

Flood Risk Modeling Aided by Machine Learning Techniques

Using a Treed Gaussian Process for a Case Study in Charleston, South Carolina

L.J.R. Terlinden-Ruhl



Flood Risk Modeling Aided by Machine Learning Techniques

Using a Treed Gaussian Process for a Case Study in Charleston, South Carolina

by

Lucas Terlinden-Ruhl

in partial fulfillment of the requirements for the degree of

Master of Science

Civil Engineering

at the Delft University of Technology

to be defended publicly on Friday June 28th 2024 at 12:00 PM in CEG lecture hall F.

Supervised by

J.A.Á. Antolínez PhD, Delft University of Technology
P. Mares-Nasarre PhD, Delft University of Technology
Ir. G.G. Hendrickx, Delft University of Technology
Ir. D. Eilander PhD, Deltares
Ir. A. Couasnon PhD, Deltares

Cover: A circular ceramic called "Imaginaire" by Isabelle Nothomb, color edit by Antoine François-Poncet.

When I saw this ceramic in person, it made me think of a satellite image of a coastal system. For me, the color blue represents the origin of the flood. The color white represents the non-spatially uniform level of protection provided by flood protection measures. Other colors represent different extents of exposure.

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.

Deltares


TU Delft

Abstract

Compound floods, which can be attributed to different drivers (pluvial, fluvial, surge, tide, and waves), generate a larger flood hazard when drivers co-occur than when they occur in isolation of each other. Current compound flood risk assessments are affected by a curse of dimensionality, where a larger number of events need to be numerically simulated to understand the response of risk to drivers. This research aims to create a methodology that improves the quantification of compound flood risk by using a Treed Gaussian Process (TGP) for the case study of Charleston. A TGP can actively learn from the response of damages to drivers to reduce the number of events that need to be simulated. By comparing this approach with a state-of-the-art approach, the research shows a reduction of the computational cost by a factor of 4, an improvement in the root mean square error by a factor of 8, and an improvement in the estimate of Expected Annual Damages (EAD) by a factor of 20. This reduction in computational cost allows for the inclusion of random variables that are normally assumed constant such as the duration and time lag of drivers. A sensitivity analysis demonstrates these variables produce a statistically significant difference in the estimate of EAD, which increases its value from 172 to 219 Million USD. The research also shows the combination of events caused by drivers leading to extreme damage changes when including these additional random variables, although surge is always found to be dominant. By applying the TGP to multiple outputs, the research demonstrates the TGP is not only applicable to the case study, which shows a TGP can be implemented in current flood risk assessments.

Layman Summary

Flooding is a frequent natural disaster that can cause large social, economic, environmental, and political damages. To reduce the damages associated with flooding, the concept of risk is used. Risk is defined by the probability and the consequences associated with a flood. Frequent and less damaging floods can be weighed against infrequent and highly damaging floods, which allows for a cost-benefit analysis of how much should be spent on flood protection measures. Floods are caused by drivers which can take different forms. Examples of these drivers are precipitation, fluvial water level, and the coastal water level. These can be defined by their magnitude, duration, and time lag. When these co-occur they worsen the effect of the flood. This is called a compound flood. To correctly understand the effect of these drivers on economic damages, a large number of artificial floods need to be simulated numerically. However, this can lead to large computational costs due to the time required to simulate these. The current state-of-the-art approach is to simulate a subset and interpolate linearly. However, it is unknown if enough simulations have been performed. To this end, a Treed Gaussian Process (TGP) is used. For the case study of Charleston, it reduces the cost of modeling artificial floods by a factor of 4, while increasing the accuracy in the estimate of consequences of these floods by a factor of 8 when compared to the state-of-the-art approach. By reducing the computational costs, more resources become available to investigate the effect of the duration and time lag of the drivers on the flood risk. A sensitivity analysis shows that the duration and time lag of drivers create significant differences in the estimate of risk, and should be included in compound flood risk assessments. It also shows that not including all probabilistic variables can underestimate the risk. By including the duration and time lag, this research also shows which driver dominates the large economic damages. For Charleston, this was found to be the coastal water level. Finally, by applying the TGP in different sub-counties of Charleston County, it is shown that the economic damages to the same events change. Therefore, the TGP should not only apply to Charleston, and can be included in current compound flood risk assessments.

Acknowledgments

When I read the initial scope for this thesis 7 months ago, it peaked my interest. I knew my lack of experience with machine learning techniques and vine copulas might slow my progress during the thesis. I would therefore like to thank both Gijs and Dirk for allowing me to research this topic while barely knowing me.

I would like to thank the entire graduation committee for being approachable at any moment during the thesis and providing quick and insightful comments. Thank you José for chairing the committee, and providing advice not only on the thesis, but also on skills that are not taught at the university. Thank you Patricia for joining the committee after the thesis proposal, and for providing insightful ideas in progress meetings where I had little time to explain certain components of the thesis. Thank you Gijs for giving me detailed feedback at different stages of the thesis, and helping me not fall into rabbit holes. Thank you Dirk and Anaïs for meeting with me on a weekly basis, showing enthusiasm for compound floods during any discussion we had together, and looking out for my well-being while at Deltares.

I would also like to thank the different people at Deltares who made my internship over the past 5 months easier. Thank you to the FloodAdapt team for providing me with the SFINCS and DELFT-FIAT models. Their support and availability during the thesis helped me save a lot of time. Thank you to the members of the HYD department and other interns I met during my internship. Your friendliness made coffee and lunch breaks a lot more enjoyable.

Thank you to the numerous students that I crossed paths with while at TU Delft. Some of you made these past 2 years easier to live through and more memorable. I will miss the in-class banter and the Thursday nights at PSOR. Hopefully, this isn't the last time we see each other, and I wish you luck with your future careers.

Merci Anne-Marie for hosting me these past 2 years. Your happiness made coming home after a hard day always easier. Finally, I would like to thank my family members, especially Maman, Papa, Pierre and Julie. You have taught me life skills that I can never truly thank you for. These guided me throughout my time at TU Delft and during my thesis.

Lucas Terlinden-Ruhl
Delft, June 2024

Contents

Abstract	i
Layman Summary	ii
Acknowledgments	iii
List of Abbreviations	viii
1 Introduction	1
1.1 Research Context	1
1.2 Research Problem	2
1.3 Research Objectives and Questions	2
1.3.1 Objectives	2
1.3.2 Questions	2
1.4 Thesis Outline	3
2 Literature Review	4
2.1 Definition of Risk	4
2.2 Compound Flood Drivers	5
2.2.1 Fluvial Discharge	5
2.2.2 Pluvial	5
2.2.3 Coastal Water Level	5
2.2.3.1 Tidal	5
2.2.3.2 Non-Tidal Residual	6
2.2.4 Local Meteorological Effects	6
2.2.5 Obtaining the Time Record of Drivers in Data Scarce Environments	6
2.3 Extreme Value Analysis (EVA)	7
2.3.1 Block Maxima (BM)	7
2.3.2 Peak Over Threshold (POT)	7
2.3.3 Identification of Co-Occurring Random Variables	8
2.4 Dependence Structure	8
2.4.1 Choice of Correlation Coefficient	8
2.4.2 Explicit Methods for Dependence Modeling	9
2.4.2.1 Vine Copulas	9
2.4.3 Implicit Methods for Dependence Modeling	10
2.4.4 Closing Thoughts on the Dependence Structure	10
2.5 Reduction in Computational Time	11
2.5.1 Improve the Computational Resources	11
2.5.2 Numerical Modeling Assumptions	11
2.5.2.1 Super-Fast INundation of CoastS (SFINCS)	12
2.5.3 Selection of Simulations	12
2.5.3.1 Structured Grid	12
2.5.3.2 Sampling from the Input Space	13
2.5.3.3 Sampling Based on the Response of the Output Space	13
2.5.3.4 Interpolation Techniques	16
2.5.4 Closing Thoughts on Computational Time	17
2.6 Conclusion	17
3 Case Study and Data	19
3.1 Charleston	19

3.2	Data Related to the Time Records of Drivers	20
3.2.1	Downstream Water Level.	20
3.2.2	Upstream Discharge	21
3.2.3	Precipitation	21
3.2.4	Wind and Pressure	22
3.3	Historical Compound Floods	22
3.4	FloodAdapt	22
3.4.1	Hydrodynamic Model	22
3.4.2	Impact Model	23
4	Methods	24
4.1	Damage Modeling	25
4.1.1	Hazard Modeling with SFINCS.	25
4.1.2	Impact Modeling with DELFT-Fast Impact Assessment Tool (DELFT-FIAT)	25
4.2	Preprocessing the Time Records of Different Drivers	26
4.3	Defining a Historical Event Set for Compound Floods.	27
4.3.1	Identifying the Dominant Driver for Extreme Value Analysis.	28
4.3.2	Identifying Co-Occurring Random Variables and Defining Their Correlation	28
4.3.3	Extreme Value Analysis on Dominant Driver	29
4.3.4	Historical Event Set	30
4.4	Generating Unseen Compound Flood Event Boundary Conditions	30
4.4.1	Modeling the Dependence Structure Between Random Variables	31
4.4.2	Generating Stochastic Event Sets.	32
4.4.2.1	Defining Non Extreme Marginal Probability Distribution Functions	32
4.4.2.2	Constants in Event Sets	33
4.4.3	Defining Boundary Conditions for the SFINCS Model	33
4.4.4	Event Sets	34
4.5	Selection of Simulations and Surrogate Model	35
4.5.1	A Priori Selection of Simulations and Surrogate Model.	35
4.5.1.1	Sampling with a Maximum Dissimilarity Algorithm (MDA)	35
4.5.1.2	Surrogate Damage Modeling with Linear Scatter Interpolation	36
4.5.2	A Posteriori Selection of Simulations and Surrogate Model.	36
4.5.2.1	Initialization with a Maximum Dissimilarity Algorithm (MDA)	37
4.5.2.2	Active Learning with a Treed Gaussian Process (TGP)	37
4.5.2.3	Defining a Stopping Criterion	37
4.5.2.4	Sampling from Multiple Outputs	38
4.5.3	Comparing Both Approaches	38
4.6	Risk Modeling Under Different Approaches.	38
4.6.1	Defining Risk.	38
4.6.2	Measuring Differences in Risk Estimates.	39
4.6.3	Identifying the Driver Dominating Large Return Periods.	39
5	Results	41
5.1	Comparing a Priori and a Posteriori Methods	41
5.1.1	Description of a Priori Sampling and Surrogate Modeling	41
5.1.2	Description of a Posteriori Sampling and Surrogate Modeling	42
5.1.3	Comparing the Accuracy of Both Approaches	43
5.1.4	Comparing the Computational Cost of Both Approaches	45
5.2	A Posteriori Approach on Multivariate Input Spaces in Different Dimensions and Outputs	45
5.2.1	Computational Cost of Using the a Posteriori Approach	45
5.2.1.1	Samples Required to Reach the Stopping Criterion	46
5.2.1.2	Computational Time Required to Reach the Stopping Criterion	47
5.2.2	Significance of Additional Dimensions on Risk Curves and EAD	48
5.3	Identifying the Driver Dominating the Large Return Periods	52
5.3.1	Different Sub-County's for Two Random Variables	52
5.3.2	Different Outputs and Different Number of Random Variables.	53

6	Discussion	55
6.1	Limitations of the Study	55
6.1.1	General Limitations	55
6.1.2	Current Compound Flood Risk Assessments	56
6.1.3	Use of a Treed Gaussian Process (TGP)	57
6.2	Sampling and Surrogate Modeling	57
6.2.1	A Priori Sampling and Interpolation	57
6.2.2	Applicability of a Treed Gaussian Process (TGP)	58
6.2.2.1	Early Stopping	58
6.2.2.2	Multiple Outputs	59
6.2.3	Comparing the Computational Cost of Both Approaches	59
6.3	Sensitivity Analysis of Additional Random Variables	60
6.4	Identifying the Driver Dominating the Large Return Periods	60
7	Conclusion and Recommendations	62
7.1	Conclusions	62
7.2	Recommendations	63
	Appendices	65
A	Supplementary Information For Damage Modeling	66
A.1	Hazard Model Governing Equations	66
A.2	Removing Permanently Flooded Buildings From Impact Model	66
B	Time Records of Drivers	67
B.1	Stationarity of the Non-Tidal Residual	67
B.2	Daily Inequality Caused by the Beating of Tidal Constituents	68
B.3	Dam Management of Upstream Discharge	69
B.4	ERA5 Precipitation	70
C	Sensitivity Analysis of Three Compound Flood Drivers	71
D	Sensitivity of Statistical Models to Peak Over Threshold (POT)	73
D.1	Quantifying the Total Water Level	73
D.1.1	Non-Extreme Marginal Probability Distribution Functions	75
D.2	Applying POT to Non-Dominant Drivers	75
D.3	Multivariate Sensitivity of POT	76
D.3.1	Multivariate Sensitivity When Applying POT to the Total Water Level	78
D.4	Univariate Sensitivity of POT	79
D.5	Conclusion of Sensitivity Analysis	79
E	Correlation Analysis	80
E.1	Historical Event Set	80
E.2	Splitting Event Set Based on Precipitation Lag	81
F	Regular Vine And Pair Copulas	85
G	Defining Marginal Probability Distribution Functions (PDFs) for Non-Dominant Magnitudes	88
H	Stochastic Event Sets	91
H.1	Training Sets	91
H.2	Test Sets	93
I	Defining a Stopping Criterion for Compound Floods	96
J	Computational Cost For Components Of Damage Modeling	101
K	Sensitivity of Additional Dimensions on Risk Curves and EAD for the Classified Model	103
K.1	Inland	104
K.2	Coast	107
K.3	Discussion	109
L	Observed Binary Partition in a TGP	110

Bibliography

111

List of Abbreviations

1D NSWE	Saint Venant Equations
AIC	Akaike Information Criterion
AICc	Akaike Information Criterion corrected
ALM	Active Learning Mackay
BIC	Bayesian Information Criterion
BM	Block Maxima
BTGPLLM	Bayesian Treed Gaussian Process with a Linear Limiting Model
CDF	Cumulative Distribution Function
COV	Coefficient Of Variation
DA-NSWE	Depth-Averaged Non-linear Shallow Water Equations
EAD	Expected Annual Damages
EPDF	Empirical Probability Distribution Function
EVA	Extreme Value Analysis
FIAT	Fast Impact Assessment Tool
GTSM	Global Tide and Surge Model
GP	Gaussian Process
GPD	Generalized Pareto Distribution
GPU	Graphics Processing Unit
HH	Higher High
HPC	High-Performance Computing
i.i.d.	independent and identically distributed
KS	Kolmogorov-Smirnov
LIE	Local Inertia Equations
MCMC	Markov Chain Monte Carlo
MDA	Maximum Dissimilarity Algorithm
MSL	Mean Sea Level
MMSL	Monthly Mean Sea Level
MWU	Mann-Whitney U rank
NOAA	National Oceanic and Atmospheric Organization
NS	Navier-Stokes
NSI	National Structure Inventory
NSWE	Non-linear Shallow Water Equations
PDF	Probability Distribution Function

P Dur	Precipitation Duration
P Lag	Precipitation Lag
P Mag	Precipitation Magnitude
POT	Peak Over Threshold
RBF	Radial Basis Function
RNG	Random Number Generator
RP	Return Period
RANS	Reynolds-Averaged Navier-Stokes
RMSE	Root Mean Square Error
SFINCS	Super-Fast INundation of CoastS
SOM	Self-Organizing Maps
S Dur	Skew surge Duration
S Mag	Skew surge Magnitude
SSE	Sum Square Error
TGP	Treed Gaussian Process
T Mag	Tidal Magnitude
USGS	United States Geological Survey
UTC	Universal Coordinated Time

1

Introduction

1.1. Research Context

Flood events can cause large amounts of damage, which are not only economic but also social, environmental, and political. Between 1990 and 2006, approximately half of the individuals affected by natural disasters were impacted by floods [Birkholz et al., 2014]. Governing bodies tackle this issue by managing their flood protection infrastructure with the concept of risk. Risk can be defined as a function of hazard, exposure, and vulnerability [Klijn et al., 2015]. This facilitates the comparison between different flood events, allowing for economic optimization of flood protection measures [Baan and Klijn, 2004; Klijn et al., 2015].

