G	3.926	3.927 (0.003)	3.927 (0.002)	4.028 (0.012)	**4.043 (0.008)**	4.002 (0.018)	4.105 (0.007)	3.994 (0.004)
$S_2^-\|H_S \sim$ Weibull	Exp k, Exp λ	**3.745**	**3.744 (0.001)**	**3.744 (0.001)**	**3.99 (0.003)**	**3.99 (0.002)**	3.992 (0.008)	4.075 (0.007)	**3.897 (0.002)**
$S_2^-\|H_S \sim$ Weibull	Exp k, Lin λ	3.679	3.666 (0.017)	3.672 (0.006)	3.647 (0.258)	3.843 (0.035)	3.656 (0.238)	3.809 (0.088)	3.71 (0.047)
$S_2^-\|H_S \sim$ Weibull	Exp k, Qua λ	3.711	3.726 (0.005)	3.721 (0.003)	3.955 (0.01)	3.956 (0.009)	3.942 (0.013)	4.019 (0.015)	3.861 (0.004)
$S_2^-\|H_S \sim$ Weibull	Lin k, Exp λ	3.743	3.743 (0.002)	3.743 (0.001)	3.982 (0.009)	3.987 (0.004)	**3.993 (0.008)**	**4.081 (0.006)**	3.896 (0.002)
$S_2^-\|H_S \sim$ Weibull	Lin k, Lin λ	3.578	3.584 (0.017)	3.576 (0.02)	3.828 (0.017)	3.855 (0.008)	3.829 (0.014)	3.852 (0.015)	3.729 (0.004)
$S_2^-\|H_S \sim$ Weibull	Lin k, Qua λ	3.697	3.695 (0.002)	3.695 (0.002)	3.885 (0.006)	3.891 (0.007)	3.807 (0.014)	3.897 (0.015)	3.795 (0.003)
$S_2^-\|H_S \sim$ Weibull	Qua k, Exp λ	3.732	3.734 (0.007)	3.735 (0.005)	3.944 (0.016)	3.975 (0.009)	3.95 (0.021)	4.061 (0.012)	3.876 (0.004)
$S_2^-\|H_S \sim$ Weibull	Qua k, Lin λ	3.718	3.715 (0.003)	3.716 (0.002)	3.905 (0.011)	3.909 (0.009)	3.859 (0.021)	3.948 (0.021)	3.824 (0.005)
$S_2^-\|H_S \sim$ Weibull	Qua k, Qua λ	3.712	3.697 (0.014)	3.691 (0.014)	3.87 (0.034)	3.852 (0.028)	3.832 (0.067)	3.944 (0.027)	3.8 (0.013)
$S_2^-\|H_S \sim$ Lognormal	Exp μ_L, Exp σ_L	3.887	3.855 (0.04)	3.864 (0.026)	3.765 (0.118)	3.924 (0.054)	3.664 (0.177)	3.987 (0.108)	3.849 (0.037)
$S_2^-\|H_S \sim$ Lognormal	Exp μ_L, Lin σ_L	3.889	3.842 (0.04)	3.802 (0.049)	3.94 (0.039)	3.928 (0.042)	3.93 (0.076)	3.909 (0.066)	3.891 (0.018)
$S_2^-\|H_S \sim$ Lognormal	Exp μ_L, Qua σ_L	3.891	3.886 (0.021)	3.877 (0.029)	3.976 (0.006)	3.979 (0.011)	3.997 (0.003)	4.082 (0.004)	3.955 (0.007)
$S_2^-\|H_S \sim$ Lognormal	Lin μ_L, Exp σ_L	3.894	3.893 (0.002)	3.893 (0.002)	3.958 (0.008)	3.964 (0.005)	3.967 (0.01)	4.052 (0.006)	3.946 (0.002)
$S_2^-\|H_S \sim$ Lognormal	Lin μ_L, Lin σ_L	**3.896**	**3.897 (0.001)**	**3.897 (0.001)**	**3.981 (0.004)**	**3.983 (0.003)**	**3.999 (0.003)**	**4.085 (0.002)**	**3.963 (0.001)**
$S_2^-\|H_S \sim$ Lognormal	Lin μ_L, Qua σ_L	3.857	3.873 (0.01)	3.868 (0.008)	3.932 (0.026)	3.911 (0.015)	3.874 (0.038)	3.935 (0.032)	3.893 (0.010)
$S_2^-\|H_S \sim$ Lognormal	Qua μ_L, Exp σ_L	3.880	3.862 (0.037)	3.874 (0.017)	3.884 (0.063)	3.843 (0.054)	3.836 (0.057)	3.875 (0.075)	3.865 (0.019)
$S_2^-\|H_S \sim$ Lognormal	Qua μ_L, Lin σ_L	3.891	3.892 (0.003)	3.892 (0.001)	3.97 (0.004)	3.972 (0.003)	3.976 (0.004)	4.063 (0.002)	3.951 (0.001)
$S_2^-\|H_S \sim$ Lognormal	Qua μ_L, Qua σ_L	3.881	3.879 (0.008)	3.877 (0.006)	3.925 (0.032)	3.919 (0.035)	3.95 (0.037)	4.03 (0.019)	3.923 (0.010)
$S_2^-\|H_S \sim$ Gamma	Exp α, Exp β	3.801	3.858 (0.009)	3.854 (0.009)	4.016 (0.023)	3.998 (0.014)	3.917 (0.035)	4.079 (0.011)	3.932 (0.007)
$S_2^-\|H_S \sim$ Gamma	Exp α, Lin β	3.775	3.815 (0.021)	3.821 (0.016)	3.928 (0.021)	3.923 (0.015)	3.832 (0.027)	3.921 (0.037)	3.859 (0.007)
$S_2^-\|H_S \sim$ Gamma	Exp α, Qua β	**3.876**	3.864 (0.006)	3.863 (0.004)	3.971 (0.02)	3.965 (0.015)	3.915 (0.023)	3.982 (0.02)	3.919 (0.005)
$S_2^-\|H_S \sim$ Gamma	Lin α, Exp β	3.872	3.872 (0.002)	3.873 (0.001)	**4.02 (0.005)**	**4.021 (0.005)**	3.991 (0.007)	4.08 (0.006)	3.961 (0.002)
$S_2^-\|H_S \sim$ Gamma	Lin α, Lin β	3.871	**3.872 (0.002)**	**3.873 (0.001)**	4.008 (0.005)	4.01 (0.008)	3.974 (0.006)	4.06 (0.005)	3.953 (0.002)
$S_2^-\|H_S \sim$ Gamma	Lin α, Qua β	3.866	3.868 (0.001)	3.868 (0)	4.001 (0.005)	4.004 (0.003)	3.959 (0.004)	4.046 (0.004)	3.945 (0.001)
$S_2^-\|H_S \sim$ Gamma	Qua α, Exp β	3.823	3.826 (0.004)	3.825 (0.003)	3.94 (0.052)	3.947 (0.01)	3.881 (0.022)	3.955 (0.016)	3.885 (0.009)
$S_2^-\|H_S \sim$ Gamma	Qua α, Lin β	3.875	3.866 (0.007)	3.866 (0.006)	4.017 (0.019)	4.019 (0.016)	**4.015 (0.017)**	**4.081 (0.025)**	**3.963 (0.006)**
$S_2^-\|H_S \sim$ Gamma	Qua α, Qua β	3.837	3.844 (0.007)	3.845 (0.006)	3.876 (0.024)	3.918 (0.026)	3.823 (0.051)	3.9 (0.046)	3.863 (0.012)
$S_2\|H_S \sim$ GEV	Exp μ_G, Exp σ_G	3.567	3.446 (0.035)	3.479 (0.021)	**3.822 (0.044)**	**3.862 (0.023)**	**3.919 (0.01)**	**3.973 (0.009)**	3.724 (0.010)
$S_2\|H_S \sim$ GEV	Exp μ_G, Lin σ_G	3.495	3.399 (0.023)	3.42 (0.013)	3.586 (0.026)	3.561 (0.031)	3.677 (0.013)	3.734 (0.009)	3.553 (0.007)
$S_2\|H_S \sim$ GEV	Exp μ_G, Qua σ_G	3.555	3.532 (0)	3.531 (0)	3.82 (0.001)	3.816 (0.001)	3.875 (0.001)	3.954 (0)	**3.726 (0.000)**
$S_2\|H_S \sim$ GEV	Lin μ_G, Exp σ_G	3.370	3.305 (0.016)	3.316 (0.012)	3.537 (0.105)	3.25 (0.075)	3.025 (0.178)	2.674 (0.088)	3.211 (0.032)
$S_2\|H_S \sim$ GEV	Lin μ_G, Lin σ_G	3.655	3.551 (0.009)	3.555 (0.013)	3.645 (0.052)	3.619 (0.041)	3.29 (0.172)	3.482 (0.032)	3.542 (0.028)
$S_2\|H_S \sim$ GEV	Lin μ_G, Qua σ_G	**3.688**	**3.697 (0.002)**	**3.694 (0.001)**	3.558 (0.003)	3.554 (0.002)	3.237 (0.004)	3.305 (0.002)	3.533 (0.001)
$S_2\|H_S \sim$ GEV	Qua μ_G, Exp σ_G	3.638	3.436 (0.067)	3.567 (0.008)	3.67 (0.019)	3.673 (0.017)	3.575 (0.032)	3.586 (0.024)	3.592 (0.012)
$S_2\|H_S \sim$ GEV	Qua μ_G, Lin σ_G	3.287	3.104 (0.213)	3.256 (0.089)	1.966 (0.087)	1.686 (0.037)	1.344 (0.017)	1.323 (0)	2.281 (0.038)
$S_2\|H_S \sim$ GEV	Qua μ_G, Qua σ_G	3.356	2.89 (0.129)	3.094 (0.06)	2.365 (0.205)	2.535 (0.122)	1.979 (0.149)	2.567 (0.06)	2.684 (0.047)
$S_2\|H_S \sim$ Weibull	Exp k, Exp λ	3.864	3.867 (0.002)	3.868 (0.002)	3.966 (0.007)	3.979 (0.005)	3.87 (0.011)	3.978 (0.013)	3.913 (0.003)
$S_2\|H_S \sim$ Weibull	Exp k, Lin λ	3.880	3.885 (0.002)	3.885 (0.002)	4.022 (0.006)	4.024 (0.005)	3.961 (0.009)	4.05 (0.007)	3.958 (0.002)
$S_2\|H_S \sim$ Weibull	Exp k, Qua λ	3.897	3.9 (0.002)	3.9 (0.001)	4.04 (0.006)	4.041 (0.004)	3.995 (0.02)	4.088 (0.005)	3.98 (0.003)
$S_2\|H_S \sim$ Weibull	Lin k, Exp λ	3.872	3.873 (0.001)	3.874 (0.001)	3.984 (0.005)	3.989 (0.004)	3.902 (0.004)	3.994 (0.006)	3.927 (0.002)
$S_2\|H_S \sim$ Weibull	Lin k, Lin λ	3.888	3.891 (0.003)	3.891 (0.002)	4.008 (0.012)	4.013 (0.009)	3.96 (0.007)	4.022 (0.015)	3.953 (0.003)
$S_2\|H_S \sim$ Weibull	Lin k, Qua λ	3.901	**3.902 (0.006)**	**3.903 (0.003)**	4.032 (0.006)	4.038 (0.006)	**4.006 (0.007)**	**4.096 (0.008)**	**3.983 (0.003)**
$S_2\|H_S \sim$ Weibull	Qua k, Exp λ	3.871	3.872 (0.001)	3.871 (0.001)	3.978 (0.009)	3.986 (0.006)	3.892 (0.009)	3.989 (0.006)	3.923 (0.003)
$S_2\|H_S \sim$ Weibull	Qua k, Lin λ	3.889	3.89 (0.002)	3.89 (0.001)	4.018 (0.005)	4.021 (0.002)	3.957 (0.005)	4.047 (0.004)	3.959 (0.001)
$S_2\|H_S \sim$ Weibull	Qua k, Qua λ	**3.903**	3.88 (0.044)	3.9 (0.01)	**4.046 (0.003)**	**4.043 (0.002)**	**4.006 (0.003)**	**4.096 (0.005)**	3.982 (0.007)
$S_2\|H_S \sim$ Lognormal	Exp μ_L, Exp σ_L	3.629	3.632 (0.007)	3.636 (0.005)	3.872 (0.021)	3.869 (0.01)	3.8 (0.022)	3.891 (0.025)	3.761 (0.006)
$S_2\|H_S \sim$ Lognormal	Exp μ_L, Lin σ_L	3.485	3.485 (0.006)	3.497 (0.006)	3.803 (0.009)	3.807 (0)	3.861 (0)	3.941 (0)	3.697 (0.002)
$S_2\|H_S \sim$ Lognormal	Exp μ_L, Qua σ_L	3.563	3.562 (0.002)	3.562 (0.001)	3.896 (0.004)	3.897 (0.002)	3.926 (0.006)	**4.008 (0.004)**	3.773 (0.001)
$S_2\|H_S \sim$ Lognormal	Lin μ_L, Exp σ_L	3.651	3.649 (0.003)	3.65 (0.002)	3.914 (0.008)	3.92 (0.007)	3.878 (0.011)	3.961 (0.016)	3.803 (0.003)
$S_2\|H_S \sim$ Lognormal	Lin μ_L, Lin σ_L	3.591	3.59 (0.003)	3.588 (0.003)	3.69 (0.016)	3.705 (0.012)	3.584 (0.027)	3.7 (0.015)	3.635 (0.005)
$S_2\|H_S \sim$ Lognormal	Lin μ_L, Qua σ_L	3.612	3.613 (0.001)	3.613 (0.001)	3.747 (0.004)	3.754 (0.003)	3.563 (0.004)	3.646 (0.006)	3.65 (0.001)
$S_2\|H_S \sim$ Lognormal	Qua μ_L, Exp σ_L	**3.669**	**3.657 (0.007)**	**3.661 (0.004)**	**3.942 (0.01)**	3.925 (0.016)	**3.947 (0.013)**	3.97 (0.028)	**3.824 (0.005)**
$S_2\|H_S \sim$ Lognormal	Qua μ_L, Lin σ_L	3.659	3.653 (0.011)	3.653 (0.01)	3.895 (0.062)	**3.928 (0.01)**	3.895 (0.055)	3.901 (0.326)	3.798 (0.046)
$S_2\|H_S \sim$ Lognormal	Qua μ_L, Qua σ_L	3.185	3.227 (0.021)	3.212 (0.016)	2.642 (0.074)	2.626 (0.035)	2.191 (0.099)	2.37 (0.061)	2.779 (0.765)
$S_2\|H_S \sim$ Gamma	Exp α, Exp β	3.608	3.606 (0.008)	3.605 (0.005)	3.694 (0.011)	3.685 (0.01)	3.66 (0.017)	3.773 (0.011)	3.662 (0.003)
$S_2\|H_S \sim$ Gamma	Exp α, Lin β	3.669	3.643 (0.006)	3.645 (0.005)	3.795 (0.01)	3.802 (0.006)	3.752 (0.013)	3.843 (0.016)	3.736 (0.004)
$S_2\|H_S \sim$ Gamma	Exp α, Qua β	3.604	3.602 (0.015)	3.601 (0.013)	3.974 (0.014)	3.974 (0.015)	3.96 (0.016)	4.038 (0.028)	3.822 (0.004)
$S_2\|H_S \sim$ Gamma	Lin α, Lin β	**3.710**	**3.709 (0.004)**	**3.71 (0.002)**	3.975 (0.002)	3.975 (0.003)	3.972 (0.007)	**4.058 (0.006)**	**3.873 (0.002)**
$S_2\|H_S \sim$ Gamma	Lin α, Exp β	3.668	3.668 (0.003)	3.668 (0.003)	3.85 (0.008)	3.857 (0.005)	3.816 (0.007)	3.902 (0.007)	3.776 (0.002)
$S_2\|H_S \sim$ Gamma	Lin α, Qua β	3.692	3.665 (0.025)	3.662 (0.024)	3.91 (0.024)	3.911 (0.019)	3.898 (0.045)	3.995 (0.018)	3.819 (0.009)
$S_2\|H_S \sim$ Gamma	Qua α, Exp β	3.599	3.487 (0.138)	3.433 (0.112)	3.904 (0.071)	3.938 (0.011)	3.922 (0.01)	3.987 (0.026)	3.753 (0.027)
$S_2\|H_S \sim$ Gamma	Qua α, Lin β	3.671	3.67 (0.007)	3.668 (0.005)	3.908 (0.008)	3.898 (0.005)	3.836 (0.015)	3.993 (0.011)	3.806 (0.003)
$S_2\|H_S \sim$ Gamma	Qua α, Qua β	3.622	3.618 (0.005)	3.618 (0.004)	**3.985 (0.004)**	**3.985 (0.003)**	**3.974 (0.005)**	4.045 (0.014)	3.835 (0.003)

Table 3.7.4: AIC and cross validation scores of all model combinations for $S_2|H_S$ and $S_2^-|H_S$, for all considered distribution parameter functional forms. The highest scoring model per distribution per column is marked in bold. Sample standard errors are given to three decimal places.

Chapter 4

Sequential Design for the Efficient Estimation of Offshore Structure Failure Probability

4.1 Introduction

4.1.1 Background

An offshore structure (such as an oil platform or wind turbine) is subject to environmental loading, e.g., from winds, waves and currents. The ocean engineer seeks to evaluate the risk posed to structural integrity by the environment, enabling the structure to be designed and maintained to the required level of reliability. Often, this involves computationally demanding fluid loading and structural response calculations. Therefore, the design of computationally efficient approaches for assessment of structural risk is a topic of considerable importance.

Take an environmental variable $\mathbf{X}$ (such as significant wave height H_S) characterising the long term metocean environment on space $\mathcal{E}_{\mathbf{X}}$. The short term environment variable $\mathbf{Y}$ (such as individual wave height H) defined on space $\mathcal{E}_{\mathbf{Y}}$, depends on $\mathbf{X}$. This $\mathbf{Y}$ is

stochastic given $\mathbf{X}$, in the sense that many values of $\mathbf{Y}$ may be summarised by a single value $\mathbf{x}$ of $\mathbf{X}$, in terms of the distribution for $\mathbf{Y}|\{\mathbf{X} = \mathbf{x}\}$. Given complete knowledge of the short term conditions $\mathbf{Y}$, along with a physical model for the response $\mathbf{R} \in \mathcal{E}_{\mathbf{R}}$ induced on the structure by $\mathbf{Y}$, it is possible to characterise the *multivariate* structural response induced on the structure fully. Typically, $\mathcal{E}_{\mathbf{R}} = \mathbb{R}^d$ for some dimension $d > 0$.

In our setting, we assume the existence of a deterministic function $g_{\mathbf{R}}(\mathbf{y}) : \mathcal{E}_{\mathbf{Y}} \mapsto \mathcal{E}_{\mathbf{R}}$ for the structural response $\mathbf{R} = \mathbf{r}$ induced by the environment $\mathbf{Y} = \mathbf{y}$. Practitioners do not typically have knowledge of the full short term environment $\mathbf{Y}$, but instead have information on the long term summary variable $\mathbf{X}$. Since $\mathbf{Y}$ is not a deterministic function of $\mathbf{X}$, practitioners estimate the density function $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) : \mathcal{E}_{\mathbf{Y}} \times \mathcal{E}_{\mathbf{X}} \mapsto \mathbb{R}^+$ of the short term environment $\mathbf{Y}|\{\mathbf{X} = \mathbf{x}\}$. In practice, evaluation of $g_{\mathbf{R}}$ and $f_{\mathbf{Y}|\mathbf{X}}$ can be computationally expensive, involving complex load calculations and the simulation of 3-dimensional wave and wind fields.

Given the functions $g_{\mathbf{R}}$ and $f_{\mathbf{Y}|\mathbf{X}}$, we can evaluate the density $f_{\mathbf{R}|\mathbf{X}}(\mathbf{r}|\mathbf{x}) : \mathcal{E}_{\mathbf{R}} \times \mathcal{E}_{\mathbf{X}} \mapsto \mathbb{R}^+$ as

$$f_{\mathbf{R}|\mathbf{X}}(\mathbf{r}|\mathbf{x}) = \int_{\mathcal{E}_{\mathbf{Y}}} g_{\mathbf{R}}(\mathbf{y}) f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \mathrm{d}\mathbf{y}, \tag{4.1.1}$$

for $\mathbf{R}|\{\mathbf{X} = \mathbf{x}\}$, the multivariate response conditioned on the long term environment. Again, evaluating $f_{\mathbf{R}|\mathbf{X}}$ can be prohibitively expensive, due the potential complexities of $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{x})$ and $g_{\mathbf{R}}(\mathbf{y})$. A natural approach to quantify the risk to a structure is then to estimate the probability of failure p due to response $\mathbf{R}$ and environment $\mathbf{X}$. For $\mathbf{R} = (R_1, \ldots, R_d) \in \mathcal{E}_{\mathbf{R}}$, this can be written

$$p = \mathbb{P}\left(\bigcup_{i=1}^{d} (R_i > r_{\mathrm{Cr}}^{(i)})\right) = 1 - \mathbb{P}\left(\bigcap_{i=1}^{d} (R_i < r_{\mathrm{Cr}}^{(i)})\right),$$

the probability that at least one response component R_i, $i = 1, \ldots, d$, exceeds its critical

level $r_{\mathrm{Cr}}^{(i)} \in \mathbb{R}$. This can be written using (4.1.1) as

$$p = \int_{\mathcal{E}_{\mathbf{X}}} \left\{ \int_{\mathcal{E}_{\mathbf{R}}} \left[1 - \left(\prod_{i=1}^{d} I(R_i < r_{\mathrm{Cr}}^{(i)} | \{\mathbf{X} = \mathbf{x}\}) \right) \right] f_{\mathbf{R}|\mathbf{X}}(\mathbf{r}|\mathbf{x}) \mathrm{d}\mathbf{r} \right\} f_{\mathbf{X}}(\mathbf{x}) \mathrm{d}\mathbf{x}, \tag{4.1.2}$$

where I is the indicator function, which takes the value unity when its argument is true, and the value zero otherwise. The integral evaluated numerically by repeated Monte Carlo sampling from models for $\mathbf{R}|\{\mathbf{X} = \mathbf{x}\}$ and $\mathbf{X}$. Throughout, we assume the density $f_{\mathbf{X}}$ of the long term environment $\mathbf{X}$ is either known or estimable, possibly using extreme value techniques (e.g., as in Section 4.4). Evaluation of (4.1.2) is thus solely made difficult by the computational expense required to obtain draws of $\mathbf{R}|\{\mathbf{X} = \mathbf{x}\}$.

We aim to minimise the uncertainty in estimating (4.1.2) given a budget of a set number of realisations of $\mathbf{R}|\{\mathbf{X} = \mathbf{x}\}$. We use the available budget efficiently by making informed choices about the values of $\mathbf{X}$ at which to sample from $\mathbf{R}|\{\mathbf{X} = \mathbf{x}\}$. Typically, methods for the efficient evaluation of (4.1.2) should target values of $\mathbf{X}$ contributing most to the integral. In the simplest terms, this is achieved by targetting regions where the value of the integrand

$$\begin{aligned} \tilde{f}_{\mathbf{X}}(\mathbf{x}; \mathbf{r}_{\mathrm{Cr}}) &= \left\{ \int_{\mathcal{E}_{\mathbf{R}}} \left[1 - \left(\prod_{i=1}^{d} I(R_i < r_{\mathrm{Cr}}^{(i)} | \{\mathbf{X} = \mathbf{x}\}) \right) \right] f_{\mathbf{R}|\mathbf{X}}(\mathbf{r}|\mathbf{x}) \mathrm{d}\mathbf{r} \right\} \times f_{\mathbf{X}}(\mathbf{x}) \\ &= \mathbb{P}\left(\text{`failure'} | \{\mathbf{X} = \mathbf{x}\}\right) f_{\mathbf{X}}(\mathbf{x}), \end{aligned} \tag{4.1.3}$$

in (4.1.2) is large, where $\mathbf{r}_{\mathrm{Cr}} = \left(r_{\mathrm{Cr}}^{(1)}, \ldots, r_{\mathrm{Cr}}^{(d)} \right)$ is the vector of critical values of responses. That is, it is beneficial to target values of the long-term environmental variables that are both likely to occur (large $f_{\mathbf{X}}(\mathbf{x})$) and to induce structural failure (large $\mathbb{P}\left(\text{`failure'} | \{\mathbf{X} = \mathbf{x}\}\right)$). We subsequently refer to $\tilde{f}_{\mathbf{X}}(\mathbf{x}; \mathbf{r}_{\mathrm{Cr}})$ defined in (4.1.3) as the conditional density of the environment (CDE), as by Bayes theorem it is the unnormalised long-term environment density conditional on the occurrence of structural failure; we use $\tilde{f}$ (rather than f) to indicate an unnormalised density.

Peherstorfer et al. (2016), Yang et al. (2018) and Wang et al. (2021) show that minimising the uncertainty in (4.1.2) can be achieved for an arbitrary multi-dimensional response. We restrict ourselves to $d = 1$, with $\mathbf{R} = R$ and $\mathbf{r}_{\mathrm{Cr}} = r_{\mathrm{Cr}}$, for brevity and ease of presentation. In this case, equation (4.1.3) reduces to

$$\tilde{f}_{\mathbf{X}}(\mathbf{x}; r_{\mathrm{Cr}}) = \mathbb{P}\left(R > r_{\mathrm{Cr}} | \{\mathbf{X} = \mathbf{x}\}\right) \times f_{\mathbf{X}}(\mathbf{x}). \tag{4.1.4}$$

Existing methodologies to reduce uncertainty in (4.1.2) by targetting (4.1.4) include: *sampling* methods such as importance sampling (see e.g., Castellon et al. 2022) and bridge sampling (Meng and Wong, 1996); *adaptive Gaussian emulation* (e.g., Gramstad et al. 2020 and Lystad et al. 2023); and approaches combining sampling and adaptive emulation (e.g., Castellon et al. 2023 and Xiao et al. 2020). Good sampling methods reduce the variance of a target integral for a given sampling budget, whilst emulation provides a cheaper route to otherwise expensive complex function evaluation by approximation of the function. Relevant recent reviews are given by Moustapha et al. (2022), Wang et al. (2022), Tabandeh et al. (2022) and Marrel and Iooss (2024).

In simple cases, we might expect that the CDE $\tilde{f}_{\mathbf{X}}$ is approximately elliptically-contoured (e.g., Speers et al. 2024), and therefore well-approximated by a unimodal Gaussian-like density in $\mathcal{E}_{\mathbf{X}}$. However, in reality there are good reasons to expect this not to be the case in general, due to e.g., the presence of multiple failure modes or resonant responses. In the current work, we are particularly interested in investigating methodologies to estimate such complex CDE structures well.

We choose to investigate the efficient estimation of (4.1.4) in the context of designing monopile structures. We choose this structural type for two reasons: firstly, because it provides a useful template structure for generic studies of fluid loading; and secondly, it is of itself a relevant structural type for e.g., offshore wind applications. This thinking motivates the synthetic study of Section 4.3, and the wind turbine application of Section 4.4.

4.1.2 Objectives and outline

The objective of the current work is to explore methodologies based on efficient sampling or adaptive Gaussian emulation, to estimate the conditional density of the environment (CDE) and thereby failure probabilities for synthetic and real-world monopile structures. In Section 4.2, we first describe an approach, termed IS-PT, coupling importance sampling with parallel tempering Markov chain-Monte Carlo (MCMC) (Earl and Deem, 2005) for estimation of multi-modal CDEs, a scenario which has received little attention in the offshore reliability literature. Secondly, building on Gramstad et al. (2020) and Cohn (1993), we consider two variants of an alternative approach, termed AGE, based on adaptive Gaussian emulation, adopting an acquisition function promoting sampling that balances exploration and exploitation of regions of $\mathcal{E}_{\mathbf{X}}$ contributing to the CDE. In Section 4.3, the approaches from Section 4.2 are applied for a synthetic monopile structure exhibiting different resonant responses, to evaluate their respective performance. We find that all approaches provide good estimation of failure probability, but that AGE approaches require fewer expensive function evaluations provided that the required balance between exploitation and exploration of $\mathcal{E}_{\mathbf{X}}$ is assumed known. If this balance is unknown, IS-PT provides a more reliable procedure. In Section 4.4, we demonstrate good performance of all approaches in a more realistic setting, estimating the structural failure probability for oscillating monopiles, with harmonic response modelled using the T-FNV (Transformed - Faltinsen, Newman and Vinje, Faltinsen et al. 1995) model of Taylor et al. (2024). Our findings here regarding the relative computational complexities of IS-PT and AGE approaches are similar to those for the synthetic case. Discussion and conclusions are provided in Section 4.5. Supplementary material (SM) in Section 4.6 provides supporting description of methodology and results.

4.2 Methodology

4.2.1 Overview of methodologies

We begin by discussing two methods for the efficient evaluation of integral (4.1.2). In Section 4.2.2, we introduce an importance sampling scheme coupled with an adaptive parallel tempering MCMC algorithm, designed for scenarios where the CDE $\tilde{f}_{\mathbf{X}}$ is multimodal. In Section 4.2.3, we describe an emulator replacing expensive draws of the structural response $R|\{\mathbf{X} = \mathbf{x}\}$ with predictions from a Gaussian process, and provide methods for the adaptive design of the emulator training set.

We emphasise that these methods are introduced as *alternative* options for the efficient estimation of (4.1.2), both seeking to minimise the target error within some set budget of expensive function evaluations. These approaches will then be compared in (4.3) to see which performs better.

4.2.2 MCMC-informed importance sampling

In offshore reliability, importance sampling methods select values $\mathbf{x}_1^*, \ldots, \mathbf{x}_{n_{\mathrm{IS}}}^*$, $n_{\mathrm{IS}} > 0$, of $\mathbf{X}$ at which to evaluate $R|\{\mathbf{X} = \mathbf{x}\}$, to make efficient use of limited computational resources (see e.g., Castellon et al. 2022). Possible approaches include traditional importance sampling techniques, or extensions such as bridge sampling (e.g., Meng and Wong 1996). In this article, we focus our attention on the former, since our initial investigations of bridge sampling showed no improvement in performance, despite increased computational cost.

Evaluation points are drawn from a proposal distribution g_{Pr}, chosen so that values $\mathbf{x}$ with higher density $g_{\mathrm{Pr}}(\mathbf{x})$ are more informative to the target quantity. Our target

quantity is the marginal structural failure probability (4.1.2), which may be written

$$\begin{aligned} p &= \int_{\mathcal{E}_{\mathbf{X}}} \frac{\tilde{f}_{\mathbf{X}}(\mathbf{x}; r_{\mathrm{Cr}})}{g_{\mathrm{Pr}}(\mathbf{x})} g_{\mathrm{Pr}}(\mathbf{x}) \mathrm{d}\mathbf{x} \\ &= \int_{\mathcal{E}_{\mathbf{X}}} \mathbb{P}\left(R > r_{\mathrm{Cr}} | \{\mathbf{X} = \mathbf{x}\}\right) \frac{f_{\mathbf{X}}(\mathbf{x})}{g_{\mathrm{Pr}}(\mathbf{x})} g_{\mathrm{Pr}}(\mathbf{x}) \mathrm{d}\mathbf{x}, \end{aligned}$$

approximated by the importance sampling estimate

$$\hat{p}_{\mathrm{IS}} = \frac{1}{n_{\mathrm{IS}}} \sum_{i=1}^{n_{\mathrm{IS}}} \mathbb{P}\left(R > r_{\mathrm{Cr}} | \{\mathbf{X} = \mathbf{x}_i^*\}\right) \frac{f_{\mathbf{X}}(\mathbf{x}_i^*)}{g_{\mathrm{Pr}}(\mathbf{x}_i^*)}, \tag{4.2.5}$$

for $\mathbf{x}_1^*, \ldots, \mathbf{x}_{n_{\mathrm{IS}}}^* \sim g_{\mathrm{Pr}}$. The variance of $\hat{p}_{\mathrm{IS}}$ is dependent on the proposal density g_{Pr}, with the optimal choice of proposal minimising the variance in the estimate with respect to a fixed budget of $\mathbf{x}^*$ of size n_{IS}. Here, the choice of g_{Pr} minimising this variance is given by the CDE (4.1.4, Rubinstein and Kroese 2016), so methods typically attempt to find proposal densities approximately equal to the CDE, either by using MCMC (e.g., Xiao et al. 2020) or surrogate modelling of the response function (e.g., Lystad et al. 2023).

We choose to develop methodology to estimate the CDE for subsequent use as the proposal density g_{Pr}, by creating an MCMC scheme with the CDE $\tilde{f}_{\mathbf{X}}(\mathbf{x}; r_{\mathrm{Cr}})$ as the posterior target distribution. From Bayes' rule $\tilde{\pi}(\mathbf{x}|\theta) = \pi(\theta|\mathbf{x}) \times \pi(\mathbf{x})$, where $\pi(\theta|\mathbf{x})$ is an empirical estimate of the probability $\mathbb{P}\left(R > r_{\mathrm{Cr}} | \{\mathbf{X} = \mathbf{x}\}\right)$ obtained by repeated sampling of $R|\{\mathbf{X} = \mathbf{x}\}$, and $\pi(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x})$. Using this approach, we obtain a sample from $\tilde{f}_{\mathbf{X}}(\mathbf{x}; r_{\mathrm{Cr}})$, and adopt a Gaussian smoothed version of $\tilde{\pi}(\mathbf{x}|\theta)$ as the proposal density g_{Pr}, see supplementary material SM3.1.

In simple scenarios, with lower dimensional environment space $\mathcal{E}_{\mathbf{X}}$ and unimodal, approximately elliptically-contoured $\tilde{f}_{\mathbf{X}}(\mathbf{x}; r_{\mathrm{Cr}})$, MCMC samples can be obtained using traditional algorithms such as Metropolis-Hastings (see e.g., Chib and Greenberg 1995). In practice, however, the posterior $\tilde{f}_{\mathbf{X}}(\mathbf{x}; r_{\mathrm{Cr}})$ may be more complex, e.g., exhibiting multi-modality or obvious departures from an elliptically-contoured density. Here we adopt parallel tempering MCMC as a more robust approach to estimate CDEs

of arbitrary complexities.

Parallel tempering MCMC allows jumps between disjoint positive-density regions of the target distribution $\tilde{\pi}$ by combining some $n_{\text{Tm}} > 1$ MCMC chains, each targetting scaled forms of $\tilde{\pi}$. These chains are evaluated at different 'temperatures' $T_1, \ldots, T_{n_{\text{Tm}}} > 0$, with the jth chain, $j = 1, \ldots, n_{\text{Tm}}$, sampling from $\tilde{\pi}^{1/T_j}$; chains with a higher temperature target a 'flatter' form of the target posterior density $\tilde{\pi}$, allowing movement between otherwise disjoint regions of positive density. Individual chains are sampled using a Metropolis-Hastings scheme with proposal density $\mathbf{x}'|\mathbf{x} \sim N(\mathbf{x}, \sigma^2_{\text{MH}})$, $\sigma_{\text{MH}} > 0$, and acceptance probability α_{MH}. Swaps between chains i and j, $(i, j = 1, \ldots, n_{\text{Tm}}\ , i \neq j)$, are periodically proposed with acceptance probability $\alpha_{\text{Sw}}(i, j)$, allowing chains of lower temperature to move between disjoint high-density regions in $\mathcal{E}_{\mathbf{X}}$. Sambridge (2013) shows that, for a parallel tempering algorithm to satisfy detailed balance, individual-chain moves from $\mathbf{x}$ to $\mathbf{x}'$ should be accepted with probability $\alpha_{\text{MH}} = \min\{1, \tilde{\pi}(\mathbf{x}'|\theta)/\tilde{\pi}(\mathbf{x}|\theta)\}$, and swaps between the ith chain at state $\mathbf{x}_i$ and the jth chain at state $\mathbf{x}_j$ should be accepted with probability

$$\alpha_{\text{Sw}}(i, j) = \min\left\{1, \left(\frac{\tilde{\pi}(\mathbf{x}_j|\theta)}{\tilde{\pi}(\mathbf{x}_i|\theta)}\right)^{1/T_i} \left(\frac{\tilde{\pi}(\mathbf{x}_i|\theta)}{\tilde{\pi}(\mathbf{x}_j|\theta)}\right)^{1/T_j}\right\}.$$

Typically swaps are proposed only between adjacent chains, at predetermined set intervals. We use the approach of Vousden et al. (2015), adaptively selecting the temperature ladder $T_1, \ldots, T_{n_{\text{Tm}}}$, as well as the step size standard deviation σ_{MH} for each individual chain. This method is implemented in the Python `pyPESTO` module (Schälte et al., 2023), employed for all MCMC sampling in this work.

A schematic of the resulting sequential design algorithm, henceforth referred to as importance sampling-parallel tempering , is given in Figure 4.2.1.

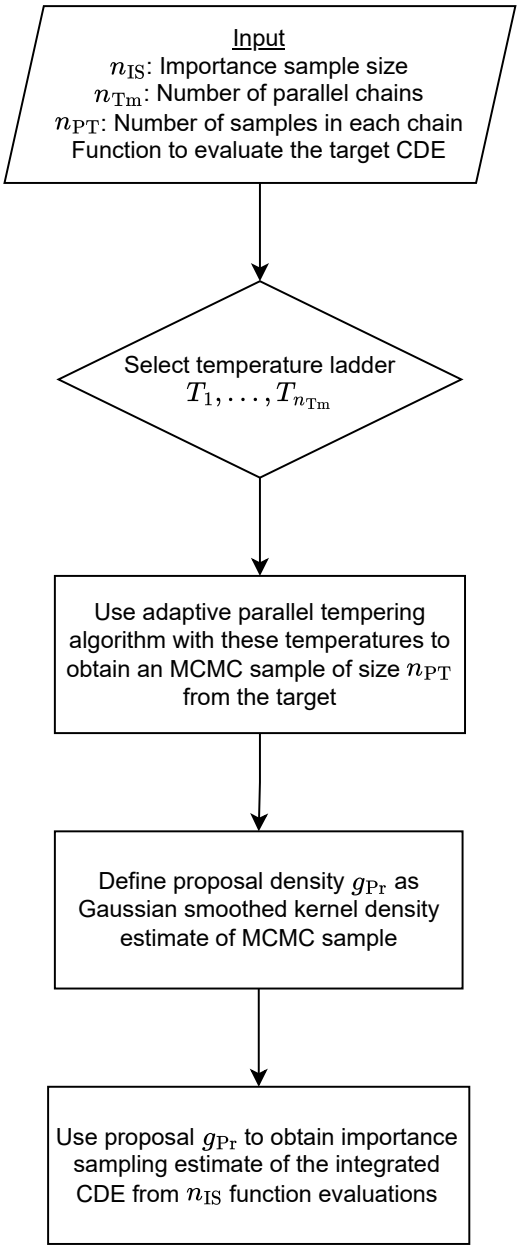


Figure 4.2.1: Schematic for the importance sampling-parallel tempering (IS-PT) sequential design algorithm of Section 4.2.2.

4.2.3 Adaptive Gaussian emulation

Gaussian emulation

Importance sampling reduces the number of evaluations of the expensive response function needed for the calculation of failure probability (4.1.2). It does, however, still require some number of expensive evaluations, with this number being dependent on the convergence rate of the MCMC required for proposal distribution estimation. An alternative approach further reducing the need for computationally expensive evaluations is to replace draws from the true response function with estimates provided by a surrogate model, such as a Gaussian process emulator.

Various approaches to Gaussian process (GP) emulation have been reported in the offshore literature. Gramstad et al. (2020), Castellon et al. (2023) and Lystad et al. (2023) assume a parametric form for the distribution of the structural response $R|\{\mathbf{X} = \mathbf{x}\}$, and so model realisations as draws from this parametric distribution, with unknown parameters modelled as a GP. In our case, we choose to target the (logarithm of the) CDE (4.1.4). To do so, we make repeated draws from $R|\{\mathbf{X} = \mathbf{x}\}$ at each of n_{Tr1} training points $\mathbf{x}_1, \ldots, \mathbf{x}_{n_{\text{Tr1}}}$, obtaining estimates of the conditional failure probability $\mathbb{P}\left(R > r_{\text{Cr}}|\{\mathbf{X} = \mathbf{x}\}\right)$ at each of these values for $\mathbf{X}$. These values form the training set $\mathcal{D} \subset \mathcal{E}_{\mathbf{X}}$, the selection of which is discussed in this Section. After training the GP emulator on $\mathcal{D}$, we can then emulate the CDE at un-observed values $\mathbf{x} \notin \mathcal{D}$.

We define the GP emulator for the log-CDE as

$$\begin{aligned} w(\mathbf{x}) &= \log\left(\tilde{f}_{\mathbf{X}}(\mathbf{x}; r_c)\right) \\ &= \log\{\mathbb{P}\left(R > r_{\text{Cr}}|\{\mathbf{X} = \mathbf{x}\}\right) f_{\mathbf{X}}(\mathbf{x})\} \sim \mathcal{GP}(\mu_{\text{GP}}(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad w : \mathcal{E}_{\mathbf{X}} \mapsto \mathbb{R}, \end{aligned} \tag{4.2.6}$$

for mean and covariance functions

$$\mu_{\text{GP}}(\mathbf{x}) = \mathbb{E}[w(\mathbf{x})], \quad \mu_{\text{GP}} : \mathcal{E}_{\mathbf{X}} \to \mathbb{R},$$
$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}\left[(w(\mathbf{x}) - \mu(\mathbf{x}))\left(w(\mathbf{x}') - \mu(\mathbf{x}')\right)\right], \quad k(\mathbf{x}, \mathbf{x}') : \mathcal{E}_{\mathbf{X}} \times \mathcal{E}_{\mathbf{X}} \to \mathbb{R},$$

where the log transform is used to ensure positivity of the predicted CDE. Compared to direct emulation of the failure probability (see SM1.2) this approach is advantageous when the density $f_{\mathbf{X}}$ is itself not modelled continuously; see for instance the gridded density estimated in Section 4.4.4. Under this model, estimation of (4.1.2) is an application of Bayesian quadrature using a Gaussian process prior (e.g., Hennig et al. 2022).

For the kernel function k, we use the Matérn kernel (e.g., Genton 2001)

$$k_{\text{Mt}}(\mathbf{x}, \mathbf{x}') = \sigma^2_{\text{Mt}} \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{\|\mathbf{x} - \mathbf{x}'\|}{\ell}\right)^{\nu} K_{\nu}\left(\sqrt{2\nu}\frac{\|\mathbf{x} - \mathbf{x}'\|}{\ell}\right),$$

with variance and length scale parameters $\sigma^2_{\text{Mt}}, \ell > 0$, and smoothness parameter $\nu > 0$, where $||\cdot||$ is the Euclidean norm, $\Gamma : \mathbb{R} \mapsto \mathbb{R}$ is the gamma function, and $K_{\nu} : \mathbb{R}^+ \mapsto \mathbb{R}^+$ is the modified Bessel function of the second kind (e.g., Abramowitz and Stegun 1965); this kernel is chosen for its improved ability to capture sudden changes in the target function relative the squared exponential kernel. We combine the Matérn kernel (weighted by a multiplicative constant $C_{\text{Kr}} > 0$) with the additive white noise kernel $k_{\text{WN}}(\mathbf{x}, \mathbf{x}') = \sigma^2_{\text{WN}}$ when $\mathbf{x} = \mathbf{x}'$, and 0 otherwise, for constant white noise variance $\sigma^2_{\text{WN}} > 0$, yielding the full kernel function

$$k(\mathbf{x}, \mathbf{x}') = C_{\text{Kr}} k_{\text{Mt}}(\mathbf{x}, \mathbf{x}') + k_{\text{WN}}(\mathbf{x}, \mathbf{x}').$$

The addition of white noise allows the model to perform well when evaluations of the true target are made with some uncertainty, as is typically the case in application,

see Section 4.4.4. Kernel parameters $\sigma^2_{\text{Mt}}, \ell, C_{\text{Kr}}$ and σ^2_{WN} are jointly estimated via maximum likelihood at each posterior update (see (4.2.7) below) using the L-BFGS-B algorithm (see Pedregosa et al. 2011). The remaining parameter ν is taken to be fixed to avoid the computational expense associated with its optimisation; a brief sensitivity analysis suggests $\nu = 2.5$ as a sensible choice.

We assume a flat prior $\mu_{\text{GP}}(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{E}_{\mathbf{X}}$. Given the covariance function k as defined above, and training data $\mathbf{w} = (w(\mathbf{x}_1), \ldots, w(\mathbf{x}_{n_{\text{Tr1}}}))$, the posterior predictive mean μ^*_{GP} and covariance function k^* for regression (4.2.6) can be found as,

$$\begin{aligned} \mu^*_{\text{GP}}(\mathbf{x}) &= k(\mathcal{D}, \mathbf{x})^T (k(\mathcal{D}, \mathcal{D}) + \alpha_{\text{Ng}} I_{n_{\text{Tr1}}})^{-1} \mathbf{w}, \\ k^*(\mathbf{x}, \mathbf{x}') &= k(\mathbf{x}, \mathbf{x}') - k(\mathcal{D}, \mathbf{x})^T (k(\mathcal{D}, \mathcal{D}) + \alpha_{\text{Ng}} I_{n_{\text{Tr1}}})^{-1} k(\mathcal{D}, \mathbf{x}'), \end{aligned} \tag{4.2.7}$$

with substitution of $\mu_{\text{GP}}(\mathbf{x}) = 0$, where α_{Ng} is an assumed observational nugget variance. In practice, we take $\alpha_{\text{Ng}} = 10^{-5}$. Given this trained GP emulator, the target marginal failure probability estimate $\hat{p}_{\text{GP}}$ can be calculated using

$$\begin{aligned} \hat{p}_{\text{GP}} &= \mathbb{E}_{W,\mathbf{X}}(\{\exp(w(\mathbf{x}))\}) \\ &= \int_{\mathcal{E}_{\mathbf{X}}} \left\{ \int_{\mathbb{R}} \exp(w) \phi(w; \mu^*_{\text{GP}}(\mathbf{x}), k^*(\mathbf{x}, \mathbf{x})) \mathrm{d}w \right\} \mathrm{d}\mathbf{x} \\ &= \int_{\mathcal{E}_{\mathbf{X}}} \exp\left(\mu^*_{\text{GP}}(\mathbf{x}) + \frac{k^*(\mathbf{x}, \mathbf{x}')}{2} \right) \mathrm{d}\mathbf{x}, \end{aligned} \tag{4.2.8}$$

using the expression for the expectation of log-normal random variable $\exp(w)$, for $W|\{\mathbf{X} = \mathbf{x}\} \sim N(\mu^*_{\text{GP}}(\mathbf{x}), k^*(\mathbf{x}, \mathbf{x}))$ with parameters obtained from (4.2.6). For display purposes in the figures of Section 4.3, we evaluate the performance of our GP regression (4.2.6) in terms of the absolute difference Δ_{GP} between the true failure probability (4.1.2) and this estimate

$$\Delta_{\text{GP}} = |p - \hat{p}_{\text{GP}}|.$$

Active learning

The surrogate model (4.2.6) must be trained to provide reliable estimates of the CDE. Often, this training is carried out iteratively, with iteration n training the surrogate against true evaluations of $w(\mathbf{x})$ for all $\mathbf{x}$ in a training set $\mathcal{D}_n$, chosen inductively: at iteration $n+1$, we update training set $\mathcal{D}_n$ to $\mathcal{D}_{n+1} = \{\mathcal{D}_n, \mathbf{x}_{n+1}\}$, where $\mathbf{x}_{n+1} = \text{argmax}_{\mathbf{x}\in\mathcal{E}_\mathbf{X}} U_n(\mathbf{x})$, for acquisition function U_n taking the form

$$U_n(\mathbf{x};\lambda) = \lambda\Sigma_n(\mathbf{x}) + (1-\lambda)\mathrm{M}_n(\mathbf{x}), \tag{4.2.9}$$

for $\lambda \in [0,1]$. Specification of the initial training set $\mathcal{D}_0$ is discussed below. Here $\Sigma_n : \mathcal{E}_\mathbf{X} \mapsto \mathbb{R}$ is an exploration term encouraging sampling at points far from existing members of $\mathcal{D}_n$, and $\mathrm{M}_n : \mathcal{E}_\mathbf{X} \mapsto \mathbb{R}$ is an exploitation term encouraging sampling close to high values of the target function; see Pollatsek and Tversky (1970) for an early discussion of this utility form. In our setting, $\mathrm{M}_n : \mathcal{E}_\mathbf{X} \mapsto \mathbb{R}$ is large at values $\mathbf{x}$ with high contributions to the integral (4.2.8), motivating our first acquisition function

$$U_n^{(1)}(\mathbf{x};\lambda) = \lambda \log k_n^*(\mathbf{x},\mathbf{x}) + (1-\lambda)\log \hat{f}_\mathbf{X}^{(n)}(\mathbf{x}; r_{\mathrm{Cr}}), \tag{4.2.10}$$

where k_n^* is the posterior GP kernel function obtained via (4.2.7) with training set $\mathcal{D}_n$, and $\hat{f}_\mathbf{X}^{(n)}(\mathbf{x}; r_{\mathrm{Cr}})$ is the estimate of CDE $\tilde{f}_\mathbf{X}(\mathbf{x}; r_{\mathrm{Cr}})$ at iteration n. Gramstad et al. (2020), Lystad et al. (2023) and Wang et al. (2024) provide examples of iterative schemes using Gaussian process emulation with acquisition functions similar to (4.2.10), for their respective forms of GP emulator (4.2.6). Following (4.1.2) and (4.1.3), we estimate the CDE $\hat{f}_\mathbf{X}^{(n)}(\mathbf{x}; r_{\mathrm{Cr}})$ as the integrand in (4.2.8), namely

$$\hat{f}_\mathbf{X}^{(n)}(\mathbf{x}; r_{\mathrm{Cr}}) = \exp\left(\mu_n^*(\mathbf{x}) + \frac{k_n^*(\mathbf{x},\mathbf{x}')}{2}\right), \tag{4.2.11}$$

where μ_n^* is the posterior GP mean function obtained via posterior update (4.2.7) with training set $\mathcal{D}_n$.

