**ONR-RRR-079**

**Investigation into the application of multivariate hazard curves for external hazard contributions.**

**C. Murphy-Barltrop**

**Lancaster University**

**November 2023**

**Summary Report for ONR-RRR 079**

This document provides a non-technical overview of the research conducted for ONR-RRR-079. For this project, a methodology was developed for quantifying risks from joint extreme events. Specifically, a risk measure known as a ‘return curve’ was considered; this measure provides a robust and interpretable summary of jointly extreme events for combinations of environmental variables. Statistical techniques for estimating return curves are presented, alongside practical methods for evaluating uncertainty and goodness-of-fit.

**Introduction**

The modelling of extreme values is important for a wide variety of sectors, including actuarial science, insurance, environmental science, and structural engineering. One area of structural engineering in which this modelling is particularly crucial is the nuclear industry; extreme natural hazards, such as floods, solar storms and earthquakes, can pose risks to the safety of nuclear sites. Due to the potential risks, nuclear facilities are built to be able to withstand very extreme natural events, and such events must be estimated during the engineering design process.

By definition, historical data sets will contain few extreme observations with which to guide estimation. Extreme value theory provides a mathematical framework for modelling such values, allowing one to extrapolate beyond the observed range of data. This allows for a more robust inference from historical extreme events and, for many practical applications, fitted models are often used to obtain estimates of extreme risk measures. Such measures are vital for informed decision-making and risk assessment.

In many cases, it does not suffice to consider the extremes of a single hazard. For example, during the 1999 flood of the Blayais nuclear power plant in France, a combination of extreme winds and sea levels resulted in the plant's flood defences being overwhelmed, causing damage to the plant's power supply and cooling facilities (Mattei, Vial, Rebour, Liemersdorf, & Turschmann, 2001). Cooling is one of the fundamental safety functions that needs to be delivered to keep nuclear reactors safe. This event was ranked as a level 2 on the International Nuclear Event Scale (International Atomic Energy Agency, 2013), and while no major accident occurred, it exposed key weaknesses within the safety measures and procedures in place at the time for many nuclear sites (Institute for Nuclear Protection and Safety, 2000) including how hazard combinations were considered. This led to the development of new methods for evaluating flood risk, resulting in a significant cost to upgrade many existing sites.

In contrast to the Blayais incident, the direct impact of the 2011 Fukushima nuclear disaster in Japan, ranked as a level 7 on the International Nuclear Event Scale, was huge. Damage to the plant's cooling facilities led to the meltdown of three reactors, resulting in $187 billion in damages and 154,000 people being evacuated. As in Blayais, this event was also triggered by the combination of two extreme events: in this case, an earthquake and a tsunami (International Atomic Energy Agency, 2015). Prior to the accident, regulatory and safety procedures for the Fukushima-Daiichi plant had only considered the separate risks from these two hazards. Owing to this shortcoming, a government report released following the accident stated that it was a “profoundly man-made disaster – that could and should have been foreseen and prevented” (National Diet of Japan, 2012). The Fukushima disaster also led to wide-spread changes in nuclear safety assessments and procedures, both in Japan and internationally. In the most drastic cases, certain countries, such as Germany and Belgium, opted to phase-out all nuclear energy operations in response to the events of Fukushima.

As a consequence of these two incidents, international nuclear regulatory practices now require that the impact from combined extreme events be taken into consideration when designing nuclear facilities. The Office for Nuclear Regulation’s (ONR’s) technical assessment guide for external hazards states that licensees “should take into account combinations of external hazards that could reasonably be expected to occur at a given site” (Office for Nuclear Regulation, 2018). Moreover, the ONR expects sites to be built such that they can withstand external hazards at a 10-4 annual exceedance probability (Office for Nuclear Regulation, 2014). This corresponds to events we would expect to observe once, on average, every 10,000 years. Similar standards are in place for other nuclear regulatory bodies, such as the US Nuclear Regulatory Commission (NRC) and the French Nuclear Safety Authority (ASN).

These regulatory expectations raise two important questions. Firstly, how does one define an extreme event for a combination of two variables? Unlike in the one variable (univariate) setting, two variable (bivariate) observations have no natural ordering, and the most extreme values of two variables need not occur simultaneously. Moreover, in many practical settings there exist wide ranges of bivariate events which can be high impact (e.g. where both variables are extreme, or where only one of the two variables is extreme), implying the definition of a joint extreme event may not be unique (Ross, et al., 2020).