Different methods are available to compute the probabilities associated with these flood events. One method is to use an extreme value analysis on water level [Arns et al., 2013]. However, this method depends on the events observed during the time record to create a probability distribution. This distribution is then extrapolated to estimate the probabilities of water levels that have not been observed yet [Hegnauer et al., 2014]. Moreover, the distribution is only valid at the location where observations are available. This is caused by the non-linear, non-steady, and non-uniform effect the bathymetry, and the drivers have on the water level [Wahl et al., 2015].

A second method is to fit statistical distributions to the drivers causing the flood event and simulate their effect on the water level using numerical models [Couasnon et al., 2022; Eilander et al., 2023c; Gori et al., 2020]. This allows for water levels to be observed within the entire model region. In a riverine environment, the water level is primarily dominated by a single driver, the upstream discharge. To reduce the number of simulations, the probability associated with a discharge can be assumed equal to the probability of the water level [Hegnauer et al., 2014].

In low-lying coastal areas, floods can be caused by multiple drivers: fluvial discharge, coastal water level (surge, waves, tide), and precipitation [Eilander et al., 2023c; Parker et al., 2023]. When drivers interact with one another, this leads to a compound flood [Wahl et al., 2015]. If these drivers are assumed statistically independent, it may change the estimate of the joint probability [Ward et al., 2018]. Moreover, if these drivers are assumed never to co-occur it will underestimate the peak water level for a given return period [Moftakhari et al., 2019], underestimating the hazard associated with an event. This is caused by the non-linear interaction of different drivers [Wahl et al., 2015]. Thus, the assumption made for riverine environments is no longer valid. To correctly quantify the risk associated with an event in a compound flood-prone area, multiple combinations of events need to be simulated [Couasnon et al., 2022; Eilander et al., 2023c].

For an accurate representation of the Expected Annual Damages (EAD), which is the exceedance probability integral of the flood impact, an estimation for all possible events is desired [Couasnon et al., 2022; Eilander et al., 2023c]. However, simulating all possible combinations is not feasible as each sample is computationally expensive to process [Hendrickx et al., 2023]. There are different methods to reduce the computational cost associated with simulating large event sets, nested in multiple dimensions. These can be split into three distinct categories: improve computational resources [Apel et al., 2016], increase the number of numerical modeling assumptions [Bates et al., 2010; Leijnse et al., 2021], and reduce the number of simulations [Eilander

et al., 2023c; Hendrickx et al., 2023].

Reducing the number of simulations is the most effective computational cost-reduction technique [Hendrickx et al., 2023], but it must be combined with an interpolation technique to obtain the results of the non-simulated events. One solution is to construct a structured grid of pre-defined simulations to run [Couasnon et al., 2022; Eilander et al., 2023c]. However, this method is affected by the curse of dimensionality: the number of required simulations increases exponentially with the number of probabilistic variables. Another solution is to use sampling algorithms that choose samples based on the input space (independent variables) [Camus et al., 2011]. However, the number of simulations must be chosen a priori, making it unclear if enough simulations have been performed. A final solution is to apply machine learning techniques, in the form of a Treed Gaussian Process (TGP), which chooses samples based on the output space (dependent variable) by using active learning [Gramacy and Lee, 2009; Hendrickx et al., 2023].

In hydraulic engineering, machine learning techniques have shown promise in reducing the number of simulations while giving a robust estimate of the output space [Hendrickx et al., 2023], but have yet to be applied to compound floods.

1.2. Research Problem

For compound floods, the research problem is related to this multidimensional input space, which makes it difficult to characterize the vulnerability related to different return periods. This is caused by: the computational cost of using numerical simulations, and the dependence structure of probabilistic variables associated with drivers. Researchers have tackled this problem by using input reduction to reduce the number of probabilistic variables, making certain sampling strategies easier to apply. For example, durations of drivers and time lags between drivers are often assumed to be constant [Couasnon et al., 2022; Eilander et al., 2023c; Jane et al., 2022]. Moreover, the current sampling strategies require an interpolation technique, often taken as linear [Couasnon et al., 2022; Eilander et al., 2023c; Jane et al., 2022]. It is unclear whether the combination of input reduction and bi-linear interpolation provides the most accurate representation of compound floods for the computational cost.

1.3. Research Objectives and Questions

1.3.1. Objectives

This research aims to create a methodology that improves the quantification of compound flood risk by using a Treed Gaussian Process. To do so, the objectives listed below are defined:

1. Define a stopping criterion for a Treed Gaussian Process that uses active learning and compare the computational cost and estimate of the EAD with a current approach.
2. Identify the driver dominating the economic damages.
3. Demonstrate the effect of different extents of input reduction on the EAD and its associated uncertainty, as well as computational cost.
4. Apply a Treed Gaussian Process to a case study and discuss its applicability for other locations.

1.3.2. Questions

To what extent does a Treed Gaussian Process improve the quantification of compound flood risk?

1. To what extent can a Treed Gaussian Process improve the estimate of compound flood risk while limiting the increase in computational cost?
2. To what extent can a Treed Gaussian Process give insight into the dominant driver when drivers are defined by multiple random variables?
3. To what extent is the value of risk, its associated uncertainty, and the computational cost of a Treed Gaussian Process sensitive to the number and choice of random variables when modeling compound flood events?
4. How are the results associated with the case study applicable to other locations?

1.4. Thesis Outline

Chapter 2 provides a background knowledge on compound flood risk, and the different tools required to conduct one. Then, the case study, accompanying data, and numerical models are presented in chapter 3. The methodology used to obtain stochastic event sets, samples for the surrogate model, and statistical tests are shown in chapter 4. Chapter 5 presents the results obtained from the methodology. These are discussed in chapter 6. Conclusions are drawn in chapter 7, and recommendations for further research are provided.

2

Literature Review

This chapter presents the background knowledge and the tools required to perform a compound flood risk assessment. In section 2.1 the definition of risk is given. Compound flood drivers, and random variables associated with them are identified in section 2.2. Information on identifying historical flood events is given in section 2.3. Section 2.4 shows the tools required to model the dependence between different random variables that can affect a compound flood-prone location. Section 2.5 shows the current state-of-the-art knowledge to decrease the computational time associated with compound flood risk assessments. A conclusion of these sections can be found in section 2.6.

2.1. Definition of Risk

The concept of risk provides governing bodies with a metric to compare the effect of different flood events. This facilitates the economic optimization of flood protection measures. Different definitions of risk exist, but these are usually functions that are dependent on hazard, exposure, and vulnerability [Klijn et al., 2015].

The hazard associated with a flood event can be defined by its probability, but also the spatially varying maximum water depth associated with it [Klijn et al., 2015]. The probability refers to the frequency of flood events, which is equivalent to the reciprocal of the Return Period (RP). The RP is the average time between subsequent flood events that leads to equal or larger magnitudes of a given variable [Gumbel, 1941]. This variable can take many forms. Previous studies have used water level (or depth) [Jane et al., 2022], discharge [Hegnauer et al., 2014], or damages (economic or social) [Eilander et al., 2023c]. Economic damage is preferred as it allows for the calculation of the Expected Annual Damages (EAD).

Exposure takes into account elements and activities that may be affected by flood hazards [Klijn et al., 2015]. These elements can be private property such as homes and businesses or public property such as roads and hospitals. Due to socio-economic developments, exposure can change over time [Baan and Klijn, 2004].

Vulnerability can be seen as the potential economic, social, environmental, and political damages caused by a hazard (flood event) to an exposed area [Klijn et al., 2015]. Here, it is assumed that vulnerability is defined solely by the economic damages of an event, creating a singular output per flood event.

The modeling of hazard requires a statistical model (section 2.4), and a hydrodynamic model (section 2.5.2). The modeling of exposure and vulnerability requires an impact model (section 4.1.2).

EAD provides an estimate of the economic damages if they were spread out over time. It allows to weigh less damaging but frequent floods with high damaging but rare floods. To obtain an estimate of EAD, the integral of the damage-frequency curve is needed [Couasnon et al., 2022; Eilander et al., 2023c]. EAD can be used to perform a cost-benefit analysis to explain how much money should be spent on flood protection measures [Oliveri and Santoro, 2000].

2.2. Compound Flood Drivers

Compound floods are caused by flood drivers co-occurring, leading to a larger flood hazard than if they occurred in isolation [Wahl et al., 2015]. Drivers co-occur because of the underlying weather patterns they are caused by [Khanal et al., 2019]. These can take many forms such as tropical cyclones and extra-tropical storms [Parker et al., 2023]. For a risk assessment, if a location is affected by different weather patterns, they should be considered separately, because the probability and consequences are dependent on the weather pattern [Diermanse et al., 2023; Eilander et al., 2023c; Nederhoff et al., 2024; Parker et al., 2023]. However, for short-time records (of data), it can be difficult to get a statistical understanding of weather patterns that have large RPs. Two solutions have been proposed in the literature. Firstly, they can be simulated with global models (section 2.2.5) to extend the time record but are computationally expensive to use (section 2.4.3). Secondly, they can be grouped to simplify the risk assessment [Jane et al., 2022].

In a coastal environment, a multitude of drivers can generate compound floods [Couasnon et al., 2020; Eilander et al., 2023c; Hendry et al., 2019; Parker et al., 2023; Wahl et al., 2015]. These can be grouped according to their effect on a hydrodynamic model. These models (section 2.5.2) are required to simulate the flood hazard map associated with a compound flood event. In the case of subcritical flow, two boundary conditions need to be defined to reach a well-posed problem: fluvial (upstream) discharge and coastal (downstream) water level. Precipitation and other meteorological conditions are not explicitly required but are defined for each grid cell. Waves are also not required but are defined at the coastal water level boundary condition. Waves contribute to flood risk in two ways, with wave set-up and run-up. Wave run-up can be ignored as it does not cause sustained flooding [Parker et al., 2023].

The effect a driver has on the overall model is not only defined by its magnitude but also by duration, location, and spatial spread [Apel et al., 2016]. In compound floods, the lag between the peaks of different drivers also becomes important [Eilander et al., 2023c; Moftakhari et al., 2019].

2.2.1. Fluvial Discharge

Discharge is defined as the volumetric flow rate of water. In a riverine environment, it is affected by the properties of the upstream catchment [Hendry et al., 2019]. These include the surface area, elevation gradient, and soil characteristics of the catchment, which dictate how much base flow (only supplied by groundwater flow) and surface runoff (generated by precipitation events) occurs. Surface runoff is the mechanism that leads to extremes in measured data [Hendry et al., 2019; Wahl et al., 2015].

In a coastal environment, backwater effects either cause an increase or decrease in discharge as the flow approaches the river mouth. The type of backwater effect depends on the magnitude of both discharge and the coastal water level.

If local data cannot be used, global models can estimate discharges [Eilander et al., 2023b].

2.2.2. Pluvial

Precipitation leads to a local increase in water level causing local flooding through surface runoff [Wahl et al., 2015]. For relatively flat topographies, it can be assumed to be spatially uniform [Eilander et al., 2023c] or Gaussian distributed [Apel et al., 2016].

2.2.3. Coastal Water Level

2.2.3.1. Tidal

Gravitational bodies (sun and moon) create bulges in ocean water masses through tangential forces generated by their differential pull [Bosboom and Stive, 2021]. The diurnal rotation of the earth creates a semi-diurnal tide which has two peaks and two troughs in water level over a day. The tidal period is dependent on the synodic rotation of the celestial body [Bosboom and Stive, 2021]. Thus, the moon and the sun both generate a semi-diurnal component as their synodic rotations are not the same. The beating of these signals creates the spring-neap tidal cycle, causing oscillations in the tidal range.

The declination angle of the earth's axis generates diurnal (and other semi-diurnal) components [Bosboom and Stive, 2021]. The interaction of diurnal and semi-diurnal components generates a daily inequality, causing the two higher and two lower tides to have unequal magnitudes.

The magnitude of tidal constituents is dependent on a multitude of factors such as the latitude or the shape

of the continental shelf [Bosboom and Stive, 2021]. The tidal beating is thus not constant around the world. Moreover, tidal constituents can have periods that are longer than a day. The longest of which is 18.6 years, caused by the variation of the lunar declination [Bosboom and Stive, 2021]. Therefore, to model tidal behavior, a harmonic analysis can be applied. If a gauge has measured more than 20 years of data, the frequency of all constituents should be identifiable.

2.2.3.2. Non-Tidal Residual

The non-tidal residual can be defined as the resulting water level which cannot be attributed to sea level rise or tides. This definition therefore encompasses Monthly Mean Sea Level (MMSL) changes, wave setup, and storm surges. This driver is generated by offshore meteorological effects. If local data is not available, the Global Tide and Surge Model (GTSM) can be used [Parker et al., 2023].

For total water level measurements, a trend in Mean Sea Level (MSL) occurs and must be removed from the data. Different methods are available, but they introduce biases and do not lead to the same results. It is therefore recommended to remove the 1-year moving average [Arns et al., 2013]. However, this will result in tidal components that have a period longer than a year to be removed from the non-tidal residual (if they are present). The one-year moving average still causes seasonality on a scale smaller than a year to be present.

Increased total water levels induce a phase shift in the tidal signal, creating peaks in the non-tidal residual that are an artifact of small changes in the timing of the predicted tide [Williams et al., 2016]. To correctly capture the magnitude of events, skew surge can be used [Williams et al., 2016]. It is defined as the absolute difference between the peak total water level and peak tide during a tidal cycle (fig. 2.1).

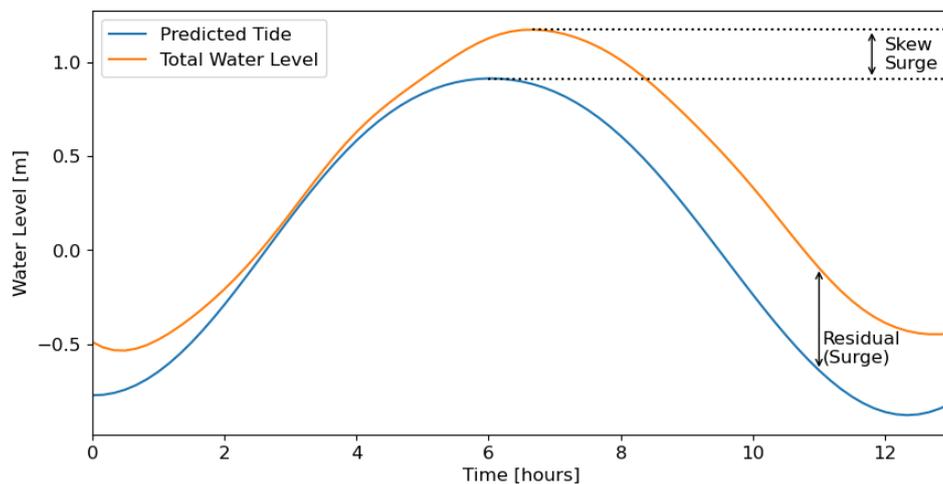


Figure 2.1: Definition of skew surge. Figure is adapted from Williams et al. [2016] and uses the matplotlib python package [Hunter, 2007].

2.2.4. Local Meteorological Effects

The non-tidal residual component is predominantly affected by meteorological effects happening offshore [Bakker et al., 2022]. However, local meteorological effects can influence the spatially and time-varying water level through pressure and wind setup as well as wind-driven waves. For certain locations, wind and pressure have to be taken into account to reach the best description of the water level associated with a compound flood [Leijnse et al., 2021]. This is not always the case, because the effect wind and pressure have on the water level are dependent on (i) the water depth, (ii) the geometric constraints and size of the local water bodies, (iii) the time over which they act [Bosboom and Stive, 2021].

