NARSIS

New Approach to Reactor Safety ImprovementS

WP1: Characterization of potential physical threats due to different external hazards and scenarios

Del 1.3 - Improved methodologies for extreme weather and flooding hazard assessment

This project has received funding from the Euratom research and training programme 2014-2018 under Grant Agreement No. 755439.

Project Acronym:	NARSIS
Project Title:	New Approach to Reactor Safety Improvements
Deliverable:	Del 1.3 - Improved methodologies for extreme weather and flooding hazard assessment
Month due:	36 **Month delivered:** 37
Leading Partner:	IRSN
Version:	Final

Primary Author: IRSN: Lucie PHEULPIN, Vito BACCHI, Vincent REBOUR, Bernard CHAUMONT, Yves GUIGENO, Christian LINCOT, Emmanuel RAIMOND

Other contributors:

- TU Delft: Pieter VAN GELDER, Phil VARDON, Varenya MOHAN, Michael HICKS
- EDF UK: Hugo WINTER
- KIT: James DANIELL, Andreas SCHAEFER

Deliverable Review:

- **Reviewer #1:** Slawomir POTEMPSKI (NCBJ) **Date:** 08/20
- **Reviewer #2:** Anne DUTFOY (EDF) **Date:** 25/09/20

Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Table of contents

List of Figures

List of Tables

List of Abbreviations

Acronym	Definition	Page
ACF	Autocorrelation Function	30
AD	Asymptotic Dependence	27
AFDA	Approximate Full Distribution Approach	53
AI	Asymptotic Independence	27
ANCOVA	Analysis of covariance	36
ANOVA	Analysis of Variance	40
ARMA	Autoregressive-moving-average models	22
ASAMPSA_E	Advanced Safety Assessment Methodologies: Extended PSA	12
ASCE	American Society of Civil Engineers	55
ASL	Above Sea Level	62
CDF	Cumulative Distribution Function	23
CMIP	Coupled Model Intercomparison Project	32
CoC	Consequence Class	57
CORDEX	Coordinated Regional Downscaling Experiment	32
DWD	Deutsche Wetter Dienst (German Weather Service)	18
eCDF	Empirical Cumulative Distribution Function	37
ECMWF	European Centre for Medium-Range Weather Forecasts	33
EEA	Extreme Event Analyser	17
EGU	European Geosciences Union	12
E-OBS	ENSEMBLES daily gridded observational dataset	33
ERA	ERA atmospheric reanalysis	33
ERSST	Extended Reconstructed Sea Surface Temperature	34
ESWD	European Severe Weather Database	17
ETCCDI	JCOMM Expert Team on Climate Change Detection and Indices	22
EU	European Union	11
EVA	Extreme Value Analysis	13
EVS	Extreme Value Statistics	19
FSA	Fault Sequence Analysis	18
GCM	Global Climate Model	32
GEV	Generalised Extreme Value	24
GPD	Generalised Pareto Distribution	23
GSA	Global Sensitivity Analysis	11
HADEX2	Hadley Centre Gridded Station-Based Indices (EX2)	33
HDMR	High-Dimensional Model Resolution	40
HVAC	Heating, ventilation, air conditioning	15
IAEA	International Atomic Energy Agency	16
IID	Independent and Identically Distributed	25
IRSN	L'Institut de Radioprotection et de Sûreté Nucléaire	11
JCOMM	Joint Technical Commission for Oceanography and Marine Meteorology	22
MCMC	Markov Chain Monte Carlo Methods	77
NOAA	National Oceanic and Atmospheric Administration	21
NPP	Nuclear Power Plant	11
PACF	Partial Autocorrelation Function	30
PDF	Probability Density Function	37

POTs	Peaks Over Threshold	12
PRIMAVERA	PRIMAVERA EU Project	11
PSA	Probabilistic Safety Assessment	11
QQ	Quantile-Quantile	66
RC	Reliability Class	56
RCM	Regional Climate Models	33
RCP	Representative Concentration Pathways	32
SA	Sensitivity Analyses	40
SI	Sensitivity Indices	40
SSC	Structure, System and Component	15
SSP	Shared Socioeconomic Pathways	33
UK	United Kingdom	13
UQ	Uncertainty Quantification	11
US	United States	21
WCRP	World Climate Research Programme	22
WENRA	Western European Nuclear Regulators' Association	14

1 Executive Summary

A number of methodologies have been presented for potential use within PSAs within Europe looking at extreme weather and flooding. When looking at multi-hazard, the duration of the events is extremely important in terms of their potential impacts when overlapping or coinciding hazards are analysed. Durations and correlation matrices have been reexamined as part of this deliverable.

We are likely to miss important information if we only estimate return levels (or equivalent metrics) focusing on the instantaneous severity of events. Several different approaches were outlined which are available to utilize information about the duration of events to improve our estimation of natural hazard risk. Our focus is on approaches from within the class of extreme value models as these are models with mathematical justification for extrapolating beyond the range of the data and are the most commonly used approaches for this type of analysis within the nuclear industry.

In terms of standard approaches for the application of climate change results to NPP design, guidelines have been put forward in various forms. PRIMAVERA-H2020 is the latest EU project to provide high resolution analyses on a smaller scale appropriate to NPP site levels. This downscaling and higher resolution allows for a better application of possible effects of climate change. As the PRIMAVERA models and other downscaled models become available at a reasonable resolution, they should be integrated in conjunction with the usual station data which is usually collected at the site, and from long-run stations around the site.

At present structural design codes are based on the assumption of stationary climate conditions. The reliability targets are typically specified, or understood, in terms of the annual probability of failure over the design life of the structure, typically 50 to 75 years. In the reliability-based design framework, the nominal design loads are specified in terms of return period or an upper percentile, e.g., 95^{th} percentile, of the annual maximum load distribution. Further, load factors are specified to calculate the factored design load. The load factors are calibrated with respect to the target reliability level (i.e., annual probability of failure).

As the climate change effects are becoming evident, national code authorities are interested in evaluating the effect of non-stationary climate conditions on the structural design provisions. More specifically, it is being recognized that the reliability-based calibration approaches used under the traditional assumption of stationary climate condition cannot be directly extended to non-stationary climate cases. For example, the concept of the return period is inapplicable to non-stationary climate, as the distribution of the time between occurrences of the load events is no longer invariant. Similarly, the annual probability of failure is no longer constant in the non-stationary climate. A technical discussion was developed and such issues were addressed using the theory of stochastic processes.

The assessment of external hazard, as flooding or earthquake, often relies on numerical models which allow assessing the variables of interest (*e.g.* water depths, ground acceleration, *etc.*). However, these models are affected by uncertainties which can be quantified through studies called “Uncertainty Quantification” (UQ) and “Global Sensitivity Analysis” (GSA). The UQ attempts to describe the whole set of possible outputs considering the inputs not perfectly known. The GSA aims to determine the most influent inputs to an output behaviour, as the non-influential ones. These two types of analyses are complementary and both classically suppose that the input parameters of the numerical models are independent. Especially, in the literature, most studies consider the inputs independent which may not be the case. A global review of the different methods used for UQ and GSA when model inputs are considered dependent was undertaken and was applied to hydraulic studies (Table 2). In the field of flooding hazard, this kind of analysis is still missing. For the interested reader, an IRSN report on existing methodologies for UQ and GSA with dependent inputs followed by an application to a simplified case of inundation has been written on this topic (Pheulpin, 2020). The major points from this report were reported in the sections 5.1 and 5.2. In addition a research article on the application of these methods to the large hydraulic model of the Loire

River is in progress. Some of the results have been presented in the EGU 2020 conference (Pheulpin and Bacchi, 2020).

The main conclusions of the work have shown that:

- There is a strong dependence between the hydrograph parameters (d, qmax, tm) and this dependency can be taken into account by using copulas.
- The use of metamodels is very useful for uncertainty analysis studies (almost all with limited computation resources).
- There is a limited impact of inputs dependency in uncertainty quantification in this study.
- The duration and time to peak inputs have strong influence on the outputs.
- The hydrograph shape should not be ignored in hydraulic studies.

It is important to note that this deliverable focusses on only a few of the key issues on the list of ASAMPSA_E indicating various issues associated with extreme weather analyses, and this field is continuing to evolve and requires more work in the coming years to address the other issues:

- Limitations in modelling and forecasting the physical phenomena and conditions leading to extreme hazard;
- Uncertainties in estimation of the impact of climate change on extreme meteorological events;
- Lack of site-specific data and limitations of spatial modelling and downscaling methods;
- Difficulties in quantification of uncertainties for common-cause failures;
- Difficulties in integrated modelling of hazard internal and external impact assessment;
- Limitation in determining the occurrence frequency of extreme weather conditions;
- Correlation among an extreme weather event induced failure modes and on the quantification of correlation coefficients.

The use of an output of a multivariate extreme value model to improve upon standard univariate extreme value analysis techniques that require declustering to be undertaken to extract independent peaks over the threshold (POTs) was presented with various benefits to this approach:

- We utilize more of the available information from the data by not having to discard exceedances which aren’t peaks but are still extreme;
- Since we don’t have to decluster the data, there is no need to pick a declustering parameter (e.g. the run length), which is a subjective choice;
- Fitting the model in the Bayesian framework permits the estimation of the predictive return level which rolls all uncertainty about the parameters and randomness in future observations into a single estimate.
- Standard frequentist approaches with declustering lead to higher estimates than the Bayesian posterior and predictive estimates;
- Uncertainty estimates for the fits to POTs are wider than those using the Bayesian approach;
- The subjective choice of declustering parameter (here the run length) can have a large impact on the best estimates and uncertainty intervals.

The complexity of extreme weather and flood modelling as well as the non-stationary nature and lack of data makes analysis difficult with major uncertainties. Some of the tools and methods presented within this report can address some of these issues, however the reader is directed to the key documents section at the start of this report.

2 Background and Introduction

2.1 Scope and Objectives of the Deliverable

This task focuses on flooding hazards and in particular aims at defining methods to address combination of phenomena in the framework of hazard curves assessment. Flooding hazards are an ideal candidate for this analysis because different natural mechanisms, dependent or independent, have to be taken into account. For example, a major flooding event could be caused by joint flooding of two rivers, or by a high storm singe breaching a coastal defence, but could as easily be caused by extreme waves occurring at the same time as severe local rainfall. This aspect poses difficulties to standard statistical analysis approaches which use univariate extreme value analysis to consider extremes of different natural hazards separately

To develop a better understanding of how a set of different hazards can combine to result in a major flooding event, multivariate extreme value models are required. These models are flexible and can be used to model: (i) dependence through time. i.e. model events that last for a prolonged period of time; (ii) dependence across space, i.e. events that might hit multiple locations simultaneously. A simple example of a multivariate model is the Joint Probability Method used to convolute tide and singe distributions to obtain estimates of extreme sea levels. More advanced recent studies using multivariate extreme value models have investigated the relationship between significant wave heights and wave period along with the directional effects on this relationship. Multivariate extreme value theory on hazard combinations and Atlantic storms have been examined. Novel extreme value models to estimate the track and occurrence probability of extreme storms which strike the UK and other parts of Western Europe. In particular, these models should be able to provide new insights into the occurrence of hazard combinations, including events that are worse than any previously seen in the observational record.

Uncertainties associated to those methods will be defined and studied in detail, in order to give a global uncertainty assessment of input and output parameters of the hazards combination step. In practice. Uncertainty Analysis (UA) and a Global Sensitivity Analysis (GSA) will be performed. UA will lead to quantify uncertainties. GSA will go one step further, allowing ranking uncertain parameters effects on model output variability. The GSA approach (using screening method or variance decomposition based methods) allow to identify parameters that highly influence the model output variability. The UA and GSA are highly computational resources demanding. High Performance Computer will be used, warping a parametric environment developed by IRSN (Prométhée). with statistical and hydraulic codes, to launch computation.

The probabilistic assessment of failures of dams and other geotechnical safety structures by external hydraulic effects (e.g. floods), will be incorporated by utilizing stochastic methodologies.

2.2 Organisation of the Deliverable

The first part of this deliverable is dedicated to a quick introduction of key concepts with significant uncertainties going beyond NARSIS deliverables Del1.1 and 1.6 for extreme weather and flooding. Here, concepts of the interactions between hazard types and duration are shown, and provides the background for the multi-hazard interactions in the software. Aspects of climate change results are also discussed and the potential need for implementation in the multi-hazard environments that are present.

A methodology for duration analysis of events is then presented in section 4, with empirical approaches and time-series models as well as univariate and multivariate extreme value analysis methodologies being put forward. The drawbacks of univariate EVA are shown, and multivariate EVA is then presented with an application to temporal dependence being shown including Markov models and simulating events. This is then shown with the application of

improving estimates of return periods. Climate change aspects are then shown with the latest models discussed and their need for integration.

In section 5, methodologies on uncertainty quantification and global sensitivity analysis, with dependent inputs are examined with a synopsis of existing methodologies being given before examining a simplified and complex case of inundation modelling. A full hydraulic model of the Loire River is explored.

Probabilistic models for non-stationary climate conditions and its effects on reliability targets are discussed in section 6 where the stationary climate condition models are shown for target reliability taking into account different acceptable levels of risks – either personal, social or economic. The target reliability under non-stationary conditions are then examined.

An application of the approaches for assessing the duration of extreme events are then examined with respect to heatwaves, windstorms in section 7, where examples for decommissioned plant locations are used with respect to multivariate extreme value analysis (EVA).

2.3 Existing Guidance: Key Documents

There exist a large number of documents which are key reading for this analysis pertaining to the various hazard models associated with this analysis. Single hazards have been explored in this

IAEA, 2003. Meteorological Events in Site Evaluation for Nuclear Power Plants. Safety Guide NS-G-3.4, 34pp.

IAEA, 2003. Flood Hazard for Nuclear Power Plants on Coastal and River Sites. Safety Guide No. NS-G-3.5, Vienna.

IAEA, 2011. Meteorological and Hydrological Hazards in Site Evaluation for Nuclear Installations Specific Safety Guide SSG-18, 146pp.

WENRA-RHWG, 2013. Report Safety of new NPP designs. Study by Reactor Harmonization Working Group RHWG, 52pp.

IAEA, 2003. Extreme External Events in the Design and Assessment of Nuclear Power Plants. Tecdoc 1341, 109pp.

IAEA, 2003. Flood Hazard for Nuclear Power Plants on Coastal and River Sites. Safety Guide NS-G-3.5, 83pp.

The following ASAMPSA_E deliverables (www.asampsa-e.eu/deliverables-library) are important:

D50.16 Report 2 – Guidance document on practices to model and implement FLOODING hazards in extended PSA (final version)

D50.17 Report 3 – Guidance document on practices to model and implement EXTREME WEATHER hazards in extended PSA (final version)

D50.18 Report 4 – Guidance document on practices to model and implement LIGHTNING hazards in extended PSA (final version)

3 Existing Extreme Weather and Flooding methodologies around Europe as a synthesis

The existing extreme weather and flooding methodologies have been covered as part of the NARSIS deliverables Del1.1 and 1.2 studies, however a few specific items have not been covered which should be mentioned within this deliverable.

It is important to revisit the list which was listed as “open issues” as part of the PSA review during the ASAMPSA_E project D22.2:

- Limitations in modelling and forecasting the physical phenomena and conditions leading to extreme hazard;
- Uncertainties in estimation of the impact of climate change on extreme meteorological events;
- Lack of site-specific data and limitations of spatial modelling and downscaling methods;
- Unclear application of complex probabilistic models, like mixed distributions;
- Difficulties in quantification of uncertainties for common-cause failures;
- High uncertainties in tornado data and tornado frequencies estimation;
- Validation limitations for tornado physical phenomena and its impact models;
- Difficulties in integrated modelling of hazard internal and external impact assessment;
- Lack of applicable experiments for the improvement of the extreme wind consequence models;
- Limitation in determining the occurrence frequency of extreme weather conditions;
- Scope of the fragility curves is limited (especially for extreme temperatures influence);
- Adequate and practically applicable methodology for assessing the failure probability of indoor SSC’s (taking into account outside temperature and HVAC capacity);
- Correlation among an extreme weather event induced failure modes and on the quantification of correlation coefficients;
- Uncertainties in operational strategy under harsh weather conditions;
- Winds and tornadoes induced missiles estimation and modelling.

3.1 Motivation – Duration and variability of natural hazard events

Extreme natural hazard events have the potential to cause disruption and safety issues on many different levels with impacts ranging from built-infrastructure and economies to public health and societal shifts. Natural hazards can manifest themselves in several ways, ranging from short and sharp shocks to a system (e.g. tornado) through to longer-term chronic issues (e.g. sea level rise, droughts). Figure 1 (taken directly from Gill and Malamud, 2014) provides a summary of the different spatial and temporal scales associated with different natural hazards.

Figure 1 makes it clear that if we do not consider the duration (or spatial) aspect of natural hazards then we shall not properly consider the risks posed to any infrastructure or societal systems. This motivates the necessity of considering spatio-temporal aspects. In this work we will focus on different approaches that are available for modelling long duration extreme events. Section 3.2 and NARSIS deliverable Del1.1 provide some motivational case studies of recent extreme events.

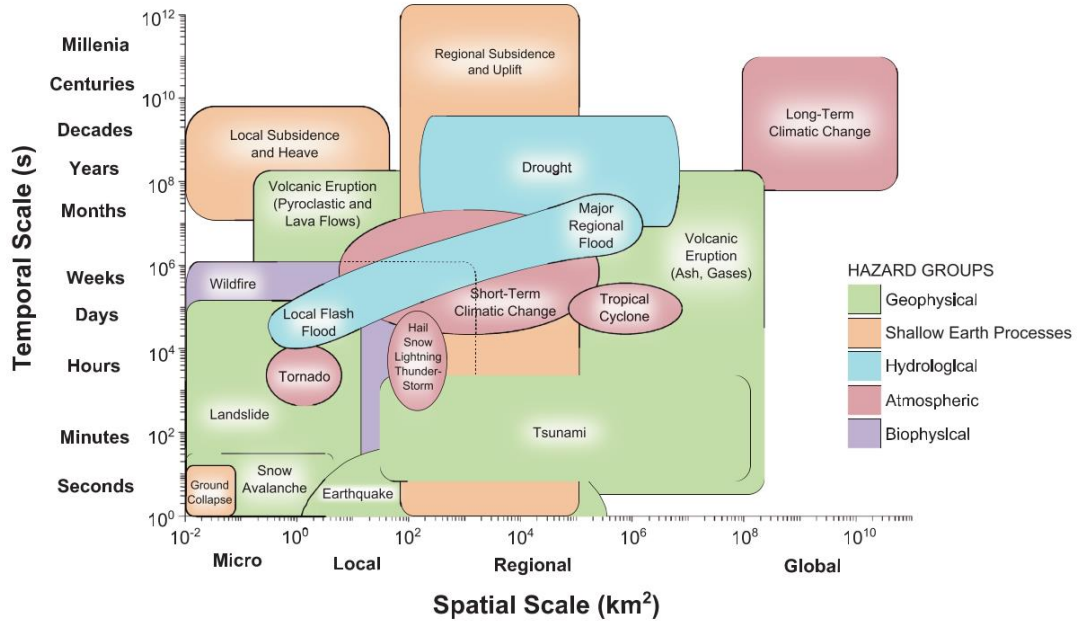


Figure 1: Summary of different spatial and temporal scales associated with various natural hazard phenomena (Gill and Malamud, 2014)

Using the base of the ASAMPSA-E with respect to climatological and hydrological parameters and their interactions, adjustments have been made to the original table for certain interactions. The interaction and correlation matrix shows the directionality of the hazards. For any location, the interactions must be examined as to what impacts they could have on the NPP site and components.

The crosses in the above interaction matrix refer to possible combinations which can be excluded under the IAEA (2011) criteria which state that they can be excluded if (a) the hazard combinations are not physically possible; (b) that they have the same joint effect on the NPP as the single effects; and (c) the annual frequency of the hazard interaction is equal or less than a threshold.

As these combinations differ markedly across Europe, most interactions are possible although would likely be screened out. Some hazard parameter thresholds can indeed be used as defined through existing projects for certain criteria for each of these above hazard types. Within ASAMPSA_E 50.17, it was shown that each of these parameters have different cutoff levels depending on the country for critical thresholds with examples of Bulgaria, Lithuania and Hungary and Sweden given.

The methodology detailed in NARSIS Del1.1 in section 5.2 can be used for the choice of hazards taking into account dependency, however, the nature of cascades is such that care must be taken to not screen out plausible combinations.

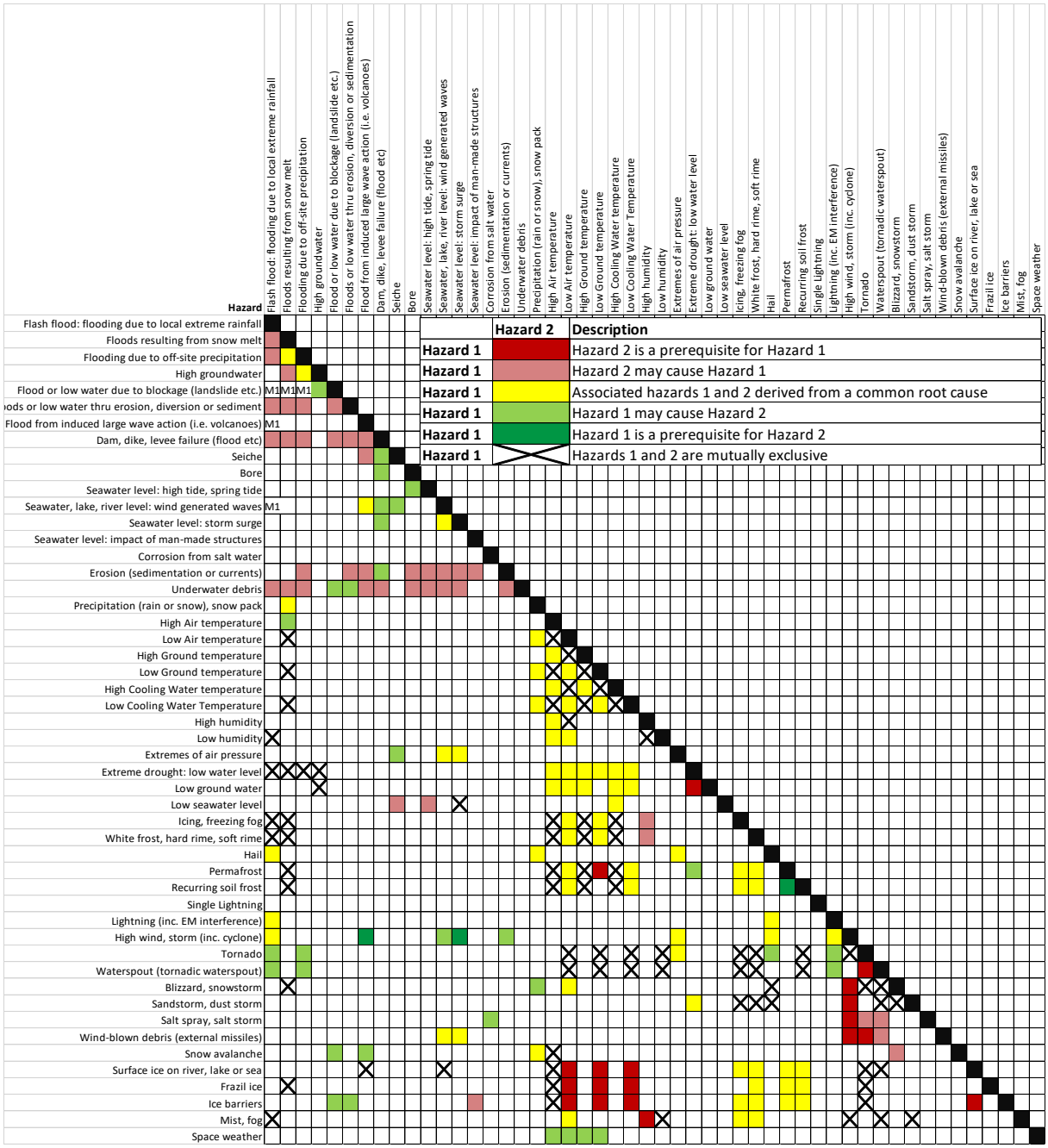


Figure 2: Interactions between different meteorological and hydrological hazards adapted from ASAMPSA_E showing the directionality and causal links between them.

Each of the hazard types has a duration associated with the event. These event overlaps form the basis for the possible interactions of the hazards. Unfortunately, there exists no historical database with enough completeness to cover secondary hazard interactions with ESWD being the closest as discussed in NARSIS Del1.1 and 1.6.

During the discussion of assessment methods for extreme values and hazard combinations: two methods are commonly used: the IAEA Fault Sequence Analysis methodology (Kuzmina et al., 2013), and the ExtremeEventAnalyzer (EEA) tool of the Lloyds Register with IAEA.

The EEA tool is further explained in ASAMPSA_E 50.17 so will not be further explained here, but needs the initiating events which will trigger a certain damage threshold at the site, upon which combinations of the hazard events can then be established. This is done after the initial hazard selection, thus the need to examine all possible combinations of hazards is required

in advance. The Fault sequence analysis (FSA) methodology is rerun using the hazard susceptibility limits of the site response analyses in order to consider the combined occurrence of these hazards.

Other national datasets such as DWD provide long-range station and interpolated raster data for various parameters, however do not provide enough details on each of the below parameters to undertake an analysis on a detailed scale of hazard combinations.

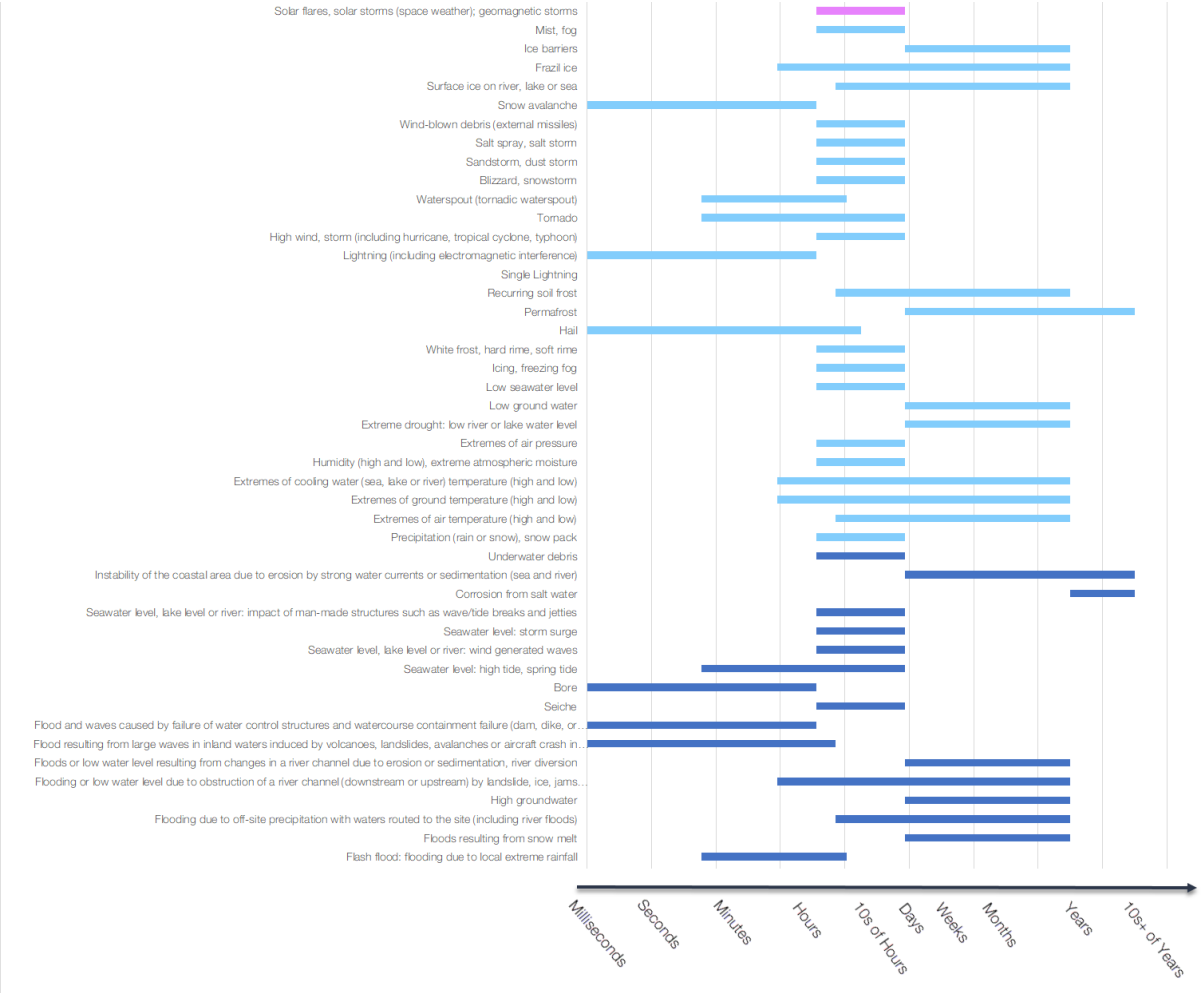


Figure 3: Normal durations of meteorological and hydrological hazards.

These time periods should also be seen as being the minimum interaction time for a particular hazard. In terms of cascades, site-specific analysis is required with the interactions to ensure that appropriate durations are defined for the interaction period of the cascades.

An example of this is the use of low-water data from stream gauges for the potential of quake lakes or volcano related lakes. The seasonal variability of the low water vs. the probability of a major rockfall provide a key issue. Topographic analysis above, and below a potential site should be examined for potential locations where such a lake could form, and what the potential impacts would be when dam break occurs, or backfilling occurs.

For such an example, the duration of the event is taken from the time period upon which the low-water is below a certain threshold and the overlap with a potential volcanic eruption or earthquake during that time period. Given that traditional landslides are correlated to rainfall in many cases, copula approaches (discussed in section 4.4.2) should likely be applied to

work out the multivariate distribution of rainfall, landslide probability and flow of the river. Given that less than 80 years of data are generally available for most stretches of river, then extreme events from history as well as various EVS methods (discussed in section 5 and in Del1.6) are needed to examine maximum possible durations associated with such an event. However, the effect of climate change on parameters such as flow gauge records (i.e. as shown here in Figure 4) is a further uncertainty, given the fact that the record from 1930-2016 is likely non-stationary.