A similar approach is the active learning Cohn (ALC) scheme of Cohn (1993), aiming to reduce the overall variance of the GP surrogate on $\mathcal{E}_{\mathbf{X}}$. They find the deduced reduction in variance across the *entire space* $\mathcal{E}_{\mathbf{X}}$, given the addition of a new query point $\mathbf{x}$ to the training set $\mathcal{D}_n$ at training iteration n. This is approximated over a reference set $\mathcal{P} = \{\mathbf{x}_j\}_{j=1}^{n_{\mathrm{Rf}}}$ on $\mathcal{E}_{\mathbf{X}}$ as

$$\mathrm{ALC}(\mathbf{x}) = \frac{1}{n_{\mathrm{Rf}}} \sum_{j=1}^{n_{\mathrm{Rf}}} \left(k_n^*(\mathbf{x}_j, \mathbf{x}_j) - \tilde{k}_{n+1}^*(\mathbf{x}_j, \mathbf{x}_j; \mathbf{x}) \right), \quad \mathbf{x} \in \mathcal{P}, \tag{4.2.12}$$

for positive semi-definite $\mathrm{ALC}(\mathbf{x})$, where $\tilde{k}_{n+1}^*(\mathbf{x}_i, \mathbf{x}_i; \mathbf{x})$ is the variance of the GP (4.2.6) at iteration $n+1$, given that query point $\mathbf{x}$ is chosen as the next training point, thereby making $\mathcal{D}_{n+1} = \{\mathcal{D}_n, \mathbf{x}\}$. The summand in (4.2.12) can be written

$$k_n^*(\mathbf{x}_j, \mathbf{x}_j) - \tilde{k}_{n+1}^*(\mathbf{x}_j, \mathbf{x}_j; \mathbf{x}) = \frac{(\mathbf{k}_{n,j}^* {\mathbf{C}_n^*}^{-1} \mathbf{m}_n^* - k_n^*(\mathbf{x}, \mathbf{x}_j))^2}{(k_n^*(\mathbf{x}, \mathbf{x}) - {\mathbf{m}_n^*}^T {\mathbf{C}_n^*}^{-1} \mathbf{m}_n^*)}, \tag{4.2.13}$$

see Seo et al. 2000, where $\mathbf{C}_n^* = k_n^*(\mathcal{D}_n, \mathcal{D}_n)$ is the covariance matrix over current design points, $\mathbf{k}_{n,j}^* = k_n^*(\mathcal{D}_n, \mathbf{x}_j)$ is the vector of covariances between the training data and reference point $\mathbf{x}_j$ and $\mathbf{m}_n^* = k_n^*(\mathcal{D}_n, \mathbf{x})$ is the covariance vector between the training data and the query point $\mathbf{x}$.

Seo et al. (2000) recommend selecting the best next query point $\mathbf{x}$ by maximising a *weighted* sum of (4.2.13) over the reference grid $\mathcal{P}$. Instead, we employ an acquisition function of the form (4.2.9) utilising the ALC criterion. We find that the acquisition function

$$U_n^{(2)}(\mathbf{x}; \lambda) = \lambda \log \mathrm{ALC}(\mathbf{x}) + (1 - \lambda) \log \hat{f}_{\mathbf{X}}^{(n)}(\mathbf{x}; r_{\mathrm{Cr}}), \tag{4.2.14}$$

performs well for careful choice of λ. This is similar to the acquisition function (4.2.10), except that in (4.2.14) the exploration term $\Sigma_n(\mathbf{x}) = \log \mathrm{ALC}(\mathbf{x})$ considers the effect

of including a query point $\mathbf{x}$ in reducing error on the whole candidate space $\mathcal{E}_{\mathbf{X}}$, rather than just at the query point itself.

A schematic of the adaptive Gaussian emulation (AGE) procedure for sequential design is given in Figure 4.2.2.

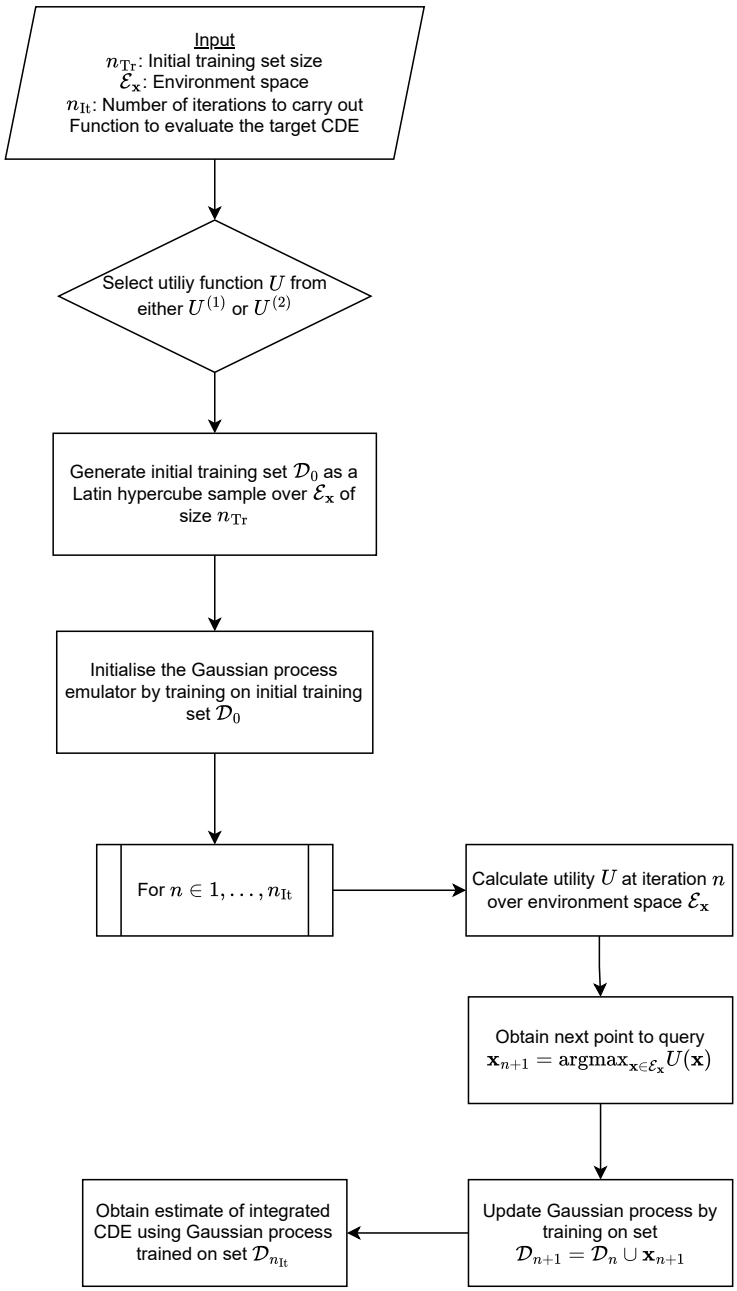


Figure 4.2.2: Schematic for the adaptive Gaussian emulation (AGE) sequential design algorithm of Section 4.2.3.

4.3 Synthetic simulation study

4.3.1 Synthetic scenario design

We now compare the methods introduced in Section 4.2 under a synthetic test scenario, designed to be simple enough to yield a valuable comparison, whilst being sufficiently methodologically challenging, and representative of a real-world structure. We construct a synthetic response with an artificially bimodal CDE, intended to represent the extreme case of non-convex failure regions discussed in Section 4.1. We follow the approaches of the likes of Gramstad et al. (2020) and Castellon et al. (2023), who model structural responses (approximately) by some parametric distribution function. The structural response R given long term environment $\mathbf{X}$ is modelled as a Weibull random variable, with distribution function

$$F_{R|\mathbf{X}}(r|\mathbf{x}) = 1 - \exp\left\{-\left(\frac{r}{\eta(\mathbf{x})}\right)^k\right\}, \quad r > 0,$$

for fixed shape parameter $k = 2$ and scale parameter $\eta : \mathcal{E}_{\mathbf{X}} \mapsto \mathbb{R}$ dependent on the long term environment. Adoption of this conditional Weibull form allows straightforward sampling from $R|\{\mathbf{X} = \mathbf{x}\}$, as well as exact evaluation of the conditional failure probability $\mathbb{P}\left(R > r_{\mathrm{Cr}}|\{\mathbf{X} = \mathbf{x}\}\right)$. The environment $\mathbf{X}$ is assumed bivariate $\mathbf{X} = (X_1, X_2)$, with density function $f_{\mathbf{X}}(\mathbf{x}) = f_{(X_1,X_2)}(x_1, x_2) = f_{X_2|X_1}(x_2|x_1)f_{X_1}(x_1)$, where

$$f_{X_1}(x_1) = \frac{x_1}{\sigma_R^2}\exp\left(-\frac{x_1^2}{2\sigma_R^2}\right), \quad x_1 \geq 0 \tag{4.3.15}$$

is the Rayleigh density with scale parameter $\sigma_R > 0$ set to $\sigma_R = 12$ in this case, and

$$f_{X_2|X_1}(x_2|x_1) = \frac{1}{x_2\sigma_{\mathrm{LN}}(x_1)\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{\log(x_2) - \mu_{\mathrm{LN}}(x_1)}{\sigma_{\mathrm{LN}}(x_1)}\right)^2\right\}$$

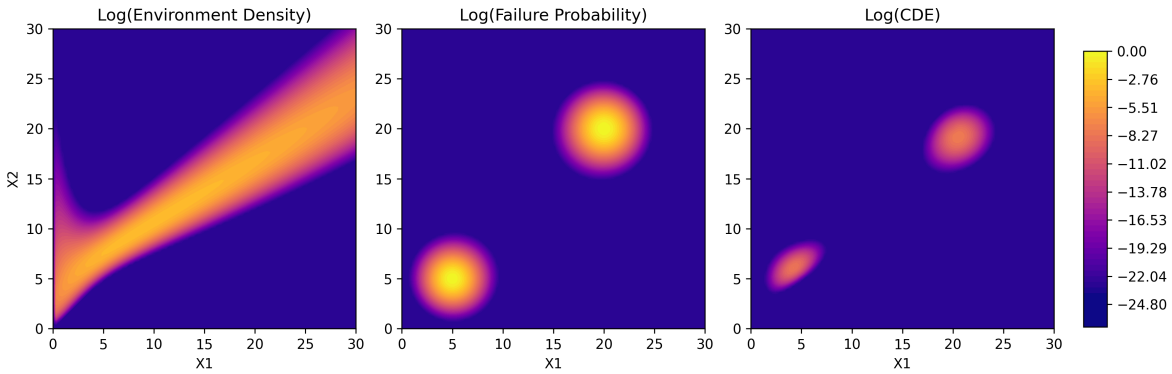


Figure 4.3.3: Panels summarising the synthetic response case study used in this section. From left to right the panels show: bivariate environment log-density (4.3.15); structural log-failure probability as Weibull exceedance probability of $r_{\mathrm{Cr}} = 175$ (4.3.16); and log-CDE (4.1.4) obtained by multiplying failure probability by environment density.

is the log-normal density with

$$\mu_{\mathrm{LN}}(x_1) = 0.933 + 0.578 x_1^{0.395},$$
$$\sigma_{\mathrm{LN}}(x_1) = 0.055 + 0.336 + \exp(-0.585 x_1).$$

This is a typical form for the joint distribution of wave period and significant wave height, see e.g., Mathisen and Bitner-Gregersen (1990). for wave period given significant wave height. The reader might therefore choose to consider this environmental specification as an example of an extreme environment of significant wave height (Rayleigh) and conditional significant wave period (log-normal), although this interpretation is not necessary.

The function η is constructed to provide a structural response with the desired multimodal behaviour. To achieve this, we define a scenario with scale $\eta(\mathbf{x})$ increasing around values $\mathbf{x}_{\mathrm{Pk1}}$ and $\mathbf{x}_{\mathrm{Pk2}}$, modelling the scale parameter using the multimodal function

$$\eta(\mathbf{x}) = C\left\{A \max(||\mathbf{x} - \mathbf{x}_{\mathrm{Pk1}}||, \nu) \; + B \max(||\mathbf{x} - \mathbf{x}_{\mathrm{Pk2}}||, \nu)\right\},$$

for scaling parameters $A, B, C > 0$, and peak radius $\nu > 0$ surrounding 'resonant' $\mathbf{X}$

values $\mathbf{x}_{\text{Pk1}} = (5, 5)$ and $\mathbf{x}_{\text{Pk2}} = (20, 20)$. We set constants, $A = 1.3$, $B = 1.5$, $C = 100$, $\nu = 0.5$, and critical response $r_{\text{Cr}} = 175$, chosen in order to yield a true 'synthetic' failure probability

$$p_{\text{Sn}} = \mathbb{P}\left(R > r_{\text{Cr}}\right) = \int_{\mathcal{E}_{\mathbf{X}}} \exp\left\{-\left(\frac{r_{\text{Cr}}}{\eta(\mathbf{x})}\right)^2\right\} f_{\mathbf{X}}(\mathbf{x})\mathrm{d}\mathbf{x} = 1.3 \times 10^{-3}. \tag{4.3.16}$$

This yields a failure probability in order of magnitude comparable to the failure probabilities discussed in Section 4.4.4. Figure 4.3.3 shows the environment density, failure probability and CDE for this synthetic scenario, over the bivariate environment space $\mathcal{E}_{\mathbf{X}} = [0, 30]^2$.

In practice, estimates of conditional failure probability $\mathbb{P}\left(R > r_{\text{Cr}} | \{\mathbf{X} = \mathbf{x}\}\right)$ are found empirically using realisations of $R|\{\mathbf{X} = \mathbf{x}\}$. We introduce further uncertainty in this synthetic case in the form of the conditional distribution for $R|\{\mathbf{X} = \mathbf{x}\}$ by making η stochastic, with

$$\eta_\delta(\mathbf{x}) = \eta(\mathbf{x}) \cdot (1 + \epsilon_\delta), \quad \text{for} \quad \epsilon_\delta \sim N(0, \delta^2), \tag{4.3.17}$$

where we set $\delta = 0.05$. This applies an additive white noise to the scale of our observations with variance proportional to the value of the scale function, meaning that larger values of the scale function will correspond to 'more uncertain' observations. In the absence of uncertainty in η_δ (i.e., with $\delta = 0$), the expected distribution of $R|\{\mathbf{X} = \mathbf{x}\}$ is relatively easily identified from a smaller number of realisations of fluid loading simulation. However, for uncertain η_δ, the number of realisations required to be confident about the expected distribution of $R|\{\mathbf{X} = \mathbf{x}\}$ increases. That is, particularly with $\delta > 0$, we expect to need to sample from the same regions of $\mathcal{E}_{\mathbf{X}}$ multiple times to build confidence in our estimate of the CDE.

Note that the IS-PT and AGE procedures are expected to give good performance for any combination of values for parameters σ_R, μ_{LN}, σ_{LN}, A, B, C, ν, x_{Pk1} and x_{Pk2}

which would yield an observably bivariate response surface in the environment space (as in Figure 4.3.3), since the parallel tempering MCMC algorithm is able to adapt to the size and location of each mode. Therefore, the actual values of parameters used here are of little direct relevance. The critical feature of the current simulation study is that the general characteristics of the distribution of environmental variables and those of the structural response to the environment, reflect the general characteristics of actual environments and structural responses. It is for this reason that we refer to the environmental variables in this section as X_1 and X_2 rather than e.g., H_s and T_p.

4.3.2 Results of synthetic study

Overview

Here we apply the methods introduced in Section 4.2 to the synthetic scenario discussed above. We first present the results of the importance sampling-parallel tempering (IS-PT) approach of Section 4.2.2 in Section 4.3.2, followed by those of the adaptive Gaussian emulation (AGE) procedure of Section 4.2.3 in Section 4.3.2. We adjust the number of expensive function evaluations n_{Ev} used for each of the IS-PT and AGE methods so they yield the same order of magnitude of root mean squared error

$$\mathrm{RMSE}(\hat{p}) = \sqrt{\sum_{r=1}^{n_{\mathrm{Rp}}} \frac{(\hat{p}_r - p_{\mathrm{Sn}})^2}{n_{\mathrm{Rp}}}}, \tag{4.3.18}$$

over some number n_{Rp} of replicate analyses, where $\hat{p}_r$ is the estimate provided by either IS-PT or AGE at replicate r. We also evaluate the bias

$$\mathrm{Bias}(\hat{p}) = \sum_{r=1}^{n_{\mathrm{Rp}}} \frac{(\hat{p}_r - p_{\mathrm{Sn}})}{n_{\mathrm{Rp}}},$$

for each of the methods.

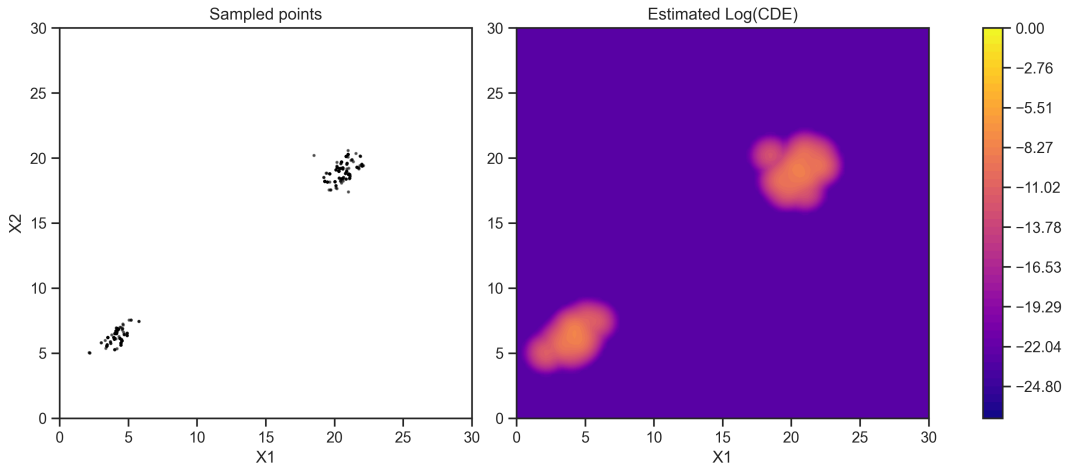


Figure 4.3.4: Example sample from CDE (4.1.4) under the synthetic structural scenario obtained using the adaptive parallel tempering MCMC algorithm of Vousden et al. (2015) (left), and corresponding smoothed log-CDE estimated using Gaussian kernel bandwidth selected according to Scott (2015) (right).

Importance sampling coupled with parallel tempering MCMC (IS-PT)

We apply the IS-PT framework of Section 4.2.2 under the synthetic scenario in three stages: first (a) parallel tempering MCMC sampling with the CDE (4.1.4) as target posterior density; followed by (b) kernel smoothing of the resulting sample to obtain proposal density p_{Pr}; and finally (c) evaluation of importance sampling estimate (4.2.5) using $n_{\mathrm{IS}} = 100$ draws from this proposal. Steps (a)-(c) are repeated $n_{\mathrm{Rp}} = 100$ times, to estimate $\hat{p}_{\mathrm{IS}}$.

Step (a) is achieved using the adaptive parallel tempering algorithm of Vousden et al. (2015) implemented in the `pyPESTO` module. We run $n_{\mathrm{Tm}} = 5$ parallel chains, supplying an initial temperature ladder $T_1, \ldots, T_5$ geometrically spaced between $T_1 = 1$ and $T_5 = 20$, with initial proposal variance $\sigma^2_{\mathrm{MH}} = 1$. The MCMC algorithm then adaptively tunes the temperature spacing and proposal variance, targetting equal acceptance probability of swaps between adjacent chains. Each of the five chains is run for $n_{\mathrm{PT}} = 400$ time steps, with periodic swaps between chains proposed according to Vousden et al. (2015), requiring $n_{\mathrm{Tm}} \times n_{\mathrm{PT}} = 2000$ expensive function evaluations in total. An example trace plot from the T_1 chain is given in SM3.1. The chain at temperature $T_1 = 1$

is retained, and burn-in length n_{Br} automatically chosen using Geweke's diagnostic (Geweke, 1991). When $n_{\mathrm{Br}} < n_{\mathrm{PT}}$, this burn-in period is discarded, leaving a sample of length $n_{\mathrm{PT}} - n_{\mathrm{Br}}$. For step (b), the sample is then used to provide a Gaussian kernel smoothed estimate of the CDE, with kernel bandwidth chosen according to Scott's rule of thumb (Scott, 2015), see SM3.1 for details. Step (c) consists of evaluating importance sampling probability estimate $\hat{p}_{\mathrm{IS}}$ given by (4.2.5), using $n_{\mathrm{IS}} = 100$ draws from proposal density g_{Pr} found in step (b), requiring a further 100 expensive function evaluations. Figure 4.3.4 shows a typical sample obtained using this approach together with resulting CDE estimate g_{Pr}. The (root mean square error) RMSE (4.3.18) is estimated to be $\mathrm{RMSE}(\hat{p}_{\mathrm{IS}}) = 2.20 \times 10^{-4}$, using $n_{\mathrm{Rp}} = 100$ replicates of the IS-PT analysis, with *each* of the n_{Rp} IS-PT estimates requiring $n_{\mathrm{Ev}} = n_{\mathrm{Tm}} \times n_{\mathrm{PT}} + n_{\mathrm{IS}} = 2100$ expensive function evaluations. The bias in the $\hat{p}_{\mathrm{IS}}$ estimate over the 100 replicates is small, equal to $\mathrm{Bias}(\hat{p}_{\mathrm{IS}}) = 5.32 \times 10^{-5}$.

Adaptive Gaussian emulation (AGE)

The GP emulator (4.2.6) is used to model the log-CDE under this synthetic scenario, following the AGE procedure of Section 4.2.3. It is iteratively trained as in (4.2.7) on training sets $\mathcal{D}_1, \ldots, \mathcal{D}_{n_{\mathrm{It1}}}$ for n_{It1} iterations, with training set $\mathcal{D}_{n+1} = \{\mathcal{D}_n, \mathbf{x}^*\}$, $n > 1$, constructed with $\mathbf{x}^*$ chosen according to either $U^{(1)}$ (4.2.10, Variance case) or $U^{(2)}$ (4.2.14, ALC case). In each case, the initial training set $\mathcal{D}_0$ is a simple space-filling Latin hypercube design of $n_{\mathrm{Tr1}} = 144$ points, chosen as a low, but adequate, number of starting points found to provide stable kernel parameter convergence at iteration zero. At each subsequent iteration, we begin the kernel parameter optimisation at the previous iteration's estimates.

Figure 4.3.5 shows an example of how utility $U^{(1)}$ is constructed using the emulator (4.2.6) trained on initial set $\mathcal{D}_0$, for two example values of λ. The upper panels shows the exploration Σ_0 and exploitation M_0 terms as defined in (4.2.10), with

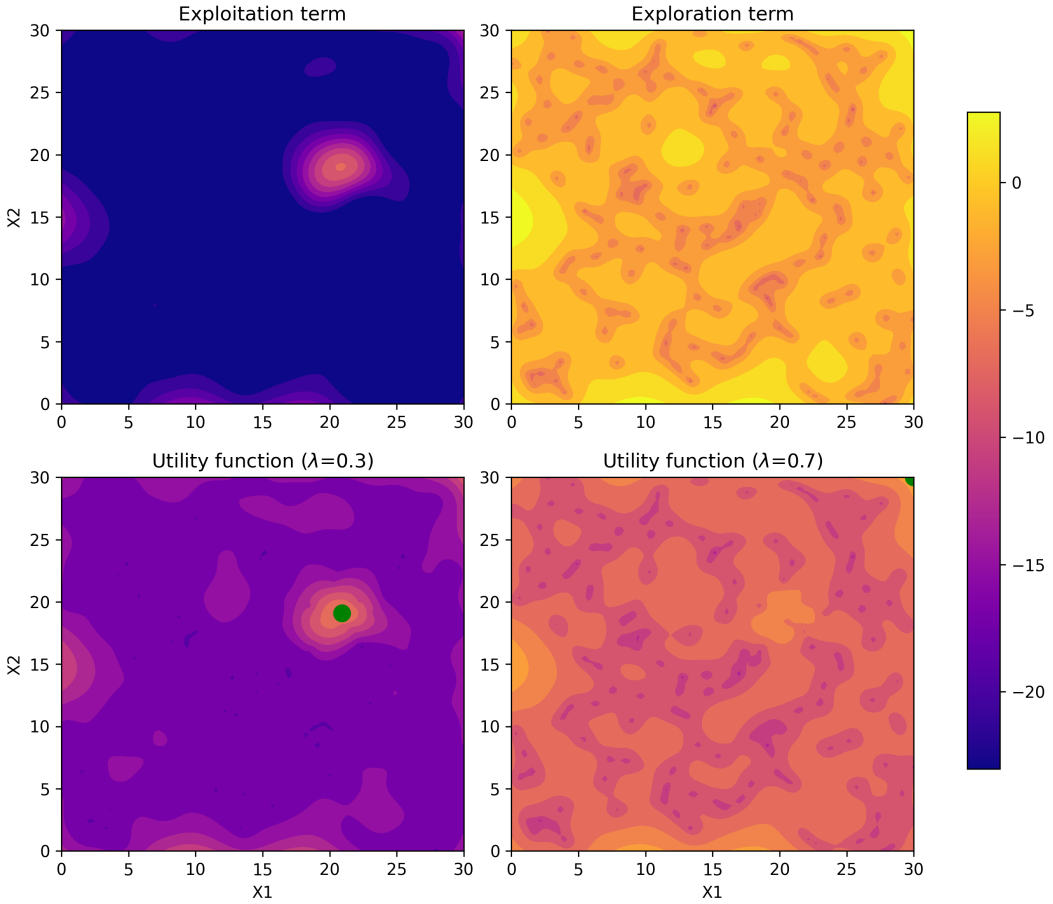


Figure 4.3.5: Behaviour of utility function $U^{(1)}(\mathbf{x};\lambda)$ over the environment space $\mathcal{E}_{\mathbf{X}}$, for synthetic scenario. Upper panels show exploitation and exploration terms obtained from GP emulator (4.2.6) trained on initial Latin hypercube set $\mathcal{D}_0$ of size $n_{\text{Tr1}} = 144$. Lower panels show resulting utility functions for weights $\lambda = 0.3$ and $\lambda = 0.7$. In each lower panel, the optimal sampling point $\mathbf{x}^* = \text{argmax}_{\mathbf{x}\in\mathcal{E}_{\mathbf{X}}} U^{(1)}(\mathbf{x};\lambda)$ is indicated in green. In the lower right hand panel, $\mathbf{x}^*$ is located in the upper right corner of $\mathcal{E}_{\mathbf{X}}$.

lower panels showing utility functions obtained by prioritising exploration ($\lambda = 0.7$) or exploitation ($\lambda = 0.3$). Green points in the lower panels indicate the maximum $\mathbf{x}^* = \text{argmax}_{\mathbf{x}\in\mathcal{E}_{\mathbf{X}}} U^{(1)}(\mathbf{x};\lambda)$, illustrating that the choice of this tuning parameter can alter the design of the training set $\mathcal{D}_1 = \{\mathcal{D}_0, \mathbf{x}^*\}$ (and thus subsequent training sets $\mathcal{D}_2, \mathcal{D}_3, \ldots$). In the lower right panel, the maximum is located on the edge of the environment space, due to the Latin hypercube sampling used to construct $\mathcal{D}_0$ placing no points on the boundary. (This can be prevented by adding initial training points along the boundary, however, this isn't necessary as subsequent iterations move away from

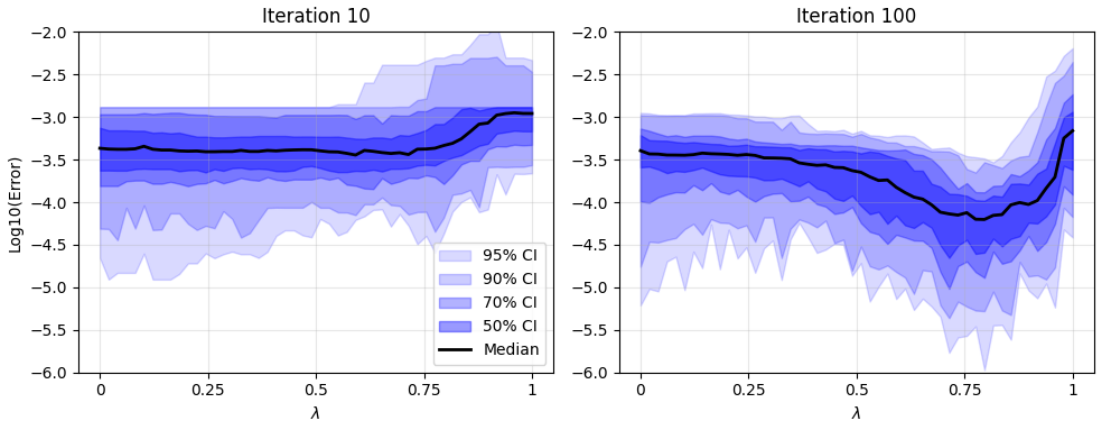


Figure 4.3.6: Log-scale absolute error Δ_{GP} of the GP probability estimate $\hat{p}_{\text{GP}}$ at specified iterations, for emulator (4.2.6) trained using $U^{(1)}$ over the range of weight $\lambda \in [0.01, 0.99]$ for the synthetic scenario. At iteration 100, the weight that minimises median error is $\lambda^* = 0.80$.

the edge of the space once it has been explored.) See SM3.2 for an equivalent example for $U^{(2)}$.

The GP emulator is trained using both utility functions $U^{(1)}$ and $U^{(2)}$, for a range of $n_\lambda = 50$ values of weight parameter λ, equally spaced on the interval $[0.01, 0.99]$. For each value of λ, we perform $n_{\text{It1}} = 100$ iterations of the GP update (4.2.7) for each utility. This yields posterior estimates μ_n^* and k_n^* for $n = 1, \ldots, n_{\text{It1}}$. This analysis is replicated $n_{\text{Rp}} = 100$ times, with randomised initial set $\mathcal{D}_0$ and conditional response scale η (4.3.17) at each replicate.

Each of the n_{Rp} replicate analyses produces $n_\lambda \times n_{\text{It1}}$ values of the failure probability estimate $\hat{p}_{\text{GP}}$ and error Δ_{GP} for each utility. Figure 4.3.6 shows the distribution of the resulting Δ_{GP} values under variance utility $U^{(1)}$ (4.2.10) with respect to λ, at iterations 10 and 100. Errors are plotted on the log scale, with 50%, 70%, 90% and 95% confidence bands indicated in different shades of blue. The median log-error trend with respect to λ is given as a black line. The GP emulator converges to the truth for weights in the interval I^*, which in this case corresponds approximately to $[0.5, 1]$. The location of I^* on the unit interval is determined by the bimodal nature of the synthetic response. For some initial training sets $\mathcal{D}_0$, at iteration zero, the emulator detects one peak in response

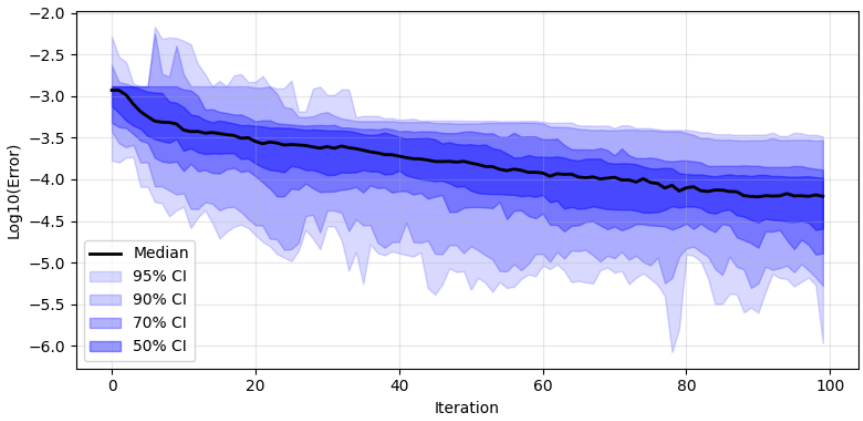


Figure 4.3.7: Distribution of log-scale absolute error Δ_{GP} in the GP probability estimate with respect to iteration, trained using $U^{(1)}$ with $\lambda = \lambda^*$. The trend in median error is indicated in black, with various confidence intervals shown in blue.

but fails to detect the other; this can be seen in the top left panel of Figure 4.3.5, where the mode at $\mathbf{x}_{\text{Pk2}} = (20, 20)$ is found, but that at $\mathbf{x}_{\text{Pk1}} = (5, 5)$ is not. For low values of λ, the utility function $U^{(1)}$ sometimes does not place enough weight on the exploration term for the algorithm to detect the second peak in subsequent iterations (e.g., the lower left panel of Figure 4.3.5 shows a low value for utility at $\mathbf{x}_{\text{Pk1}}$, whereas the lower right panel has a higher utility there). That is, for low values of λ, the iterative algorithm tends not to allow the GP to 'discover' the second mode. The value of λ minimising the median error at the final iteration is $\lambda^* = 0.80$. Figure 4.3.7 shows the distribution of Δ_{GP} across all iterations when $\lambda = \lambda^*$. In general, there is a decrease in error with iteration, with a 'spike' at around iteration 10 for some replicates; these spikes shows where the algorithm tends to detect the second mode, causing a temporary increase in bias due to the uncertainty in (4.3.17). Figures corresponding to Figures 4.3.6 and 4.3.7 for ALC utility $U^{(2)}$ can be found in SM3.2. For $U^{(2)}$, a minimum of Δ_{GP} is found in $I^* = [0.2, 0.5]$, and comparison of errors Δ_{GP} at the final iteration indicates that Δ_{GP} for $U^{(2)}$ is somewhat larger than for $U^{(1)}$.

Figure 4.3.8 shows an example GP emulator at the final iteration, trained on set $\mathcal{D}_{100}$ selected using $U^{(1)}$ with $\lambda^* = 0.80$. The left panel shows the posterior GP mean $\mu^*_{100}(\mathbf{x})$

and the right the posterior GP standard deviation $k^*_{100}(\mathbf{x},\mathbf{x})^{1/2}$, both over $\mathbf{x} \in \mathcal{E}_{\mathbf{X}}$. The initial Latin hypercube training set $\mathcal{D}_0$ is shown as dark green crosses, and the iteratively selected new training points $\mathcal{D}_{100} \setminus \mathcal{D}_0$ are shown as light green crosses. The light green crosses, iteratively selected using the utility function, mostly cluster around the high-density regions of the synthetic CDE, whilst allowing some exploration into low-density regions of $\mathcal{E}_{\mathbf{X}}$.

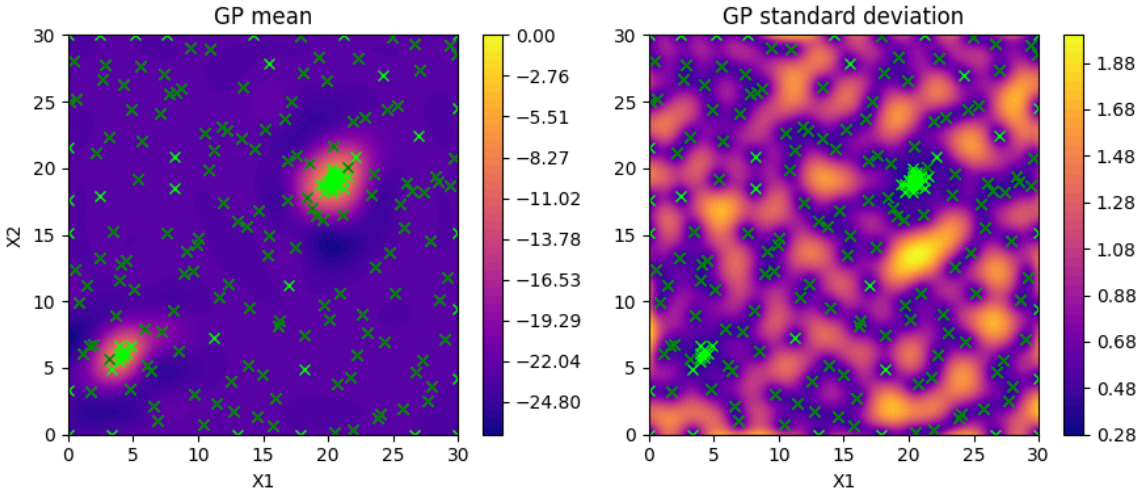


Figure 4.3.8: GP emulator at iteration 100 for variance utility $U^{(1)}$ (4.2.10), trained using the optimal value $\lambda^* = 0.80$ minimising median of error Δ_{GP}. The panels from left to right show: the posterior GP mean $\mu^*_{100}(\mathbf{x})$ over $\mathbf{x} \in \mathcal{E}_{\mathbf{X}}$; the posterior GP standard deviation $k^*_{100}(\mathbf{x},\mathbf{x})^{1/2}$. The initial random Latin hypercube training set $\mathcal{D}_0$ is shown as dark green crosses, and the iteratively selected new training points $\mathcal{D}_{100} \setminus \mathcal{D}_0$ are shown as light green crosses.

We evaluate the RMSE (4.3.18) of the AGE approach using $\hat{p}_{\text{GP}}$ obtained from emulators trained under $U^{(1)}$ with $\lambda = \lambda^*$ over $n_{\text{Rp}} = 100$ replicate analysis. The resulting estimates yield $\text{RMSE}(\hat{p}_{\text{GP}}) = 1.16 \times 10^{-4}$, a similar value to $\text{RMSE}(\hat{p}_{\text{IS}})$ reported in Section 4.3.2. For the AGE approach, each replicate analysis involves a total of $n_{\text{Ev}} = |\mathcal{D}_0| + n_{\text{Itr1}} = 244$ expensive function evaluations, assuming λ^* is known. The bias in the $\hat{p}_{\text{GP}}$ estimate over the 100 replicates is $\text{Bias}(\hat{p}_{\text{GP}}) = 3.18 \times 10^{-5}$, comparable in size to that of $\hat{p}_{\text{IS}}$. Corresponding results using $U^{(2)}$ are similar, and summarised in the next section.

Comparison of IS-PT and AGE results

Figure 4.3.9 shows the distribution of the IS-PT estimate $\hat{p}_{\text{IS}}$ and the AGE estimates $\hat{p}_{\text{GP}}$ (from $U^{(1)}$ and $U^{(2)}$ at iteration 100, $\lambda = \lambda^*$) around the target failure probability p_{Sn}. A summary of the RMSEs and biases for these estimates can be seen in Table 4.3.1, along with the number of expensive function evaluations n_{Ev} required for each replicate analysis. Both variants of the $\hat{p}_{\text{GP}}$ estimate show an equivalent performance to $\hat{p}_{\text{IS}}$ for around 12% of required expensive function evaluations, provided we have knowledge of the optimal weight parameter λ^*. The AGE approach with utility $U^{(2)}$ is computationally somewhat more demanding than that using $U^{(1)}$, due to the required calculation of ALC (4.2.12) at each iteration.

However, if λ^* is unknown, and cannot be reliability estimated, we see that IS-PT provides a useful if computationally more demanding alternative. The current analysis shows that approximately 2000 expensive function iterations using IS-PT are sufficient to estimate a bimodal CDE well in two dimensions, avoiding the need to specify problematic hyperparameters such as λ. To provide some context, we also considered estimating failure probability using a simple standard Monte Carlo sampling algorithm, with a total number of 2100 Monte Carlo samples to match that used for IS-PT and AGE variants in Table 1. We found corresponding RMSE and bias values for failure probability of 3.61×10^{-4} and 7.29×10^{-5} respectively. Clearly in terms of RMSE, standard Monte Carlo sampling provides considerably poorer estimates of failure probability than both PI-IS and the two AGE variants considered here. We explore the relative merits of IS-PT and AGE methodologies further for the monopile structure scenario of Section 4.4.

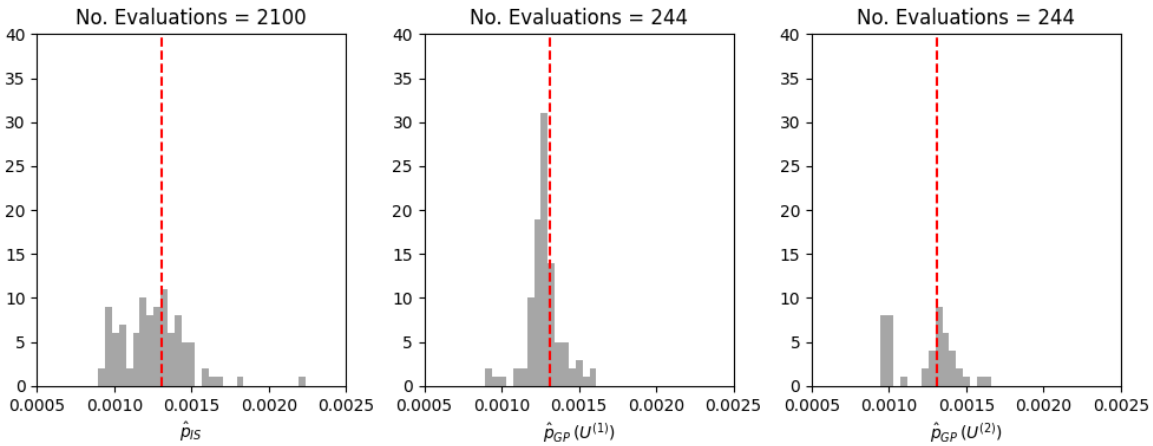


Figure 4.3.9: Distribution of $n_{\text{Rp}} = 100$ estimates $\hat{p}_{\text{IS}}$ (left), $\hat{p}_{\text{GP}}$ for $U^{(1)}$ iteration 100 with $\lambda = \lambda^*$ (centre) and $\hat{p}_{\text{GP}}$ for $U^{(2)}$ iteration 100 with $\lambda = \lambda^*$ (right), for true failure probability p_{Sn} (red). The number of function evaluations required for a single replicate analysis is indicated in the panel titles.

	IS-PT	$U^{(1)}$ AGE	$U^{(2)}$ AGE
RMSE	2.20×10^{-4}	1.16×10^{-4}	2.40×10^{-4}
Bias	5.32×10^{-5}	3.18×10^{-5}	1.50×10^{-4}
Number of function evaluations, n_{Ev}	$n_{\text{Tm}} \times n_{\text{PT}} + n_{\text{IS}} = 2100$	$\lvert\mathcal{D}_0\rvert + n_{\text{Itr1}} = 244$	$\lvert\mathcal{D}_0\rvert + n_{\text{Itr1}} = 244$

Table 4.3.1: RMSEs and biases of $\hat{p}_{\text{IS}}$ and $\hat{p}_{\text{GP}}$ when targetting failure probability p_{Sn}, calculated for $n_{\text{Rp}} = 100$ replicate analyses. The true value of probability of failure is $p_{\text{Sn}} = 1.3 \times 10^{-3}$. Also shown is the number of expensive response function evaluations required for a single replicate analysis for each of IS-PT and AGE.

4.4 Application to monopile response models

4.4.1 Overview of case study

We now apply the IS-PT and AGE methodologies of Section 4.2 to a real-world case study, using hindcast data from a location around 1km offshore of Albany, Western Australia, produced by the Centre for Australian Weather and Climate Research, see Section 4.4.2 for details. We consider a model monopile structure situated in this environment, subject to wave-induced loading, which in turn induces some resonant effect. To construct the test scenario, we first use the extreme value methods of Davison and Smith (1990) and Heffernan and Tawn (2004) to model the joint behaviour of a bivariate ocean $\mathbf{X}$ at this location, see Section 4.4.3. This is followed in Section 4.4.4

with numeric simulation from the T-FNV model of Taylor et al. (2024) to approximate the inertial load placed on an offshore wind turbine at this location. Finally, this load is propagated through the linear response function for a damped harmonic oscillator, yielding realisations of the harmonic response on our model structure. The results of these simulations are given in Section 4.4.5 to provide a 'baseline' estimate of the CDE. This baseline is then used to assess the performance of IS-PT and AGE methodologies in Sections 4.4.6 and 4.4.7. These methods are then compared in Section 4.4.8.

The reader is referred to articles including Riise et al. (2018), He and Zhu (2019), Wan et al. (2023), Liu et al. (2025) and Yang et al. (2025), as well as Orszaghova et al. (2025), for relevant discussion and illustration of wind turbine response characteristics.

4.4.2 Albany hindcast data

The data includes hourly hindcast observations over the period 1980-2017, consisting of sea state variables significant wave height H_s, peak wave period T_p, energy wave period T_e, and mean wave period T_m. There are a total of 333120 observations. We preprocess the data by isolating storm peak values of the sea state H_s. Given storm events that are sufficiently well spaced in time, this removes any temporal correlation in the storm peak data, simplifying the modelling process whilst retaining the observations most likely to induce structural failure.

To isolate the storms peaks, we follow the procedure of Ewans and Jonathan (2008). Firstly, a wave height h_{St} (in metres) is chosen as the storm threshold, such that an upcrossing above this height is considered the beginning of a storm event. The subsequent downcrossing of this height is considered the end of the storm event. We also merge any two storm events that occur within 48 hours of one another, retaining only the largest storm peak value. The value of h_{St} is determined by assessing the number of storm peaks recovered from the dataset using a given threshold against the realism of observed storm lengths; this creates a trade-off between retaining enough points for

statistical modelling, whilst avoiding identifying storms of unrealistically long duration. We choose to limit occurrences of storms lasting longer than three days, selecting $h_{\mathrm{St}} = 4$ yielding a total of 976 storm peak observations, with around 6% of identified storm durations exceeding three days.

Figure 4.4.10 illustrates this process, with the left panel showing identified storms. Given selection of storm peak H_s values, these can be matched with the corresponding T_p, T_e or T_m values to obtain a joint storm peak environment. We focus our attention on storm peak significant wave height H_s and (significant) wave steepness

$$S_e = \frac{2\pi H_s}{g T_e^2},$$

for gravitational acceleration $g = 9.81\mathrm{ms}^{-2}$, modelling the 2-dimensional environment (H_s, S_e). We choose to model S_e over T_e (or other wave period variables) because the most extreme sea states tend to be the steepest. Using S_e, our interest therefore lies in characterising the pair of positively valued variables (H_s, S_e), when at least one of the pair is very large, an appropriate setting for application of the conditional extremes method of Heffernan and Tawn (2004). Going forward, we let $\mathbf{X} = (H_s, S_e)$, referring to the joint storm peak values seen in the right panel of Figure 4.4.10, rather than the original hourly data.

4.4.3 Joint storm peak variable modelling

Outline of long term environment model

We describe the joint behaviour a long term environment, using the conditional extreme value model of Heffernan and Tawn (2004). This approach facilitates modelling of the joint extremes of $\mathbf{X}$, facilitating the extrapolation of joint behaviour beyond the range of the sample data. This asymptotically justified framework has been widely applied in capturing the tail dependence of environmental data due to its flexibility in capturing

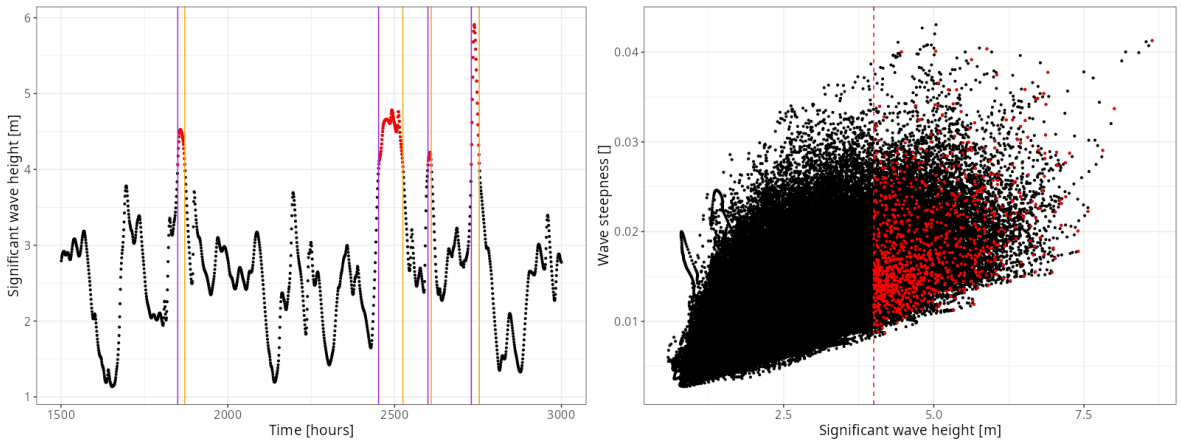


Figure 4.4.10: Illustration of storm peak isolation for storm threshold $h_{\text{St}} = 4$. The left panel shows H_s value against hourly index, with the beginning of each storm (defined as the first upcrossing of 4m) indicated in purple. The end of each storm (defined as the first downcrossing of 4m) is shown in orange. Sequences of within-storm H_s values are highlighted in red. In the right panel, the entire hindcast data of S_e against H_s are shown in black, with chosen storm peak values indicated in red. The storm threshold $h_{\text{St}} = 4$ is indicated as a dashed red vertical line.

different extremal dependence types and its ease of use (e.g., Jonathan et al. 2014b; Towe et al. 2019; Shooter et al. 2021b; Tendijck et al. 2023a). It is also simple to extend this model to account for seasonality or long term trends in an environment through the addition of covariates (e.g., Ewans and Jonathan 2008); however, we omit the inclusion of covariate effects as this is not the focus of this work. Furthermore, whilst our synthetic environment and example data are of dimension $d = 2$, we present the conditional extremes method in the general case of $d > 1$.

The conditional extremes method consists of a two stage modelling process: first, the transformation of the environment variable $\mathbf{X}$ to a standard marginal scale (typically Laplace); and second, modelling of the joint structure of the standardised variables. The first step is achieved using univariate extreme value techniques (e.g., Davison and Smith 1990), and the second via a series of d pairwise non-linear regressions. Details of these steps are discussed below.

Marginal modelling and transformation

We use the peaks over threshold method of Davison and Smith (1990) when modelling the marginal distributions of $X_1, \ldots, X_d$. That is, for X_j, $j = 1, \ldots, d$, we fit a generalised Pareto distribution (GPD) to sample exceedances of X_j above some high threshold u_j, modelling non-exceedances empirically. The full model for the marginal distribution F_{X_j} of X_j is

$$F_{X_j}(x) = \begin{cases} \tilde{F}_{X_j}(x) & x \leq u_j \\ \tilde{F}_{X_j}(u_j) + \{1 - \tilde{F}_{X_j}(u_j)\} F_{\mathrm{GPD},j}(x; u_j, \sigma_j, \xi_j) & x > u_j, \end{cases} \tag{4.4.19}$$

for empirical distribution $\tilde{F}_{X_j}$ of X_j, and GPD distribution function

$$F_{\mathrm{GPD},j}(x; u_j, \sigma_j, \xi_j) = 1 - \left(1 + \frac{\xi_j(x - u_j)}{\sigma_j}\right)_+^{-1/\xi_j}, \quad x > u_j,$$

for scale and shape parameters $\sigma_j > 0$ and $\xi_j \in \mathbb{R}$, with $y_+ = \max(y, 0)$ for $y \in \mathbb{R}$. The conditioning thresholds u_j, $j = 1, \ldots, d$, are chosen so the asymptotic behaviour justifying the use of the GPD tail distribution holds approximately. Appropriate values of these thresholds are typically selected by either manually examining the stability of σ_j and ξ_j when fitting to exceedances above candidate values for u_j, or using automated methods such as those of Varty et al. (2021) and Murphy et al. (2025). We find parameter stability tests to be satisfactory for our data, see SM4.1. Given a choice of u_j, parameters σ_j and ξ_j are found using maximum likelihood techniques.