Assuming one has a means of defining a joint extreme event, an important question remains: how does one estimate the risk from such events? This requires an assessment of how the extremes of one variable depend on the extremes of another. Formally, this is known as the extremal dependence structure, and many frameworks exist for its estimation. In practice, accurate estimation of the extremal dependence structure is crucial for ensuring joint extremal behaviour can be properly evaluated.

This report summarises work to answer these questions. A risk measure, known as a return curve, provides an interpretable summary of joint extreme events for two variables. These curves are used to evaluate extreme responses in a wide range of sectors. Using models from the field of extreme value theory (EVT), we introduce novel statistical techniques for estimating extremal dependence structures and return curves. Furthermore, we provide uncertainty quantification and diagnostic tools for assessing the accuracy of return curves, allowing for robust and detailed practical analyses.

**Defining extreme events for one variable**

In the univariate setting, we typically define extreme events as follows: given some observed variable $X$, we say that all observations exceeding some high threshold, $u$, are extreme. For example, if $X$ represents temperate for a location in the UK, we could set $u=30°C$ and say that all temperatures exceeding this value are extreme.

In practice, such events can be modelled using the univariate peaks-over-threshold approach, whereby a generalised Pareto distribution (GPD) is fitted to the threshold-exceeding observations. The fitted model can then be used to approximate extreme behaviour and extrapolate, in a principled manner, outside of the observed range of data; this latter feature is not possible with standard (non-EVT) statistical techniques. The peaks-over-threshold approach is illustrated in Figure 1; the solid line represents the threshold $u$, while the orange points represent the threshold-exceeding observations, to which a GPD is fitted.

**Figure 1:** Illustration of the peaks-over-threshold modelling approach. The green points represent the threshold-exceeding observations, while the red dotted line denotes the fixed threshold value.

A wide range of literature is available for peaks-over-threshold modelling (Davison & Smith, 1990; Dupuis, 1999; Coles, 2001). Moreover, to quantify extreme risks in practice, practitioners typically calculate return levels from the fitted GPD models. Given a small probability $p$, the return level $x\_{p}$ is defined as the value we would the variable $X$ to exceed with probability $p$, i.e., $Probability(X>x\_{p})=p$. For example, setting $p=0.1$ and letting $X$ again represent temperature, the probability of $X$ exceeding $x\_{p}$ is $10\%$. Equivalently, we would expect $X$ to exceed temperature $x\_{p}$ once, on average, every $1/p$ observations, with the quantity $1/p$ termed the return period.

In many practical applications, return level estimates are obtained for return periods that correspond to some number of years. For example, when considering external hazards, nuclear regulators often specify a $10,000$ year time period for design standards. This corresponds to a very extreme event that we would expect a hazardous variable to exceed once every $10,000$ years, on average. Since historical data records are never available for this length of time, extreme value theory is therefore a crucial and necessary tool for ensuring one can approximate extreme behaviour at this return period in a principled fashion.

Once return level estimates have been obtained, uncertainty for this risk measure can be easily quantified using standard statistical methods. A wide range of techniques are also available for checking goodness-of-fit in model fits. Combined, these techniques allow us to approximate return levels at different return periods and perform robust risk analyses.

**Defining extreme events for two variables**

Unlike in the univariate setting, there is no standard practice for defining extreme events in the bivariate setting. This is in part due to the lack of a natural ordering for bivariate observations, i.e., given two observations $\left(x\_{1},y\_{1}\right),\left(x\_{2},y\_{2}\right)$, the statement $\left(x\_{1},y\_{1}\right) < \left(x\_{2},y\_{2}\right)$ lacks proper meaning.

To illustrate this point, suppose we have two variables $X$ and $Y$ for which we want to understand the joint extremal behaviour. Example of such variables have been plotted together in Figure 2. As can be observed, there are three different regions of significance: the region where only $X$ is extreme, the region where only $Y$ is extreme, and the region where both variables are extreme. Clearly to be able to perform a robust and detailed risk analysis for these two variables, we need some means of summarising the joint behaviour across these three different regions.