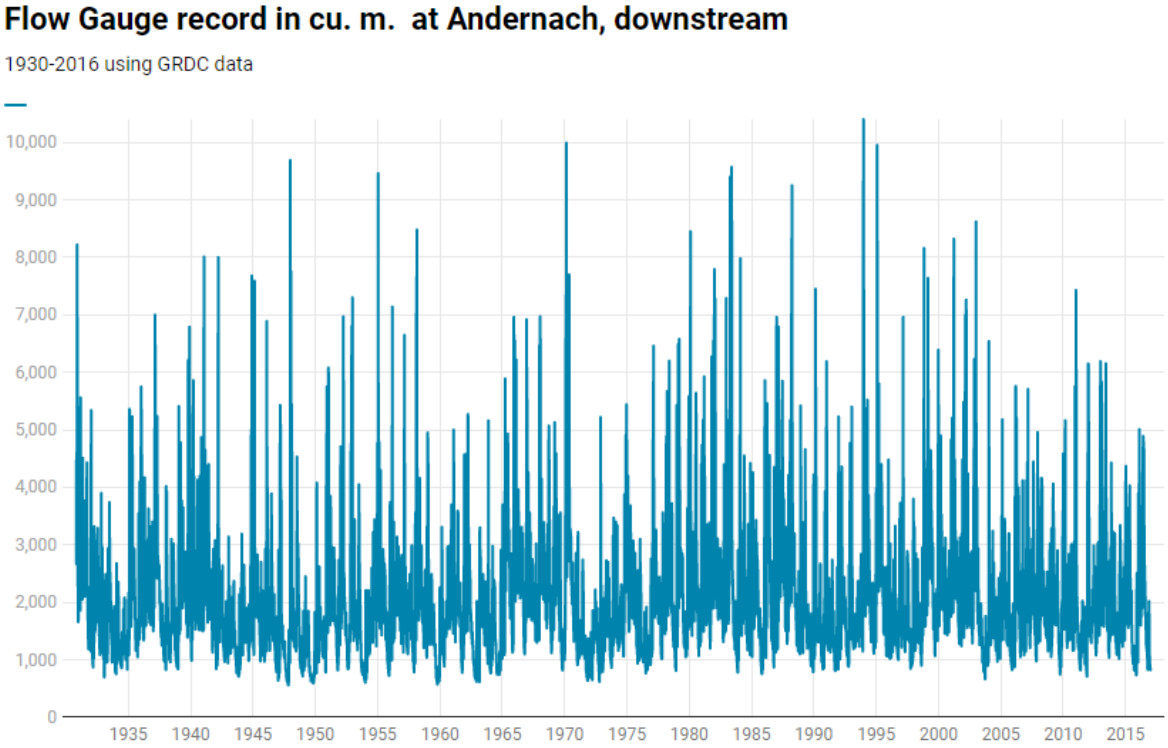


Figure 4: Example figure of flow gauge data from Andernach in Germany via GRDC

3.2 Case studies

In this section, three recent natural hazard events are outlined, including: (i) the European heatwave of 2003; (ii) Storm Xynthia 2010; (iii) Hurricane Harvey 2017. The risks posed by each of these events have been exacerbated by the fact that they have persisted over many days to weeks.

3.2.1 European heatwave in 2003

The European heatwave of 2003 lasted for a period from July to August, striking many countries across western and central Europe. The event was caused by large-scale anticyclonic conditions that started during May and persisted until August (Black et al., 2004). The average temperature across Europe exceeded that of any previous summer for the last 500 years, although the main impact on mortality was the exceptionally hot fortnight at the start of August (Trigo et al., 2005). Temperature extremes were broken in many locations across Europe; as an example, the UK temperature record was broken with 38.5ºC observed at Brogdale in Kent on 10th August.

There were many impacts of the heatwave event, including:

- Human effects: about 15,000 excess deaths were attributed to the heatwave in France. Other excess heat-related deaths were observed in the UK (2,000), Portugal (2,100), Italy (3,100), Holland (1,500) and Germany (300).
- Agriculture: much livestock was killed during the event with an estimated cost to European farming of €13.1 billion.
- Transport: in some locations, railways buckled in the heat and road surfaces melted.
- Water supplies: hosepipe bans were introduced in some parts of the UK.
- Energy: two nuclear power plants in Germany were forced to close temporarily.

3.2.2 Storm Xynthia 2010

Storm Xynthia was a windstorm that crossed western Europe from 27th February to 1st March 2010. Xynthia developed in the Atlantic off the coast of Madeira, crossing the Canary Islands and initially striking the Iberian Peninsula. The greatest impacts of the storm were observed in mainland France, where the storm intensified across the Bay of Biscay. This led to extreme wind speeds, storm surges and wave heights across much of the Atlantic coast of France. The event continued into Germany where it eventually dissipated on 1st March.

The most intense observed wind gusts observed at the coast were 160 km/h (with inland wind speeds reaching gusts of 120-130 km/h) (Meteo France, 2020); these wind speeds were not as intense as those observed during previous storms Martin and Lothar. However, the co-occurrence of extreme storm surges on a high tide did lead to unprecedented flooding in some areas. For example, in La Rochelle on February 28th the total water height reached 8.02 m (an increase of 1.53 m), unheard of since the start of tide gauge observations in 1997.

Storm Xynthia had multiple different impacts:

- At least 51 people were killed in France with a further 12 people killed in across Europe.
- The storm cut power to over a million homes in France and Portugal respectively.
- In the Vendée, 11,000 ha of agricultural lands were impacted by the salt seawater.

3.2.3 Hurricane Harvey 2017

Although not a hazard directly impacting Europe, Hurricane Harvey in 2017 provides an interesting case study of an event with greater impacts due to an extended duration. Harvey was first picked up as a weak tropical storm on 17th August and then rapidly moved and intensified to a category 4 hurricane by 26th August when it made landfall near Corpus Christi in Texas. Having made landfall, the hurricane stayed over land for 2 days and then returned into Matagorda Bay on 28th August, tracking north and making final landfall on 30th August before weakening to a tropical cyclone and eventually dissipating. Figure 5 shows the best track positions for Hurricane Harvey from its genesis to the point when it dissipated.

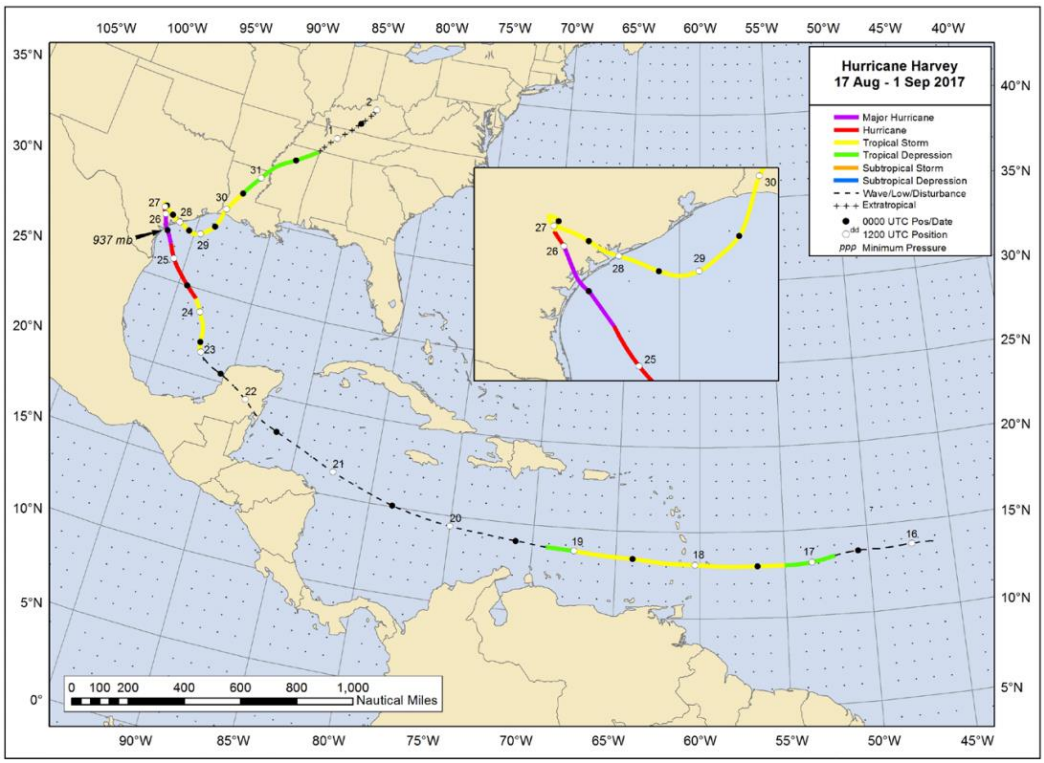


Figure 5: Best track positions for Hurricane Harvey (Blake and Zelinsky, 2018).

Some meteorological statistics are provided in Blake and Zelinsky (2018) for wind speed, storm surge and extreme rainfall. The maximum winds were generated in the first few hours after the landfall in Texas with an estimated intensity of 115 knots (59 m/s). The combined effects of storm surge and high tide led to inundation levels of 6-10 ft (2-3 m) across various parts of coastal Texas. The real significant hazard during the Harvey event was extreme precipitation with records observed in terms of both peak rainfall and the area flooded. The highest storm rainfall amount was reported as 60.58 inches (1.54 m) near Nederland in Texas; this exceeded the previous record of 52 inches (1.32 m) in Hawaii with the previous continental US record of 48 inches (1.22 m) at Medina in Texas. Such extreme values were mainly caused by the unique meteorology of the event, which moved quite slowly across the region (getting caught on a slow-moving frontal system) and moved back over the sea after making initial landfall in the US. This meant that the system persisted for longer in an area conducive to continued rainfall and therefore the duration of the event led to the very extreme accumulations observed.

Hurricane Harvey led to several impacts:

- 68 direct deaths in the US;
- Approximate NOAA damage estimate from Harvey of $125 billion;
- 300,000 structures flooded with 336,000 customers losing power;
- 40,000 people needed to evacuate their homes.

4 Methodology for duration analysis of events

The motivating examples from section 3.2 illustrate that a better understanding the duration of natural hazard events can lead to a much better appreciation of the risks posed by extreme events. We are likely to miss important information if we only estimate return levels (or equivalent metrics) focusing on the instantaneous severity of events. In this section, we outline different approaches available to utilize information about the duration of events to improve our estimation of natural hazard risk. Our focus is on approaches from within the class of extreme value models as these are models with mathematical justification for extrapolating beyond the range of the data and are the most commonly used approaches for this type of analysis within the nuclear industry.

In section 4.1.1 we outline some empirical approaches and time-series models for modelling the duration of events. This leads onto an initial discussion of EVA approaches in section 4.1.2 which seeks to provide a framework for assessing the duration of events using EVA. Specifically, there is a recap of univariate EVA (section 4.2) and a discussion about the drawbacks of using such standard approaches for estimating extremes of natural hazards (section 4.3). Finally, the theory behind multivariate extreme value models is outlined in section 4.4 and their use here is described in sections 4.5 and 4.6.

4.1.1 Empirical approaches and time-series models

The simplest approaches to understanding the duration of events involve empirical measures applied directly to observation data. In the extremes this is likely to involve the setting of a critical threshold and then counting the number of observations that exceed this threshold and whether these exceedances tend to cluster. There are a variety of different climate indices which seek to summarise the duration of events for different hazards. The joint CCl/WCRP-Clivar/JCOMM Expert Team on Climate Change Detection and Indices (ETCCDI) has defined a core set of descriptive indices of extremes. The indices describe particular characteristics of extremes, including frequency, amplitude and persistence. Some examples of heatwave indices are given in Dupuis (2012) and Alexander et al. (2006).

One approach commonly used to assess the duration of extreme events within the field of extreme value theory is the extremal index. This is a quantity (often denoted as θ) which summarises the amount of clustering within the extremes of a data series. The measure θ takes a value between 0 and 1 and the reciprocal of the value is broadly defined as the average number of exceedances within a cluster. This measure can be estimated empirically from a data series, using a variety of different non-parametric approaches. We shall return to this measure in section 4.5 and highlight how this can be estimated using an extreme value model.

The main issue with any of the empirical approaches mentioned above is that we are constrained by the information contained within the observational record. It is not possible to estimate the duration of events that are more severe or longer than those in the observational record using these approaches. For this purpose, it is necessary to fit a statistical model to the observation data to generate inferences. Standard time-series models such as ARMA models are often used in this situation. However, a key drawback of this type of model is that they are not well justified for extrapolating beyond the range of the observed data. Methods based upon extreme value theory will be better justified for doing this and are thus preferred for this type of analysis. This type of approach is expanded further through sections 4.1.2 to 4.6.

4.1.2 Extreme value analysis

Extreme value analysis can be used to estimate the duration of extreme events – this approach can help to get around some of the issues with empirical approaches and time-series modelling. However, a more detailed approach is required than when using univariate EVA to estimate return levels.

Specifically, it is necessary to fit a multivariate extreme value model to capture the

dependence between consecutive extreme values within the time series. The general approach can be formulated in the following steps:

- Fit a univariate EVA model to the data to estimate the marginal distribution (here assumed to be a GPD above some high threshold and the empirical CDF below the threshold) – section 4.2.
- Choose the order of the Markov model that is going to be fitted – section 4.5.1
- Fit preferred dependence model to extremes of $X_{t+k}|X_{t+k-1}, \dots, X_t$ (where k is the order of the Markov chain) to estimate the parameters of the chosen multivariate extreme value model – section 4.4
- Use the dependence parameters of the multivariate extreme value model to construct a simulated event (called the tail chain) with the same statistical properties as the original data – section 4.5.2
- Repeat the simulation N times (for a large N, usually greater than 10,000). Obtain estimation of the duration of an event by obtaining relevant summary statistics across the simulations – section 4.5.2

The rest of this section outlines the different models and approaches required to undertake the different steps outlined above. More details on base theory are present in Del1.1 and 1.6.

4.2 Recap of univariate extreme value analysis

4.2.1 Motivation and methodology

To understand the impact that a weather hazard may have on a nuclear power plant, it is necessary to estimate the frequency and intensity of the hazard of interest. In most situations we are not interested in the 'normal' levels for a hazard (e.g. the daily mean air temperature), but instead the largest events within the data series. This presents a couple of problems when striving to model this type of phenomena using statistical approaches:

1. Rare events (by definition) don't occur very often and as such, if we focus on these values then we won't have much data available;
2. Most common statistical models (e.g. Normal distribution) provide a good fit to 'normal' values but are poor when it comes to extrapolating beyond the range of the dataset.

In our situation it is very important to be able to extrapolate beyond the range of our data as we are seeking to design and protect nuclear infrastructure. Industry regulation across different countries often requires a level of protection between the 100-year and 10,000-year return level (depending on the hazard of interest), with levels up to the 10,000,000-year return level sometimes required for beyond design basis studies. Extrapolation is also very important given the occurrence of climate change and the likelihood that future hazards (such as extreme air temperature) are likely to be more intense that those observed before (this type of extrapolation will be possible using covariates to introduce non-stationarity into statistical models).

For this reason, we need to use a set of statistical models that are specially built for the purpose of extrapolating and that can be used to draw inferences from a small sample of data. Extreme value analysis (EVA) is a commonly used statistical technique for the estimation of extremes that may be possible in the future but may not necessarily have been previously observed. The general principle of EVA is to fit a statistical model to extremes within observed data and use this model to extrapolate to higher levels to estimate very rare extreme levels.

There are two EVA approaches for determining extremes within the observational data record: (i) block maxima; (ii) threshold exceedances. The block maxima approach divides the data into sections or blocks, which commonly represent time (e.g. years) and then the maximum

value in each block is retained. Threshold exceedance methods set a high threshold, often taken as a quantile of the observational data (for example the 95th quantile) and exceedances above the threshold are retained. Threshold exceedance approaches are often regarded the superior methods as they allow more data points to be retained, leading to more robust statistical inferences. However, the choice of threshold is subjective and should be made carefully to ensure that there are sufficient values falling above the threshold to permit a robust analysis, but also high enough so that focus is still given to the extremes.

Once the extremes have been determined a distribution is fitted. The generalised extreme value (GEV) distribution is fitted to extremes defined by the block maxima approach, while a generalised Pareto distribution (GPD) is fitted to extremes defined by the threshold exceedance method. The distribution is then used to extrapolate from the observed data to extreme levels and estimate return levels of rare events.

For the rest of this section and subsequent analyses we focus on threshold exceedance methods. EVA is based upon asymptotic theorems which only hold true for "sufficiently" extreme values. The choice of what constitutes a "sufficiently" extreme value is often difficult and subjective. Many different approaches exist for choosing a suitable threshold for an extreme value study. Within statistical literature, mean residual life plots and parameter stability plots are used for threshold choice. Broadly, these plots aim to suggest levels above which the fitted extreme value model is stable. If this is the case, extrapolations obtained using the fitted model will be valid. Other approaches are based upon physical constraints, for example setting a threshold such that there is on average one extreme event each year.

Modern studies into the estimation of extreme environmental variables all use threshold-based approaches. Let X be a random variable with distribution function F. The distribution of the exceedances of a high threshold u, i.e. $X|X > u$, follow a GPD(σ_u, γ) which has the form:

$$G(x) = P(X \leq x|X > u) = 1 - \left(1 + \gamma \frac{x-u}{\sigma_u}\right)_+^{-\frac{1}{\gamma}}, \gamma \neq 0$$

for $x > u$, where $\sigma_u > 0$ is the scale parameter and $-\infty < \gamma < \infty$ is the shape parameter . The GPD is fitted above the threshold u and an empirical distribution function is fitted to all value that fall below u. As such the probability distribution is given as:

$$F(x) = \begin{cases} 1 - \delta_u[1 - G(x)], & x \geq u, \\ \tilde{F}(x), & x < u, \end{cases}$$

where $\delta_u = P(X > u)$ is the rate parameter and $\tilde{F}(x)$ is the empirical distribution function. The T-year return level x_T can be obtained directly as:

$$x_T = \begin{cases} u + \frac{\sigma_u}{\gamma}[(Tn_T\delta_u)^\gamma - 1] & \text{if } \gamma \neq 0, \\ u + \sigma_u \log(m\delta_u) & \text{if } \gamma = 0, \end{cases}$$

where n_T is the number of observations within a year. EVA is usually conducted with packages built in the R statistical programming language; popular packages are *ismev*, *extRemes* and *texmex* packages.

4.2.2 Estimating confidence intervals

It is important to provide estimates of the uncertainty associated with estimates of parameters and return levels. This is especially important when estimating very extreme return levels as these extrapolations can be very uncertain. There are three main approaches to deriving confidence intervals:

1. Delta method;
2. Bootstrapping;
3. Profile likelihood.

The delta method is commonly used as the default approach to confidence interval estimation within most extreme value packages in R. Broadly, the sampling variability (standard error) in the quantity of interest (i.e. parameter or return level) is estimated and this is used to create a symmetric interval around the best estimate. So, for a parameter φ, the (1−ω)×100% confidence interval can be estimated as:

$$\hat{\varphi} \pm z_{1-\omega/2} se(\hat{\varphi})$$

where $\hat{\varphi}$ is the best estimate of the parameter φ and $z_{1-\omega/2}$ is the $1-\omega/2$ quantile of the standard normal distribution. This approach for estimating confidence intervals is relatively easy computationally but does always lead to symmetric confidence intervals. There is no restriction to the value that these confidence intervals can take and as such, given the correct conditions, it is possible to estimate confidence intervals which contain negative extreme hot temperature or negative precipitation for example.

Bootstrapping is an alternate approach for estimating confidence intervals and it is the method applied in this study. This approach models the underlying uncertainty in the observed data by creating multiple bootstrap data sets (often at least 1000 data sets need to be created) by resampling the original data set. Extreme value models are then fitted to each separate bootstrapped data set and the relevant parameter or return level is estimated for each data set with replacement. The estimates obtained from each bootstrapped data set are then collected and percentiles of these values are taken; for example, to get a 95% confidence interval the 2.5 and 97.5 percentiles are taken. Since each model is fitted to a resampled version of the original data, this approach avoids unrealistic estimates falling within the confidence intervals. It is not necessary for the intervals to be symmetric around the best estimate.

The profile likelihood approach is not commonly used within applied extreme value analysis studies and as such is not discussed in this report.

Bootstrapping was performed with 10,000 iterations to minimise the variation between bootstraps with the aim of making the results as replicable as possible. Attention is drawn to the fact that although the differences between confidence intervals obtained from different runs have been minimised, some variance may remain in replications, mostly in the high return level estimations (10^{-5}, 10^{-6} and 10^{-7}) for the high confidence intervals (e.g. 99%).

4.3 Drawbacks of univariate EVA and needs for extension

The main aim of univariate EVA models is to estimate extreme return periods (e.g. the 100-year return level, the event that occurs on average once every 100 years). As discussed in section 4.2.1 this can be done using models for block maxima (GEV) or exceedances of a high threshold (GPD). A key assumption when using the GPD is that all exceedances of a high threshold are independent and identically distributed (IID). However, when modelling most types of environmental data, we observe that consecutive exceedances are unlikely to be independent (e.g. periods of several consecutive hot days are often observed in summer). If we do not account for this clustering within our return level estimates then we are likely to estimate confidence bounds around our best estimate that are too narrow; as such, an adjustment needs to be made.

The standard approach for ensuring extremes are independent is to use declustering to extract the independent peaks over the threshold (POTs). There is often a level of subjectivity when using declustering since all approaches require some internal parameter choices to be made. However, even in the situation where we make a good selection of POTs, the process of declustering reduces the amount of data available for fitting the GPD model. As a result, uncertainty bounds can be quite wide when using this approach (especially when estimating very rare return levels).

A better approach is to directly account for the extremal dependence within a time-series

rather than having to do any declustering. To do this it is necessary to build a model to explicitly account for the extremal dependence between several variables. This uses a class of models called multivariate extreme value models (which will be outlined in more detail in section 4.4).

These multivariate models are used to model the temporal dependence within time-series and the end results can be used in a couple of different ways:

- The multivariate models can be used to directly assess the duration of extreme events above some high threshold and therefore provide an assessment of the occurrence frequency and severity of persistent extreme events (e.g. heatwaves and wind storms). See case study in section 7.1.
- Estimates of the uncertainty associated with return levels can be reduced by fitting a GPD to all exceedances and then using the temporal dependence estimate to readjust the confidence intervals for the fact that all the exceedances are not actually independent. See case study in section 7.2.

Both approaches above require the use of multivariate extreme value models which are outlined below.

4.4 Multivariate extreme value theory

To develop an understanding of the risks posed by multi-hazards, it is necessary to build statistical models which can account for the behavior of more than one hazard. It is clearly an over-simplification to assume that all different natural hazards can be treated as independent of one another and thus modelled separately using univariate statistical models. As an example, extra-tropical cyclones that track across the Atlantic Ocean and strike Europe are likely to bring extreme rainfall, wind speeds, storm surges, wave heights amongst other hazards.

Different modelling approaches are available for modelling extremes of more than one hazard (i.e. multi-hazards) or joint extremes at several locations / consecutive times. An interesting summary of the different approaches for multi-hazards is given in Tilloy et al. (2019); the paper outlines many different classes of model (defined as stochastic, empirical or mechanistic) that are often used to model combinations of hazards. For the rest of this section we focus on the class of stochastic models, which will be used to model the extremes of two variables (bivariate case chosen for notational simplicity).

4.4.1 Dependence and extremal dependence

When modelling multiple variables, it is important to quantify the sign and strength of the relationship between the different variables of interest. If two (or more) hazards are positively related to one another, then this increases the chance of both hazards occurring at once; a negative relationship means that the occurrence of one hazard will decrease the likelihood of another hazard occurring.

A standard measure of dependence is the correlation coefficient. Although there are several different definitions, the most commonly used is Pearson's correlation coefficient. Given a pair of variables (X, Y), Pearson's correlation coefficient is defined as:

$$\rho(\mathrm{X},\mathrm{Y}) = \frac{\mathrm{cov}(\mathrm{X},\mathrm{Y})}{\sigma_{\mathrm{X}}\sigma_{\mathrm{Y}}}$$

where $\rho \in (-1,1)$, σ_{X} and σ_{Y} are the standard deviations of X and Y respectively and:

$$\mathrm{cov}(\mathrm{X},\mathrm{Y}) = \mathrm{E}[(\mathrm{X}-\mu_{\mathrm{X}})(\mathrm{Y}-\mu_{\mathrm{Y}})]$$

where μ_{X} and μ_{Y} are the means of μ_{X} and μ_{X} respectively. The value of the correlation coefficient represents the level of dependence in the data being modelled. When $\rho = 1$, the two variables are said to be perfectly positively correlated; $\rho = -1$ corresponds to perfect negative dependence and $\rho = 0$ to independence.

The correlation coefficient summarises the level of dependence in the body of the data (i.e. values that are close to the mean) well, but **this is not the quantity that we are necessarily interested in when looking at the extremes of multiple variables**. It is often more important to know how the extremes of different variables are related, not just the relationship of 'average' values. To investigate this further it is necessary to define the concept of *extremal* dependence. The Spearman coefficient or Kendall rate give details of the dependence only being determined from only the copula.

A simple definition of *extremal* dependence is provided by the following measure for a pair of variables (X,Y):

$$\chi(v)=P(Y>v|X>v)$$

where $\chi(v) \in (0,1)$ and v is some threshold set at a sufficiently high level of interest. The value of the extremal depdence measure varies depending on the strength of dependence in the two variables above v. A value of $\chi(v) = 1$ implies total dependence whereas $\chi(v) = 0$ is the case of independence.

The measure explained above should actually be referred to as the sub-asymptotic extremal dependence measure; there is an equivalent asymptotic version which is defined in the limit as $v \to v^*$ (where v^* is the upper end point of the common marginal distribution), i.e.:

$$\chi = \lim_{v \to v^*} \chi(v).$$

Here, when $\chi = 0$ this implies asymptotic independence (AI) whereas $\chi > 0$ implies asymptotic dependence (AD) (Coles et al., 1999). For asymptotically dependent data the measure χ provides a measure of the strength of the dependence. However, for asymptotically independent data, the measure χ does not provide any information about the strength of the sub-asymptotic dependence.

An additional dependence measure $\bar{\chi} \in (-1,1)$ was defined in Coles et al. (1999) with different values of $\bar{\chi}$ defining the level of asymptotic independence between the pair of variables (X, Y); a value of $\bar{\chi} \in (0,1)$ implies asymptotic independence with positive association, $\bar{\chi} \in (-1,0)$ implies asymptotic independence with negative association, $\bar{\chi} = 0$ implies independence and $\bar{\chi} = 1$ implies asymptotic dependence.

From a practical perspective, the importance of the AD/AI dichotomy comes when specifying a statistical model for investigating the occurrence and severity of joint extremes. Different statistical models will have varying asymptotic behavior and as such will result in different extrapolations. If we use a model which can only account for AI and our data is AD, then any extrapolations that are made will underestimate the probability of observing a joint extreme and leave us underprotected; the opposite is true if we pick an AD model when the data are AI.

The main difficulty arises from the fact that it is hard to estimate whether data are AD or AI from an observed data sample. In practice, we do not have an infinite sample of data and as such cannot directly observe the asymptotic behavior of the variables we are modelling. In most situations, the sub-asymptotic dependence measure $\chi(v)$ can be estimated with confidence at high quantiles. However, as v is increased less data is available in the observed sample and it becomes hard to get a reliable estimate for $\chi(v)$ (in a similar way as if the threshold is set too high with univariate extreme value models). Given the context of nuclear infrastructure safety, it is vital to provide estimates of joint risk beyond the range of the observational record; as such we cannot rely on empirical estimates of $\chi(v)$ and need to specifiy a model to estimate extremal dependence for more extreme values than previously observed. Of course there are large uncertainties as part of this process.

The asymptotic dependence structure of the model we choose to fit thus becomes very important and shall be discussed in more detail for the models outlined below. Some approaches (e.g. most copulas) require the user to specify the asymptotic dependence

structure in advance of modelling the data which can lead to issue with mis-specification. Other approaches are more flexible and provide parameterisations that permit both AD and AI. As a result, the model can be fitted and can be used to diagnose the asymptotic dependence structure of the variables under consideration.

4.4.2 Copulas

Copulas represent a commonly used class of multivariate models for modelling multiple variables. Copulas are essentially defined as statistical models for the joint dependence between variables once the marginal dependence has been accounted for, building upon Sklar's theorem. Many different copulas exist for modelling environmental extremes and there are many books that provide a comprehensive summary of the different approaches such as Joe (1997) and Cressie (1993). A recent summary of the use of different copulas for multi-hazard events is provided in Tilloy et al. (2020).

There are some drawbacks with using copulas to model multivariate extremes. Copulas tend to require much data at more than three dimensions to scale appropriately and as such if it is necessary to model extremes of multiple variables it can be hard to fit these models. Also, most copulas are either asymptotically dependent or asymptotically independent and as such it is necessary to pick the type of copula a priori to match the extremal behavior of the observation data. Although it is possible to estimate whether a bivariate dataset is AD or AI using empirical approaches, sometimes these approaches do not provide clear guidance. Extreme value copulas are required. As such, having to choose a type of asymptotic dependence structure before fitting a copula increases the chances of mis-specifying the dependence structure; this is something avoided by the models outlined in sections 4.4.3 and 4.4.4.

4.4.3 Joint tail model

A different approach to modelling the extremal dependence structure is outlined in Ledford and Tawn (1997). Taking a pair of variables (X,Y) with common marginal distribution, the joint tail model of Ledford and Tawn (1997) models the asymptotic form of the joint survivor function directly. As a result, we explicitly model only the points for both variables that fall above a high level $\mathrm{v} = \mathrm{v}_\mathrm{p}$, often defined as the (1-p)th quantile i.e.:

$$\mathrm{P}(\mathrm{X}>\mathrm{v}_\mathrm{p}) = \mathrm{p}.$$

The form of the joint tail is given in Leford and Tawn (1996) on Fréchet margins, but more generally is given as:

$$P\left(X > \mathrm{v}_\mathrm{p}, Y > \mathrm{v}_\mathrm{p}\right) \sim L(1/p)p^{1/\tau}$$

as $\mathrm{p} \to 0$, where L is a slowly varying function and τ is named the coefficient of tail dependence; defined over the range $\tau \in (0,1]$. The slowly varying function L satisfies:

$$\frac{\mathrm{L}(\mathrm{t}/\mathrm{p})}{\mathrm{L}(1/\mathrm{p})} \to 1$$

as $\mathrm{p} \to 0$ for all fixed $\mathrm{t} > 0$. The value of the coefficient of tail dependence τ gives the level of extremal dependence between the marginal variables where $0 < \tau < 1/2$ implies negative association, $\tau = 1/2$ implies independence and $1/2 < \tau < 1$ implies positive association. If $\tau = 1$ and $\mathrm{L}(1/\mathrm{p})$ does not tend to 0 as $\mathrm{p} \to 0$ the pair of variables (X,Y) are asymptotically dependent. Any values of $\tau < 1$ imply asymptotic independence. A common modelling assumption is $\mathrm{L}(1/\mathrm{p}) \to c$ as $\mathrm{p} \to 0$, where $\mathrm{c} \in (0,1]$. In the limit this assumption could introduce a small amount of bias, but at the subasymptotic levels we are interested in the slowly varying function tends to constant. As such, at the levels we are interested in the likelihood is not mis-specified, even if this may be the case in the limit. Under the above assumption, the problem of calculating extremal dependence comes down to the estimation of parameters (c, τ) which fully explain the dependence in the joint tail.