The marginal model F_{X_j} is used to map X_j onto X'_j with Laplace margins, via the probability integral transform

$$X'_j = \begin{cases} \log\left\{2F_{X'_j}\left(X'_j\right)\right\} & X'_j \leq F^{-1}_{X'_j}(0.5) \\ -\log\left\{2\left[1 - F_{X'_j}\left(X'_j\right)\right]\right\} & X'_j > F^{-1}_{X'_j}(0.5), \end{cases} \tag{4.4.20}$$

for $j = 1, \ldots, d$, obtaining the multivariate Laplace-scale environment variable $\mathbf{X}' = (X'_1, \ldots, X'_d)$.

Joint dependence modelling

We now apply the conditional extremes framework to the Laplace-scale environment variable $\mathbf{X}' \in \mathbb{R}^d$. This requires specifying a conditioning environmental variable $X'_j \in \mathbb{R}$, followed by modelling the remaining variables $\mathbf{X}'_{-j} \in \mathbb{R}^{d-1}$, conditional on the event $X'_j > v_j$ for $v_j > 0$, $j = 1, \ldots, d$. Fitting this model allows simulation of new multivariate events with extremal dependence structure representative of the original process $\mathbf{X}'$, facilitating estimation of joint extreme event set probabilities.

Broadly following Keef et al. (2013), Heffernan and Tawn (2004) assume that, for $j = 1, \ldots, d$, there exist unique values $\boldsymbol{\alpha}_{|j} \in [-1, 1]^{d-1}$, $\boldsymbol{\beta}_{|j} \in (-\infty, 1]^{d-1}$ and $\mathbf{z}_{|j} \in \mathbb{R}^{d-1}$, such that

$$\lim_{v_j \to \infty} \mathbb{P}\left(\frac{\mathbf{X}'_{-j} - \boldsymbol{\alpha}_{|j} X'_j}{X_j'^{\boldsymbol{\beta}_{|j}}} < \mathbf{z}_{|j}, X'_j - v_j > x | X'_j > v_j \right) = e^{-x} G_{|j}(\mathbf{z}_{|j}), \tag{4.4.21}$$

for $x > 0$ and distribution function $G_{|j} : \mathbb{R}^{d-1} \mapsto \mathbb{R}$ with non-degenerate marginals, where componentwise operations are assumed. In practice, the limit (4.4.21) is assumed to hold for some suitably large finite threshold v_j, yielding the regression

$$\mathbf{X}'_{-j} | \{X'_j = x\} = \boldsymbol{\alpha}_{|j} x + x^{\boldsymbol{\beta}_{|j}} \mathbf{Z}_{|j}, \tag{4.4.22}$$

for $x > v_j$, and residual random variable $\mathbf{Z}_{|j}$ independent of X'_j given $X'_j > v_j$, where element-wise operations are assumed. Regression (4.4.22) is then used to model all data in the region $\{\mathbf{X}'_j \in \mathbb{R}^d : X'_j > v_j\}$, and parameters $\boldsymbol{\alpha}_{|j}$ and $\boldsymbol{\beta}_{|j}$ are estimated using standard maximum likelihood estimation (MLE) techniques. For this estimation we utilise the additional parameter constraints of Keef et al. (2013) ensuring consistency of conditional return level values between extremal dependence types. For model fitting

only, it is assumed that $G_{|j}$ corresponds to independent Gaussian distributions with unknown means and variances. Once parameter estimates have been obtained, we follow Winter and Tawn (2017) and model $G_{|j}$ using the Gaussian kernel smoothed density estimate of the observed values of residual

$$\mathbf{Z}_{|j} = \frac{\mathbf{X}'_{-j} - \boldsymbol{\alpha}_{|j} X'_j}{X_j^{\boldsymbol{\beta}_{|j}}},$$

for $X'_j > v_j$, smoothed using kernel bandwidth $\delta_{\text{HT}}\hat{\sigma}_j > 0$, for empirical residual standard deviation $\hat{\sigma}_j$ and scale factor δ_{HT}. The conditioning threshold v_j is chosen by studying parameter stability above candidate values, see SM4.1. The selection of δ_{HT} is considered in SM4.2. This model is fitted for all choices of the conditioning variable $\mathbf{X}'_j$, allowing simulation of $\mathbf{X}'$ in each of the corresponding regions $\{\mathbf{X}' \in \mathbb{R}^d : X'_j > v_j\}$ as described by Heffernan and Tawn (2004).

The environment density $f_{\mathbf{X}}$ is then estimated as in Speers et al. (2024), using prediction from fitted models in each of the the upper tail regions $\mathcal{E}_{\mathbf{X}}{}^{(j)} = \{\mathbf{X}' \in \mathbb{R}^d : X'_j > v_j\}$, $j = 1, \ldots, d$, followed by empirical estimation in the remaining lower region $\mathcal{E}_{\mathbf{X}}{}^{\text{Lw}} = \{\mathbf{X}' \in \mathbb{R}^d : X_j \leq v_j \forall j\}$. In a given upper region $\mathcal{E}_{\mathbf{X}}{}^{(j)}$, we make Laplace-scale simulations from the joint dependence model (4.4.21) (using the parameter estimates found via MLE), followed by marginal transformation back to the physical scale using the inverse of transformation (4.4.20). During simulation in $\mathcal{E}_{\mathbf{X}}{}^{(j)}$, we reject realisations for which $\max_{j':j' \neq j} X'_{j'} > X'_j$. This simulation yields a set of n_{Sm} realisations of $\mathbf{X}$ within $\mathcal{E}_{\mathbf{X}}{}^{(j)}$, from which we may empirically estimate the probability density $f_{\mathbf{X}}$ over a gridded set of subregions of $\mathcal{E}_{\mathbf{X}}{}^{(j)}$. Specifically, for a set D of feasible values of $\mathbf{X}$ such that $\mathbb{P}(\mathbf{X} \in \mathcal{E}_{\mathbf{X}} \setminus D) \approx 0$, we partition D using grid $(D_1, \ldots, D_{n_{\text{Gr}}})$. We then assume that each $|D_i|$, $i = 1, \ldots, n_{\text{Gr}}$, is small enough for the approximation

$$\mathbb{P}(\mathbf{X} \in D_i) = \int_{\mathbf{s} \in D_i} f_{\mathbf{X}}(\mathbf{s}) \mathrm{d}\mathbf{s} \approx |D_i| f_{\mathbf{X}}(\mathbf{x}), \tag{4.4.23}$$

to be suitable, assuming that $f_{\mathbf{X}}$ is reasonably constant for all $\mathbf{x} \in D_i$. For any D_i within an upper tail region ${\mathcal{E}_{\mathbf{X}}}^{(j)}$, $j = 1, \ldots, d$, we estimate the joint density with

$$\hat{f}_{\mathbf{X}}^{(j)}(\mathbf{x}) = \frac{n_{\mathrm{Sm}}^{(i)}}{n_{\mathrm{Sm}}|D_i|}, \quad \mathbf{x} \in D_i \subset {\mathcal{E}_{\mathbf{X}}}^{(j)},$$

where $n_{\mathrm{Sm}}^{(i)}$ is the number of simulated values of $\mathbf{x}$ in D_i. Combining these estimates with the empirical density $\tilde{f}_{\mathbf{X}}$ used in the lower region ${\mathcal{E}_{\mathbf{X}}}^{\mathrm{Lw}}$, the full density estimate for $\mathbf{x} \in \mathcal{E}_{\mathbf{X}}$ is

$$\hat{f}_{\mathbf{X}}(\mathbf{x}) = \begin{cases} \tilde{f}_{\mathbf{X}}(\mathbf{x}) & \mathbf{x} \in {\mathcal{E}_{\mathbf{X}}}^{\mathrm{Lw}} \cap D, \\ \hat{f}_{\mathbf{X}}^{(j)}(\mathbf{x}) & \mathbf{x} \in {\mathcal{E}_{\mathbf{X}}}^{(j)} \cap D, \quad j = 1, \ldots, d, \\ 0 & \mathbf{x} \notin D. \end{cases} \tag{4.4.24}$$

Estimate of the environment density

Figure 4.4.11 shows the resulting estimate (4.4.24) of the environment density $f_{\mathbf{X}}$, found using the conditional extremes model. We take marginal thresholds $u_1 = \tilde{F}_{H_s}^{-1}(0.7)$ and $u_2 = \tilde{F}_{S_e}^{-1}(0.7)$, where $\tilde{F}_{H_s}$ and $\tilde{F}_{S_e}$ are the empirical distribution functions of H_s and S_e. The conditioning threshold for H_s is chosen as $v = \tilde{F}_{H_s}^{-1}(0.6)$. When simulating joint values of $\mathbf{X}$ conditional on $H_s > v$, we take $\delta_{\mathrm{HT}} = 0.4$, $n_{\mathrm{Sm}} = 10^5$, $D = [3, 12] \times [0.01, 0.05]$ and $n_{\mathrm{Gr}} = 90 \times 45$; these values yield $|D_i|$ small enough for approximation (4.4.23) to be reasonable, for negligible computational cost when estimating density $f_{\mathbf{X}}$. We choose not to fit the conditional extremes model to the region where S_e is large since in typical offshore applications, large values of H_s rather than S_e dominate structural failure. We model the density empirically for values of H_s below the conditioning threshold v, and apply a Gaussian kernel smoother with bandwidth chosen according to Scott (2015).

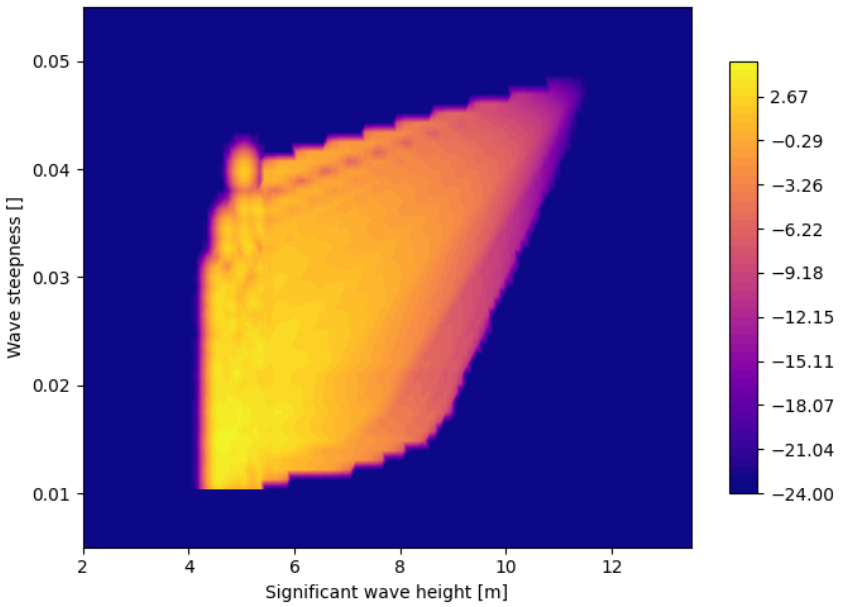


Figure 4.4.11: Estimate of the log of the joint density $f_{\mathbf{X}}$ of environment variable $\mathbf{X} = (H_S, S_e)$, using the conditional extremes model of Heffernan and Tawn (2004) fitted to storm peak data from a location 1km offshore of Albany.

4.4.4 Non-linear harmonic structural response simulation

Overview of response simulation

We now obtain empirical distributions of the structural response $R|\{\mathbf{X} = \mathbf{x}\}$, $\mathbf{X} = (H_s, S_e)$, in our model monopile scenario, given a fixed environment $\mathbf{x} \in D$. This is achieved using realisations $R_j|\{\mathbf{X} = \mathbf{x}\}$, $j = 1, \ldots, n_{\text{Rl}}$, of the structural response, obtained via repeated direct simulation using physical models of environmental loading on the monopile. For each of n_{Rl} realisations of fluid loading, this requires: (a) simulation of hour-long time series realisations of the stochastic linear wave elevation $\{E_j(t;\mathbf{x}) : t \in [0, 60^2]\}$ for an underlying environment $\mathbf{x}$; (b) conversion of the surface elevation $E_j(t;\mathbf{x})$ to load $L_j(t;\mathbf{x})$ induced on a monopile; (c) transformation of this load via a linear response function to resonant response $R_j(t;\mathbf{x})$ time series observed on the model structure; and (d) isolation of maximum response $R_j|\{\mathbf{X} = \mathbf{x}\}$ from the time series $R_j(t;\mathbf{x})$. Note that all physical quantities are given in SI units throughout this section. Further, $R_j(t;\mathbf{x})$ refers to a time series of response whereas $R_j|\{\mathbf{X} = \mathbf{x}\}$ is the

maximum response observed over the time series,

$$R_j|\{\mathbf{X}=\mathbf{x}\} = \max_{t\in[0,60^2]} R_j(t;\mathbf{x}). \tag{4.4.25}$$

Step (a) is achieved by modelling the surface elevation according to linear wave theory (see e.g., Holthuijsen 2010). We obtain the load in (b) using linear surface elevation as input to the methods of Taylor et al. (2024) and Orszaghova et al. (2025), outputting an approximation to the non-linear inertial load $L_j(t;\mathbf{x})$ that $E_j(t;\mathbf{x})$ induces on a monopile. For (c) we pass this load through the linear response function for a damped harmonic oscillator, obtaining a realisation of harmonic response time series $R_j(t;\mathbf{x})$. Steps (a)-(d) are repeated for all n_{Rl} realisations, yielding a numerical estimate of the environment conditioned response $R|\{\mathbf{X}=\mathbf{x}\}$.

Simulation details

Under linear wave theory, the surface elevation $E_j(t;\mathbf{x})$ at time $t>0$, in a sea state with parameter $\mathbf{X}=\mathbf{x}$, is modelled as the finite sum of Fourier components at n_{Fr} evenly spaced frequencies $f_1,\ldots,f_{n_{\mathrm{Fr}}}>0$, $f_2-f_1=\Delta_{\mathrm{Fr}}$, with contributions determined by underlying wave spectrum $S(f;\mathbf{x})$. We take $n_{\mathrm{Fr}}=60^2$, $f_1=10^{-3}$ and $f_{n_{\mathrm{Fr}}}=1$, and the JONSWAP (Hasselmann et al., 1973) parametric form for $S(f;\mathbf{x})$ (see SM2) and then model the surface elevation at the location of the structure as

$$E_j(t;\mathbf{x}) = \sum_{i=1}^{n_{\mathrm{Fr}}} \Big\{ A_i|\{\mathbf{X}=\mathbf{x}\}\cdot\cos(2\pi f_i t) + B_i|\{\mathbf{X}=\mathbf{x}\}\cdot\sin(2\pi f_i t)\Big\}, \quad t>0, \tag{4.4.26}$$

where $A_i|\{\mathbf{X}=\mathbf{x}\},\ B_i|\{\mathbf{X}=\mathbf{x}\}\sim N(0,\Delta_{\mathrm{Fr}}S(f_i;\mathbf{x}))$, $i=1,\ldots,n_{\mathrm{Fr}}$, are random Gaussian coefficients with variance equal to the wave energy in frequency band $(f_i-\Delta_{\mathrm{Fr}}/2, f_i+\Delta_{\mathrm{Fr}}/2)$ of the discretised wave spectrum. Model (4.4.26) assumes the monopile is placed at the spatial origin and is concentrated at this point with no spatial dimensions. The wave surface elevation (4.4.26) is stochastic due to the random

Gaussian coefficients, requiring multiple realisations of hourly time series $\{E_j(t;\mathbf{x}) : t \in [0, 60^2]\}$ to capture the full behaviour of the wave surface when $\mathbf{X} = \mathbf{x}$.

The method of Orszaghova et al. (2025) takes these linear surface elevation time series $E_j(t;\mathbf{x})$, recovers the non-linear higher-order harmonics of the wave signal (see SM4.3) and outputs a time series of non-linear horizontal monopile loading $L_j(t;\mathbf{x})$ using the T-FNV model of Taylor et al. (2024). We omit the full details of this methodology here as it is beyond the scope of our case study; in short, the method allows evaluation of structural load without the need to calculate full wave-kinematic profiles (see e.g., Speers et al. 2024), greatly increasing computational efficiency. For this reason, the T-FNV approach is well-suited for our example scenario, as it provides physically-accurate model output (thus testing our methodology in a realistic setting) at a low computational cost (allowing us to generate the full true CDE (4.1.4) as the target). This method requires specifying a water depth d, for which we take $d = 30$.

To approximate the effect of wave-induced oscillation on the model monopile, we pass the load $L_j(t;\mathbf{x})$ through the linear response, or transfer, function of a damped harmonic oscillator. For input signal $L_j(t;\mathbf{x})$, the output signal $R_j(t;\mathbf{x})$ is then defined as

$$\chi_{R_j}(f) = \chi_T(f;\gamma)\chi_{L_j}(f), \quad f > 0,$$

where transfer function $\chi_T(f;\gamma)$, the ratio of Fourier transform of the output to the input, takes the form

$$\chi_T(f;\gamma) = \frac{1}{f_0 - f^2 + i\gamma f}. \tag{4.4.27}$$

Alternatively,

$$R_j(t;\mathbf{x}) = \mathcal{F}^{-1}\{\chi_T(f) \cdot \mathcal{F}(L_j(t;\mathbf{x})\}, \tag{4.4.28}$$

where $\mathcal{F} : \mathbb{R} \mapsto \mathbb{R}^+$ is the Fourier transform mapping functions in the time domain to the frequency domain. See SM4.3 for further discussion of the transfer function (4.4.27). From time series (4.4.28), we obtain a realisation of maximum response (4.4.25). Steps

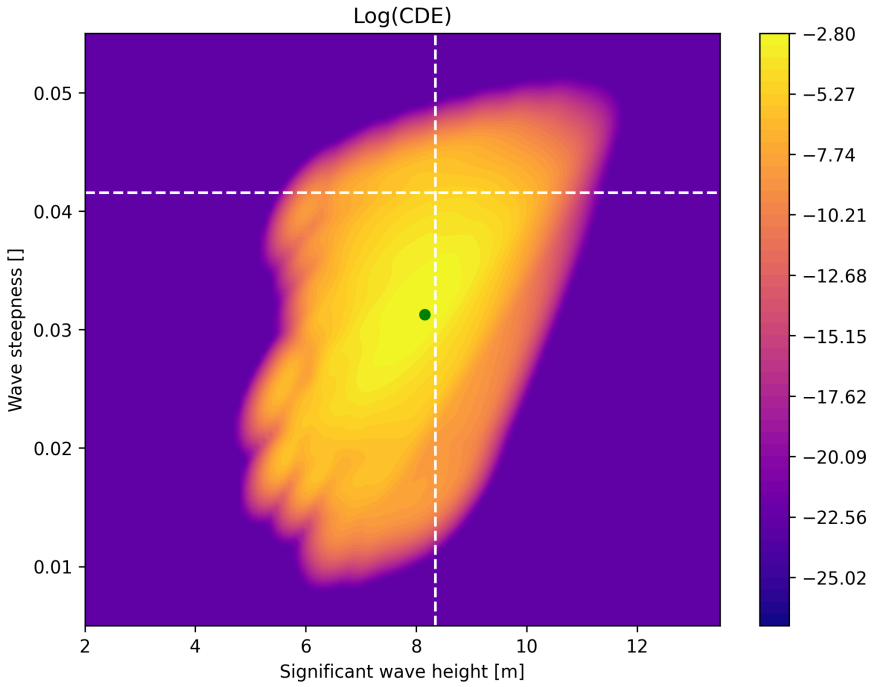


Figure 4.4.12: Conditional density of the environment (CDE) for the oscillating monopile scenario, conditioned on exceedance of the 50-year response event. White lines show the marginal 50-year events. The mode of the CDE is indicated in green.

(a)-(d) are carried out over a grid of environment values $\mathbf{x}_1, \ldots, \mathbf{x}_{n_{\mathrm{Gr}}}$, chosen as the centre points of the cells D_i, $i = 1, \ldots, n_{\mathrm{Gr}}$, of $\mathcal{E}_{\mathbf{X}}$ used to estimate the environment density via (4.4.24). At each grid point $\mathbf{x}_i$, we obtain $n_{\mathrm{Rl}} = 1000$ realisations of the response $R|\{\mathbf{X} = \mathbf{x}_i\}$, for centre value $\mathbf{x}_i$ of D_i. This procedure provides an empirical estimate $\tilde{F}_{R|\mathbf{X}}(\cdot|\mathbf{x})$ of the distribution of response $R|\{\mathbf{X} = \mathbf{x}\}$, for $\mathbf{x} = \mathbf{x}_1, \ldots, \mathbf{x}_{n_{\mathrm{Gr}}}$, from $n_{\mathrm{Rl}} = 1000$ realisations $R_j|\{\mathbf{X} = \mathbf{x}\}$, $j = 1, \ldots, n_{\mathrm{Rl}}$, for each $\mathbf{x}$.

4.4.5 Benchmarking: obtaining a good estimate of CDE and probability of failure

We choose to exploit knowledge of the distribution of $R|\{\mathbf{X} = \mathbf{x}\}$ over the full grid of $\mathbf{x}_1, \ldots, \mathbf{x}_{n_{\mathrm{Gr}}} \in \mathcal{E}_{\mathbf{X}}$ in order to obtain a good estimate of CDE. In practical application, we would not attempt this estimation, since it requires a prohibitively expensive total of $n_{\mathrm{Rl}} \times n_{\mathrm{Gr}}$ function evaluations of $R|\{\mathbf{X} = \mathbf{x}\}$. However, with this good estimate of

CDE, we are able to evaluate the performance of the IS-PT and AGE procedures, the main objective of this section, reported in Section 4.4.7. The CDE is estimated using the simulation of Section 4.4.4 as

$$\tilde{f}_{\mathbf{X}}(\mathbf{x}; r_{\mathrm{Cr}}) = \left\{1 - \tilde{F}_{R|\mathbf{X}}(r_{\mathrm{Cr}}|\mathbf{x})\right\} \times \hat{f}_{\mathbf{X}}(\mathbf{x}), \tag{4.4.29}$$

for empirical distribution $\tilde{F}_{R|\mathbf{X}}(\cdot|\mathbf{x})$. Further, $\hat{f}_{\mathbf{X}}$ is the estimated environment density (4.4.24) and r_{Cr} is the critical response. The resulting CDE estimate, smoothed using a Gaussian kernel smoother with Scott (2015) bandwidth, is shown in Figure 4.4.12. The white dashed lines show the marginal 50-year events for both H_s and S_e, found using the marginal extreme value models (4.4.19). The modal point of the estimated CDE (4.4.29) is indicated in green.

In practice, the critical response r_{Cr} is specified by domain experts from detailed knowledge of the structure, and the corresponding failure probability then estimated as discussed in Section 4.1. Here, we set the value of r_{Cr} in order to yield a known failure probability for testing purposes. The critical response r_{Cr} is set to $r_{\mathrm{Cr}} = \tilde{F}_{R_A}^{-1}(1-1/50)$, the 50-year response event, where

$$\tilde{F}_{R_A} = \sum_{m=0}^{\infty} [\tilde{F}_R(r)]^m \frac{\rho_{\mathrm{St}} e^{-\rho_{\mathrm{St}}}}{m!} = \exp\big[-\rho_{\mathrm{St}}(1 - \tilde{F}_R(r)\big],$$

is the empirical distribution of the annual maximum response R_A. Further, $\rho_{\mathrm{St}} = 26$ is the expected number of storms per annum estimated empirically from the data, and

$$\tilde{F}_R(r) = \int_{\mathcal{E}_{\mathbf{X}}} \tilde{F}_{R|\mathbf{X}}(r|\mathbf{x}) \hat{f}_{\mathbf{X}}(\mathbf{x}) \mathrm{d}\mathbf{x},$$

the empirical distribution of the marginal response R for a single storm event. See Section 3.1.2 of Speers et al. (2024) for further discussion of annual response distribution estimation. The resulting 'single storm' failure probability in this case becomes $p_{\mathrm{TFNV}} =$

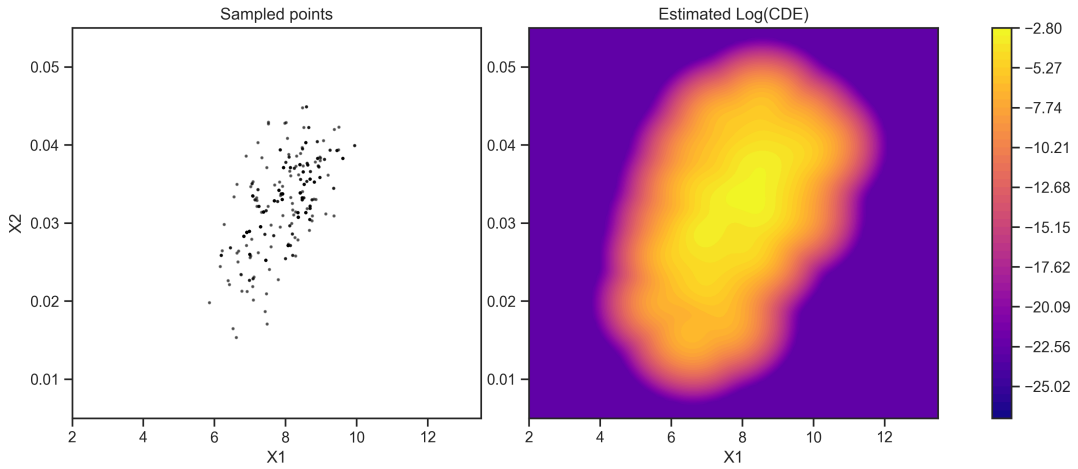


Figure 4.4.13: Example sample from CDE (4.1.4) under the synthetic structural scenario obtained using the adaptive parallel tempering MCMC algorithm of Vousden et al. (2015) (left), and corresponding smoothed log-CDE estimated using Gaussian kernel bandwidth selected according to Scott (2015) (right).

1.1×10^{-3}.

4.4.6 IS-PT results

We use the IS-PT approach of Section 4.2 to emulate the CDE (4.4.29), taking sampling parameter values $n_{\mathrm{Tm}} = 5$, $n_{\mathrm{PT}} = 400$, $n_{\mathrm{IS}} = 100$, initial proposal variance $\sigma^2_{\mathrm{MH}} = 1$, and bounding temperatures $T_1 = 1$, $T_5 = 20$ seen to perform well in Section 4.3.2. For $n_{\mathrm{Rp}} = 100$ replicates, the adaptive algorithm of Vousden et al. (2015) is used to obtain a sample of size $n_{\mathrm{PT}} = 400$ (minus burn-in length n_{Br}) from the CDE. A Gaussian kernel smoothed estimate with Scott's bandwidth is then used as proposal density p_{Pr} in importance sampling estimate (4.2.5). Figure 4.4.13 shows an example MCMC sample and resulting proposal estimate of the CDE at a single replicate. Over all $n_{\mathrm{IS}} = 100$ replicates, we obtain $\mathrm{RMSE}(\hat{p}_{\mathrm{IS}}) = \left(\sum_{i=1}^{n_{\mathrm{Rp}}} (\hat{p}_{\mathrm{IS}}^{(i)} - p_{\mathrm{TFNV}})^2 / n_{\mathrm{Rl}}\right)^{1/2} = 5.10 \times 10^{-5}$ where $\hat{p}_{\mathrm{IS}}^{(i)}$ is the probability estimate (4.2.5) obtained at replicate i, for true $p_{\mathrm{TFNV}} = 1.1 \times 10^{-3}$. The corresponding bias $\mathrm{Bias}(\hat{p}_{\mathrm{IS}}) = 1.83 \times 10^{-6}$ is also small.

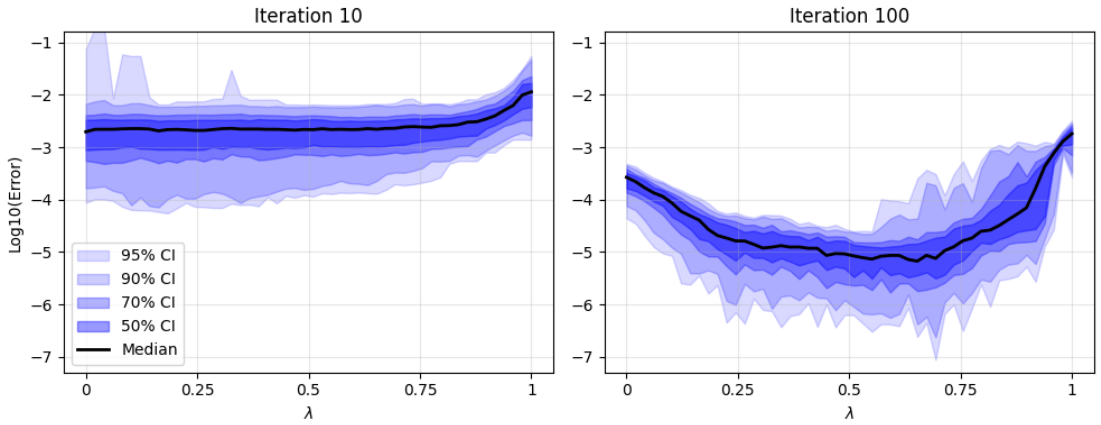


Figure 4.4.14: Log-scale absolute error Δ_{GP} of the GP probability estimate $\hat{p}_{\text{GP}}$ at specified iterations, for emulator (4.2.6) trained using $U^{(1)}$ over the range of weight $\lambda \in [0.01, 0.99]$ for the TFNV scenario. At iteration 100, the weight that minimises median error is $\lambda^* = 0.67$.

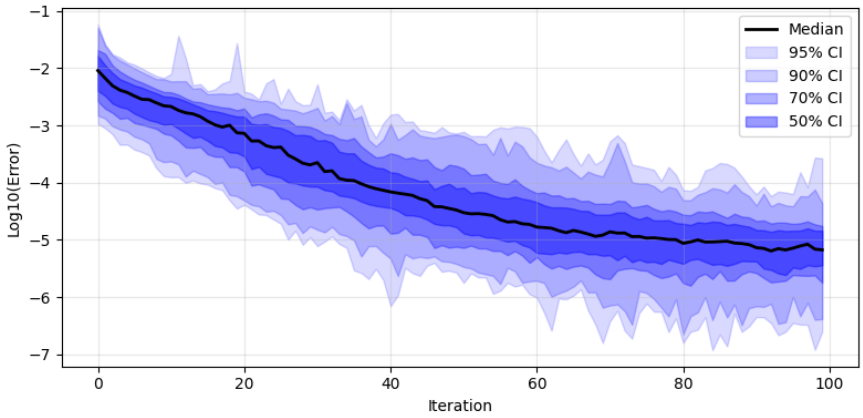


Figure 4.4.15: Distribution of log-scale absolute error Δ_{GP} in the GP probability estimate with respect to iteration, trained using $U^{(1)}$ with $\lambda = \lambda^*$. The trend in median error is indicated in black, with various confidence intervals shown in blue.

4.4.7 AGE results

We now use the AGE methods of Section 4.2 to emulate the log-CDE, seeking a reasonable estimate of probability of failure with considerably fewer than the $n_{\text{Gr}} \times n_{\text{Rl}}$ function evaluations of $R|\{\mathbf{X} = \mathbf{x}\}$ used for the IS-PT estimate in Section 4.4.5. The emulator for CDE (4.4.12) is defined as in (4.2.6). For $n_{\text{Rp}} = 100$ replicates, it is initialised using Latin hypercube sample $\mathcal{D}_0$, $|\mathcal{D}_0| = 100$, then trained inductively over n_{Rl} realisations

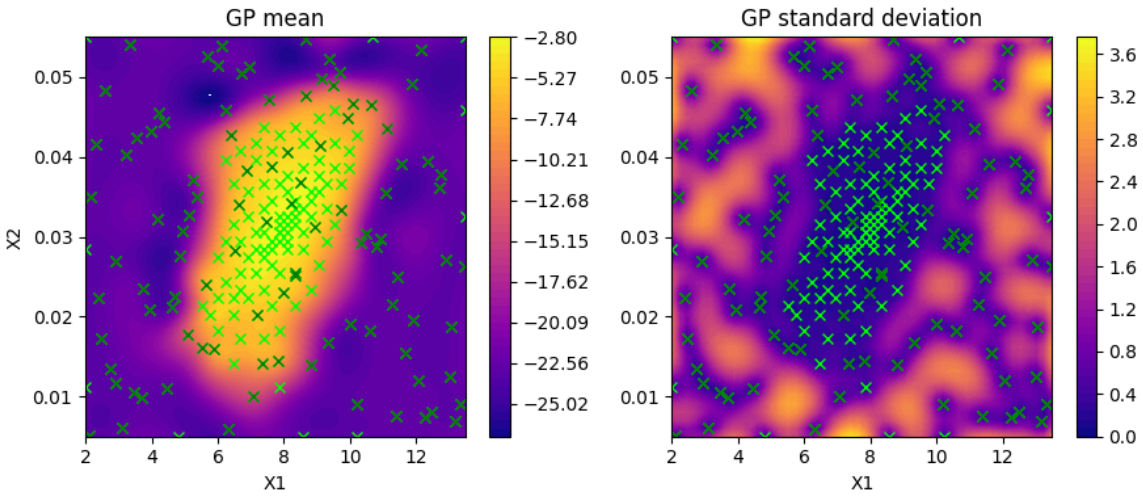


Figure 4.4.16: GP emulator at iteration 100 for variance utility $U^{(1)}$ (4.2.10), trained using the optimal value $\lambda^* = 0.67$ minimising median of error Δ_{GP}. The panels from left to right show the posterior GP mean $\mu^*_{100}(\mathbf{x})$ over $\mathbf{x} \in \mathcal{E}_{\mathbf{X}}$, and the posterior GP standard deviation $k^*_{100}(\mathbf{x}, \mathbf{x})^{1/2}$. The initial random Latin hypercube training set $\mathcal{D}_0$ is shown as dark green crosses, and the iteratively selected new training points $\mathcal{D}_{100} \setminus \mathcal{D}_0$ are shown as light green crosses.

of $n_{\text{It2}} = 100$ iterations, using $U^{(1)}$ and $U^{(2)}$ for a range of weights $\lambda \in [0.01, 0.99]$. For utility $U^{(1)}$, Figure 4.4.14 shows the relationship between error Δ_{GP} and the value of λ, by comparison with the estimate of p_{TFNV} from Section 4.4.5. The weight $\lambda^* = 0.67$ is found to minimise the median value of error Δ_{GP} at iteration 100. Figure 4.4.15 shows Δ_{GP} with iteration for λ^*, and Figure 4.4.16 shows the emulator trained using λ^* at iteration 100. For $|\mathcal{D}_{100}| = 200$ total function evaluations at $\mathbf{x} \in \mathcal{D}_{100} \setminus \mathcal{D}_0$ chosen by $U^{(1)}$ with $\lambda = \lambda^*$, we obtain $\text{RMSE}(\hat{p}_{\text{GP}}) = \left(\sum_{r=1}^{n_{\text{Rp}}} (\hat{p}^{(r)}_{\text{GP}} - p_{\text{TFNV}})^2 / n_{\text{Rl}} \right)^{1/2} = 6.99 \times 10^{-5}$ where $\hat{p}^{(r)}_{\text{GP}}$ is the probability estimate (4.2.8) obtained at iteration 100 and replicate r, for true $p_{\text{TFNV}} = 1.1 \times 10^{-3}$. The corresponding bias $\text{Bias}(\hat{p}_{\text{GP}}) = 1.57 \times 10^{-5}$ is also small. Results using utility $U^{(2)}$ are reported in 4.6.6, and summarised in the next section.

4.4.8 Comparison of IS-PT and AGE performance

Figure 4.4.17 shows the distribution of the IS-PT estimate $\hat{p}_{\text{IS}}$ and the AGE estimates $\hat{p}_{\text{GP}}$ (based on variance and ALC utilities $U^{(1)}$ and $U^{(2)}$ at iteration 100, $\lambda = \lambda^*$) around the target failure probability p_{TFNV}. These results are summarised in Table 4.4.2. For the given budgets of expensive function evaluation set, as in Section 4.3.2, methods demonstrate essentially equivalent performance. Again, the key issue is specification of λ for AGE procedures. With λ known, AGE procedures are computationally more efficient. However, specification of λ is in general problematic, suggesting that IS-PT is a more reliably applicable approach.

	IS-PT	AGE $U^{(1)}$	AGE $U^{(2)}$
RMSE	5.10×10^{-5}	6.99×10^{-5}	4.99×10^{-5}
Bias	1.83×10^{-6}	1.57×10^{-5}	9.40×10^{-6}
Number of function evaluations, n_{Ev}	$n_{\text{Tm}} \times n_{\text{PT}} + n_{\text{IS}} = 2100$	$\lvert\mathcal{D}_0\rvert + n_{\text{Itr1}} = 244$	$\lvert\mathcal{D}_0\rvert + n_{\text{Itr1}} = 244$

Table 4.4.2: RMSEs and biases of $\hat{p}_{\text{IS}}$ and $\hat{p}_{\text{GP}}$ when targetting failure probability p_{TFNV}, calculated for $n_{\text{Rp}} = 100$ replicate analyses. The true value of probability of failure is $p_{\text{TFNV}} = 1.3 \times 10^{-3}$. Also shown is the number of expensive response function evaluations required for a single replicate analysis for each of IS-PT and AGE.

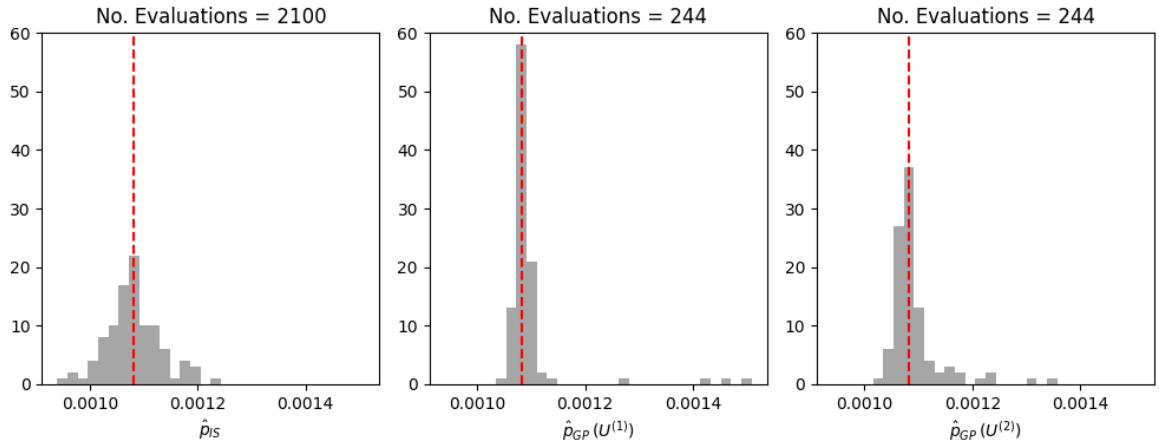


Figure 4.4.17: Distribution of $n_{\text{Rp}} = 100$ estimates $\hat{p}_{\text{IS}}$ (left), $\hat{p}_{\text{GP}}$ for $U^{(1)}$ iteration 100 with $\lambda = \lambda^*$ (centre), and $\hat{p}_{\text{GP}}$ for $U^{(2)}$ iteration 100 with $\lambda = \lambda^*$ (right) for true failure probability p_{TFNV} (red). The number of function evaluations required for a single replicate analysis is indicated in the panel titles.

4.5 Discussion

Estimation of failure probability for marine structures can be a computationally demanding task. In earlier work (Speers et al., 2024) we showed that the conditional density of the environment (CDE) for a structure is a useful design diagnostic, preferable to design contours. Moreover, CDE also provides a natural starting point for estimation of failure probability: the integral of the CDE over the environment space is the probability of structural failure.

The mode of the CDE represents the combination of long term environmental conditions most likely to induce structural failure at the 50-year level. In practice, the location of the mode depends both on the extremal dependence characteristics of the environment variables, and the nature of the fluid-structure interaction. Interestingly, for the oscillating monopile application discussed in Section 4.4, the location of the mode, calculated by brute force in Section 4.4.5, corresponds approximately to the combination of the 50-year storm peak significant wave height and the 1-year storm peak significant wave steepness.

In this work, we develop, demonstrate and compare two methods to estimate the CDE and hence failure probability for simple monopile structures. The first methodology (IS-PT) incorporates parallel tempering MCMC to estimate the CDE, together with importance sampling to estimate failure probability. The second methodology uses adaptive Gaussian emulation (AGE) to estimate the CDE and thence Bayesian quadrature to estimate failure probability.

Whereas use of either methodology requires the specification of hyperparameters, the AGE approach is particularly problematic, necessitating the specification of a key weight (λ in e.g., (4.2.9)) to control the extent to which adaptive emulation is encouraged to explore the environmental space as opposed to exploiting already-identified informative structure in that space. Specification of λ is in general case dependent.

The computational complexity of each methodology is typically dictated by the num-

ber of expensive evaluations of the structural response R given long term environmental conditions $\mathbf{X}$. If the value of explore-exploit λ is known, then procedures adopting AGE provide a good estimate of failure probability requiring an order of magnitude fewer expensive function evaluations than IS-PT. We demonstrate the good performance of IS-PT and two AGE procedures on a simple synthetic structure with complex fluid loading behaviour (and bimodal CDE), and on a more realistic monopile structure (with more straightforward unimodal CDE). Good performance for AGE procedures requires knowledge of the optimal choice of explore-exploit λ, which was evaluated by us in this work using assumed knowledge of the true structural response; obviously, this information will not generally be available to the structural designer. Nevertheless, if it is anticipated that the CDE is likely to be unimodal, we speculate that the choice of λ is likely to be less critical than for more complex CDEs. This can be seen, e.g., by comparison of the intervals I^* of minimum median error in the right hand panels of Figure 4.3.6 (bimodal CDE, AGE with variance utility $U^{(1)}$, $I^* \approx [0.6, 0.9]$), Figure 4.4.14 (unimodal, $U^{(1)}$, $I^* \approx [0.25, 0.75]$), Figure 4.6.21 (bimodal, AGE with ALC utility $U^{(2)}$, $I^* \approx [0.2, 0.35]$) and Figure 4.6.28 (unimodal CDE, $U^{(2)}$, $I^* \approx [0.1, 0.5]$). Intervals I^* of acceptable values for λ are wider for unimodal CDEs, and moreover the intervals corresponding to utilities $U^{(1)}$ and $U^{(2)}$ overlap. However, for bimodal CDEs, the intervals I^* corresponding to $U^{(1)}$ and $U^{(2)}$ are disjoint. Given the importance of selecting λ well, investigation into the performance acquisition functions not reliant on the specification of weight parameter λ (e.g., Osborne et al. 2012, Gunter et al. 2014) is warranted. These methods, however, incur additional theoretical assumptions and computational complexity, which may limit their usefulness in general offshore applications.

More generally, sampling from high-dimensional, multimodal distributions is receiving increasing attention in the academic literature, since it is a fundamental issue in modern statistical learning. Tempering methods are often used when multimodality is suspected, yet these can in principle be difficult to tune as noted by e.g., Miasojedow

et al. (2013), especially in high dimensional settings. Solutions have been proposed, including Graham and Storkey (2017), Luo et al. (2018) and Park (2025) in the context of Hamiltonian Monte Carlo sampling. Other authors (e.g., Qiu and Wang 2024) consider sampling using tempered distribution flows. In fact, the parallel tempering package exploited in the current work has already been applied successfully in five dimensions by Vousden et al. (2015); this may be sufficient for less complex metocean design studies. Nevertheless, exploration of the performance of IS-PT in higher dimensions for more complex structural types is an interesting area for future work; studies along these lines are currently being conducted by the authors.

If the optimal value of λ for AGE is unknown (as will generally be the case), IS-PT provides a reliable general-purpose approach to estimation of CDE and failure probability useful even for challenging multimodal CDEs. In the current work, we consider univariate responses in a two-dimensional environment. We anticipate that, for higher-dimensional responses and environments, the structure of the CDE will be more complex (and multimodal) in general. Given this, it appears reasonable to assume that IS-PT will prove a more reliable route to estimation of CDE and probability of structural failure. In contrast, we view AGE as a method requiring further development before it can be considered a mature engineering solution.

The current work should be considered in the context of the growing body of academic literature on the application of machine learning and related techniques in ocean engineering, in particular to the design and reanalysis of offshore structures. ElHamahmy (2025) provides a recent review, comparing different methodologies with examples of specific use cases, and a discussion of current and future challenges. Alves Ribeiro et al. (2025) provides a review specific to the design of offshore wind facilities. One popular area of application is the use of machine learning to estimate surrogate models (e.g., using deep neural networks, Gaussian processes) to emulate the output of computationally-complex physical model simulators, e.g., for fluid loading and soil

mechanics (Yu et al. 2024, Alves Ribeiro et al. 2025, Martzikos et al. 2024, Kirby et al. 2023, Mert et al. 2025). The review of Zhou et al. (2025) lists a number of outstanding issues that need to be addressed before modern machine learning tools can be adopted reliably in practical application. These include: (a) the need for adequate databases to train machine learning and deep learning models in offshore engineering contexts; (b) know-how to guide the appropriate choice of machine learning model architectures and to estimate the often large number of hyperparameters involved; (c) the need to develop interpretable machine learning models that can be understood and diagnosed by the ocean engineering practitioner; and (d) improved collaboration between specialists in machine learning and structural engineering, bridging the gap between machine learning theory and academic application, and the real world.

Acknowledgments

The work was completed while Matthew Speers was part of the EPSRC funded STOR-i centre for doctoral training (grant no. EP/S022252/1), with part-funding from the ARC TIDE Industrial Transformational Research Hub at the University of Western Australia. The authors wish to acknowledge the support of Jana Orszaghova and Paul Taylor at the University of Western Australia, and David Randell at Shell.

4.6 Supplementary material

4.6.1 Alternative methodology to that presented in Section 4.2

Gaussian process-informed importance sampling

Referring to Section 4.2.2, authors Xiao et al. (2020) and Lystad et al. (2023), use Gaussian emulation to inform their choice of importance sampling density. We briefly propose a similar approach using the GP emulators of the main article. This avoids

the need to run the MCMC sampler described in the main article, allowing either (a) more budget allocation to the evaluation of importance sampling estimate $\hat{p}_{\text{IS}}$, or (b) a reduction in total computational cost.

Our approach mirrors that of Lystad et al. (2023), who create a uniform proposal density with support informed by a GP estimate of the CDE. Given an estimate $\hat{f}^{(n)}_{\mathbf{X}|R>r_{\text{Cr}}}$ of the CDE, found via (4.2.11) using the nth-iterate GP emulator defined in the main article, we define a proposal density

$$g^{(n)}_{\text{Pr}}(\mathbf{x}) = \frac{1}{A_n} \quad \text{for} \quad A_n = \int_{\mathcal{E}_{\mathbf{X}}} \mathbb{I}\left\{\hat{f}^{(n)}_{\mathbf{X}|R>r_{\text{Cr}}}(\mathbf{x}) > \delta\right\} \mathrm{d}\mathbf{x},$$

for some $\delta \in [0, 1]$. That is, we draw a proposal sample $\mathbf{x}^*_1, \ldots, \mathbf{x}^*_N$ uniformly on the region where the estimate of the CDE at the current iteration n is greater than some δ.

Gaussian process emulation of failure probability

We consider an alternate emulator construction to that shown in Section 4.2.3. Here, we emulate only the conditional failure probability $\Pr\left(R > r_{\text{Cr}} | \{\mathbf{X} = \mathbf{x}\}\right)$, rather than the entire integrand $\Pr\left(R > r_{\text{Cr}} | \{\mathbf{X} = \mathbf{x}\}\right) f_{\mathbf{X}}(\mathbf{x})$. Since probabilities must always be observed on the unit interval, we map the Gaussian emulator output $w(\mathbf{x}) \in \mathbb{R}$ onto the range $[0, 1]$ via the logistic function $g_{\text{Lg}} : \mathbb{R} \mapsto [0, 1]$, $g_{\text{Lg}}(w) = e^w/(1 + e^w)$, modelling

$$w(\mathbf{x}) = g^{-1}_{\text{Lg}}(\Pr(G_R(\mathbf{x}) > r_{\text{Cr}})) \sim \mathcal{GP}(\mu_{\text{GP}}(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad w : \mathcal{E}_{\mathbf{X}} \mapsto \mathbb{R}, \tag{4.6.30}$$

for mean and covariance functions

$$\begin{aligned}
\mu_{\text{GP}}(\mathbf{x}) &= \mathbb{E}[w(\mathbf{x})], \quad \mu_{\text{GP}} : \mathcal{E}_{\mathbf{X}} \to \mathbb{R}, \\
k\left(\mathbf{x}, \mathbf{x}'\right) &= \mathbb{E}\left[\left(w(\mathbf{x}) - \mu(\mathbf{x})\right)\left(w(\mathbf{x}') - \mu\left(\mathbf{x}'\right)\right)\right], \quad k(\mathbf{x}, \mathbf{x}') : \mathcal{E}_{\mathbf{X}} \times \mathcal{E}_{\mathbf{X}} \to \mathbb{R}.
\end{aligned}$$

This emulator may then be trained according to the posterior update steps defined in the main article. The target marginal failure probability estimate $\hat{p}_{\mathrm{GP}}$ can then be summarised using

$$\begin{aligned}\hat{p}_{\mathrm{GP}} &= \mathbb{E}_{W,\mathbf{X}}(\{g_{\mathrm{Lg}}(w(\mathbf{x}))\}) \\ &= \int_{\mathcal{E}_{\mathbf{X}}} \left\{ \int_{\mathbb{R}} g_{\mathrm{Lg}}(w)\phi(w; \mu^*_{\mathrm{GP}}(\mathbf{x}), k^*(\mathbf{x},\mathbf{x}))\mathrm{d}w \right\} f_{\mathbf{X}}(\mathbf{x})\mathrm{d}\mathbf{x}, \end{aligned} \tag{4.6.31}$$

for $W|\{\mathbf{X} = \mathbf{x}\} \sim N(\mu^*_{\mathrm{GP}}(\mathbf{x}), k^*(\mathbf{x},\mathbf{x}))$ with parameters obtained from (4.6.30). The estimate (4.6.31) can be written

$$\hat{p}_{\mathrm{GP}} \approx \int_{\mathcal{E}_{\mathbf{X}}} g_{\mathrm{Lg}} \left(\frac{\mu^*_{\mathrm{GP}}(\mathbf{x})}{\sqrt{1 + \pi k(\mathbf{x},\mathbf{x}')/8}} \right) f_{\mathbf{X}}(\mathbf{x})\mathrm{d}\mathbf{x}, \tag{4.6.32}$$

by the approximation for the convolution of a logistic sigmoid function with a Gaussian density given in Section 4.5.2 of Bishop and Nasrabadi (2006). Values for (4.6.32) may then be obtained via numerical integration. In this setting, the CDE estimate of the GP emulator at iteration n becomes

$$\hat{f}^{(n)}_{\mathbf{X}|R>r_{\mathrm{Cr}}}(\mathbf{x}) = \frac{g_{\mathrm{Lg}}\left(\mu^*_n(\mathbf{x})(1 + \pi k^*_n(\mathbf{x},\mathbf{x}')/8)^{-\frac{1}{2}}\right) f_{\mathbf{X}}(\mathbf{x})}{\int_{\mathcal{E}_{\mathbf{X}}} g_{\mathrm{Lg}}\left(\mu^*_n(\mathbf{x})(1 + \pi k^*_n(\mathbf{x},\mathbf{x}')/8)^{-\frac{1}{2}}\right) f_{\mathbf{X}}(\mathbf{x})\mathrm{d}\mathbf{x}}.$$

4.6.2 JONSWAP wave spectrum discussed in Section 4.4

The JONSWAP spectral density (Hasselmann et al., 1973) is used to simulate linear random waves in the monopile application of the main article. It is defined, in terms of angular frequency $\omega = 2\pi f$, as

$$S(\omega;\mathbf{x}) = \alpha\omega^{-r} \exp\left\{ -\frac{r}{4}\left(\frac{|\omega|}{\omega_p(\mathbf{x})} \right)^{-4} \right\} \gamma^{\delta(\omega;\mathbf{x})},$$

for $\omega > 0$, where $\mathbf{X} = (H_s, S_e)$ and $\omega_p(\mathbf{x}) = 2\pi/t(\mathbf{x})$, where $t(\mathbf{x})$ is the observed value of the second spectral moment wave period $T_2 = \sqrt{(2\pi H_S)/(gS_e)}$ in sea state $\mathbf{X} = \mathbf{x}$,

with

$$\delta(\omega;\mathbf{x}) = \exp\left\{-\frac{1}{2\left(0.07+0.02\cdot I\{\omega_p(\mathbf{x})>|\omega|\}\right)^2}\left(\frac{|\omega|}{\omega_p(\mathbf{x})}-1\right)^2\right\},$$

and constants $r, \alpha, \gamma > 0$. The Phillips constant α is chosen so that

$$4\cdot\left\{\int_{-\infty}^{\infty} S(\omega;\mathbf{x})\mathrm{d}\omega\right\}^{\frac{1}{2}} = h(\mathbf{x}),$$

where $h(\mathbf{x})$ is the observed value of significant wave height H_S in sea state $\mathbf{X}=\mathbf{x}$.