2.2.5. Obtaining the Time Record of Drivers in Data Scarce Environments

The method used in a research paper is often dependent on the case study location [Chen and Liu, 2014; De Angeli et al., 2018; Kumbier et al., 2018; Webster et al., 2014]. Eilander et al. [2023b,c]; Parker et al. [2023] create frameworks that are dependent on global models, which standardizes the approach. Parker et al. [2023] use ERA5 climate reanalysis as a boundary condition to GTSM. They were able to determine that tides, wave setup, and non-tidal residual were the dominant contributions to extreme water levels along the

Atlantic coast of the United States. It was also demonstrated that the relative influence of the non-tidal residual component increased for more extreme events, driven by hurricanes. They also showed that the skill of the framework was similar to more complex models. Eilander et al. [2023c] include GTSM forced with ERA5 as the downstream boundary condition but also uses ERA5 to compute precipitation and discharge. They couple a hydrodynamic and impact model to reach a value for risk.

Using global models is attractive as it makes it easier to create a local model without spending time on data scraping. However, the studies are affected by the uncertainty and the resolution of global datasets. For example, ERA5 can capture the location and shape of extreme precipitation events but is unable to capture extreme precipitation totals [Lavers et al., 2022]. This propagates to the outcome as irreducible uncertainties are introduced when computing extreme flows [Bates et al., 2021]. Therefore, if local data sources are available, these would be preferred.

2.3. Extreme Value Analysis (EVA)

Flood hazard is driven by extremes in compound flood driver magnitudes [Couasnon et al., 2022; Eilander et al., 2023c; Hegnauer et al., 2014; Jane et al., 2022]. These events can be identified through an Extreme Value Analysis (EVA) [Arns et al., 2013], which assumes that the time series is stationary and that each event is independent and identically distributed (i.i.d.). L-moments can be used to fit parameters and choose the best theoretical distribution according to a certain metric. For instance the Akaike Information Criterion (AIC), Akaike Information Criterion corrected (AICc), or Bayesian Information Criterion (BIC) score can be used [Eilander et al., 2023a]. The AIC estimates prediction error and provides a score, which can be computed with eq. (2.1). BIC (eq. (2.2)) further penalizes model complexity as smaller scores are associated with models that have fewer parameters. For all three metrics, the best-fitting model has the lowest score. Two different approaches can be used to apply EVA, which are Block Maxima (BM) (section 2.3.1) and Peak Over Threshold (POT) (section 2.3.2).

$$AIC = -2 \ln L + 2k \quad (2.1)$$

$$BIC = k \ln n - 2 \ln L \quad (2.2)$$

In eq. (2.1) and eq. (2.2), L is the maximized value of the likelihood function, k is the number of parameters in the model, and n is the number of data points used to fit the distribution.

2.3.1. Block Maxima (BM)

This approach divides the time record of a random variable into equal blocks of time. Since flood drivers show seasonality on the scale of a year, a common approach is to take blocks that span a year [Eilander et al., 2023c]. In certain locations, the seasonality may be shorter than a year. In these cases, a block that spans a month can also be used [Couasnon et al., 2022]. The largest event within the block is retained to fit an extreme value distribution. There are numerous limitations to this approach. Firstly, the start and end of each block need to be defined to ensure the i.i.d. assumption is not violated. Secondly, the number of available historical events is equal to the number of years of coinciding time records. This means that if two extremes occur within a block, only the largest is used, and if no extremes occur within a year, the largest value within the block is still retained.

2.3.2. Peak Over Threshold (POT)

This approach sets a threshold, any event above this threshold can be defined as an extreme. The validity of the i.i.d. assumption is dependent on a de-clustering time. The main limitation of this approach is defining a robust threshold and an accurate de-clustering time [Bocharov, 2023], which can be subjective. This can be reduced by using mean residual life, return value, and parameter stability plots [Bocharov, 2023]. For a range of thresholds, the mean residual life plot should show an approximately linear behavior, and both the return value and parameters should be approximately constant. POT has been shown to be more stable than BM in predicting return values when varying the length of the time record, which makes it advantageous to use when the time record is short [Arns et al., 2013].

2.3.3. Identification of Co-Occurring Random Variables

The above EVA methods describe how to obtain extreme events and distributions for random variables, but they do not explain how to identify event pairs to model the joint probability distribution. For both approaches, there exist two methods. Firstly, an EVA can be fit to all random variables representing the magnitude of flood drivers [Eilander et al., 2023c]. For BM, event pairs are identified by occurring within the same block. This will intrinsically contain both compound and non-compound events. Non-compound events are expected to contribute heavily to the joint, introducing a bias [Eilander et al., 2023c]. For POT, drivers exhibit a different number of extremes for a given year, the method used for BM cannot be used. The solution is to identify the co-occurring magnitudes of other drivers for the different random variables, which ensures all events are compound. However, since not all events will be extreme, the probability distributions will not be valid for the entire joint distribution. Secondly, to ensure all events are compound, and to ensure the probability distributions associated with an EVA are valid, an EVA can be fitted to a dominant driver. The values associated with the different random variables can then be identified by investigating the co-occurring magnitudes [Jane et al., 2022]. The issue here is there is an omission of events that can lead to significant damages associated with extremes of other drivers, this method is therefore case study specific.

2.4. Dependence Structure

Drivers can be considered to act independently [De Angeli et al., 2018; Webster et al., 2014]. However, this may misinterpret the joint probability of these drivers if they are in reality dependent [Ward et al., 2018]. A dependence structure is therefore required to best represent the RPs of a compound event. Compound flood risk studies differ in the way this dependence structure is defined, which causes the return value associated with a RP to change [Jane et al., 2022; Ward et al., 2018].

2.4.1. Choice of Correlation Coefficient

There are different ways to characterize the correlation between drivers. The Pearson correlation coefficient assumes a linear relationship between bivariate variables. This is a significant limitation, as not all correlations are linear. Moreover, its parametric nature can make it susceptible to outliers, which can heavily affect the estimation of correlation. Kendall's tau (τ) or Spearman's rho (ρ) take a more robust approach to computing the correlation of extreme events in a bivariate space [Wahl et al., 2015]. These are ranked-based approaches, which are not dependent on their marginal distributions. For example, Kendall's τ is defined as the probability of concordance minus the probability of discordance of two random variables [Czado, 2019]. There are different variations of Kendall's τ . τ_A will not make adjustments for ties (eq. (2.3)), while τ_B does (eq. (2.4)) [Czado, 2019].

$$\tau_A = \frac{N_c - N_d}{\binom{n}{2}} \quad (2.3)$$

$$\tau_B = \frac{N_c - N_d}{\sqrt{N_c + N_d + N_1} \cdot \sqrt{N_c + N_d + N_2}} \quad (2.4)$$

In eq. (2.3) and eq. (2.4), N_c represents the number of concordant pairs, N_d the number of discordant pairs. In eq. (2.3), n represents the total number of pairs. In eq. (2.4), N_1 the number of extra x_1 pairs, and N_2 the number of extra x_2 pairs.

Regardless of its variations, Kendall's τ has been used in previous compound flood studies and can show interesting outcomes. It has shown climate change is causing a larger incidence of joint events in the upper tail region [Wahl et al., 2015]. It has also shown the correlation is dependent on the time lag associated between drivers [Ward et al., 2018], and the season (i.e. tropical cyclones are only relevant during a certain part of the year) [Couasnon et al., 2022]. This is because of the underlying physical processes that cause the drivers. One example is the weather pattern first affects the coast, which causes a combination of surge and wind-driven waves, before causing precipitation inland, causing a flood wave and urban precipitation. Another example is subsequent weather patterns cause precipitation to occur inland while simultaneously affecting the coast [Khanal et al., 2019].

2.4.2. Explicit Methods for Dependence Modeling

Multivariate probability density functions are used to represent the dependence structure between different random variables. Examples of methods that can be applied to higher dimensional problems include copula functions or the Heffernan and Tawn model [Couasnon, 2023].

In compound floods, one popular method is to use bivariate copulas on local boundary conditions [Couasnon et al., 2022; Gori et al., 2020; Moftakhari et al., 2019, 2017; Phillips et al., 2022]. Sklar's Theorem states that for a 2-dimensional random vector with joint distribution F , and marginal Cumulative Distribution Functions (CDFs) (F_i, F_j) , the joint distribution can be expressed with eq. (2.5). For continuous distributions, the copula C is unique [Czado, 2019]. The copula is fitted in the uniformly ranked space for variables u and v .

$$F(x_i, x_j) = C(F_i(x_i), F_j(x_j)) = C(u, v) \quad (2.5)$$

In two dimensions, isolines can be drawn to represent the bivariate CDF giving the return values for which a RP can occur [Moftakhari et al., 2019]. Different scenarios exist [Serinaldi, 2014], but Moftakhari et al. [2019] only investigate two of these. One represents the 'AND' case where both random variables exceed the specified RP. The other represents the 'OR' case where only one random variable exceeds the specified RP. A choice must be made to discriminate between the two. The 'AND' case is preferred as pairs associated with the 'OR' case at the extremities of the isoline exceed the values associated with the marginal distribution for the same RP [Moftakhari et al., 2019]. An assumption can be made only to simulate this event, which reduces the number of simulations [Moftakhari et al., 2019]. However, this assumes that a singular design event best represents the return value for a given RP.

Another method is to generate events with the copula [Couasnon et al., 2022]. This assumes each generated event has a constant frequency of occurrence per year, f_{oc} , which is dependent on the extreme rate of the historical event set, E_{rate} (eq. (2.6)). An empirical CDF of the dependent variable can be created once all generated events are simulated/estimated [Couasnon et al., 2022], effectively removing the assumption mentioned above. For large event sets and large computational costs per simulation, this method becomes dependent on the sampling and interpolation methods used (see section 2.5).

$$f_{oc} = \frac{1}{E_{rate}} \quad (2.6)$$

2.4.2.1. Vine Copulas

In a higher dimensional space, vine copulas can be used to characterize the dependence structure and generate events [Eilander et al., 2023c]. A regular vine is defined as a set of trees, $V = (T_1, \dots, T_{d-1})$ on d elements where each tree, $T_j = (N_j, E_j)$, is connected. Moreover, T_1 is a tree with node set $N_1 = 1, \dots, d$ and edge set E_1 . For $j \geq 2$, T_j is a tree with node set $N_j = E_{j-1}$ and edge set E_j . Finally, for $2 \leq j \leq d-1$, and $a, b \in E_j$, $|a \cap b| = 1$, where a and b are nodes in T_j . Therefore, in T_{j-1} , they are edges and must share a common node [Czado, 2019].

The number of regular vines grows at a rate (eq. (2.7)). This is caused by an enumerating problem when counting all possible edges [Morales-Nápoles et al., 2023].

$$\frac{d!}{2} \cdot 2^{\binom{d-2}{2}} \quad (2.7)$$

A vine copula is defined by three components when applied to a d dimensional random vector [Czado, 2019]. Firstly, F marginal distributions for each d dimension are needed. Secondly, a V vine tree sequence on d elements is required. Finally, a set of B bivariate pair copulas is needed. The number of copulas follows (eq. (2.8)), because each edge, e in V is defined by a copula. A simplifying assumption is made where the copula for an arbitrary edge is not dependent on the specific values of conditioning variables [Czado, 2019].

$$\frac{d \cdot (d-1)}{2} \quad (2.8)$$

To select bivariate copula families (and their associated parameters), **AIC** and **BIC** can be used. They are more reliable than other tests such as the Cramer-von Mises statistic [Czado, 2019]. The best-fitting model is the one with the lowest score.

The pair copula family set and their associated parameters depend on the vine structure [Czado, 2019]. Moreover, eq. (2.7) makes it difficult to use a brute force approach on high-dimensional datasets [Morales-Nápoles et al., 2023]. For this reason, a sequential top-down approach called Dißmann's algorithm [Dißmann et al., 2013] is often used [Czado, 2019; Nagler and Vatter, 2023]. This approach selects the first tree T_1 , then constructs the following trees sequentially until the last tree, T_{d-1} , is reached. Each tree needs to be defined for all nodes and must satisfy the definition of a regular vine. Moreover, for each tree T_i , the best tree can be obtained by maximizing a certain weight, w . This weight can be defined with different parameters [Czado, 2019]. One approach is to use the absolute empirical Kendall's tau [Nagler and Vatter, 2023]. Dißmann's algorithm only makes local optimizations for each tree. Therefore, there is no guarantee the algorithm will find the best set of trees, V [Morales-Nápoles et al., 2023]. However, brute force can require significant computational cost for high dimensional datasets, and has not yet been used when there are more than 8 random variables [Morales-Nápoles et al., 2023].

Events can be generated from a fitted copula with an inverse Rosenblatt transform [Nagler and Vatter, 2023]. For each event, d , **i.i.d.** samples are generated from a uniform distribution on the interval $[0, 1]$ (**U**). For a copula C , these independent uniform variables are transformed in a random vector \mathbf{V} of length d using eq. (2.9) [Czado, 2019].

$$V_i = C_{i|i-1, \dots, 1}^{-1}(U_i | U_{i-1}, \dots, U_1), i = 1, \dots, d \quad (2.9)$$

Once the marginal distributions F for each d dimension are fitted (using methods such as **AIC**, **BIC**, or Sum Square Error (**SSE**)), they can be used to transform \mathbf{V} from a uniformly ranked space to the variable space. This uses the inverse **CDF** of F_i and takes V_i as a quantile. Combining this process with eq. (2.9) M times generates an event set of size M .

2.4.3. Implicit Methods for Dependence Modeling

Implicit methods, such as a weather generator partially solve the dependence structure. This is because individual drivers are inherently linked as they are caused by the same or correlated weather pattern(s) [Khanal et al., 2019]. One limitation is that the multivariate time record of multiple drivers is reduced to a univariate hazard (water depth) time series [Couasnon, 2023]. To obtain this time series, computationally intensive numerical models (hydrological and hydrodynamic) need to be run to convert boundary conditions from the weather generator [Couasnon, 2023; Khanal et al., 2019]. Since changes in coastal water levels and discharges are influenced by regional/continental changes in meteorological conditions, these models need to be defined on a similar scale. This further increases the computational cost.

For extratropical weather events, weather generators can be used to create a synthetic time record for water levels and discharge [Breinl et al., 2015; Khanal et al., 2019]. A hydrological model is needed to compute the effect of precipitation on discharge. A hydrodynamic model is needed to compute the effect of discharge and meteorological conditions on water levels. These weather generators can demonstrate why drivers are dependent on each other, and that there exists a range of time lags where both drivers are dependent on one another. The land use and catchment area causes this. For large **RPs**, this dependence increases, as the weather pattern causing the compound event becomes more dominant [Khanal et al., 2019].

For tropical cyclones, the computational cost can be reduced by creating a subset of the meteorological forcings using a sampling algorithm (see section 2.5). By only sampling relevant physical parameters, only a small event set needs to be simulated. This method reduces the uncertainty in extreme value predictions and is useful in areas with limited data on tropical cyclones [Bakker et al., 2022]. However, the model lacks important independent variables in a compound flood environment, such as precipitation, which means the current iteration can only be used to assess coastal floods.

2.4.4. Closing Thoughts on the Dependence Structure

With a large number of probabilistic variables, an explicit method such as vine copulas seems to be the most attractive to define the dependence structure. They are a versatile option as they are flexible and in-

interpretable. Since they create an event set of local boundary conditions, the computational cost is reduced (no hydrological model, and the hydrodynamic model domain is smaller). To find the best vine, brute force is expected to perform better than Dißmann's algorithm. However, brute force can be computationally expensive for high dimensional datasets and has not yet been used for more than 8 random variables [Morales-Nápoles et al., 2023]. For compound floods, if the magnitude, duration, lag, location, and spread are made probabilistic for each driver, it is expected there will be more than 8 random variables. Therefore, Dißmann's algorithm is preferred. For correlation, the advantages of Kendall's τ cannot be overlooked, especially if it is used to apply Dißmann's algorithm [Nagler and Vatter, 2023].