**Figure 2:** Plot of variable $Y$ against variable $X$. Three regions of extremal significance are also illustrated, with the blue, brown, and orange data points illustrating the region where only $X$ is extreme, the region where only $Y$ is extreme, and the region where both variables are extreme, respectively.

**Return curves**

Return curves are the natural extension of univariate return levels and are also interpreted in terms of a return period. Given a small probability $p$, the $p$-probability return curve is given by the set:

$RC(p) :=\left\{ \left(x,y\right) : Probability\left(X>x,Y>y\right) = p \right\}$.

This corresponds to all values $\left(x,y\right)$ that satisfy the equation $Probability\left(X>x,Y>y\right) = p$, defining a curve in the two-dimensional plane. Furthermore, values of $p$ close to zero correspond to rare joint exceedance events; consequently, the return curve provides a summary of joint extremal behaviour across three regions of interest illustrated in Figure 2. Within the literature, this set has a variety of labels including level curves (Salvadori & Michele, 2004), quantile curves (Chebana & Ouarda, 2011), joint probability curves (Gouldby, et al., 2017) and isolines (Cooley, Thibaud, Castillo, & Wehner, 2019). In analogy to return levels, we also define the return period to be $1/p$; given any point $\left(x,y\right)$ on the return curve, we would expect to observe the joint exceedance event $\{X>x,Y>y\}$ once, on average, each return period.

An example return curve is given in Figure 3 for the same data set illustrated as in Figure 2. One can observe how the return curve provides a boundary that intersects all three regions of significance. Furthermore, the probability of observing data points in the blue shaded region is equal to $p$; this is true for all equivalent regions whose corner point lies on the return curve.



**Figure 3:** Example of a return curve $RC(p) $ for variables $X$ and $Y$. The probability of observing data in sets of the form depicted by the shaded blue region is constant and equal to $p$ for any point on the return curve.

**Return curve estimation**

Return curve estimation has not been well studied in practice. Few methods have been proposed for their estimation, and of the existing approaches, the majority are limited in the forms of extremal dependence they can capture. Furthermore, similar to return levels, it is crucial to evaluate uncertainty and goodness-of-fit when estimating return curves in practice. The development of estimation, uncertainty quantification and diagnostic tools for return curves therefore represents an important and necessary area of research. Our goal for the thesis was to provide these tools to allow return curves to be used in practice.

Alongside tools for return curves, many of the available techniques for estimating extremal dependence structures are limited in the forms of dependence they can capture. Specifically, most approaches can only be used for modelling data sets exhibiting a class of dependence known as asymptotic dependence, even though another class, known as asymptotic independence, is frequently observed in practice. Furthermore, trends in environmental variables arising from climate change, for example, can result in extremal dependence structures that are not fixed in time. Therefore, another goal for the thesis was to provide flexible estimation techniques for the extremal dependence structure, including in cases where this structure is not fixed.

**Summary of research completed during the PhD**

The PhD thesis is made up of three independent papers detailed in Chapters 3, 4 and 5. As of June 2023, one paper has been accepted in the journal Environmetrics, while the other two of which have been submitted to other statistical journals. Below, we briefly outline the contributions of each paper in the context of return curve estimation.

* **Chapter 3**

Chapter 3 details the paper titled “New estimation methods for extremal bivariate return curves”. In this paper, we proposed two novel return curve estimation techniques. Unlike existing techniques, our proposed methods are not limited in the forms of extremal dependence they can capture; specifically, they are able to capture both asymptotically dependent and asymptotically independent data structures. Moreover, we compared our proposed estimation techniques to an existing approach, ultimately finding the technique based on the WT13 (Wadsworth & Tawn, 2013) model to be preferable. Our two novel curve estimation techniques are illustrated in Figure 4 for two metocean data sets. The considered metocean variables are of particular relevance for the structural reliability of offshore and coastal structures, and risk measures such as return curves are commonly used to inform the design basis for such structures (Jonathan, Ewans, & Flynn, 2014; Haselsteiner, et al., 2021). These data sets provide realistic examples with which to illustrate the utility of return curve estimates.