The joint tail model is commonly fitted using maximum likelihood estimation. In the standard fitting approach, only data points which are extreme in both variables are used within the model fitting (i.e. data values which are extreme in one variable but not the other are not used). This can lead to issues with fitting when there are few joint extremes in the observed sample. The approach can be extended to multivariate problems, but again scaling can be difficult due to the reliance on only using data from the joint exceedance region.

4.4.4 Conditional extremes model

The conditional extremes approach Heffernan and Tawn (1999) models multivariate extreme data in a different way to assess the strength of extremal dependence. This approach avoids the limiting arguments of the joint tail methods in which all variables must become large at the same rate. Standard copula-based methods can typically only handle one form of extremal dependence, either asymptotic dependence or asymptotic independence. As such the form of the dependence structure must be chosen in advance before the model is fitted. The conditional extremes approach estimates the form of the extremal dependence structure as part of the fitting procedure so removes the need to choose the form of the dependence structure in advance. The conditional extremes approach can also be used to model high-dimensional data with greater ease than for copula-based methods, although all theory in this section is given for the bivariate case.

To estimate the dependence structure of (X,Y) using the conditional extremes approach both variables must be transformed onto common marginal distributions; the best choice is to transform the data onto Laplace margins (Keef et al., 2013) to give (X_L,Y_L), which them permits the following form for the limiting conditional distribution:

$$P\left(\frac{Y_L - \alpha X_L}{{X_L}^{\beta}} \leq z, X_L - u > x \mid X_L > u\right) \rightarrow \exp(-x)G(z)$$

as $u \rightarrow \infty$, where $G(z)$ is a non-degenerate distribution function, $\alpha \in [-1,1]$ and $\beta \in (-\infty,1)$. Different values of α and β characterise different forms of tail dependence. In the case where $\alpha = 1$ and $\beta = 0$, variables (X_L,Y_L) exhibit asymptotic positive dependence. Due to the exponential lower tail specified by the Laplace margins, the case of asymptotic negative dependence is given when $\beta = 0$, $\alpha = -1$ and $\alpha = -1$. If $\alpha = \beta = 0$ and $G(z)$ is the Laplace distribution function the variables are independent.

Again, the standard way of fitting the conditional extremes model is via maximum likelihood estimation. Further discussion about the semi-parametric approach for estimating the model parameters is provided in Keef et al. (2013). It is also possible to easily simulate directly from the fitted model to understand the dependence between variables at extreme levels, especially at levels beyond previously observed levels.

The conditional extremes approach has been widely used for several different studies on joint extremes, including extreme river flows at multiple locations (Lamb et al., 2010) and persistence of extreme air temperatures (Winter and Tawn, 2016). The conditional extremes approach has also been used to model combinations of different hazards, for example: significant wave height and wave period (Jonathan et al., 2013); wave height and sea level (Gouldby et al., 2017). Recent studies have also started to incorporate covariates into the conditional extremes approach; some examples include Jonathan et al. (2013) and (Winter et al., 2016).

4.5 Applying multivariate extreme value models for temporal dependence

The multivariate models outlined above can now be used to model the temporal dependence within the extremes of time series. In this situation (under the assumption of a first-order Markov process to make the notation simpler) we use the multivariate models to investigate the joint distribution of (X_t, X_{t+1}), where t and $t+1$ are consecutive time steps (e.g.

consecutive hours for hourly data or consecutive days for daily data). The fit of the multivariate model can then be used to construct replicate events with the same statistical properties in the extremes. By simulating a large amount of replicate events, it is possible to estimate the extremal index or other aspects of the duration of events. For the rest of this section we choose to outline how to use the conditional extremes model within this framework (as in Winter and Tawn, 2016); other models (such as copulas or the joint tail model) could be used if desired.

4.5.1 Choosing the order of Markov model

An important first step is to pick the order of the Markov model which we want to fit. Broadly, the order of the Markov model controls how much memory the simulation process will have. Higher order Markov models will incorporate more information from previous time-steps. As an example, the assumption of a first-order Markov model will lead us to focus on the pair (X_t, X_{t+1}) whereas a second-order Markov model will model (X_t, X_{t+1}, X_{t+2}) and so on.

There is a trade-off when choosing the order of the Markov model that we wish to fit. In theory, fitting a higher-order Markov model is likely to lead to a better fit as we are incorporating more information from more consecutive time-steps. However, there is likely to be a limit to the benefit of increasing the order of the Markov model above a certain level. We also need to appreciate that every time we increase the order of the Markov model, we shall need to estimate more parameters within the multivariate extreme value model using the same amount of data which will lead to a natural increase in the uncertainty when estimating model parameters. There is also a distinct possibility of overfitting the model to a specific dataset if we choose too high an order.

The standard approach for estimating the order of the Markov process is to use autocorrelation and partial autocorrelation functions (ACF and PACF respectively). The autocorrelation function is defined as:

$$R(\tau) = E(X_t \overline{X_{t+\tau}}),$$

where $R(\tau) \in [0,1]$, $E(.)$ is the expectation function and τ is a time lag. When $R(\tau) = 0$ then the process X_t is independent at lag τ; a value of $R(\tau) = 1$ implies perfect dependence in the process X_t at lag τ. The partial autocorrelation $R_p(\tau)$ measures the autocorrelation between two time-steps that are lag τ apart once the autocorrelation between all intermediate time-steps $(X_{t+1}, \dots, X_{t+\tau-1})$ has been accounted for.

The order of the Markov process is often chosen using the PACF plotted using standard functions. When the value of the PACF is statistically close to zero then this lag is chosen as the order of the Markov chain as any lags beyond this value are independent and thus increasing the order of the Markov process doesn't add more information. It should be noted that the ACF and PACF are applied over the whole data series and thus incorporates information about the non-extreme values as well as the extremes; this is not found to be a large problem in Winter and Tawn (2016).

For the rest of this section we assume a first-order Markov process for notational simplicity. For more information on fitting a higher-order Markov model then see Winter and Tawn (2017).

4.5.2 Simulating events and drawing inferences

Having chosen the order of the Markov process it is now necessary to fit the conditional extremes model to $X_{t+1}|X_t > u$. This can be done using the approach in Winter and Tawn (2016) and this provides parameter estimates $(\hat{\alpha}, \hat{\beta})$ for the conditional extremes model parameters as well as $\hat{G}$ which is a non-parametric estimate of the limiting distribution G from section 4.4.4. These outputs from the fitted conditional extremes model can be used to simulate tail chains using the following algorithm (below $T(.)$ is defined as the Laplace distribution function):

1. Set critical level of interest $v > u$ and simulate an initial exceedance using a GPD(σ_v, γ)

distribution

2. Set this exceedance as X_1^*

For i in 1: $k-1$:

3. Make draw Z_i^* with replacement from $\hat{G}$
4. Set $X_{i+1}^* = T^{-1}\{\hat{\alpha}T(X_i^*) + T(X_i^*)^{\hat{\beta}} Z_i^*\}$

End

The output of this algorithm is a tail chain $(X_1^*, \ldots, X_k^*)$ with a dependence structure given by $(\hat{\alpha}, \hat{\beta})$. It should also be noted that another simulation approach is available (peak value tail chain estimation); this approach is detailed in Winter and Tawn (2016) but is not used further here for simplicity.

Irrespective of approach, it is possible to use the algorithm multiple times (e.g. 10,000) to generate a large set of simulated tail chains which can then be used to estimate quantities of interest. One such quantity is the duration distribution of events, often denoted as $\pi(i, u)$. This quantity broadly relates to the probability of having exactly i exceedances above a threshold u within a cluster. The duration distribution can be estimated from the tail chains (obtained using the simulation approach above) using theory detailed in Winter and Tawn (2016). Another important quantity is the extremal index θ which can be estimated from the duration distribution using the relationship:

$$\theta^{-1} = \sum_{i}^{\infty} i\,\pi(i)$$

Since this is a simulation approach, there is the possibility of investigating other important metrics. Some other potential metrics are outlined in Winter and Tawn (2016), alongside theory that can be used to convert within cluster results to across cluster results.

4.6 Using temporal dependence to improve return level estimates

Aside from the clear benefits when using the multivariate extreme models to estimate the duration of extreme events, the output can also be used to reduce the uncertainty associated with univariate return level estimation as well. An approach for incorporating temporal dependence into return level estimation is outlined in Fawcett and Walshaw (2012); a brief summary is provided below.

When using a threshold exceedance approach to model extreme events, the process of declustering is often undertaken to pick out the independent peaks over threshold (POTs) to ensure that the data being fed into the model are independent. However, as discussed in section 4.3, the declustering approach discards a lot of useful data and requires some specific choices of parameters (e.g. the run length) which can lead to increased uncertainty in the end results.

An alternate approach is to fit a GPD model to all the exceedances and then make an adjustment to the return level estimates using the extremal index θ as defined in section 4.5.2. This adjustment considers the fact that the extremes fed into the GPD fit are not independent.

The formula for estimating the return level is modified from the version in section 4.2.1 to be:

$$x_T = u + \frac{\sigma_u}{\gamma}\left[\left(\delta_u{}^{-1}\left\{1 - [1 - 1/(Tn_T)]^{\theta}\right\}\right)^{-\gamma} - 1\right],$$

where all the terms have equivalent definitions as in section 4.2.1. The extremal index can be estimated in several different ways, Fawcett and Walshaw (2012) highlight a variety of parametric and non-parametric approaches for doing this. Here, we use the multivariate extreme value model constructed in sections 4.4 and 4.5 to estimate the extremal index and

thus to see how it can be used to improve estimates of return levels.

It should also be noted that the implementation of the model in Fawcett and Walshaw (2012) is undertaken using Bayesian statistical modelling as opposed to the frequentist approach outlined in section 4.2.1. Both approaches for inference will lead to similar results. The main difference revolves around the fact that in frequentist statistics the data are assumed to be drawn from a random process and the model parameters are fixed whereas the Bayesian paradigm assumes the data are fixed and the model parameters can vary. For our application to return level estimation, the use of Bayesian statistics permits the estimation of the T-year predictive return level, which is broadly defined as an estimate of the return level with all uncertainty in parameters and randomness in future observations rolled in. For an explanation of this quantity see Sharkey and Winter (2019).

4.7 Climate Change studies

Climate change impacts have been discussed partially within Del1.1, however a greater insight will now be given as to the non-stationary nature of climatic parameters as well as data and trends of climate change in the EU, as well as the application for evaluating the different interactions of climate change.

There are a number of key parameters that have been derived at various scales to differing levels of quality, which are always dependent on model inputs. Given the particular baselines and variability of the parameters, it is not always clear which climate scenario should be used.

Climate change has generally been assessed at various scales across the world, however the key climate models are housed within the CMIP (Climate Model Incomparison Project) which was established in 1995. The program has had 6 phases/iterations so far and has various research institutions globally running climate models with standardised outputs and input parameters. This has allowed for checks of various forcing models in terms of variability of the range of responses and how well they recreate recent climate data.

Scenario experiments RCP2.6, RCP4.5, RCP6.0, RCP8.5 were the pathways chosen in CMIP5: Future projections (2006-2100) forced by RCP2.6, 4.5, 6.0, and 8.5. RCPs (representative concentration pathways) approximately result in radiative forcings of 2.6, 4.5, 6.0 and 8.5 W m-2 at the year 2100 respectively, relative to pre-industrial conditions.

The CMIP5 ensemble calculations (Taylor et al., 2012) have 3 types of sensitivity studies:

- realisation (r), which change only the initial conditions of the model through random numbers thus indicating the internal variability;
- initialisation (i), which is the variation in initial conditions and sensitivity;
- physics (ii), which are the differences of the sub-grid scale processes.

Only through these 3 components can an appropriate choice of climate parameter be made for use within a PSA.

Regional CMIP5 downscaling has been done by Rohini et al. (2019). CORDEX was the Coordinated Regional Climate Downscaling Experiment, which provides climate change parameters at finer detail than what the global climate models (GCMs) can do. They are more representative of small-scale regional climate characteristics and can represent extreme events better, thus are more applicable for NPP applications.

In CORDEX, RCPs are run along the ensemble of CORDEX climate projection experiments, which uses the ERA-Interim analysis by ECMWF for the past, as well as the historical climate observations.

For the global models (GCMs) within the CMIP6 scenarios (Eyring et al., 2016), the SSPs have been tied to a particular radiative forcing, similarly RCMs (regional climate models) can be applied.

In Europe, the E-OBS gridded dataset of the European Climate Assessment and Dataset (Cornes et al., 2018) interpolate station-derived meteorological observations. As of June 30, 2020 this is at v21.0. ETCCDI parameters used observed data from stations in order to derive various parameters which relate to the aforementioned hazards to produce long-term trends in order to get a coreset of 27 precipitation and temperature parameters describing frequency, amplitude and persistence. In total 76 indices of cold, heat, multi-element, drought and rain type indices for historical climate data from 1950-2019 are presented. HADEX2 was one range of this (Donat et al. 2013a) which however was derived at a very coarse resolution.

PRIMAVERA-H2020 project is the latest EU project to provide high resolution analyses on a smaller scale appropriate to NPP site levels. This downscaling and higher resolution allows for a better application of possible effects of climate change. For the various hazards, climate change impacts will not be explored, not in terms of variation at a site (given that the CMIP6 and PRIMAVERA models are still being produced), but in terms of the potential adjustment per hazard type.

For one example, the following CMIP6 scenarios are presented for Trino Vercellese in terms of baseline temperature and precipitation indicators.

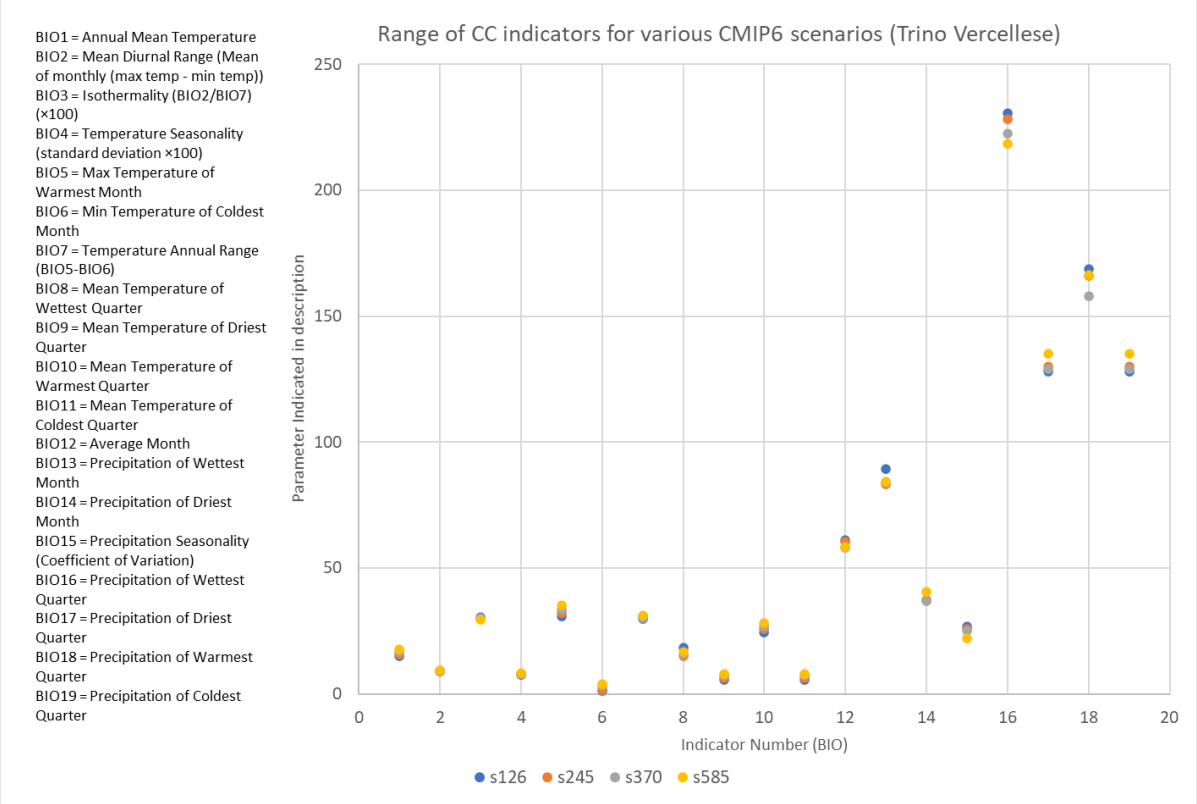


Figure 6: Difference in absolute terms for different weather parameters (temperature in degrees Celsius, precipitation in mm)

In terms of differences between different parameters vs. the lowest scenario of SSP1-2.6, there is significant differences between the different indicators with significant spread seen depending on the scenario which can make a large difference in final determination of combined hazard analyses if not taken into account. The stationary vs. non-stationary nature of climate scenarios is discussed in section 6.3/6.4.

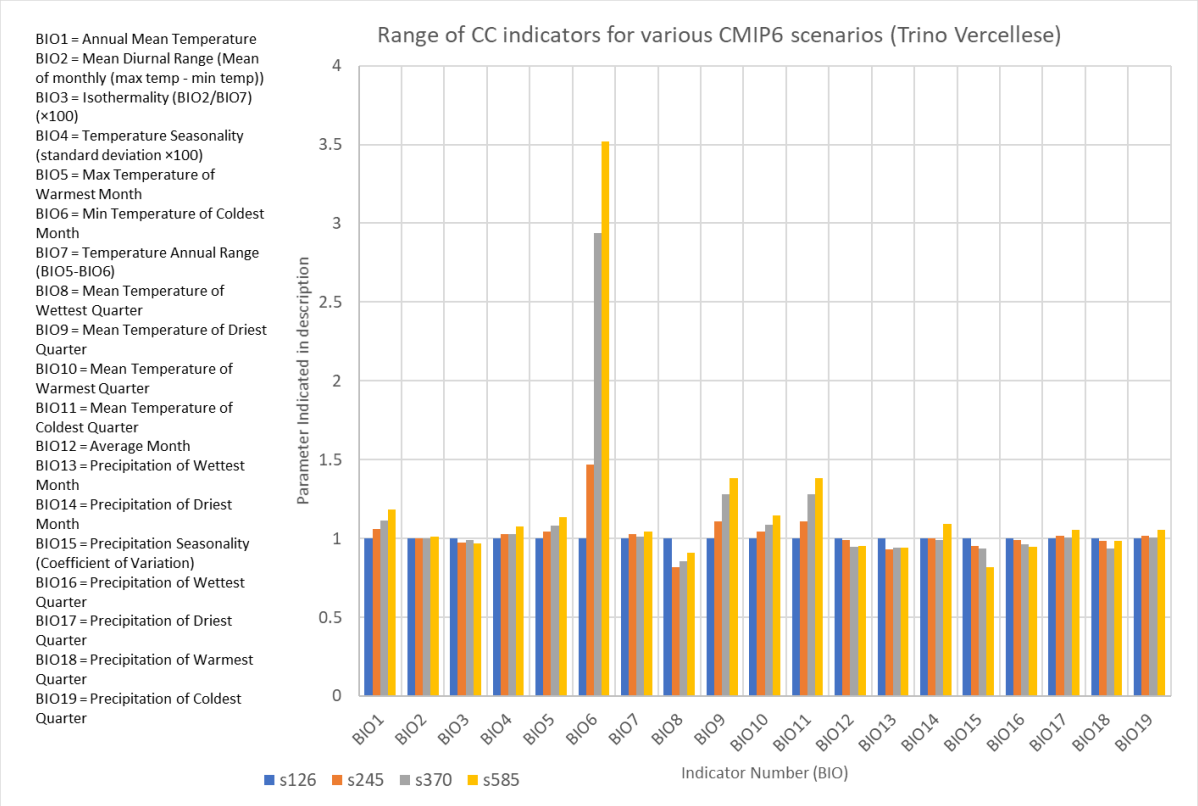


Figure 7: Difference in relative terms for different weather parameters as a ratio vs. the S126 scenario using WorldClim data from CMIP6 (downscaled for Trino Vercellese)

ERSSTv5 has been used for determination of monthly sea surface temperatures. The observed daily temperature and rainfall dataset is publicly available at https://eca.knmi.nl/download/ensembles/download.php#months.

ERA-5 reanalysis and ERSSTv5 datasets are publicly available at:

https://cds.climate.copernicus.eu/cdsapp#!/dataset/, and

ftp://ftp.ncdc.noaa.gov/pub/data/cmb/ersst/v5/.

The simulations from CMIP6 and CMIP5 are publicly available at:

https://esgf-node.llnl.gov/projects/cmip6, and

https://esgf-node.llnl.gov/search/cmip5/, respectively.

An adaption of the usual flow chart for extreme weather hazards in PSA shows the time step at which the key climate analysis should be included.

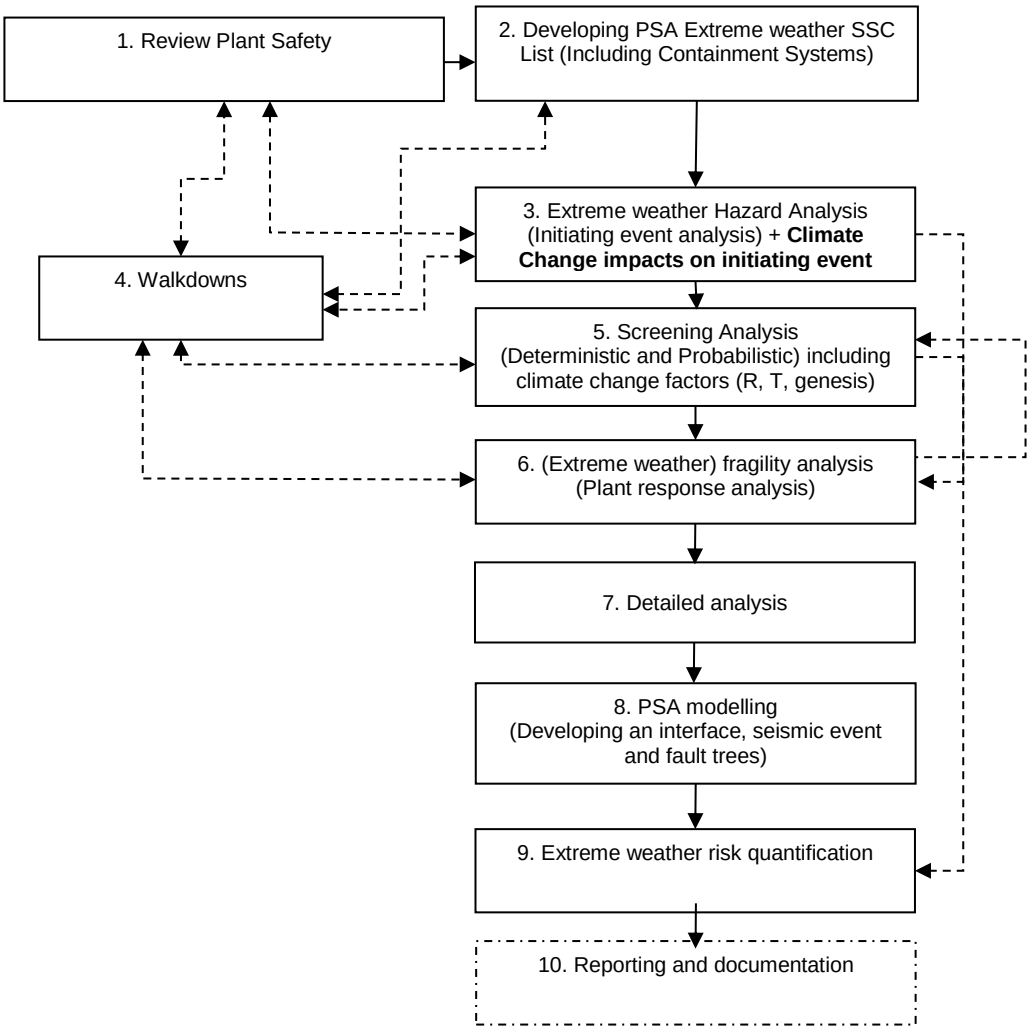


Figure 8: Flow chart for extended extreme weather hazards in PSA

In terms of standard approaches for the application of climate change results to NPP design, guidelines have been put forward in various forms. As the PRIMAVERA models and other downscaled models become available at a reasonable resolution, they should be integrated in conjunction with the usual station data which is usually collected at the site, and from long-run stations around the site. Examples of the EVS applications are shown in section 7: Application of approaches for assessing duration of extreme events.

5 Methodologies on uncertainty quantification and global sensitivity analysis, with dependent inputs (IRSN)

The assessment of external hazard, as flooding or earthquake, often relies on numerical models which allow assessing the variables of interest (*e.g.* water depths, ground acceleration, *etc.*). However, these models are affected by uncertainties which can be quantified through studies called “Uncertainty Quantification” (UQ) and “Global Sensitivity Analysis” (GSA). The UQ attempts to describe the whole set of possible outputs considering the inputs not perfectly known. The GSA aims to determine the most influent inputs to an output behaviour, as the non-influential ones. These two types of analyses are complementary and both classically suppose that the input parameters of the numerical models are independent. Especially, in the literature, most studies consider the inputs independent which may not be the case. In this context, the objective of the research activity conducted by IRSN is to perform a global review of the different methods used for UQ and GSA when model inputs are considered dependent and to apply them to hydraulic studies. In the field of flooding hazard, this kind of analysis is still missing. For the interested reader, an IRSN report on existing methodologies for UQ and GSA with dependent inputs followed by an application to a simplified case of inundation has been written on this topic (Pheulpin, 2020). The major points from this report are reported in the sections 5.1 and 5.2. In addition a research article on the application of these methods to the large hydraulic model of the Loire River is in progress. Some of the results have ever been presented in the EGU (Pheulpin and Bacchi, 2020) and the major points are reported in section 5.3.

5.1 Existing methodologies

A traditional UQ is conducted by randomly choosing a set of input values inside the associated univariate distribution functions (*e.g.* with a Monte-Carlo sampling method). Then, the inputs are passed through a numerical model to obtain the distributions of the resulting outputs. This type of sampling is available with independent inputs. In case of dependent inputs, copulas can be used to describe the dependencies between some inputs. A copula is a joint distribution defined in a d-dimensional space $[0,1]^d$ with uniform marginal distributions (Nelsen, 2007), d being the number of parameters inside the given copula. To build the copulas, the inputs are classified in groups of dependent inputs and correlation coefficients between pairs of dependent inputs are calculated from observed data. The most suitable copulas are selected by testing different copulas on the observed data, with the help of statistic tests (*e.g.* Cramér-von Mises or Kolmogorov-Smirnov). Once the copulas selected, input parameters, having their own probability distributions, are randomly drawn inside the copulas and then the outputs are calculated through the original model. In the literature, a few studies dealing with dependent inputs and using copulas exists but none is related to large scale hydraulic modelling. Most of these works are related to hydroclimatic studies (Maity and Nagesh Kumar, 2008) or flood frequency analyses (Michailidi and Bacchi, 2017; Sraj et al., 2015).

Some traditional methods of GSA initially designed for models with independent inputs have been adapted to models with dependent inputs. We identified more than 20 methods (reported in Table 2) to deal with dependent parameters. Most of them are variance-based methods allowing to compute sensitivity indices, useful to rank model inputs. For instance, some methods are based on a replicated Latin Hypercube Sampling (McKay, 1996), some others use an ANCOVA decomposition technique. It allows to compute various SI, as the uncorrelated contribution and the correlated one ((Chastaing et al., 2012)). The Shapley effects can also be computed with dependent inputs by using copulas (Iooss and Prieur, 2018). The last method (Table 2) is based on a screening technique. The classical screening Morris method, based on a discretization of the inputs in levels, has been adapted to models with dependent inputs by introducing copulas. This method is generally used for initial screening to eliminate the non-influential inputs before to apply, for instance, a variance-based method which is much more costly in terms of computation time. Finally five of these methods

(in blue in Table 2) have been applied to a very simple case of inundation before to apply them to large hydraulic model.

5.2 Application to a simplified case of inundation

To illustrate the methods previously cited, we used a very simple case of application representing a river section lined by a dyke. The model equation (eq. 1) simulates the height of a river (h in meters) and compares it to the height of a dyke (H_d). There are 8 uncertain inputs (Table 1), including 3 copulas of 2 parameters (Q/K_s, Z_v/Z_m and L/B). For each application (UQ and GSA), two cases have been tested: one with input parameters considered to be independent and one with certain parameters considered to be dependent. All the calculations are made with the R software and the scripts are provided in the appendices of the document (Pheulpin, 2020):

$$S = Z_v + h - H_d - C_b \qquad \text{with} \qquad h = \left(\frac{Q}{BK_S \sqrt{\frac{Z_m - Z_v}{L}}} \right)^{0.6} \tag{1}$$

Table 1: Model inputs description

Inputs	Symbols	Units	PDF	Parameters
Maximal annual flow rate	Q	m^3/s	Truncated Gumbel	**G**(scale=1013, loc=558, min=500, max=3000)
Strickler coefficient	K_s	-	Truncated Normal	**N**(mean=30, sd=8, min=15, max=60)
River downstream level	Z_v	m	Triangle	**T**(min=49, mode=50, max=51)
River upstream level	Z_m	m	Triangle	**T**(min=54, mode=55, max=56)
Dike height	H_d	m	Uniform	**U**(min=7, max=9)
Bank level	C_b	m	Triangle	**T**(min=55, mode=55.5, max=56)
Length of the river stretch	L	m	Triangle	**T**(min=4990, mode=5000, max=5010)
River width	B	m	Triangle	**T**(min=295, mode=300, max=305)

Concerning the UQ, as it can be seen in Figure 9, there is almost no difference between the results coming from the two different methods. The range of outputs is smaller in case of dependent inputs and it can be further reduced with higher correlation coefficients. The eCDF (empirical Cumulative Distribution Function) plot shows a slightly different repartition of the outputs (regarding the tail distribution) and it can be increased with higher correlation coefficients. For this simple case of application, taking into account the dependencies or not gives essentially the same results. We also tried different copulas (*e.g.* Clayton copula or t-copula) and different correlation coefficients; the results change a bit more but non-significantly.