2.5. Reduction in Computational Time

It is computationally infeasible to get a holistic overview of compound floods for large event sets, nested in multiple dimensions [Hendrickx et al., 2023]. The explanation is twofold. Firstly, the number of required simulations scales with the number of dimensions, which is equivalent to the number of random variables. This is similar to the curse of dimensionality in deep learning [Prince, 2023]. Secondly, flood propagation is both non-linear and non-uniform, which means hydrodynamic models are required to understand the spatio-temporal varying water depths for different events. Three different solutions exist to this problem, and they are not mutually exclusive from one another (i.e, they can be combined). Firstly, the computational resources can be improved (section 2.5.1). Secondly, the numerical problem related to the computation of the flood hazard can be simplified (section 2.5.2). Finally, the number of samples can be reduced by only simulating a smaller subset of the event set and using interpolation techniques to obtain all values for the event (section 2.5.3).

2.5.1. Improve the Computational Resources

The computational resources can be improved, reducing the computational time associated with a computational cost. One example is a Graphics Processing Unit (GPU) [Apel et al., 2016]. Another example is a High-Performance Computing (HPC) cluster [Morales-Nápoles et al., 2023]. However, these resources are not always readily available. Moreover, in the case of the GPU, its use is software-specific.

2.5.2. Numerical Modeling Assumptions

The best possible description of a moving fluid can be obtained by solving the Navier-Stokes (NS) equations. However, this is computationally infeasible for a large model domain, as the grid cell size needs to be sufficiently small to capture the effect of turbulence to achieve a high accuracy [Pourmostafa and Ghadimi, 2020]. Moreover, certain terms contained within these equations can be assumed negligible as they do not contribute extensively to the computation of the flood hazard [Leijnse et al., 2021]. The numerical problem can therefore be simplified by neglecting certain physical processes.

In the context of regional flooding, the effect of turbulence can be assumed to be negligible. A reduction in the number of grid cells can be achieved by using the Reynolds-Averaged Navier-Stokes (RANS) equations. RANS is reached by averaging over the turbulent time scale, which smooths the effect of turbulence. Additional terms, in the form of Reynolds stresses, are created to represent the effect of unresolved turbulence. These terms are solved by using turbulence closure models, which are semi-empirical [Stewart, 2008]. Moreover, incompressible flow can be assumed to reduce the number of terms in the governing equations [Stewart, 2008].

For low-lying coastal areas, the effect of vertical velocity can be ignored. This is achieved by reducing the vertical momentum balance to a hydrostatic balance, reducing the number of terms that need to be solved for the Non-linear Shallow Water Equations (NSWE). This assumes the vertical velocities are smaller than the horizontal velocities, which is valid if the horizontal length scale is larger than the vertical length scale. To further reduce the number of grid cells, and the complexity of the numerical scheme, the vertical direction is averaged over. This leads to the Depth-Averaged Non-linear Shallow Water Equations (DA-NSWE). Different numerical solvers can solve both the NSWE and the DA-NSWE [Leijnse et al., 2021]. These solvers can capture the relevant hydrodynamic processes and have previously been used in compound flood risk assessments: MIKE 11 + MIKE 21 [Webster et al., 2014]; Delft3D [Kumbier et al., 2018]. They are expected to bring significant accuracy, but have a large computational cost associated with them [Leijnse et al., 2021].

To further simplify the numerical scheme, and reduce the computational cost, the Saint Venant Equations (1D NSWE) further average over a lateral direction (eq. (2.10) and eq. (2.11)). This is the starting point for the

numerical solver implemented in LISFLOOD-FP [Bates et al., 2010].

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0 \quad (2.10)$$

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x} \left[\frac{Q^3}{A_c} \right] + g A_c \frac{\partial h}{\partial x} + c_f \frac{|Q|Q}{A_c R} = 0 \quad (2.11)$$

Where A is the channel area [m^2], t is time [s], Q is the discharge [$m^3 s^{-1}$], x is the spatial coordinate [m], A_c is the conveyance area [m^2], g is the gravitational constant [$m^2 s^{-1}$], h is the water depth [m] and R is the hydraulic radius [m]. In eq. (2.11) (conservation of momentum in the x direction), the first term represents acceleration, the second advection, the third water slope and the final term is friction slope. An equation analogous to eq. (2.11) can be defined for the y direction. To reach the Local Inertia Equations (LIE), Bates et al. [2010] neglect advection, assumes a rectangular channel and approximates R with h . This creates an explicit scheme, which can cause instabilities at shallow depths. These can be reduced by modifying the discharge in the term that dictates friction slope, creating an equation that is linear in discharge at the next time step.

2.5.2.1. Super-Fast INundation of CoastS (SFINCS)

Super-Fast INundation of CoastS (SFINCS) [Leijnse et al., 2021] is a model based on the LIE, but further optimized for coastal flooding. When compared to the SFINCS, it includes a wind drag term, a source term (allows for precipitation to be included), and an optional advection term that is useful for wave-related problems (but approximately doubles the computational cost). SFINCS also does not resolve morphodynamics or the surf zone due to the increase in computational cost but enables coupling with empirical relationships. SFINCS also achieves faster simulation times by computing fewer grid cells (deeper parts of the domain can be ignored), and performs parallel computations.

While validating SFINCS, Leijnse et al. [2021] show that the model underestimates wave heights as it is a more dissipative numerical scheme. When applying the model for a compound flood event and comparing it to a process-based model (Delft3D), SFINCS is over 100 times faster while having a difference of 5.2 cm [Leijnse et al., 2021]. This error is likely attributed to the resolution of the data [Bates et al., 2021], and the lack of a sub-grid schematization [Eilander et al., 2023b], which has since been implemented.

Eilander et al. [2023b] validate SFINCS for a case study in Mozambique. When comparing it to a global model (CaMa-Flood), SFINCS shows a better skill, but there are still significant amounts of false positives and negatives. This can have a large effect on the estimation of damages for a given event as the model may inundate an area with high vulnerability, when in reality it remains dry (false positive).

As mentioned above, SFINCS does not take into account morphodynamics and does not resolve the swash zone. However, if morphological effects are not taken into account, water depths predicted by the model may be inaccurate for a compound event [Kumbier et al., 2018]. Empirical relations can be used to get a value for wave run up [Leijnse et al., 2021]. However Parker et al. [2023] argued wave run-up could be ignored as it does not cause sustained flooding.

Thus, SFINCS may not provide the best representation of maximum water depths across the model region. Nonetheless, its reduction in computation cost is extremely attractive for first-order risk assessments [Eilander et al., 2023c].

2.5.3. Selection of Simulations

The number of samples can be reduced by only simulating a smaller subset of the event set and using interpolation techniques to obtain all values for the event. These can take various forms but can be placed in three distinct categories: creating a structured grid [Couasnon et al., 2022; Eilander et al., 2023c], sampling from the input space [Camus et al., 2011], and/or sampling based on the output space by using active learning [Gramacy and Lee, 2009; Hendrickx et al., 2023].

2.5.3.1. Structured Grid

This approach disregards the event set by creating a grid based on the exceedance probabilities of all driver magnitudes [Couasnon et al., 2022; Eilander et al., 2023c]. In the case of Eilander et al. [2023c], the grid

is based on one non-extreme condition and six extreme univariate conditions (2, 5, 10, 50, 100, and 500-year return values). With four drivers, this means $7^4 = 2401$ simulations need to be run, clearly showing the effect of the curse of dimensionality. This method is highly dependent on the type of interpolation technique used to obtain the consequences related to the event set (see section 2.5.3.4). By being independent of the joint probability, changes in dependence could be investigated. A structured grid also allows for a sensitivity analysis to be performed, making it easier to understand changes in the risk with respect to changes in the dependence structure of the flood drivers modeled [Eilander et al., 2023c].

2.5.3.2. Sampling from the Input Space

This approach uses algorithms to choose a subset of samples based on the input space, which can be defined by a grid or an event set. Possibilities include Monte Carlo sampling, Latin hypercube sampling, directional sampling, stratified sampling, importance sampling, Self-Organizing Maps (SOM), k-means clustering, numerical integration, and Maximum Dissimilarity Algorithm (MDA) [Camus et al., 2011; Diermanse et al., 2023; Hendrickx et al., 2023]. Sampling from the input space is not efficient as it can generate clusters in the output space that have a high correlation with one another. Moreover, the number of samples needs to be defined a priori. This number can be difficult to choose when information on the output is unknown.

For compound floods, interpolation is required. For this purpose, Camus et al. [2011] demonstrate that a MDA is the best algorithm to use when compared to k-means clustering and SOM. This is because it is not affected by the disparity in density of the input, and thus best explores the boundaries of the input with a limited number of samples. Based on an initialization, and a dissimilarity measure, which are often taken as the maximum of an arbitrary dimension and Euclidean distance respectively [Camus et al., 2011; Hendrickx et al., 2023], a MDA [Kennard and Stone, 1969] repeats two steps until the number of samples is reached. Firstly, given the current subset, it assigns all remaining points to the sampled point to which it is the least dissimilar. Secondly, based on this assignment, the point which has the largest remaining dissimilarity is the next point added to the subset.

An example of using a MDA to collect 6 samples on a regular spaced grid can be seen in fig. 2.2. The color blue represents available samples, the color orange represents samples currently contained in the subset, and solid black lines are used to partition the input space and show which available samples are the closest to the current samples in the subset. In this case, the maximum of both dimensions is used to initialize the MDA, and Euclidean distance is used as a measure of dissimilarity. Based on this particular initialization, the first four samples will contain the corners of the input space. The subsequent samples will create an evenly spaced unstructured grid, which has a finer resolution with additional samples.

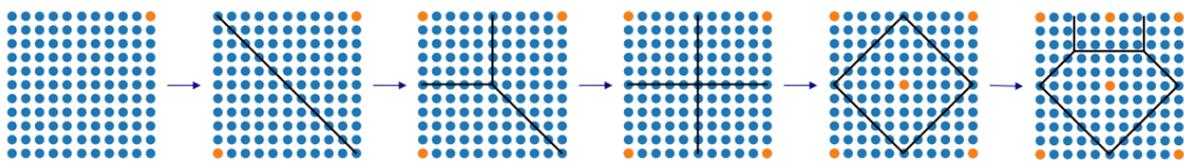


Figure 2.2: Using a MDA to collect 6 samples from a regularly spaced grid in 2 dimensions. Plots are created using the Matplotlib Python package Hunter [2007].

2.5.3.3. Sampling Based on the Response of the Output Space

This approach uses machine learning techniques (regression) to fit a model to the current subset of samples. These models are regularized, removing the need for a validation/test set, which would be needed for other machine learning algorithms that use stochastic gradient descent (a neural network would be an example). The model can provide estimates of the dependent variable(s) for the input space, and can be combined with an acquisition function to select samples, but requires additional computational resources. These estimates allow for a stopping criterion to be defined, which dictates when a sufficient number of samples have been simulated to reach an accurate estimate for all the dependent variable(s) corresponding to the input space. Finding an optimal stopping criterion for sampling algorithms is an active area of research. There are different classifications. Accuracy and confidence-based criteria use a test dataset to evaluate the current model concerning known values. Stability-based criteria evaluate the difference between values before and after obtaining new data [Ishibashi and Hino, 2021]. Only the last criterion is appealing as it circumvents a

test dataset. A test dataset is not desired as it requires more simulations, increasing the computational cost. Moreover, picking these samples would require a separate sampling algorithm.

One such regression model is a Gaussian Process (GP). An extension is a Treed Gaussian Process (TGP) [Gramacy and Lee, 2009], which has been applied in hydraulic engineering to reduce the number of simulations for an event set in 13 dimensions [Hendrickx et al., 2023].

Gaussian Process (GP)

A GP is defined by two components, its mean ($m(x)$) and covariance (or kernel) ($k(x, x')$) (eq. (2.12), where x and x' represent 2 points in the input space), and uses Bayes theorem to actively learn from sequential data arrivals [Bishop, 2006]. After each data point, a new model is trained, where the previous posterior is set as the new prior (eq. (2.13), where t represents target, \hat{t} is the posterior target, \hat{m} is the posterior mean (eq. (2.14)), $\hat{\sigma}^2$ is the posterior variance (eq. (2.15)), \hat{x} are inputs where prediction occurs, \mathbf{X} are the sampled inputs, β^{-1} represents the confidence in the observed data (related to measurement error), and \mathbf{I} is an identity matrix).

$$y(x) \sim GP(m(x), k(x, x')) \quad (2.12)$$

$$p(t|\hat{t}) = N(\hat{t}|\hat{m}, \hat{\sigma}^2) \quad (2.13)$$

$$\hat{m} = K(\hat{x}, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \beta^{-1}\mathbf{I}]^{-1}\mathbf{t} \quad (2.14)$$

$$\hat{\sigma}^2 = k(\hat{x}, \hat{x}) - K(\hat{x}, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \beta^{-1}\mathbf{I}]^{-1}K(\mathbf{X}, \hat{x}) + \beta^{-1} \quad (2.15)$$

Each trained model provides predictions and modeling uncertainties for the dependent variable(s). This assumes the local Gaussian approximation is valid [Gramacy and Lee, 2009]. These modeling uncertainties can be used as an acquisition function to actively learn from the output space. One metric is to use the quantile difference of the modeling uncertainty, also known as the Active Learning Mackay (ALM) statistic, which is associated with the largest gain in expected information [MacKay, 1992]. For a structured grid, this may lead to inefficient simulations as this metric is expected to be the largest at the extremities of the grid. This is caused by the GP using the prior to fit the input space outside of its sampled information. For an unstructured grid, the number of available samples is reduced at the extremities of the input space, reducing this effect. Moreover, if these points are simulated, the information gained from these samples will be useful as they are part of the stochastic event set. The computational cost associated with a GP is related to inverting a matrix when computing the posterior predictive surface. This is proportional to the number of sampled events cubed ($O(N^3)$) [Bishop, 2006; Gramacy and Lee, 2009].

A common initial prior for a GP is to assume a mean with a value of zero, a squared exponential kernel (eq. (2.16), where σ_f is a scaling factor and ℓ is a length scale), and a constant β [Bishop, 2006]. Therefore, if a prior does not include information on any data points, it will initially sample randomly from the input space, as the quantile difference would be spatially uniform (fig. 2.3). To prevent this, and initially simulate data points that are expected to have meaningful information associated with them, Hendrickx et al. [2023] use a MDA with a small subset size.

$$k(x, x') = \sigma_f^2 \cdot \exp\left(-\frac{\|x - x'\|^2}{2 \cdot \ell^2}\right) \quad (2.16)$$

Treed Gaussian Process (TGP)

The advantage of using a treed model is that it partitions the input, allowing different models to fit different variations in the data. This allows the TGP to be non-stationary and account for heteroskedasticity. Heteroskedasticity occurs when the noise/residual for the observations is not constant for the entire input space. Therefore partitions prevent uncertainty from propagating from feature-rich sections, which would result in large ALM statistics in uninteresting areas [Gramacy, 2023]. Therefore, in theory, a TGP enables the sampling to only occur in feature-rich sections of the input space. A feature can be defined as a measurable quantity

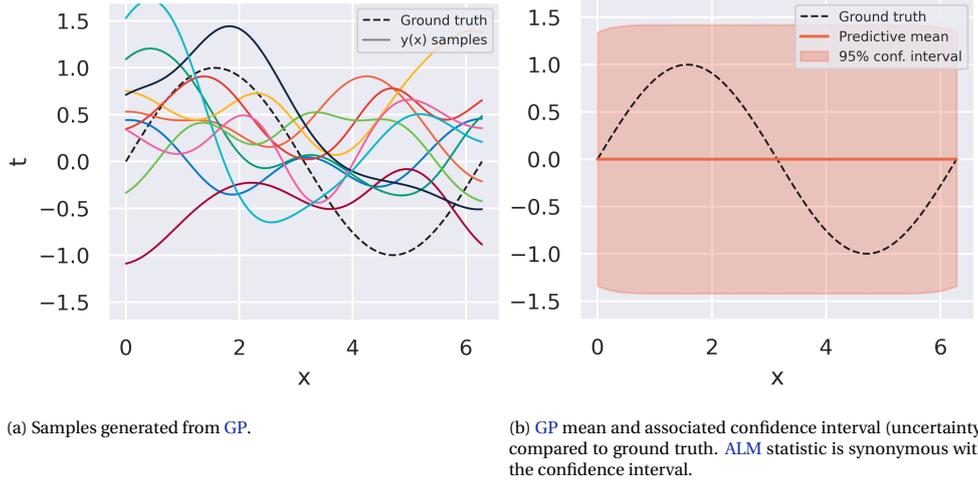


Figure 2.3: GP behavior under prior assumptions. Figures obtained from DSAIE [2024].

in the data, which the regression model needs to identify to provide an accurate estimate of the dependent variable(s). An example can be the locations, magnitudes, and sparsity of local maxima/minima.