**Figure 4:** Return curves estimated for measured (left) and hindcast (right) metocean data sets, with $p=10^{-4}$.

Alongside developing and comparing curve estimation techniques, we also proposed a framework for evaluating uncertainty in curve estimates. This framework allows one to compute meaningful confidence intervals for return curves using a bootstrapping procedure, whereby a data set is resampled a large number of times to account for sampling variability (Politis & Romano, 1994). In practice, confidence intervals are crucial for ensuring the underlying variability in curve estimates is properly represented. Example $95\%$ return curve confidence regions are given for both metocean data sets in Figure 5, along with mean and median curve estimates.



**Figure 5:** Median (orange) and mean (brown) curve estimates, along with $95\%$ (black dotted) confidence regions obtained using bootstrapping for the measured (left) and hindcast (right) metocean data sets.

Furthermore, we proposed a diagnostic tool for assessing goodness-of-fit in return curve estimates. This tools checks that the estimated return curve accurately represents the underlying probability $p$ by comparing this value to empirical probability estimates computed at different points on the curve. To account for sampling variability, bootstrapping is again used to compute confidence intervals for these empirical estimates. If the true probability $p$ lies within the empirical confidence intervals for the majority of points on the curve, this indicates accuracy within the curve estimate, giving us confidence in the underlying estimation procedure. This diagnostic tool is illustrated in Figure 6 for the first metocean data set; the red line here represents the true probability, while the grey regions represent the confidence intervals for empirical probabilities at different points on the curve. As can be observed, the red line lies within the grey region for most points on the curve, indicating accuracy of the curve estimate.



**Figure 6:** Diagnostic plots for measured metocean data set. Solid red and black lines denote true and mean empirical estimates, respectively, and grey shaded region between dotted blues lines describe empirical 95% CI estimates.

The published version of Chapter 3 is freely available online (Murphy-Barltrop, Wadsworth, & Eastoe, 2023).

* **Chapter 4**

Chapter 4 details the paper titled “Modelling non-stationarity in asymptotically independent extremes”. In practice, we are often interested in estimating return curves for combinations of environmental variables. However, such variables typically exhibit non-stationarity, meaning they are not identically distributed in time. We can observe non-stationarity both within the magnitudes of individual variables (e.g., temperatures increasing over time) as well as the relationships between variables (e.g., the joint extremes of temperature and humidity changing over time); we refer to these trends as non-stationarity within the margins and extremal dependence structure, respectively. These trends in turn lead to non-stationary return curves which are not fixed in time. Accounting for non-stationarity in return curve estimation is therefore crucial for ensuring joint extremes in future climates can be properly evaluated.

To account for such trends, we proposed a modelling framework for capturing non-stationarity in return curve estimation. Very few approaches have been proposed in the literature that allow for the simultaneous modelling of non-stationarity within the margins and extremal dependence structure. Moreover, prior to our work, no approaches had been developed for estimating non-stationary return curves. Our novel modelling framework allows for this estimation, while simultaneously expanding upon existing modelling approaches.

We compared our approach to an existing method and found it to capture non-stationary extremal dependence structures more accurately over a wide range of examples. We also applied our methodology to a data set comprised of temperature and relative humidity projections obtained from the UKCP18 data base for the period 1981-2080 under emissions scenario RCP8.5 (Met Office Hadley Centre, 2018). Our modelling framework was able to accurately capture the observed non-stationary trends, and return curve estimates were obtained for the entire observation period: these estimates are illustrated in Figure 7. These curve estimates suggest that both the magnitudes and dependence between extremes of temperature and relative humidity change significantly over the time frame, illustrating the importance of accounting for non-stationarity when evaluating joint risks.

**Figure 7:** Return curve estimates at the regulatory standard over the observation period. Time is illustrated using a colour transition, with the curves for the start and end of the time frame labelled.

The preprint of Chapter 4 is freely available online (Murphy-Barltrop & Wadsworth, 2022).

* **Chapter 5**

Chapter 5 details a paper titled “Improving estimation for asymptotically independent bivariate extremes via global estimators for the angular dependence function”. Recall that in Chapter 3, we found that the WT13 model gave the best return curve estimates overall. However, implementation of the WT13 model had not been well studied. In particular, estimation of a key model quantity, known as the angular dependence function, has historically been carried our using a pointwise estimation technique. This results in non-smooth, unrealistic estimates of this function, which in turn alter the reliability and accuracy of the resulting return curve estimates.