4.6.3 Supplementary results to case studies of Section 4.3

Importance sampling coupled with parallel tempering MCMC (IS-PT)

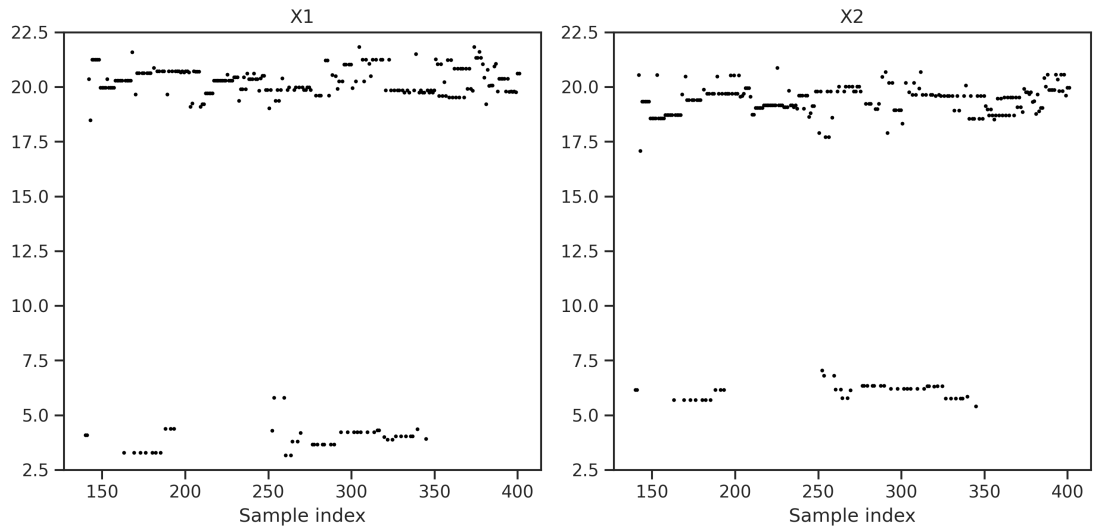


Figure 4.6.18: Example trace plot of sample from synthetic CDE discussed in Section 3, obtained from adaptive parallel tempering algorithm of Vousden et al. (2015) at temperature $T_1 = 1$, used in IS-PT approach for estimation of proposal density. Sample is initially of length $n_{\mathrm{PT}} = 400$, with $n_{\mathrm{Br}} = 144$ discarded (not plotted).

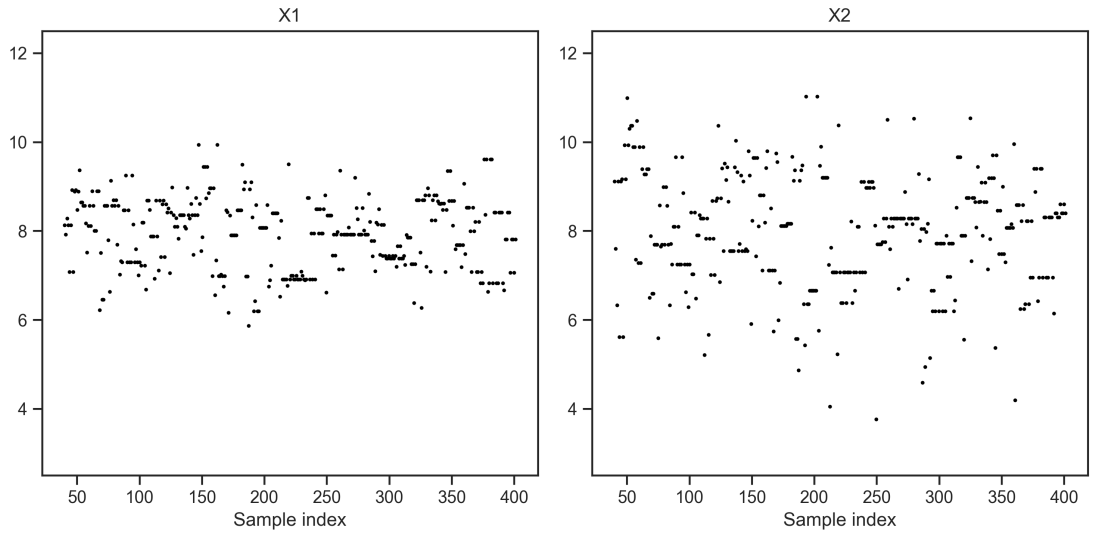


Figure 4.6.19: Example trace plot of sample from monopile CDE discussed in Section 4, obtained from the adaptive parallel tempering algorithm of Vousden et al. (2015) at temperature $T_1 = 1$, used in IS-PT approach for estimation of proposal density. Sample is initially of length $n_{\mathrm{PT}} = 400$, with $n_{\mathrm{Br}} = 40$ discarded (not plotted).

Adaptive Gaussian emulation (AGE)

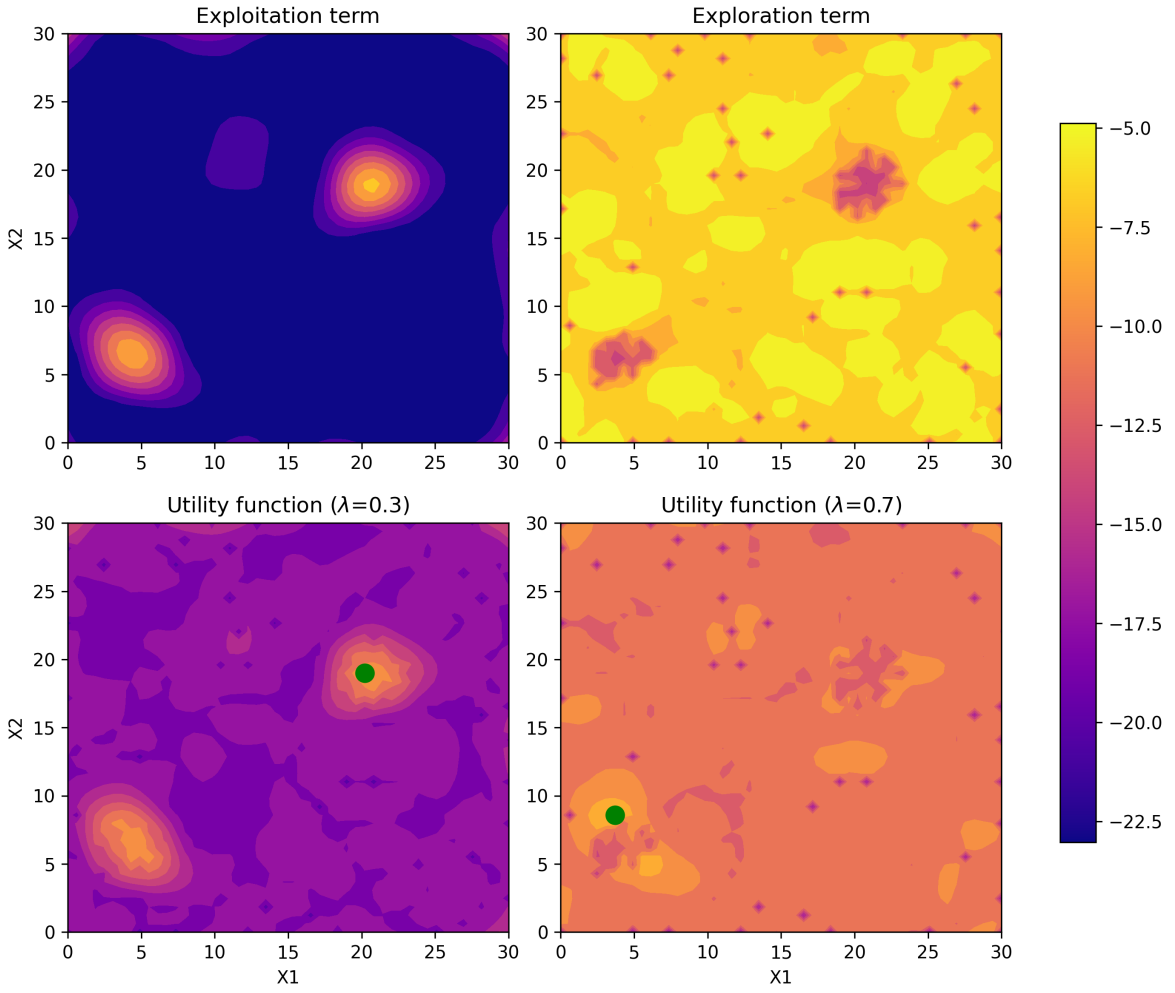


Figure 4.6.20: Behaviour of utility function $U^{(2)}(\mathbf{x};\lambda)$ over the environment space $\mathcal{E}_{\mathbf{X}}$, for synthetic scenario. Upper panels show exploitation and exploration terms obtained from GP emulator (10) trained on initial Latin hypercube set $\mathcal{D}_0$ of size $n_{\text{Tr1}} = 144$. Lower panels show resulting utility functions for weights $\lambda = 0.3$ and $\lambda = 0.7$. In each lower panel, the optimal sampling point $\mathbf{x}^* = \text{argmax}_{\mathbf{x}\in\mathcal{E}_{\mathbf{X}}} U^{(2)}(\mathbf{x};\lambda)$ is indicated in green. To be compared with Figure 5.3.5.

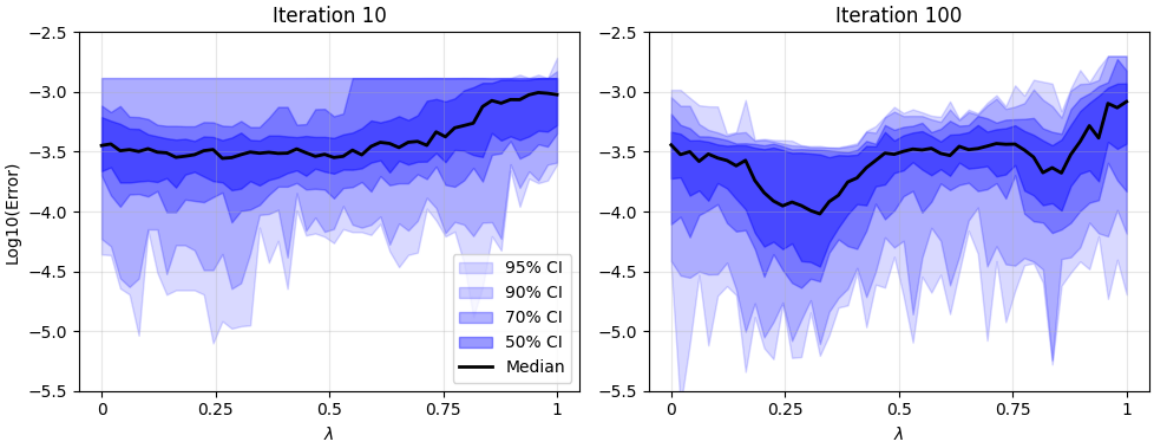


Figure 4.6.21: Log-scale absolute error Δ_{GP} of the GP probability estimate $\hat{p}_{\text{GP}}$ at specified iterations, for emulator (10) trained using $U^{(2)}$ over the range of weight $\lambda \in [0.01, 0.99]$ for the synthetic scenario. At iteration 100, the weight that minimises median error is $\lambda^* = 0.33$. To be compared with Figure 4.3.6.

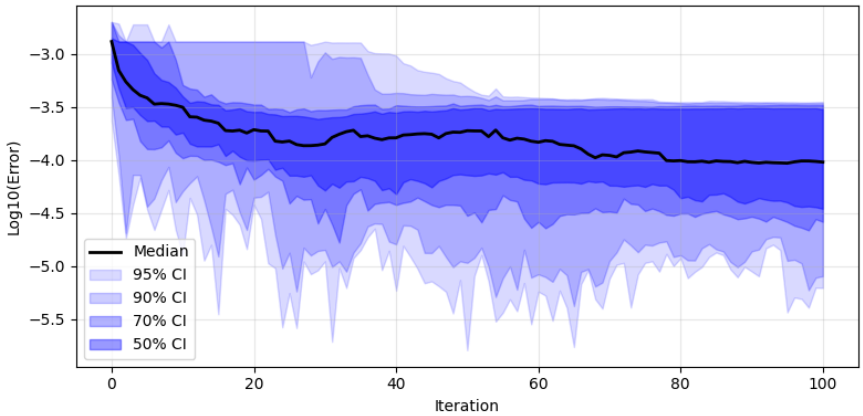


Figure 4.6.22: Distribution of log-scale absolute error Δ_{GP} in the GP probability estimate with respect to iteration, trained using $U^{(2)}$ with $\lambda = \lambda^*$. The trend in median error is indicated in black, with various confidence intervals shown in blue. To be compared with Figure 4.3.7.

4.6.4 Supplementary results to monopile case study of Section 4.4

Extreme value model threshold selection supporting the discussion of Section 4.4.3

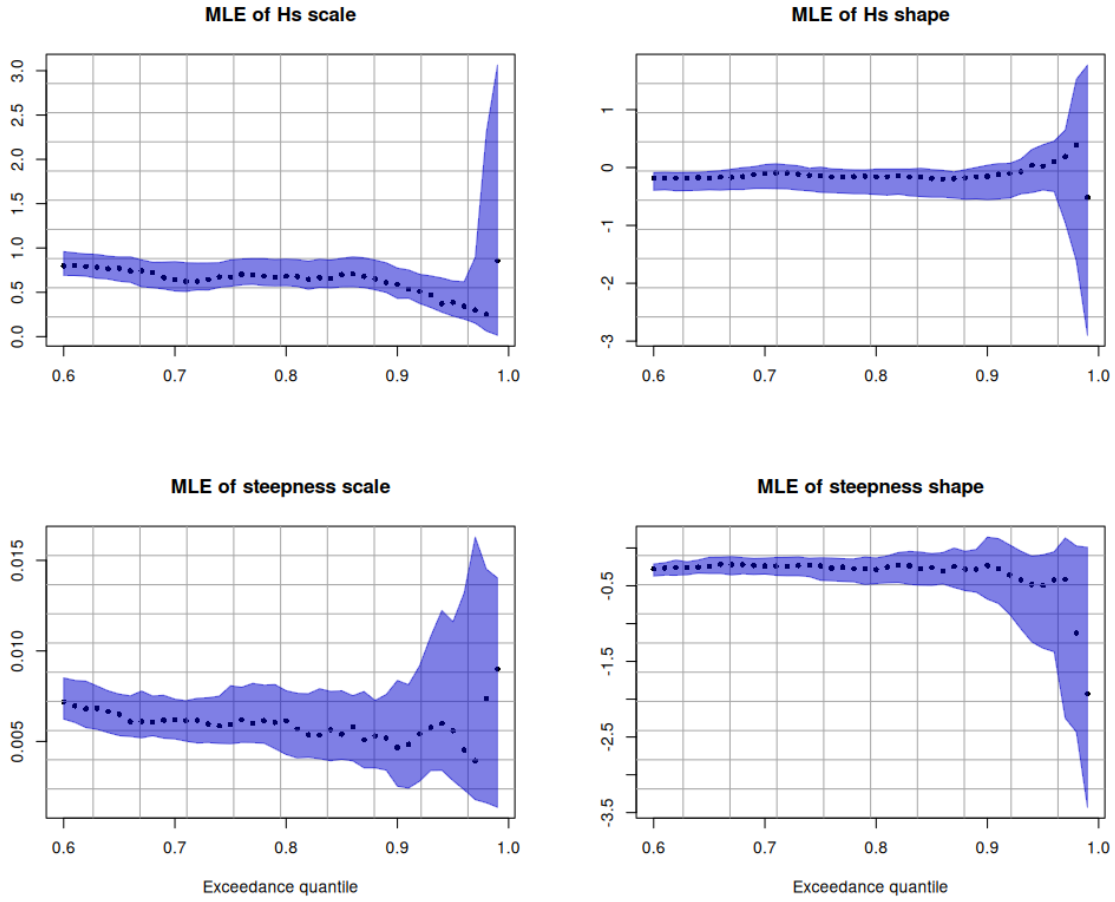


Figure 4.6.23: Threshold stability plots for estimates of the generalised Pareto scale and shape parameters σ and ξ when fitted to H_s and S_e above a range of values of the conditioning threshold $u > 0$. The estimates of σ and ξ are given on the y-axes, with the respective threshold quantile $q_u = \tilde{F}_{H_s}^{-1}(u)$ and $q_u = \tilde{F}_{S_e}^{-1}(u)$ indicated on the x-axes, for empirical distributions $\tilde{F}_{H_s}$ of H_s and $\tilde{F}_{S_e}$ of S_e. Point estimates from original Albany data are given in black, and bootstrapped 95% confidence intervals are shown as a blue region. Stability of estimates for ξ and linearity of estimates of σ above a threshold quantile q_u indicates that u is a suitable choice for GPD model threshold. Following visual analysis of the four panels we select threshold $u_1 = \tilde{F}_{H_s}^{-1}(0.7)$ and $u_2 = \tilde{F}_{S_e}^{-1}(0.7)$ for marginal modelling of H_s and S_e respectively.

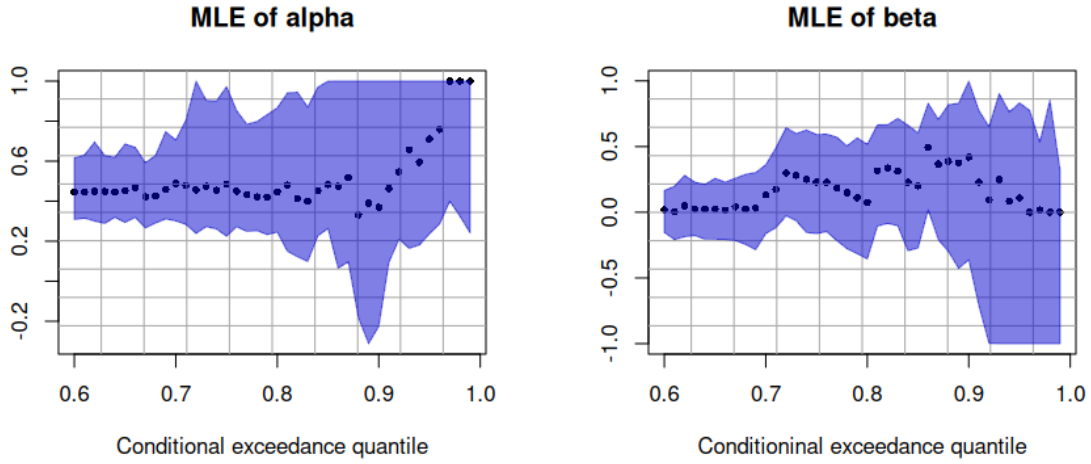


Figure 4.6.24: Threshold stability plots for estimates of the conditional extremes parameters α and β when fitted to $(H_s, S_e)|\{H_s > v\}$, for a range of values of the conditioning threshold $v > 0$. The estimates of α and β are given on the y-axes and the respective quantile $q_v = \tilde{F}_{H_s}^{-1}(u)$ on the x-axes. Point estimates from original Albany data are given in black, and bootstrapped 95% confidence intervals are shown as a blue region. Stability of parameter estimates above a threshold quantile q_v indicates that v is a suitable choice for conditional model threshold. Following visual analysis of the two panels, we select threshold $v = \tilde{F}_{H_s}^{-1}(0.6)$ for joint modelling of H_s and S_e.

The Heffernan-Tawn model assumes Gaussian residual form to facilitate model fitting only, but adopts the empirically-evaluated residuals from the model fit to the sample for inferences under the model. For this reason, assessment of model adequacy using residual diagnostics using e.g., Q-Q plots is of no relevance for the Heffernan-Tawn model.

Extreme value density estimation supporting the discussion of Section 4.4.3 of the main text.

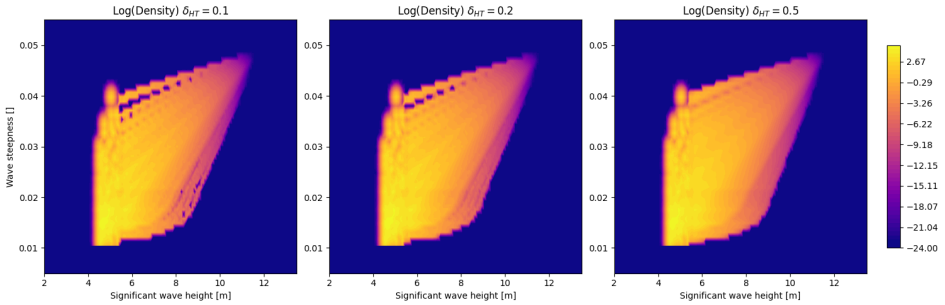


Figure 4.6.25: Sensitivity analysis of estimated joint density $\hat{f}_{\mathbf{X}}$ of $\mathbf{X} = (H_s, S_e)$ with conditional extremes smoothing parameter δ_{HT}. We aim to obtain the smallest value of δ_{HT} which eliminates 'gaps' in the extrapolated region. Following visual inspection of the three panels and Figure 4.4.11, we take $\delta_{\text{HT}} = 0.4$.

4.6.5 Non-linear harmonic response simulation

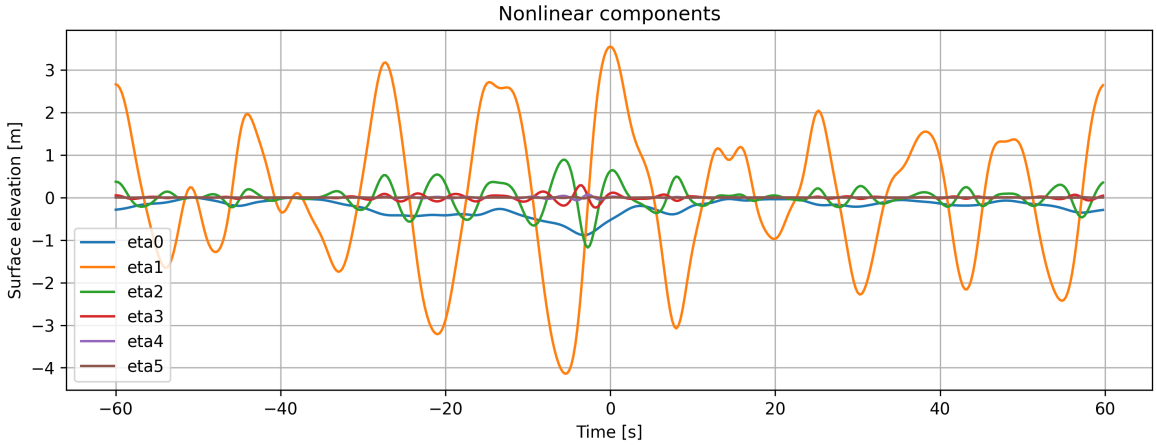


Figure 4.6.26: Harmonic signals constructed using the method of Orszaghova et al. (2025), from a linear surface elevation input. The 0-5th order harmonics are shown. To support the discussion in Section 4.4.4.

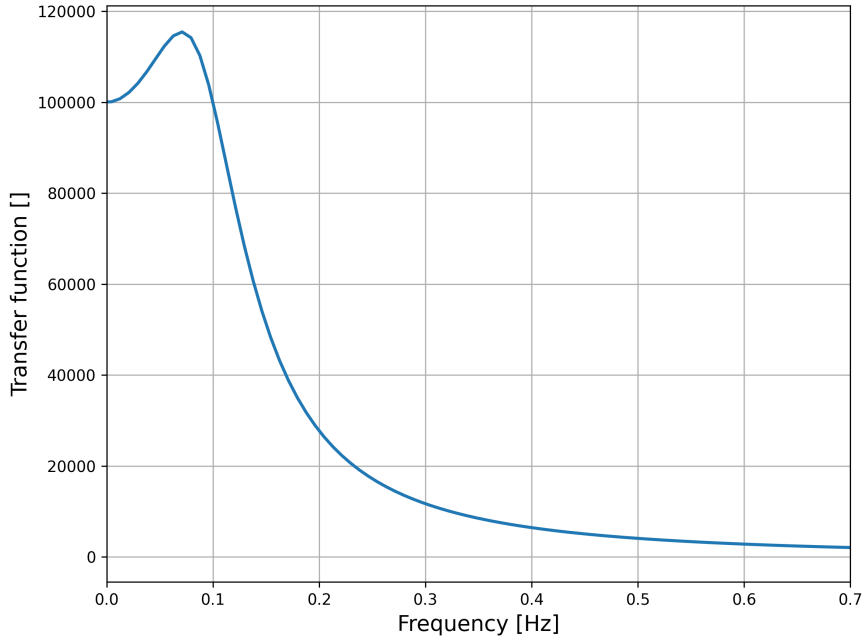


Figure 4.6.27: Transfer function of a damped harmonic oscillator (49), used in the case study of Section 4.4. The transfer function is plotted against input frequency [Hz], with resonant frequency $f_0 = 1/10$. The output of the transfer function is assumed unit-less for our case study.

4.6.6 Supplementary results to the AGE results of Section 4.4.7.

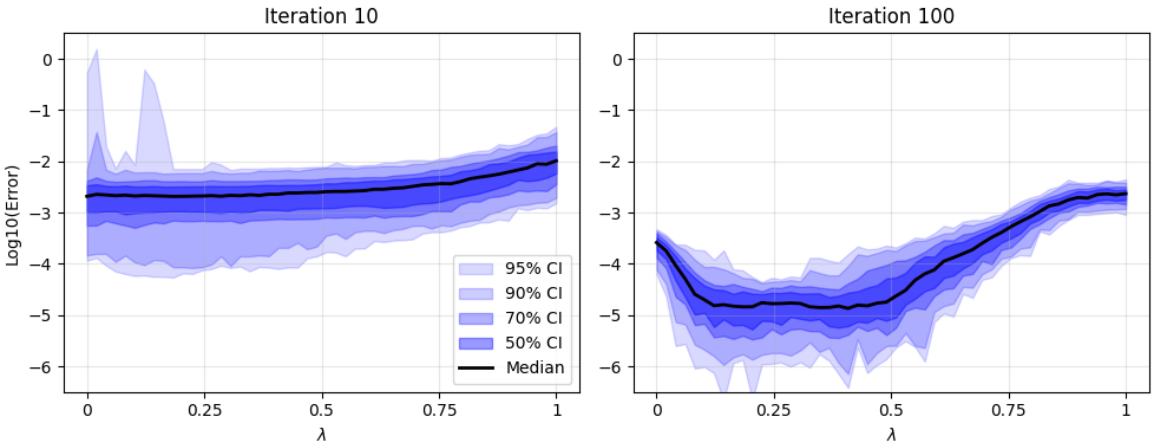


Figure 4.6.28: Log-scale absolute error Δ_{GP} of the GP probability estimate $\hat{p}_{\text{GP}}$ at specified iterations, for emulator (10) trained using $U^{(2)}$ over the range of weight $\lambda \in [0.01, 0.99]$ for the TFNV scenario. At iteration 100, the weight that minimises median error is $\lambda^* = 0.41$. To be compared with Figure 4.4.14.

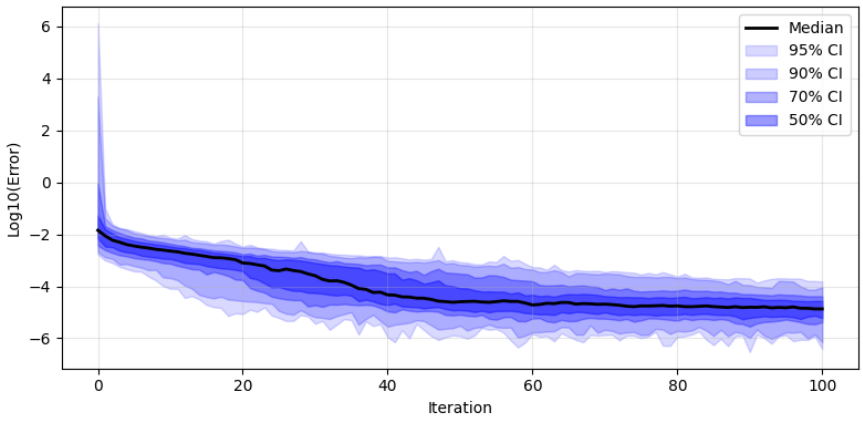


Figure 4.6.29: Distribution of log-scale absolute error Δ_{GP} in the GP probability estimate with respect to iteration, trained using $U^{(2)}$ with $\lambda = \lambda^*$. The trend in median error is indicated in black, with various confidence intervals shown in blue. To be compared with Figure 4.4.15.

Chapter 5

Automated Threshold Selection for Conditional Multivariate Extreme Value Models

5.1 Introduction

5.1.1 Background

Extreme value models for the metocean environment at a structure's location are crucial to the estimation of structural failure probability, since it is extreme storm events that will induce the largest structural response. In particular, it is essential to correctly characterise the extremal dependence structure between environment variables, as the interaction between variables can lead to larger structural response when both variables are extreme simultaneously, in comparison to when only one is marginally extreme, or in the case of $d \geq 3$ only a subset are extreme. For instance, an offshore platform may be at higher risk of failure during jointly extreme wave height or wave steepness events, than during marginal wave height and wave steepness events of the same (or greater) magnitude.

It is therefore essential to accurately estimate the probability of jointly extreme events via models which have the flexibility to capture the extremal dependence structure well. This is particularly true for environments with dimension $d \geq 3$, as different subsets of variables can exhibit both asymptotic dependence and asymptotic independence, as discussed in Section 2.2.4. This model provides a flexible framework for modelling the joint behaviour of variables, given that one is large, and is capable of capturing both asymptotic dependence and asymptotic independence within different subsets of variables.

The conditional multivariate extreme value model of Heffernan and Tawn (2004) has been successfully applied to the joint modelling of both metocean environments (Jonathan et al., 2010; Tendijck et al., 2019; Towe et al., 2019; Shooter et al., 2021b) and other environmental variables (Quinn et al., 2019; Brown, 2020; Jane et al., 2020; Shooter et al., 2021a; Bates et al., 2023).

It is not always clear when carrying out inference using extreme value models how to define an extreme observation. For instance, when using the univariate threshold approach of Davison and Smith (1990), an exceedance threshold u must be specified, such that any observed value greater than u is considered an extreme event, i.e., which is consistent with asymptotic theory of extremes, and hence is included in inference. Setting the value of u too high will result in too few observations remaining for inference, unnecessarily increasing parameter estimation variance. Conversely, setting u too low results in non-extreme observations being included in inference, biasing parameter estimates. There is an extensive literature on the optimisation of this bias-variance trade off and threshold selection for univariate extremes, see e.g., the review of Scarrott and MacDonald (2012). Of particular relevance to this thesis are the methods of Varty et al. (2021) and Murphy et al. (2025), who construct an automated approach for the selection of the univariate exceedance threshold to be used with the generalised Pareto distribution (GPD), see Davison and Smith (1990). For a given candidate value of

the univariate exceedance threshold, and a set of observations exceeding this candidate threshold, they obtain estimates of GPD scale and shape parameters via maximum likelihood inference. Using these parameter estimates, they compare theoretical quantiles derived from the GPD model (see Section 2.1.3) with empirical quantiles calculated using the set of observed exceedances, at a range of pre-specified probability levels. They then construct a metric based on the distance between the empirical and model quantiles, e.g., mean absolute or mean squared difference. Sampling uncertainty in the metric is accounted for by non-parametric bootstrapping of the observed exceedance data to obtain new sample exceedances, calculating the metric for each of these samples, and averaging the metrics across the bootstrapped samples to provide a final metric. The optimal threshold is then chosen as the member of a candidate threshold set with the smallest bootstrap averaged metric value. Wan and Davis (2019) also develop a metric which aims to assess model fit for a given threshold, and an associated threshold selection method for the bivariate setting, under the assumption of asymptotic dependence between any subsets of variables that are not completely independent.

Whilst there exist automated threshold selection methods in the univariate (Varty et al., 2021; Murphy et al., 2025) and multivariate asymptotically dependent (Wan and Davis, 2019) settings, there are currently no approaches automating the selection of the conditioning threshold for use in the method of Heffernan and Tawn (2004), or any other multivariate extreme value model capable of capturing both asymptotic dependence and asymptotic independence. We propose a method for the selection of the conditioning threshold used in the approach of Heffernan and Tawn (2004), taking inspiration from Murphy et al. (2025) and defining a metric which, when minimised, gives the threshold choice that suitably balances the bias-variance trade off.

5.1.2 Objectives and outline

In Section 5.2.1, we provide an overview of the conditional extreme value model in the bivariate case. All methodology developed here is presented in the bivariate case, with speculation on a potential extension to the multivariate setting left to the discussion. In Section 5.2.2, we construct a metric that acts as a proxy for the bias present in the conditional extremes model fit, given a particular choice of conditioning threshold. We take inspiration from the approaches of Varty et al. (2021) and Murphy et al. (2025), who develop a bias metric that assesses the model error associated with a given threshold for univariate GPD models. Our metric can be evaluated for a range of candidate thresholds and then the candidate threshold with the smallest metric value is selected for use in inference.

In Section 5.2.3, we extend this metric by adding a variance term, which penalises picking higher thresholds when lower thresholds show adequate bias reduction. A third metric is defined as the convex weighted combination of the bias and variance metrics. Again, the optimal threshold is chosen as the candidate threshold with smallest metric value. We investigate the effect of the weight on the choices of threshold yielded by our threshold selection method.

We demonstrate the advantages of the new metric method using a two stage preliminary analysis. In Section 5.3.2, we compare probability estimates found using a 'baseline' version of the conditional extreme value model (with randomly selected thresholds) to those found using the approach of Wan and Davis (2019), which includes automatic threshold selection but is not capable of capturing positively dependent variables exhibiting asymptotic independence. Additionally, the estimation method of Heffernan and Tawn (2004) (discussed in Section 2.2.5), for finding probabilities of extreme multivariate events is able to reflect the type of dependence structure inferred by the model, unlike that of Wan and Davis (2019) (discussed in Section 2.2.6), which assumes asymptotic dependence. We show that, even without automatic threshold selection, the condi-

tional extreme value model outperforms the multivariate threshold selection method of Wan and Davis (2019), thus motivating the use of conditional extremes as the preferred model. Notably, the model of Heffernan and Tawn (2004) outperforms that of Wan and Davis (2019) even in the asymptotic dependence setting, despite its threshold selection and inference methods assuming no prior knowledge of asymptotic dependence, unlike those of Wan and Davis (2019) which do assume asymptotic dependence.

In Section 5.3.3, we demonstrate the performance of the conditional extremes model utilising an automatically selected threshold versus the conditional extremes model utilising a randomly selected threshold, showing that a reduction in error in conditional extreme value model parameter estimates can be achieved when using an optimised threshold. We conclude with a discussion of the implications of these initial findings and of our ideas for future development.

5.2 Methodology

5.2.1 Bivariate conditional extreme value model

We first describe the model of Heffernan and Tawn (2004) for the behaviour of a bivariate random variable, conditional on one component being large. Whilst the model is developed and presented by Heffernan and Tawn (2004) for arbitrary dimension d, we present it in the case of $d = 2$ since this is the setting in which our threshold selection method is currently developed. For a detailed discussion of the model in the $d > 2$ setting, see the literature review of Section 2.2.5. Key notation given in Section 2.2.5 is re-defined here in the $d = 2$ setting, with detailed justification and background to the model being left to Section 2.2.5.

Consider a bivariate random variable (X, Y) on $\mathbb{R}^2$ with Laplace margins, such that $X \sim F_L$ and $Y \sim F_L$ for standard Laplace distribution function F_L. Heffernan and Tawn (2004) and Keef et al. (2013) show that, for a wide range of parametric copula

examples, there exist parameters $\alpha \in [-1, 1]$ and $\beta \in (-\infty, 1]$, and residual random variable $Z = (Y - \alpha X)/X^{\beta}$, such that

$$\lim_{u \to \infty} \Pr\left(Z < z, X - u > x \mid X > u\right) = e^{-x} G(z), \tag{5.2.1}$$

for $x > 0$, $z \in \mathbb{R}$, and non-degenerate distribution G. That is, the joint distribution of X and Z, conditional on $X > u$, converges to the product of two independent marginals, where the marginal distribution of $X' = (X - u)|(X > u)$ for $u > 0$ is standard exponential, and the limiting marginal distribution of Z is G. In practice, this model is employed by choosing some finite threshold $u > 0$ of X, large enough such that limit (5.2.1) approximately holds, meaning we can write $X' = (X - u) \sim \text{Exp}(1)$ when $X > u$, and

$$Y \mid \{X = x\} = \alpha_u x + x^{\beta_u} Z_u, \quad \text{for} \quad x > u, \tag{5.2.2}$$

where Z_u is non-degenerate and Z_u is independent of X', and the subscript in the residual variable Z_u and the parameters α_u and β_u indicate their dependence on the choice of finite threshold u. We refer to u as the conditional exceedance threshold. Given a set of observations for which $X > u$, and by assuming a Gaussian distribution for the residual Z_u, maximum likelihood estimates $\hat{\alpha}_u$ and $\hat{\beta}_u$ for parameters α_u and β_u can be obtained. Event prediction and joint probability estimation can then be carried out using estimates $\hat{\alpha}_u$ and $\hat{\beta}_u$, along with an empirical model for the distribution G of residual Z_u. Specifically, the model $\hat{G}$ for G is calculated empirically from the observed values of residual

$$Z_u = \frac{Y - \hat{\alpha}_u X}{X^{\hat{\beta}_u}}, \tag{5.2.3}$$

for sampled values of $(X, Y)|(X > u)$. It has been shown by Winter and Tawn (2017) that estimation can be improved via the use of kernel density estimation for Z_u (Silver-

man, 1986), instead of the empirical distribution, see Section 2.2.5 for details. In this section, and the remainder of this chapter, we present our method for $(X, Y)|X > u$. All approaches can be similarly applied to the case of $(X, Y)|Y > u$, with appropriate change of notation.

5.2.2 Comparing bias across threshold candidates

We propose a method to automatically select the conditional exceedance threshold u in model (5.2.2), by defining a metric that is minimised for the best value of the threshold u. Our metric acts as a proxy for the bias present in the model fitted to observations above a particular value of u, and is defined using a measure of the uniformity of a sample obtained via transformation of $(X, Y)|(X > u)$ to a uniform point process with random variables (V_u, W_u), where the transformation is carried out using model (5.2.2) fitted to observations above the candidate value of u. If the sample of variables (V_u, W_u) is sufficiently close to being jointly uniform, then the model is a deemed to be good fit to the data and therefore the candidate value of u is appropriate, though potentially a more optimal threshold may be lower than such a threshold u. Conversely, if the sample of (V_u, W_u) is clearly non-jointly uniform, this suggests the model (5.2.2) is mis-specified for this candidate value of u, and so a higher threshold value should be chosen.

The transformation $(X, Y)|(X > u) \mapsto (V_u, W_u)$ for a particular value of u is carried out as follows. Consider parameter estimates $\hat{\alpha}_u$ and $\hat{\beta}_u$, obtained from fitting the regression (5.2.2) to the subset of observations $\{(x_i, y_i)\}_{i=1}^{n}$ such that $X > u$ (as described in Section 2.2.5). The observed excesses are given by

$$x'_j = x_j - u, \quad \text{for } j = 1, \ldots, n_u,$$

where j indexes over the n_u observations where the observed value of X exceeds u. Let $\{y_j\}_{j=1}^{n_u}$ denote the corresponding observed values of Y for which $X > u$. Assuming

that u is sufficiently large for the limit in (5.2.1) to approximately hold, the distribution $F_{X'}$ of observed values $\{x'_j\}_{j=1}^{n_u}$ of the excess X' is taken to be the standard exponential distribution. Additionally, following model (5.2.2), the conditional distribution $F_{Y|X}$ of $Y|(X = x)$, for $x > u$, can be estimated (independently from $F_{X'}$) with

$$\hat{F}_{Y|X}(y|x) = \hat{G}\left(\frac{y - \hat{\alpha}_u x}{x^{\hat{\beta}_u}}\right), \quad y \in \mathbb{R}, \tag{5.2.4}$$

where $\hat{G}$ is an estimate of the distribution G, which we obtain empirically from observed values of residual (5.2.3). In this case, the transformed observations

$$(v_{u,j}, w_{u,j}) = \left[1 - \exp(-x'_j), \hat{F}_{Y|X}(y_j|\{x'_j + u\})\right], \quad j = 1, \ldots, n_u, \tag{5.2.5}$$

with $\hat{F}_{Y|X}$ as in (5.2.4), should be a sample of bivariate jointly uniform random variables (V_u, W_u) (Rosenblatt, 1952). The performance of this transformation relies on: (a) accuracy of parameter estimates $\hat{\alpha}_u$ and $\hat{\beta}_u$; (b) suitability of the empirical model $\hat{G}$ of G; and (c) the independence of sampled values of residual Z_u and excesses X'. Testing for joint uniformity in the sample given by transformation (5.2.5) thus assesses the cumulative bias in parameter estimates $\hat{\alpha}_u$ and $\hat{\beta}_u$, the fit of the estimated distribution $\hat{G}$ to the sample residuals, as well as the validity of the assumption of independence between the excesses and residuals expressed in model (5.2.2).

In order to design an automated threshold selection approach based on assessing the joint uniformity of the sample (5.2.5), we require a test for uniformity that is automated (i.e, not manual inspection) and capable of capturing any type of departure from a uniform point process. This second requirement makes the use of standard measures such as Spearman and Kendall correlations unsuitable, as these only measure monotonic association. We therefore assess the uniformity of the point process of sample $\{(v_{u,j}, w_{u,j})\}_{j=1}^{n_u}$ in expression (5.2.5) with Ripley's K-function (see Ripley 1976). We

achieve this using the variance-standardised L_u function

$$L_u(t) = \left(\frac{K_u(t)}{\pi} \right)^{1/2},$$

where K_u (Ripley, 1988) is given by

$$K_u(t) = \nu^{-1} \sum_{i \neq j} \frac{I\left(d_{ij} < t\right)}{n_u}, \quad t > 0,$$

for a set of n_u data points on some window $A \subset \mathbb{R}^2$ (in our case $A = [0,1]^2$), with average intensity $\nu = n_u/|A|$, where d_{ij} is the Euclidean distance between points i and j, and I is the identity function. The function $K_u(t)$ provides the average number of observations lying within a window of radius t surrounding a point in the set. Under uniformity, $K_u(t) = \pi t^2$ for all $t > 0$.

When calculating $K_u(t)$ over any fixed window A, care must be taken when handling edge effects caused by points within t of the window boundary. If these effects are not accounted for, the average number of observations around these boundary points will be incorrectly evaluated as 'non-uniform'. To resolve this issue, we employ the isotropic edge effect correction of Ripley (1988).

It is shown by Ripley (1988) that $\mathbb{E}\left[L_u(t)\right] = t$, for all $t \geq 0$, for jointly uniform data. The uniformity of a sample can thus be assessed by inspection of $L_u(t)$ against t; if points lie close to the line of equality of $L_u(t) = t$ the sample is considered to be jointly uniform. Before doing so, however, we account for the effect of varying the number of threshold exceedances as the threshold changes by scaling by n_u, defining the $\tilde{L}_u(t)$ function

$$\tilde{L}_u(t) = n_u(L_u(t) - t), \tag{5.2.6}$$

that is, the n_u-standardised difference between the L_u function and radius t. From $\tilde{L}_u$, we can quantify the uniformity of the sample (5.2.5) for a given threshold u using the

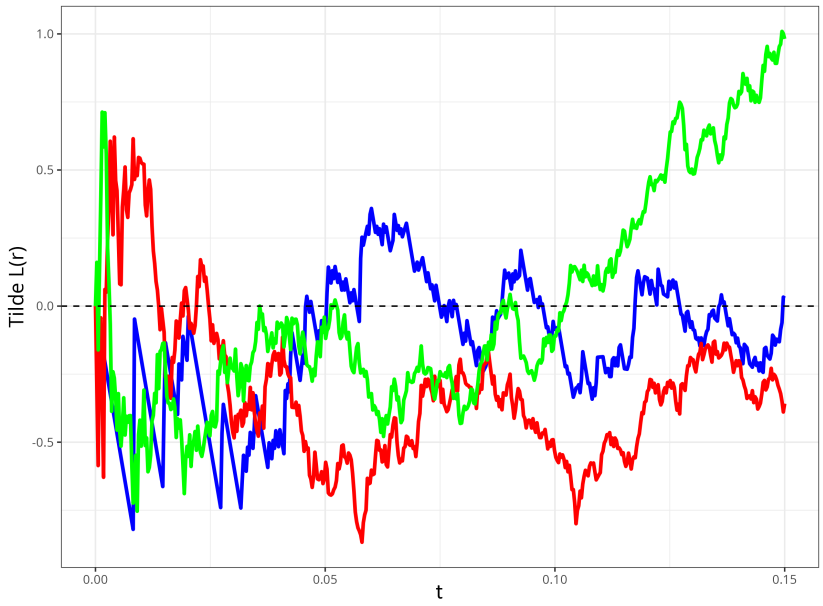


Figure 5.2.1: The n_u-standardised difference between the L_u function and radius t, given by (5.2.6), calculated for uniform samples of size 10^3 (red), 10^4 (blue) and 10^5 (green).

metric

$$T_u = \frac{n_u}{k} \sum_{i=1}^{k} \tilde{L}_u(t) = \frac{n_u}{k} \sum_{i=1}^{k} |L_u(t_i) - t_i|, \tag{5.2.7}$$

where n_u is the size of the sample with $X > u$, for a set of $k > 1$ radii $(t_1, \ldots, t_k)$. That is, we calculate mean absolute error, scaled by n_u, between the locus of points $\{t_i, L(t_i)\}_{i=1}^k$ and the line $y = x$ up to some radius t_k. Initial experimentation suggested it is suitable to take $\{t_i\}_{i=1}^k$ equally spaced with $t_1 = 0$ and $t_k = 0.15$, and $k = 50$, with little sensitivity to k within a sensible range around this value. We demonstrate the effect of scaling by n_u in Figure 5.2.1, which shows $\tilde{L}_u(t)$ against t for uniform samples of size 10^3, 10^4 and 10^5. The figure shows the scaling by n_u has ensured the variance of $\tilde{L}_u$ is stable over n_u, since all curves are of the same order of magnitude despite n_u varying by a factor of 100.

A large value of the metric $T_u \geq 0$ indicates clustering of points within $[0, 1]^2$, i.e., exhibiting behaviour that is a departure from a jointly uniform sample, and a small value of the metric indicates higher dispersion of points, which is more consistent with

the sample being from jointly uniformly distributed variables. A truly joint uniform sample would have no clustering, and so the metric thus provides a measure of the average deviation of the density of points from what would be expected for a true jointly uniform sample. Therefore, for a given set of candidate values of conditional exceedance threshold u, the candidate value with the smallest corresponding metric value of T_u should be chosen to ensure the smallest bias.

To capture sampling uncertainty (necessary because a user of this method is likely to have only a single sample of few joint extremes), in the values of metric (5.2.7) we employ a bootstrapping procedure, as with the approach of Murphy et al. (2025). For a candidate value of u and corresponding set of observed exceedances $\{(x'_j, y_j)\}_{j=1}^{n_u}$, the bootstrapping is carried out by:

(a) Fitting the conditional extremes model to $\{(x'_j, y_j)\}_{j=1}^{n_u}$, obtaining conditional extremes parameter estimate $\hat{\mathbf{\Theta}}_u^{(0)} = (\hat{\alpha}_u^{(0)}, \hat{\beta}_u^{(0)})$, and empirical estimate $\hat{G}^{(0)}$ of the residual distribution function G;

(b) Simulating some $B > 1$ new sets of exceedance observations of size n_u using the conditional extremes model with parameter $\hat{\mathbf{\Theta}}_u^{(0)}$ and residual distribution $\hat{G}^{(0)}$;

(c) Fitting the conditional extremes model to each of these B bootstrapped exceedance samples, obtaining parameter estimates $\hat{\mathbf{\Theta}}_u^{(1)}, \dots \hat{\mathbf{\Theta}}_u^{(B)}$ and empirical residual distributions $\hat{G}^{(1)}, \dots, \hat{G}^{(B)}$;

(d) Using these B parameter and residual distribution estimates to transform the *original sample* $\{(x'_j, y_j)\}_{j=1}^{n_u}$ to realisations of (V_u, W_u) via (5.2.5);

(e) Calculating the metric (5.2.7) from these B samples on the (V_u, W_u) space and obtaining $T_u^{(1)}, \dots, T_u^{(B)}$; and

(f) Averaging over these metrics to obtain the bootstrapped metric

$$\bar{T}_u = \sum_{b=1}^{B} T_u^{(b)}. \tag{5.2.8}$$

This provides a more robust estimate of the metric value of a particular candidate value of threshold u than when using a single sample. In step (d), we transform the original set of exceedance observations, not the parametric bootstrap exceedances since these will automatically be standard exponential and the simulated residuals will be distributed according to the empirical residual distribution, making assessment of the exponential distribution and empirical residual distribution as models meaningless.