A TGP works by recursively partitioning the input space by making binary splits, which are parallel to coordinate axes. To prevent a partition from occurring at every data point (overfitting), a meaningful prior for the size of data points within a tree is required. Starting with a tree that has no partitions, $v \in T$, the probability of a split can be calculated using eq. (2.17), where q_v is the depth of the tree, a and b are parameters that dictate the size and spread of the trees [Gramacy and Lee, 2009]. a and b can be assumed to be 0.5 and 2 respectively [Gramacy and Lee, 2009].

$$P_{split} = a(1 + q_v)^{-b} \quad (2.17)$$

For each v region, a hierarchical GP model is used which follows eq. (2.18) [Gramacy and Lee, 2009].

$$\begin{aligned} \mathbf{Z}_v | \boldsymbol{\beta}_v, \sigma_v^2, \mathbf{K}_v &\sim N_{n_v}(\mathbf{F}_v \boldsymbol{\beta}_v, \sigma_v^2 \mathbf{K}_v), & \boldsymbol{\beta}_0 &\sim N_m(\boldsymbol{\mu}, \mathbf{B}) \\ \boldsymbol{\beta}_v | \sigma_v^2, \tau_v^2, \mathbf{W}, \boldsymbol{\beta}_0 &\sim N_m(\boldsymbol{\beta}_0, \sigma_v^2 \tau_v^2 \mathbf{W}) & \tau_v^2 &\sim IG(\alpha_\tau/2, q_\tau/2), \\ \sigma_v^2 &\sim IG(\alpha_\sigma/2, q_\sigma/2), & \mathbf{W}^{-1} &\sim W((\rho \mathbf{V})^{-1}, \rho), \end{aligned} \quad (2.18)$$

Each v region contains data $D_v = (\mathbf{X}_v, \mathbf{Z}_v)$ (\mathbf{Z} is analogous to \mathbf{t} in eq. (2.13)). Linear trend coefficient $\boldsymbol{\beta}_v$ comes from an unknown mean $\boldsymbol{\beta}_0$ and region specific variance $\sigma_v^2 \tau_v^2$. N is a multivariate normal distribution. σ_v^2 is the variance with a Kernel \mathbf{K}_v . $\mathbf{F}_v = (\mathbf{1}, \mathbf{X}_v)$. \mathbf{W} is an $m \times m$ matrix, where m represents the number of independent variables and an intercept. IG and W represent the Inverse-Gamma and Wishart distributions respectively. All other variables are hyperparameters which can be treated as known (see section 2.5.3.4). For datasets with a small amount of noise, an improper prior, where $\mathbf{W} = \infty$ can be used to simplify eq. (2.18) [Gramacy, 2023].

In eq. (2.18), nothing prevents the mean of the overall model from being discontinuous at the location of a partition [Gramacy and Lee, 2009]. This can cause the highest uncertainty to be located along the partition, causing a sample near it to be simulated.

GP regression is more flexible than linear regression, but if the data portrays linear trends, a linear model is more efficient [Gramacy and Lee, 2009]. Therefore, eq. (2.18) can be augmented by using a limiting linear model. This is done by introducing a boolean which either selects a GP or a linear model for a partition or a dimension [Gramacy and Lee, 2009].

If no partitions occur, the additional hyperparameters and equations in eq. (2.18) increase the computational cost of a TGP when compared to a GP without bringing additional information. However, if a partition oc-

curs, splitting the input space reduces the main computational cost of using a GP, as it is proportional to the number of sample points that it is modeling ($O(N^3)$).

2.5.3.4. Interpolation Techniques

Only simulating a subset of the event set means interpolation techniques are required to get a value of consequences for the entire event set. In compound floods, this is often taken as linear: structured grid [Couasnon et al., 2022; Eilander et al., 2023c]; unstructured grid [Jane et al., 2022]). Different approaches are available: GP or TGP [Gramacy and Lee, 2009], Radial Basis Functions (RBFs) [Bishop, 2006; Jane et al., 2022], or a neural network [Hendrickx et al., 2023]. The main issue with these latter approaches is an additional computational cost is required to train these models.

When comparing interpolation techniques for compound floods, Jane et al. [2022] show linear scatter interpolation leads to the smallest Root Mean Square Error (RMSE). However, they did not consider a GP, TGP, or neural networks.

Linear interpolation requires limited additional computational cost as it is deterministic and has no hyperparameters. Its accuracy depends on the sparsity of simulated data points, which should be minimized.

As shown earlier, The GP or TGP needs to compute the posterior predictive surface. Additionally, a GP, and a TGP only have hyperparameters. To optimize these, two methods are available. For a GP, empirical Bayes (also known as Bayesian inference) [Bishop, 2006] can be used (fig. 2.4), where the log-likelihood is maximized (eq. (2.19), dependencies for σ_f and ℓ come from \mathbf{K}), which requires an iterative process [Bishop, 2006]. For a TGP, Gramacy and Lee [2009] use Markov Chain Monte Carlo (MCMC). A MCMC works by initializing a value for the hyperparameters and sampling from the distributions (eq. (2.18)). The sampling is repeated for a pre-defined number of steps [Gramacy, 2023]. The accuracy of the estimate for the hyperparameters is dependent on how many samples are used to represent the distribution.

RBFs have hyperparameters and weights associated with them. These weights need to be marginalized to reach a likelihood function that only depends on hyperparameters. Then empirical Bayes can be used to optimize the hyperparameters [Bishop, 2006].

$$\ln p(\mathbf{t}|\sigma_f, \ell, \beta) = -\frac{1}{2} \ln |\mathbf{K} + \beta^{-1} \mathbf{I}| - \frac{1}{2} \mathbf{t}^T (\mathbf{K} + \beta^{-1} \mathbf{I})^{-1} \mathbf{t} - \frac{N}{2} \ln 2\pi \quad (2.19)$$

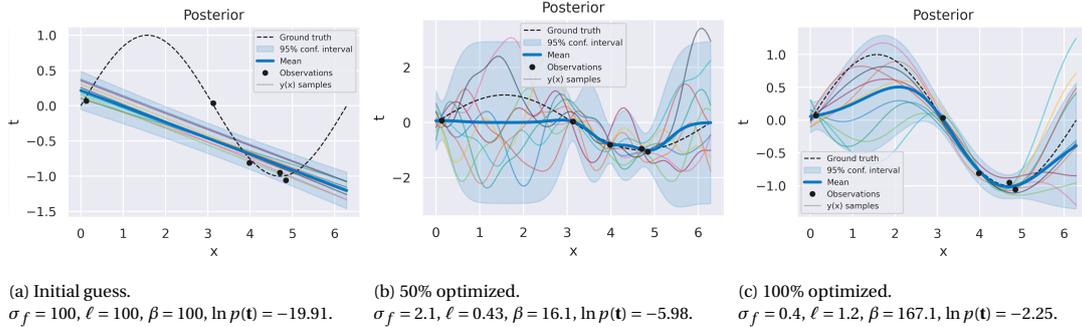


Figure 2.4: Applying empirical Bayes for a GP with a squared exponential kernel (eq. (2.16)) and 5 sampled points [DSAIE, 2024].

A neural network can take various forms. It can be used as a regression technique to estimate a small number of outputs [Ayyad et al., 2022; Hendrickx et al., 2023] or multiple outputs, such as pixels of an image [Bentivoglio et al., 2023]. Since they are affected by the curse of dimensionality, more weights are required to correctly map and understand all of the dimensions [Prince, 2023]. This exponential process makes it computationally expensive to tackle problems with a large dimensionality. The curse of dimensionality can be solved by applying inductive biases, which are prior assumptions of a trend/feature existing in the data. These reduce the number of weights required [Bentivoglio et al., 2023]. However, neural networks are not regularized. Neural networks are therefore not an attractive option for minimizing the computational cost. To ensure they do not overfit the data, they require the dataset to be split into two or three components: training, validation, and testing. For a small number of samples, this becomes problematic. Moreover, they

can require significant computational resources to train the models, as a large number of weights need to be optimized.

2.5.4. Closing Thoughts on Computational Time

The computational time can be reduced in three ways: improving computational resources, reducing the complexity of numerical models by increasing the number of assumptions, and using a sampling strategy coupled with an interpolation technique.

SFINCS has already been applied in the context of compound floods [Diermanse et al., 2023; Eilander et al., 2023c] and has shown that its accuracy is not equivalent to more complex numerical models [Leijnse et al., 2021]. Nonetheless, its reduction in computational cost makes it attractive to apply in compound flood risk assessments.

A TGP has shown its applicability in reducing the number of simulations to reach an accurate estimate of the dependent variable(s) [Gramacy and Lee, 2009; Hendrickx et al., 2023]. Its non-stationarity and heteroskedasticity make it more appealing than a GP. A computational cost is required to obtain the posterior predictive surface, which is only beneficial when a substantial number of samples are not required to simulate. This is dependent on the choice of a stopping criterion. Current literature has shown that stability-based methods are the most appealing. However, nothing indicates the best variable to monitor, this will need to be defined during the research. Finally, a TGP heavily depends on an initial prior, but Hendrickx et al. [2023] have shown that a MDA can be used to provide an initial set of data points, reducing the computational cost of using a TGP for a certain amount of samples.

2.6. Conclusion

In conclusion, this literature review has presented insights into the assumptions and tools required to model and quantify compound flood risk in estuarine areas.

Firstly, risk is defined as a function of hazard, exposure, and vulnerability. Hazard is defined by the probability of the event and its associated maximum water depth. The probability refers to the frequency at which a value of economic damages is exceeded. The maximum water depth can be modeled using a hydrodynamic model. Different options are available but SFINCS is preferred as it significantly reduces the computational cost of running simulations. However, its reduced complexity can introduce uncertainty in the quantification of flood hazard. Exposure and vulnerability can be modeled using an impact model. EAD can be used as a metric to perform a cost-benefit analysis on flood protection but requires quantification of risk for different RPs.

Secondly, a compound flood is defined as the interaction of multiple drivers leading to a larger flood hazard than if they occurred in isolation. These drivers are fluvial discharge, pluvial, coastal water level, and local meteorological effects. These drivers influence the flood hazard with their magnitude, duration, spatial pattern, and time lag with other drivers. Depending on the case study location, some of these drivers can be ignored as the flood hazard will not be sensitive to them. This simplifies the risk assessment in both the statistical modeling and overall computational cost. Different underlying weather patterns have distinct effects and they should be taken into account separately, but a small time record may make this difficult.

Thirdly, using an EVA, historical compound flood events can be identified. These should be applied to the magnitude of the dominant driver to ensure all identified events are expected to lead to an extreme hazard. A POT approach is preferred as it is expected to perform better on short-time records.

Then, to reduce the computational cost of numerically simulating events, an explicit method of modeling the dependence structure shows the most promise. These can create stochastic event sets of local boundary conditions reducing the modeling extent required. A vine copula can be used to model the dependence structure of the random variables associated with different drivers. Fitting a vine copula to a high-dimensionality dataset is difficult, as an enumeration problem makes it difficult to choose the vine copula that fits the data with the smallest error. To reduce the computational cost of using a brute force approach (fitting all structures and choosing the one with the smallest error), Dißmann's algorithm is available, but it cannot ensure that it fits the best statistical model.

Finally, using machine learning techniques such as a GP or TGP has been shown to reduce the number of samples required to reach an accurate estimate of the dependent variable(s). However, a model needs to be

trained after each simulation. This introduces a computational cost that is proportional to the number of sampled points. A stopping criterion will need to be defined in this research to minimize the computational cost.

3

Case Study and Data

For this research, a location affected by compound floods was required. Furthermore, these events preferably had to happen at a relatively high frequency, making it easier to model the dependence structure and validate hydrodynamic numerical models. Additionally, it is easier to correctly model the dependence structure if a reliable and long-term record of coinciding variables is available. This can be achieved by using local data sources in areas that have the resources to install and monitor the necessary measuring devices.

The case study is introduced in section 3.1. Information on the raw data related to flood drivers is given in section 3.2. A brief description of past historical compound flood event is provided in section 3.3. Finally, a description of the current numerical models is given in section 3.4.

3.1. Charleston

Charleston, South Carolina is located along the Atlantic coast of the United States. This region has been identified as an area prone to compound floods. These are driven by combinations of precipitation, surge, tide, waves, and discharge [Couasnon et al., 2020; Parker et al., 2023; Wahl et al., 2015]. Tropical cyclones often cause the largest amount of damage in this region as individual events generate co-occurring extremes in drivers which lead to compound flood events. In the Atlantic, tropical cyclones are known as hurricanes and are classified based on their wind speed [NOAA, 2012]. The hurricane season takes place between June 1 and November 30 and has return periods that range between 5 and 15 years for the South Atlantic coast [NOAA, 2021]. This has caused numerous numerical models to be validated in this region in an attempt to understand flood risk and reduce the annual expected damages by improving flood protection infrastructure.

The Atlantic coast is also affected by extra tropical storms that occur during the winter. These events have a return period that is approximately one year and generate significant storm surges that often coincide with significant precipitation, which are not necessarily extreme. These cause more frequent but less severe compound flood events, which should still be accounted for [Parker et al., 2023].

The United States also has sources of reliable local data, which are open source. National Oceanic and Atmospheric Organization (NOAA) have gauges in important cities across the United States that measure water level, wind, pressure, and temperature. They are also able to provide an estimate for the tide by performing a spectral analysis on the 37 most important constituents [NOAA, 2019]. United States Geological Survey (USGS) provide data related to the quantity, quality, distribution, and movement of surface and underground water [USGS, 2009]. Gauges can be placed along relatively important rivers, which provide information on variables such as discharge and precipitation.

The city of Charleston is situated on the estuary created by the tidally dominated Ashley and Cooper Rivers, and its historical center is located approximately 8 kilometers from the tidal inlet. Since its settlement, it has enabled its growth through land reclamation of wetlands, normally flooded during high tides [Butler, 2020]. This has placed the population of the city at risk of floods as these reclamations are low-lying pieces of land, which are affected by relative sea level rise [Holtz et al., 2014]. EVA on total water levels demonstrate that return periods of events have decreased significantly over the past 70 years [Diermanse et al., 2023], while

flood defense/protection measures have not been improved to accommodate these changes [Holtz et al., 2014].

Upstream of Charleston, the Pinopolis Dam, located along the Cooper River was constructed to supply rural communities in South Carolina with electricity. This effectively removed the natural variability of the Cooper River, as flow is predominantly dictated by the demand for electricity. This also had the indirect effect of reducing the effect of compound floods as extremes in discharge caused by intense precipitation in the upstream catchments can be stored in Lake Moultrie, which is the dam reservoir.

3.2. Data Related to the Time Records of Drivers

This section presents the sources, and locations (fig. 3.1) of data to obtain the time record of the different compound flood drivers identified in chapter 2. These are fluvial discharge, pluvial, coastal water level, and local meteorological effects. In some cases, it also explains why the drivers are not investigated further. The combined time record of the considered compound flood drivers is 24 years and 4 months.