In this paper, we proposed a range of smooth estimators for the angular dependence function. We compared these estimators to existing approaches, finding the smooth estimators to give less bias and variance for a wide range of examples. We also use these smooth estimators to obtain return curve estimates for river flow data sets from the North of England, UK. These return curves are illustrated in Figure 8 for five river gauge pairings, alongside the return curve estimates obtained using the pointwise, non-smooth technique. Model diagnostics indicate good performance of the estimated curves, and we can accurately capture each of the observed extremal dependence structures. This in turn allows us to better assess the probability of observing simultaneous floods at multiple river locations, allowing for a more robust and detailed risk analysis.

**Figure 8**: Estimated return curves for five pairings of river gauges. The purple, pink and green lines represent the curve estimates from three different angular dependence function estimators. In particular, the purple lines represent the historical pointwise estimation technique, while the pink and green lines represent the novel smooth techniques.

The preprint of Chapter 5 is freely available online (Murphy-Barltrop, Wadsworth, & Eastoe, 2023).

**Discussion**

The aim of this thesis was to address two questions: how does one define extreme risks in a bivariate setting, and how does one estimate such risks? These questions have been addressed through the development of novel estimation techniques for extremal dependence structures and return curves. Several of the modelling problems considered had previously been given little or no attention within the literature, and where methodology existed, our proposed methods were shown to outperform existing techniques in many cases.

As discussed in the introduction, accounting for joint risks from two (or more) variables is crucial for ensuring safe nuclear operations. In this context, the methods introduced in this thesis could allow for a more realistic and robust evaluation of joint extremes in a changing climate. For example, return curve estimates could be used to inform the design basis for future nuclear installations.

There are several ways in which our work could be extended. Perhaps most obviously, we have restricted attention to the bivariate setting throughout. This choice was made for several practical reasons. For example, visualisation and modelling become increasingly complex as the number of dimensions increases. Furthermore, little consideration had been given to extremal return curve estimation in the first place; applying the principle of parsimony, it is therefore best to first consider this problem thoroughly in the bivariate setting.

Given careful consideration, we believe all of the proposed methodology could be extended to the general multivariate setting. However, prior to extending any of the proposed methodology, we recommend considering the following questions: when is it beneficial to obtain extremal risk measures for three (or more) variables, and how can one estimate, utilise and interpret such measures? Of course, the answers to these questions are likely to be highly context dependent, and also require some degree of expert judgement.

Alongside extensions to higher dimensions, much of the proposed methodology for estimating extremal dependence structures could be developed further. These methods are not limited exclusively to return curve estimation and could offer utility more generally within applications of multivariate extreme value theory. Furthermore, we have made novel theoretical advancements and obtained interesting insights about how various modelling frameworks perform in practice. We therefore hope this thesis can provide both the motivation and basis for future research on the estimation of extremal dependence structures.

Finally, throughout this thesis, we chose to exclusively consider the bivariate risk measure known as a return curve. This was mainly because return curves provide a natural extension to return levels, which are the most widely used risk measure in the univariate setting. However, as described in Chapter 2.4 of the thesis, many alternative risk measures have been proposed within the literature (Serinaldi, 2015; Haselsteiner, et al., 2021). Similar to return curves, literature on the estimation of many of these risk measures for extreme scenarios is sparse, and similar techniques from bivariate extreme value theory are likely to be applicable. Many avenues for future research therefore exist - though we recommend that developments are made with caution to ensure the resulting estimates offer utility within practical applications.

# **Bibliography**

Chebana, F., & Ouarda, T. B. (2011, February). Multivariate quantiles in hydrological frequency analysis. *Environmetrics, 22*(1), 63-78. doi:10.1002/ENV.1027

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Springer London. doi:10.1007/978-1-4471-3675-0

Cooley, D., Thibaud, E., Castillo, F., & Wehner, M. F. (2019). A nonparametric method for producing isolines of bivariate exceedance probabilities. *Extremes, 22*(3), 373-390. doi:10.1007/s10687-019-00348-0