The upper panels of Figure 5.2.2 show two illustrative datasets of size $n = 10^5$, used here to demonstrate the performance of the metric (5.2.8). The upper left panel shows a dataset generated from the conditional extreme value model under the following scheme. The the 'true' threshold is set at $u_{\mathrm{T}} = F_L^{-1}(0.8)$ for standard Laplace distribution function F_L. Points for which $X < u_{\mathrm{T}}$ are drawn from a conditional extreme value model with parameters $\alpha = 1$ and $\beta = 0$, and Student's t, $df = 1$, residual distribution $Z \sim \mathrm{Stt}(1)$[1]. Points for which $X > u_{\mathrm{T}}$ are drawn from a conditional extreme value model with parameters $\alpha = 1$ and $\beta = 0$, and standard Gaussian residual distribution $Z \sim N(0, 1)$. We simulate from this model by first sampling $X \sim \mathrm{Exp}(1)$. If $X < u$, then we draw $Y = X + Z$ for $Z \sim \mathrm{Stt}(1)$. Conversely, if $X > u$ then we draw $Y = X + Z$ with $Z \sim N(0, 1)$. The resulting sample provides an idealised dataset on which to test the method's ability to select a threshold when there is a known 'true' threshold quantile, in the asymptotically dependent setting. The upper right panel shows the Gaussian dataset with $\rho = 0.5$, which provides a more challenging case of asymptotic independence with no known threshold.

The lower panels of Figure 5.2.2 show $\bar{T}_u$ against u for both the idealised case with

[1]We note that this level of tail behaviour for Z is not feasible for Laplace distributed Y, however, this example is purely for illustrative purposes and the methods do not exploit that $Y \sim$ Laplace.

known true threshold $u_{\mathrm{T}} = F_L^{-1}(0.8)$, and the Gaussian case with unknown threshold. The blue regions indicate 95% bootstrapped confidence intervals for the metric. In the lower left panel, corresponding to the idealised case, we see a sudden drop in metric value for candidates larger than u_{T}. When fitting the method to thresholds lower than u_{T}, the change in residual distribution results in a poor fit of regression (5.2.2), invalidating transformation (5.2.5) and resulting in a non-jointly uniform sample, and therefore a large value of metric (5.2.8). When considering threshold values beyond u_{T}, there is no change in residual distribution in the sample since all residual values are drawn from $Z \sim N(0, 1)$, and so the sample obtained by transformation (5.2.5) is jointly uniform for all $u \geq u_{\mathrm{T}}$, and thereby the value of metric (5.2.8) will be small for this range of u.

The effects of threshold choice on the fitting of regression (5.2.2), and the resulting metric values in the idealised case are shown in Figures 5.2.3, 5.2.4 and 5.2.5. Figure 5.2.3 shows Heffernan and Tawn (2004) parameter estimates and empirical residual distributions when fitting to excesses of the idealised case above the 0.6 marginal Laplace quantile (upper panels) and 0.95 marginal Laplace quantile (lower panels), over 50 replicates, representing cases with a threshold choice above and below the true threshold of $u_{\mathrm{T}} = F_L^{-1}(0.8)$ respectively. When using either the lower or higher threshold, the distributions of both parameter estimates are sensible, with $\hat{\alpha}_u$ having a mode around the true value of $\alpha = 1$ and $\hat{\beta}_u$ a mode around the true value of $\beta = 0$; this is as we would expect from our construction of the sample. However, when using a threshold value lower than the true threshold u_{T}, there is a mixture of residual distributions caused by the combination of Gaussian and Student's t used in the sample generation. This can be seen in the panels on the right of Figure 5.2.3 where, when fitting to excesses above the 0.6 marginal Laplace quantile, the sampled residuals are no longer standard Gaussian, as they are when fitting to excesses above the 0.95 marginal Laplace quantile.

Additionally, Figure 5.2.4 shows two examples of sample (5.2.5) obtained at candidate thresholds corresponding to marginal Laplace 0.6 (left) and 0.95 (right) quantiles, for the idealised cases with known true threshold at the 0.8 marginal Laplace quantile. The sample obtained using the threshold below u_{T} is clearly non-uniform, whereas the sample obtained using the threshold above u_{T} appears uniform. This visual assessment is reinforced by Figure 5.2.5, which shows the n_u-standardised difference between the $L_u(t)$ function and radius t, given by expression (5.2.6), calculated for the above samples. As expected, the $\tilde{L}_u$ function lies closer to the line $\tilde{L}_u(t) = 0$ (or equivalently $L_u(t) = t$) when calculated using the sample corresponding to the higher threshold than when calculated using the sample corresponding to the lower threshold.

The impact on the metric value can be seen in the lower left panel of Figure 5.2.2, with a clear elbow present at the conditioning quantile value of 0.8. Additionally, the bootstrapped 95% confidence interval for the metric is narrow, since parameter estimates have low variability because the data used for inference comes from the conditional extreme value model itself, and so regression (5.2.2) correctly captures the properties of the data.

Conversely, the plot of $\bar{T}_u$ against u for the Gaussian data does not show a clear elbow, and therefore no clear choice of threshold u. This is because the reduction in bias in the conditional model fit caused by finite approximation of the limit in (5.2.1) is gradual, with no clear transition point from biased to unbiased such as that seen in the idealised case. Whilst the metric correctly reflects that the bias decreases with an increase in u, there is no natural way to choose the threshold value from this figure, even after accounting for the bootstrapped evaluated uncertainties.

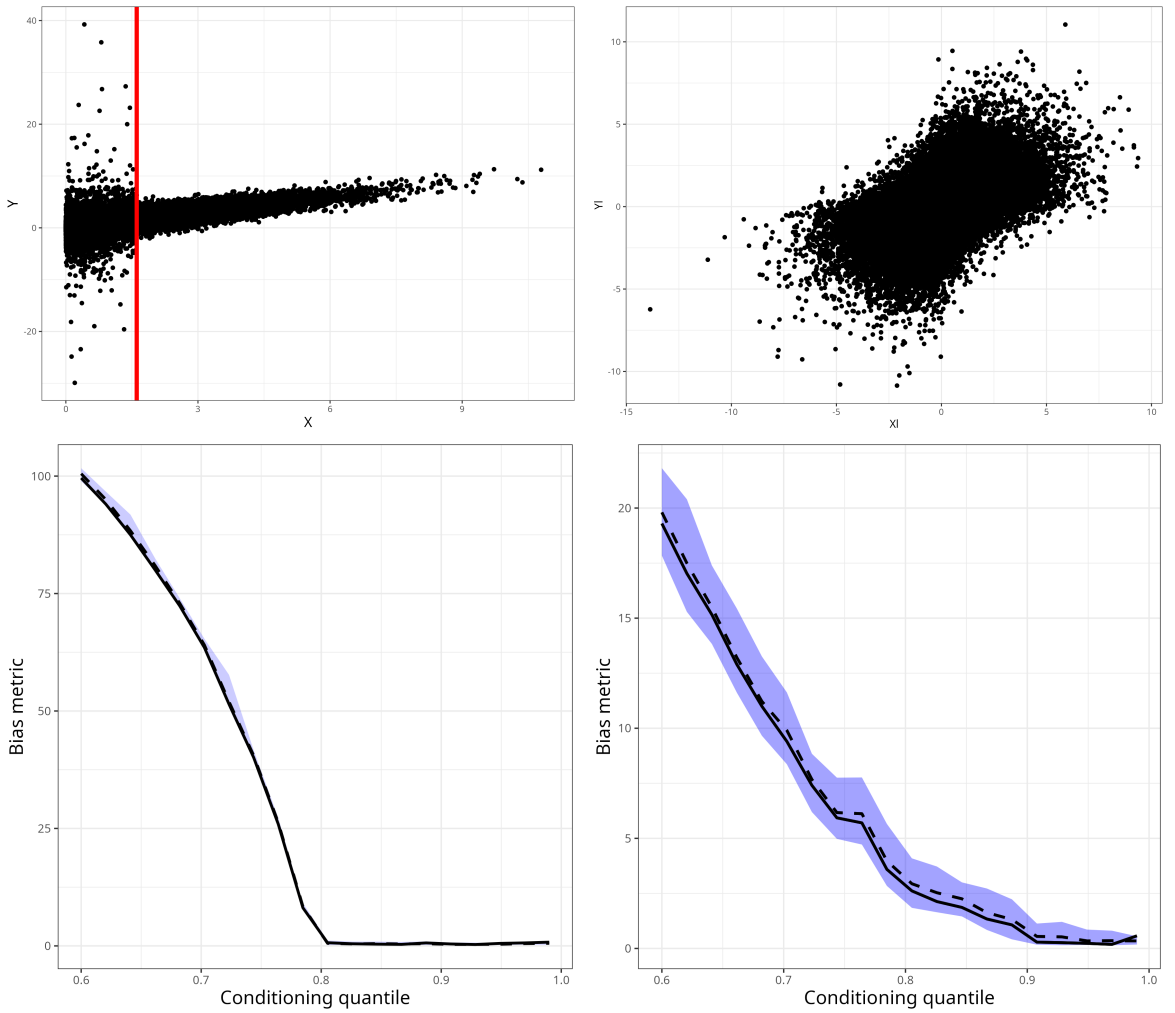


Figure 5.2.2: Example data sets (top) from the conditional model (left) and bivariate Gaussian copula (right). Associated plots of $\tilde{T}_u$ against u (bottom) show the metric calculated from the original sample (full line), the median value of the metric calculated from bootstrapped samples (dashed line) and 95% bootstrapped confidence intervals (blue region).

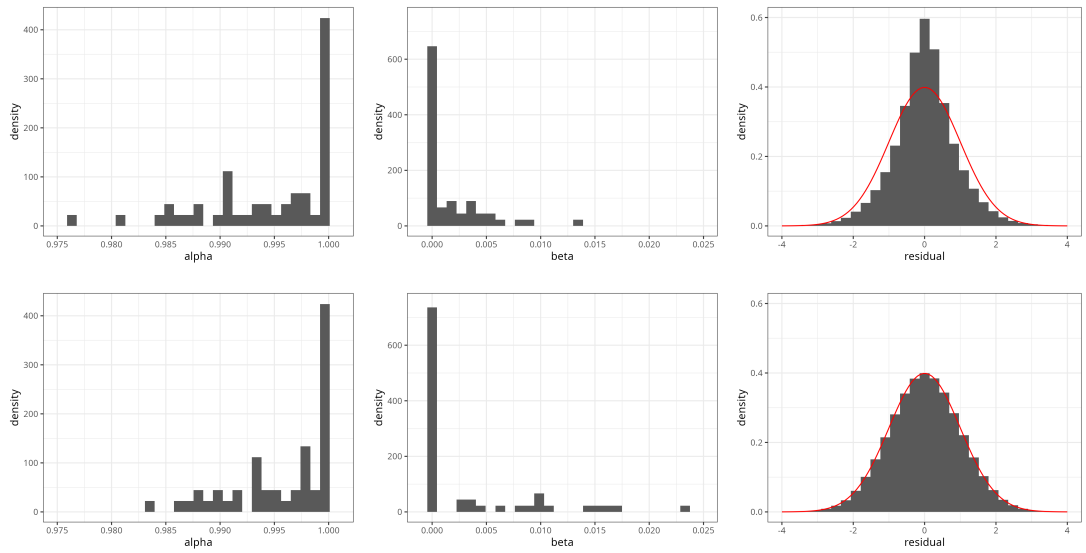


Figure 5.2.3: Heffernan and Tawn (2004) parameter estimates and empirical residual distributions when fitting to excesses of the idealised case above the 0.6 marginal Laplace quantile (upper panels) and 0.95 marginal Laplace quantile (lower panels), over 50 replicates. The red curve overlaid on residual histograms shows the standard Gaussian density.

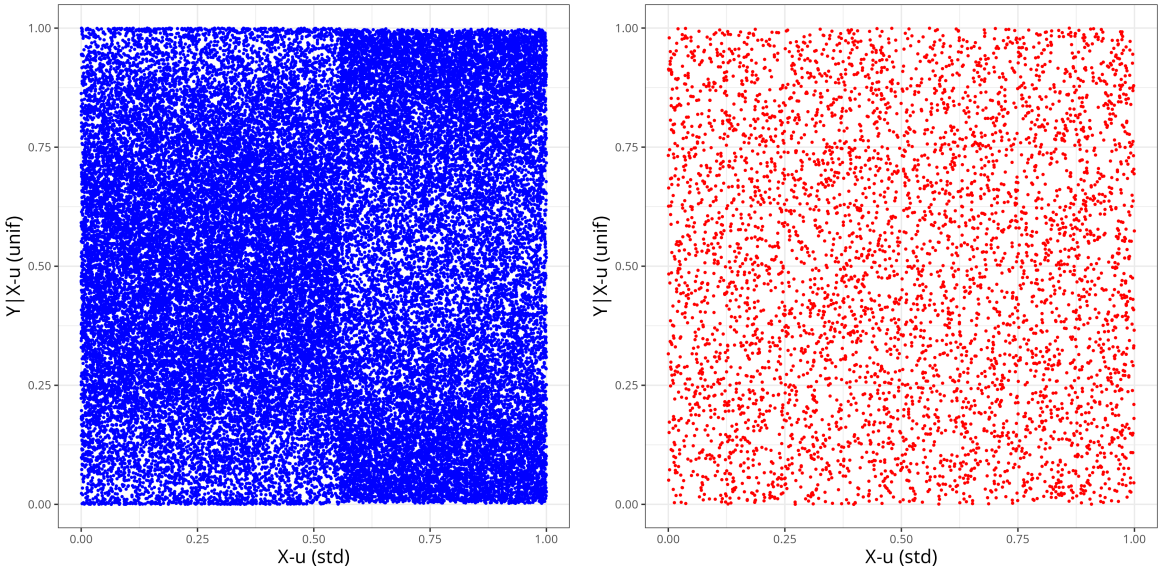


Figure 5.2.4: Examples of sample (5.2.5) obtained at candidate thresholds corresponding to marginal Laplace 0.6 (left) and 0.95 (right) quantiles, for the idealised cases with known true threshold at the 0.8 marginal Laplace quantile.

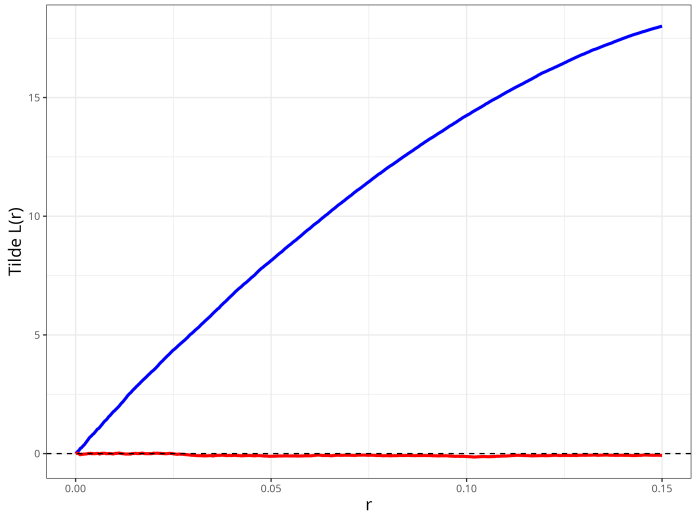


Figure 5.2.5: The $\tilde{L}_u(t)$ function given by expression (5.2.6), calculated for the samples shown in Figure 5.2.4 and plotted in red (higher threshold) and blue (lower threshold). The dashed line shows the x-axis, where the plot of $\tilde{L}_u(t)$ against t should lie in the case of perfect uniformity.

5.2.3 Variance weighted measure for threshold selection

The metric (5.2.7) provides a measure of quality of fit of the conditional extremes model; however, it does not account for the increase in parameter variance seen at higher thresholds. As seen in Figure 5.2.2, the metric will therefore always favour the highest candidate threshold, as this will provide the best approximation to the limiting behaviour in (5.2.1). To allow the metric to consider lower thresholds, we propose the variance-weighted metric $T_u(\lambda)$ defined by

$$T_u(\lambda) = \lambda \log(|\hat{\Sigma}_{(\hat{\alpha}_u,\hat{\beta}_u)}|) + (1-\lambda)\log(\bar{T}_u), \tag{5.2.9}$$

where $|\hat{\Sigma}_{(\hat{\alpha}_u,\hat{\beta}_u)}|$ is the determinant of the estimated covariance matrix $\hat{\Sigma}_{(\hat{\alpha}_u,\hat{\beta}_u)}$ of the conditional extremes model parameter estimates $(\hat{\alpha}_u, \hat{\beta}_u)$, i.e., from a fit for $X > u$, and $\lambda \in [0, 1]$ is a bias-variance weighting parameter. We choose the determinant of the covariance matrix as the measure of model uncertainty since it directly captures the uncertainty in parameter estimates and is also easily estimated obtained from the set of

bootstrapped parameter estimates $(\hat{\alpha}_u^{(1)}, \hat{\beta}_u^{(1)}), \ldots, (\hat{\alpha}_u^{(B)}, \hat{\beta}_u^{(B)})$. We take the logarithm of each term to ensure sensible optimal values of the weight parameter, and since using the standard linear combination is not appropriate here as our bias term is a metric of bias, not bias itself.

The value of λ determines the balance of metric (5.2.9) between penalising bias and penalising variance. If λ is close to 0, then more weight is placed on the bias term $\bar{T}_u$, meaning the metric will favour model fits resulting in 'more jointly uniform' samples (5.2.5); in this case higher thresholds tend to be chosen since these provide better approximations to the limiting behaviour in (5.2.1). For λ close to 1, the variance term $|\hat{\Sigma}_{(\hat{\alpha}_u, \hat{\beta}_u)}|$ will have higher weighting, meaning the metric will favour models with smaller parameter variability; in this case lower thresholds tend to be chosen, since these provide more observations and thus minimise inference uncertainty.

Figure 5.2.6 shows the metric $T_u(\lambda)$ for the idealised and Gaussian dataset discussed in Section 5.2.2, plotted across a range of candidate threshold quantiles in the range $[0.6, 0.99]$. We compare $\lambda = 0.1$ and $\lambda = 0.9$. In the case $\lambda = 0.1$, the weighted metric that correctly selects the true threshold in the idealised sample case, and provides a sensible choice of threshold quantile in the Gaussian case. In the case $\lambda = 0.9$, we see that choosing a value of λ too large can over-penalise the model based on its parameter variance, leading to a metric that chooses the lowest supplied candidate threshold in order to utilise the most data. A preliminary investigation into the sensitivity of chosen threshold and resulting parameter estimates with changes in λ is given in Section 5.3.3.

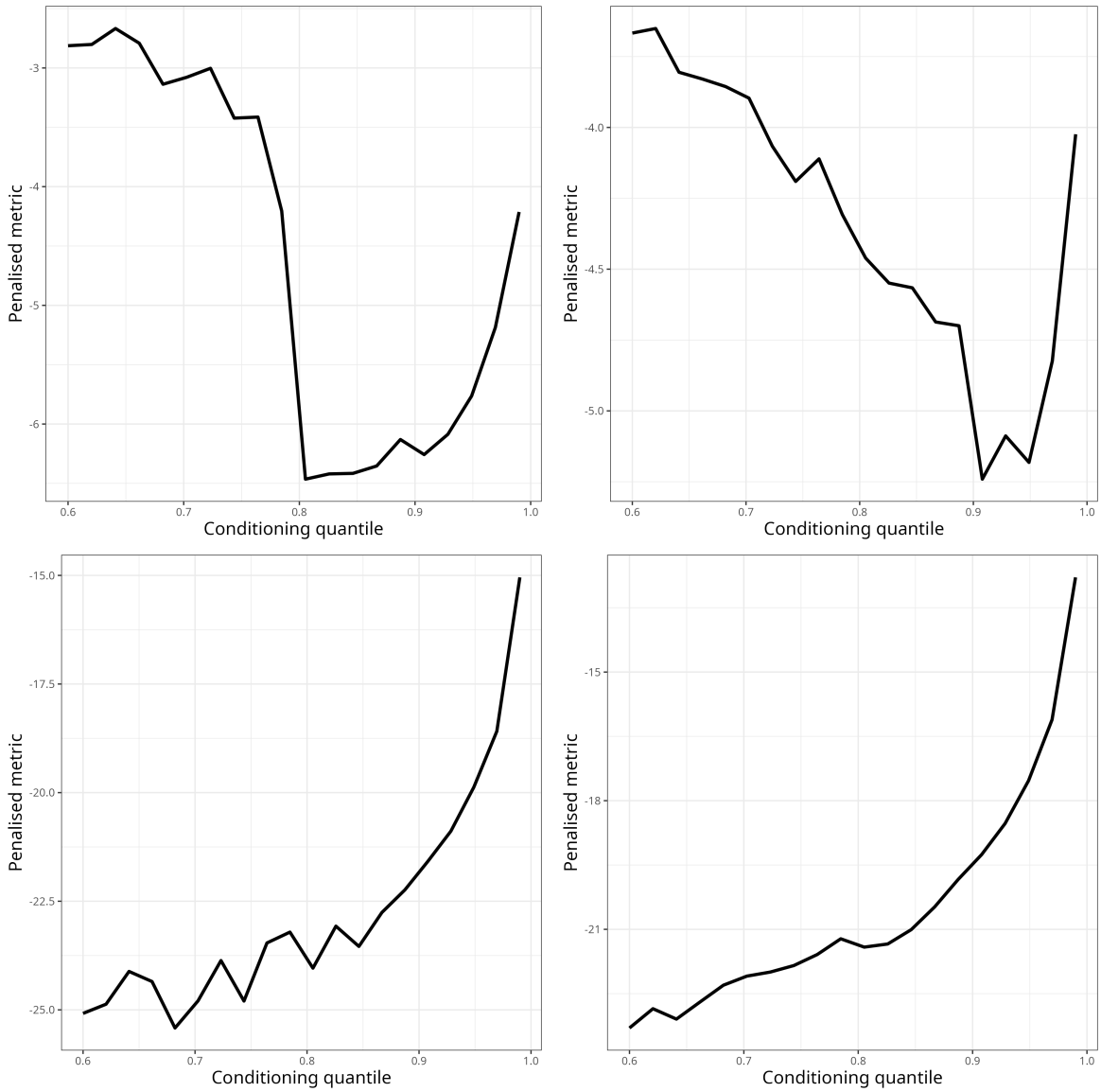


Figure 5.2.6: Metric $T_u(\lambda)$ for the idealised (left) and Gaussian (right) dataset discussed in Section 5.2.2, for $\lambda = 0.1$ (top) and $\lambda = 0.9$ (bottom). Plots are across a range of candidate threshold quantiles in the range $[0.6, 0.99]$.

5.3 Simulation study

5.3.1 Overview

We conduct a preliminary simulation study with the aim to demonstrate the potential of this approach and motivate future development. The study is divided into two parts. In Section 5.3.2, we compare estimates of probabilities of extreme events that are obtained from the conditional extreme value model of Heffernan and Tawn (2004) with those obtained from the approach of Wan and Davis (2019). The results of this section show that the model of Heffernan and Tawn (2004) provides a smaller error in probability estimates than the approach of Wan and Davis (2019), despite the latter's use of an automated threshold selection method. In Section 5.3.3, we investigate the sensitivity of the error in conditional extreme value model parameter estimates relative to the choice of weight parameter λ, in representative asymptotically dependent and asymptotically independent cases with known limiting parameter values, and compare these errors to those obtained using randomly chosen exceedance thresholds. We show that, for a suitable choice of λ, use of our method over a randomly selected threshold reduces the error in parameter estimates for both test cases.

5.3.2 Comparison with an existing multivariate threshold selection method

The method of Wan and Davis (2019), reviewed in Section 2.2.6, provides an approach for the automatic selection of an exceedance threshold in the asymptotically dependent setting. They work with Fréchet marginal random variables i.e, $(X_F, Y_F) \in \mathbb{R}^2_+$, with $F_X(x) = F_Y(x) = e^{-1/x}$, and consider the radial-angular variables

$$R = X_F + Y_F \quad \text{and} \quad W = X_F/(X_F + Y_F).$$

Approaches developed for multivariate block maxima and multivariate regular variation by the likes of de Haan and Resnick (1977), discussed in Section 2.2.2, can be used to model the joint extremal dependence structure of (X_F, Y_F) in the region $\{\mathbb{R}^2_+ : R > r\}$, for a finite threshold $r > 0$, where, for a sufficiently large r, the variables R and W can be assumed to be independent, see Section 2.2.6 for details. Given a choice of r, joint survivor probabilities can be estimated by again assuming an asymptotically dependent dependence structure, and following the probability estimation procedure reviewed in Section 2.2.6. Wan and Davis (2019) test this independence assumption using their own metric, and construct a corresponding method for the automatic selection of the finite threshold r.

The approach of Heffernan and Tawn (2004) differs from that of Wan and Davis (2019) in the sense that it does not assume an asymptotically dependent dependence structure, allowing the dependence structure inferred by fitting regression (5.2.2) to observed exceedances to inform both the probability estimation steps, and our threshold selection method based on this approach; see Section 2.2.5 for details.

We compare probability estimates obtained using the approach of Wan and Davis (2019), utilising their automated threshold selection approach and corresponding probability estimation framework, to those obtained using the conditional extreme value model of Heffernan and Tawn (2004) and their corresponding probability estimation framework, utilising randomly selected values of exceedance threshold u in the range [0.6, 0.99]. We select a randomly chosen threshold for each sample to capture the effect of the full range of sensible threshold choices.

We carry out the comparison of methods by using bivariate samples of size $n = 10^5$ drawn from two distributions; these are the bivariate logistic with dependence parameter $\theta = 0.5$, representing an asymptotically dependent case, and the bivariate Gaussian with $\rho = 0.5$, representing an asymptotically independent case. Example samples shown on Laplace margins in Figure 5.3.7. For each dependence structure,

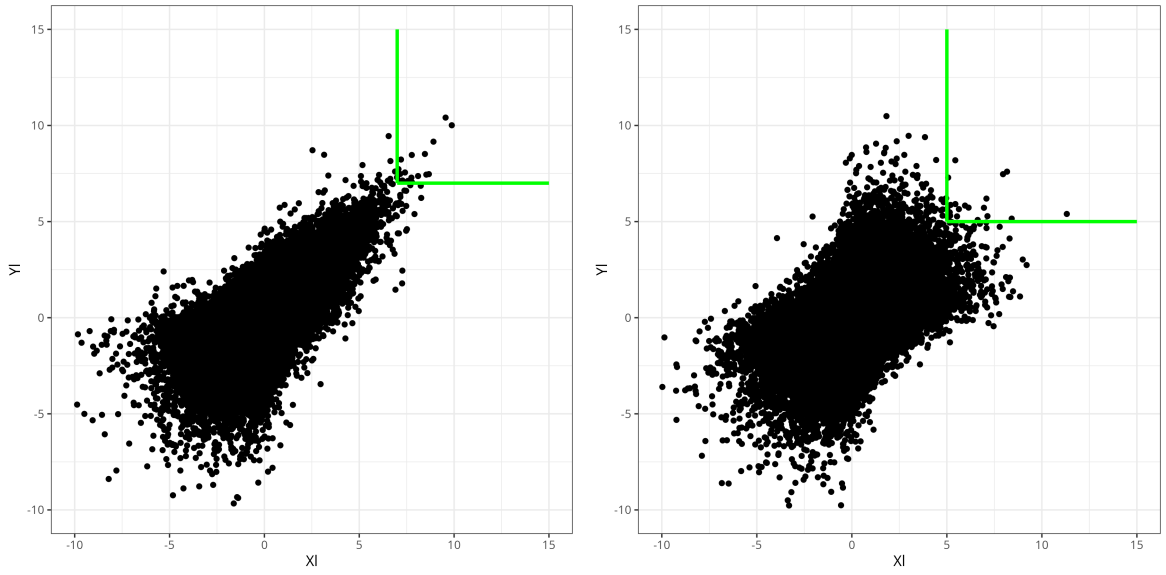


Figure 5.3.7: Target exceedance sets for example Logistic (left) sample and Gaussian (right) samples of size $n = 10^5$, providing representative asymptotic dependence and asymptotic independence test cases respectively. The logistic sample has dependence parameter $\theta = 0.5$, and the Gaussian sample has correlation parameter $\rho = 0.5$.

we use the two methods to estimate the joint exceedance probability corresponding to the jointly extreme region marked in green on both panels, where the boundary of the region is chosen to yield true joint exceedance probabilities $p_{\text{Log}} = 2.67 \cdot 10^{-4}$ and $p_{\text{Gaus}} = 2.88 \cdot 10^{-4}$ respectively.

We simulate 50 replicate samples of each dependence structure, and obtain probability estimates $\hat{p}_{i,j}^{(1)}, \ldots, \hat{p}_{i,j}^{(50)}$ for dependence structures $i \in \{\text{Log}, \text{Gaus}\}$, using methods $j \in \{\text{Wan}, \text{HT}\}$. Given the known true probabilities p_{Log} and p_{Gaus}, we can then calculate the root mean squared error (RMSE)

$$RMSE(\hat{p}_{i,j}) = \sqrt{\frac{1}{50} \sum_{r=1}^{50} (\hat{p}_{i,j}^{(r)} - p_i)^2},$$

for each combination of dependence structure i and method j. The calculated RMSEs

are reported as RMSE ratios

$$R_{\text{Log}} = \frac{RMSE(\hat{p}_{\text{Log,Wan}})}{RMSE(\hat{p}_{\text{Log,HT}})} \quad \text{and} \quad R_{\text{Gaus}} = \frac{RMSE(\hat{p}_{\text{Gaus,Wan}})}{RMSE(\hat{p}_{\text{Gaus,HT}})},$$

to highlight the reduction error in using the 'HT' (Heffernan and Tawn, 2004) method over the 'Wan' (Wan and Davis, 2019) method. We obtain values $R_{\text{Log}} = 1.6$ and $R_{\text{Gaus}} = 30$.

As discussed in Section 2.2.6, the method of Wan and Davis (2019) is developed for the asymptotically dependent setting, in both threshold selection and probability estimation steps. Therefore, an RMSE ratio that is much in excess of 1 in the asymptotically independent case of Gaussian dependence is to be expected, as their method is not developed for application to this setting (in the sense that all asymptotic independence cases collapse to the case of complete independence). However, even in the asymptotically dependent setting of logistic dependence, the conditional extreme value method with randomly chosen threshold values outperforms that of Wan and Davis (2019). This suggests that the conditional extreme value method, despite using randomly chosen threshold values and covering both asymptotic dependence and asymptotic independence possibilities, is the preferable approach for a range of extremal dependence types for which the method of Wan and Davis (2019) is designed to perform well in.

5.3.3 Sensitivity of parameter estimate error to weight

With the results of Section 5.3.2 suggesting that the approach of Heffernan and Tawn (2004) outperforms that of Wan and Davis (2019) when using randomly selected conditional exceedance thresholds, we now demonstrate how our threshold selection approach further improves the performance of the threshold selection and associated inference from using the Heffernan and Tawn (2004) approach. To do so, we consider condi-

tional extremes parameter estimates $\hat{\alpha}_\lambda$ and $\hat{\beta}_\lambda$, obtained using sample exceedances with conditional exceedance threshold chosen by our variance-weighted threshold selection method (5.2.9) with weight parameter λ. For a given sample $\{(x_i, y_i)\}_{i=1}^n$, we retain n_λ observations $\{(x'_j, y_j)\}_{j=1}^{n_\lambda}$ of values where X exceeds a threshold u_λ, with $x'_1, \dots x'_{n_\lambda}$ being the excesses of u_λ by X. Here, for a given value of λ, u_λ is chosen via minimisation of metric $T_u(\lambda)$ with respect to u, over 20 equally spaced values of u in the interval [0.6, 0.99]. The regression (5.2.2) is then fitted to the observations $\{(x'_j, y_j)\}_{j=1}^{n_\lambda}$, yielding maximum likelihood estimates $\hat{\alpha}_\lambda$ and $\hat{\beta}_\lambda$ for α and β. We measure the error in these parameter estimates, repeated over 50 replicated samples.

We again consider bivariate samples of size $n = 10^5$ with Gaussian and logistic dependence structures. For these two dependence structures, the values of α and β in the limit (5.2.1) are known, with $(\alpha_{\text{Gaus}}, \beta_{\text{Gaus}}) = (\rho^2, 1/2)$ for Gaussian dependence structure with covariance $\rho > 0$, and $(\alpha_{\text{Log}}, \beta_{\text{Log}}) = (1, 0)$ for logistic dependence structure with any value of parameter $\theta \in (0, 1]$; see, e.g., Simpson and Tawn (2024). We consider 50 samples from these dependence structures, replicating the parameter estimation procedure described above for each sample to obtain estimates $\{(\hat{\alpha}^{(r)}_{\text{Log},\lambda}, \hat{\beta}^{(r)}_{\text{Log},\lambda})\}_{r=1}^{50}$ for $(\alpha_{\text{Log}}, \beta_{\text{Log}})$ and $\{(\hat{\alpha}^{(r)}_{\text{Gaus},\lambda}, \hat{\beta}^{(r)}_{\text{Gaus},\lambda})\}_{r=1}^{50}$ for $(\alpha_{\text{Gaus}}, \beta_{\text{Gaus}})$. We then calculate the RMSE, for dependence structure $i \in \{\text{Log}, \text{Gaus}\}$ and weight parameter $\lambda \in [0, 1]$, as

$$RMSE(\hat{\alpha}_{i,\lambda}, \hat{\beta}_{i,\lambda}) = \sqrt{\frac{1}{50} \sum_{r=1}^{50} ||(\hat{\alpha}^{(r)}_{i,\lambda}, \hat{\beta}^{(r)}_{i,\lambda}) - (\alpha_i, \beta_i)||^2},$$

where we define the error in a parameter estimate as the Euclidean distance between an estimate and the true value.

We compare RMSE of these parameters to the RMSE of corresponding parameter estimates found using thresholds selected randomly from the same set of candidate values within the interval [0.6, 0.99]. The resulting RMSE ratios are given in Table 5.3.1, presented as the ratio of RMSE when using random thresholds over RMSE when using

Dependence structure	$\lambda = 0$	$\lambda = 0.1$	$\lambda = 0.3$	$\lambda = 0.5$	$\lambda = 0.7$	$\lambda = 0.9$
Logistic (AD)	1.38	1.38	1.36	1.26	1.13	1.07
Gaussian (AI)	1.44	1.49	1.38	1.27	0.94	0.83

Table 5.3.1: RMSE of parameter estimates when using random thresholds over RMSE of parameter estimates when using our method, for logistic dependence type with $\theta = 0.5$ and Gaussian dependence type with $\rho = 0.5$, for a range of λ values. A ratio greater than 1 indicates that the conditional extreme value model using our threshold selection method outperforms the conditional extreme value model using randomly chosen thresholds.

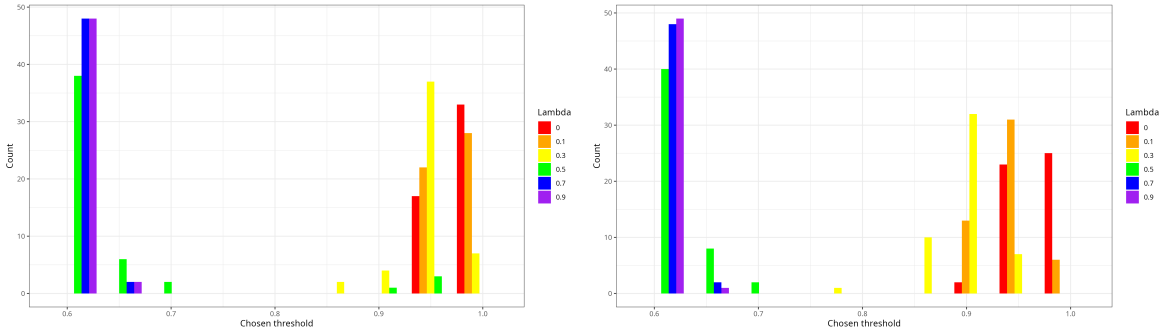


Figure 5.3.8: Distribution of chosen thresholds over the 50 replicate samples with Logistic (left) and Gaussian (right) dependence structures, for the range of λ values in Table 5.3.1.

our approach, meaning a value greater than 1 indicates that our method outperforms randomly chosen thresholds. These ratios are presented for five illustrative values of the weight parameter λ, ranging from cases of higher bias term weighting ($\lambda = 0$) and higher variance term weighting ($\lambda = 0.9$). Additionally, Figure 5.3.8 shows the distribution of chosen thresholds over the 50 replicate samples with Logistic (left) and Gaussian (right) dependence structures, for the range of λ values in Table 5.3.1.

For both dependence types, when setting $\lambda \geq 0.7$, the metric places more weight on the variance term, and so chooses thresholds to the lower end of the interval [0.6, 0.99], which retain more threshold exceedances for inference and thus reduce parameter estimate variability. The larger error in the parameter estimates in comparison to using a random threshold seen for these higher values of λ suggest the resulting threshold choices to be inappropriately low. The results for $\lambda = 0.3$ and $\lambda = 0.5$ show an

interesting pattern, in that the RMSE ratios for each dependence type are similar, but the chosen thresholds are very different. It could be that these values perform similarly because they tend to lead to threshold choices away from the boundaries of the interval [0.6, 0.99], despite the difference in the thresholds. When setting $\lambda = 0.1$, the weighted metric (5.2.9) appears to provide a good balance between the bias and variance terms, leading to a sensible choice of threshold, as suggested by the smaller error in parameter estimates compared to using a random threshold, for both dependence types, as seen in Table 5.3.1. When setting $\lambda = 0$, the metric reverts to the bias metric (5.2.7) with no variance term, and so chooses thresholds to the upper end of the interval [0.6, 0.99], as these provide a better approximation to the limiting behaviour in expression (5.2.1). For the Logistic dependence structure, the error when using $\lambda = 0$ is the same as that when using $\lambda = 0.1$, suggesting that the variance term is not needed in this case. However, for the Gaussian dependence structure, the method shows better performance for $\lambda = 0.1$ than for $\lambda = 0$, suggesting that inclusion of the variance term has value.

5.4 Conclusion

5.4.1 Preliminary findings

We have introduced a bias metric that gives a sensible measure of the conditional extreme value model's quality of fit to excesses above a candidate threshold, by assessing the performance of the transformation (5.2.5) to joint uniformity and thereby if it is a good model fit when using observations above the candidate threshold. Sampling variability in the bias metric is accounted for using a parametric bootstrapping procedure, making the method more robust than when using a single sample. In cases with a true threshold for which the model of Heffernan and Tawn (2004) is exact above, e.g., the idealised case explored in Section 5.2.2, the metric can successfully identify a sudden drop in bias at the true threshold, and thus be used to select a value for the

conditional exceedance threshold. In cases with no true threshold for which the model is exact above, e.g., for a Gaussian dependence structure, the metric correctly identifies the decrease in bias expected with an increase in exceedance threshold, however, it does not offer a clear choice of threshold value.

To make better use of the bias metric in cases with no true threshold, we introduce a variance-weighted extension of the metric, consisting of a convex combination of the logarithm of the bias metric and the logarithm of the determinant of the covariance matrix of conditional extreme value model parameters, weighted by a parameter $\lambda \in [0, 1]$. The covariance matrix is estimated using the parameter estimates yielded from the parametric bootstrap. Adding this variance penalty to the metric creates a bias-variance trade-off, where larger values of λ lead to the metric being minimised for lower candidate thresholds which increase the number of exceedances for use in inference and thus reduce parameter estimate variability, and smaller values of λ leading to the metric being minimised for higher candidate thresholds which improve the finite approximation to the limit (5.2.1).

Preliminary experiments demonstrate the improved estimation performance for extreme event probabilities of the method of Heffernan and Tawn (2004) over the existing threshold selection and probability estimation approach of Wan and Davis (2019), the latter of which performs well for asymptotically dependent settings only. This motivates the development of our novel threshold selection method for conditional extreme value models, whose model specification and estimation procedure for extreme event probabilities are able to correctly capture both asymptotic dependence and asymptotic independence.

An analysis of error in conditional extreme value model parameters show that performance of the variance-weighted metric is dependent on the value of λ. Smaller values of λ (e.g., $\lambda \approx 0.1$) result in good performance of the metric, choosing threshold values in the middle or upper end of the candidate interval. This results in a reduction

in parameter estimate error over those found using randomly selected thresholds. For larger values of λ, the metric over-penalises the increased variability in parameter estimates seen at higher candidate threshold values, leading to low choices of threshold value. This leads to an increase in parameter estimate error over those found using randomly selected thresholds, due to the increased bias in the model caused by poor approximation to the limit (5.2.1) at low thresholds.

5.4.2 Further work

Threshold selection methods using alternative metrics could be explored. Motivated by the univariate threshold selection method of Wadsworth (2016), we briefly investigated a method that measured the change in residual distribution with respect to excesses of X above a candidate value of u, testing if the assumption of independence between residuals and excesses of Heffernan and Tawn (2004) is satisfied for this value of u. By using hypothesis testing approaches such as the Kolmogorov-Smirnoff test (e.g., Berger and Zhou 2014) to obtain p-values for the similarity between distributions of residuals in fixed 'bands' of excesses above u, and then using changepoint detection methods to find at which candidate threshold these p-values indicate the residuals and excesses are independent, we constructed an alternative automated threshold selection approach. Our investigation into this particular method, however, found that the approach was unable to detect sensible threshold choices for all test cases, due to an apparent lack of statistical power in the hypothesis test.

The results of our experiments motivate further investigation into the selection of the weight parameter λ, in order to better guarantee the performance of our metric in general cases. We believe that this investigation could move in two potential directions. The first direction is an empirical assessment of the effect of varying λ over a comprehensive set of test cases, representing the full range of extremal dependence types, for different sample sizes. The performance would need to be evaluated against a range

of measures, including probability and quantile estimate errors. The idea behind this direction would be to find a sensible range of λ values for which the metric reliably offers improvement over randomly chosen thresholds, allowing a user to be confident the metric is performing well when using a λ value in this range. We have carried out preliminary work into this first direction, exploring the improvements in probability and quantile estimate error offered by the method over using randomly chosen thresholds, for a range of representative dependence structures. The second direction is development of a tuning approach (such as cross validation, see e.g., Hastie et al. 2009) for the value of λ in arbitrary settings, allowing a user to find an 'optimal' value of λ for any data, even if this value varies between settings.

Other options for future development include modifications of the current metric, or use of the metric on different versions of the conditional extreme value model. For instance, consider the bias metric discussed in Section 5.2.2. We might withhold the top $\phi\%$ of observations of X when calculating this metric, where the exceedance threshold u_ϕ corresponding to these observations is larger than any candidate threshold considered. We can then calculate the bias metric (5.2.7) using only observations in the interval (u, u_ϕ) at each candidate threshold value u in the model fit, thus providing a metric less sensitive to the largest values in our sample. The withheld observations, or potentially a subset of retained observations, could then be used as part of a cross-validation procedure. This mirrors the strategy taken by Collings et al. (2025), who in the univariate setting modify the metric of Murphy et al. (2025) to place increased weight on observations within a specified (typically extreme) range, prioritising model fit in the part of the distribution considered most important.

Also, it would be interesting to apply our metric to conditional extreme value model frameworks other than those of Heffernan and Tawn (2004) and Keef et al. (2013). For example, we are aware of work being carried out currently, separately by Aiden Farrell (Farrell et al., 2025) and Kristina Bratkova, both at Lancaster University, that uses a

form of asymmetric generalised Gaussian distribution model for the residual distribution G, rather than the empirical estimator that we employ. If a stricter form is imposed on the distribution G, then the assessment of fit of the conditional extreme value model may become more powerful in aiding the choice of threshold quantile to use, potentially leading to the bias metric (5.2.7) showing a clearer choice of threshold value in general cases.

When employing such models for the residual distribution, our method should be extended to the general $d > 2$ case, requiring extension of the transformation (5.2.5) to jointly uniform variables (V_u, W_u) to dimension $d > 2$. This would require knowledge of the marginal distributions for each component of the d-dimensional random variable being modelled, other than the conditioning component. Additionally, the suitability of the test of uniformity using Ripley's K function Ripley (1988) should be considered. The method of Ripley (1988) is defined for arbitrary dimension $d \geq 2$, and has seen successful application for $d = 3$ (see e.g., Baddeley et al. 2016). However, complications surrounding edge correction techniques and the need for large amounts of data in $d > 3$ suggest that alternative tests for uniformity should be utilised. For example, we may consider the method of Friedman and Rafsky (1979), which tests for similarity between the distributions of two samples in arbitrary dimension $d \geq 2$ by considered the minimal spanning tree of the pooled sample points.

Chapter 6

Extreme value methods for estimating rare events in Utopia: EVA (2023) conference data challenge

6.1 Introduction

This chapter details an approach to the data challenge organised for the Extreme Value Analysis (EVA) 2023 Conference. The objective of the challenge was to estimate extremal probabilities, or their associated quantiles, for simulated environmental data sets for various locations in a fictitious country called Utopia. The data challenge is split into 4 challenges; challenges C1 and C2 focus on a setting where data is obtained from a single location while challenges C3 and C4 concern multivariate data sets, where data is obtained simultaneously from multiple locations.

Challenge C1 requires estimation of the 0.9999-quantile of the distribution of the environmental response variable Y conditional on a covariate vector $\boldsymbol{X}$, for 100 realisa-

tions of covariates. To do so, we model the tail of $Y \mid \boldsymbol{X} = \boldsymbol{x}$ using a generalised Pareto distribution (GPD; Pickands, 1975) and employ the extreme value generalised additive modelling (EVGAM) framework, first introduced by Youngman (2019a), to account for the non-stationary data structure. We consider a variety of model formulations and select our final model using cross-validation. Furthermore, central 50% confidence intervals are estimated via a non-stationary bootstrapping technique, and the final model performance is assessed using the number of times the true conditional quantile lies in the confidence intervals (Rohrbeck et al., 2023). For Challenge C2, we are interested in estimating the value of q that satisfies $\Pr(Y > q) = 1/(300T)$, where $T = 200$.

Challenges C3 and C4 concern the estimation of probabilities for extreme multivariate regions, subsets of $\mathbb{R}^d$, where some or all of the components are so large that we seldom observe any data in them. Such estimates require techniques for modelling and extrapolating within the joint tail. For challenge C3, we want to estimate two joint tail probabilities for three unknown non-stationary environmental variables.

To achieve this, we propose a non-stationary extension of the model introduced by Wadsworth and Tawn (2013). Lastly, for challenge C4, we wish to estimate the probability that 50 variables (locations) jointly exceed prespecified extreme thresholds.

Based on an initial analysis, we separate the variables into five independent groups, and obtain distinct probability estimates for each group using the conditional extremes approach of Heffernan and Tawn (2004).

The remainder of the chapter is structured as follows. A suitable background to EVA is provided in Section 6.2, introducing concepts required throughout our work. Section 6.3 covers our approach to the univariate challenges C1 and C2, and the multivariate challenges C3 and C4 are considered in Sections 6.4 and 6.5, respectively. The chapter ends with a discussion of the results of all challenges in Section 6.6.

6.2 EVA background

6.2.1 Univariate modelling

Univariate EVA methods are concerned with capturing the behaviour of the tail of a distribution which allows for extreme quantities to be estimated. A common univariate approach is the peaks-over-threshold framework. Consider a continuous, independent and identically distributed (IID) random variable Y with distribution function F and upper endpoint $y^F := \sup\{y : F(y) < 1\}$. Pickands (1975) shows that, for some high threshold $v < y^F$, the excesses $(Y - v) \mid Y > v$, after suitable rescaling, converge in distribution to a GPD as $v \to y^F$. Davison and Smith (1990) provide an overview of the properties of the GPD, and also propose an extension of this framework to the non-stationary setting: given a non-stationary process Y with associated covariate(s) $\boldsymbol{X}$, the authors propose the following model

$$\Pr(Y > y + v \mid Y > v, \boldsymbol{X} = \boldsymbol{x}) = \left(1 + \frac{y\xi(\boldsymbol{x})}{\sigma(\boldsymbol{x})}\right)_+^{-1/\xi(\boldsymbol{x})}, \tag{6.2.1}$$

for $y > 0$, where $\sigma(\cdot), \xi(\cdot)$ are the covariate-dependent scale and shape parameters, respectively. Recent extensions of the Davison and Smith (1990) framework include allowing the threshold to be covariate-dependent, i.e., $v(\boldsymbol{x})$ (Kyselý et al., 2010; Northrop and Jonathan, 2011), and using generalised additive models (GAMs; Chavez-Demoulin and Davison, 2005; Youngman, 2019a) to capture the functions $\sigma(\cdot)$ and $\xi(\cdot)$ in a flexible manner.

6.2.2 Extremal dependence measures

In addition to analysing marginal tail behaviours, multivariate EVA methods are concerned with quantifying the dependence between extremes of the individual components. An important classification of this dependence is obtained through the measure χ (Joe, 1997): given a d-dimensional random vector $\boldsymbol{Z}$, with $d \geq 2$ and $Z_i \sim F$ for all

$i \in \{1, \ldots, d\}$,

$$\chi(u) := \left(\frac{1}{1-u}\right) \Pr(F(Z_1) > u, \ldots, F(Z_d) > u), \tag{6.2.2}$$

with $u \in [0, 1)$. Where the limit exists, we set $\chi := \lim_{u \to 1} \chi(u) \in [0, 1]$. When $\chi > 0$, we say that the variables in $\boldsymbol{Z}$ exhibit asymptotic dependence, i.e., can take their largest values simultaneously, with the strength of dependence increasing as χ approaches 1. If $\chi = 0$, the variables cannot all take their largest values together. In particular, for $d = 2$, we refer to the case $\chi = 0$ as asymptotic independence.