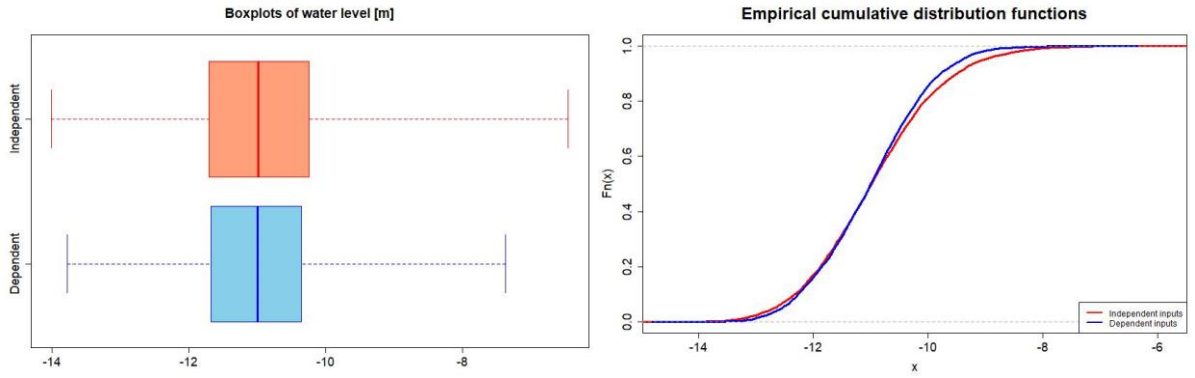


Figure 9: Boxplots and eCDF of the outputs (x is the water level), considering independent inputs or not

Regarding the GSA, the Morris method adapted to dependent inputs has been applied and the standard deviation and elementary mean calculated (Figure 10). *L* and *B* have no influence on the outputs. Z_m seems to have influence in the case where it is linked to Z_v, which is consistent with the studied physical phenomenon (flooding) which is driven by the topographic gradient. The other parameters are influent but they have no interactions between them because the absolute standard deviation (which is an indicator of the interaction between parameters) is well below the absolute mean (which is an indicator of the individual impact of each parameter) in both cases.

Four variance-based methods have been applied to this case study. We compared the first order sensitivity indices coming from these methods (Figure 11). For the four tested methods (Li, McKay, Jacques and Shapley), the ranking parameter is different if the inputs are considered to be independent or not. It proves that it is important to take into account the correlations between inputs. In any cases, the inputs C_b, *L* and *B* never influence the output distribution. In general, in case of independent parameters the sensitivity indices are very similar for all methods and in case of dependent inputs, there are more differences. Except for the Jacques method, the most influent parameter is always *Q* in case of independent parameters and always H_d in case of dependent inputs. Moreover, it seems that the Li method is the more efficient because it is very low time consuming and the model equations are not needed to proceed the GSA (supposing that the inputs and outputs are available).

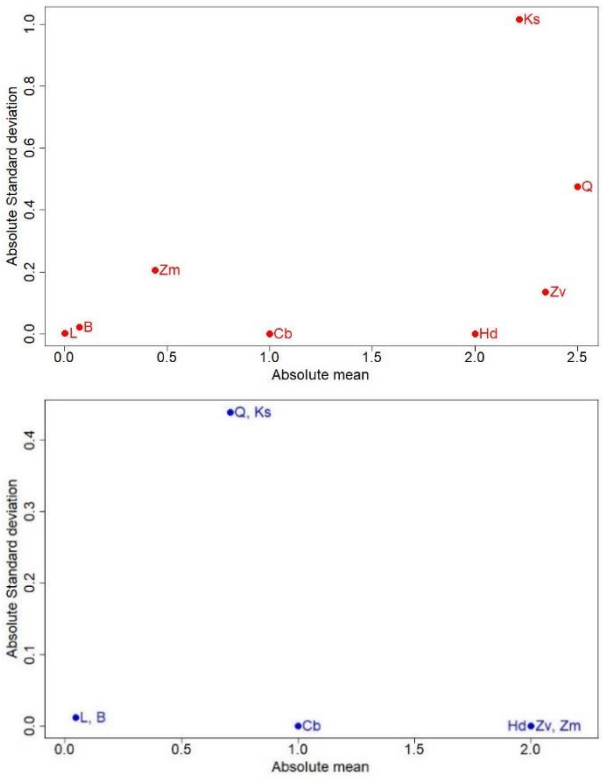


Figure 10: Morris SI considering independent inputs (red indices on the left) and groups of dependent inputs (blue indices on the right).

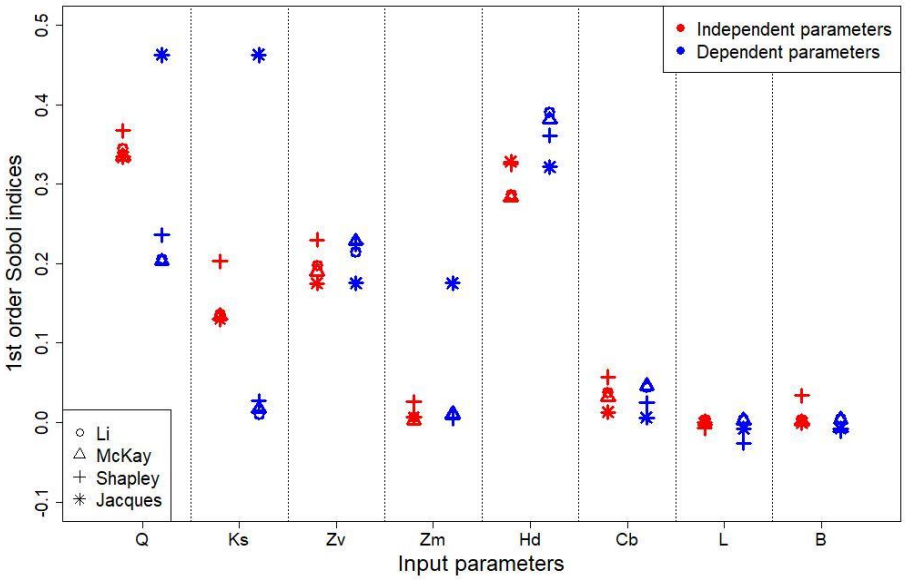


Figure 11: Comparison between the first order SI calculated with four different methods (Li, McKay, Jacques and Shapley), and considering all the inputs to be independent or not (with the Jacques method, the SI are estimated for groups of dependent inputs)

Table 2: Main methods to perform GSA with dependent parameters

	Source	Computation of	Notes
1	Iman and Hora, 1990	Importance measure based on a metric distance between two distributions	
2	Chun et al., 2000		Improvement and extension of method n°**1**
3	Borgonovo, 2007; Borgonovo et al., 2011	δ moment-free importance measure	
4	López-Benito and Bolado-Lavín, 2017		Application of the method n°**3** to a natural gas transmission model
5	Fang et al., 2004	SI estimated with a method based on sequential random sampling	Development of an algorithm based on multi-expressions of multinormal distribution to generate sequential random samples
6	McKay, 1995	Sobol' SI using LHS sampling	
7	Saltelli et al., 2004		
8	Ratto et al., 2005		
9	Oakley and O'Hagan, 2004	SI computed analytically through multidimensional integrals	Bayesian SA aiming to approximate the modelling function by a so-called "kriging" response surface
10	Jacques et al., 2006	Multidimensional SI	Extension of SI definition by taking into account blocks of correlations among the inputs
11	Da Veiga et al., 2009	SI based on local polynomial approximations for conditional moments	Intermediate method between n°**7** and n°**8**
12	Xu and Gertner, 2008	Various SI: Total contribution, uncorrelated contribution and correlated contribution	
13	Li et al., 2010		Extension of method n°**11**: ANCOVA decomposition technique (partial variance decomposed in a variance part and a covariance part) and use of HDMR
14	Pan et al., 2011		Application of the method n°**11** to a unsaturated flow and contaminant transport model
15	Li et al., 2011		Extension of method n°**11**: Decomposition of variance contribution of the correlated inputs based on the state dependent parameters (SDP)
16	Mara and Tarantola, 2012		Alternative method of n°**12**: Decorrelation of the inputs with the Gram-Schmidt procedure and ANOVA-HDMR
17	Kucherenko et al., 2012		Alternative method of n°**12**: Computation of Sobol' SI based on Gaussian-copulas
18	Caniou, 2012		Extension of method n°**12**: Coupling of the ANCOVA technique with polynomial chaos expansion
19	Chastaing et al., 2012		Alternative method of n°**12** defining the SI using a Hoeffding-Sobol decomposition (inclusion of the covariance taking into account the dependency)
20	Zhou et al., 2013		Alternative method of n°**11** using a new algorithm avoiding the sampling procedure
21	Grandjacques et al., 2015	Sobol' SI for static and dynamic inputs	Use of Pick and Freeze sampling method
22	Li and Mahadevan, 2016	First-order Sobol' SI	Direct Sobol' SI estimation based only on input/output samples (model equations are not needed)
23	Owen and Prieur, 2016		

	Iooss and Prieur, 2018	Shapley effects and Sobol' SI	SI computation allocates the mutual contribution (due to correlation and interaction) of a group of inputs to each individual input within the group
24	Wang et al., 2018	Derivative-based SI using copula	Derivative-based sensitivity of variance contribution with respect to the distribution parameters. The dependence between inputs is described through copulas
25	Ţene et al., 2018	Elementary mean and standard deviation	Extension of the Morris' screening method allowing to take into account the dependencies between model parameters through copulas

5.3 Application to the hydraulic model of the Loire River

5.3.1 Introduction

Once the different R scripts of UQ and GSA for dependent parameters tested on the simple model equation (part 5.2), the objective is to apply them to a large hydraulic model. The selected model represents the Loire River between Gien and Jargeau over a distance of 50 km and it includes the NPP of Dampierre-en-Burly (Figure 12). The model was built by IRSN with the open-source software, Telemac-2D. It counts 262,800 meshes and it takes, on average (depending on the flood duration), one and a half hours for one run with 38 parallel processors.

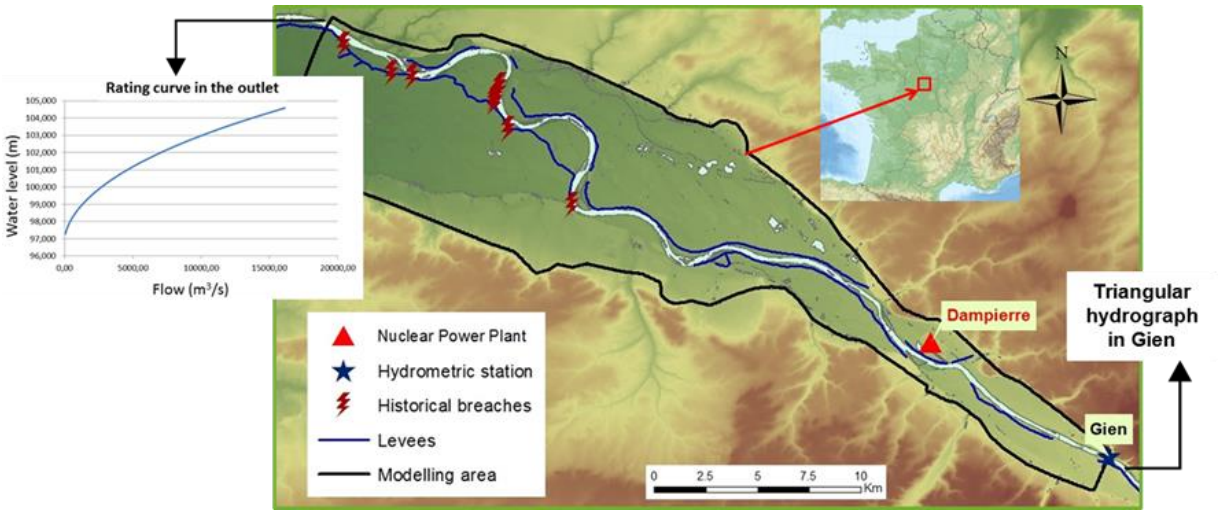


Figure 12: Modelling area of the Loire River

For this application case, we propose a methodology to take into account the dependencies between the model inputs in UQ and GSA. Here we only focus on hydrograph parameters. The objective is to evaluate the dependence load on the model outputs corresponding to the maximum water level inside the floodplain areas.

5.3.2 Methodology and results

The methodology can be summarised in 10 steps described hereafter (Figure 13).

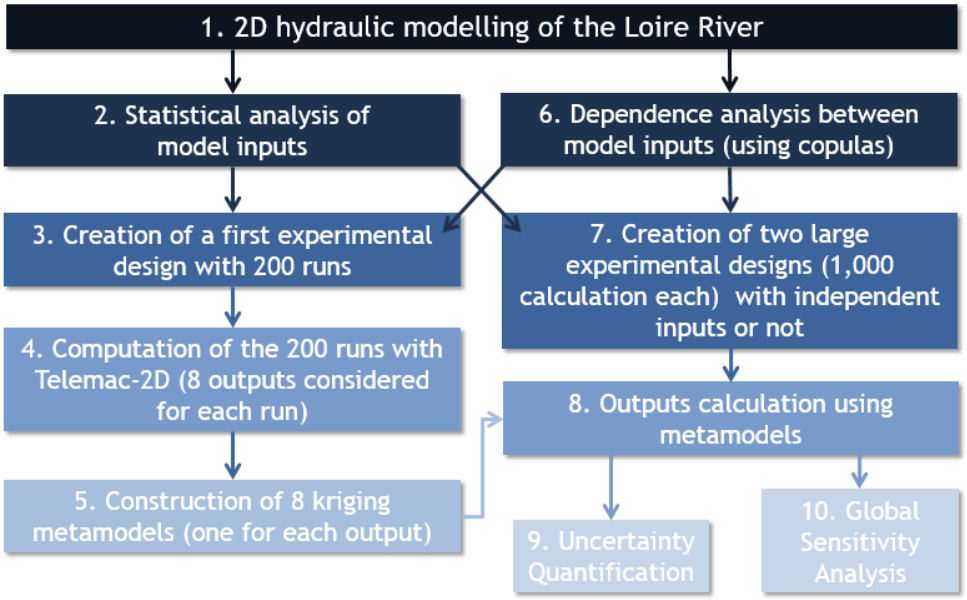


Figure 13: Methodology of the study

1. **Construction of the 2D hydraulic model and definition of the uncertain inputs and the desired outputs:** We considered 8 uncertain inputs (5 roughness coefficients (Strickler coefficients, K_s1 to K_s5) and 3 hydrograph parameters which are the total duration of the flood (*d*), the rise time (*tm*) and the maximum flow (*qmax*)) and 8 outputs corresponding to the maximum water level in some points of the floodplain (*P1* to *P8* in Figure 14).
2.

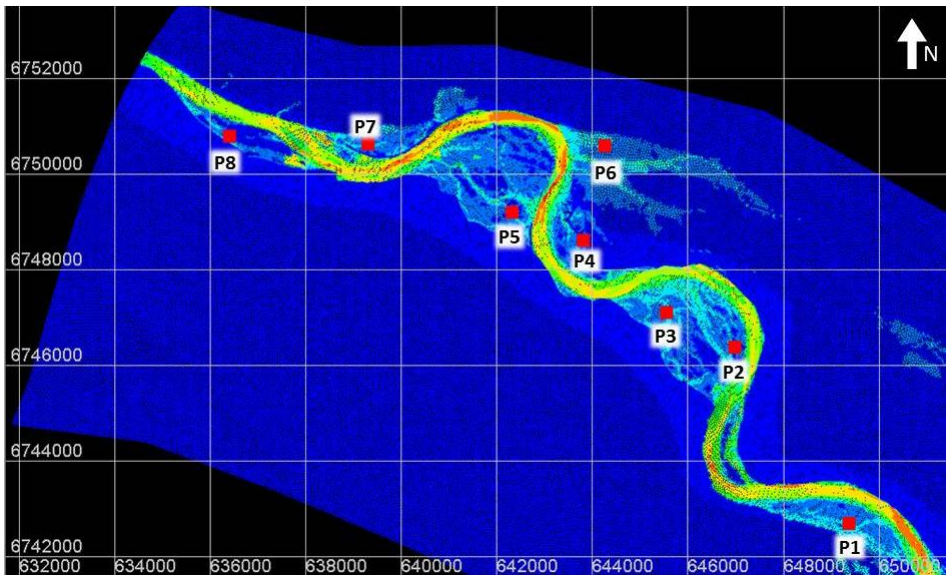


Figure 14: Outputs location

3. **Statistical analysis of model inputs and definition of the probability density functions (PDF) of all inputs using field data:** We consider that the 5 roughness coefficients (K_s1 to K_s5) follow uniform distributions ($25<K_s<45$). As there are calibration

coefficients, non-measurable in the River, we have chosen minimal and maximal bounds from Strickler roughness table. Regarding the hydrograph parameters, we extracted floods (182 floods between 1953 and 2019) from the flow data in the Loire River in Gien. For each flood, the maximum flow, the total duration and the rise time of the flood have been extracted. We researched the most accurate PDF of the 3 parameters: for the maximum flow it is the Gumbel distribution and for the maximum flow and the rise time it is the log normal distribution.

4. **Creation of a first experimental design with 200 runs:** In this step, all the inputs are considered to be independent and the values of each parameter are randomly sampled inside the PDF defined in step 2.
5. **Computation of the 200 runs with Telemac-2D and extraction of the maximum water levels in the 8 points defined in step 1:** To run a wide range of successive calculations with different inputs, Telemac-2D is coupled to the parametric computing environment developed in IRSN: Funz (https://github.com/Funz).
6. **Construction of 8 kriging metamodels:** As one simulation lasts between 36 minutes and 2 hours and a half (with 38 parallel processors), we had to use metamodels to launch a wide range of calculations. The idea is to replace the original model by a mathematical function to reduce the computational time. In this case we constructed and validated 8 kriging metamodels (one for each output).
7. **Dependence analysis between inputs (hydrograph parameters):** From the 182 floods extracted in step 2, the most accurate copula is selected. We have chosen a normal copula (class of meta-elliptical copula) to model the dependence between the 3 hydrograph parameters.
8. **Creation of two large experimental designs with 1,000 runs each:** For the first experimental design, all the inputs are considered to be independent so the inputs are sampled inside their own PDF. For the second experimental design, the 5 Strickler coefficients are randomly sampled inside their own PDF and the 3 hydrograph parameters, which are considered to be dependent, are randomly sampled inside a multivariate distribution defined by the combination between the normal copula previously defined and the PDF of each input.
9. **Outputs calculation using metamodels:** The computation time lasts less than 10 seconds with the metamodels, instead of 2,600 hours with the hydraulic model.
10. **Uncertainty Quantification:** We compared the outputs distribution considering dependent inputs or not, using boxplots (Figure 15) and empirical Cumulative Density Functions (eCDF) (Figure 16). Regarding the boxplots, it seems that there is almost no differences between the 2 cases (dependent or not). However, regarding the eCDF, the behaviour of the tail distribution is different between the independent and the dependent case for the downstream outputs (red circle in Figure 16). A good measure of the model dependence would be the calculation of the probability of the output variable exceeding a given threshold for both cases.

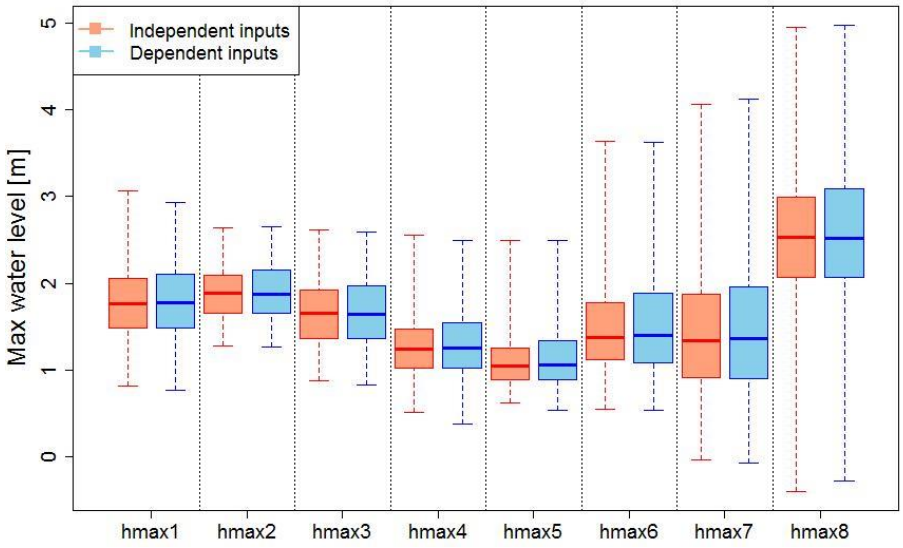


Figure 15: Boxplots of the 8 outputs considering dependent inputs or not

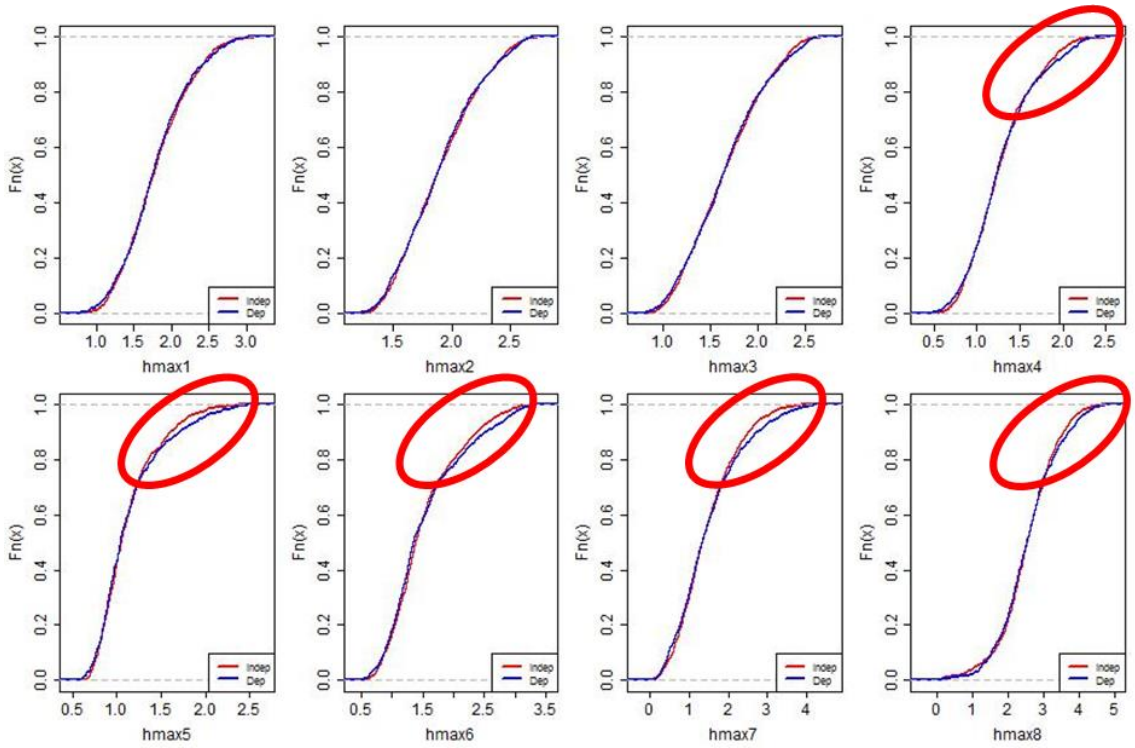


Figure 16: eCDF of the outputs considering dependent inputs or not (the blue curves corresponds to the case with dependent inputs and the red ones to the case with independent inputs)

Global Sensitivity Analysis: Here we used 3 methods to compute sensitivity indices (Looss and Prieur, 2018; Li and Mahadevan, 2016; McKay, 1996) for the 8 outputs and considering certain inputs dependent or not. The Figure 17 allows to compare the differences between the 8 outputs considering dependent inputs or not for the Li method. First, depending on the location of an output, the influence of each K_s coefficients differs. Then we observe that for all the inputs except *d* and *tm*, the indices are almost equal considering inputs dependency or not.

Finally, for *d* and *tm*, the indices considering certain dependent inputs are much higher than considering only independent inputs which results in a changing in parameter ranking. The Figure 18 shows the differences between the 3 methods previously cited. Finally we notice that there are few differences between the 3 methods. In addition the Li method is the fastest and with the Iooss and Prieur method (called "Shapley" in Figure 18) it is also possible to compute total order SI which are slightly higher than the 1st order sensitivity index. The dependence between variables from 2nd order Sobol indices can account for the dependence between variables.

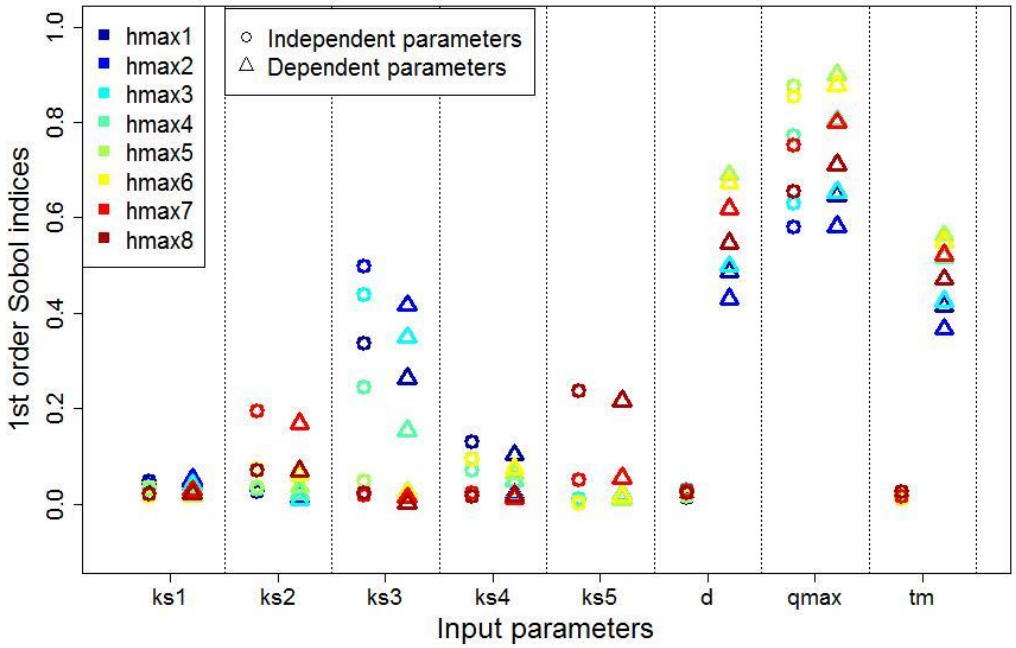


Figure 17: First order Sobol' indices computed for the 8 outputs with the Li method, considering dependent inputs or not

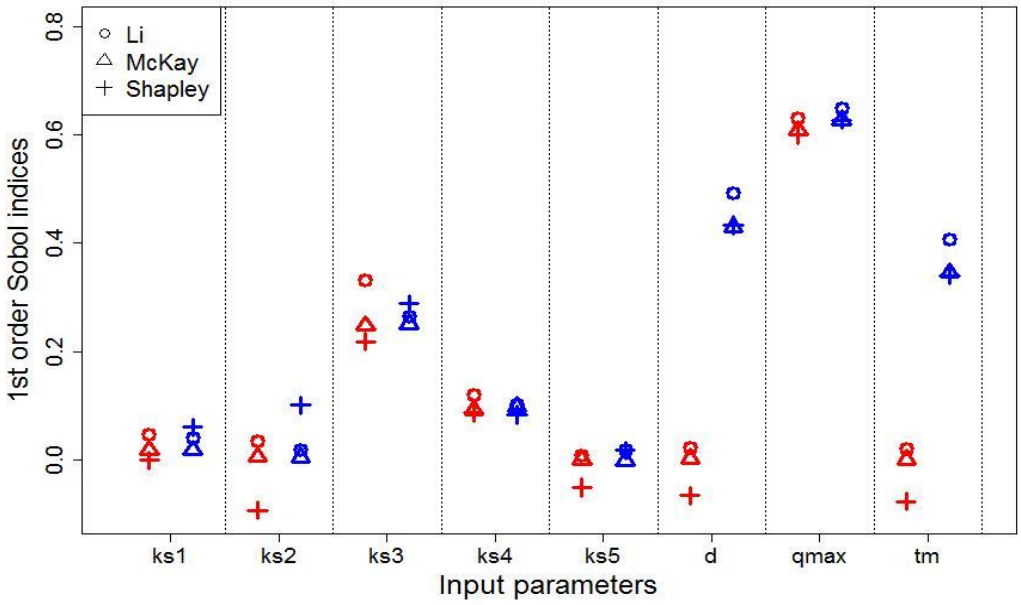


Figure 18: Comparison between the first order Sobol' indices computed with 3 different methods and considering dependent inputs or not

5.3.3 Conclusion and perspectives

The main conclusions of this work are:

- There is a strong dependence between the hydrograph parameters (d, qmax, tm) and this dependency can be taken into account by using copulas.
- The use of metamodels is very useful for uncertainty analysis studies (almost all with limited computation resources).
- There is a limited impact of inputs dependency in uncertainty quantification in this study.
- The duration and time to peak inputs have strong influence on the outputs.
- The hydrograph shape should not be ignored in hydraulic studies.

Currently this study is still under progress and will be published as soon as possible. We are trying to use Vine copulas (Joe and Kurowicka, 2011) to model the dependencies and we would like to study the influence of other hydraulic parameters dependencies (*i.e.* breach levee parameters).

6 Probabilistic models for non-stationary climate conditions and its effects on reliability targets (TU Delft)

6.1 Introduction

At present structural design codes are based on the assumption of stationary climate conditions. The reliability targets are typically specified, or understood, in terms of the annual probability of failure over the design life of the structure, typically 50 to 75 years. In the reliability-based design framework, the nominal design loads are specified in terms of return period or an upper percentile, e.g., 95^{th} percentile, of the annual maximum load distribution. Further, load factors are specified to calculate the factored design load. The load factors are calibrated with respect to the target reliability level (i.e., annual probability of failure). This is the essence of design approach adopted for wind and snow load cases, which are focus of this stage of investigation.

As the climate change effects are becoming evident, national code authorities are interested in evaluating the effect of non-stationary climate conditions on the structural design provisions. More specifically, it is being recognized that the reliability-based calibration approaches used under the traditional assumption of stationary climate condition cannot be directly extended to non-stationary climate cases. For example, the concept of the return period is inapplicable to non-stationary climate, as the distribution of the time between occurrences of the load events is no longer invariant. Similarly, the annual probability of failure is no longer constant in the non-stationary climate. This chapter will provide a technical discussion of such issues and they can be addressed using the theory of stochastic processes.