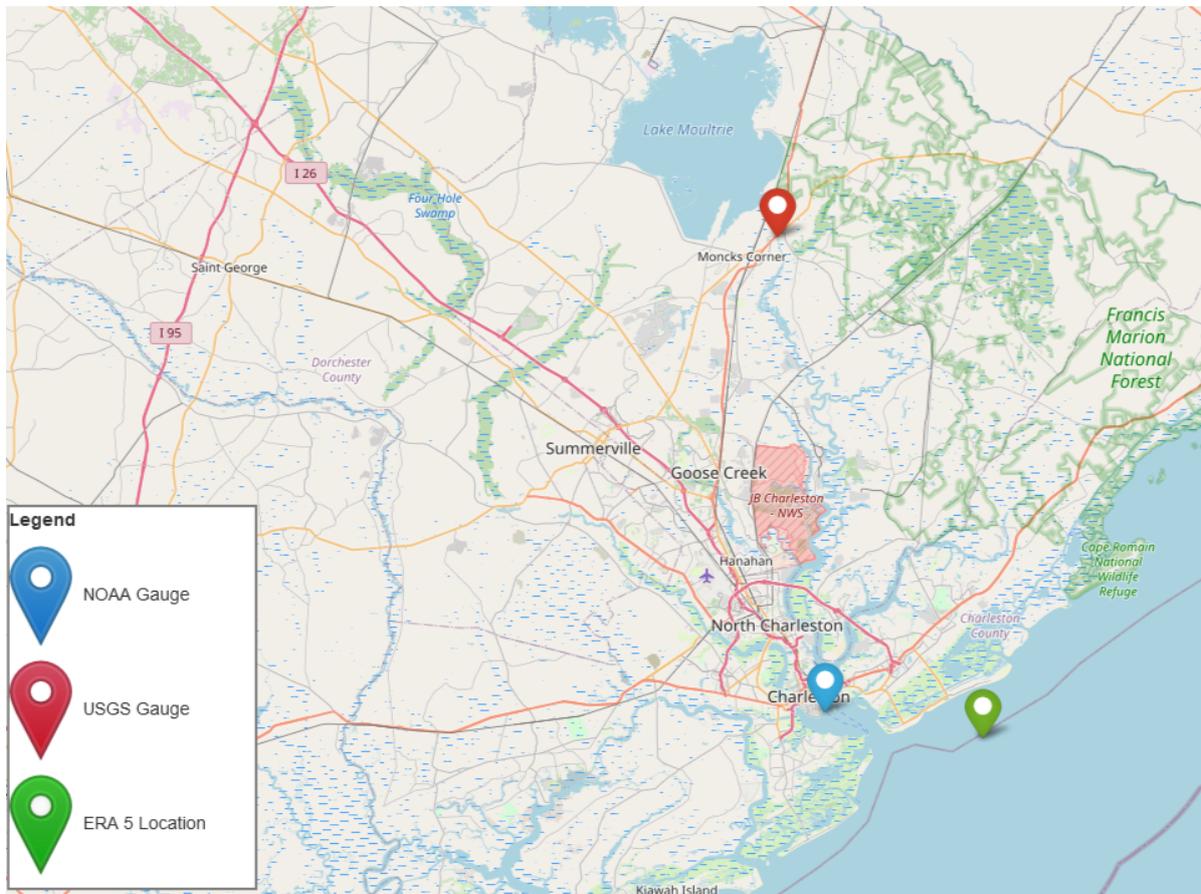


Figure 3.1: Gauge locations used to obtain the time record of compound flood drivers.

3.2.1. Downstream Water Level

Information on the tide and water level can be obtained from NOAA. The gauge is located in Charleston Harbor near the city center. This is the blue marker in fig. 3.1. The tidal time record is computed using a harmonic analysis and is stationary. NOAA only includes 37 components, none of which have a period longer than a year.

Data on water level is available from 1921 to the present day with a sampling rate of 1 hour. The data is non-stationary and needs to be detrended to apply an extreme value analysis. This is caused by the yearly trend of MSL as well as the seasonality in MMSL. The total water level time record contains the effect of tides, sea level rise, and a non-tidal residual component (see section 2.2.3.2 for the definition of the non-tidal residual

component).

3.2.2. Upstream Discharge

Information on the historical discharge of the Cooper River can be obtained from the [USGS](#). The gauge is located directly downstream of Lake Moultrie. Data is available from October 1998 onwards, and there are periods of missing data, which can last on the scale of months. Data is measured at 15-minute intervals in cubic feet per second.

3.2.3. Precipitation

Local data for precipitation is available, but only until 2007. ERA5 reanalysis will therefore be used instead. Using the HydroMT data catalog [[Eilander et al., 2023a](#)], ERA5 data can be obtained from January 1979 to March 2023. The sampling rate is 1 hour.

ERA5 has a coarse resolution for its precipitation reanalysis, with a grid based on 0.25° arcs on longitude and latitude, meaning precipitation is not directly measured at Charleston. A choice therefore has to be made on where to measure precipitation. One option is to take the spatial average [[Eilander et al., 2023c](#)]. However, this could diffuse the magnitude of a historical event if the peak occurs at slightly different time steps, underestimating the actual magnitude of the event. Another option is to take data from a single coordinate and assume it represents all the other coordinates.

The second option reduces peak attenuation and is more conservative. To choose the coordinate, [NOAA](#) creates return period maps of precipitation (fig. 3.2), which show locations near the open coast are susceptible to larger magnitudes of precipitation. Therefore, to be conservative, the coordinate with the smallest Euclidean distance to Charleston, whilst also located near the open coast was chosen. With the resolution of ERA5, this resulted in 32.75° North, 79.75° West, which can be seen in fig. 3.1.

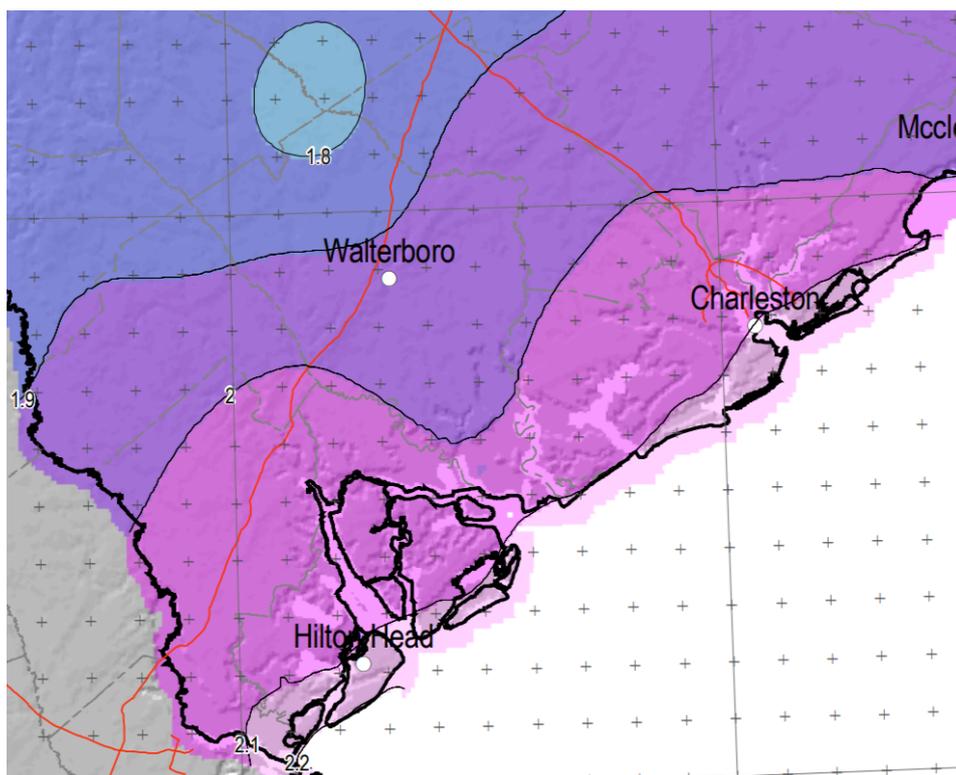


Figure 3.2: 2-year return period map for rainfall from the NOAA. '+' markers represent differences of 0.1° in longitude and latitude. Isolines create bounding regions of equal precipitation magnitude. The SFINCS model region can be encapsulated within three of these regions. In metric, the difference between the return values of regions with the largest and smallest return values is approximately 10 millimeters per hour.

3.2.4. Wind and Pressure

Data related to local meteorological conditions can be obtained locally from various sources (such as the NOAA). However, similar to precipitation the time record is too short. Moreover, based on geometric constraints and the size of the estuary (fig. 3.1), it is assumed that wind and pressure setup will have negligible effects on flood risk, and this will not be investigated further. This is advantageous as it reduces the dimensionality, but is also a limitation as it does not provide the best representation of compound flood risk.

3.3. Historical Compound Floods

For the coinciding time record of available data between October 1998 and March 2023, two hurricanes had a noticeable effect on the non-tidal residual water level in Charleston. Matthew in October 2016 and Irma in September 2017.

Hurricane Matthew caused the largest skew surge at the NOAAs gauge for the record of coinciding data. Accompanied by a significant amount of precipitation, Charleston County suffered from widespread floods and power outages. In South Carolina, Matthew is estimated to have killed 4 individuals and caused 249.6 million USD in damages [Wikipedia, 2023].

Hurricane Irma caused the second-largest skew surge at the NOAAs gauge over the same time record. Accompanied by a large higher high tide and extreme precipitation, multiple roads were inundated. However, the short duration of the skew surge meant that little damage was caused by the event [Wikipedia, 2024].

3.4. FloodAdapt

The numerical models described in this section were obtained from FloodAdapt, an open-source graphical user interface that combines accurate numerical modeling with an interpretable user interface to show the effect of land use patterns and climate change on flood risk [Deltares, 2024]. Charleston was used as a pilot study to show the different applications of the software. It combines a hydrodynamic model (section 3.4.1) and an impact model (section 3.4.2).

3.4.1. Hydrodynamic Model

During a previous study, a SFINCS model (fig. 3.3) was validated for the county of Charleston [Diermanse et al., 2023]. In summary, the computational cost was minimized by reducing the model extent. Since computational time is correlated to the grid size, a 200-meter grid was used. A sub-grid can be used to improve the accuracy of the model while limiting the increase in computational expense. Levees and dunes are modeled as weirs. Two pumps are included to represent existing infrastructure. Finally, an infiltration curve is also included.

For validation, four scenarios were evaluated against historical water levels using an R-squared metric (R^2) against the NOAAs gauge in Charleston Harbor [Diermanse et al., 2023]. The closer R^2 is to 1, the better the model performs. The findings related to three of these scenarios can be found below. There is little relevance in reporting the outcomes of the fourth scenario (hurricane Joaquin) as the results were affected by a bug, which has since been resolved [Diermanse et al., 2023].

1. Long-term tide and surge water levels: The model achieved $R^2 = 0.96$ for tide and $R^2 = 0.64$ for surge.
2. King Tide event: The model achieved a value of $R^2 = 0.92$. It also correctly predicted which roads were inundated.
3. Hurricane Irma: The model overestimated the water levels by approximately 0.5 feet as it does not take into account seasonality in mean sea level. During the hurricane 166 homes were flooded. In the model, 162 of these homes were located in inundated areas.

While the model can be considered validated, it is important to mention that it does not represent reality exactly. This is caused by a multitude of factors such as model choice, accuracy of data, and model schematization. Moreover, for scenario 3, Diermanse et al. [2023] only mentions the ratio of true positives to actual positives. However, there is no consideration of false positives, which should occur if the water levels are overestimated by 0.5 feet, leading to an overestimation of flooded homes.

Additionally, water level boundaries are also forced with water level data from the NOAA gauge, located in the Charleston Harbor. Since the gauge is not located on an open coast, the water level is affected by river

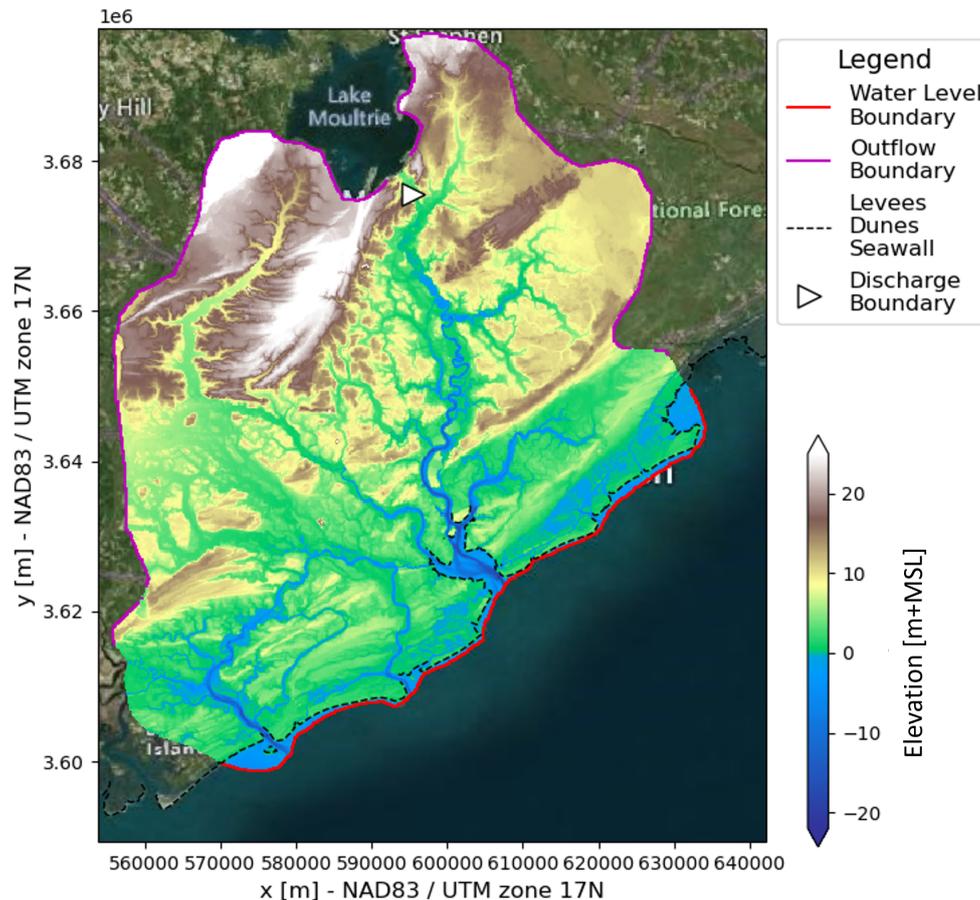


Figure 3.3: SFINCS model showing water level boundaries, outflow boundaries, levee/dune/seawall locations, discharge boundaries modeled as point sources, and the digital elevation model [Diermanse et al., 2023; Eilander et al., 2023a].

discharge, local meteorological effects, and tidal amplification, not providing the best representation of the water level at the boundaries. Another limitation of this model is concerning the types of failure. Currently, flood hazards can only be caused by overflow. It disregards other levee failures such as wave overtopping, piping, or micro instability, which reduces the risk associated with a flood event.

3.4.2. Impact Model

The DELFT-Fast Impact Assessment Tool (FIAT) model combines a flood hazard map, exposure, and vulnerability data to obtain consequences related to an event. These consequences can be economic or social. The model is only defined for Charleston County [Diermanse et al., 2023], which means that the impact model is only defined in the downstream region of the SFINCS model (fig. 3.3). The model uses 224 census block groups to subdivide the model. Damages occur if the water level at the centroid of the footprint of a building exceeds its ground elevation. Damage for a building is calculated by combining the maximum water depth (hazard), maximum damage (exposure), and depth damage curves (vulnerability). SFINCS calculates the spatially varying maximum water depth. Exposure is defined by the National Structure Inventory (NSI) [Diermanse et al., 2023]. Vulnerability is defined by various data sources based on the category of the modeled building [Diermanse et al., 2023]. The model includes roads, but these do not have depth-damage curves associated with them, they are only used to show accessibility constraints, a metric to measure social consequence.