We also consider the coefficient of tail dependence proposed by Ledford and Tawn (1996). Using the formulation given in Resnick (2002), let

$$\eta(u) := \frac{\log(1-u)}{\log \Pr\left(F(Z_1) > u, \ldots, F(Z_d) > u\right)},$$

with $u \in [0, 1)$. When the limit exists, we set $\eta := \lim_{u \to 1} \eta(u) \in (0, 1]$. The cases $\eta = 1$ and $\eta < 1$, correspond to $\chi > 0$ and $\chi = 0$, respectively. For $\eta < 1$, this coefficient quantifies the form of dependence for random vectors that do not take their largest values simultaneously.

Since χ and η are limiting values, they are unknown in practice and must be approximated using numerical techniques. Therefore, when quantifying extremal dependence, we approximate χ (η) using empirical estimates of $\chi(u)$ $\big(\eta(u)\big)$ for some high threshold u.

6.3 Challenges C1 and C2

Both challenges concern 70 years of daily data for the capital city of Amaurot. Each year has 12 months of 25 days and two seasons (season 1 for months 1-6, and season 2 for months 7-12). Suppose Y is an unknown response variable, and $\boldsymbol{X} = (V_1, \ldots, V_8)$ is a vector of covariates, (V_1, V_2, V_3, V_4) denoting unknown environmental variables and (V_5, V_6, V_7, V_8) denoting season, wind direction (radians), wind speed (unknown scale), and atmosphere (recorded monthly), respectively.

For C1, we build a model for $Y \mid \boldsymbol{X}$ and estimate the 0.9999-quantile, with associated 50% confidence intervals, for 100 different covariate combinations denoted $\tilde{\boldsymbol{x}}_i$ for $i \in \{1, \ldots, 100\}$. Note $\tilde{\boldsymbol{x}}_i$ are not covariates observed within the data set, but new observations provided by the challenge organisers.

For C2, we estimate the marginal quantile q such that $\Pr(Y > q) = (6 \times 10)^{-4}$, which corresponds to a once in 200-year event in the IID setting; in particular, q is obtained subject to a predefined loss function. We first estimate the marginal distribution $F_Y(y)$ using Monte-Carlo techniques; see for instance, Eastoe and Tawn (2008). Since we have a large sample size, $n = 21,000$, it is reasonable to assume that the observed covariate sample is representative of $\boldsymbol{X}$. Thus, we can approximate the marginal distribution $F_Y(y)$ as follows,

$$\hat{F}_Y(y) = \int_{\boldsymbol{X}} F_{Y|\boldsymbol{X}}(y \mid \boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \approx \frac{1}{n} \sum_{t=1}^{n} F_{Y_t|\boldsymbol{X}_t}(y_t \mid \boldsymbol{x}_t). \tag{6.3.3}$$

where $F_{Y|\boldsymbol{X}}(\cdot)$ is the conditional distribution function of $Y \mid \boldsymbol{X}$ and $f_{\boldsymbol{X}}(\cdot)$ denotes the joint probability density of the covariates $\boldsymbol{X}$.

We incorporate the following loss function provided by the challenge organisers,

$$\mathcal{L}(q, \hat{q}) = \begin{cases} 0.9(0.99q - \hat{q}) & \text{if } 0.99q > \hat{q}, \\ 0 & \text{if } |q - \hat{q}| \leq 0.01q, \\ 0.1(\hat{q} - 1.01q) & \text{if } 1.01q < \hat{q}, \end{cases} \tag{6.3.4}$$

where q and $\hat{q}$ are the true and estimated marginal quantiles, respectively. This loss function penalises under-estimation more heavily than an over-estimation.

We conduct the same exploratory data analysis for both challenges given the same covariates are used; this is outlined in Section 6.3.1. In Section 6.3.2 we introduce our techniques for modelling $Y \mid \boldsymbol{X}$, which is then used for modelling Y via (6.3.3). Our approach for uncertainty quantification is outlined in Section 6.3.3, and we give our results for both challenges in Section 6.3.4.

6.3.1 Exploratory data analysis

Given the covariate vector $\boldsymbol{X_t} = \{V_{1,t}, \ldots, V_{8,t}\}$, the environmental response variable $Y_t,\ t \in \{1, \ldots, n\}$, is temporally independent (Rohrbeck et al., 2023). However, it is not clear which covariates affect Y, and what form these covariate-response relationships take. In what follows, we aim to explore these relationships so we can account for them in our modelling framework.

To begin, we explore the dependence between all variables to understand the relationships between covariates, as well as the relationships between individual covariates and the response variable. We investigate dependence in the main body of the data using Kendall's τ measure, while for the joint tails, we use the pairwise extremal dependence coefficients χ and η defined in Section 6.2; values for all pairs are shown in Figure 6.3.1, with the threshold u set at the empirical 0.95-quantile for the extremal measures.

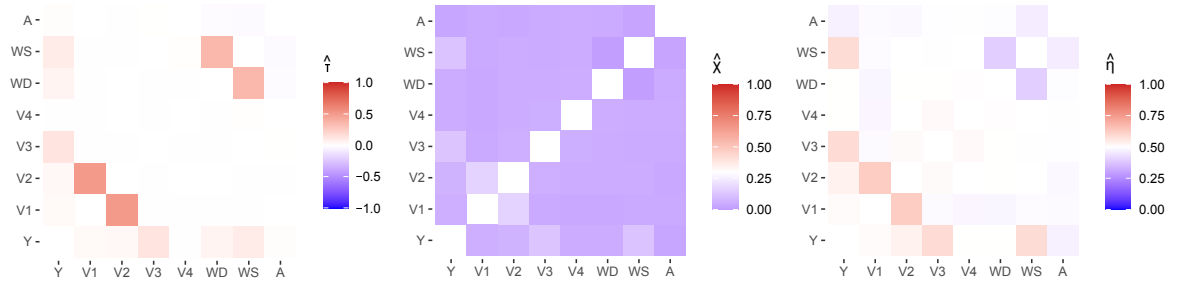


Figure 6.3.1: Heat maps for dependence measures for each pair of variables: Kendall's τ (left), χ (middle) and η (right). Note the scale in each plot varies, depending on the support of the measure, and the diagonals are left blank, where each variable is compared against itself.

The response variable Y has the strongest dependence with V_3 in the body of the distribution (see $\hat{\tau}$ in Figure 6.3.1), followed by V_6 (wind speed) then V_7 (wind direction). For the tail of the distribution, Y has strongest dependence with V_2, V_3 and V_6 (see $\hat{\chi}$ and $\hat{\eta}$ in Figure 6.3.1). We also find strong dependence between V_6 and V_7 in the body, but evidence of weak dependence in the tail (dark blue for $\hat{\chi}$ and $\hat{\eta}$). There is also strong dependence between V_1 and V_2 in both the body and tail (see dark

red for $\hat{\eta}$). We find very similar dependence relationships when the data are split into seasons. In the Supplementary Material, we show scatter plots of each covariate against the response variable; these demonstrate a highly non-linear relationship for each explanatory variable with Y.

Next, we explore temporal relationships for the response variable Y. We first find temporal non-stationarity as the distribution of Y varies significantly with V_5 (season); see the Supplementary Material for more detail. The mean and range of Y is higher in season 1 than season 2, with greater extreme values observed in season 1. However, within each season, across months, there is little temporal variation in the distribution of Y. We also find that Y exhibits temporal independence at all lags, with auto-correlation function (acf) values close to zero; see the Supplementary Material.

As noted in Rohrbeck et al. (2023), 11.7% of the observations have at least one value missing completely at random (MCAR). A detailed breakdown of the pattern of missing predictor observations is provided in the Supplementary Material. Since we can assume the data are MCAR, ignoring the observations that have a missing predictor covariate will not bias our inference, however, a complete case analysis is undesirable due to the amount of data loss. To mitigate against this, we attempted to impute the observations where predictors are missing but ultimately could not find an imputation method that satisfactorily retained the dependence structure between the response and covariates, particularly in the tails of the distribution. Therefore, we use a case analysis approach, whereby an observation is only removed if a predictor covariate of interest is missing. This results in only 4% of observations being removed for our final model.

6.3.2 Methods

Due to the complex nature of the data, we consider various non-stationary GPD models, as in equation (6.2.1), that are formulated as GAMs to fit $Y \mid \boldsymbol{X}$. For threshold selection, we extend the method proposed by Murphy et al. (2025) to select a thresh-

old for non-stationary, covariate-dependent GPD models; the details are provided in Section 6.3.2. Our inference and model selection procedures are then provided in Sections 6.3.2 and 6.3.2, respectively. We note that the same model formulation is used for both C1 and C2 with a small adjustment to the parameter estimation procedure for C2 to incorporate the provided loss function given in (6.3.4). We utilise equation (6.3.3) to obtain the marginal distribution of Y.

General model formulation

Let $\tilde{\boldsymbol{X}}_t$ denote the set of predictor covariates with $t \in \{1, \ldots, n\}$. Then y_t and $\tilde{\boldsymbol{x}}_t$ denote the observations of the response variable and predictive covariates, respectively. We consider models with the following form,

$$F_{Y_t|\tilde{\boldsymbol{X}}_t}(y_t|\tilde{\boldsymbol{X}}_t = \tilde{\boldsymbol{x}}_t) = 1 - \zeta(\tilde{\boldsymbol{x}}_t)\left[1 + \xi(\tilde{\boldsymbol{x}}_t)\left(\frac{y_t - v(\tilde{\boldsymbol{x}}_t)}{\sigma(\tilde{\boldsymbol{x}}_t)}\right)\right]_+^{-1/\xi(\tilde{\boldsymbol{x}}_t)}, \tag{6.3.5}$$

where $v(\tilde{\boldsymbol{x}}_t)$ and $\zeta(\tilde{\boldsymbol{x}}_t)$ are a covariate-dependent threshold and rate parameter, respectively, such that the rate parameter corresponds to the probability of exceeding the threshold.

Our analysis in Section 6.3.1 indicates that V_3, V_5 (season), and V_6 (wind speed) exhibit non-trivial dependence relationships with the response variable. Therefore we assume these variables can be used as predictor variables for modelling Y, and set $\tilde{\boldsymbol{x}} := (\boldsymbol{V}_j)_{j\in\{3,5,6\}}$. Although V_7 (wind direction) also exhibits strong dependence with Y, we do not consider it here since it is highly correlated with wind speed so would involve adding complex interaction terms to the model formulation, and V_6 has a stronger relationship with Y compared to V_7 (see Figure 6.3.1).

Owing to the complex covariate structure observed in the data, as described in Section 6.3.1, we employ the flexible EVGAM framework proposed in Youngman (2019a) for modelling tail behaviour. Under this framework, GAM formulations are used to capture non-stationarity in the threshold, scale and shape functions given in equation

(6.3.5). Without loss of generality, consider the scale function $\sigma(\cdot)$. We assume that

$$h(\sigma(\tilde{\boldsymbol{x}})) = \psi_\sigma(\tilde{\boldsymbol{x}}), \quad \text{with} \quad \psi_\sigma(\tilde{\boldsymbol{x}}) = \beta_0 + \sum_{\kappa=1}^{K} \sum_{p=1}^{P_\kappa} \beta_{\kappa p} b_{\kappa p}(\tilde{\boldsymbol{x}}), \tag{6.3.6}$$

where $h(x) := \log(x)$ denotes the link function which ensures the correct support, with coefficients $\beta_0, \beta_{\kappa p} \in \mathbb{R}$ and basis functions $b_{\kappa p}$ for $p \in \{1, \ldots, P_\kappa\}, \kappa \in \{1, \ldots, K\}$, where K is the number of splines in the GAM formulation and P_κ is the basis dimension relating to spline κ. The basis functions can be in terms of individual covariates, i.e., $b_{\kappa p} : \mathbb{R} \mapsto \mathbb{R}$, or multiple covariates, i.e., $b_{\kappa p} : \mathbb{R}^m \mapsto \mathbb{R}$, $1 < m \leq 8$. Analogous forms can be taken for $v(\cdot)$ and $\xi(\cdot)$, adjusting the link function $h(\cdot)$ as appropriate, although these are not considered here for reasons detailed below.

To select an appropriate threshold, we employ the threshold selection method of Murphy et al. (2025) and extend this approach to select a threshold for non-stationary, covariate-dependent GPD models. The method selects a threshold based on minimising the expected quantile discrepancy (EQD) between the sample quantiles and fitted GPD model quantiles. When fitting a non-stationary model, the excesses will not be identically distributed across covariates. Thus, to utilise the EQD method in this case, we use the fitted non-stationary GPD parameter estimates to transform the excesses to common standard exponential margins and compare sample quantiles against theoretical quantiles from the standard exponential distribution. This transformation is a common approach for checking the model fit of a non-stationary GPD (Coles, 2001).

We use a stepped-threshold according to season as there is clear variation in the distribution, and thereby the extremes, of Y between seasons; see the Supplementary Material for more details. Specifically, we set $v(\tilde{\boldsymbol{x}}_t) := \mathbb{1}(\tilde{x}_{2,t} = 1)v_1 + \mathbb{1}(\tilde{x}_{2,t} = 2)v_2$, $v_1, v_2 \in \mathbb{R}$, with corresponding rate parameter $\zeta(\tilde{\boldsymbol{x}}_t) := \mathbb{1}(\tilde{x}_{2,t} = 1)\zeta_1 + \mathbb{1}(\tilde{x}_{2,t} = 2)\zeta_2$, where $\zeta_1, \zeta_2 \in [0, 1]$ denote the probabilities of exceeding the threshold for seasons 1 and 2, respectively, and $\tilde{x}_{r,t}$ are realisations of the r^{th} component of $\tilde{\boldsymbol{x}}$ for $r \in \{1, 2, 3\}$. This seasonal threshold significantly improves model fits; see the Supplementary Material for further details. GAM forms for the threshold were also explored, but did not offer

significant improvement. Furthermore, the smooth GAM formulation of the GPD scale parameter adequately captures any residual variation in the response arising due to covariate dependence.

Inference

For all GAM formulations, we only consider basis functions of singular covariates, since specifying basis functions of multiple variables requires a detailed understanding of covariate interactions and can significantly increase the computational complexity of the modelling procedure (Wood, 2017). We keep the shape function $\xi(\boldsymbol{x}) := \xi \in \mathbb{R}$ constant across covariates; this is common in non-stationary analyses, since this parameter is difficult to estimate (Chavez-Demoulin and Davison, 2005). Within the GAM formulation, we consider several parametric forms to account for the predictive covariates in the scale parameter using linear models, indicator functions and splines.

When using splines, we are required to select a basis dimension $P_\kappa \in \mathbb{N}$; this determines the number of coefficients to be estimated. Basis dimension is the most important choice within spline modelling procedures and directly corresponds with the flexibility of the framework (Wood, 2017). We only consider splines for V_3 and V_6. For each $\tilde{X}_r$, $r \in \{1, 3\}$, we determine the basis dimension P_1 and P_2, respectively, by first building a model for $Y_t \mid \tilde{X}_{r,t}$, to allow us to consider the effect of this predictor on the response directly. We vary the basis dimension and compare the resulting models using cross validation (CV), detailed in the following section. We set $P_1 = 4$ and $P_2 = 3$ for V_3 and V_6, respectively.

For C2, we incorporate the loss function of equation (6.3.4) into the estimation procedure. Let $\mathcal{I}_v := \{t \in \{1, \ldots, n\} \mid y_t > v(\tilde{\boldsymbol{x}}_t)\}$ denote the set of temporal indices corresponding to threshold exceedances and $n_v := |\mathcal{I}_v|$. We consider the objective function

$$S(\boldsymbol{\theta}) := -l_R(\boldsymbol{\theta}) + \sum_{i \in \mathcal{I}_v} \mathcal{L}(q_i^*, \hat{q}_i)/n_v, \tag{6.3.7}$$

where $l_R(\boldsymbol{\theta})$ denotes the penalised log-likelihood function of the restricted maximum likelihood estimation (REML) approach (Wood, 2017), $\boldsymbol{\theta}$ denotes the parameter vector associated with the GPD formulation of equation (6.3.6), and $\sum_{i\in\mathcal{I}_v} \mathcal{L}(q_i^*, \hat{q}_i)/n_v$ denotes the average loss between the sample quantiles of the transformed excesses and the theoretical standard exponential quantiles. Specifically, we transform the excesses, $(y_t - v(\tilde{\boldsymbol{x}}_t))_{t\in\mathcal{I}_v}$, to standard exponential margins using the fitted non-stationary GPD parameter estimates and compare the ordered excesses, $\boldsymbol{q}^*$, to the theoretical quantiles, $\hat{\boldsymbol{q}}$, from a standard exponential distribution evaluated at probabilities $\{p_i = i/(n_v + 1), i = 1, \ldots, n_v\}$. Minimising the objective function $S(\boldsymbol{\theta})$ ensures that the parameter estimates also account for and minimise the loss function, $\mathcal{L}$. We use this formulation to adjust the GPD parameters for challenge C2 once a threshold is selected.

Model selection

To determine the best-fitting model, we use a forward selection process and aim to minimise the model's CV score. For each model, we apply k-fold CV (Hastie et al., 2001, Ch 7.) utilising the continuous ranked probability score (CRPS, Gneiting and Katzfuss, 2014) as our goodness-of-fit metric. CRPS describes the discrepancy between the predicted distribution function and observed values without the specification of empirical quantiles. We explore model ranking by taking both $k = 10$ and 50, and find that both give an equivalent ranking; we present results for the latter. We also provide the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values to aid in model selection. A subset of models used in the forward selection process are detailed in Table 6.3.1 where, for each model, we provide the change in the CRPS, AIC and BIC relative to model 1. The parameterisation of model 7 achieves the largest reduction for all three metrics relative to the baseline model.

Table 6.3.1: Table of selected models considered for challenge C1. $\mathbb{1}(\cdot)$ denotes an indicator function, $s_i(\cdot)$ for $i \in \{1, 2\}$ denote thin-plate regression splines, β_0, β_1 are coefficients to be estimated, and $\tilde{x}_{r,t}$ is defined as in the text. All values have been given to one decimal place.

Model	$\sigma(\tilde{\boldsymbol{x}}_t)$	ΔCRPS	ΔAIC	ΔBIC
1	β_0	0	0	0
2	$\beta_0 + \beta_1 \mathbb{1}(\tilde{x}_{2,t} = 1)$	-0.5	-33.4	-26.1
3	$\beta_0 + s_1(\tilde{x}_{1,t})$	-0.9	-408.5	-379.2
4	$\beta_0 + s_2(\tilde{x}_{3,t})$	-0.5	-284.3	-276.8
5	$\beta_0 + \beta_1 \mathbb{1}(\tilde{x}_{2,t} = 1) + s_1(\tilde{x}_{1,t})$	-0.9	-425.8	-388.1
6	$\beta_0 + s_1(\tilde{x}_{1,t}) + s_2(\tilde{x}_{3,t})$	-1.0	-752.7	-717.2
7	$\beta_0 + \beta_1 \mathbb{1}(\tilde{x}_{2,t} = 1) + s_1(\tilde{x}_{1,t}) + s_2(\tilde{x}_{3,t})$	**-1.1**	**-780.0**	**-735.3**

6.3.3 Uncertainty

For each of the 100 different covariate combinations, $\tilde{\boldsymbol{x}}_i$ for $i \in \{1, \ldots, 100\}$, we need to construct central 50% confidence intervals. We use a bootstrapping procedure to avoid making potentially inaccurate assumptions such as the asymptotic normality approximation of maximum likelihood estimates, for example. Traditional bootstrap approaches are non-parametric and randomly resample the data with replacement. However, in Section 6.3.1 we find that the response variable is dependent on covariates, and these covariates exhibit temporal dependence. A standard bootstrap procedure would therefore not retain this dependence. Instead, we preserve the temporal dependence structure of covariates and their relationship with the response variable by approximating our confidence intervals using the stationary, semi-parametric bootstrapping procedure adopted by D' Arcy et al. (2023).

First, the response variable Y_t is transformed to Uniform(0,1) margins to preserve its non-stationary behaviour; denote this sequence $U_t^Y = F_{Y_t|\tilde{\boldsymbol{X}}_t}(Y_t|\tilde{\boldsymbol{X}}_t = \tilde{\boldsymbol{x}}_t)$ where $F_{Y_t|\tilde{\boldsymbol{X}}_t}$ is the estimated model given in equation (6.3.5). We then adopt the stationary bootstrap procedure of Politis and Romano (1994) to retain the temporal dependence in the response and explanatory variables by sampling blocks of consecutive observations. The block length L is random and simulated from a Geometric$(1/l)$ distribution, where

the mean block length $l \in \mathbb{N}$ is carefully selected based on the autocorrelation function. This was selected at 50 days, the maximum lag for which the autocorrelation was significant across all variables; see the Supplementary Material. Denote this bootstrapped sequence on Uniform margins by U_t^B. We transform U_t^B back to the original scale using our fitted model, preserving the original structure of Y_t; we denote this series Y_t^B. Then we fit our model to Y_t^B to re-estimate all of the parameters and thus the quantile of interest. We repeat this procedure to obtain 200 bootstrap samples.

6.3.4 Results

For C1, we use our final model of Section 6.3.2 to estimate the 0.9999-quantile of $Y \mid \tilde{\boldsymbol{X}} = \tilde{\boldsymbol{x}}_i$, $i \in \{1, \ldots, 100\}$, for the set of 100 covariate combinations. The left panel of Figure 6.3.2 shows the quantile-quantile (QQ) plot for our model. There is general alignment between the model and empirical quantiles; however, there is some over-estimation in the upper tail, and our 95% tolerance bounds do not contain some of the most extreme response values. The right panel of Figure 6.3.2 shows our predicted quantiles, and their associated confidence intervals, compared to their true quantiles. As expected, our predictions tend to over-estimate the true quantiles. We note this figure is different from the one presented by Rohrbeck et al. (2023) due to an error in our code being fixed after submission. In this scenario, our estimated confidence intervals lead to a 14% coverage of the true quantiles, which does not alter our ranking for this challenge. Our performance and model improvements are discussed in Section 6.6.

For challenge C2, we estimate the quantile of interest as $\hat{q} = 213.1$ $(209.3, 242.1)$. A 95% confidence interval for the estimate is given in parentheses based on the bootstrapping procedure outlined in Section 6.3.2. Due to a coding error, this value differs from the original estimate submitted for the EVA (2023) Conference Data Challenge. The updated value over-estimates compared to the truth ($q = 196.6$).

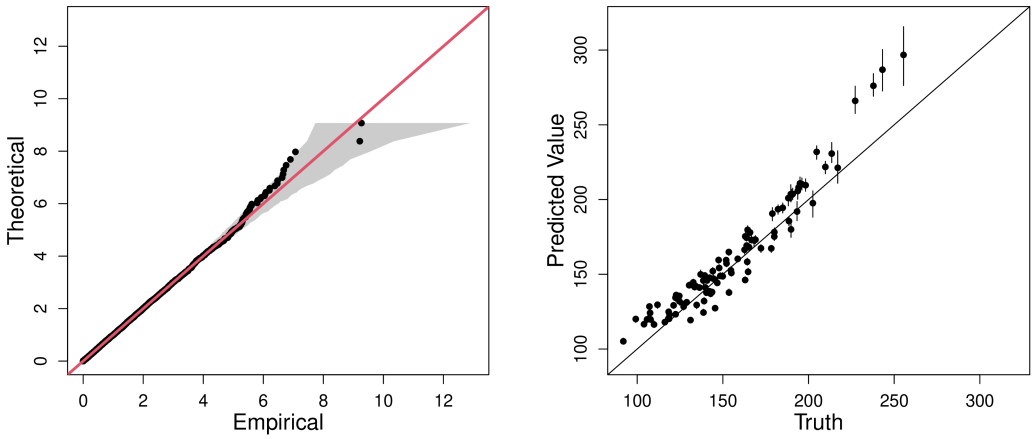


Figure 6.3.2: QQ plot for our final model (model 7 in Table 6.3.1) on standard exponential margins. The $y = x$ line is given in red and the grey region represents the 95% tolerance bounds (left). Predicted 0.9999−quantiles against true quantiles for the 100 covariate combinations. The points are the median predicted quantile over 200 bootstrapped samples and the vertical error bars are the corresponding 50% confidence intervals. The $y = x$ line is also shown (right).

6.4 Challenge C3

6.4.1 Exploratory data analysis

For challenge C3, we are provided with 70 years of daily data of an environmental variable for three towns on the island of Coputopia. These data are denoted by $Y_{i,t}$, $i \in \{1, 2, 3\}$, $t \in \{1, \ldots, n\}$, where i is the index of each town and t is the point in time. Each year consists of 12 months, each lasting 25 days, resulting in $n = 21,000$ observations for each location.

We are also provided with daily covariate observations $\boldsymbol{X}_t = (S_t, A_t)$, where S_t and A_t denote seasonal and atmospheric conditions, respectively. Season is a binary variable, taking values in the set $\{1, 2\}$, with each year of observations exhibiting both seasons for exactly 150 consecutive days. Atmospheric conditions are piecewise constant over months, with large variation in the observed values between months. A descriptive

figure of both covariates is given in the Supplementary Material.

In Rohrbeck et al. (2023), we are informed that $Y_{i,t}$ are distributed identically across all sites and over time, with standard Gumbel margins. However, it is not known whether the covariates $\boldsymbol{X}_t$ influence the dependence structure of $\boldsymbol{Y}_t := (Y_{1,t}, Y_{2,t}, Y_{3,t})$. We are also informed that, conditioned on covariates, the process is independent over time, i.e., $(\boldsymbol{Y}_t \mid \boldsymbol{X}_t) \perp\!\!\!\perp (\boldsymbol{Y}_{t'} \mid \boldsymbol{X}_{t'})$ for any $t \neq t'$. In this section, we examine what influence, if any, the covariate process $\boldsymbol{X}_t$ may have on the dependence structure of $\boldsymbol{Y}_t$.

We begin by transforming the time series $Y_{i,t}$ to standard exponential margins, denoted by $\boldsymbol{Z}_{i,t}$, via the probability integral transform. This transformation is common in the study of multivariate extremes and can simplify the description of extremal dependence (Keef et al., 2013). To explore the extremal dependence in the Coputopia time series, we consider all 2- and 3-dimensional subvectors of the process, i.e., $\{Z_{i,t}, i \in I, t \in \{1, \ldots, n\}\},\ I \in \mathcal{I} := \{\{1,2\},\{1,3\},\{2,3\},\{1,2,3\}\}$. This separation is important to ensure the overall dependence structure is fully understood, since intermediate scenarios can exist where a random vector exhibits $\chi = 0$, but $\chi > 0$ for some 2-dimensional subvector(s) (Simpson et al., 2020).

Furthermore, to explore the impact of covariates on the dependence structure, we partition the time series into subsets using the covariates. For the seasonal covariate, let $G^S_{I,j} := \{Z_{i,t}, i \in I, S_t = j\}$ for $j = 1, 2$, and for the atmospheric covariate, let $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ denote the permutation associated with the order statistics of A_t, defined so that ties in the data are accounted for. We then split the data into 10 equally sized subsets corresponding to the atmospheric order statistics, i.e., $G^A_{I,k} := \left\{Z_{i,t}, i \in I, t \in \Sigma^k\right\}$ for $k = 1, 2, \ldots, 10$, where $\Sigma^k := \{t \mid (k-1)n/10 + 1 \leq \pi(t) \leq kn/10\}$. Thus, the atmospheric values associated with each subset $G^A_{I,k}$ will increase over k.

The idea behind these subsets is to examine whether altering the values of either covariate impacts the extremal dependence structure. Consequently, we set $u = 0.9$ and

estimate $\chi(u)$ using the techniques outlined in Section 6.2, with uncertainty quantified through bootstrapping with 200 samples. The bootstrapped χ estimates for $G^A_{I,k}$ with $I = \{1, 2, 3\}$ are given in Figure 6.4.3. The plots for the remaining index sets in $\mathcal{I}$, along with the subsets associated with the seasonal covariate, are given in the Supplementary Material. The estimates of χ appear to vary, in the majority of cases, across both subset types (seasonal and atmospheric), suggesting both covariates have an impact on the dependence structure. For the atmospheric process in particular, the values of χ tend to decrease for higher atmospheric values, suggesting a negative association between the strength of positive extremal dependence and the atmospheric covariate. We also observe that across all subsets, χ appears consistently low in magnitude, suggesting the extremes of some, if not all, of the sub-vectors are unlikely to occur simultaneously. As such, for modelling the Coputopia time series, we require a framework that can capture such forms of dependence. We also consider pointwise estimates of the function $\lambda(\cdot)$, as defined later in equation (6.4.8), over $G^S_{I,j}$ and $G^A_{I,k}$ for fixed simplex points; these results are given in the Supplementary Material. Similar to χ, estimates of $\lambda(\cdot)$ vary significantly across subsets, providing additional evidence of non-stationarity within extremal dependence structure.

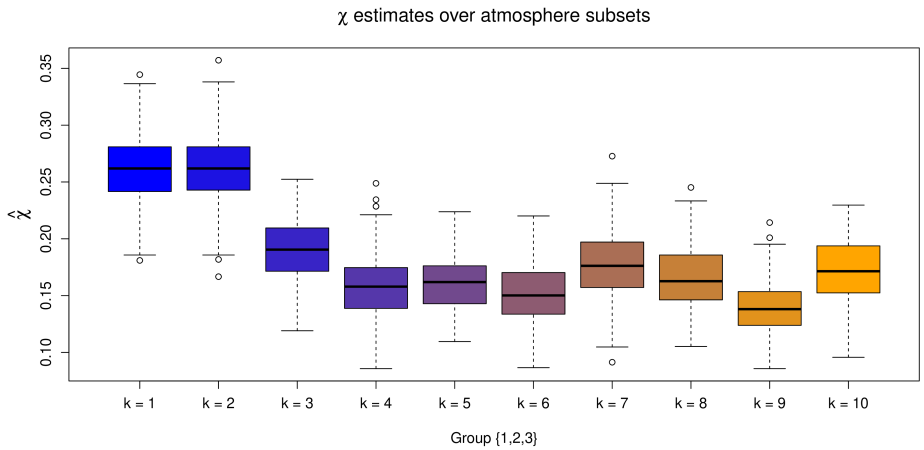


Figure 6.4.3: Boxplots of empirical χ estimates obtained for the subsets $G^A_{I,k}$, with $k = 1, \ldots, 10$ and $I = \{1, 2, 3\}$. The colour transition (from blue to orange) over k illustrates the trend in χ estimates as the atmospheric values are increased.

6.4.2 Modelling of joint tail probabilities under asymptotic independence

For challenge C3, we are required to estimate probabilities $p_1 := \Pr(Y_1 > y, Y_2 > y, Y_3 > y)$ and $p_2 := \Pr(Y_1 > v, Y_2 > v, Y_3 < m)$, with $y = 6$, $v = 7$ and $m = -\log(\log(2))$. Note that p_1 and p_2 are independent of the covariate process and correspond to different extremal regions in $\mathbb{R}^3$; we refer to p_1 and p_2 as parts 1 and 2 of the challenge, respectively. For the remainder of this section we will consider the transformed exponential variables (Z_1, Z_2, Z_3), omitting the subscript t for ease of notation. Observe that $F_{(-Z_3)}(z) = e^z$, for $z < 0$; setting $\tilde{Z}_3 := -\log(1 - \exp(-Z_3))$, we have

$$p_2 = \Pr(Z_1 > \tilde{v}, Z_2 > \tilde{v}, Z_3 < \tilde{m}) = \Pr\left(Z_1 > \tilde{v}, Z_2 > \tilde{v}, \tilde{Z}_3 > \tilde{m}\right),$$

where $\tilde{v}$ and $\tilde{m}$ denote the values v and m transformed to the standard exponential scale, e.g., $\tilde{v} := -\log(1 - \exp(-\exp(-v)))$. Similarly, we have $p_1 = \Pr(Z_1 > \tilde{y}, Z_2 > \tilde{y}, Z_3 > \tilde{y})$. Consequently, both p_1 and p_2 can be considered as joint survivor probabilities.

Since not all extremes of Z_1, Z_2 and Z_3 are observed simultaneously, we employ the framework by Wadsworth and Tawn (2013), which is a generalisation of the approach proposed in Ledford and Tawn (1996). The model of Wadsworth and Tawn (2013) assumes that for any ray $\boldsymbol{\omega} \in \boldsymbol{S}^2 := \{(\omega_1, \omega_2, \omega_3) \in [0,1]^3 : \omega_1 + \omega_2 + \omega_3 = 1\}$, where $\boldsymbol{S}^2$ denotes the standard 2-dimensional simplex,

$$\begin{aligned} \Pr(Z_1 > \omega_1 r,\, Z_2 > \omega_2 r,\, Z_3 > \omega_3 r) &= \Pr(\min\{Z_1/\omega_1,\, Z_2/\omega_2,\, Z_3/\omega_3\} > r) \\ &= \mathcal{L}(e^r; \boldsymbol{\omega}) e^{-r\lambda(\boldsymbol{\omega})}, \end{aligned} \tag{6.4.8}$$

as $r \to \infty$, where $\lambda(\boldsymbol{\omega}) \geq \max(\boldsymbol{\omega})$ is known as the angular dependence function (ADF). Asymptotic dependence occurs at the lower bound, i.e., $\lambda(\boldsymbol{\omega}) = \max(\boldsymbol{\omega})$ for all $\boldsymbol{\omega} \in \boldsymbol{S}^2$, and the coefficient of tail dependence is related to the ADF via $\eta = 1/\{3\lambda(1/3, 1/3, 1/3)\}$. In practice, equation (6.4.8) can be used to evaluate extreme joint survivor probabilities; in particular, probabilities p_1 and p_2 can be identified

with the rays $\boldsymbol{\omega}^{(1)} := (\tilde{y}, \tilde{y}, \tilde{y})/r^{(1)}$ and $\boldsymbol{\omega}^{(2)} := (\tilde{v}, \tilde{v}, \tilde{m})/r^{(2)}$ in $\boldsymbol{S}^2$, respectively, where $r^{(1)} := \tilde{y} + \tilde{y} + \tilde{y}$ and $r^{(2)} := \tilde{v} + \tilde{v} + \tilde{m}$. See Section 6.4.4 for further details.

6.4.3 Accounting for non-stationary dependence

In the stationary setting, pointwise estimates of $\lambda(\cdot)$ can be obtained via the Hill estimator (Hill, 1975), from which tail probabilities can be approximated. However, alternative procedures are required for data exhibiting trends in dependence, such as the Coputopia data set. Existing approaches for capturing non-stationary dependence structures are sparse in the extremes literature, and most approaches are limited to asymptotically dependent data structures. For the case when data are not asymptotically dependent, Mhalla et al. (2019) and Murphy-Barltrop and Wadsworth (2024) propose non-stationary extensions of the Wadsworth and Tawn (2013) framework, while Jonathan et al. (2014c) and Guerrero et al. (2023) propose non-stationary extensions of the Heffernan and Tawn (2004) model (see Murphy-Barltrop and Wadsworth (2024) for a detailed review).

To account for non-stationary dependence in C3, we propose an extension of the Wadsworth and Tawn (2013) framework. With $\boldsymbol{Z}_t = (Z_{1,t}, Z_{2,t}, Z_{3,t})$ and $\boldsymbol{X}_t$, defined as in Section 6.4.1, we define the structure variable $T_{\boldsymbol{\omega},t} := \min\{Z_{1,t}/\omega_1, Z_{2,t}/\omega_2, Z_{3,t}/\omega_3\}$, for any $\boldsymbol{\omega} \in \boldsymbol{S}^2$; we refer to $T_{\boldsymbol{\omega},t}$ as the min-projection variable at time t. From Section 6.4.1, we know that the joint distribution of $\boldsymbol{Z}_t$ is not identically distributed over t; which implies non-stationarity in the distribution of $T_{\boldsymbol{\omega},t}$. To account for this, Mhalla et al. (2019) and Murphy-Barltrop and Wadsworth (2024) assume the following model given the vector of covariates $\boldsymbol{x}_t$:

$$\Pr\left(T_{\boldsymbol{\omega},t} > u \mid \boldsymbol{X}_t = \boldsymbol{x}_t\right) = \mathcal{L}\left(e^u \mid \boldsymbol{\omega}, \boldsymbol{x}_t\right) e^{-\lambda(\boldsymbol{\omega};\boldsymbol{x}_t)u} \text{ as } u \to \infty, \tag{6.4.9}$$

for all t, where $\lambda\left(\cdot; \boldsymbol{x}_t\right)$ denotes the non-stationary ADF. Note that this assumption is very similar in form to equation (6.4.8), with the primary difference being the function

$\lambda(\cdot; \boldsymbol{x}_t)$ is non-stationary over t. From equation (6.4.9), we have

$$\Pr\left(T_{\boldsymbol{\omega},t} - u > z \mid T_{\boldsymbol{\omega},t} > u, \boldsymbol{X}_t = \boldsymbol{x}_t\right) = e^{-\lambda(\boldsymbol{\omega};\boldsymbol{x}_t)z} \text{ as } u \to \infty, \tag{6.4.10}$$

for $z > 0$. Consequently, equation (6.4.9) is equivalent to assuming $(T_{\boldsymbol{\omega},t} - u) \mid \{T_{\boldsymbol{\omega},t} > u, \boldsymbol{X}_t = \boldsymbol{x}_t\} \sim \text{Exp}(\lambda(\boldsymbol{\omega}; \boldsymbol{x}_t))$ as $u \to \infty$.

We found that equation (6.4.9) was not flexible enough to capture the tail of $T_{\boldsymbol{\omega},t}$ for the Coputopia data; see Section 6.4.3 for further discussion. Thus, we propose the following model: given any $z > 0$ and a fixed $\boldsymbol{\omega} \in \boldsymbol{S}^2$, we assume

$$\Pr\left(T_{\boldsymbol{\omega},t} - u > z \mid T_{\boldsymbol{\omega},t} > u, \boldsymbol{X}_t = \boldsymbol{x}_t\right) = \left(1 + \frac{\xi(\boldsymbol{\omega}; \boldsymbol{x}_t)\, z}{\sigma(\boldsymbol{\omega}; \boldsymbol{x}_t)}\right)^{-1/\xi(\boldsymbol{\omega};\boldsymbol{x}_t)} \text{ as } u \to \infty, \tag{6.4.11}$$

where $\sigma(\cdot; \boldsymbol{x}_t), \xi(\cdot; \boldsymbol{x}_t)$ are non-stationary scale and shape parameter functions, respectively. This is equivalent to assuming $(T_{\boldsymbol{\omega},t} - u) \mid \{T_{\boldsymbol{\omega},t} > u, \boldsymbol{X}_t = \boldsymbol{x}_t\} \sim \text{GPD}(\sigma(\boldsymbol{\omega}; \boldsymbol{x}_t), \xi(\boldsymbol{\omega}; \boldsymbol{x}_t))$ as $u \to \infty$, and equation (6.4.10) is recovered by taking the limit as $\xi(\boldsymbol{\omega}; \boldsymbol{x}_t) \to 0$ for all t.

Our proposed formulation in equation (6.4.11) allows for additional flexibility within the modelling framework by including a GPD shape parameter $\xi(\boldsymbol{\omega}; \boldsymbol{x}_t)$, which quantifies the tail behaviour of $T_{\boldsymbol{\omega},t}$. Given the wide range of distributions in the domain of attraction of a GPD (Pickands, 1975), it is reasonable to assume that the tail of $T_{\boldsymbol{\omega},t}$ can be approximated by equation (6.4.11). For the Coputopia time series, this assumption appears valid, as demonstrated by the diagnostics in Section 6.4.3.

Model fitting

To apply equation (6.4.11), we first fix $\boldsymbol{\omega} \in \boldsymbol{S}^2$ and assume that the formulation holds approximately for some sufficiently high threshold level from the distribution of $T_{\boldsymbol{\omega},t}$; we denote the corresponding quantile level by $\tau \in (0, 1)$. For simplicity, the same quantile level is considered across all t. Further, let $v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t)$ denote the corresponding threshold function, i.e., $\Pr(T_{\boldsymbol{\omega},t} \leq v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t) \mid \boldsymbol{X}_t = \boldsymbol{x}_t) = \tau$ for all t. Under our assumption, we

have $(T_{\boldsymbol{\omega},t} - v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t)) \mid \{T_{\boldsymbol{\omega},t} > v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t), \boldsymbol{X}_t = \boldsymbol{x}_t\} \sim \text{GPD}(\sigma(\boldsymbol{\omega}; \boldsymbol{x}_t), \xi(\boldsymbol{\omega}; \boldsymbol{x}_t))$. We emphasise that $v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t)$ is not constant in t, and we would generally expect $v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t) \neq v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_{t'})$ for $t \neq t'$.

As detailed in Section 6.4.2, both p_1 and p_2 can be associated with points on the simplex $\boldsymbol{S}^2$, denoted by $\boldsymbol{\omega}^{(1)}$ and $\boldsymbol{\omega}^{(2)}$, respectively. Letting $\boldsymbol{\omega} \in \{\boldsymbol{\omega}^{(1)}, \boldsymbol{\omega}^{(2)}\}$, our estimation procedure consists of two stages: estimation of the threshold function $v_\tau(\boldsymbol{\omega}, \boldsymbol{z}_t)$ for a fixed $\tau \in (0, 1)$, followed by estimation of GPD parameter functions $\sigma(\boldsymbol{\omega}; \boldsymbol{x}_t), \xi(\boldsymbol{\omega}; \boldsymbol{x}_t)$. For both steps, we take a similar approach to Section 6.3.2 and use GAMs to capture these covariate relationships. To simplify our approach, we falsely assume that the atmospheric covariate A_t is continuous over t; this step allows us to utilise GAM formulations containing smooth basis functions. Given the significant variability in A_t between months, discrete formulations for this covariate would significantly increase the number of model parameters and result in higher variability.

Let $\log(v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t)) = \psi_v(\boldsymbol{x}_t)$, $\log(\sigma(\boldsymbol{\omega}; \boldsymbol{x}_t)) = \psi_\sigma(\boldsymbol{x}_t)$ and $\xi(\boldsymbol{\omega}; \boldsymbol{x}_t) = \psi_\xi(\boldsymbol{x}_t)$ denote the GAM formulations of each function, where ψ_- denotes the basis representation of equation (6.3.6). Exact forms of basis functions are specified in Section 6.4.3. As in Section 6.3.2, model fitting is carried out using the `evgam` software package (Youngman, 2022). For the first stage, $v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t)$ is estimated by exploiting a link between the loss function typically used for quantile regression and the asymmetric Laplace distribution (Yu and Moyeed, 2001). The spline coefficients associated with ψ_σ and ψ_ξ are estimated subsequently using the obtained threshold exceedances.

Selection of GAM formulations and diagnostics

Prior to estimation of the threshold and parameter functions, we specify a quantile level τ and formulations for each of the GAMs. To begin, we fix $\tau = 0.9$ and consider a variety of formulations for each ψ_v, ψ_σ and ψ_ξ. By comparing metrics for model selection, namely AIC, BIC and CRPS, we found the following formulations to be

sufficient

$$\psi_v(\boldsymbol{x}_t) = \beta_u + s_v(a_t) + \beta_s \mathbb{1}(s_t = 2), \quad \psi_\sigma(\boldsymbol{x}_t) = \beta_\sigma + s_\sigma(a_t) \quad \text{and} \quad \psi_\xi(\boldsymbol{x}_t) = \beta_\xi, \tag{6.4.12}$$

for parts 1 and 2, where $\beta_u, \beta_\sigma, \beta_\xi \in \mathbb{R}$ denote constant intercept terms, $\mathbb{1}$ denotes the indicator function with corresponding coefficient $\beta_s \in \mathbb{R}$, and s_u, s_σ denote cubic regression splines of dimension 10. The shape parameter is set to constant for the reasons outlined in Section 6.3.2. Cubic basis functions are used for ψ_v and ψ_σ since they have several desirable properties, including continuity and smoothness (Wood, 2017). A dimension of size 10 appears more than sufficient to capture the trends relating to the atmosphere variable. Alternative formulations were tested for both parts, but this made little difference to the resulting model fits.

We remark that the seasonal covariate is only present with the formulation for ψ_v. Once accounted for in the non-stationary threshold, the seasonal covariate appeared to have little influence on the fitted GPD parameters. More complex GAM formulations were tested involving interaction terms between the seasonal and atmospheric covariates, which showed little to no improvement in model fits. Thus, we prefer the simpler formulations on the basis of parsimony.

With GAM formulations selected, we now consider the quantile level $\tau \in (0, 1)$. To assess sensitivity in our formulation, we set $\mathrm{T} := \{0.8, 0.81, \ldots, 0.99\}$ and fit the GAMs outlined in equation (6.4.12) for each $\tau \in \mathrm{T}$. Letting $\delta_{\boldsymbol{\omega},t}$ and $\mathcal{T}_\tau := \{t \in \{1, \ldots, n\} \mid \delta_{\boldsymbol{\omega},t} > v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t)\}$ denote the min-projection observations and indices of threshold-exceeding observations, respectively, we expect the set $\mathcal{E} := \{-\log\{1 - F_{GPD}(\delta_{\boldsymbol{\omega},t} - v_\tau(\boldsymbol{\omega}, \boldsymbol{x}_t)) \mid \sigma(\boldsymbol{\omega}; t \in \mathcal{T}_\tau\}$ to follow a standard exponential distribution.

With all exceedances transformed to a unified scale, we compare the empirical and model exponential quantiles using QQ plots, through which we assess the relative performance of each $\tau \in \mathrm{T}$. We selected τ values for which the empirical and theoretical quantiles appeared most similar in magnitude. From this analysis, we set $\tau = 0.83$ and $\tau = 0.85$ for parts 1 and 2, respectively. The corresponding QQ plots are given in Fig-

ure 6.4.4, where we observe reasonable agreement between the empirical and theoretical quantiles. However, whilst these values appeared optimal within T, we stress that adequate model fits were also obtained for other quantile levels, suggesting our modelling procedure is not particularly sensitive to the exact choice of quantile. Furthermore, we also tested a range of quantile levels below the 0.8-level, but were unable to improve the quality of model fits.

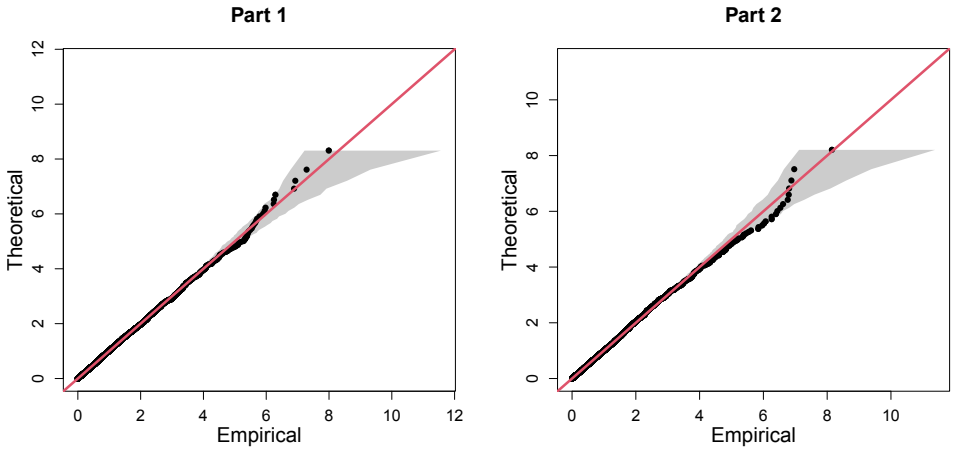


Figure 6.4.4: Final QQ plots for parts 1 (left) and 2 (right) of C3, with the $y = x$ line given in red. In both cases, the grey regions represent the 95% bootstrapped tolerance bounds.

Plots illustrating the estimated GPD scale parameter functions are given in the Supplementary Material, with the resulting dependence trends in agreement with the observed trends from Section 6.4.1. We also remark that the estimated GPD shape parameters obtained for parts 1 and 2 were $0.042\,(0.01, 0.075)$ and $0.094\,(0.059, 0.128)$, respectively, where the brackets denote 95% confidence intervals obtained using posterior sampling (Wood, 2017). These estimates, which indicate slightly heavy-tailed behaviour within the min-projection variable, provide insight into why the original exponential modelling framework is not appropriate for C3.

Overall, these results suggest different extremal dependence trends exist for the two simplex points $\boldsymbol{\omega}^{(1)}$ and $\boldsymbol{\omega}^{(2)}$, illustrating the importance of the flexibility in our model. These findings are also in agreement with empirical trends observed in Section 6.4.1,

suggesting our modelling framework is successfully capturing the underlying extremal dependence structures.

6.4.4 Results

Given estimates of threshold and parameter functions, probability estimates can be obtained via Monte Carlo techniques. Taking p_1, for instance, we have

$$\begin{aligned} p_1 &= \Pr(Z_1 > \tilde{y}, Z_2 > \tilde{y}, Z_3 > \tilde{y}) \\ &= \Pr\left(\min\left(Z_1/\omega_1^{(1)}, Z_2/\omega_2^{(1)}, Z_3/\omega_3^{(1)}\right) > r^{(1)}\right) \\ &= \int_{\boldsymbol{X}_t} \Pr\left(T_{\boldsymbol{\omega}^{(1)},\,t} > r^{(1)} \mid \boldsymbol{X}_t = \boldsymbol{x}_t\right) f_{\boldsymbol{X}_t}(\boldsymbol{x}_t)\mathrm{d}\boldsymbol{x}_t \\ &= (1-\tau)\int_{\boldsymbol{X}_t} \Pr(T_{\boldsymbol{\omega}^{(1)},\,t} > r^{(1)} \mid T_{\boldsymbol{\omega}^{(1)},\,t} > v_\tau(\boldsymbol{\omega}^{(1)}, \boldsymbol{x}_t), \boldsymbol{X}_t = \boldsymbol{x}_t) f_{\boldsymbol{X}_t}(\boldsymbol{x}_t)\mathrm{d}\boldsymbol{x}_t \\ &\approx \frac{1-\tau}{n}\sum_{t=1}^{n}\left(1 + \frac{\xi(\boldsymbol{\omega}^{(1)}; \boldsymbol{x}_t)\left(r^{(1)} - v_\tau(\boldsymbol{\omega}^{(1)}, \boldsymbol{x}_t)\right)}{\sigma\left(\boldsymbol{\omega}^{(1)}; \boldsymbol{x}_t\right)}\right)^{-1/\xi\left(\boldsymbol{\omega}^{(1)}; \boldsymbol{x}_t\right)}, \end{aligned}$$

assuming $\{\boldsymbol{x}_t : t \in \{1, \ldots, n\}\}$ is a representative sample from $\boldsymbol{X}_t$. The procedure for p_2 is analogous. We note that this estimation procedure is only valid when $r^{(1)} > v_\tau(\boldsymbol{\omega}^{(1)}, \boldsymbol{x}_t)$, or $r^{(2)} > v_\tau(\boldsymbol{\omega}^{(2)}, \boldsymbol{x}_t)$, for all t: however, for each $\tau \in \mathrm{T}$, this inequality is always satisfied, owing to the very extreme nature of the probabilities in question. Through this approximation, we obtain $\hat{p}_1 = 1.480 \times 10^{-5}$ and $\hat{p}_2 = 2.461 \times 10^{-5}$.