6.2 Probabilistic models

The application of probabilistic techniques to the structural design method was pioneered by Freudenthal (1956) in the U.S., and in parallel by Pugsley (1951) in the U.K., with the objective that structures achieve a minimum required level of reliability in a consistent and economical manner. Adaptation of a probabilistic approach to modernize structural design codes also inspired the development of probabilistic safety assessment methods in nuclear, aircraft, electrical, and instrumentation engineering fields.

A central question in a reliability-based design code is about the target or minimum required level of reliability of the structure that is designed using the code procedures. Initially, the target reliability (or probability of failure, P_f) values were not stated explicitly, and efforts were made to reproduce the level of reliability implied by the old working stress design methods. However, in recent times, the design standards have begun to include reliability targets in a more explicit manner (Baravalle and Kohler, 2019).

Failure probabilities of structures such as flood defences can be determined by analysing historical failure data or by probabilistic calculation of the limit states. For most cases there is not enough specific failure data available so we have to determine the failure probabilities by computation.

The probabilistic computation uses the reliability function and the probability density function of the variables as the base for the determination of the failure probability. A reliability function is a function of the strength and the load for a particular failure mode. In general the formulation of the reliability function is: Z=R-S in which R is the strength and S is the load. The failure mode

will not occur as long as the reliability function is positive. The graph of

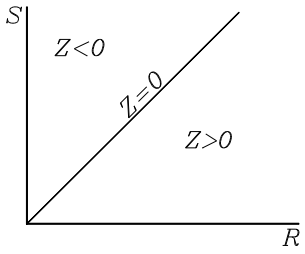


Figure ***19*** shows the reliability function. The line Z=0 is a limit state. This line represents all the combinations of values of the strength and the loading for which the failure mode will just not occur. So it is a boundary between functioning and failure.

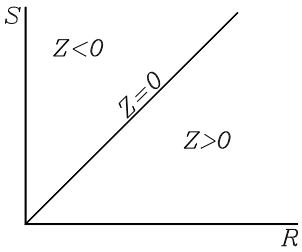


Figure 19: Reliability function

In the reliability function the strength and load variables are assumed to be stochastic variables. A stochastic variable is a variable which is defined by a probability distribution and a probability density function.

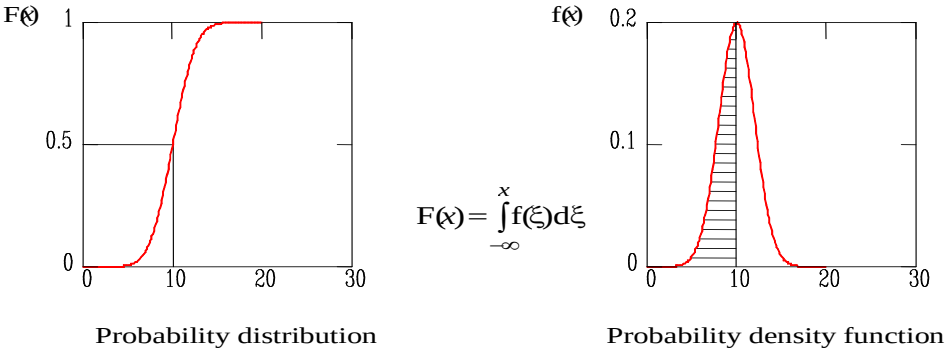


Figure 20: Probability distribution and probability density

The probability distribution F(*x*) returns the probability that the variable is less than *x*. The probability density function is the first derivative of the probability distribution. If the distribution an the density of all the strength and load variables are known it is possible to estimate the probability that the load has a value *x or higher* and that the strength has a value less than *x* (see Figure 21).

The failure probability is the probability that S=x and R<x for every value of x. So we have to compute the sum of the probabilities for all possible values of x:

$$\left.\begin{aligned} P(S=x) &= f_S(x)\,dx \\ P(R\le x) &= F_R(x) \end{aligned}\right\} \Rightarrow P(S=x \cap R\le x) = f_S(x)\,F_R(x)\,dx$$

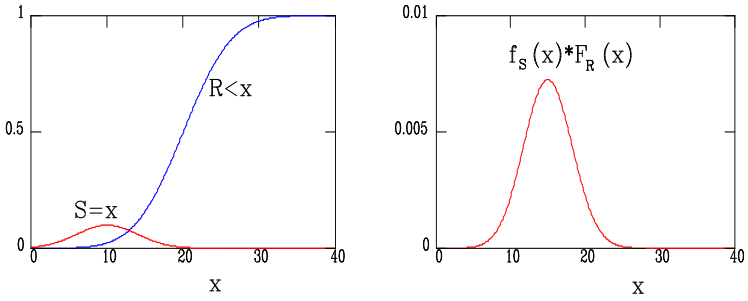


Figure 21: Components of the failure probability

$$P_f = \int_{-\infty}^{\infty} f_S(x)\,F_R(x)\,dx$$

This method can be applied when the strength and the load are independent of each other.

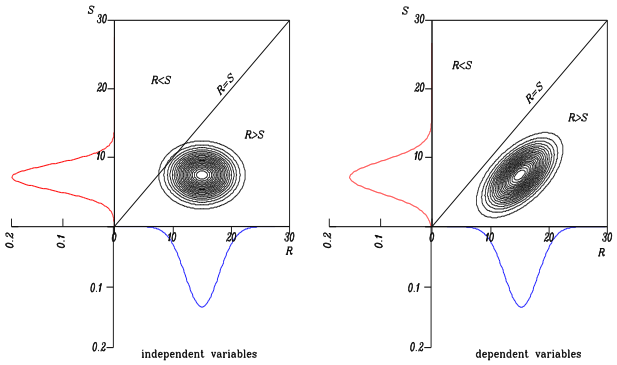


Figure 22: Joint probability density function

Figure 22 gives the joint probability density of the strength and the load for a certain failure mode in which the strength and the load are not independent.

The strength is plotted on the horizontal axis and the load is plotted on the vertical axis.

The contours give the combinations of the strength and the load with the same probability density. In the area (Z<0) the value of the reliability function is less then zero and the element will fail.

The failure probability can by determined by summation of the probability density of all the combinations of strength and load in this area:

$$P_f = \iint_{Z<0} f_{RS}(r,s)\,dr\,ds$$

In a real case the strength and the load in the reliability function are nearly always functions of multiple variables. For instance the load can consist of the water level and the significant wave height. In this case the failure probability is less simple to evaluate.

Nevertheless with numerical methods like numerical integration and Monte Carlo simulation, it is possible to solve the integral:

$$P_f = \iint_{Z<0} .. \iint f_{r_1, r_2, ..., r_n, s_1, s_2, ..., s_m}(r_1, r_2, ..., r_n, s_1, s_2, ..., s_m) dr_1 dr_2 ... r_n ds_1 ds_2 ... ds_m$$

These methods which take into account the real distribution of the variables are called level III probabilistic methods. In the Monte Carlo simulation method a large sample of values of the basic variables is generated and the number of failures is counted. The number of failures equals:

$$N_f = \sum_{j=1}^{N} 1\left(g\left(\mathbf{x}_j\right)\right)$$

In which N is the total number of simulations. The probability of failure can be estimated by:

$$P_f \approx \frac{N_f}{N}$$

The coefficient of variation of the failure probability can be estimated by:

$$V_{P_f} \approx \frac{1}{\sqrt{P_f N}}$$

In which P_f denotes the estimated failure probability.

The accuracy of the method depends on the number of simulations. The relative error made in the simulation can be written as:

$$\varepsilon = \frac{\frac{N_f}{N} - P_f}{P_f}$$

The expected value of the error is zero (estimator is unbiased). The standard deviation is given as:

$$\sigma_\varepsilon = \sqrt{\frac{1 - P_f}{N P_f}}$$

For a large number of simulations, the error is Normal distributed. Therefore the probability that the relative error is smaller than a certain value E can be written as:

$$P(\varepsilon < E) = \Phi\left(\frac{E}{\sigma_\varepsilon}\right)$$

$$N > \frac{k^2}{E^2}\left(\frac{1}{P_f} - 1\right)$$

The probability of the relative error E being smaller than $k\sigma_\varepsilon$ now equals $\Phi(k)$. For desired values of k and E the required number of simulations is given by:

Requiring a relative error of E = 0.1 lying within the 95 % confidence interval (k = 1.96) results in:

$$N > 400\left(\frac{1}{P_f} - 1\right)$$

The equation shows that the required number of simulations and thus the calculation time depend on the probability of failure to be calculated. Most structures in coastal- and river engineering possess a relatively high probability of failure (i.e. a relatively low reliability) compared to structural elements/systems, resulting in reasonable calculation times for Monte Carlo simulation. The calculation time is independent of the number of basic variables and therefore Monte Carlo simulation should be favoured over the Riemann method in case of a large number of basic variables (typically more than five). Furthermore, the Monte Carlo method is very robust, meaning that it is able to handle discontinuous failure spaces and reliability calculations in which more than one design point are involved (see below).

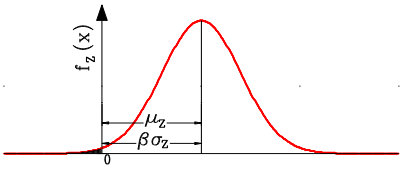
The problem of long calculation times can be partly overcome by applying importance sampling. This is not elaborated upon here.

If the reliability function (Z) is a sum of a number of normal distributed variables then Z is also a normal distributed variable. The mean value and the standard deviation can easily be computed with these equations:

$$Z = \sum_{i=1}^{n} a_i x_i \;,\quad \mu_Z = \sum_{i=1}^{n} a_i \mu_{x_i} \;,\quad \sigma_Z = \sqrt{\sum_{i=1}^{n} \left(a_i \sigma_{x_i}\right)^2} \;.$$

This is the base of the level II probabilistic calculation. The level II methods approximate the distributions of the variables with normal distributions and they estimate the reliability function with a linear first order Taylor polynomial , so that the Z-function is normal distributed.

If the distribution of the Z-function is normal and the mean value and the standard deviation are known it is rather easy to determine the failure probability:



$$P_f = P(Z < 0) = \Phi(-\beta)$$

$$\text{with:} \quad \beta = \frac{\mu_Z}{\sigma_Z}$$

Figure 23: Probability density of the Z-function

By computing β as μ divided by σ, it is possible to use the standard normal distribution to estimate the failure probability. There are tables available of the standard normal distribution in the handbooks for statistics.

In case of a non linear Z-function it will be estimated with a Taylor polynomial:

$$Z(\vec{x}) \approx Z(\vec{x}^*) + \sum_{i=1}^{n} \frac{\partial Z}{\partial x_i} \cdot (x_i - x_i^*)$$

The function is depends on the point where it will be linearised. The mean value and the standard deviation of the linear Z-function can be determined easily. If the reliability function is

estimated by a linear Z-function in the point where all the variables have their mean value ($x_i^* = \mu_{x_i}$) we speak of a Mean Value Approach.

The so-called design point approach estimates the reliability function by a linear function in a point on Z=0 where the joint PDF has a maximum. Finding the design point is a maximisation problem. For this problem there are several numerical solutions which will not be discussed here.

If the basic variables of the Z-function are not normally distributed the Z-function will be unknown and probably non normally distributed. To cope with this problem the non normally distributed basic variables in the Z-function can be replaced by a normally distributed variable (Box Cox transformation). In the design point the adapted normal distribution must have the same value as the real distribution. Because the normal distribution has two parameters (μ and σ) one condition is not enough to find the right normal distribution. Therefore the value of the adapted normal probability density function must also have the same value as the real probability density function (see Figure 24).

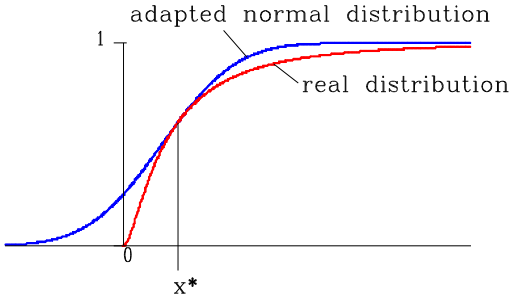


Figure 24: Adapted normal distribution

The two conditions give a set of two equations with two unknown which can be solved:

$$\left.\begin{aligned} F_N(x^*) &= F_x(x^*) \\ f_N(x^*) &= f_x(x^*) \end{aligned}\right\} \Rightarrow \mu_N, \sigma_N$$

This method is known as the Approximate Full Distribution Approach (AFDA).

6.3 Target reliability under stationary conditions

In the design of flood protection structures, we very often work with the design loads; i.e. the loads with a return period of T years. We call this way of designing: design by given probability of exceedance (Smith, 1993). Different choices for the assumed distribution of the loads will lead to different estimates in the T-year loads. In order to find the best distribution, we ought to specify some criteria. We would like to use a kind of criterion which is in relation with the consequences of a possible “wrong fit” (Slack et al., 1975).

A philosophy for the target reliability of flood defence structures and civil structures in general which takes into account the cost/benefit and the voluntariness aspect, was developed by the Technical Advisory Committee for Water Retaining Structures (TAW, 1984). The safety standard consists of a flexible evaluation of the individual and the societal acceptable risk but adds to these an economic approach taking the material damage into account. The latter provides the link with the safety philosophy of the Dutch dikes that was developed after the 1953 flood in the south-west of the Netherlands.

6.3.1 Personally acceptable level of risk

The smallest component of the socially accepted level of risk is the personal assessment of risks by the individual. As an attempt to model this appraisal procedure quantitatively is not feasible, it is proposed to look with the insight gained to the preferences revealed in the accident statistics.

The actual personal risk levels inherent to various activities show statistical stability over the years and are approximately equal for the Western countries, indicating a consistent pattern of preferences. The probability of losing ones life in normal daily activities such as driving a car or working in a factory appears to be one or two orders of magnitude lower than the overall probability of dying. Only a purely voluntary activity such as mountaineering entails a higher risk. This observation of public tolerance of 1000 times greater risks from voluntary than from involuntary activities with the same benefit was already made in the early 1970's.

In view of the consistency and the stability of the death risks presented, apart from a slightly downward trend due to technical progress, it would appear permissible to deduce therefrom a guideline for decisions with regard to the personally acceptable probability of failure:

$$P_{fi} = \frac{\beta_i \ 10^{-4}}{P_{d|fi}}$$

where $P_{d|fi}$ denotes the probability of being killed in the event of an accident.

In this expression the policy factor β_i varies with the degree of voluntariness with which an activity is undertaken and with the benefit perceived. It ranges from 100 in the case of complete freedom of choice like mountaineering, to 0.01 in case of an imposed risk without any perceived direct benefit. This last case includes the individual risk criterion for the siting of a hazardous installation near a housing area without any direct benefit to the inhabitants.

6.3.2 Socially acceptable level of risk

The basis of the framework with respect to societal risk is an evaluation of risks due to a certain activity on a national level. The risk evaluation on a national level has to be translated to local installations or activities in order to support a systematic appraisal by the local authorities. If a risk criterion is defined on a local level, the height of the national risk criterion is determined by the number of locations, where the activity takes place and the type of probability density function of the consequences of an accident. The resulting national norm has to be evaluated, as it was not intentionally formulated. It seems preferable to start with a risk criterion on a national level and to evaluate the acceptable local risk level in view of the actual number of installations, the cost/benefit aspects of the activity and the general progress in safety in an iterative process with say a ten year cycle.

6.3.3 Economically optimal level of risk

The problem of the acceptable level of risk can also be formulated as an economic decision problem. The expenditure I for a safer system is equated with the gain made by the decreasing present value of the risk (Figure 25).

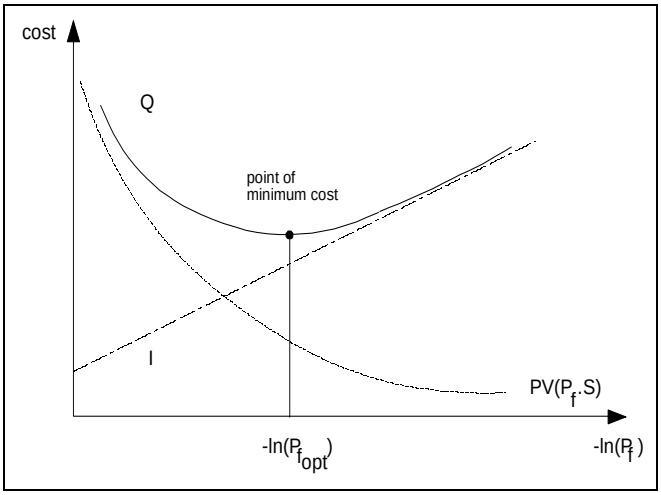


Figure 25: The economically optimal probability of failure of a structure

The optimal level of safety indicated by P_{fopt} corresponds to the point of minimal cost.

$$\min(Q) = \min(I(P_f) + PV(P_f \; S))$$

where :

Q = total cost

$I(P_f)$ = investment to obtain a safety level corresponding P_f

PV = present value operator

S = total damage in case of failure

If despite of ethical objections, the value of a human life is rated at s, the amount of damage is increased to:

$$P_{d|fi} \; N_{pi} \; s + S$$

where:

N_{pi} = number of participants in activity I .

This extension makes the optimal failure probability a decreasing function of the expected number of deaths.

The valuation of human life is chosen as the present value of the net national product per inhabitant. The advantage of taking the possible loss of lives into account in economic terms is that the safety measures are affordable in the context of the national income. Risk aversion may also be included in the economic approach as shown by Van Gelder and Vrijling (1997).

Standards for acceptable levels of risk

The Standard ASCE 7-16 specifies the target annual probability of failure for different categories of consequence, which are given in Table 3.

Table 3: Target probabilities of failure specified in ASCE 7-16 Standard

Table 1.3-1 Target Reliability (Annual Probability of Failure, P_F) and Associated Reliability Indices (β)[1] for Load Conditions That Do Not Include Earthquake, Tsunami, or Extraordinary Events[2]

	Risk Category			
Basis	**I**	**II**	**III**	**IV**
Failure that is not sudden and does not lead to widespread progression of damage	$P_F = 1.25 \times 10^{-4}$/yr, β = 2.5	$P_F = 3.0 \times 10^{-5}$/yr, β = 3.0	$P_F = 1.25 \times 10^{-5}$/yr, β = 3.25	$P_F = 5.0 \times 10^{-6}$/yr, β = 3.5
Failure that is either sudden or leads to widespread progression of damage	$P_F = 3.0 \times 10^{-5}$/yr, β = 3.0	$P_F = 5.0 \times 10^{-6}$/yr, β = 3.5	$P_F = 2.0 \times 10^{-6}$/yr, β = 3.75	$P_F = 7.0 \times 10^{-7}$/yr, β = 4.0
Failure that is sudden and results in widespread progression of damage	$P_F = 5.0 \times 10^{-6}$/yr, β = 3.5	$P_F = 7.0 \times 10^{-7}$/yr, β = 4.0	$P_F = 2.5 \times 10^{-7}$/yr, β = 4.25	$P_F = 1.0 \times 10^{-7}$/yr, β = 4.5

[1]The target reliability indices are provided for a 50-year reference period, and the probabilities of failure have been annualized. The equations presented in Section 2.3.6 are based on reliability indices for 50 years because the load combination requirements in Section 2.3.2 are based on the maximum loads for the 50-year reference period.
[2]Commentary to Section 2.5 includes references to publications that describe the historic development of these target reliabilities.

Table 4: Risk categories defined in Table 1.5.1 of ASCE 7-16 Standard

Use or Occupancy of Structure	Risk Category
Low risk to human life in the event of failure	I
All structures except those listed in Risk categories I, III and IV	II
(i) Failures could pose a substantial risk to human life, (ii) Failures with potential to cause a substantial economic impact and mass disruption of day-to-day civilian life, (iii) Facilities dealing with hazardous substance with quantity exceeding above a threshold	III
Structures designated as essential facilities	IV

In this Standard, a structure is classified into one of 4 categories based on the risk to human life and welfare resulting from their damage or failure. These categories are given in Table 4. The target annual probability of failure varies from 10^{-4} to 10^{-7} per year. The Standard stipulates 50 years as the reference life of the structure.

Table 5: Recommended minimum values for target reliability (ultimate limit state) specified in the European Standard EN 1990:2002

Reliability Class	1 year reference period	50 year reference period
RC3	10^{-7}	$8.5\ 10^{-6}$
RC2	$1.3\ 10^{-6}$	$7.2\ 10^{-5}$
RC1	$1.3\ 10^{-5}$	$4.8\ 10^{-4}$

Table 6: Description of reliability classes in the European Standard EN 1990:2002

Reliability Class	Consequence Class	Description
RC3	CoC3	High consequence for loss of human life, or economic, social or environmental consequences very great
RC2	CoC2	Medium consequence for loss of human life, economic, social or environmental consequences considerable
RC1	CoC1	Low consequence for loss of human life, and economic, social or environmental consequences small or negligible

The European standard EN 1990:2002 recommends values of target reliability for the annual and 50 year reference period (Table B2, page 59), which are presented in Table 5. Reliability classes associated with these targets are described in Table 6. In this Standard, Table C2 (page 65) explicitly acknowledges that these targets are meant for the structural member.

The National Building Code of Canada (NBCC) does not specify explicit reliability targets. A study evaluated that NBCC 1995 and 2005 imply on average a target probability of failure of a structural member as 1.3×10^{-3} over a 50 year reference period (Bartlett et al, 2003).

The IEC has developed several Standards for reliability-based design of electrical, mechanical, and electronic components used in a variety of applications. As an example, target reliability levels given in the standard IEC 61508, “Functional Safety of Electrical/Electronic/ and Programmable Electronic”, are summarized in Table 7 for four Safety-Integrity Levels (known as SILs), which are somewhat similar to reliability classes used in structural engineering standards. Functional safety is aimed at the active prevention of failure of a system from causing harm to people and property.

Table 7: Target failure rates specified in the IEC Standard 61508

Safety integrity level	High demand rate (dangerous failures/hr)	Low demand rate (probability of failure on demand)
4	$>= 10^{-9}$ to $< 10^{-8}$	$>= 10^{-5}$ to $< 10^{-4}$
3	$>= 10^{-8}$ to $< 10^{-7}$	$>= 10^{-4}$ to $< 10^{-3}$
2	$>= 10^{-7}$ to $< 10^{-6}$	$>= 10^{-3}$ to $< 10^{-2}$
1	$>= 10^{-6}$ to $< 10^{-5}$	$>= 10^{-2}$ to $< 10^{-1}$

This Standard uses two different reliability measures, namely, the failure rate for high demand systems and the probability of failure on demand for low demand systems. The high demand definition is used when the demand on a safety related function is greater than once per annum and the low demand definition is used when it is less frequent (IEC 61508, clause 3.5.14 of part 4).

The four SILs are defined as follows. SIL4 is the highest target and most onerous to achieve, requiring state of the art techniques (usually avoided). SIL3 is less onerous than SIL4 but still requiring the use of sophisticated design techniques. SIL2 requires good design and operating practice to a level such as that found in an ISO 9001 managements system. SIL 1 is the minimum level but still implies good design practice.

6.4 Target reliability under non-stationary conditions

Sea-level rise rates have become important drivers for policy makers dealing with the long-term protection of coastal populations. Scenario studies suggest that an acceleration in sea-level rise is imminent. The anticipated acceleration is hard to detect because of spatial and temporal variability, which consequently, have become important research topics. A known decadal-scale variation is the 18.6-year nodal cycle. Baart et al (2012) show how failing to account for the nodal cycle resulted in an overestimation of Dutch sea-level rise.

Similar types of time-dependent analyses are necessary for wind speed data, which trigger high water levels on sea, lakes and rivers.The conventional approach is to estimate the parameters of the Gumbel distribution from annual maximum wind speed observations and to estimate the corresponding quantiles according to:

$$f(x) = \alpha^{-1} \exp\{-(x-\xi)/\alpha\} \exp[-\exp\{-(x-\xi)/\alpha\}]$$
$$F(x) = \exp[-\exp\{-(x-\xi)/\alpha\}]$$
$$x(F) = \xi - \alpha \cdot \log(-\log F)$$

In order to account for time dependency, a time-dependent location parameter can be added to the Gumbel model (redefining theta):

$$F_{\max}(x,t) = \exp\left[-\exp\left(-\frac{x-\mu(t)}{\beta}\right)\right]$$

Where the location parameter is a function of time, such as a linear function:

$$\mu(t) = \alpha_0 + \alpha_1 t$$

The parameter estimation can take place by maximum likelihood method.

This section presents key definitions of the time-dependent reliability theory with appropriate analytical formulas to minimize any ambiguity in the interpretation of results.

The lifetime, X, of a structure is defined as the duration from the time of in-service to the occurrence of structural failure.

The lifetime is treated as a random variable with an appropriate continuous probability distribution. The lifetime distribution describes, in terms of probability, the ability of a component to survive up to a certain age.

In this section, the lifetime is denoted as X, which has the probability density, $f_X(x)$, and cumulative distribution, $F_X(x)$. The reliability at any age x is given by the reliability function, $R_X(x) = 1 - F_X(x)$. The probability that the component lifetime (X) will be in a range: $x_1 \leq X < x_2$ is given as:

$$p_{12} = F_X(x_2) - F_X(x_1)$$

The most precise definition of time-dependent reliability of a component is given in more technical texts of reliability, as follows:

“Reliability is a *conditional probability* at a given *confidence* level that the structure will perform its intended function satisfactorily at a *given age* for a specified *length of time* (mission) while operating under the specified application and the environment”

Here conditioning is on the fact that the structure is in a reliable state at the start of the operational period (or mission). Even if the operation starts at time 0, it is implicit that the structure is in working state at this time. From the definition of reliability, the word “conditional” has been unfortunately omitted by commonly used texts, which has led to many conceptual errors of interpretation in this field.

In a non-repairable problem, the reliability of a component in the time interval (x_1, x_2) essentially means that a component of age x_1 will not fail during an operational interval $(x_2 - x_1)$, $x_2 > x_1$, which is now defined as:

$$R_X(x_1, x_2) = P[X > x_2 | X > x_1] = \frac{1 - F_X(x_2)}{1 - F_X(x_1)} = \frac{R_X(x_2)}{R_X(x_1)}$$

This formula is correct for all values of x_1 and x_2.

A complement of the mission reliability is the mission probability of failure given as:

$$P_f(x_1, x_2) = 1 - R_X(x_1, x_2) = \frac{F_X(x_2) - F_X(x_1)}{1 - F_X(x_1)} = \frac{p_{12}}{R_X(x_1)}$$

where $p_{12} = F_X(x_2) - F_X(x_1)$.

The annual probability of failure is a popular metric of reliability, which in general varies with the age, x_1, of the component (or system). The annual probability of failure can be computed as:

$$P_{fA}(x_1, x_1 + 1) = \frac{F_X(x_1 + 1) - F_X(x_1)}{1 - F_X(x_1)}$$

Only in a special case of $F_X(x_1) \approx 0$, or $R_X(x_1) \approx 1$, the annual probability of failure can be approximated as i.e., $P_{fA}(x_1, x_1 + 1) \approx F_X(x_1 + 1) - F_X(x_1)$. Although this can happen in the early life of a highly reliable component, this approximation can be quite erroneous in the mid to late life of the component.

In the non-repairable problem, propensity to failure with age is quantified in terms of the failure rate or hazard rate, $h(x)$, which should be precisely written as “conditional” failure rate to distinguish it from the failure rate of a repairable system. Weibull is often used in reliability analysis for this purpose. It is conditional on the fact that the component is reliable at age x and its probability of failure in the next short interval, $(x, x + dx)$, is given as $h(x)dx$. The failure rate is given as:

$$h_X(x_1) = \frac{f_X(x_1)}{1 - F_X(x_1)} = \frac{f_X(x_1)}{R_X(x_1)}$$

From a reliability point of view, the failure rate is a continuous form of the annual probability of failure. In biostatistics, it is referred to as “mortality rate”, or more precisely, age-specific mortality rate.

In a group of *n* components, the number of failures in a time interval, (x_1, x_2), $x_1 < x_2$, is denoted as $N(x_1, x_2)$, which is a binomially distributed random variable with an expected (or average) value of:

$$E[N(x_1, x_2)] = n\big(F_X(x_2) - F_X(x_1)\big) = n\, p_{12}$$

The binomial distribution is valid under the assumption that failure events in the population are probabilistically independent. The annual failure frequency can be calculated for a one-year interval at an age t, which is essentially the following quantity:

$$E[N(t,t+1)] = n(F_X(t+1) - F_X(t))$$

A more precise definition of the expected number of failures per unit time at an age t can be given as:

$$\frac{d}{dt}E[N(t)] = \theta(t) = n \lim_{dt\to 0} \frac{\left(F_X(t+dt) - F_X(t)\right)}{dt} = n\, f_X(t)$$

Thus, the failure frequency in a non-repairable problem is equivalent to the probability density scaled by the original number of components in the population. In the long term, this failure frequency will approach zero since $f_X(x) \to 0$ as time $x \to \infty$.

It is important to note that neither the expected number of failures nor the failure frequency is a measure of the reliability of a component in the time interval, (x_1, x_2).

In general, the reliability of a component is uniquely related to the failure rate function as:

$$R_X(t) = e^{-H(t)}, \text{ or } \ln[R_X(t)] = -H(t)$$

where $H(t)$ is the integrated failure (hazard) rate, given as: $H(t) = \int_0^t h(u)du$.

The relation above, which shows that the reliability depends on the integrated failure rate, plays a key role in setting up the target reliability requirements in structural design codes, which consist of two independent limits:

(1) on the overall reliability over a stipulated design life (t_d), e.g., $R_X(50)$. This is called end of life reliability target, and
(2) on failure rate, $h(t)$, less than a limit for all $t \leq t_d$.

These two constraints ensure that (1) a minimum required safety level is not compromised at any point in time in the service life of the structure, and (2) a structure cannot be used indefinitely without a thorough condition assessment and life extension work at the end of first life.

It should be stressed that specifying only a single target limit on overall reliability, $R_X(50)$, or the failure rate will not ensure satisfying the two requirements discussed above. The reason is that the reliability depends on the integrated value of the failure rate, such that it cannot constrain a point value of $h(t)$. Thus, it is possible that a reliability requirement is satisfied while the failure rate could reach an unacceptable level in a time interval within the service life of the structure.

This concept can be further explained by an example. In mountainous regions highway traffic is generally closed when a section of the road is exposed to a higher risk of land slide (or avalanche), even though the condition of most of the highway is trouble free. This precaution is taken to ensure nobody is affected by a high, concentrated risk, even in a short portion of travel.

This section has presented technically precise definitions of various terms used in the reliability engineering field.

Terms such as: annual probability of failure, failure rate, and failure frequency are not synonymous in a “non-repairable” equipment reliability problem.

Nevertheless, in case of a highly reliable system in the reference evaluation period of the component, the annual probability of failure and failure rate are numerically equal to the probability density, i.e., annual failure frequency.