Davison, A. C., & Smith, R. L. (1990, July). Models for Exceedances Over High Thresholds. *Journal of the Royal Statistical Society. Series B: Statistical Methodology, 52*(3), 393-425. doi:10.1111/j.2517-6161.1990.tb01796.x

Dupuis, D. J. (1999). Exceedances over High Thresholds: A Guide to Threshold Selection. *Extremes, 1*(3), 251-261. doi:10.1023/A:1009914915709

Gouldby, B., Wyncoll, D., Panzeri, M., Franklin, M., Hunt, T., Hames, D., . . . Pullen, T. (2017). Multivariate extreme value modelling of sea conditions around the coast of England. *Proceedings of the Institution of Civil Engineers: Maritime Engineering, 170*(1), 3-20. doi:10.1680/jmaen.2016.16

Haselsteiner, A. F., Coe, R. G., Manuel, L., Chai, W., Leira, B., Clarindo, G., . . . Huseby, A. B. (2021, September). A benchmarking exercise for environmental contours. *Ocean Engineering, 236*, 1-29. doi:10.1016/J.OCEANENG.2021.109504

Institute for Nuclear Protection and Safety. (2000). Report on the flooding of the Blayais site on 27 December 1999. *Report on the flooding of the Blayais site on 27 December 1999*.

International Atomic Energy Agency. (2015). The Fukushima Daiichi Accident: Report by the Director General. *The Fukushima Daiichi Accident: Report by the Director General*.

Jonathan, P., Ewans, K., & Flynn, J. (2014). On the estimation of ocean engineering design contours. *Journal of Offshore Mechanics and Arctic Engineering, 136*(4), 1-8. doi:10.1115/1.4027645

Mattei, J. M., Vial, E., Rebour, V., Liemersdorf, H., & Turschmann, M. (2001). Generic results and conclusions of re-evaluating the flooding protection in French and German nuclear power plants. *Generic results and conclusions of re-evaluating the flooding protection in French and German nuclear power plants*.

Met Office Hadley Centre. (2018). UKCP18 Regional Climate Model Projections for the UK. *UKCP18 Regional Climate Model Projections for the UK*.

Murphy-Barltrop, C. J., & Wadsworth, J. L. (2022). Modelling non-stationarity in asymptotically independent extremes. *arXiv, 2203.05860*.

Murphy-Barltrop, C. J., Wadsworth, J. L., & Eastoe, E. F. (2023). Improving estimation for asymptotically independent bivariate extremes via global estimators for the angular dependence function. *arXiv, 2303.13237*.

Murphy-Barltrop, C. J., Wadsworth, J. L., & Eastoe, E. F. (2023, February). New estimation methods for extremal bivariate return curves. *Environmetrics, e2797*, 1-22. doi:10.1002/env.2797

National Diet of Japan. (2012). The official report of The Fukushima Nuclear Accident Independent Investigation Commission - Executive summary. *The official report of The Fukushima Nuclear Accident Independent Investigation Commission - Executive summary*.

Office for Nuclear Regulation. (2014). Safety Assessment Principles - 2014 edition (Revision 1, January 2020). *Safety Assessment Principles - 2014 edition (Revision 1, January 2020)*, 1-226.

Office for Nuclear Regulation. (2018). NS-TAST-GD-013. *NS-TAST-GD-013*, 1-84.

Politis, D. N., & Romano, J. P. (1994, December). The Stationary Bootstrap. *Journal of the American Statistical Association, 89*(428), 1303-1313. doi:10.1080/01621459.1994.10476870

Ross, E., Astrup, O. C., Bitner-Gregersen, E., Bunn, N., Feld, G., Gouldby, B., . . . Jonathan, P. (2020, January). On environmental contours for marine and coastal design. *Ocean Engineering, 195*, 106194. doi:10.1016/J.OCEANENG.2019.106194

Salvadori, G., & Michele, C. D. (2004). Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research, 40*(12), 1-17. doi:10.1029/2004WR003133

Serinaldi, F. (2015). Dismissing return periods! *Stochastic Environmental Research and Risk Assessment, 29*(4), 1179-1189. doi:10.1007/s00477-014-0916-1

Wadsworth, J. L., & Tawn, J. A. (2013). A new representation for multivariate tail probabilities. *Bernoulli, 19*(5 B), 2689-2714. doi:10.3150/12-BEJ471