6.5 Challenge C4

6.5.1 Exploratory data analysis

Challenge C4 entails estimating survival probabilities across 50 locations on the island of Utopula. As stated in Rohrbeck et al. (2023), the Utopula island is split in two administrative areas, for which the respective regional governments 1 and 2 have collected data concerning the variables $Y_{i,t}$, $i \in I = \{1, \ldots, 50\}$, $t \in \{1, \ldots, 10,000\}$. Index i denotes the i^{th} location, with locations $i \in \{1, \ldots, 25\}$ and $i \in \{26, \ldots, 50\}$ belonging

to the administrative areas of governments 1 and 2, respectively. Index t denotes the time point in days; however, since $Y_{i,t}$ are IID for all i, we drop the subscript t for the remainder of this section.

Since many multivariate extreme value models are only applicable in low-to-moderate dimensions, we consider dimension reduction based on an exploration of the extremal dependence structure of the data. In particular, we analyse pairwise estimates of the extremal dependence coefficient $\chi(u)$, introduced in equation (6.2.2), for all possible pairwise combinations of sites; the resulting estimates, using $u = 0.95$, are presented in the heat map of Figure 6.5.5. Identification of any dependence clusters is achieved through visual investigation, which seems appropriate for this data. We note, however, that should visual considerations not suffice, alternative more sophisticated clustering methods are available and can be applied; see for example Bernard et al. (2013).

Figure 6.5.5 suggests the existence of 5 distinct subgroups where all variables within each subgroup have similar extremal dependence characteristics, while variables in different subgroups appear to be approximately independent of each other in the extremes. It is worth mentioning that the same clusters are identified when we analyse pairwise estimates of the extremal dependence coefficient $\eta(u)$; the resulting estimates can be found in the Supplementary Material. Moreover, examining the magnitudes of $\chi(\cdot)$ and $\eta(\cdot)$ estimates, it does not appear reasonable to assume asymptotic dependence between variables in the same group. We therefore consider models that can be applied to data structures that do not take their extreme values simultaneously. The indices of the five aforementioned subgroups are $G_1 = \{4, 14, 19, 28, 30, 38, 43, 44\}$, $G_2 = \{3, 10, 15, 18, 22, 29, 45, 47\}$, $G_3 = \{8, 21, 25, 26, 32, 33, 34, 40, 41, 42, 48, 49, 50\}$, $G_4 = \{1, 2, 5, 7, 9, 17, 20, 31, 46\}$ and $G_5 = \{6, 11, 12, 13, 16, 23, 24, 27, 35, 36, 37, 39\}$. Groups G_1 and G_2 include the most strongly dependent variables (shown by the darkest color blocks in Figure 6.5.5), followed by group G_3, while groups G_4 and G_5 contain the most weakly dependent variables. We henceforth assume independence between

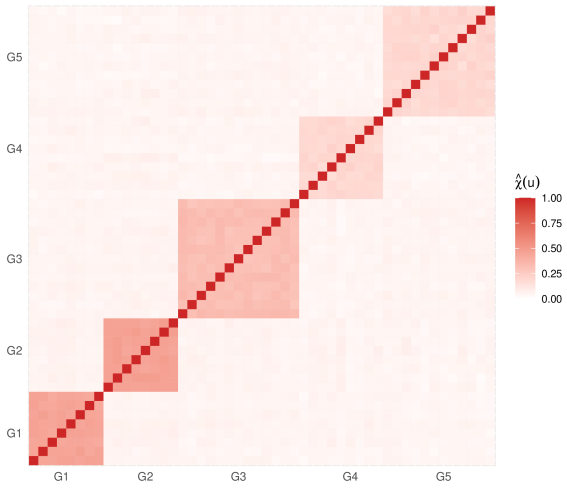


Figure 6.5.5: Heat map of estimated empirical pairwise $\chi(u)$ extremal dependence coefficients with $u = 0.95$.

these groups of variables, i.e., $\Pr((Y_i)_{i\in G_k} \in A_k, (Y_i)_{i\in G_{k'}} \in A_{k'}) = \Pr((Y_i)_{i\in G_k} \in A_k) \Pr((Y_i)_{i\in G_{k'}} \in A_{k'})$, $A_k \subset \mathbb{R}^{|G_k|}, A_{k'} \subset \mathbb{R}^{|G_{k'}|}$, for any $k \neq k' \in \{1, \ldots, 5\}$.

Challenge C4 requires us to estimate the probabilities $p_1 = \Pr(Y_i > s_i;\ i \in I)$ and $p_2 = \Pr(Y_i > s_1;\ i \in I)$, where $s_i := \mathbb{1}(i \in \{1, 2, \ldots, 25\})s_1 + \mathbb{1}(i \in \{26, 27, \ldots, 50\})s_2$ and s_1 (s_2) denotes the marginal level exceeded once every year (month) on average. Under the assumption of independence between groups, the challenge can be broken down to 5 lower-dimensional challenges involving the estimation of joint tail probabilities for each G_k, $k \in \{1, \ldots, 5\}$. These can then be multiplied together to obtain the required overall probabilities due to (assumed) between-group independence; specifically, we have $p_1 = \prod_{k=1}^{5} \Pr(Y_i > s_i;\ i \in G_k)$ and $p_2 = \prod_{k=1}^{5} \Pr(Y_i > s_1;\ i \in G_k)$.

6.5.2 Conditional extremes

The conditional multivariate extreme value model (CMEVM) of Heffernan and Tawn (2004) provides a flexible multivariate extreme value framework capable of capturing a range of extremal dependence forms without making assumptions about the specific form of joint dependence structure. Consider a d-dimensional random variable $\boldsymbol{W} = (W_1, \ldots, W_d)$ on standard Laplace margins. For $i \in \{1, \ldots, d\}$, the CMEVM approach

assumes the existence of parameter vectors $\boldsymbol{\alpha}_{-|i} \in [-1,1]^{d-1}$ and $\boldsymbol{\beta}_{-|i} \in (-\infty,1]^{d-1}$ such that

$$\lim_{u_i \to \infty} \Pr\left\{ \boldsymbol{W}_{-i} \leq \boldsymbol{\alpha}_{-|i} W_i + W_i^{\boldsymbol{\beta}_{-|i}} \boldsymbol{z}_{|i}, W_i - u_i > w \mid W_i > u_i \right\} = e^{-w} H_{|i}\left(\boldsymbol{z}_{|i}\right), \quad w > 0,$$

with non-degenerate distribution function $H_{|i}(\cdot)$, vector operations being applied componentwise, and conditional threshold u_i. The vector $\boldsymbol{W}_{-i}$ denotes $\boldsymbol{W}$ excluding its i^{th} component and $\boldsymbol{z}_{|i}$ is within the support of the residual random vector $\boldsymbol{Z}_{|i} = (\boldsymbol{W}_{-i} - \boldsymbol{\alpha}_{-|i} w_i)/w_i^{\boldsymbol{\beta}_{-|i}} \sim H_{|i}(\cdot)$. We apply this model to data where $W_i > u_i$, for some finite conditioning threshold u_i, to estimate the probabilities p_1 and p_2 defined in Section 6.5.1, using the inference procedure of Keef et al. (2013).

6.5.3 Results

Let $\boldsymbol{W} := (W_1, \ldots, W_{50})$ denote the random vector after transformation to standard Laplace margins. This vector is divided into the five subgroups identified in Section 6.5.1, and the subgroup probabilities are estimated using predictions obtained from the sampling method of Heffernan and Tawn (2004). We condition on the first variable of each subgroup being extreme, and simulate 10^8 predictions from each of the resulting fitted conditional extremes models. To account for uncertainty in the estimates, we perform a parametric bootstrapping procedure with 100 samples.

Sensitivity analyses of the estimated probabilities to the choice of conditioning variable suggest no significant effect. Furthermore, we consider a range of conditioning thresholds; the corresponding estimates of subgroup probabilities defined in Section 6.5.1 appear relatively stable with respect to the conditioning threshold quantile. We ultimately select 0.85-quantiles for the conditioning thresholds of our final probability estimates. These are given by $\hat{p}_1 = 1.094 \times 10^{-26}$ $(2.150 \times 10^{-36}, 1.359 \times 10^{-24})$ and $\hat{p}_2 = 1.076 \times 10^{-31}$ $(1.596 \times 10^{-46}, 1.850 \times 10^{-29})$, with 95% confidence intervals obtained from parametric bootstrapping given in parentheses.

6.6 Discussion

In this paper, we have proposed a range of statistical methods for estimating extreme quantities for challenges C1-C4. For the univariate challenge C1, we estimated the 0.9999-quantile, and the associated 50% confidence intervals, of $Y \mid \boldsymbol{X} = \boldsymbol{x_i}$, $i \in \{1, \ldots, n\}$. For challenge C2, we estimated a quantile, corresponding to a once in 200 year level, of the marginal distribution Y whilst incorporating the loss function in equation (6.3.4). Overall we ranked 6^{th} and and 4^{th} for challenges C1 and C2, respectively.

For challenge C1, our final model (model 7 in Table 6.3.1) was chosen to minimise the model selection criteria; however, QQ plots showed over-estimation of the most extreme values of the response (see Figure 6.3.2). As a result, the conditional quantiles calculated for C1 are generally over-estimated when compared with the true quantiles. If we ignored the model selection criteria and chose the model based on a visual assessment of QQ plots, we would have chosen model 5 in Table 6.3.1 and this would have covered the true quantile on fewer occasions than our chosen model. Therefore, the main issue with our results concerns the width of the confidence intervals.

Narrow confidence intervals are an indication of over-fitting and this could have arisen in several places. For instance, Rohrbeck et al. (2023) suggested all the seasonality is captured in the threshold, while our model includes a seasonal threshold and a covariate for seasonality in the scale parameter of the GPD model. As well as over-fitting, the model may not have been flexible enough; this could be, in part, due to our model missing covariates. For instance, the true model contained V_2 as a covariate (Rohrbeck et al., 2023) whilst our model did not. In addition, the basis dimensions for our splines are low. In practice, a higher dimension than we would expect should be considered and, although we chose the dimension using a model-based approach, it may have resulted in the splines not being flexible enough to capture all of the trends in the data.

Narrow confidence intervals may have also resulted from the choice of uncertainty quantification procedure. Changing the average block length l in our stationary bootstrap procedure would alter the confidence interval widths, although this was carefully chosen to reflect the temporal dependence in the data. Alternative methods, such as the standard bootstrap procedure or the delta method, could be implemented to investigate how this affects the confidence interval widths. We expect that such confidence intervals will be wider than those presented here since the dependence in the data is not accounted for, but assuming temporal independence would be inaccurate. Therefore, whilst adopting an alternative procedure may widen confidence intervals, thus improving our performance, such intervals may not be well calibrated for this data set.

The over-fitting and over-estimation issues encountered in C1 are carried through to C2 since the same model is used for both challenges. However, one aspect specific to C2 is the choice of quantile evaluation within the loss function. Many methods exist for evaluating the non-stationary quantiles which feed into the loss function term of the objective function $S(\boldsymbol{\theta})$ in equation (6.3.7). As the loss function will be dominated by the log-likelihood in $S(\boldsymbol{\theta})$, we chose to transform to standard exponential margins when evaluating the quantiles in order to give more importance to the loss function. Since the data is light tailed ($\xi < 0$) this transformation elongates the tail and therefore inflates any deviations between the model and theoretical quantiles which in turn, inflates the contribution of the average loss function to $S(\boldsymbol{\theta})$. However, this approach means that the objective function will have a preference to minimise the deviations in the upper-tail of the distribution, leading to potential over-fitting to the upper-tail and possibly, a poor fit in the rest of the tail. This may not necessarily be undesirable since the loss function penalises under-estimation more than over-estimation, however, since the model in C1 already over-fits, this method may only exacerbate the problem for C2.

For the first multivariate challenge C3, we employed an extension of the method proposed by Wadsworth and Tawn (2013) to estimate probabilities of three variables

lying in extremal sets. Our extension accounts for non-stationarity in the extremal dependence structure, with GAMs used to represent covariate relationships. The QQ plots for the resulting model suggested reasonable fits. For this challenge, we ranked 5^{th} and our estimates are on the same order of magnitude as the truth (Rohrbeck et al., 2023).

We note similarities in the methodologies presented for the challenges C1, C2, and C3. Specifically, each of the proposed methods used the EVGAM framework for capturing non-stationary tail behaviour via a generalised Pareto distribution. We acknowledge that the model selection tool proposed for C1 and C2 could also be applied for C3. However, we opted not to use this tool for several reasons. Firstly, unlike the univariate setting, there is no guarantee of convergence to a GPD in the limit, and the GPD tail assumption thereby needs to be tested. Moreover, in exploratory analysis, we tested the model selection tool for C3 but found the selected models and quantiles to not be satisfactory, particularly in the upper tail of the min-projection variable. We therefore selected a model manually, using QQ plots to evaluate performance. Exploring threshold and model selection techniques for multivariate extremes represents an important area of research.

In the final multivariate challenge C4, we estimated very high-dimensional joint survival probabilities. To do so, we split the probability into 5 lower-dimensional components which are assumed independent of each other, then estimated each using the CMEVM of Heffernan and Tawn (2004). In the final rankings of Rohrbeck et al. (2023), we ranked 3^{rd} for this challenge. A more prudent method could have been implemented, as groups of variables were never truly independent. Alternatively, although we achieve relatively stable probability estimates with respect to threshold in Section 5.2 (see Supplementary Material for details), our approach could potentially have been improved by estimating individual group probabilities across varying thresholds and taking an average value as our final result. We also do not report the effect of the choice of

the conditioning variable on our estimates. Preliminary analysis suggested this to be negligible. However, conditioning on each site in a given subgroup and then taking a weighted sum of the resulting probabilities (e.g., Keef et al., 2013) may have resulted in more robust estimates.

6.7 Supplementary material

6.7.1 Additional figures for Section 6.3

In this section, we present additional figures for Section 6.3 of the main paper, concerned with challenges C1 and C2. Figures 6.7.6-6.7.8 support the exploratory analysis for challenges C1 and C2. We explore the within-year seasonality of the response variable Y in Figure 6.7.6, looking at the distribution of Y per month and across the two seasons. This shows that there is a significant difference in the distribution of Y between seasons 1 and 2, but within each season there is little difference across months.

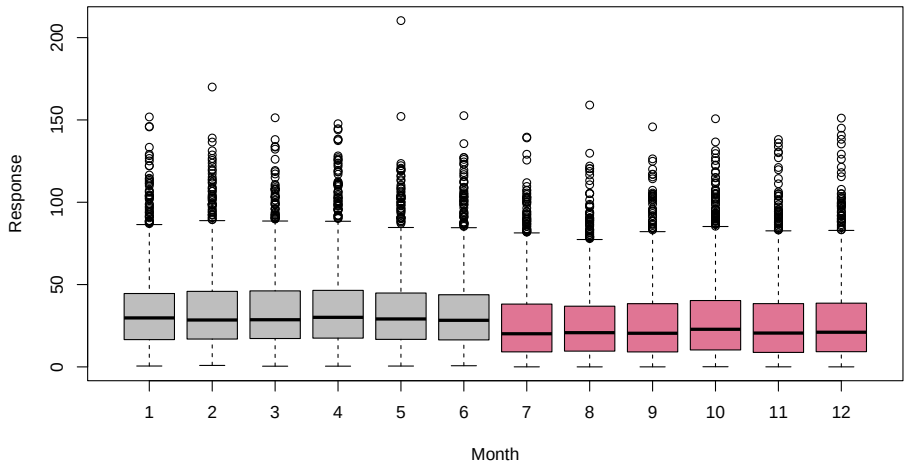


Figure 6.7.6: Box plot of the response variable Y with each month and season (season 1 in grey and season 2 in red).

Figure 6.7.7 shows a scatter plot of Y against each covariate $V_1, \ldots, V_8$, excluding V_6 which corresponds to season. Covariates V_1, V_2 and V_8 do not seem to have a relationship

with Y, whilst there seems to be dependence for the remaining covariates. These observed relationships appear complex and non-linear.

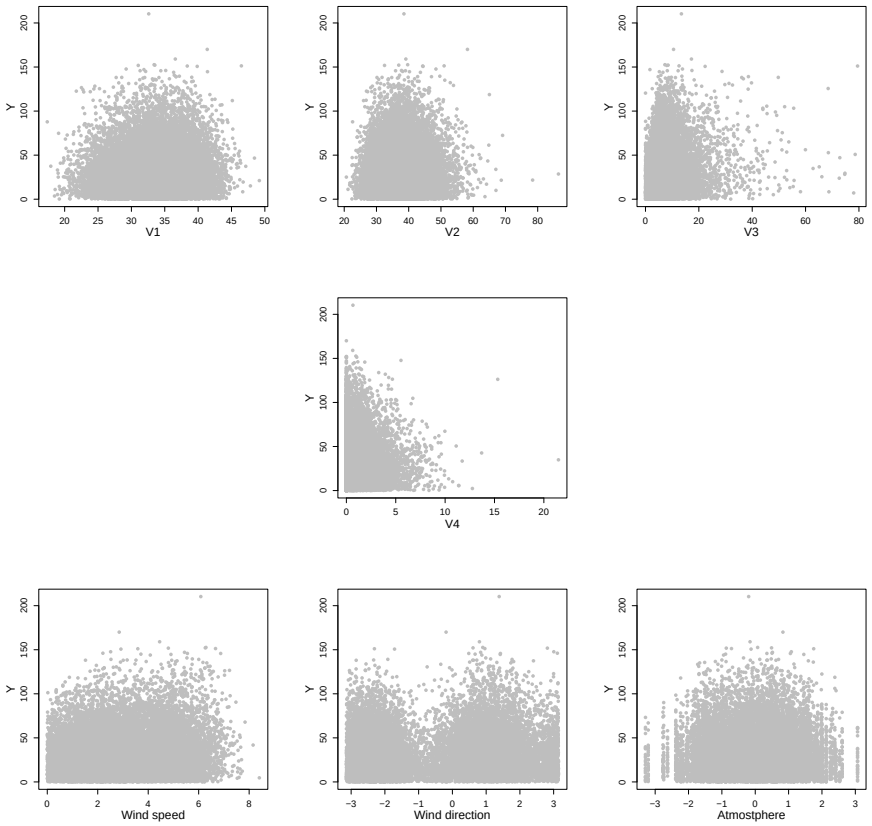


Figure 6.7.7: Scatter plots of explanatory variables $V_1, \ldots, V_4$, wind speed (V_6), wind direction (V_7) and atmosphere (V_8), from top-left to bottom-right (by row), against the response variable Y.

We also explore temporal dependence in Figure 6.7.8 that details the auto-correlation function (acf) values for the response Y and explanatory variables $V_1, \ldots, V_4, V_6, \ldots, V_8$, up to a lag of 60. All variables have negligible acf values beyond lag 0, except V_6 (wind speed), V_7 (wind direction) and V_8 (atmosphere).

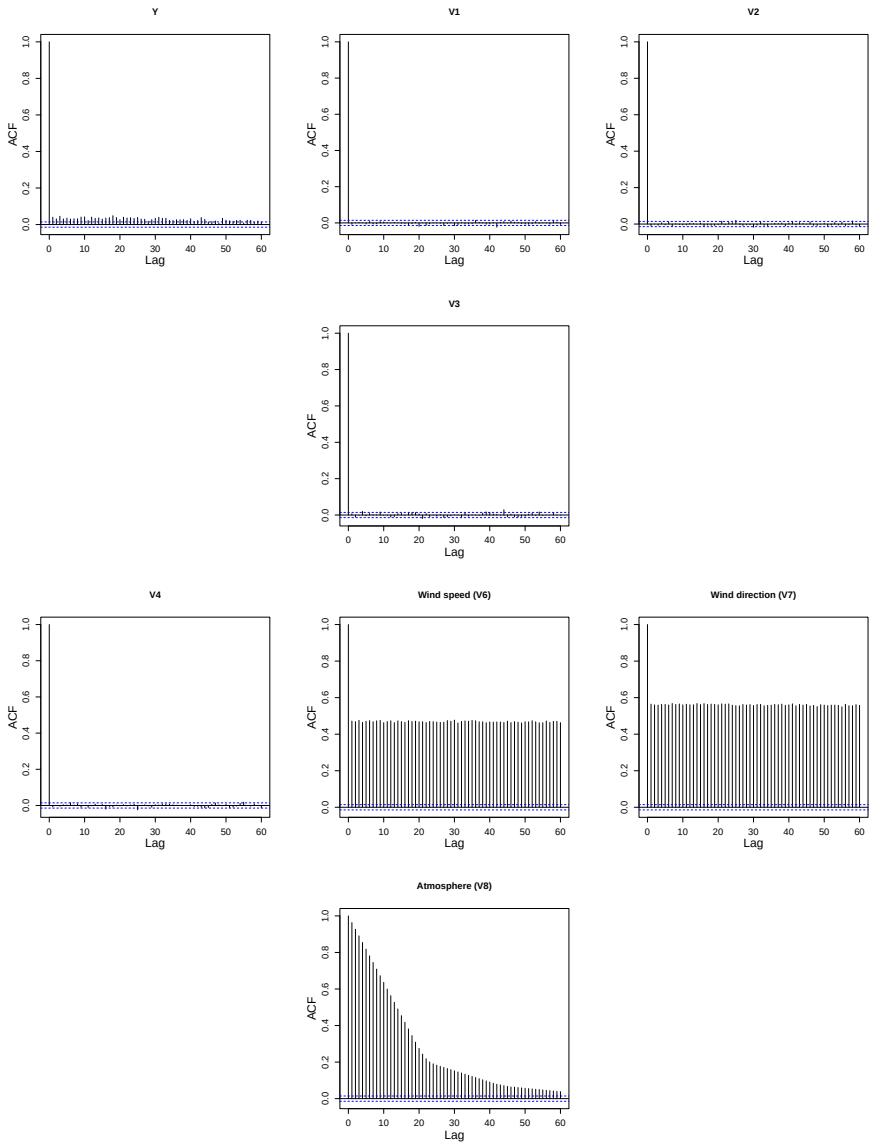


Figure 6.7.8: Autocorrelation function plots for the response variable Y and explanatory variables $V1, \ldots, V4$, wind speed ($V6$), wind direction ($V7$) and atmosphere ($V8$), from top-left to bottom-right (by row).

Figure 6.7.9 shows the QQ-plots corresponding to a standard GPD model fitted to the excesses of Y above a constant (left) and seasonally-varying threshold (right). 95% tolerance bounds (grey) show a lack of agreement between observations and the standard GPD model above a constant threshold. The second plot demonstrates a significant improvement in model fit.

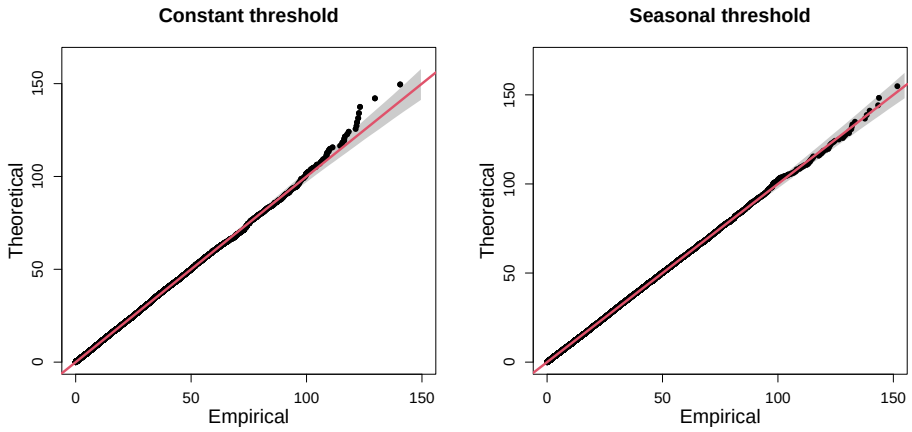


Figure 6.7.9: QQ-plots showing standard GPD model fits with 95% tolerance bounds (grey) above a constant (left) and stepped-seasonal (right) threshold.

Figure 6.7.10 shows a detailed summary of the pattern of missing data in the data and can be produced using the `missing_pattern` function in the `finalfit` package in `R` (Harrison et al., 2023). To interpret the figure note that blue and red squares represent observed and missing variables, respectively. The number on the right indicates the number of missing predictor variables (i.e., the number of red squares in the row), while the number on the left is the number of observations that fall into the row category. On the bottom, we have the number of observations that fall into the column category. For example, 18,545 observations are fully observed (denoted by the first row); there are 407 observations where only $V4$ is missing (denoted by the second row), 13 observations where both $V4$ and $V6$ are missing (denoted by the fourth row), 456 observations where $V4$ and at least one other predictor is missing (denoted by the last column in the table), etc. It can be seen that there are very few observations where more than one predictor is missing.

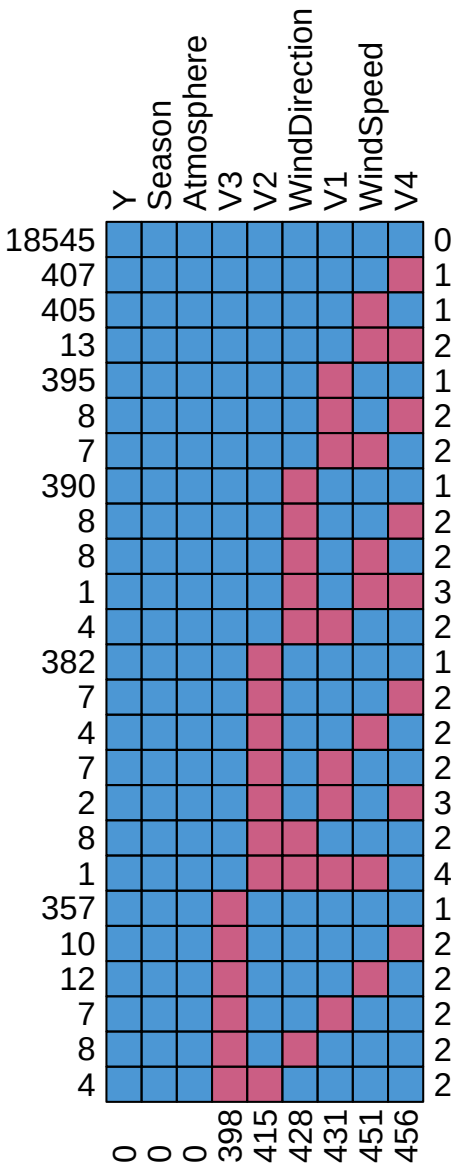


Figure 6.7.10: Detailed pattern of missing predictor variables in the Amaurot data set.

6.7.2 Additional figures for Section 6.4

In this section, we present additional plots related to Section 6.4 of the main article. Figure 6.7.11 illustrates the time series of both covariates for the first 3 years of the observation period. It can be seen how the seasons vary periodically over each year, as well as the discrete nature of the atmospheric covariate.

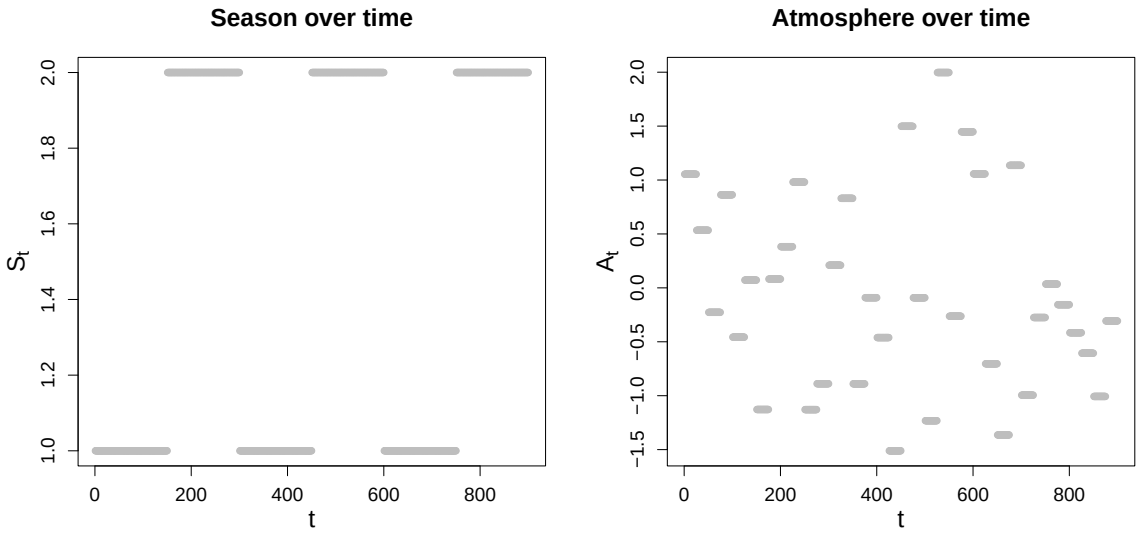


Figure 6.7.11: Plots of S_t (left) and A_t (right) against t for the first 3 years of the observation period.

Bootstrapped χ estimates for the groups $G^A_{I,k}, k \in \{1,\dots,10\}$, $I \in \mathcal{I} \setminus \{1,2,3\}$ and $G^S_{I,k}, k \in \{1,2\}$, $I \in \mathcal{I}$ are given in Figures 6.7.12 - 6.7.15. These estimates illustrate the impact of atmosphere on the dependence structure.

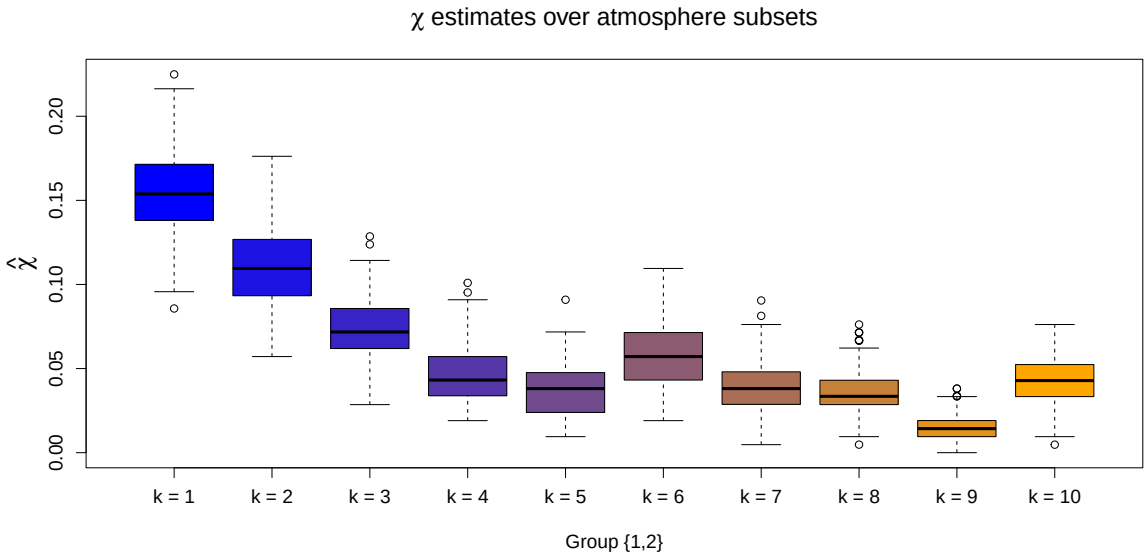


Figure 6.7.12: Boxplots of empirical χ estimates obtained for the subsets $G^A_{I,k}$, with $k = 1,\dots,10$ and $I = \{1,2\}$. The colour transition (from blue to orange) over k illustrates the trend in χ estimates as the atmospheric values are increased.

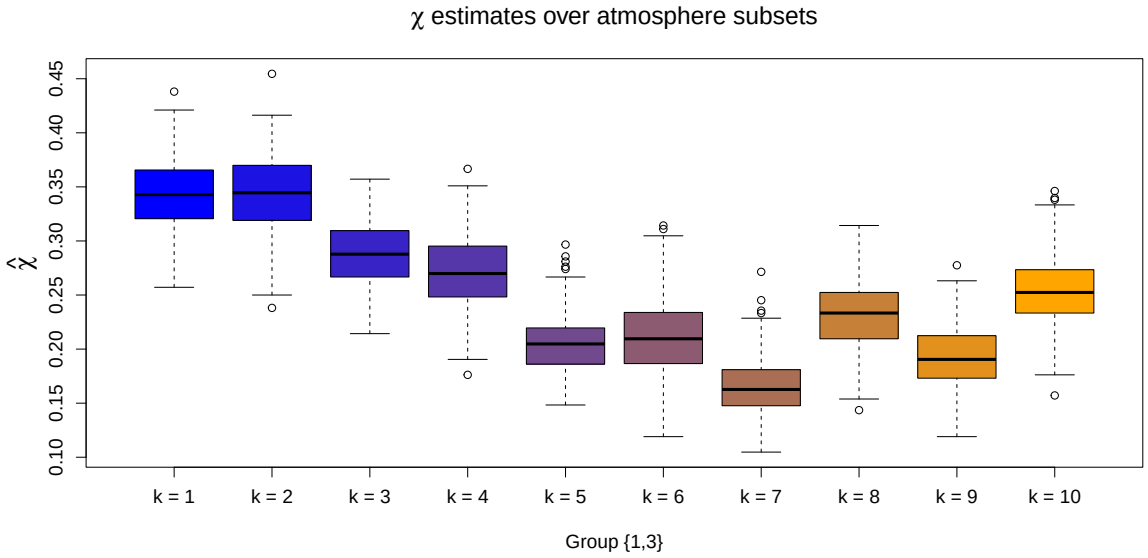


Figure 6.7.13: Boxplots of empirical χ estimates obtained for the subsets $G^A_{I,k}$, with $k = 1, \ldots, 10$ and $I = \{1, 3\}$. The colour transition (from blue to orange) over k illustrates the trend in χ estimates as the atmospheric values are increased.

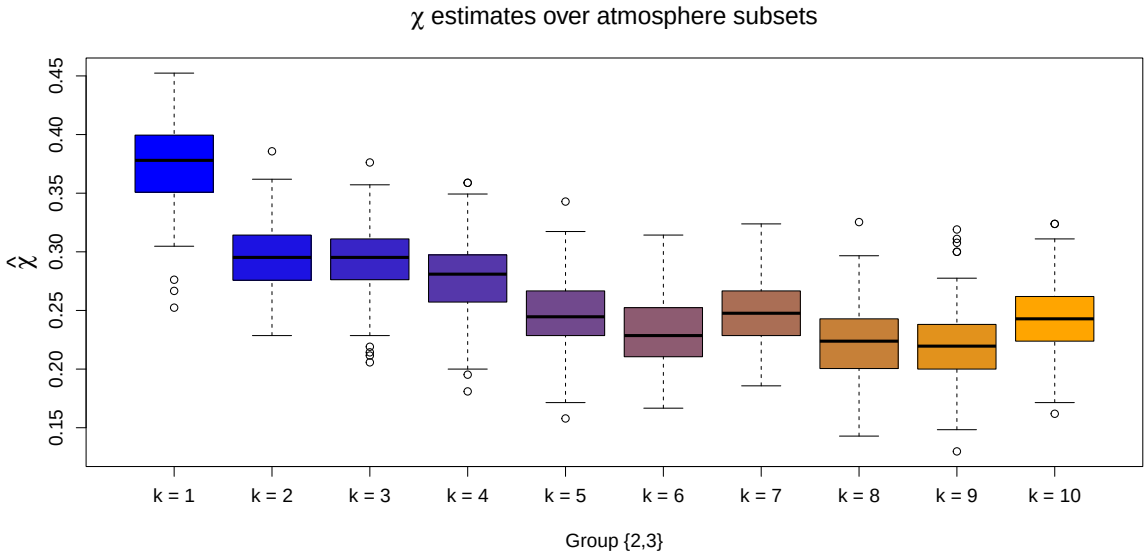


Figure 6.7.14: Boxplots of empirical χ estimates obtained for the subsets $G^A_{I,k}$, with $k = 1, \ldots, 10$ and $I = \{2, 3\}$. The colour transition (from blue to orange) over k illustrates the trend in χ estimates as the atmospheric values are increased.

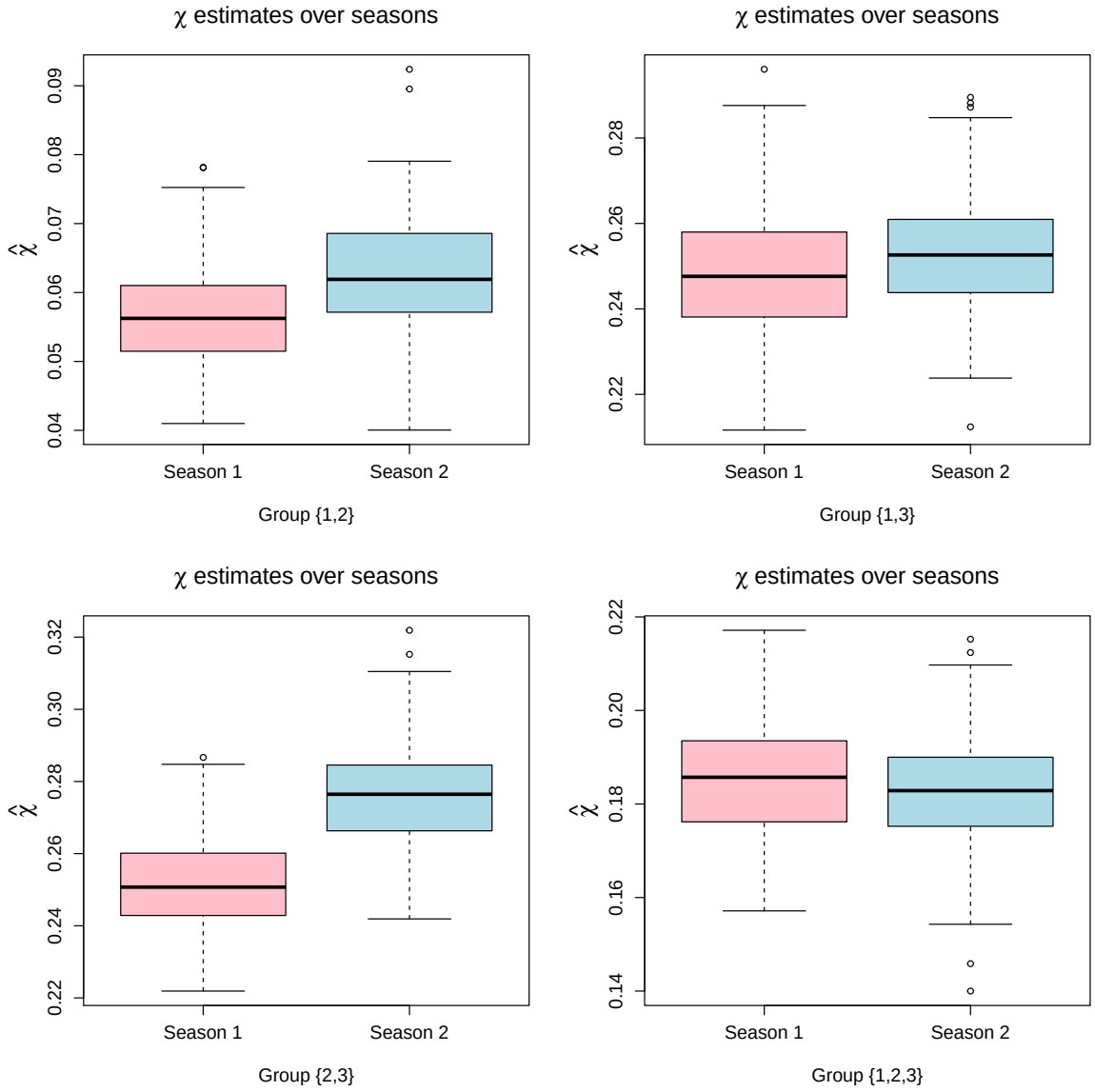


Figure 6.7.15: Boxplots of empirical χ estimates obtained for the subsets $G^S_{I,k}$, with $k = 1, 2$. In each case, pink and blue colours illustrate estimates for seasons 1 and 2, respectively. From top left to bottom right: $I = \{1, 2, 3\}$, $I = \{1, 2\}$, $I = \{1, 3\}$, $I = \{2, 3\}$.

For a 3-dimensional random vector, the angular dependence function, denoted $\lambda(\cdot)$, is defined on the unit-simplex $\boldsymbol{S}^2$ and describes extremal dependence along different rays $\boldsymbol{\omega} \in \boldsymbol{S}^2$. As noted in Section 4.2 of the main manuscript, we can associate each of the probabilities from C3, p_1 and p_2, with points on $\boldsymbol{S}^2$, denoted $\boldsymbol{\omega}^1$ and $\boldsymbol{\omega}^2$ respectively. With $I = \{1, 2, 3\}$, we consider $\lambda(\boldsymbol{\omega}^1)$ and $\lambda(\boldsymbol{\omega}^2)$ over the subsets $G^S_{I,k},\ k \in \{1, 2\}$ and $G^A_{I,k},\ k \in \{1, \ldots, 10\}$. We note that $\lambda(\boldsymbol{\omega}^1)$ is analogous with the coefficient of tail dependence $\eta \in (0, 1]$ (Ledford and Tawn, 1996), with $\eta = 1/3\lambda(\boldsymbol{\omega}^1)$; this corresponds with the region where all variables are simultaneously extreme. Furthermore, $\lambda(\boldsymbol{\omega}^2)$, which corresponds to a region where only two variables are extreme, is only evaluated after an additional marginal transformation of the third Coputopia time series; see Section 4.2 of the main manuscript.

Estimation of $\lambda(\cdot)$ for each simplex point and subset was achieved using the Hill estimator (Hill, 1975) at the 90% level, with uncertainty subsequently quantified via bootstrapping. These results are given in Figures 6.7.16 - 6.7.19. These plots provide further evidence of a relationship between the extremal dependence structure and the covariates.

To illustrate the estimated trend in dependence, Figure 6.7.20 shows the estimated scale functions, $\sigma(\boldsymbol{\omega}; \boldsymbol{x}_t)$, over atmosphere for parts 1 and 2. Under the assumption of asymptotic normality in the spline coefficients, 95% confidence intervals are obtained via posterior sampling; see Wood (2017) for more details. We observe that σ tends to increase and decrease over atmosphere for parts 1 and 2, respectively, although the trend is less pronounced for the latter. Under our modelling framework, we note that higher values of σ are associated with less positive extremal dependence in the direction $\boldsymbol{\omega}$ of interest; to see this, observe that the survivor function of the GPD with fixed ξ is negatively associated with σ. Considering the trend in $\sigma(\boldsymbol{\omega}; \boldsymbol{x}_t)$, our results indicate a decrease in dependence in the region where all variables are extreme.

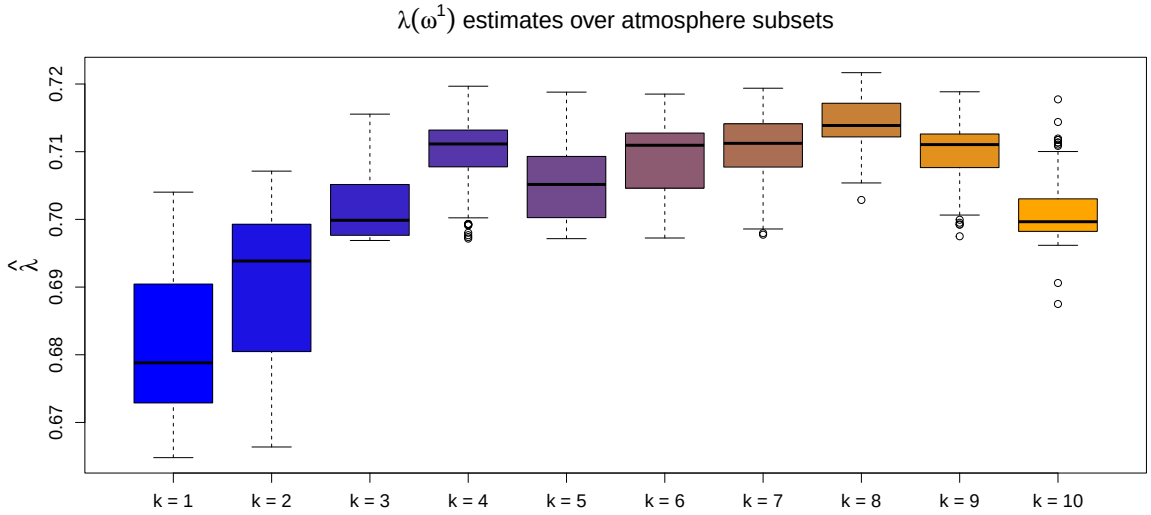


Figure 6.7.16: Boxplots of empirical $\lambda(\boldsymbol{\omega}^1)$ estimates obtained for the subsets $G^A_{I,k}$, with $k = 1, \ldots, 10$ and $I = \{1, 2, 3\}$. The colour transition (from blue to orange) over k illustrates the trend in λ estimates as the atmospheric values are increased.

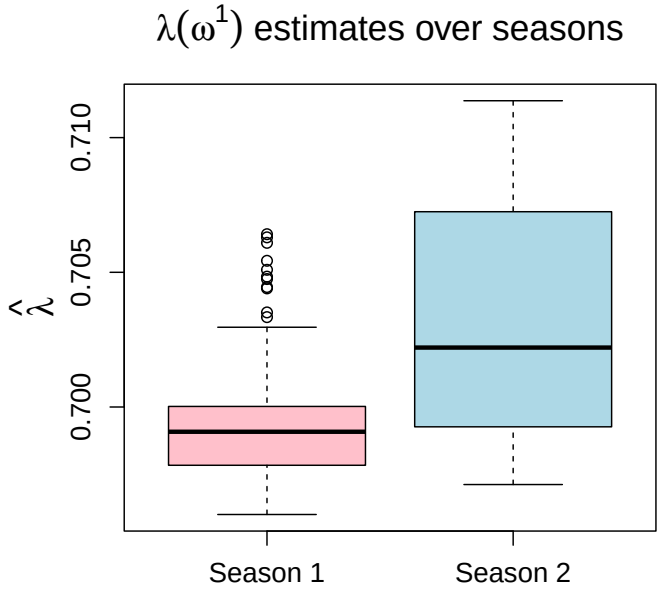


Figure 6.7.17: Boxplots of empirical $\lambda(\boldsymbol{\omega}^1)$ estimates obtained for the subsets $G^S_{I,k}$, with $k = 1, 2$ and $I = \{1, 2, 3\}$. In each case, pink and blue colours illustrate estimates for seasons 1 and 2, respectively.

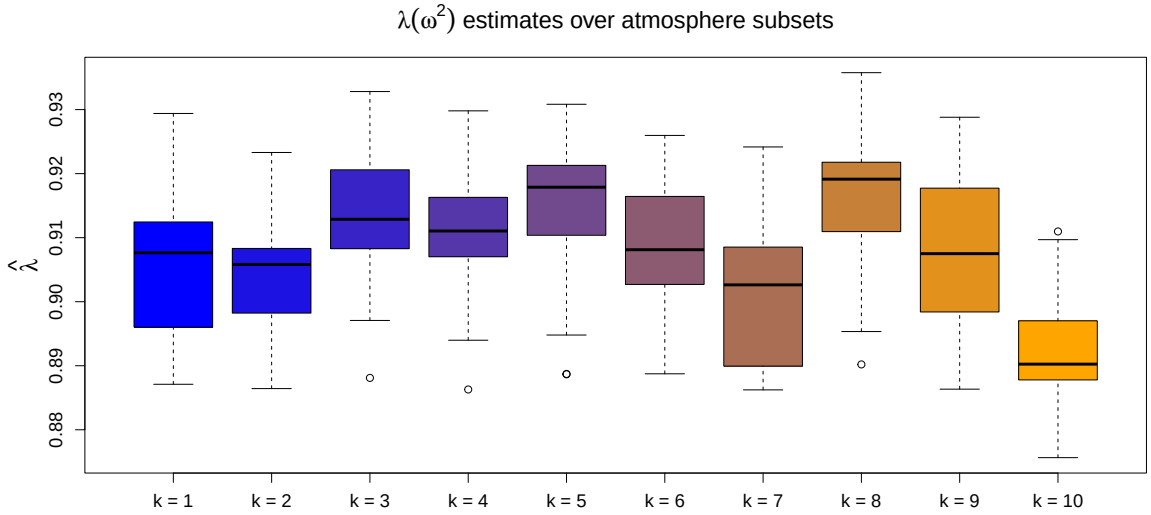


Figure 6.7.18: Boxplots of empirical $\lambda(\boldsymbol{\omega}^2)$ estimates obtained for the subsets $G^A_{I,k}$, with $k = 1, \ldots, 10$ and $I = \{1, 2, 3\}$. The colour transition (from blue to orange) over k illustrates the trend in λ estimates as the atmospheric values are increased.

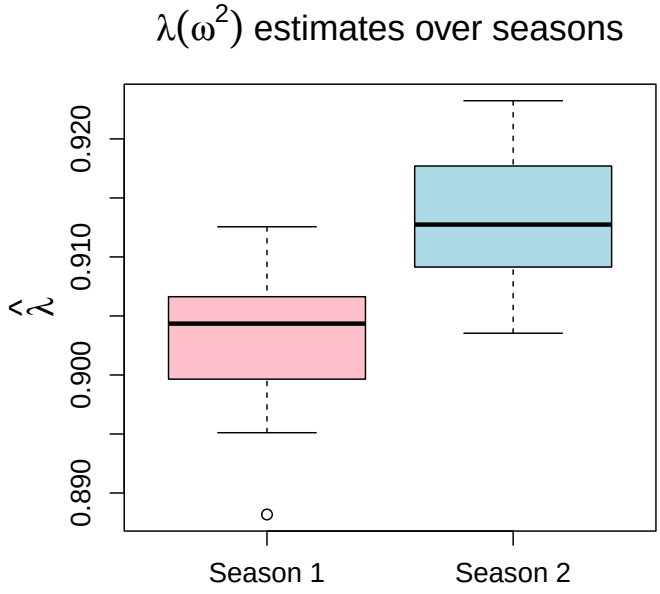


Figure 6.7.19: Boxplots of empirical $\lambda(\boldsymbol{\omega}^2)$ estimates obtained for the subsets $G^S_{I,k}$, with $k = 1, 2$ and $I = \{1, 2, 3\}$. In each case, pink and blue colours illustrate estimates for seasons 1 and 2, respectively.