It is important to note that reliability depends on the integrated value of the failure rate, which plays a key role in setting up the target reliability requirements in structural design codes. The reliability targets have to be specified both at the lifetime reliability level, as well as the failure rate level (annual P_f) to ensure that a minimum required safety level is maintained at every point in time in the service life of the structure.

7 Application of approaches for assessing duration of extreme events

We now illustrate the application of the approaches detailed in section 4 for a set of different case studies (heatwaves and wind storms). Each of these use observation data from sites in the vicinity of Mülheim-Kärlich NPP. Depending on the hazard of interest, different sites and length of data will be available and as such the different data sets are introduced at the start of each section.

7.1 Heatwaves

In this section, we take air temperature observations in the vicinity of Mülheim-Kärlich NPP (section 7.1.1) and analyse the likelihood and severity of heatwave events. This includes an initial analysis of empirical measures (section 7.1.2) and then an application of a first-order Markov model as outlined in Winter et al. (2016) (see section 7.1.3).

7.1.1 Data

Extreme air temperature observations are available at a set of sites in the vicinity of Mülheim-Kärlich NPP. The available data are quality controlled measurements and observations derived from Deutscher Wetterdienst (DWD) stations and legally and qualitatively equivalent partner stations operated for climatological and climate related applications. As we are focusing on extreme hot temperature, daily maxima air temperature during summer months (June, July and August) are kept for the analysis, as it allows us to remove the seasonality of the observations. Figure 26 shows the locations of different sites where air temperature observation data are available and Table 8 specifies the altitudes of each gauge that varies from 62 m ASL at Bonn-Friesdorf to 627 m ASL at Nurberg. As a reminder, the altitude of the NPP site under consideration is approximately 65 m ASL.

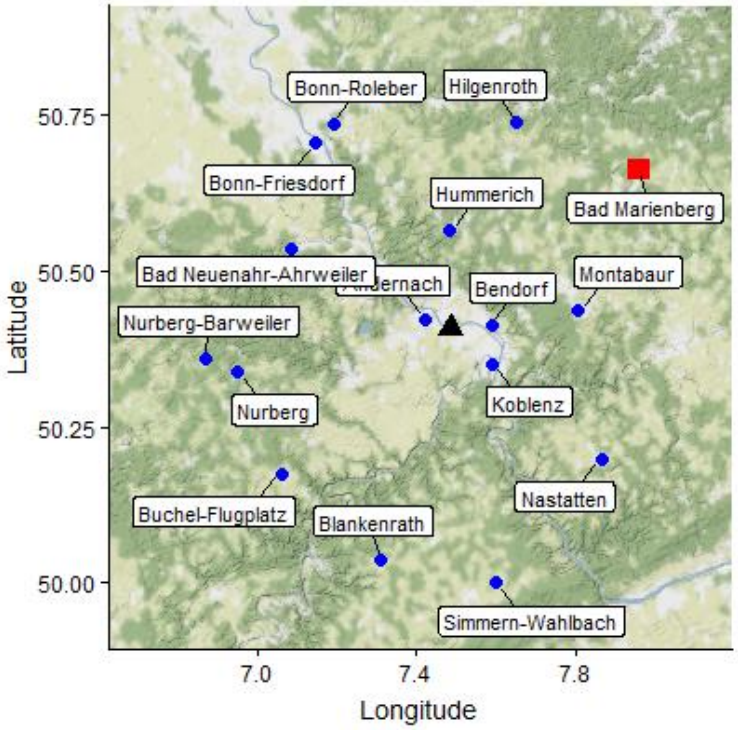


Figure 26: Location of weather gauges with air temperature data series (blue dots), representative site used for most of the study (red square) and Mülheim-Kärlich NPP (black triangle).

Table 8: Gauge information for the different weather gauges used for the air temperature analysis.

Gauge name	Gauge ID	Altitude (m)	Latitude	Longitude
Andernach	00161	75	50.4237	7.4202
Bendorf	00348	127	50.4135	7.5886
Blankenrath	00535	417	50.0372	7.3079
Bonn-Friesdorf	00599	62	50.7055	7.1467
Bonn-Roleber	00603	159	50.7349	7.1931
Buchel-Flugplatz	00766	477	50.1746	7.0595
Hilgenroth	02211	295	50.7371	7.6528
Hummerich	02362	328	50.5651	7.4843
Koblenz	02656	96	50.3519	7.5906
Bad Marienberg	03167	547	50.6620	7.9602
Montabaur	03340	265	50.4383	7.8061
Bad Neuenahr-Ahrweiler	03490	111	50.5346	7.0853
Nurberg	03659	627	50.3391	6.9501
Nurberg-Barweiler	03660	485	50.3601	6.8697
Simmern-Wahlbach	04709	445	49.9996	7.5981
Nastatten	06186	286	50.1989	7.8651

We check the amount of air temperature data available at each site (Figure 27). The earliest data records are from the 1950s, although there are only 3 series with data spanning from 1960s until the 1990s. There are many more series with data from the mid-2000s, however these series are too short to provide reliable results from extreme value analysis. We choose to use the dataset from Bad Marienberg (03167) for the rest of this study as it is the longest

continuous series. Figure 28 shows the daily maximum air temperature series for the Bad Marienberg weather gauge (noting elevation differences exist).

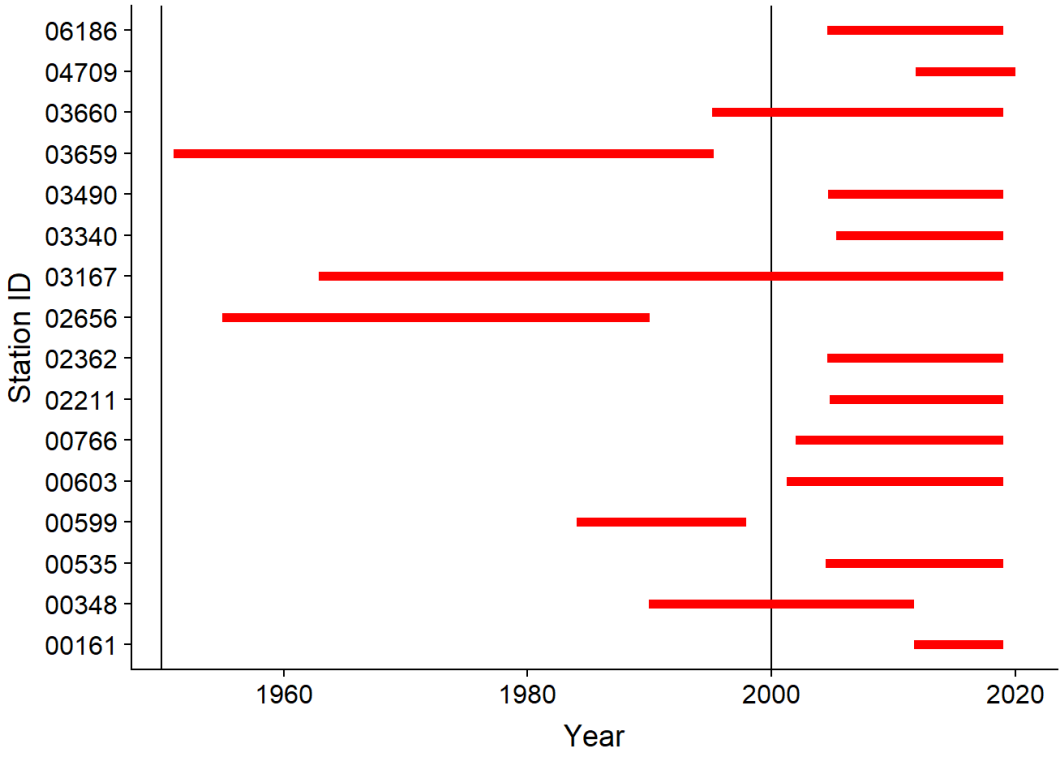


Figure 27: Time period over which air temperature data are available for each site; the plot only shows the start and end date of each of the series, there may be additional missing gaps within the series which are checked later in the analysis.

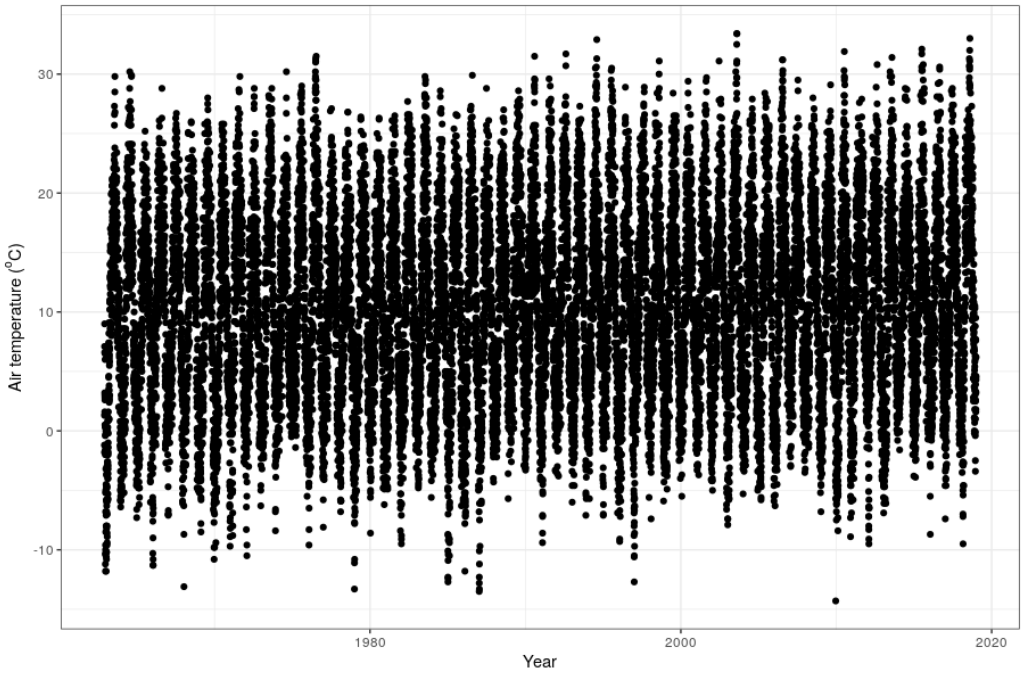


Figure 28: Daily maximum air temperatures (ºC) at the Bad Marienberg weather gauge.

7.1.2 Initial empirical analysis

An initial assessment of the duration of heatwave events can be made by looking at the observation time-series (Figure 28) and counting the number of exceedances above a set of extreme levels. Table 9 provides some summary statistics for a set of critical levels comprising of 20ºC, 25ºC, 30ºC and 35ºC. As expected, as the level is increased there are fewer exceedances and fewer heatwave events. It should be noted that the heatwave events have been defined using runs declustering with run length 1 (i.e. an event is ended once there is a single day that falls below the critical level). The heatwave events that are within the observational data also seem to last for fewer days as the critical level is increased, as indicated by the longest events above each level and the respective empirical estimates of the extremal index.

Table 9: Summary of empirical measures for assessing the duration of events applied to the daily maximum air temperature data at Bad Marienberg.

Critical level (ºC)	Number of exceedances	Number of separate events	Longest event (days)	Extremal index
20	2747	815	33	0.08
25	629	296	15	0.23
30	47	32	4	0.34
35	0	0	0	n/a

An important result from Table 9 is that as the critical level is increased, the number of available events within the observational dataset reduces sharply. As such, using the empirical estimates up to a point are reasonable but are unlikely to give adequate results at the very extreme levels that we are most interested in. For this, we need to start applying the types of models outlined in sections 4.4 and 4.5.

7.1.3 Application of multivariate EVA

Having applied empirical approaches in section 7.1.2, we now apply a framework built upon extreme value analysis to provide estimates of the durations of heatwave events. To do this we need to follow the steps outlined in section 4.1.2. We shall initially fit the model to the daily maximum air temperature data (in the June July August period) from the Bad Marienberg weather gauge (Figure 29).

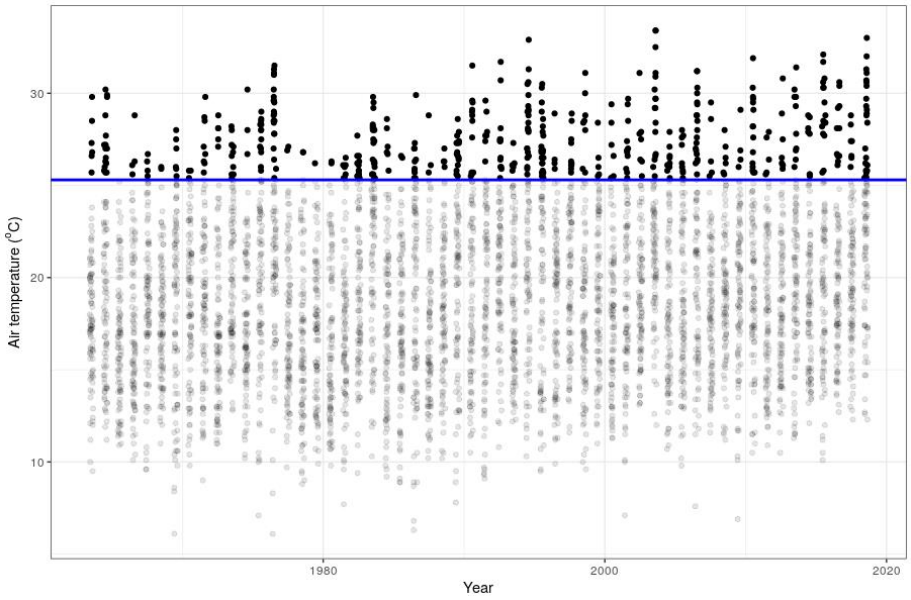


Figure 29: Summer daily maxima air temperature time series at the Bad Marienberg weather gauge. A threshold for EVA is shown (blue line) with exceedances (solid black dots) and non-exceedances (translucent black dots).

7.1.3.1 Fit univariate EVA model

The first step is to fit a univariate extreme value model; here, we choose to fit a GPD to exceedances above a high threshold. For estimating return levels, the standard type of approach would be to fit a GPD to declustered data. However, for the analysis of duration we need to make an amendment to this previous analysis. The previous analysis used a declustering scheme to obtain the peaks-over-threshold (POTs) and the GPD was fitted to these POTs. This ensured that the data used were independent of one another which is a key assumption when estimating return levels using the GPD. However, in this analysis we are aiming to explicitly capture the extremal dependence using a multivariate extreme value model and as such do not want to remove the dependence in the time-series using declustering. As a result, we proceed to fit the GPD to all the exceedances of the chosen threshold (but we shall not attempt to use this fit to estimate return levels).

The threshold selected for this analysis is the 90th quantile which is 25.3ºC; this threshold choice is supported by threshold diagnostic plots (not shown). Figure 30 provides a QQ-plot of the GPD model fitted to all the exceedances of the threshold. The model provides a good fit to the data as illustrated by the fact that the points are lined up along the diagonal.

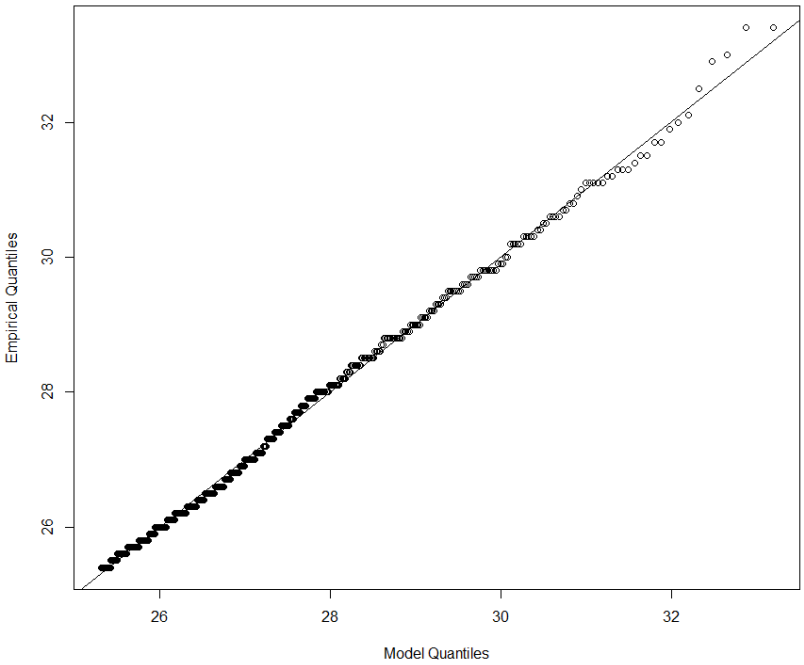


Figure 30: QQ-plot of the GPD fit to all exceedances above threshold of 25.3ºC.

Parameter estimates for the statistical model are provided in Table 10. The scale and shape parameter estimates are both reasonable, although the estimate of the shape parameter is still quite negative which will lead to a strongly bounded probability distribution. We also observe that the standard error is smaller than for a standard analysis, but this is most likely caused by our use of all the exceedances as opposed to the POTs.

Table 10: Parameter estimates for the GPD fit to all exceedances above threshold of 25.3ºC.

Threshold (˚C)	Quantity	Scale	Shape
25.3	Best estimate	2.87	-0.31
	Standard error	0.16	0.03

7.1.3.2 Choose the order of the Markov model to be fitted

Before fitting the multivariate extreme value model, it is important to understand the strength of temporal dependence within the dataset we are analyzing. We shall capture the temporal dependence through a Markov model (see section 4.5 for more information); the easiest way to estimate this is using the autocorrelation function (ACF) and partial autocorrelation function (PACF). Figure 31 shows the ACF and PACF for the daily maximum air temperature data at the Bad Marienberg weather gauge. The ACF plot suggests that there is dependence in the time-series out to approximately lag 25. However, the PACF (which measures the amount of correlation between two lags conditional upon the intervening values) seems to drop off sharply after lag 1. As such, a good first approximation is to assume a first-order Markov model. This is the model that we shall fit in section 7.1.3.3. However, there is some indication that there is some dependence out to lag 3 which suggests that we should test whether higher-order Markov models can provide a more accurate result. This is not done during this analysis but an approach is outlined in Winter and Tawn (2017).

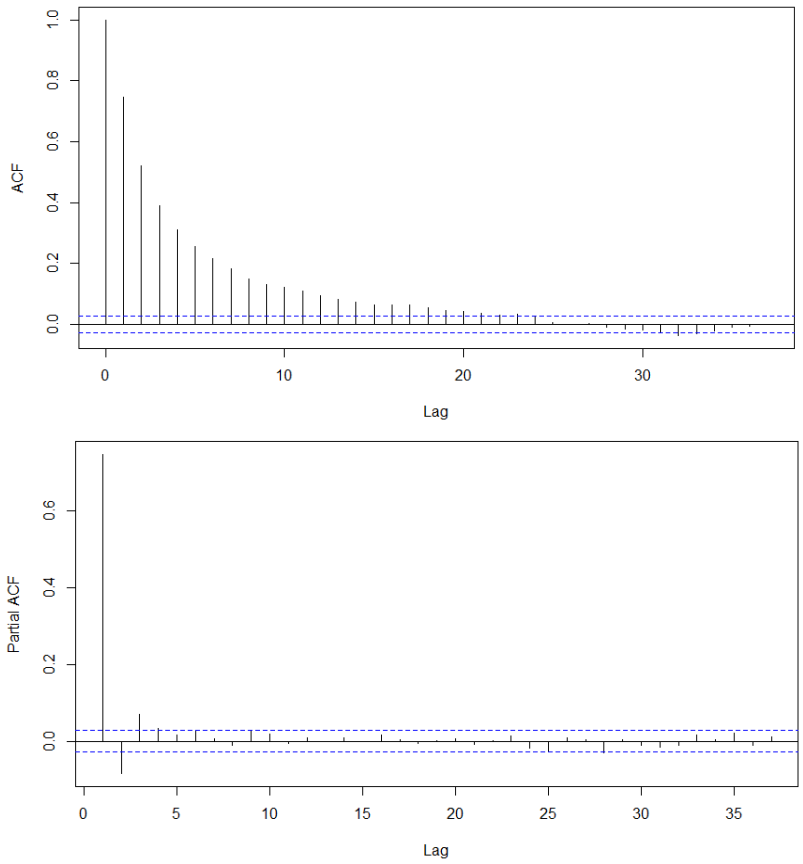


Figure 31: Plots of the autocorrelation function (ACF, top) and the partial autocorrelation function (PACF, bottom) for the summer daily maximum air temperature series at Bad Marienberg.

7.1.3.3 Fit multivariate extreme value model

The initial assumption of a first-order Markov model means that we shall model the temporal dependence using a conditional model of the form $X_{t+1}|X_t$ (i.e. the air temperature at time step $t+1$ only depends on the air temperature value at time step t). Figure 32 shows a scatter plot of consecutive daily maximum air temperature observations. As expected, we clearly observe a positive correlation between air temperatures on consecutive days.

It is now possible to make an empirical estimate of two commonly used extremal dependence metrics ($\chi(v)$ and $\bar{\chi}(\mathrm{v})$) using the data illustrated in Figure 32. Figure 33 shows the empirical estimate of $\chi(v)$ and $\bar{\chi}(\mathrm{v})$ for a range of different quantiles. Here, we are most interested in the behaviour of the quantities as the quantile approaches 1 since this is the joint extreme area of our variables. The value of $\bar{\chi}(\mathrm{v})$ is clearly contained between 0 and 1 for high quantiles which suggests asymptotic independence with positive association at subasymptotic levels. This conclusion seems to fit well with the plot in Figure 32 since there is clearly positive association within the data, but this seems to become less positive when looking at the most extreme values within our data set (e.g. looking at the spread of air temperatures on day $t+1$ given that the temperature on day t was over 30ºC).

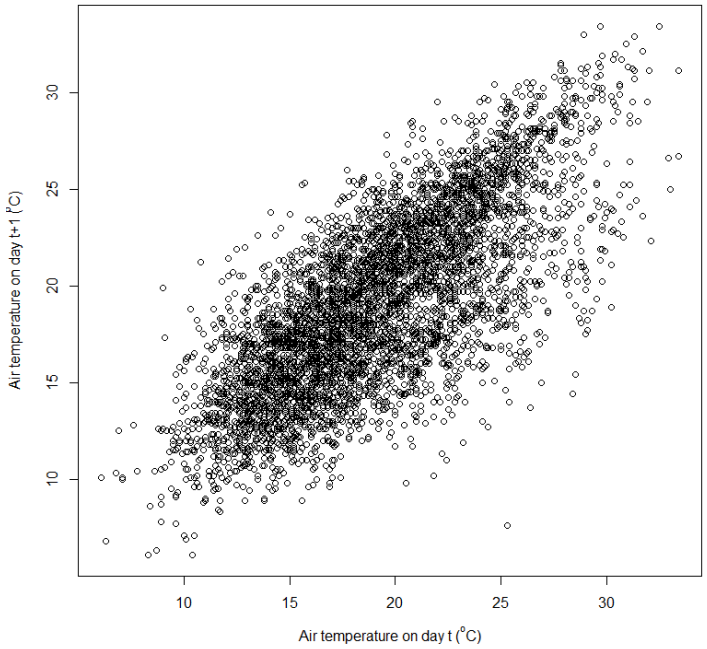


Figure 32: Scatter plot of daily maximum air temperatures (ºC) on consecutive days (i.e. day t and day t+1).

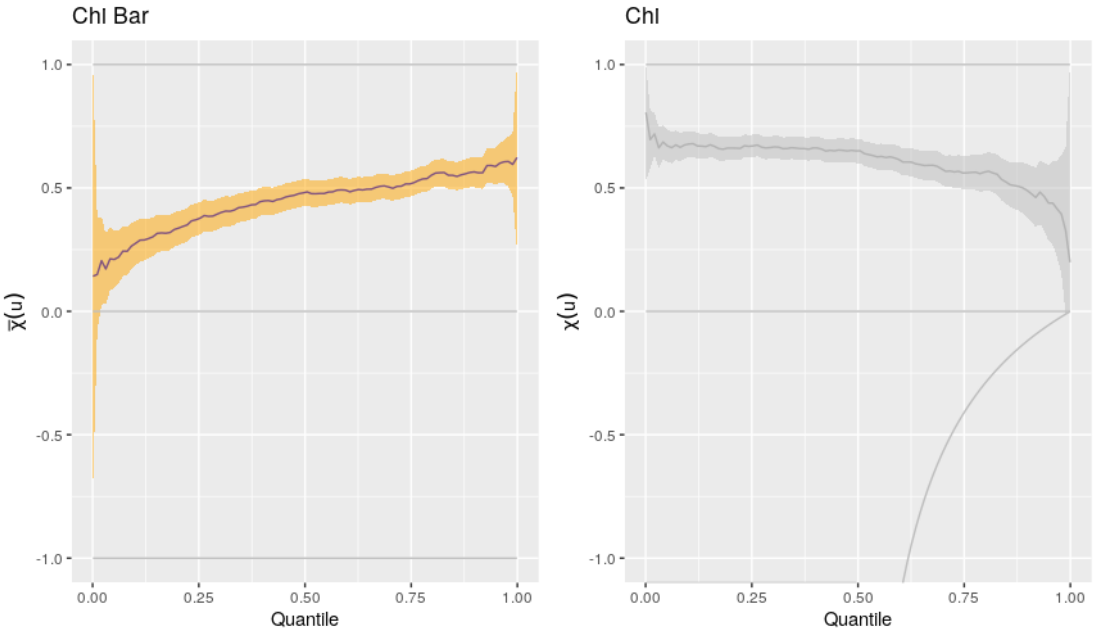


Figure 33: Empirical estimates of $\bar{\chi}(u)$ (left) and $\chi(u)$ (right) for a set of thresholds u (or equivalent quantiles).

The scatter plot on the original margins (Figure 32) contains contributions from the marginal behaviour and dependence structure. We can use the GPD fit from section 7.1.3.1 with a probability integral transform to convert from the original margins onto unit margins to remove the contribution of the marginal distribution; the result is shown in Figure 34. The important aspect of this plot is in the top right-hand corner where there are many points grouped together.

This highlights that there is still positive extremal dependence between consecutive daily maximum air temperatures when we've removed the marginal behaviour. This means it is sensible to continue to fit a dependence model to the data.

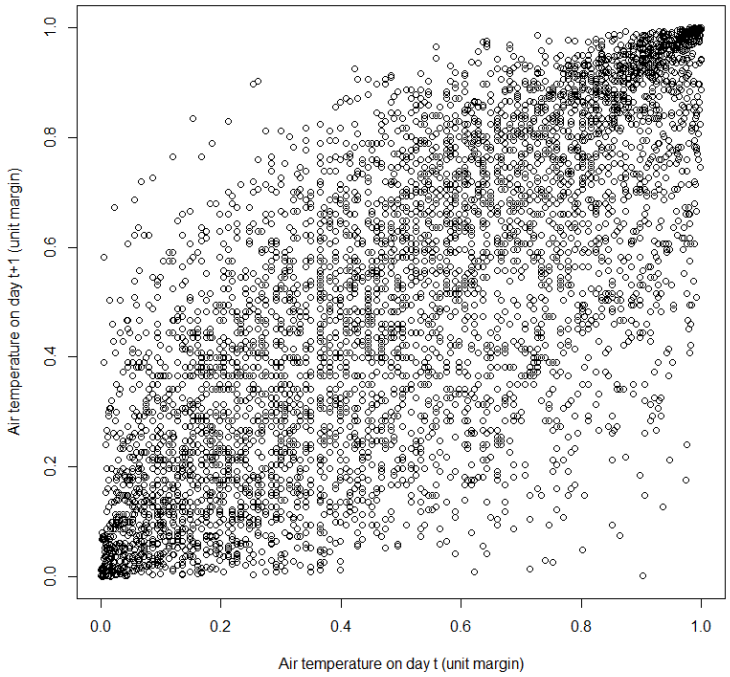


Figure 34: Scatter plot of daily maximum air temperatures (unit margin) on consecutive days (i.e. day t and day t+1).

We fit the conditional extremes model (section 4.4.4) to the data from Figure 34 to obtain an estimate of the fitted model of $X_{t+1}|X_t > u$. For consistency, in this example we shall continue to use the threshold of 25.3ºC as in the fit of the univariate extreme value model. Table 11 provides estimates of the main two parameters within the conditional extremes model (α, β). The parameters tend to trade off against one another which makes it hard to get direct inferences from the model parameter estimates (see section 4.4.4 for some examples of sets of parameters with specific meaning). It does seem that our estimates sit somewhere between the cases of positive asymptotic dependence and asymptotic independence which seems to agree with Figure 34 (i.e. there is some sort of positive dependence).

Table 11: Best estimates and standard errors (estimated using bootstrapping with 1000 resampled datasets) of the two main dependence parameters from fitting the conditional extreme model.

Threshold (°C)	Quantity	α	β
25.3	Best estimate	0.29	0.61
	Standard error	0.19	0.11

7.1.3.4 Use fitted model to construct simulated events

The fitted multivariate extreme value model can now be used to create heatwave events with similar statistical properties to the observed time series. As in section 4.5.2, we shall use tail chain simulation to create 50,000 replicate heatwave events and use these to construct the duration distribution $\pi(i, v)$ which is broadly the probability of having an event lasting for i days above a level v. The duration distribution can also be used to estimate the extremal index at a

level v (i.e. the expected number of exceedances of the level v during a heatwave event).

Figure 35 shows estimates of the extremal index obtained using the first-order Markov model and empirically from the original data series. We have chosen critical levels between 25ºC and 32ºC as this represents a good range of thresholds within the range of our available observation data. Choosing values of the critical level above 32ºC leads to very unstable estimates of the empirical extremal index as there is little data exceeding this level. The plot illustrates that the outputs from the first-order Markov model seem to match the empirical estimate well and comfortably fall within the 95% confidence intervals. This provides us with confidence that the model is representing the behaviour within the observed data well.