4

Methods

Four distinct tools are required to perform a compound flood risk assessment. Firstly, economic damages from flood events need to be modeled (section 4.1). This requires a hydrodynamic model (section 4.1.1) and an impact model (section 4.1.2). Secondly, a statistical model is needed to model the dependence between the different random variables (section 4.4.1), allowing for the generation of representative stochastic events (section 4.4.2), which are used as the hydrodynamic model's boundary conditions (section 4.4.3). To fit the statistical model, historical observations of past flood events are needed (section 4.3). This requires preprocessing the time records of the different drivers (section 4.2). Then, the dominant driver is identified (section 4.3.1), on which an Extreme Value Analysis (EVA) is performed (section 4.3.3), and corresponding random variables are defined (section 4.3.2). This generates a historical event set (section 4.3.4). Thirdly, a surrogate damage model is required to reduce the computational cost of simulating all stochastic events generated by the statistical model (section 4.5). Two different approaches are investigated: a priori (section 4.5.1) and a posteriori (section 4.5.2). The a priori approach is the current state-of-the-art recommended by Jane et al. [2022]. It consists of sampling with a Maximum Dissimilarity Algorithm (MDA) (section 4.5.1.1) and using linear scatter interpolation (section 4.5.1.2). The a posteriori approach consists of sampling and surrogate damage modeling with a Treed Gaussian Process (TGP) (section 4.5.2.2). This requires an initialization (section 4.5.2.1), a stopping criterion (section 4.5.2.3), and an explanation of how to sample and model multiple outputs (section 4.5.2.4). Finally, once a sufficiently accurate estimation of economic damages for all stochastic events is obtained, risk can be modeled (section 4.6) once it is defined (section 4.6.1). This allows for a statistical analysis (section 4.6.2) to compare different sampling and surrogate approaches, and the effect of including additional random variables. It also allows to understand which driver is contributing the most to the economic damages (section 4.3.1). The overall method is summarized in fig. 4.1.

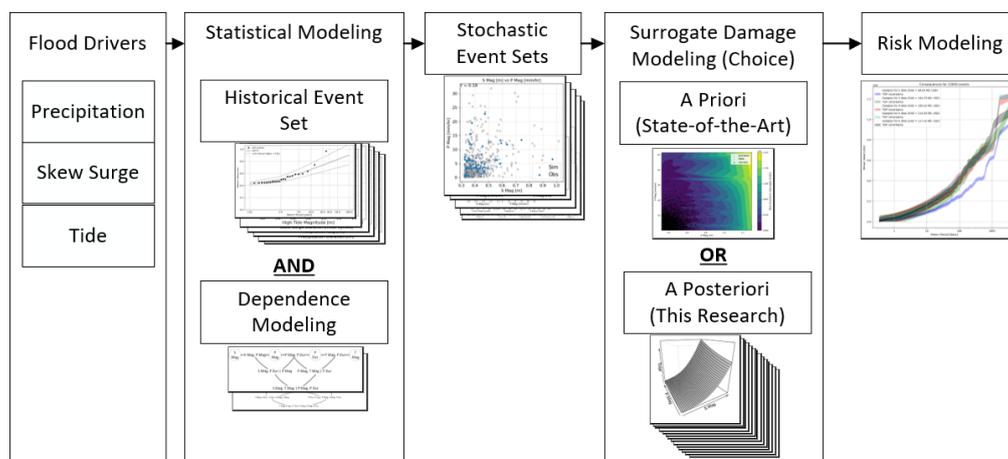


Figure 4.1: Overall methodology, showing the different tools needed to model risk. The damage (hydrodynamic and impact) model is included in the surrogate damage modeling approaches.

4.1. Damage Modeling

The hazard (section 4.1.1) and impact (section 4.1.2) numerical models are required to obtain the damages related to (compound) flood events. These models were validated in a previous study [Diermanse et al., 2023]. Only small changes were made to the impact model (section A.2). Here a description of the workflows of the models is given to explain how an economic damage value can be obtained from boundary condition inputs (fig. 4.2). Once the time series of the boundary conditions is known, it is provided to the *SFINCS*, which computes the maximum water depth for each sub-grid cell (section 4.1.1). This is known as the hazard map. This is then used as an input for *DELFT-FIAT* model, which combines it with the ground elevation, depth-damage curves, and the maximum damages associated with the buildings located in the model domain to obtain the economic damages associated with a flood event (section 4.1.2).

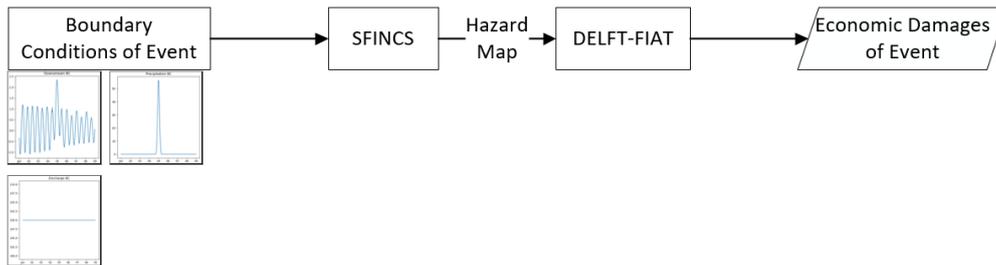


Figure 4.2: Damage modeling. Workflow used to reach total economic damages for a flood event.

4.1.1. Hazard Modeling with SFINCS

SFINCS requires a set of boundary conditions to calculate the maximum water depth of each grid cell, which is used as the hazard map. A boundary condition for upstream discharge and downstream water level is required, for precipitation it is optional. These boundary conditions need to be defined for each time step and at specific locations. Here it is assumed that the coastal water level and precipitation are spatially uniform to reduce the number of dimensions needed to characterize a driver. For the former, the water level is uniform along the downstream boundary. For the latter, the precipitation is uniform over the entire model domain (conservative according to fig. 3.2). It is also assumed that only the discharge of the Cooper River is important. Solving the governing equations (section A.1) allows *SFINCS* to compute the maximum water depth for each sub-grid over the entire model domain (fig. 3.3), which is used as the hazard map.

4.1.2. Impact Modeling with DELFT-Fast Impact Assessment Tool (DELFT-FIAT)

DELFT-Fast Impact Assessment Tool (FIAT) allows for the quantification of the exposure and vulnerability of a flood event once the hazard map is known [Diermanse et al., 2023]. The model requires three inputs. Firstly, a flood map (hazard) containing maximum water depths at all grid cells (provided by *SFINCS*). Secondly, ground elevation and depth-damage functions (vulnerability curves) dictate the damage fraction associated with a particular building (or grid cell), and water depth. These curves are often non-linear. Finally, maximum potential damages (exposure) associated with a particular building (or grid cell). When hazard, vulnerability curves, and exposure are combined, the economic damages of a building (or grid cell) can be obtained. Summing all buildings (or grid cells) together provides an estimate for total economic damages. This workflow can be visually seen in fig. 4.3.

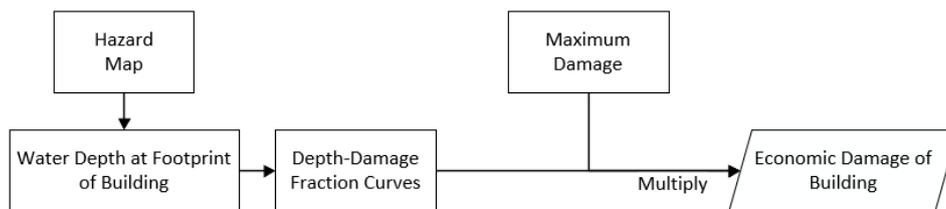


Figure 4.3: DELFT-FIAT workflow.

While the model for Charleston was validated in a previous study [Diermanse et al., 2023], some buildings were constantly flooded and had to be removed (section A.2).

The current DELFT-FIAT model is only defined for Charleston County. This first model will be called the complete model. All the buildings within the model are associated with census block groups, which are used to split the complete model. Each census block group is attributed to a sub-county according to the Census County Division [United States Census Bureau, n.d.]. These 11 sub-counties are used to define a second model which is called the sub-county model (fig. 4.4). Then, according to these sub-county definitions, a third model is defined by grouping sub-counties according to their proximity to the open coast. These groups can be classified as either coastal or inland. This third model will be called the classified model (table 4.1). The reasons for these different models will become apparent when introducing the stopping criterion for the TGP (section 4.5.2.3).

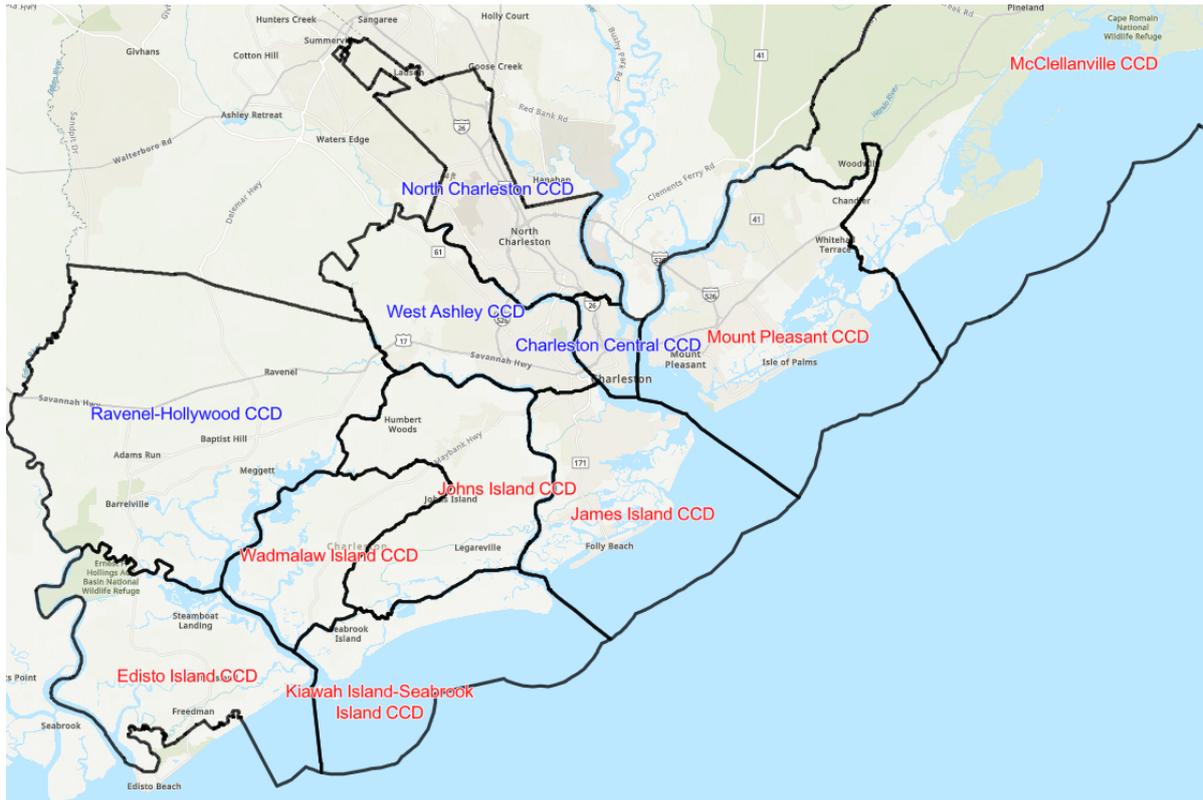


Figure 4.4: Sub-county model according to the Census County Division [Esri Federal Datasets, 2024]. The colors of names are representative of their classification in table 4.1: red represents coastal and blue represents inland.

Table 4.1: Classified model according to the proximity to the open coast of each sub-county.

Coastal	Inland
Edisto Island	Ravenel-Hollywood
Kiawah Island-Seabrook Island	West Ashley
Wadmalaw Island	North Charleston
Johns Island	Charleston Central
James Island	
Mount Pleasant	
McClellanville	

4.2. Preprocessing the Time Records of Different Drivers

In this section, the data preprocessing steps taken to reach the time records of the different flood drivers are explained. It is necessary to avoid introducing any bias in the quantification of past flood hazards. This will

skew the dependence and the magnitude of the stochastic events.

The first step in preprocessing is to ensure that all datasets are based on the same timezone. This is important for the time lag between drivers. The Universal Coordinated Time (UTC) is used as the switch to daylight saving time does not affect it. Moreover, local time is not needed for the main analysis as stochastic events will be generated (section 4.4.3). Then, the total water level, upstream discharge, and precipitation time records are further preprocessed.

For the total water level, three components are decoupled: the tide, the Mean Sea Level (MSL), and the non-tidal residual. These components can be linearly superposed to obtain the total water level. This provides a higher accuracy when schematizing the dependence structure, as the tidal magnitudes are bounded by their constituents and are not weather-related, skewing the correlation with different drivers when identifying extremes. Therefore, the EVA should only be performed on the non-tidal residual to provide a more meaningful correlation (this assumption is confirmed in section D.3.1). Since the time record is non-stationary, the trend in the MSL should first be removed to apply an EVA. Here, a 1-year moving average is used [Arns et al., 2013]. This analysis has the undesired effect of removing tidal periods that are longer than a year from the analysis. It also provides the value of the MSL component to use when linearly superposing different components to obtain the total water level, which is 0.2 m (section B.1). The concept of skew surge [Williams et al., 2016] is used to represent the non-tidal residual (fig. 2.1). Here, a tidal cycle is defined as occurring from trough to trough. The use of skew surge causes the skew surge duration to occur in 12-hour increments for the case study of Charleston (section B.1).

The upstream discharge is converted from imperial to metric. The data is resampled from 15-minute to hourly intervals by taking the maximum discharge within that hour. Since there are periods of missing data that can last for months, it is difficult to find a suitable interpolation method. For simplicity, a backfill interpolation is used, which means the last observed value is repeated until another observed value is reached. The discharge time record is affected by hydropeaking, which makes it difficult to apply an EVA on (section B.3). However, the findings in chapter C (summarized in section 4.3.1) meant this did not have to be solved for the case study of Charleston.

ERA5 provides its data for precipitation as meters per time step. Therefore, for hourly time steps, the time record is divided by 1,000 to reach units equal to millimeters per hour (recognized by SFINCS). As mentioned in chapter 2, ERA5 under-predicts extremes, and this is also the case for Charleston (section B.4 and section D.2).

4.3. Defining a Historical Event Set for Compound Floods

In this section, a description is given of how to identify historical flood events. For each flood event, random variables that define simplified parametric boundary conditions (section 4.4.3) are collected. This step is crucial in fitting a statistical model to generate stochastic events (section 4.4). Small differences in the methodology used to identify flood events and/or collect the random variables can have larger differences in the stochastic events generated, which will influence the estimate of flood risk.

The short time record of 24 years and 4 months means Peak Over Threshold (POT) is preferred to increase the number of identified extremes. To ensure the marginal Probability Distribution Functions (PDFs) are valid for the entire joint probability distribution, a dominant driver should first be identified (section 4.3.1), which requires a sensitivity analysis. Then, the co-occurring random variables (magnitudes, durations, and time lags) and their correlation should be defined (section 4.3.2). Finally, section 4.3.3 shows how POT is applied to the dominant driver. The threshold is chosen based on a univariate and multivariate sensitivity analysis. Once a stable threshold is defined, the historical event set is obtained (section 4.3.4). This workflow can be visualized in fig. 4.5.