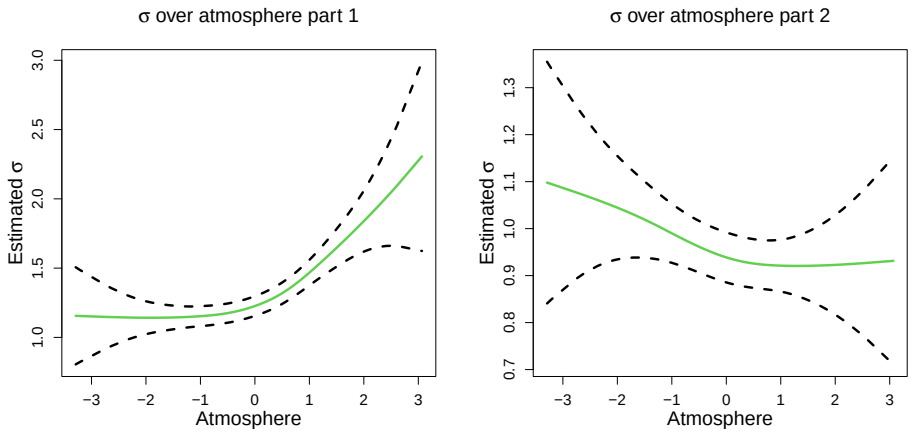


Figure 6.7.20: Estimated σ functions (green) over atmosphere for part 1 (left) and 2 (right). In both cases, the regions defined by the black dotted lines represent 95% confidence intervals obtained using posterior sampling.

6.7.3 Additional figures for Section 6.5

In this section, we present additional plots related to Section 6.5 of the main article an we refer to p_1 and p_2 as parts 1 and 2 of C4, respectively. Figure 6.7.21 shows a heat map of empirically estimated $\eta(\cdot)$ dependence coefficients and provides further evidence of the existence of the 5 dependence subgroups identified in our exploratory analysis for challenge C4. It also suggests that our modelling assumptions are reasonable; specifically that there is in-between group independence, and that the extremes within each group do not occur simultaneously.

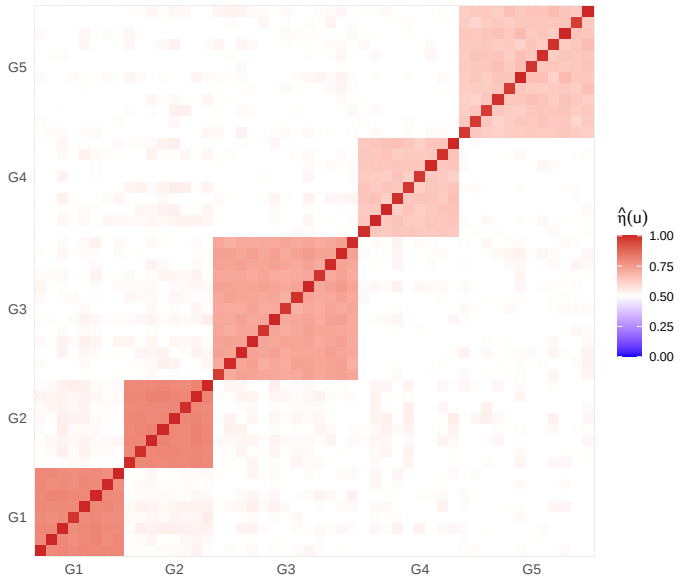


Figure 6.7.21: Heat map of estimated empirical pairwise $\eta(u)$ extremal dependence coefficients with $u = 0.95$.

Figure 6.7.22 shows the bootstrapped estimated individual group and overall probabilities with respect to conditioning threshold quantile for part 1 of challenge C4. Similarly, Figure 6.7.23 shows the bootstrapped estimated individual group and overall probabilities with respect to conditioning threshold quantile for part 2 of challenge C4.

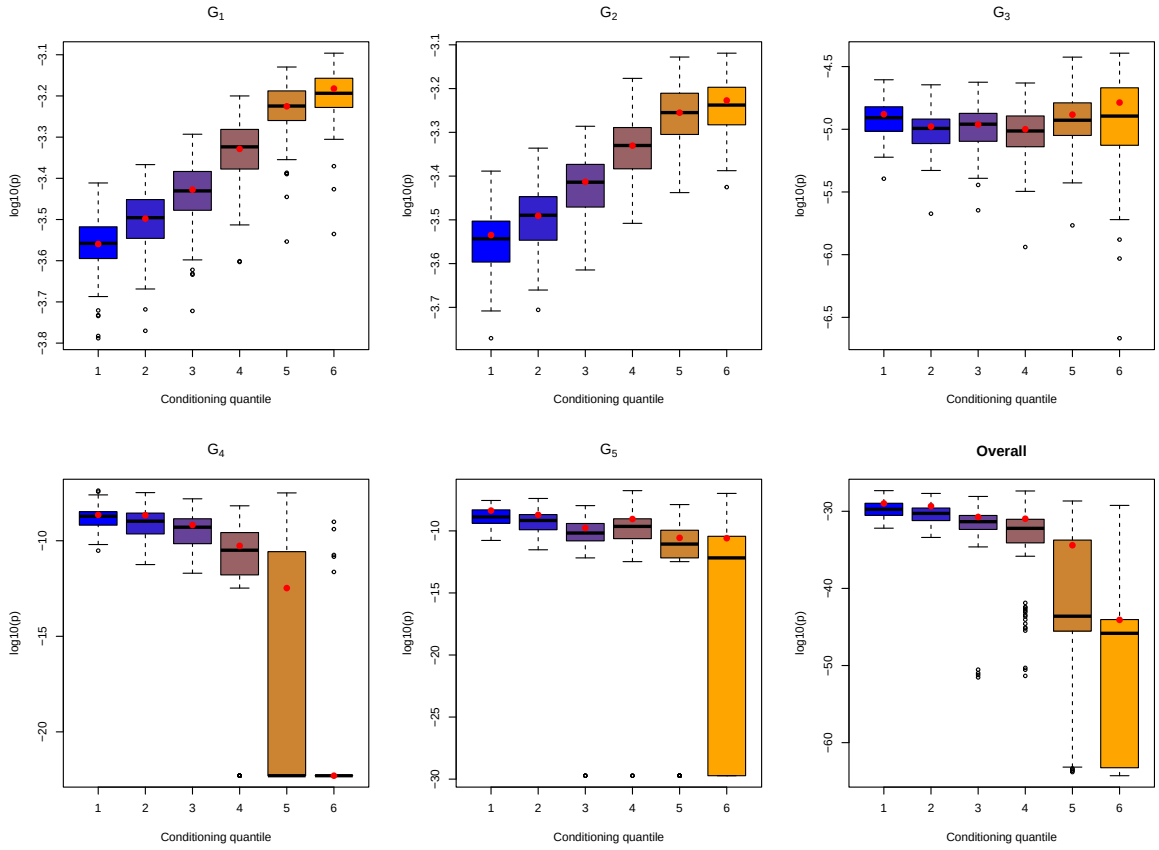


Figure 6.7.22: Part 1 subgroup and overall bootstrapped probability estimates on the log scale. The red points indicate the original sample estimates and the colouring of the boxplots indicates the choice of conditioning threshold, with the conditioning quantile indices 1-6 referring to the quantile levels $\{0.7, 0.75, 0.8, 0.85, 0.9, 0.95\}$, respectively.

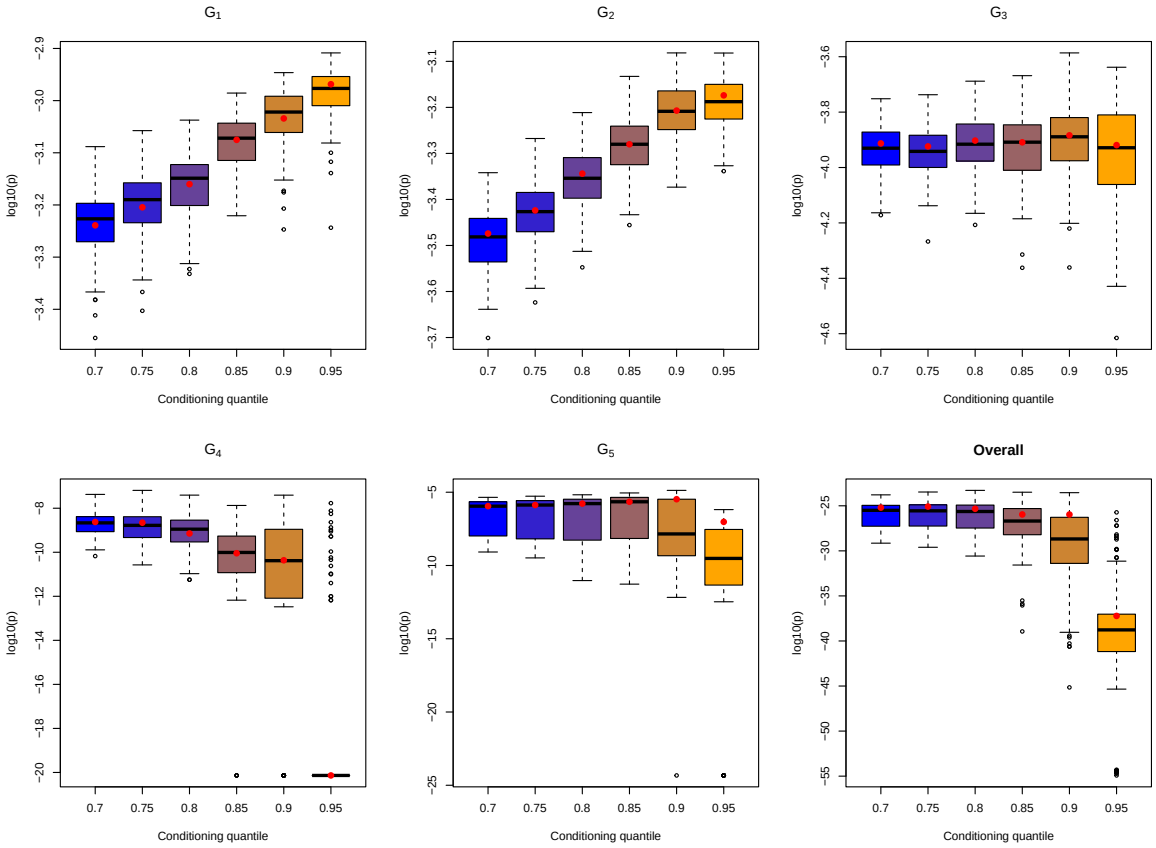


Figure 6.7.23: Part 2 subgroup and overall bootstrapped probability estimates on the log scale for C4. The red points indicate the original sample estimates and the colouring of the boxplots indicates the choice of conditioning threshold, with the conditioning quantile indices 1-6 referring to the quantile levels $\{0.7, 0.75, 0.8, 0.85, 0.9, 0.95\}$, respectively.

Chapter 7

Conclusion

7.1 Summary of contributions

In Chapter 3, we present a novel method for assessing the impact of a multivariate metocean environment on an offshore structure. We construct a forward model approach, utilising the multivariate extreme value framework of Heffernan and Tawn (2004) to model the joint extremal behaviour of storm peak significant wave height and peak wave steepness and obtain an estimate for the density of these joint conditions. This model provides a flexible framework for the modelling of jointly extreme events, facilitating joint probability and density estimation in both asymptotically dependent and asymptotically independent settings. It is therefore well-suited to the study of joint metocean variables, for which we have limited previous knowledge of the nature of their extremal dependence structure. We also assess the effect that the joint conditions can have on a vertical monopile, or 'stick-type' structure in the ocean, using simulation from physically-based models for short-term wave height, velocity and acceleration, and models for the load induced on the structure by these kinematics. Repeated simulation from these models can be used to obtain empirical distributions for the structural response at the desired environmental conditions. Combining these two stages allows

for full propagation of uncertainty through each stage of the forward model. Additionally, while previous approaches to assessing structural failure probability (e.g., Coles and Tawn 1994 and Haver and Winterstein 2009) do not consider the short-term wave variability and interaction with structures, our method fully accounts for these factors. This ensures assessment of structure failure is specific to the the type of structure being considered; for instance, the impact of a particular environment on an offshore platform will differ greatly to that on a floating wind turbine. We use a novel representation for the impact of a set of environmental conditions on an offshore structure to assess our forward approach against the widely-used environmental contour method of Haver and Winterstein (2009). We show that, for three example test structures, the environmental contour method failures to capture the full environment-structures interaction, thereby motivating our approach.

In Chapter 4, we extend this work, via an application to a more complex model for environment-structure interactions. We consider a model representing an offshore structure with resonant properties, which amplify the structural response induced by short-term wave behaviour at particular frequencies. This example further demonstrates the importance of our method which couples extreme value statistics with physically-based stochastic simulation, as a statistical analysis alone would not correctly account for these resonant effects. For more complex physically-based models, the computational cost of obtaining empirical distributions for the structural response at all possible environmental conditions can make this 'brute-force' forward approach infeasible. Instead, we introduce methods to avoid this computational burden in the form of efficient sampling and emulation techniques. Our efficient sampling method builds on work by the likes of Castellon et al. (2023), coupling importance sampling with parallel tempering MCMC. The resulting algorithm automatically chooses which joint environmental conditions at which to estimate the empirical structural response distribution, in order to reduce uncertainty in estimates for structural failure probability for a fixed computa-

tional cost. Our emulation approach uses a Gaussian process regression as a surrogate for expensive physically-based simulation, which is trained on an initial sample of 'true' evaluations from the physically-based model at a set of initial environmental conditions. The Gaussian process emulator is then updated according to a utility function designed to balance the exploration-exploitation trade-off when exploring the environment space. This balance is determined by a weight parameter, with particular values of this parameter resulting in better training of the emulator and therefore less biased estimates for the structural failure probability. We demonstrate the improved performance of these efficient sampling and emulation methods over a simple Monte Carlo approach for a synthetic structure with resonant effects. We also show that our methods offer similar improvements over Monte Carlo methods for a real-world scenario, which we model by utilising the structural loading computational model of Taylor et al. (2024) to approximate the load induced by the metocean environment on an oscillating offshore wind turbine.

In Chapter 5, we develop a novel approach for the selection of data to be used when carrying out inference using the multivariate conditional extreme value model of Heffernan and Tawn (2004). When using this approach, a conditioning threshold in one variable must be chosen, such that any observed data with conditioning variable larger than this threshold is considered extreme and can therefore be used in inference. Care must be taken when choosing this threshold in order to make best use of the data, and balance the resulting bias-variance trade-off. We develop a method which allows the automatic selection of this threshold, and thus makes optimal use of the available data. This method is particularly valuable as no other automated threshold selection method exists for this model, or any other model capable of capturing the full range of asymptotic dependence and asymptotic independence. We define a metric which captures both the bias and variance present in the model fit for a particular choice of conditioning threshold, again balanced by a weight parameter. The value

of conditioning threshold is then chosen by comparing the metric value for a set of candidate threshold values, with the candidate which minimises the metric value being considered the best choice. We first demonstrate the benefit of using the approach of Heffernan and Tawn (2004), which is capable of correctly accounting for both asymptotic dependence and asymptotic independence when estimating joint probabilities, over approaches such as that of Wan and Davis (2019) which, whilst including an approach for the automatic selection of their equivalent exceedance threshold, does not correctly account for asymptotically independent settings. We show that the method of Heffernan and Tawn (2004) provides greatly reduced error in the estimation of joint survivor probabilities when compared to the method of Wan and Davis (2019) when the variables are asymptotically independent, and performs slightly better when the variables are asymptotically dependent. We then carry out a basic analysis of the performance of our threshold selection method, showing that, in both asymptotically dependent and asymptotically independent cases, it can offer a reduction in parameter estimate error over using randomly chosen thresholds restricted to a reasonable fixed range when applying the Heffernan and Tawn (2004) model. This improved performance, however, is dependent on the value of the weight parameter, with some values resulting in higher parameter estimate error than when using random thresholds. From our preliminary analysis we identify a range of weight parameter values that result in a reduction in error for all test cases.

7.2 Further work

7.2.1 Extreme value modelling and estimation of structural failure probability in higher dimensions

The analyses, presented in Chapters 3 and 4, of the impact of joint metocean environments on offshore structures only consider a two dimensional environment (of wave

height and wave steepness). The short-term wave kinematics and associated structural responses modelled via physically-based simulation are therefore only dependent on these two variables. In reality, there are other environment factors such as wind speed, wind and wave direction, and sea-current that can affect the structure. Inclusion of these would require application of extreme value models in higher dimensions (whilst accounting for directional variables for which marginal extremes are not of interest) and simulation from models for kinematics and models for structural response that consider additional variables; for instance, directional components can be added to simulated waves by defining a joint frequency-direction spectrum (Grainger et al., 2021).

Additionally, accurate modelling of the effect of a higher dimensional environment on an offshore structure requires more detailed physically-based models. The current 'stick-type' model structures employed in this thesis are simplifications of the real-world monopile structures that they represent. In practice, more representative schemes, such as finite element models, should be used. The study of wave-structure interactions in higher dimensions requires thus complex physically-based models, and will therefore result in higher computational cost. This, however, would further motivate the use of efficient sampling and emulation techniques discussed in Chapter 4. Additionally, a higher dimensional environment and more complex wave-structure interaction could lead to a more complex response surface than those seen in Chapters 3 and 4, with multiple modes. This would further invalidate the simplifying assumptions made by environmental contour methods such as those of Haver and Winterstein (2009).

7.2.2 Tuning of the weight parameter used in emulator utility function

In Chapter 4, we introduce a utility function used to select new environmental conditions at which to evaluate the response distribution, which is then used to update the Gaussian emulator for structural response. This utility function response depends on

the value of its weight parameter. Our sensitivity analysis shows that the performance of the active learning method depends heavily on this weight parameter, and so it would be useful to develop an automated tuning approach for the selection of this parameter. For instance, we might employ a cross validation method, splitting the output of some initial 'batch' of physically-based model simulations into training and test sets.

We could also consider alternative utility functions that do not rely on weight parameter; for instance, the methods of Osborne et al. (2012) and Gunter et al. (2014), whose utility functions aim to minimise the total expected entropy in a target integral (in our case the integral used to estimate the structural failure probability). These methods, however, rely on additional theoretical assumptions and often incur their own computational cost, so care must be taken when applying them to general offshore design settings.

7.2.3 Multivariate threshold selection in higher dimensions

In Chapter 5, we propose an automated threshold selection approach for the method of Heffernan and Tawn (2004), currently only in the case of $d = 2$ dimensions. As discussed above, it may be desirable to work in higher dimension environment spaces so that the effect of additional metocean variables (such as wind speed or still water sea-level) on the structure can be accounted for. Our current threshold selection method relies on a transformation of observed excesses and residual variables to the joint bivariate uniform space, and then a test for the joint uniformity of the resulting sample. Each of these steps would need to be extended to the arbitrary $d \geq 2$ setting in order for our method to be applicable in higher dimensions. We discuss issues with such potential extensions below.

Extension of the joint uniform transform to higher dimensions can be carried out using methods such as those described by Winter and Tawn (2017), who present a copula framework for the joint residual distribution with known marginal distributions.

The test of uniformity, however, would likely need to be redesigned. We current utilise the method of Ripley (1988), which has been shown to scale well to $d = 3$ (Baddeley et al., 2016), but not for $d > 3$. It may therefore be necessary to introduce a new test of joint uniformity. For instance, the method of Friedman and Rafsky (1979), which uses minimal spanning trees to test for similarity between the distributions of two samples in arbitrary dimension $d > 2$, could be considered.

Bibliography

Abramowitz, M. and Stegun, I. A. (1965). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*, volume 55. Courier Corporation.

Airy, G. B. (1845). *Tides and waves*. B. Fellowes.

Alves Ribeiro, J., Alves Ribeiro, B., Pimenta, F., M.O. Tavares, S., Zhang, J., and Ahmed, F. (2025). Offshore wind turbine tower design and optimization: a review and AI-driven future directions. *Applied Energy*, 397:126294.

André, L. M., Campbell, R., D' Arcy, E., Farrell, A., Healy, D., Kakampakou, L., Murphy, C., Murphy-Barltrop, C. J. R., and Speers, M. (2025). Extreme value methods for estimating rare events in utopia: Eva (2023) conference data challenge: team lancopula utopiversity. *Extremes*, 28(1):23–45.

Baddeley, A., Rubak, E., Turner, R., et al. (2016). *Spatial point patterns: methodology and applications with R*, volume 1. CRC press Boca Raton.

Barnett, V. (1976). The ordering of multivariate data. *Journal of the Royal Statistical Society: Series A (General)*, 139(3):318–344.

Bates, P. D., Savage, J., Wing, O., Quinn, N., Sampson, C., Neal, J., and Smith, A. (2023). A climate-conditioned catastrophe risk model for uk flooding. *Natural Hazards and Earth System Sciences*, 23(2):891–908.

Berger, V. W. and Zhou, Y. (2014). Kolmogorov–smirnov test: Overview. *Wiley statsref: Statistics reference online*.

Bernard, E., Naveau, P., Vrac, M., and Mestre, O. (2013). Clustering of maxima: Spatial dependencies among heavy rainfall in france. *Journal of Climate*, 26:7929–7937.

Bishop, C. M. and Nasrabadi, N. M. (2006). *Pattern Recognition and Machine Learning*, volume 4. Springer.

Bortot, P., Coles, S. G., and Tawn, J. A. (2000). The multivariate Gaussian tail model: An application to oceanographic data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(1):31–049.

Bortot, P. and Tawn, J. A. (1998). Models for the extremes of Markov chains. *Biometrika*, 85(4):851–867.

Brown, S. J. (2020). Future changes in heatwave severity, duration and frequency due to climate change for the most populous cities. *Weather and Climate Extremes*, 30:100278.

Brozius, H. and de Haan, L. (1987). On limiting laws for the convex hull of a sample. *Journal of Applied Probability*, 24(4):852–862.

Capéraà, P., Fougères, A.-L., and Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84(3):567–577.

Castellon, D. F., Fenerci, A., Øiseth, O., and Petersen, Ø. W. (2022). Investigations of the long-term extreme buffeting response of long-span bridges using importance sampling monte carlo simulations. *Engineering Structures*, 273:114986.

Castellon, D. F., Fenerci, A., Petersen, Ø. W., and Øiseth, O. (2023). Full long-term buffeting analysis of suspension bridges using gaussian process surrogate modelling

and importance sampling monte carlo simulations. *Reliability Engineering & System Safety*, 235:109211.

Chai, W. and Leira, B. J. (2018). Environmental contours based on inverse SORM. *Marine Structures*, 60:34–51.

Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 54:207–222.

Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The american statistician*, 49(4):327–335.

Cohn, D. (1993). Neural network exploration using optimal experiment design. In Cowan, J., Tesauro, G., and Alspector, J., editors, *Advances in Neural Information Processing Systems*, volume 6. Morgan-Kaufmann.

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*, volume 208. Springer.

Coles, S. G., Heffernan, J. E., and Tawn, J. A. (1999). Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365.

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 53(2):377–392.

Coles, S. G. and Tawn, J. A. (1994). Statistical methods for multivariate extremes: an application to structural design (with discussion). *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 43(1):1–31.

Collings, T. P., Murphy-Barltrop, C. J., Murphy, C., Haigh, I. D., Bates, P. D., and Quinn, N. D. (2025). Automated tail-informed threshold selection for extreme coastal sea levels. *Natural Hazards and Earth System Sciences*, 25(11):4545–4562.

Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19(90):297–301.

Cormier, E., Genest, C., and Nešlehová, J. G. (2014). Using b-splines for nonparametric inference on bivariate extreme-value copulas. *Extremes*, 17(4):633–659.

Davis, R., Mulrow, E., and Resnick, S. (1987). The convex hull of a random sample in. *Stochastic Models*, 3(1):1–27.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society Series B*, 52(3):393–425.

de Haan, L. (1985). Extremes in higher dimensions: the model and some statistics. *Proceedings 45th, Session of International Statistical Institution, 1985*, page 26.3.

de Haan, L. and Resnick, S. I. (1977). Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 40(4):317–337.

de Hauteclocque, G., Mackay, E., and Vanem, E. (2022). Quantitative comparison of environmental contour approaches. *Ocean Engineering*, 245:110374.

Dean, R. G. and Dalrymple, R. A. (1991). *Water wave mechanics for engineers and scientists*, volume 2. World Scientific Publishing Company.

DNVGL-RP-C205 (2017). Environmental conditions and environmental loads. Det Norske Veritas group, Norway.

D' Arcy, E., Tawn, J. A., Joly, A., and Sifnioti, D. E. (2023). Accounting for seasonality in extreme sea-level estimation. *The Annals of Applied Statistics*, 17(4):3500–3525.

Earl, D. J. and Deem, M. W. (2005). Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910–3916.

Eastoe, E. F. and Tawn, J. A. (2008). Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 58(1):25–45.

Eastoe, E. F. and Tawn, J. A. (2012). Modelling the distribution of the cluster maxima of exceedances of subasymptotic thresholds. *Biometrika*, 99(1):43–55.

Eddy, W. F. and Gale, J. D. (1981). The convex hull of a spherically symmetric sample. *Advances in Applied Probability*, 13(4):751–763.

ElHamahmy, D. A. (2025). Integrating artificial intelligence in offshore platform design: Enhancing efficiency, safety, and structural integrity. *Safety, and Structural Integrity (June 01, 2025)*.

Ewans, K. and Jonathan, P. (2008). The effect of directionality on northern North Sea extreme wave design criteria. *Journal of Offshore Mechanics and Arctic Engineering*, 130:041604.

Faltinsen, O. M., Newman, J. N., and Vinje, T. (1995). Nonlinear wave loads on a slender vertical cylinder. *Journal of Fluid Mechanics*, 289:179–198.

Farrell, A., Eastoe, E. F., and Lee, C. (2025). Conditional extremes with graphical models.

Fawcett, L. and Walshaw, D. (2007). Improved estimation for temporally clustered extremes. *Environmetrics*, 18(2):173–188.

Feld, G., Randell, D., Wu, Y., Ewans, K., and Jonathan, P. (2015). Estimation of storm peak and intra-storm directional-seasonal design conditions in the North Sea. *Journal of Offshore Arctic Engineering*, 137:021102.

Ferro, C. A. T. and Segers, J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(2):545–556.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190.

Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Annales de la Société Polonaise de Mathématique*.

Friedman, J. H. and Rafsky, L. C. (1979). Multivariate generalizations of the wald-wolfowitz and smirnov two-sample tests. *The Annals of statistics*, pages 697–717.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, 42(6):2243 – 2281.

Geffroy, J. (1958). Contributions à la théorie des valeurs extrêmes. *Publ. Inst. Statist. Univ. Paris*, 7:37–185.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC.

Genton, M. G. (2001). Classes of kernels for machine learning: a statistics perspective. *Journal of machine learning research*, 2(Dec):299–312.

Geweke, J. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Technical report, Federal Reserve Bank of Minneapolis.

Gibson, R. (2020). Extreme Environmental Loading of Fixed Offshore Structures: Summary Report, Component 2. https://www.hse.gov.uk/offshore/assets/docs/summary-report-component2.pdf.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, 44(3):423–453.

Gneiting, T. and Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1:125–151.

Graham, M. M. and Storkey, A. J. (2017). Continuously tempered Hamiltonian Monte Carlo.

Grainger, J. P., Sykulski, A. M., Jonathan, P., and Ewans, K. (2021). Estimating the parameters of ocean wave spectra. *Ocean Engineering*, 229:108934.

Gramstad, O., Agrell, C., Bitner-Gregersen, E., Guo, B., Ruth, E., and Vanem, E. (2020). Sequential sampling method using gaussian process regression for estimating extreme structural response. *Marine Structures*, 72:102780.

Guerrero, M. B., Huser, R., and Ombao, H. (2023). Conex–connect: Learning patterns in extremal brain connectivity from multichannel eeg data. *The Annals of Applied Statistics*, 17:178–198.

Guillotte, S., Perron, F., and Segers, J. (2011). Non-parametric Bayesian inference on bivariate extremes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(3):377–406.

Gumbel, E. J. (1958). *Statistics of Extremes*. Columbia University Press.

Gumbel, E. J. (1960). Distributions des valeurs extremes en plusiers dimensions. *Publ. Inst. Statist. Univ. Paris*, 9:171–173.

Gunter, T., Osborne, M. A., Garnett, R., Hennig, P., and Roberts, S. J. (2014). Sampling for inference in probabilistic models with fast Bayesian quadrature. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.

Hafver, A., Agrell, C., and Vanem, E. (2022). Environmental contours as Voronoi cells. *Extremes*, 25(3):451–486.

Hall, P. and Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science*, 15(2):153–167.

Hall, W. and Wellner, J. A. (2020). Estimation of mean residual life. In *Statistical Modeling for Biological Systems*, pages 169–189. Springer.

Harrison, E., Drake, T., and Ots, R. (2023). *finalfit: Quickly Create Elegant Regression Results Tables and Plots when Modelling*. R package version 1.0.7.

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., and Huseby, A. B. (2021). A benchmarking exercise for environmental contours. *Ocean Engineering*, 236:109504.

Hasselmann, K., Barnett, T. P., Bouws, E., Carlson, H., Cartwright, D. E., Enke, K., Ewing, J., Gienapp, A., Hasselmann, D., Kruseman, P., Meerburg, A., Muller, P., Olbers, D., Richter, K., Sell, W., and Walden, H. (1973). Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). Ergänzungsheft zur Deutschen Hydrographischen Zeitschrift, Reihe A, Nr. 12.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The elements of statistical learning*. Springer, New York.

Hastie, T., Tibshirani, R., Friedman, J., et al. (2009). The elements of statistical learning.

Haver, S. (1987). On the joint distribution of heights and periods of sea waves. *Ocean Engineering*, 14:359–376.

Haver, S. and Winterstein, S. (2009). Environmental contour lines: a method for estimating long term extremes by a short term analysis. *Transactions - Society of Naval Architects and Marine Engineers*, 116:116–127.

He, R. and Zhu, T. (2019). Model tests on the frequency responses of offshore monopiles. *Journal of Marine Science and Engineering*, 7(12):430.

Heffernan, J. E. and Tawn, J. A. (2003). An extreme value analysis for the investigation into the sinking of the M. V. Derbyshire. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 52(3):337–354.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society Series B*, 66:497–546.

Hennig, P., Osborne, M. A., and Kersting, H. P. (2022). *Probabilistic Numerics: Computation as Machine Learning*. Cambridge University Press.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3:1163–1174.

Holthuijsen, L. H. (2010). *Waves in Oceanic and Coastal Waters*. Cambridge University Press.

Hosking, J. R. M., Wallis, J. R., and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261.

Hsing, T., Hüsler, J., and Leadbetter, M. R. (1988). On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields*, 78(1):97–112.

Huser, R. and Davison, A. C. (2014). Space–time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76:439–461.

Jane, R., Cadavid, L., Obeysekera, J., and Wahl, T. (2020). Multivariate statistical modelling of the drivers of compound flood events in south florida. *Natural Hazards and Earth System Sciences*, 20(10):2681–2699.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348):158–171.

Joe, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistics & Probability Letters*, 9(1):75–81.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. Chapman and Hall/CRC, New York.

Jonathan, P., Ewans, K., and Flynn, J. (2012). Joint modelling of vertical profiles of large ocean currents. *Ocean Engineering*, 42:195–204.

Jonathan, P., Ewans, K., and Flynn, J. (2014a). On the estimation of ocean engineering design contours. *Journal of Offshore Mechanics and Arctic Engineering*, 136(4):041101.

Jonathan, P., Ewans, K., and Randell, D. (2013). Joint modelling of extreme ocean environments incorporating covariate effects. *Coastal Engineering*, 79:22–31.

Jonathan, P., Ewans, K., and Randell, D. (2014b). Non-stationary conditional extremes of northern North Sea storm characteristics. *Environmetrics*, 25(3):172–188.

Jonathan, P., Flynn, J., and Ewans, K. (2010). Joint modelling of wave spectral parameters for extreme sea states. *Ocean Engineering*, 37(11-12):1070–1080.

Jonathan, P., Randell, D., Wu, Y., and Ewans, K. (2014c). Return level estimation from non-stationary spatial data exhibiting multidimensional covariate effects. *Ocean Engineering*, 88:520–532.

Keef, C., Papastathopoulos, I., and Tawn, J. A. (2013). Estimation of the conditional distribution of a multivariate variable given that one of its components is large:

Additional constraints for the Heffernan and Tawn model. *Journal of Multivariate Analysis*, 115:396–404.

Kirby, A., Briol, F.-X., Dunstan, T. D., and Nishino, T. (2023). Data-driven modelling of turbine wake interactions and flow resistance in large wind farms. *Wind Energy*, 26(9):968–984.

Kyselý, J., Picek, J., and Beranová, R. (2010). Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. *Global and Planetary Change*, 72:55–68.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (2012). *Extremes and Related Properties of Random Sequences and Processes*. Springer Science & Business Media.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(2):475–499.

Ledford, A. W. and Tawn, J. A. (2003). Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):521–543.

Liu, J., Xu, S., Mao, H., Han, C., and Han, J. (2025). High-frequency resonance of bottom-fixed monopile-supported offshore wind turbines under nonlinear waves. *Ocean Engineering*, 340:122355.

Luo, R., Wang, J., Yang, Y., Wang, J., and Zhu, Z. (2018). Thermostat-assisted continuously-tempered Hamiltonian Monte Carlo for Bayesian learning. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

Lystad, T. M., Fenerci, A., and Øiseth, O. (2023). Full long-term extreme buffeting response calculations using sequential gaussian process surrogate modeling. *Engineering Structures*, 292:116495.

Mackay, E. and de Hauteclocque, G. (2023). Model-free environmental contours in higher dimensions. *Ocean Engineering*, 273:113959.

Mackay, E. and Haselsteiner, A. F. (2021). Marginal and total exceedance probabilities of environmental contours. *Marine Structures*, 75:102863.

Madsen, H. O., Krenk, S., and Lind, N. C. (1986). *Methods of Structural Safety*. Englewood Cliffs: Prentice-Hall.

Majumder, R., Shaby, B. A., Reich, B. J., and Cooley, D. S. (2025). Semiparametric estimation of the shape of the limiting bivariate point cloud. *Bayesian Analysis*, TBA(TBA):1–27.

Marrel, A. and Iooss, B. (2024). Probabilistic surrogate modeling by gaussian process: A review on recent insights in estimation and validation. *Reliability Engineering & System Safety*, 247:110094.

Martzikos, N., Ruzzo, C., Malara, G., Fiamma, V., and Arena, F. (2024). Applying neural networks to predict offshore platform dynamics. *Journal of Marine Science and Engineering*, 12(11).

Mathisen, J. and Bitner-Gregersen, E. (1990). Joint distributions for significant wave height and wave zero-up-crossing period. *Applied Ocean Research*, 12(2):93–103.

Meng, X.-L. and Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica*, 6(4):831–860.

Mert, A. C., Guo, X., Shen, Z., Pan, H., and Dias, D. (2025). A reliability-based

framework for offshore monopile design using cpt data and deep learning enhanced adaptive metamodeling. *Ocean Engineering*, 342:122952.

Mhalla, L., Opitz, T., and Chavez-Demoulin, V. (2019). Exceedance-based nonlinear regression of tail dependence. *Extremes*, 22:523–552.

Miasojedow, B., Moulines, E., and Vihola, M. (2013). An adaptive parallel tempering algorithm. *Journal of Computational and Graphical Statistics*, 22(3):649–664.

Morison, J. R., O'Brien, M. P., Johnson, J. W., and Schaaf, S. A. (1950). The force exerted by surface waves on piles. *Journal of Petroleum Technology*, 189:149–154.

Moustapha, M., Marelli, S., and Sudret, B. (2022). Active learning for structural reliability: Survey, general framework and benchmark. *Structural Safety*, 96:102174.

Murphy, C., Tawn, J. A., and Varty, Z. (2025). Automated threshold selection and associated inference uncertainty for univariate extremes. *Technometrics*, 67(2):215–224.

Murphy-Barltrop, C. and Wadsworth, J. (2024). Modelling non-stationarity in asymptotically independent extremes. *Computational Statistics and Data Analysis*, 199:108025.

Murphy-Barltrop, C., Wadsworth, J., de Carvalho, M., and Youngman, B. (2025). Modelling non-stationary extremal dependence through a geometric approach. *arXiv preprint arXiv:2509.22501*.

Murphy-Barltrop, C., Wadsworth, J., and Eastoe, E. (2024). Improving estimation for asymptotically independent bivariate extremes via global estimators for the angular dependence function. *Extremes*, 27(4):643–671.

Myrhaug, D. (2018). Some probabilistic properties of deep water wave steepness. *Oceanologia*, 60(2):187–192.

Naveau, P., Guillou, A., Cooley, D., and Diebolt, J. (2009). Modelling pairwise dependence of maxima in space. *Biometrika*, 96(1):1–17.

Nolde, N. (2014). Geometric interpretation of the residual dependence coefficient. *Journal of Multivariate Analysis*, 123:85–95.

Nolde, N. and Wadsworth, J. L. (2022). Linking representations for multivariate extremes via a limit set. *Advances in Applied Probability*, 54(3):688–717.

NORSOK N-003 (2017). NORSOK Standard N-003:2017: Actions and action effects. NORSOK, Norway.

Northrop, P. J., Attalides, N., and Jonathan, P. (2016). Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 66(1):93–120.

Northrop, P. J. and Jonathan, P. (2011). Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*, 22:799–809.

Nussbaumer, H. J. (1982). *The fast Fourier transform*. Springer.

Orszaghova, J., Taylor, P. H., Wolgamot, H., McCauley, G., Kurniawan, A., Wu, Q., Tan, B., and George, A. E. (2025). Wave loads on monopile foundations revisited – new high-quality experiments for validation of a novel engineering model. In *Proc. ASME OMAE 2025*, Vancouver, British Columbia, Canada. ASME. OMAE2025-156732.

Osborne, M., Garnett, R., Ghahramani, Z., Duvenaud, D. K., Roberts, S. J., and Rasmussen, C. (2012). Active learning of model evidence using Bayesian quadrature. In Pereira, F., Burges, C., Bottou, L., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc.

Parey, S., Hoang, T. T. H., and Dacunha-Castelle, D. (2013). The importance of mean and variance in predicting changes in temperature extremes. *Journal of Geophysical Research: Atmospheres*, 118(15):8285–8296.

Park, J. (2025). Sampling from high-dimensional, multimodal distributions using automatically tuned, tempered Hamiltonian Monte Carlo.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Édouard Duchesnay (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830.

Peherstorfer, B., Cui, T., Marzouk, Y., and Willcox, K. (2016). Multifidelity importance sampling. *Computer Methods in Applied Mechanics and Engineering*, 300:490–509.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131.

Pickands, J. (1981). Multivariate extreme value distribution. *Proceedings 43th, Session of International Statistical Institution, 1981*, page 859–878.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89:1303–1313.

Pollatsek, A. and Tversky, A. (1970). A theory of risk. *Journal of Mathematical Psychology*, 7(3):540–553.

Qiu, Y. and Wang, X. (2024). Efficient multimodal sampling via tempered distribution flow. *Journal of the American Statistical Association*, 119(546):1446–1460.

Quinn, N., Bates, P. D., Neal, J., Smith, A., Wing, O., Sampson, C., Smith, J., and

Heffernan, J. (2019). The spatial dependence of flood hazard and risk in the united states. *Water Resources Research*, 55(3):1890–1911.

Randell, D., Feld, G., Ewans, K., and Jonathan, P. (2015). Distributions of return values for ocean wave characteristics in the South China Sea using directional–seasonal extreme value analysis. *Environmetrics*, 26(6):442–450.

Randell, D., Turnbull, K., Ewans, K., and Jonathan, P. (2016). Bayesian inference for non-stationary marginal extremes. *Environmentrics*, 27:439–450.

Resnick, S. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5:303–336.

Resnick, S. I. (2008). *Extreme Values, Regular Variation, and Point Processes*, volume 4. Springer Science & Business Media.

Richards, J., Tawn, J. A., and Brown, S. (2023). Joint estimation of extreme spatially aggregated precipitation at different scales through mixture modelling. *Spatial Statistics*, 53:100725.

Riise, B. H., Grue, J., Jensen, A., and Johannessen, T. B. (2018). High frequency resonant response of a monopile in irregular deep water waves. *Journal of Fluid Mechanics*, 853:564–586.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13(2):255–266.

Ripley, B. D. (1988). *Statistical inference for spatial processes*. Cambridge university press.

Rohrbeck, C., Simpson, E. S., and Tawn, J. A. (2023). Editorial: EVA (2023) Conference Data Challenge. *Extremes*, (to appear).

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23(3):470–472.

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

Rubinstein, R. Y. and Kroese, D. P. (2016). *Simulation and the Monte Carlo method*. John Wiley & Sons.

Sambridge, M. (2013). A parallel tempering algorithm for probabilistic sampling and multimodal optimization. *Geophysical Journal International*, 196(1):357–374.

Santo, H., Taylor, P., Dai, S., Day, A., and Chan, E. (2020). Wave-in-deck experiments with focussed waves into a solid deck. *Journal of Fluids and Structures*, 98:103139.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT-Statistical journal*, 10(1):33–60.

Schälte, Y., Fröhlich, F., Jost, P. J., Vanhoefer, J., Pathirana, D., Stapor, P., Lakrisenko, P., Wang, D., Raimúndez, E., Merkt, S., Schmiester, L., Städter, P., Grein, S., Dudkin, E., Doresic, D., Weindl, D., and Hasenauer, J. (2023). pyPESTO: a modular and scalable tool for parameter estimation for dynamic models. *Bioinformatics*, 39(11):711.

Scott, D. W. (2015). *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons.

Seo, S., Wallat, M., Graepel, T., and Obermayer, K. (2000). Gaussian process regression: active data selection and test point rejection. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, volume 3, pages 241–246 vol.3.

Shooter, R., Ross, E., Ribal, A., Young, I. R., and Jonathan, P. (2021a). Spatial dependence of extreme seas in the north east atlantic from satellite altimeter measurements. *Environmetrics*, 32:1–15.

Shooter, R., Tawn, J. A., Ross, E., and Jonathan, P. (2021b). Basin-wide spatial conditional extremes for severe ocean storms. *Extremes*, 24:241–265.

Sibuya, M. (1960). Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, 11(2):195–210.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, volume 26. CRC press.

Simpson, E. S. and Tawn, J. A. (2024). Estimating the limiting shape of bivariate scaled sample clouds: With additional benefits of self-consistent inference for existing extremal dependence properties. *Electronic Journal of Statistics*, 18(2):4582–4611.

Simpson, E. S., Wadsworth, J. L., and Tawn, J. A. (2020). Determining the dependence structure of multivariate extremes. *Biometrika*, 107(3):513–532.

Smith, R. L., Tawn, J. A., and Yuen, H. K. (1990). Statistics of multivariate extremes. *International Statistical Review / Revue Internationale de Statistique*, 58(1):47–58.

Smith, R. L. and Weissman, I. (1994). Estimating the extremal index. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(3):515–528.

Speers, M., Randell, D., Tawn, J. A., and Jonathan, P. (2024). Estimating metocean environments associated with extreme structural response to demonstrate the dangers of environmental contour methods. *Ocean Engineering*, 311:118754.

Speers, M., Tawn, J., and Jonathan, P. (2026). Sequential design for the efficient estimation of offshore structure failure probability. *Ocean Engineering*, 349:123990.

Standard Norge (2022). Shall NORSOK N-0031 and NORSOK N-0062 be updated as a result of findings in LOADS JIP? Conclusions from the evaluation committee. https://standard.no/globalassets/fagomrader-sektorer/petroleum/loads-jip-and-norsok-n_003.pdf.

Swan, C. (2020). Extreme Environmental Loading of Fixed Offshore Structures: Summary Report, Component 1. https://www.hse.gov.uk/offshore/assets/docs/summary-report-component1.pdf.

Tabandeh, A., Jia, G., and Gardoni, P. (2022). A review and assessment of importance sampling methods for reliability analysis. *Structural Safety*, 97:102216.

Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika*, 75(3):397–415.

Tawn, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika*, 77(2):245–253.

Taylor, P. H., Jonathan, P., and Harland, L. A. (1997). Time domain simulation of jack-up dynamics with the extremes of a Gaussian process. *Journal of Vibration and Acoustics*, 119:624–628.

Taylor, P. H., Tang, T., Adcock, T. A., and Zang, J. (2024). Transformed-FNV: wave forces on a vertical cylinder—a free-surface formulation. *Coastal Engineering*, 189:104454.

Tendijck, S., Eastoe F., E., Tawn, J. A., Randell, D., and Jonathan, P. (2023a). Modeling the extremes of bivariate mixture distributions with application to oceanographic data. *Journal of the American Statistical Association*, 118(542):1373–1384.

Tendijck, S., Ross, E., Randell, D., and Jonathan, P. (2019). A model for the directional evolution of severe ocean storms. *Environmetrics*, 30(1):e2541.

Tendijck, S., Tawn, J., and Jonathan, P. (2023b). Extremal characteristics of conditional models. *Extremes*, 26(1):139–156.

Tiago de Oliveira, J. (1959). Extremal distributions. *Univ. Lisboa Revista Fac. Ci. A.*, 7:219–227.

Towe, R., Randell, D., Kensler, J., Feld, G., and Jonathan, P. (2023). Estimation of associated values from conditional extreme value models. *Ocean Engineering*, 272:113808.

Towe, R., Ross, E., Randell, D., and Jonathan, P. (2024). covXtreme: MATLAB software for non-stationary penalised piecewise constant marginal and conditional extreme value models. *Environ. Model. Softw.*, 177:106035.

Towe, R., Zanini, E., Randell, D., Feld, G., and Jonathan, P. (2021). Efficient estimation of distributional properties of extreme seas from a hierarchical description applied to calculation of un-manning and other weather-related operational windows. *Ocean Engineering*, 238:109642.

Towe, R. P., Tawn, J. A., Lamb, R., and Sherlock, C. G. (2019). Model-based inference of conditional extreme value distributions with hydrological applications. *Environmetrics*, 30(8):e2575.

Tromans, P. S. and Vanderschuren, L. (1995). Response based design conditions in the North Sea: application of a new method. In *Offshore Technology Conference*, pages OTC–7683. OTC.

Vallis, G. K. (2017). *Atmospheric and Oceanic Fluid Dynamics*. Cambridge University Press.

Vanem, E., Zhu, T., and Babanin, A. (2022). Statistical modelling of the ocean environment - a review of recent developments in theory and applications. *Marine Structures*, 86:103297.

Varty, Z., Tawn, J. A., Atkinson, P. M., and Bierman, S. (2021). Inference for extreme earthquake magnitudes accounting for a time-varying measurement process. *arXiv preprint arXiv:2509.22501*.

Von Mises, R. (1936). La distribution de la plus grande de n valuers. *Revue de l'Union Mathématique Interbalcanique*, 1:141–160.

Vousden, W. D., Farr, W. M., and Mandel, I. (2015). Dynamic temperature selection for parallel tempering in markov chain monte carlo simulations. *Monthly Notices of the Royal Astronomical Society*, 455(2):1919–1937.

Wadsworth, J. and Tawn, J. (2022). Higher-dimensional spatial extremes via single-site conditioning. *Spatial Statistics*, 51:100677.

Wadsworth, J. L. (2016). Exploiting structure of maximum likelihood estimators for extreme value threshold selection. *Technometrics*, 58(1):116–126.

Wadsworth, J. L. and Campbell, R. (2024). Statistical inference for multivariate extremes via a geometric approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(5):1243–1265.

Wadsworth, J. L. and Tawn, J. A. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19:2689–2714.

Wan, J.-H., Bai, R., Li, X.-Y., and Liu, S.-W. (2023). Natural frequency analysis of monopile supported offshore wind turbines using unified beam-column element model. *Journal of Marine Science and Engineering*, 11(3):628.

Wan, P. and Davis, R. A. (2019). Threshold selection for multivariate heavy-tailed data. *Extremes*, 22(1):131–166.

Wang, C., Qiang, X., Xu, M., and Wu, T. (2022). Recent advances in surrogate modeling methods for uncertainty quantification and propagation. *Symmetry*, 14(6):1219.

Wang, H., Gramstad, O., Schär, S., Marelli, S., and Vanem, E. (2024). Comparison of probabilistic structural reliability methods for ultimate limit state assessment of wind turbines. *Structural Safety*, 111:102502.

Wang, J., Sun, Z., and Cao, R. (2021). An efficient and robust kriging-based method for system reliability analysis. *Reliability Engineering & System Safety*, 216:107953.

Winter, H. C. and Tawn, J. A. (2017). kth-order Markov extremal models for assessing heatwave risks. *Extremes*, 20(2):393–415.

Winterstein, S. R., Ude, T. C., Cornell, C. A., Bjerager, P., and Haver, S. (1993). Environmental parameters for extreme response: Inverse form with omission factors. In *Proceedings of the ICOSSAR-93, Innsbruck, Austria*, pages 551–557.

Wood, S. N. (2017). *Generalized additive models*. Chapman and Hall/CRC, New York.

Xiao, N. C., Zhan, H., and Yuan, K. (2020). A new reliability method for small failure probability problems by combining the adaptive importance sampling and surrogate models. *Computer Methods in Applied Mechanics and Engineering*, 372:113336.

Yang, G. L. (1978). Estimation of a biometric function. *The Annals of Statistics*, pages 112–116.

Yang, H., Zhang, Y., Lin, J., and Mao, H. (2025). Higher-harmonic resonance response of a monopile-supported offshore wind turbine in waves and wind. *Ocean Engineering*, 342:123120.

Yang, X., Liu, Y., Mi, C., and Tang, C. (2018). System reliability analysis through active learning kriging model with truncated candidate region. *Reliability Engineering & System Safety*, 169:235–241.

Youngman, B. D. (2019a). Generalized additive models for exceedances of high thresholds with an application to return level estimation for us wind gusts. *Journal of the American Statistical Association*, 114(528):1865–1879.

Youngman, B. D. (2019b). Generalized additive models for exceedances of high thresholds with an application to return level estimation for U.S. wind gusts. *Journal of the American Statistical Association*, 114:1865–1879.

Youngman, B. D. (2022). evgam: An R package for generalized additive extreme value models. *Journal of Statistical Software*, 103(3):1–26.

Yu, A., Li, Y., Li, S., and Gong, J. (2024). Construction high precision neural network proxy model for ship hull structure design based on hybrid datasets of hydrodynamic loads. *Journal of Marine Science and Application*, 23(1):49–63.

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54:437–447.

Zhou, X., Hou, C., Yu, Y., and Zhou, Y. (2025). Machine learning-based techniques for marine structures: A state-of-the-art review. *Ocean*, 1(1):9470005.