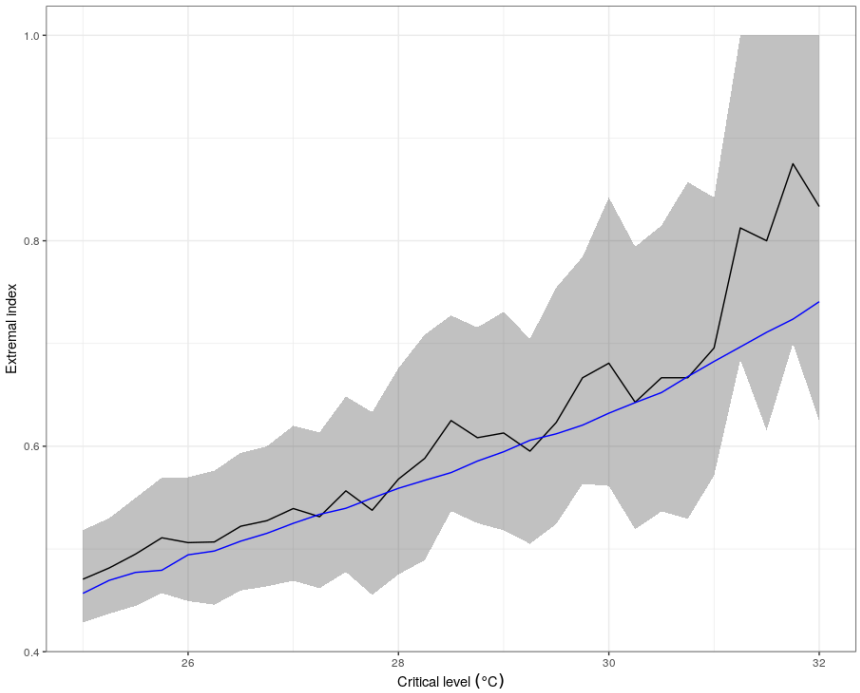


Figure 35: Estimates of the extremal index for a set of different critical levels including: empirical estimate (black line) with associated 95% confidence interval (grey shaded area); estimate from first-order Markov model (blue line).

Figure 36 shows a similar result but is showing the average length of heatwave events above the range of critical levels. This metric is obtained by taking the reciprocal of the extremal index that was illustrated in Figure 35.

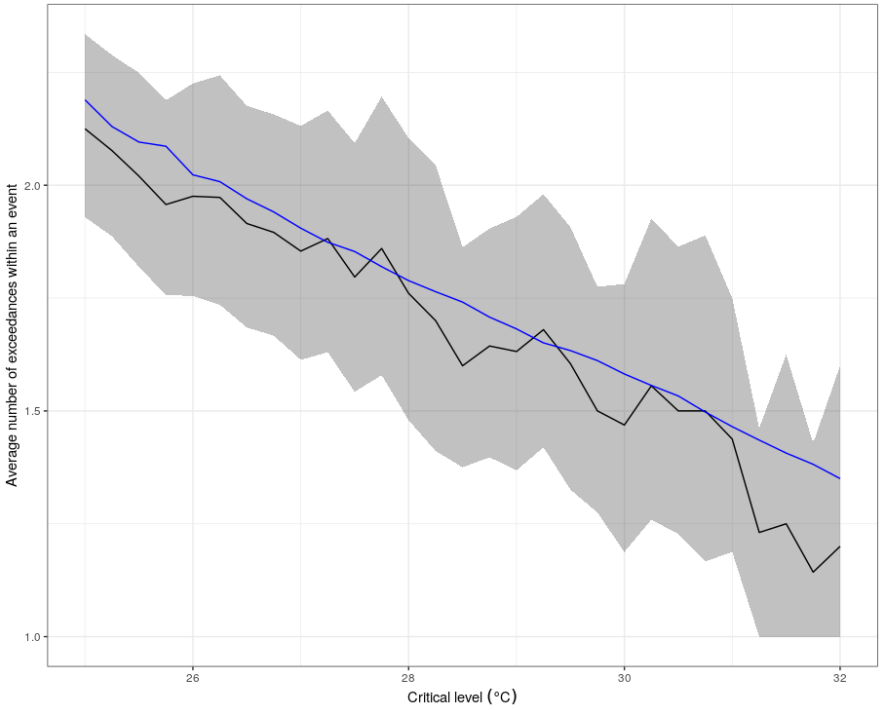


Figure 36: Average number of exceedances within a heatwave event for a set of different critical levels including: empirical estimate (black line) with associated 95% confidence interval (grey shaded area); estimate from first-order Markov model (blue line).

Having now checked that the first-order Markov model fits the observations well within the range of the data, we can use the model to estimate the extremal index with uncertainty bounds. Table 12 provides estimates of the extremal index for a range of critical levels. As expected, the estimate of the extremal index increases as the critical level is increased. Irrespective of the critical level, the average number of exceedances within an event remains between 1 and 2 days.

Table 12: Best estimates of the extremal index and average number of exceedances per heatwave event obtained using the first-order Markov model. Values in parentheses represent 95% confidence intervals obtained using 1000 bootstrap replications.

Critical level (°C)	Extremal index	Average number of exceedances per event
28	0.56 (0.52, 0.60)	1.79 (1.66, 1.94)
30	0.63 (0.59, 0.71)	1.58 (1.41, 1.68)
32	0.74 (0.65, 0.85)	1.36 (1.18, 1.55)

Although the extremal index is an interesting quantity, it only tells us about the average length of extreme events above a specific critical level. We are also interested in the probability of events that last for more days above the critical level. As such, we are also going to estimate the probability of observing an event lasting for 3 or 5 days above the respective critical levels from Table 12; these results are provided in Table 13. The estimates are presented as annual exceedance probabilities and return periods for completeness (although both results contain the same information in slightly different forms). As expected, the longer an event lasts above a critical level then the less likely it is to occur; this is also the case for increasing the critical level.

Table 13: Best estimates of annual exceedance probability (to two decimal places) and return period (rounded to the nearest year) for 3- and 5-day heatwave events above the respective critical levels. Values in parentheses represent 95% confidence intervals obtained using 1000 bootstrap replications.

Critical level (°C)	Annual exceedance probability		Return period (years)	
	3-day	5-day	3-day	5-day
28	0.09 (0.08, 0.10)	0.02 (0.02, 0.03)	11 (10, 12)	41 (33, 61)
30	0.07 (0.06, 0.09)	0.01 (0.01, 0.02)	13 (11, 18)	69 (45, 171)
32	0.05 (0.02, 0.07)	0.01 (0.00, 0.01)	21 (14, 54)	181 (71, 2030)

7.1.4 Summary of analysis

In this case study, we have used the output of a multivariate extreme value model to estimate the duration of heatwave events in the vicinity of Mülheim-Kärlich NPP. The approach outlined has permitted the explicit modelling of extremal temporal dependence which is not possible when fitting a univariate extreme value model to declustered extremes above a threshold. There are several clear benefits to this approach as opposed to empirical methods or standard statistical models:

- We can extrapolate beyond the range of the current data to estimate the duration of events that are more intense than previously observed;
- The extremal dependence is captured using a model designed for this purpose, compared to standard time-series approaches which capture typical behaviour better than extremes.

The analysis of daily maximum air temperatures has highlighted extremal dependence between consecutive extreme hot days. On average, events tend to last for 1-2 days above a set of different critical levels. However, longer events are certainly possible and well within the 10,000-year return period timescale; for example, the return period of having a 5-day event above a threshold of 32ºC is 181 years (71, 2030).

There are improvements that could be made to the modelling approach which could lead to additional insight:

- The results provided in section 7.1.3.4 are based upon the assumption that the first-order Markov model is appropriate (as justified by the PACF in Figure 31). Figure 35 shows that the model results are quite consistent with the empirical estimates for the extremal index. However, it is useful to understand whether a higher-order Markov model could provide an even better fit to the observed data.
- The daily night-time temperature is often an important measure in excess human mortality during a heatwave. Here, we have not directly modelled this, but a small adjustment to our model could permit this to be accounted for.

7.2 Wind storms

For the wind storm case study, we use hourly wind speed data close to Mülheim-Kärlich NPP to illustrate how the multivariate extreme value model can be used to improve the estimates of univariate return levels, in particular to reduce the uncertainty associated with very rare return levels.

7.2.1 Data

Average hourly wind speed observations are available at a different set of sites in the vicinity of Mülheim-Kärlich NPP than for the previous analyses (some observation gauges are

identical). Again, the available data are quality controlled measurements and observations derived from Deutscher Wetterdienst (DWD) stations and legally and qualitatively equivalent partner stations operated for climatological and climate related applications. Figure 37 shows the locations of different sites where wind speed observation data are available and Figure 38 specifies the altitudes of each gauge that varies from 62 m ASL at Bonn-Friesdorf to 627 m ASL at Nurberg. As a reminder, the altitude of the NPP site under consideration is approximately 65 m ASL.

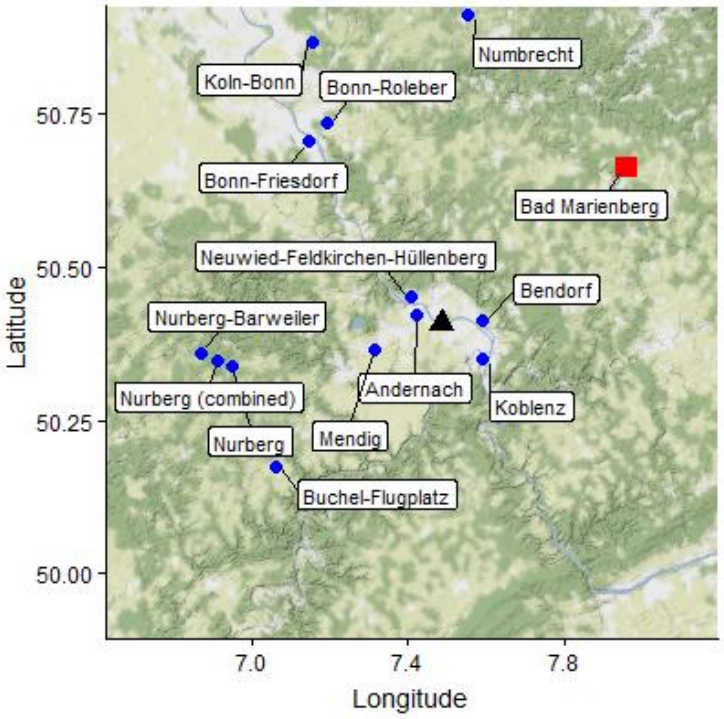


Figure 37: Location of weather gauges with wind data series (blue dots), representative site used for most of the study (red square) and Mülheim-Kärlich NPP (black triangle).

Table 14: Gauge information for the different weather gauges used for the wind speed analysis.

Gauge name	Gauge ID	Altitude (m)	Latitude	Longitude
Andernach	00161	75	50.4237	7.4202
Bendorf	00348	127	50.4135	7.5886
Bonn-Friesdorf	00599	62	50.7055	7.1467
Bonn-Roleber	00603	159	50.7349	7.1931
Buchel-Flugplatz	00766	477	50.1746	7.0595
Koblenz	02656	96	50.3519	7.5906
Köln-Bonn	02667	92	50.8646	7.1575
Bad Marienberg	03167	547	50.6620	7.9602
Mendig	03246	181	50.3667	7.3167
Nümbrecht auf dem Lindchen	03657	341	50.9081	7.5543
Nurberg	03659	627	50.3391	6.9501
Nurberg-Barweiler	03660	485	50.3601	6.8697
Neuwied-Feldkirchen-Hüllenberg	15044	199	50.4517	7.4058

Figure 38 shows the available data for weather gauges recording average wind speed measurements in the vicinity of Mülheim-Kärlich NPP. There seem to three sites recording data across a long period from 1960s to the present day; these are Bad Marienberg, Köln-Bonn and Buchel-Flugplatz. All sites are a reasonably similar distance away from the NPP site

with Köln-Bonn being closer in terms of altitude. However, further analysis of the series seems to suggest that the gauge at Bad Marienberg has good quality (from 1981 until the present day) and is chosen here.

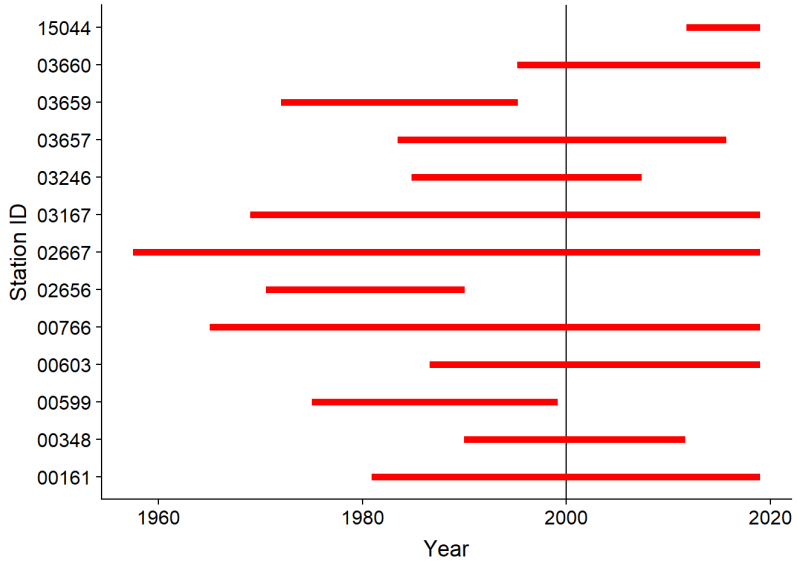


Figure 38: Time period over which wind speed data are available for each site; the plot only shows the start and end date of each of the series, there may be additional missing gaps within the series which are checked later in the analysis.

For this analysis, we extract the maximum observed average hourly wind speed within a day to provide a daily series which shall be analysed; the data cover all seasons and are shown in Figure 39.

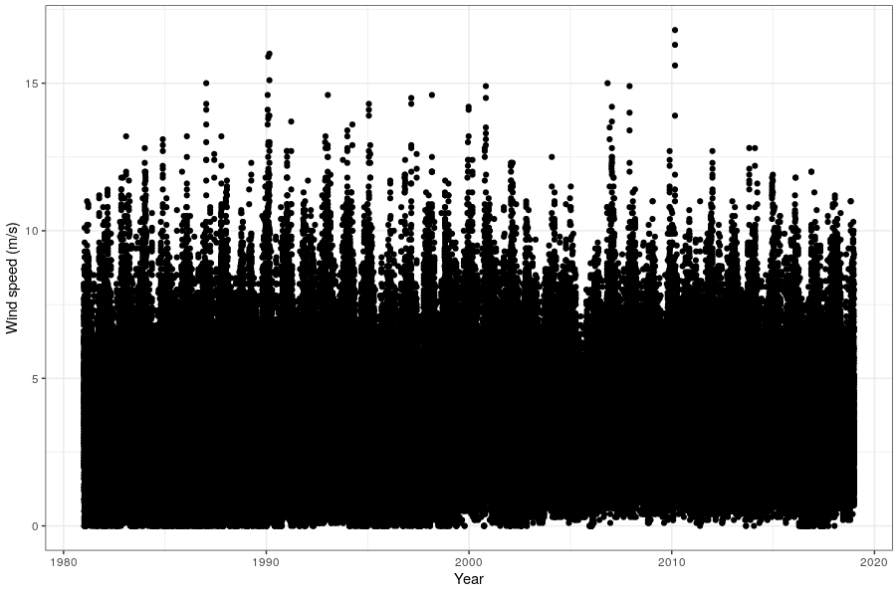


Figure 39: Average hourly wind speed time series at the Bad Marienberg weather gauge.

Data quality issues were detected in the data prior to 1981 and as such these have been removed from the dataset to be analysed.

7.2.2 Application of EVA with declustering

A threshold of 8.7 m/s (the 99th quantile of the data) is taken for our EVA using the standard threshold diagnostics. Runs declustering has been undertaken to extract out the peaks over the threshold (POTs).

Table 15 shows the GPD parameter estimates for different choices of run length parameter when using runs declustering. It is observed that there is a gradual increase in the scale parameter and a reduction in the shape parameter as the run length is increased. As expected, the number of POTs available for the model fit is reduced as the run length is increased, however this doesn't seem to have a large effect on the standard errors of the scale and shape parameter estimates. An important factor to be considered is that the shape parameter estimates are different sign when using different choices on the run length; this is likely to lead to very different extrapolations (especially at very rare levels such as the 10,000-year return level).

Table 15: Best estimates and standard errors (in parentheses) for the GPD model parameters applied to average hourly wind speeds from the Bad Marienberg weather gauge. Different run length choices have been tested to assess the sensitivity to this parameter and the induced differences in the number of POTs.

Run length	Number of POTs	Scale	Shape
1	1079	1.03 (0.05)	0.04 (0.03)
2	925	1.15 (0.05)	0.01 (0.03)
3	846	1.20 (0.06)	0.00 (0.03)
4	813	1.27 (0.06)	-0.02 (0.03)
5	791	1.29 (0.06)	-0.03 (0.03)
6	770	1.32 (0.07)	-0.04 (0.03)
7	758	1.34 (0.07)	-0.04 (0.03)
8	746	1.35 (0.07)	-0.04 (0.03)
9	732	1.38 (0.07)	-0.05 (0.03)
10	719	1.39 (0.07)	-0.05 (0.03)

Return level estimates for average hourly wind speed are given in Table 16. Different choices of the run length parameter have been chosen to highlight the sensitivity due to runs declustering. The estimates at lower return levels do not differ much, but for rarer events, the run length parameter does have a large impact on results; this is due to the change in the sign of the shape parameter which affects the extrapolation more at rarer levels. It is clear that an approach that doesn't require a subjective choice of declustering parameters would be beneficial and is explored in further detail in section 0.

Table 16: Return level estimates (with a 70% confidence interval in parentheses obtained via the delta method) with various choices of the run length parameter for runs declustering.

Run length	Return level (m/s)			
	10	**100**	**1,000**	**10,000**
1	15.2 (14.8, 15.7)	18.4 (17.3, 19.5)	21.9 (19.9, 23.9)	25.7 (22.4, 29.1)
2	15.2 (14.7, 15.6)	17.9 (16.9, 18.9)	20.7 (19.0, 22.4)	23.6 (20.9, 26.3)
3	15.1 (14.7, 15.7)	18.0 (17.0, 19.0)	20.7 (19.0, 22.5)	23.5 (20.7, 26.3)
4	15.1 (14.7, 15.5)	17.6 (16.7, 18.5)	20.0 (18.4, 21.5)	22.2 (19.9, 24.5)
5	15.1 (14.7, 15.5)	17.5 (16.7, 18.4)	19.8 (18.3, 21.3)	22.0 (19.7, 24.2)
6	15.1 (14.7, 15.5)	17.5 (16.6, 18.3)	19.7 (18.2, 21.1)	21.7 (19.5, 23.8)
7	15.1 (14.7, 15.5)	17.4 (16.6, 18.3)	19.6 (18.2, 21.0)	21.5 (19.4, 23.7)
8	15.1 (14.7, 15.5)	17.4 (16.6, 18.2)	19.5 (18.1, 20.9)	21.4 (19.3, 23.5)
9	15.0 (14.6, 15.5)	17.3 (16.5, 18.1)	19.3 (18.0, 20.7)	21.1 (19.2, 23.1)
10	15.1 (14.7, 15.5)	17.3 (16.5, 18.2)	19.3 (18.0, 20.7)	21.1 (19.1, 23.1)

7.2.3 Application of EVA taking advantage of temporal dependence

We now use the theory provided in section 4.6 to take advantage of the clustering behaviour in the data to avoid the need for declustering and see if we can reduce the uncertainty associated with extreme return level estimates. For this application, we undertake the inference using Bayesian statistics as opposed to the frequentist approaches used in previous sections. We take the Uniform distribution as our prior (chosen to be an uninformative prior) and fit the GPD to all exceedances of the threshold (kept as the 99th quantile at 8.7 m/s). Markov Chain Monte Carlo (MCMC) is used to sample from the posterior distribution; specifically, a Metropolis-Hastings algorithm is set up to do this. In total, 20,000 samples from the posterior distributions of the scale and shape parameters are created, with the first 5,000 of these discarded as burn-in. Figure 40 shows the MCMC trace plots for the scale and shape parameters for a GPD fitted to all the exceedances of the threshold.

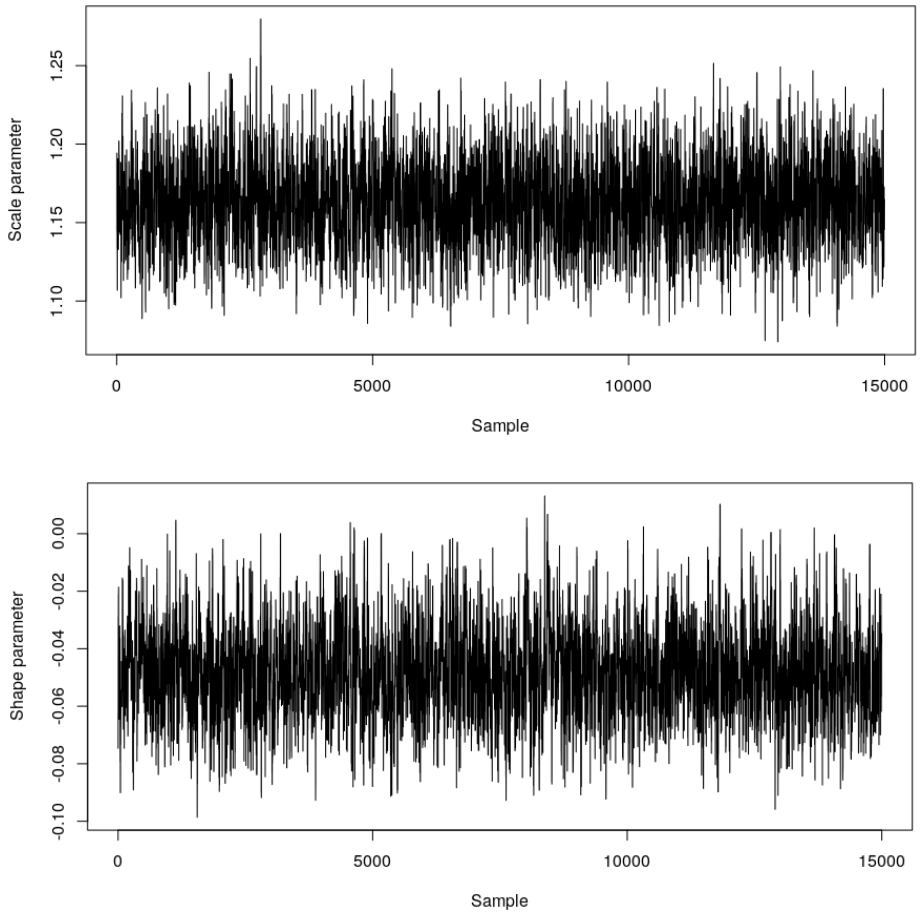


Figure 40: MCMC trace plots for the GPD scale and shape parameters from the fit to average hourly wind speed values at the Bad-Marienberg weather gauge.

Table 17 shows estimates of the posterior mean and standard deviation taken from the Bayesian fit to all the exceedances. The estimates are quite similar to those provided in

Table 15, however caution should be exercised when comparing results. The posterior mean and standard deviation are proxies for the best estimate and standard error from a frequentist analysis but not exactly equivalent.

Table 17: Estimates of the mean and standard deviation of the posterior distribution for the scale and shape parameter of the GPD fit. NB: the posterior mean and standard deviation are proxies for the best estimate and standard error from a frequentist analysis but not exactly equivalent.

	Scale	Shape
Posterior mean	1.16	-0.05
Posterior standard deviation	0.03	0.02

It is now possible to estimate posterior return levels at a variety of return periods and compare these with the values from Table 16. As noted in section 4.6, we need to use the updated version of the return level estimation formula to account for the clustering within the dataset. This requires an approach for estimating the extremal index; this is explored in more detail in

section 7.2.4 using multivariate extreme value analysis. Initially, we pick a variety of extremal index values between 0 and 1 to test the sensitivity to this value. Table 18 shows estimates of various return levels for different potential extremal index values (ranging from strong clustering $\theta = 0.01$ to no clustering $\theta = 1$). Broadly, the estimates from the posterior distribution are lower than those provided in Table 16 at all return levels; uncertainty estimates have also been reduced slightly. It is clear that the estimate of the extremal index can have a large impact on results which makes it clear that the estimation of this quantity is going to be important (see section 7.2.4).

Table 18: Return level estimates (with a 70% posterior credible interval in parentheses) with various potential estimates of the extremal index.

Extremal index	Return level (m/s)			
	10	**100**	**1,000**	**10,000**
0.01	11.1 (11.1, 11.1)	13.4 (13.3, 13.5)	15.4 (15.2, 15.7)	17.3 (16.8, 17.7)
0.5	14.8 (14.6, 15.1)	16.7 (16.3, 17.1)	18.5 (17.8, 19.1)	20.0 (19.1, 20.9)
1	15.4 (15.2, 15.7)	17.3 (16.8, 17.7)	18.9 (18.2, 19.7)	20.4 (19.4, 21.4)

Within the Bayesian context it is also possible to estimate the predictive return level, which is a single estimate of a return level with all uncertainty about parameters and potential future data rolled into the estimation. Table 19 shows predictive return level estimates for various return periods and fixed values of the extremal index. As described above, a single number is provided as we have integrated across the uncertainty in the model parameters (since we have fitted the model using a Bayesian approach and the parameters are uncertain and not assumed to be fixed). Generally, estimates are slightly higher compared to Table 18 but are quite consistent with the posterior means.

Table 19: Predictive return level estimates with various potential estimates of the extremal index.

Extremal index	Return level (m/s)			
	10	**100**	**1,000**	**10,000**
0.01	13.4	13.4	15.4	17.4
0.5	14.8	16.8	18.6	20.4
1	15.4	17.4	19.2	21.0

Figure 41 shows the comparison between the posterior and predictive estimates of the 10,000-year return level for a variety of different values of the extremal index. The result is similar to the one provided in Fawcett and Walshaw (2012) with the return level estimates increasing with an increasing extremal index (i.e. less clustering in the data). As observed in the comparison between Table 18 and Table 19, the predictive return level is slightly higher that the posterior means at all values of the extremal index. It should be noted that all estimates using this approach are lower and have smaller uncertainty than when fitting to the POTs using declustering with any value of the run length.

However, one drawback of the estimates provided above is that we have assumed that the extremal index is fixed at a set of specific levels. Ideally we would like to estimate the extremal index using a model (as opposed to fixing it at a specific value) which permits the extremal index to vary with the return level we wish to estimate; this is investigated further in section 7.2.4.

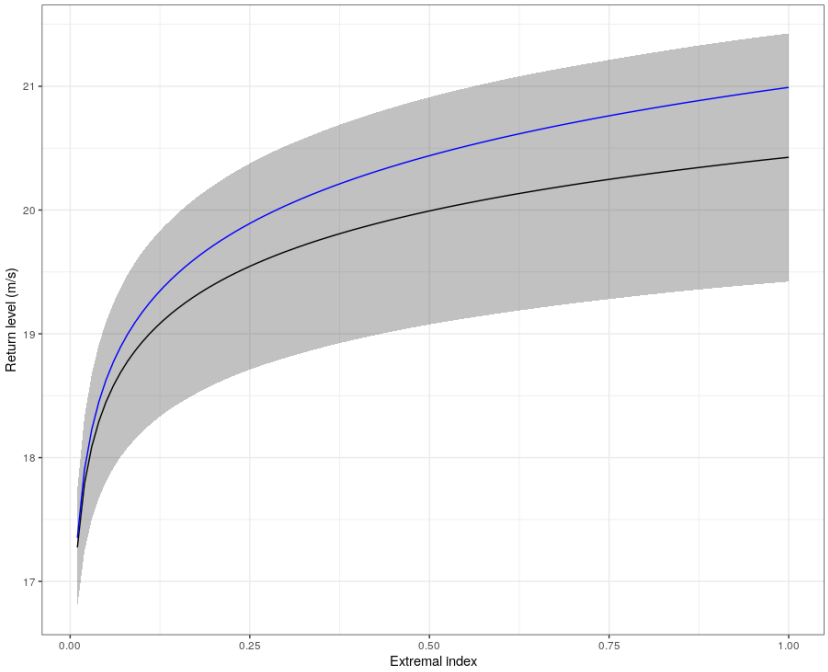


Figure 41: Estimates of the 10,000-year return level for different values of the extremal index, including: posterior mean values (black line); posterior 70% credible intervals (grey shaded area); predictive values (blue line).

7.2.4 Estimating the extremal index

To estimate the extremal index, it is necessary to build a multivariate extreme value model for consecutive extreme hourly wind speeds (as was previously done for extreme air temperature in section 7.1). This multivariate EVA model can be used to create replicate simulated heatwave events which can be used to estimate the duration distribution and thus the extremal index.

Autocorrelation and partial autocorrelation function plots are provided in Figure 42 and are used to determine the order of the Markov process used to model the extremes. The PACF seems to suggest that there is strong dependence at lag 1 with a smaller amount of dependence out to lag 3. Beyond that point, there is some dependence, but it is quite small and creating a more complicated model to capture this is unlikely to be beneficial. To start with, we use a first-order Markov process to capture the temporal dependence in the extremes. Figure 43 shows a scatter plot of average wind speeds in consecutive hours; as is expected they are strongly dependent, with a high level of dependence in the extremes also observed.

An assessment of the dependence within the extremes can be made using the empirical estimator to the extremal dependence measures $\bar{\chi}(u)$ and $\chi(u)$; see Figure 44. These plots suggest that the data could either be asymptotically dependent or asymptotically independent with a strong amount of positive dependence within the range of the data. Since it is not clear from the empirical estimator whether the data are asymptotically dependent or asymptotically independent, it makes sense to use the conditional extremes model again which can account for both types of asymptotic behaviour.

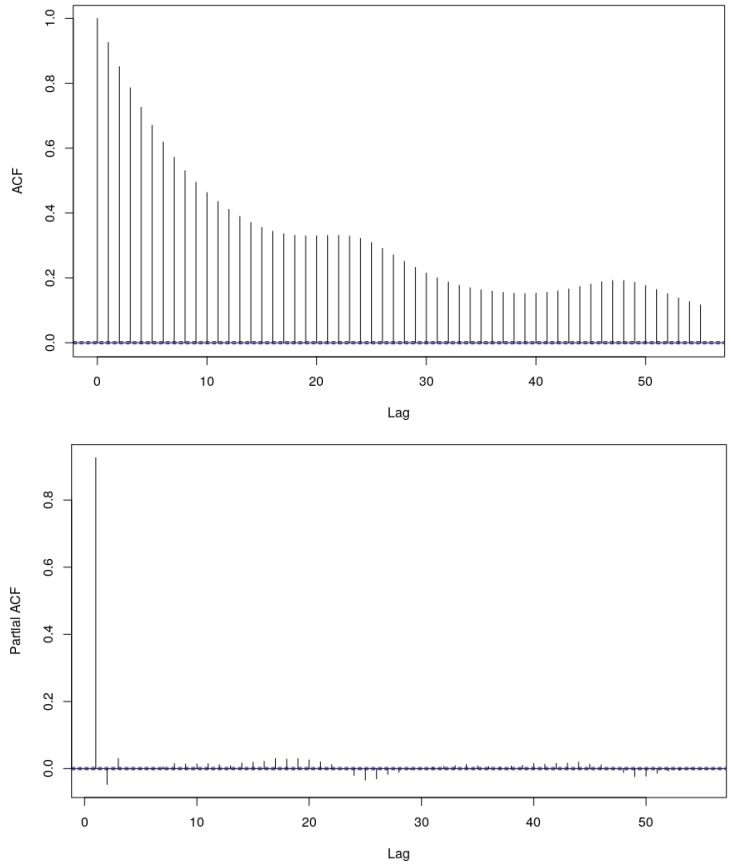


Figure 42: ACF (above) and PACF (below) plots for the hourly wind speed data at the Bad Marienberg weather gauge.

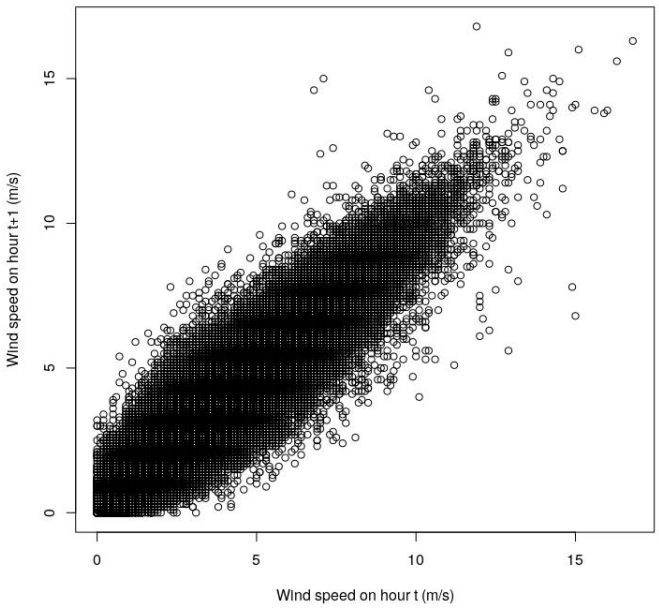


Figure 43: Scatter of average hourly wind speeds (m/s) in hour t and hour t+1.

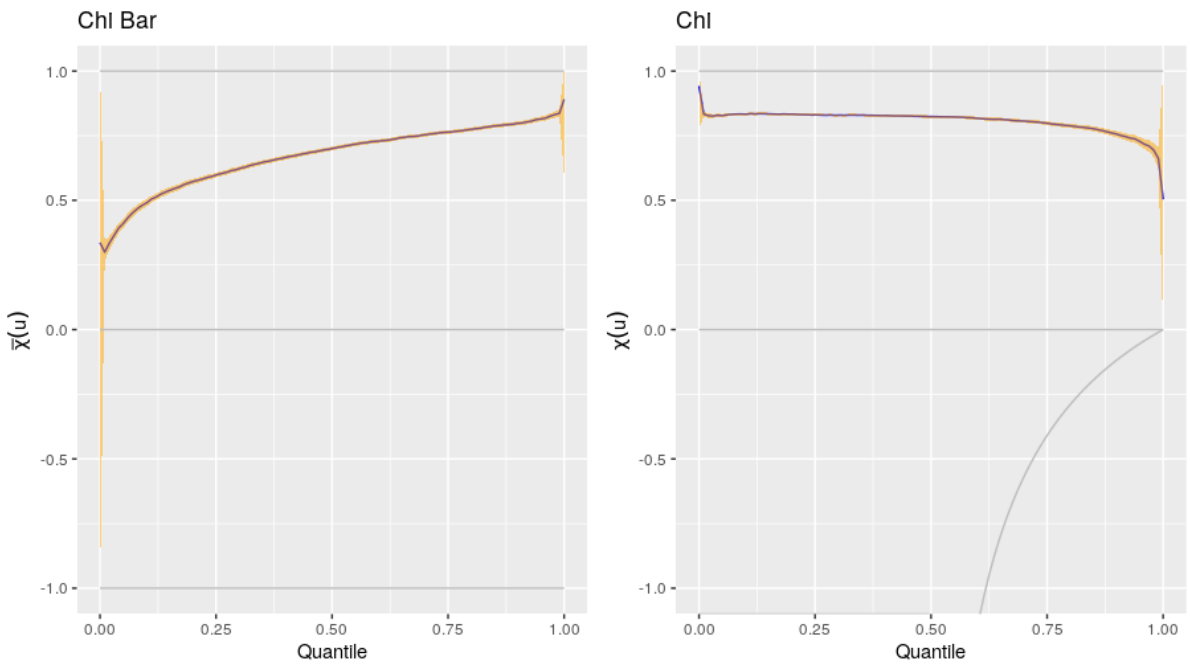


Figure 44: Empirical estimates of $\bar{\chi}(u)$ (left) and $\chi(u)$ (right) for a set of thresholds u (or equivalent quantiles).

A marginal probability integral transform is undertaken using the GPD fitted to all exceedances and then the conditional extremes model is fitted to the data on unit margins using a quantile of 0.99 (to be equivalent to the marginal model). The model parameter estimates are given as $\alpha = 0.67$ and $\beta = 0.88$.

The multivariate model is now used to simulate replicate wind events and estimate the extremal index. This is undertaken at a variety of different critical levels that are linked to return periods. In this way, it will be possible to account for changes in the extremal index with the return period within our final estimates of the posterior and predictive return level estimates (unlike the results of section 0 which assumed a fixed value of the extremal index).

Table 20 provides estimates of the subasymptotic extremal index $\theta(v)$ for various critical levels v which are associated to specific return periods. We observe that the data are less clustered at more extreme levels.

Table 20: Estimates of the subasymptotic extremal index $\theta(v)$ for various critical levels v which are associated with specific return periods of interest.

	Critical level associated return period			
	10	**100**	**1,000**	**10,000**
Extremal index	0.45	0.50	0.54	0.57

Table 21 provides a summary of the different return level estimates obtained using the different approaches outlined in section 7.2. The posterior mean and predictive estimates are obtained using the estimates of the extremal index derived from the multivariate model (Table 20). The estimates using the POTs identified using declustering are taken directly from Table 16 and have been provided for a couple of different run lengths ($r = 1$ and $r = 7$). As observed in previous results, the estimates from the posterior mean and predictive are comparable (with

the predictive very slightly higher). The estimates obtained through a standard fit to declustered data are all higher (irrespective of the run length) and have a higher uncertainty associated to the results. Caution is advised as the 70% confidence intervals from a frequentist fit and the 70% credible intervals are not directly the same quantity, but they do broadly provide a summary of the same type of quantity. A distinct benefit of the Bayesian approach outlined is that we have avoided the need to pick a declustering parameter and the subjectivity that this introduced.

Table 21: Estimates of various return levels for average hourly wind speed (m/s) using different approaches: (i) posterior mean and 70% credible interval in parentheses using modelled extremal index; (ii) predictive estimate using modelled extremal index; (iii) direct estimate on POTs using declustering with run lengths of 1 and 7 respectively with 70% confidence intervals in parentheses obtained via the delta method.

	Return level (m/s)			
	10	**100**	**1,000**	**10,000**
Posterior mean	14.7 (14.5, 15.0)	16.7 (16.3, 17.1)	18.5 (17.9, 19.2)	20.1 (19.1, 21.0)
Predictive	14.7	16.8	18.7	20.5
Declustering $r = 1$	15.2 (14.8, 15.7)	18.4 (17.3, 19.5)	21.9 (19.9, 23.9)	25.7 (22.4, 29.1)
Declustering $r = 7$	15.1 (14.7, 15.5)	17.4 (16.6, 18.3)	19.6 (18.2, 21.0)	21.5 (19.4, 23.7)

Figure 45 provides a set of hazard curves for extreme hourly average wind speeds based upon the different models presented in Table 21. It is observed again that the fit to the POTs provides much higher return level estimates with uncertainty bounds that are a lot wider than for the methods based upon Bayesian statistics. The posterior return level estimates are the lowest with the predictive value closer to the posterior estimates for less rare return periods (e.g. 1 and 10) compared to the very rare return periods (e.g. 10^6 and 10^7) where the estimate is pushed higher by the inherent uncertainty when trying to estimate out to such extreme levels.

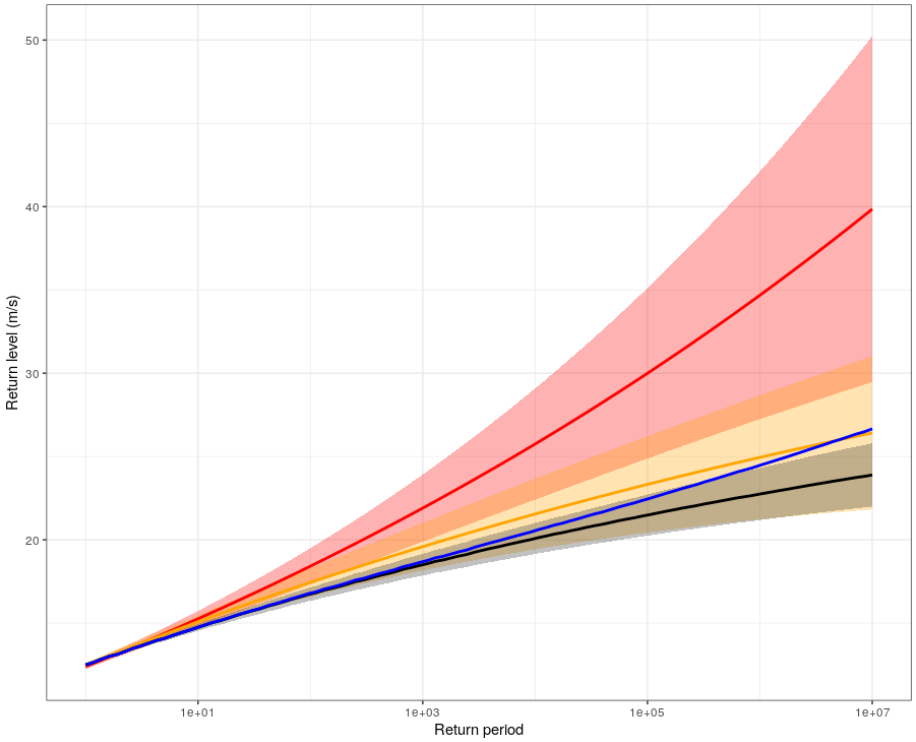


Figure 45: Hazard curves for hourly average wind speed, estimated using a variety of different approaches: maximum likelihood fit to POTs identified using runs declustering with $r = 1$ (red) and $r = 7$ (orange); posterior distribution (black) and predictive estimate (blue) using Bayesian approach. Shaded regions are 70% confidence/credible intervals (depending on approach used).

7.2.5 Summary of analysis

In this case study, we have used the output of a multivariate extreme value model to improve upon standard univariate extreme value analysis techniques that require declustering to be undertaken to extract independent peaks over the threshold (POTs). There are several clear benefits to this approach:

- We utilize more of the available information from the data by not having to discard exceedances which aren't peaks but are still extreme.
- Since we don't have to decluster the data, there is no need to pick a declustering parameter (e.g. the run length), which is a subjective choice.
- Fitting the model in the Bayesian framework permits the estimation of the predictive return level which rolls all uncertainty about the parameters and randomness in future observations into a single estimate.

The analysis of hourly average wind speeds has highlighted a high level of dependence between extremes in consecutive hours which has been modelling using the conditional extremes model. The tail chain simulation approach has then been used to create replicate events and these have been used to estimate the extremal index at a selection of different critical levels (associated with specific return levels). The extremal index estimates are then used to permit the estimation of return levels using all the data which take account of temporal clustering (as per the approach in Fawcett and Walshaw, 2012).

The results obtained using the different approaches show some marked differences, especially when attempting to estimate return levels at very rare return periods (i.e. 10,000-year and

beyond).

The main results are summarized below:

- Standard frequentist approaches with declustering lead to higher estimates than the Bayesian posterior and predictive estimates.
- Uncertainty estimates for the fits to POTs are wider than those using the Bayesian approach.
- The subjective choice of declustering parameter (here the run length) can have a large impact on the best estimates and uncertainty intervals.

8 Conclusions and Recommendations

A number of methodologies have been presented for potential use within PSAs within Europe looking at extreme weather and flooding. When looking at multi-hazard, the duration of the events is extremely important in terms of their potential impacts when overlapping or coinciding hazards are analysed. Durations and correlation matrices have been reexamined as part of this deliverable.

We are likely to miss important information if we only estimate return levels (or equivalent metrics) focusing on the instantaneous severity of events. Several different approaches were outlined which are available to utilize information about the duration of events to improve our estimation of natural hazard risk. Our focus is on approaches from within the class of extreme value models as these are models with mathematical justification for extrapolating beyond the range of the data and are the most commonly used approaches for this type of analysis within the nuclear industry.

In terms of standard approaches for the application of climate change results to NPP design, guidelines have been put forward in various forms. PRIMAVERA-H2020 project is the latest EU project to provide high resolution analyses on a smaller scale appropriate to NPP site levels. This downscaling and higher resolution allows for a better application of possible effects of climate change. As the PRIMAVERA models and other downscaled models become available at a reasonable resolution, they should be integrated in conjunction with the usual station data which is usually collected at the site, and from long-run stations around the site.

At present structural design codes are based on the assumption of stationary climate conditions. The reliability targets are typically specified, or understood, in terms of the annual probability of failure over the design life of the structure, typically 50 to 75 years. In the reliability-based design framework, the nominal design loads are specified in terms of return period or an upper percentile, e.g., 95^{th} percentile, of the annual maximum load distribution. Further, load factors are specified to calculate the factored design load. The load factors are calibrated with respect to the target reliability level (i.e., annual probability of failure).

As the climate change effects are becoming evident, national code authorities are interested in evaluating the effect of non-stationary climate conditions on the structural design provisions. More specifically, it is being recognized that the reliability-based calibration approaches used under the traditional assumption of stationary climate condition cannot be directly extended to non-stationary climate cases. For example, the concept of the return period is inapplicable to non-stationary climate, as the distribution of the time between occurrences of the load events is no longer invariant. Similarly, the annual probability of failure is no longer constant in the non-stationary climate. A technical discussion was developed and such issues were addressed using the theory of stochastic processes.

The assessment of external hazard, as flooding or earthquake, often relies on numerical models which allow assessing the variables of interest (*e.g.* water depths, ground acceleration, *etc.*). However, these models are affected by uncertainties which can be quantified through studies called “Uncertainty Quantification” (UQ) and “Global Sensitivity Analysis” (GSA). The UQ attempts to describe the whole set of possible outputs considering the inputs not perfectly known. The GSA aims to determine the most influent inputs to an output behaviour, as the non-influential ones. These two types of analyses are complementary and both classically suppose that the input parameters of the numerical models are independent. Especially, in the literature, most studies consider the inputs independent which may not be the case. A global review of the different methods used for UQ and GSA when model inputs are considered dependent was undertaken and was applied to hydraulic studies (Table 2). In the field of

flooding hazard, this kind of analysis is still missing. For the interested reader, an IRSN report on existing methodologies for UQ and GSA.

GSA with dependent inputs followed by an application to a simplified case of inundation has been written on this topic (Pheulpin, 2020). The major points from this report were reported in the following paragraphs (5.1 and 5.2). In addition a research article on the application of these methods to the large hydraulic model of the Loire River is in progress. Some of the results have been presented in the EGU (Pheulpin and Bacchi, 2020).

The main conclusions of the work showed that:

- There is a strong dependence between the hydrograph parameters (d, qmax, tm) and this dependency can be taken into account by using copulas.
- The use of metamodels is very useful for uncertainty analysis studies (almost all with limited computation resources).
- There is a limited impact of inputs dependency in uncertainty quantification in this study.
- The duration and time to peak inputs have strong influence on the outputs.
- The hydrograph shape should not be ignored in hydraulic studies.

It is important to note that this deliverable focusses on only a few of the key issues on the list of ASAMPSA_E indicating various issues associated with extreme weather analyses, and this field is continuing to evolve and requires more work in the coming years to address the other issues:

- Limitations in modelling and forecasting the physical phenomena and conditions leading to extreme hazard;
- Uncertainties in estimation of the impact of climate change on extreme meteorological events;
- Lack of site-specific data and limitations of spatial modelling and downscaling methods;
- Difficulties in quantification of uncertainties for common-cause failures;
- Difficulties in integrated modelling of hazard internal and external impact assessment;
- Limitation in determining the occurrence frequency of extreme weather conditions;
- Correlation among an extreme weather event induced failure modes and on the quantification of correlation coefficients.

The use of an output of a multivariate extreme value model to improve upon standard univariate extreme value analysis techniques that require declustering to be undertaken to extract independent peaks over the threshold (POTs) was presented with various benefits to this approach:

- We utilize more of the available information from the data by not having to discard exceedances which aren’t peaks but are still extreme.
- Since we don’t have to decluster the data, there is no need to pick a declustering parameter (e.g. the run length), which is a subjective choice.
- Fitting the model in the Bayesian framework permits the estimation of the predictive return level which rolls all uncertainty about the parameters and randomness in future observations into a single estimate.
- Standard frequentist approaches with declustering lead to higher estimates than the Bayesian posterior and predictive estimates.
- Uncertainty estimates for the fits to POTs are wider than those using the Bayesian approach.
- The subjective choice of declustering parameter (here the run length) can have a large impact on the best estimates and uncertainty intervals.

The complexity of extreme weather and flood modelling as well as the non-stationary nature and lack of data makes analysis difficult with major uncertainties. Some of the tools and methods presented within this report can address some of these issues, however the reader is directed to the key documents section at the start of this report.

9 References

Alexander, L.V. et al., “Global observed changes in daily climate extremes of temperature and precipitation,” J. Geophys. Res. Atmos., vol. 111, pp. 1–22, 2006.

Black, E., M. Blackburn, G. Harrison, B. Hoskins, and J. Methven, “Factors contributing to the summer 2003 European heatwave,” Weather, vol. 59, no. 8, pp. 217–223, 2004, doi: 10.1256/wea.74.04.

Blake, E. and D. Zelinsky, “National Hurricane Center Tropical Cyclone Report: Hurricane Harvey (AL092017),” 2018.

Borgonovo, E., 2007. A new uncertainty importance measure. Reliability Engineering & System Safety 92, 771–784. https://doi.org/10.1016/j.ress.2006.04.015

Borgonovo, E., Castaings, W., Tarantola, S., 2011. Moment Independent Importance Measures: New Results and Analytical Test Cases: Moment Independent Importance Measures. Risk Analysis 31, 404–428. https://doi.org/10.1111/j.1539-6924.2010.01519.x

Caniou, Y., 2012. Global sensitivity analysis for nested and multiscale modelling. Université Blaise Pascal - Clermont II.

Chastaing, G., Gamboa, F., Prieur, C., 2012. Generalized Hoeffding-Sobol decomposition for dependent variables - application to sensitivity analysis. Electronic Journal of Statistics 6, 2420–2448. https://doi.org/10.1214/12-EJS749

Chun, M.-H., Han, S.-J., Tak, N.-I., 2000. An uncertainty importance measure using a distance metric for the change in a cumulative distribution function. Reliability Engineering & System Safety 70, 313–321. https://doi.org/10.1016/S0951-8320(00)00068-5

Coles, S.G., J. E. Heffernan, and J. A. Tawn, “Dependence measures for extreme value analyses,” Extremes, vol. 2, no. 4, pp. 339–365, 1999.

Cressie, N.A.C., Statistics for Spatial Data. Wiley, New York, 1993.

D50.16 Report 2 – Guidance document on practices to model and implement FLOODING hazards in extended PSA (final version)

D50.17 Report 3 – Guidance document on practices to model and implement EXTREME WEATHER hazards in extended PSA (final version)

D50.18 Report 4 – Guidance document on practices to model and implement LIGHTNING hazards in extended PSA (final version)

Da Veiga, S., Wahl, F., Gamboa, F., 2009. Local Polynomial Estimation for Sensitivity Analysis on Models With Correlated Inputs. Technometrics 51, 452–463. https://doi.org/10.1198/TECH.2009.08124

Dupuis, D.J. “Modeling waves of extreme temperature: the changing tails of four cities,” J. Am. Stat. Assoc., vol. 107, no. 497, pp. 24–39, 2012.

Fang, S., Gertner, G.Z., Anderson, A.A., 2004. Estimation of sensitivity coefficients of nonlinear model input parameters which have a multinormal distribution. Computer Physics Communications 157, 9–16. https://doi.org/10.1016/S0010-4655(03)00488-0

Fawcett, L. and D. Walshaw, “Estimating return levels from serially dependent extremes,” Environmetrics, vol. 23, no. 3, pp. 272–283, May 2012.

Gill, J.C. and B. D. Malamud, “Reviewing and visualizing the interactions of natural hazards,” Rev. Geophys., vol. 52, pp. 680–722, 2014, doi: 10.1002/2013RG000445.

Gouldby, B. et al., “Multivariate extreme value modelling of sea conditions around the coast of England,” Proc. Inst. Civ. Eng. Marit. Eng., vol. 170, no. 1, pp. 3–20, 2017.

Grandjacques, M., Delinchant, B., Adrot, O., 2015. Pick and Freeze estimation of sensitivity index for static and dynamic models with dependent inputs 27.

Heffernan, J.E. and J. A. Tawn, “A conditional approach for multivariate extreme values (with discussion),” J. R. Stat. Soc. Ser. B, vol. 66, no. 3, pp. 497–546, 2004.

IAEA Safety Standards Series SSG-18, Meteorological and hydrological hazards in site evaluation for nuclear installations, Vienna, 2011.

IAEA, 2003. Extreme External Events in the Design and Assessment of Nuclear Power Plants. Tecdoc 1341, 109pp.

IAEA, 2003. Flood Hazard for Nuclear Power Plants on Coastal and River Sites. Safety Guide No. NS-G-3.5, Vienna.

IAEA, 2003. Flood Hazard for Nuclear Power Plants on Coastal and River Sites. Safety Guide NS-G-3.5, 83pp.

IAEA, 2003. Meteorological Events in Site Evaluation for Nuclear Power Plants. Safety Guide NS-G-3.4, 34pp.

IAEA, 2011. Meteorological and Hydrological Hazards in Site Evaluation for Nuclear Installations Specific Safety Guide SSG-18, 146pp.

Iman, R.L., Hora, S.C., 1990. A Robust Measure of Uncertainty Importance for Use in Fault Tree System Analysis. Risk Analysis 10, 401–406. https://doi.org/10.1111/j.1539-6924.1990.tb00523.x

Iooss, B., Prieur, C., 2018. Shapley effects for sensitivity analysis with dependent inputs: comparisons with Sobol’ indices, numerical estimation and applications 39.

Jacques, J., Lavergne, C., Devictor, N., 2006. Sensitivity analysis in presence of model uncertainty and correlated inputs. Reliability Engineering & System Safety 91, 1126–1134. https://doi.org/10.1016/j.ress.2005.11.047

Joe, H., Kurowicka, D., 2011. Dependence modeling: vine copula handbook. World Scientific.

Joe, H., Multivariate models and dependence concepts. Chapman and Hall/CRC, 1997.

Jonathan, P., K. Ewans, and D. Randell, “Joint modelling of extreme ocean environments incorporating covariate effects,” Coast. Eng., vol. 79, pp. 22–31, 2013.

Keef, C., I. Papastathopoulos, and J. A. Tawn, “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,” J. Multivar. Anal., vol. 115, pp. 396–404, 2013.

Kucherenko, S., Tarantola, S., Annoni, P., 2012. Estimation of global sensitivity indices for models with dependent variables. Computer Physics Communications 183, 937–946. https://doi.org/10.1016/j.cpc.2011.12.020

Kuzmina I., Lyubarskiy A., Hughes P., Kluegel J., Kozlik T., Serebrjakov V., The Fault Sequence Analysis Method to Assist in Evaluation of the Impact of Extreme Events on NPPs (Proceedings of the Nordic PSA Conference – Castle Meeting 2013, 10-12 April 2013, Stockholm, Sweden).

Lamb, R. et al., “A new method to assess the risk of local and widespread flooding on rivers and coasts,” J. Flood Risk Manag., vol. 3, no. 4, pp. 323–336, 2010.

Ledford, A.W. and J. A. Tawn, “Modelling dependence within joint tail regions,” J. R. Stat. Soc. Ser. B, vol. 59, no. 2, pp. 475–499, 1997.

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

Li, C., Mahadevan, S., 2016. An efficient modularized sample-based method to estimate the first-order Sobol′ index. Reliability Engineering & System Safety 153, 110–121. https://doi.org/10.1016/j.ress.2016.04.012

Li, G., Rabitz, H., Yelvington, P.E., Oluwole, O.O., Bacon, F., Kolb, C.E., Schoendorf, J., 2010. Global Sensitivity Analysis for Systems with Independent and/or Correlated Inputs. The Journal of Physical Chemistry A 114, 6022–6032. https://doi.org/10.1021/jp9096919

Li, L., Lu, Z., Zhou, C., 2011. Importance analysis for models with correlated input variables by the state dependent parameters method. Computers & Mathematics with Applications 62, 4547–4556. https://doi.org/10.1016/j.camwa.2011.10.034

López-Benito, A., Bolado-Lavín, R., 2017. A case study on global sensitivity analysis with dependent inputs: The natural gas transmission model. Reliability Engineering & System Safety 165, 11–21. https://doi.org/10.1016/j.ress.2017.03.019

Maity, R., Nagesh Kumar, D., 2008. Probabilistic prediction of hydroclimatic variables with nonparametric quantification of uncertainty. Journal of Geophysical Research 113. https://doi.org/10.1029/2008JD009856

Mara, T.A., Tarantola, S., 2012. Variance-based sensitivity indices for models with dependent inputs. Reliability Engineering & System Safety 107, 115–121. https://doi.org/10.1016/j.ress.2011.08.008

McKay, M.D., 1995. Evaluating prediction uncertainty. Nuclear Regulatory Commission.

McKay, M.D., 1996. Variance-based methods for assessing uncertainty importance.

Meteo France, “Xynthia, 10 years later,” 2020. [Online]. Available: http://www.meteofrance.fr/actualites/79755493-xynthia-10-ans-apres.

Michailidi, E.M., Bacchi, B., 2017. Dealing with uncertainty in the probability of overtopping of a flood mitigation dam. Hydrology and Earth System Sciences 21, 2497–2507. https://doi.org/10.5194/hess-21-2497-2017

Nelsen, R.B., 2007. An introduction to copulas. Springer Science & Business Media.

Oakley, J.E., O’Hagan, A., 2004. Probabilistic sensitivity analysis of complex models: a Bayesian approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66, 751–769. https://doi.org/10.1111/j.1467-9868.2004.05304.x

Owen, A.B., Prieur, C., 2016. On Shapley value for measuring importance of dependent inputs. arXiv:1610.02080 [math, stat].

Pan, F., Zhu, J., Ye, M., Pachepsky, Y.A., Wu, Y.-S., 2011. Sensitivity analysis of unsaturated flow and contaminant transport with correlated parameters. Journal of Hydrology 397, 238–249. https://doi.org/10.1016/j.jhydrol.2010.11.045

Pheulpin, L., 2020. Methodologies on uncertainty quantification and global sensitivity analysis with dependent inputs: Review and application to a simplified case of inundation (No. IRSN/2020-00321). IRSN/PSE-ENV/SCAN/BEHRIG.

Pheulpin, L., Bacchi, V., 2020. Uncertainty quantification and global sensitivity analysis with dependent inputs: Application to the 2D hydraulic model of the Loire River. Presented at the EGU 2020-18939.

Ratto, M., Tarantola, S., Saltelli, A., Young, P.C., 2005. Accelerated estimation of sensitivity indices using state dependent parameter models 10.

Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. Sensitivity analysis in practice: a guide to assessing scientific models. John Wiley & Sons.

Sharkey, P. and H. C. Winter, "A Bayesian spatial hierarchical model for extreme precipitation in Great Britain," Environmetrics, vol. 30, no. 1, pp. 1–19, 2019.

Sraj, M., Bezak, N., Brilly, M., 2015. Bivariate flood frequency analysis using the copula function: a case study of the Litija station on the Sava River: COPULA FLOOD FREQUENCY ANALYSIS FOR THE LITIJA STATION IN SLOVENIA. Hydrological Processes 29, 225–238. https://doi.org/10.1002/hyp.10145

Ţene, M., Stuparu, D.E., Kurowicka, D., El Serafy, G.Y., 2018. A copula-based sensitivity analysis method and its application to a North Sea sediment transport model. Environmental Modelling & Software 104, 1–12. https://doi.org/10.1016/j.envsoft.2018.03.002

Tilloy, A., B. D. Malamud, H. C. Winter, and A. Joly-Laugel, "Evaluating the efficacy of bivariate extreme modelling approaches for multi-hazard scenarios," Submitt. to NHESS, 2020.

Tilloy, A., B. D. Malamud, H. Winter, and A. Joly-laugel, "A review of quantification methodologies for multi-hazard interrelationships," Earth-Science Rev., vol. 196, no. May, p. 102881, 2019.

Trigo, R. M., R. García-Herrera, J. Díaz, I. F. Trigo, and M. A. Valente, "How exceptional was the early August 2003 heatwave in France?," Geophys. Res. Lett., vol. 32, pp. 1–4, 2005.

Wang, P., Lu, Z., Zhang, K., Xiao, S., Yue, Z., 2018. Copula-based decomposition approach for the derivative-based sensitivity of variance contributions with dependent variables. Reliability Engineering & System Safety 169, 437–450. https://doi.org/10.1016/j.ress.2017.09.012

WENRA-RHWG, 2013. Report Safety of new NPP designs. Study by Reactor Harmonization Working Group RHWG, 52pp.

Winter, H.C. and J. A. Tawn, "kth-order Markov extremal models for assessing heatwave risks," Extremes, Nov. 2017, doi: 10.1007/s10687-016-0275-z.

Winter, H.C. and J. A. Tawn, "Modelling heatwaves in central France: a case-study in extremal dependence," J. R. Stat. Soc. Ser. C (Applied Stat., vol. 65, no. 3, pp. 345–365, Apr. 2016, doi: 10.1111/rssc.12121.

Winter, H.C., S. J. Brown, and J. A. Tawn, "Characterising the changing behaviour of heatwaves with climate change," Dyn. Stat. Clim. Syst., vol. 1, no. 1, p. dzw006, Jan. 2016, doi: 10.1093/climsys/dzw006.

Xu, C., Gertner, G.Z., 2008. Uncertainty and sensitivity analysis for models with correlated parameters. Reliability Engineering & System Safety 93, 1563–1573. https://doi.org/10.1016/j.ress.2007.06.003

Zhou, C., Lu, Z., Li, L., Feng, J., Wang, B., 2013. A new algorithm for variance based importance analysis of models with correlated inputs. Applied Mathematical Modelling 37, 864–875. https://doi.org/10.1016/j.apm.2012.03.017