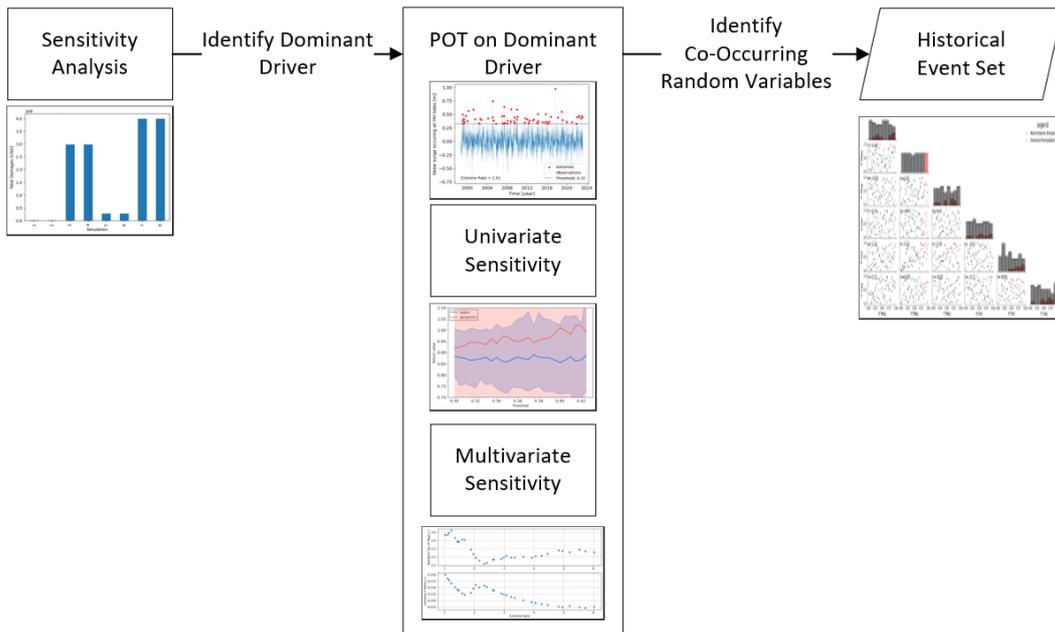


Figure 4.5: Workflow to reach historical event set.

4.3.1. Identifying the Dominant Driver for Extreme Value Analysis

To identify the dominant driver, a sensitivity analysis of the different driver magnitudes on the total economic damages of the complete *FIAT* model is performed. For each driver, two magnitudes are used: the maximum historically observed value, and the average value over the time record. This creates a grid of 2^3 simulations. The dominant driver is defined as the driver that causes the largest difference in total economic damages when it is considered extreme when compared to the event where all drivers are not extreme. For Charleston, skew surge is identified as the dominant driver, and discharge is found to not contribute to compound floods and is omitted from the rest of the research (chapter C).

4.3.2. Identifying Co-Occurring Random Variables and Defining Their Correlation

Once the dominant driver is defined, *POT* (section 4.3.3) identifies all extreme events for that driver. For each event, co-occurring random variables associated with the different drivers need to be identified. These random variables define the parametric model used to define the boundary conditions for the drivers (section 4.4.3). The choice of random variables to include is based on [Eilander et al. \[2023c\]](#). They use non-tidal driver magnitudes and recommend including duration and time lag. For this research, the tidal driver is also included as a random variable. Once all random variables are obtained, their correlation can be investigated, but this requires the definition of a correlation coefficient. The correlation between the different variables influences the dependence structure of the statistical model.

To define the random variables for each event, a window of ± 3 days is centered on the time at which the maximum magnitude of the dominant driver occurs. This window length is chosen as historical compound floods for the case study have shown to occur within a small time lag of each other. Other non-tidal driver magnitudes are taken as the maximum magnitude occurring within the window. The lag of these drivers is defined as the difference in time (hours) between the dominant driver and the other driver's maximum magnitudes.

The duration of non-tidal drivers is taken as the maximum of: 1) the length of time the magnitude is above a quantile or 2) 6 days (window length). This definition ensures two things. Firstly, the duration is bounded. Secondly, the duration will not be affected by the trivial solution of always being the upper bound. The latter could be caused by measurement inaccuracies or processes that affect the drivers on a time scale that is larger than 6 days. Therefore, the choice of quantile is dependent on the driver and its time record. For skew surge, this is defined as the largest upper quantile of the *MMSL*, which is 0.2 meters (section B.1). For precipitation, it is defined as 0.3 millimeters per hour.

The tidal driver can be defined solely by its magnitude. For Charleston, the large daily inequality of the tide may cause the linear summation between a lower high tide and a skew surge to be dwarfed by a higher high tide with no skew surge. This can occur at a range of tidal magnitudes (section B.2). To prevent this when identifying extremes in skew surge, the extremes in skew surge magnitudes are restricted to Higher High (HH) tides. This has little effect on the quantification of extreme water levels (section D.1). For an event, the tidal magnitude is defined as the magnitude of the tide which coincides with this definition of skew surge.

Kendall's τ is used to define the level of correlation between bivariate pairs of random variables. Since it assesses rank, it can understand non-linear dependence. Here τ_B (eq. (2.4)) is used [Virtanen et al., 2020] as it takes into account ties in ranks. The statistical significance of the correlation can be assessed with the associated p-value of τ_B (test for independence). Another advantage is that the absolute empirical Kendall τ is used to fit the regular vine [Nagler and Vatter, 2023]. Therefore, observations made in the correlation analysis should translate to the pair copulas and regular vine used to define the vine copula.

4.3.3. Extreme Value Analysis on Dominant Driver

POT is used to identify extremes in the magnitude of the dominant driver. With the definitions of co-occurring random variables (section 4.3.2), the historical event set can be obtained. To fit the POT model, the HydroMT package [Eilander et al., 2023a] is used. It utilizes l-moments to fit exponential and generalized Pareto distributions and chooses the distribution based on the smallest Bayesian Information Criterion (BIC) score (eq. (2.2)). The BIC score is chosen over the AIC or AICc as it penalizes overfitting models for small datasets.

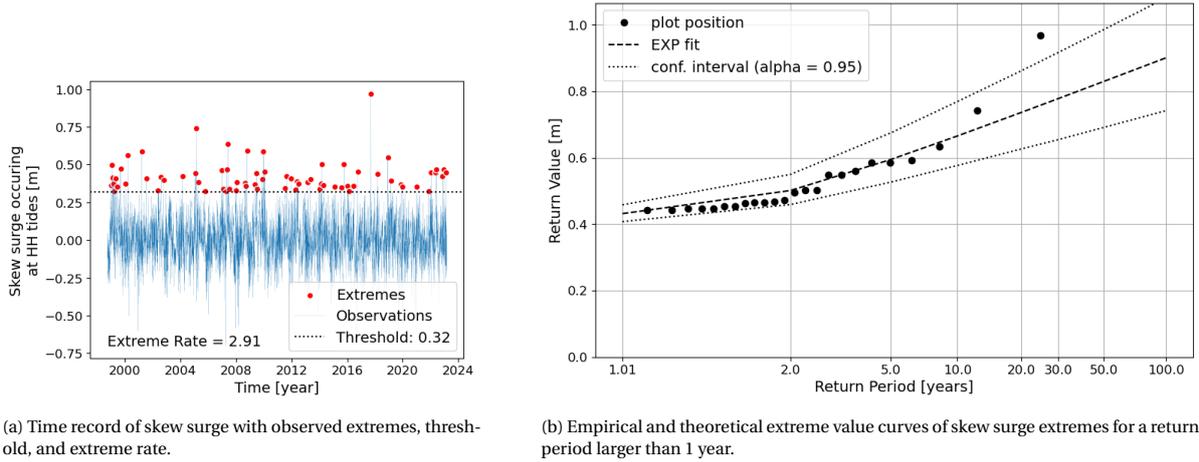
POT depends on two parameters that decide which extremes to include in the model. These are the declustering time and the threshold. The declustering time is taken as a week, which is a conservative estimate, as it is equivalent to the average duration of observed hurricanes [NOAA, 2021]. The choice of a suitable threshold is more complex, as it dictates the number of events to include in the historical event set. Different methods exist to define the threshold, but both are based on the sensitivity of a metric concerning the threshold.

Firstly, the univariate sensitivity can be investigated using mean residual life and return value stability plots [Bocharov, 2023]. The mean residual life plot shows for different thresholds the average excess of all events identified by POT above a threshold. Thresholds are stable if the mean excess is approximately linear. For different thresholds, the return value stability plot shows the stability of the return value for a given RP. Here the 100-year RP is used. Thresholds are stable if there are little changes in the return value.

Secondly, multivariate sensitivity can be investigated using the extreme driver ratio and the correlation between dominant and other driver magnitudes for different thresholds (or extreme rates). The former is defined as the ratio of events identified by a POT for other drivers included in the events identified by the threshold for the dominant driver to the total number of events. The latter is defined as the value of τ_B between the dominant driver magnitude and other driver magnitudes. In both of these cases, a period of stability is desired.

To reach the final event set with d dimensions (or random variables) for d_{driver} drivers, multivariate sensitivity is first used to identify a range of extreme rates where the metrics are stable. Then, combined with univariate stability on the dominant driver, a threshold for the dominant driver is chosen. This results in a threshold of 0.32 m for skew surge, equivalent to an extreme rate of 2.91 events per year (chapter D). The threshold corresponds to the dotted horizontal line in fig. 4.6a.

The identified extremes (red) are shown in fig. 4.6a alongside the time record of the skew surge. Figure 4.6b shows the exponential extreme value curve and the 25 events which have the largest skew surge magnitudes. One clear outlier in fig. 4.6b corresponds to Hurricane Matthew, causing the largest observed skew surge to not match the behavior of the tail of the exponential curve. A decision is taken not to fit the tail of the skew surge extremes. There are two reasons for this. Firstly, it was decided to fit statistical models to tropical and extra-tropical events. Therefore, fitting the tail would prioritize tropical cyclone events, affecting the statistical behavior expected from extra-tropical events. Secondly, because of the short time record, the largest observed event may have a RP of 25 years. If a longer record were available, this return value could correspond with a larger RP.



(a) Time record of skew surge with observed extremes, threshold, and extreme rate.

(b) Empirical and theoretical extreme value curves of skew surge extremes for a return period larger than 1 year.

Figure 4.6: Output of Extreme value analysis of skew surge using POT with a threshold of 0.32 m (extreme rate of 2.91 events per year). Performed using the HydroMT package [Eilander et al., 2023a] and code obtained from Eilander et al. [2023c].

4.3.4. Historical Event Set

The identification of skew surge extremes (fig. 4.6) and the definition of co-occurring random variables (section 4.3.2) allows for a representation of the historical event set and its correlation (section E.1). The historical event set contains 3 drivers which are defined by 6 random variables: Skew surge Magnitude (**S Mag**), Precipitation Magnitude (**P Mag**), Tidal Magnitude (**T Mag**), Precipitation Duration (**P Dur**), Skew surge Duration (**S Dur**), and Precipitation Lag (**P Lag**). The correlation analysis (section E.1) shows that four bivariate pairs have significant correlation. Here they will be presented in order of descending value of τ_B : **P Mag** and **P Dur**, **S Mag** and **P Lag**, **T Mag** and **P Dur**, and **S Mag** and **P Mag**.

4.4. Generating Unseen Compound Flood Event Boundary Conditions

This section will produce different stochastic event sets, which will contain the parameters used to define the time series of the boundary conditions for the different flood drivers. Stochastic event sets are necessary to produce a large number of events to obtain the return values of economic damages for unobserved RPs. Different stochastic event sets are necessary for 2 different reasons. Firstly, it allows the investigation of additional random variables on flood risk. If stochastic event sets have less than 6 random variables, constants will be used to define the parameters for the non-random variables. Secondly, it allows the generation of training and test event sets. Training event sets will contain 500 samples and will be applied to the lowest and largest number of dimensions. Their purpose is to define a stopping criterion (section 4.5.2.3) and calculate the accuracy of the surrogate models. Other stochastic event sets will be termed the test datasets and will be applied for all the different numbers of dimensions and will contain 10,000 events. Their purpose is to investigate the effect of including additional random variables on the computational cost, the estimate of risk, and the driver dominating the largest economic damages.

The workflow (fig. 4.7) used to generate stochastic events and boundary conditions for fig. 4.2 requires the historical event set as a starting point (section 4.3). Firstly, the statistical model (vine copula) is created. A vine copula requires a regular vine, and pair copulas to model the dependence (section 4.4.1). This is done multiple times to represent different levels of input reduction. Each model generates at least one stochastic event set (section 4.4.2). A vine copula also requires the CDFs of the random variables to transform stochastic events from the ranked to the variable space (section 4.4.2.1). Secondly, if a stochastic event set has less than 6 random variables, a constant needs to be defined for the missing random variables (section 4.4.2.2). Then, a description is given to explain how the time series of the boundary conditions are created for events within the different event sets (section 4.4.3). An example of an event set can be found in section 4.4.4.

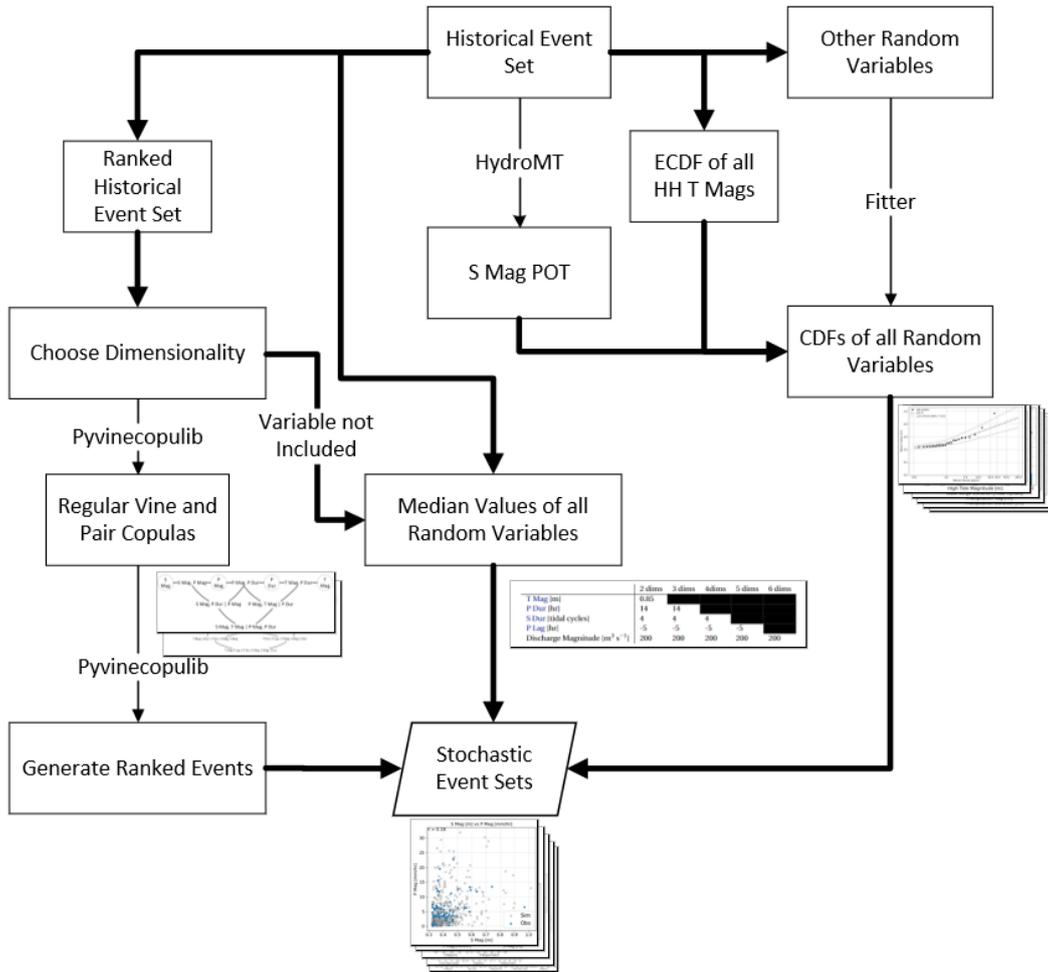


Figure 4.7: Workflow used to reach stochastic event sets. Connectors with less weight represent the use of Python packages.

4.4.1. Modeling the Dependence Structure Between Random Variables

A vine copula is fitted 4 times. For simplicity, a copula is also applied in two dimensions. These are applied in the univariate ranked space of the historical event set (generated in section 4.3). For continuous random variables with a coarse resolution (such as *S Dur*), many data points occupy the same rank. To prevent a bias when fitting the vine copula, a small amount of Gaussian noise is added to all observations.

Random variables are added sequentially to the vine copula. The starting point is the (vine) copula generated by the non-tidal driver magnitudes. This results in a copula fitted to *S Mag* and *P Mag*. The choice of variable to add is a subjective one. Here, variables are added based on the following order: *T Mag*, *P Dur*, *S Dur*, and *P Lag*. Therefore when increasing the number of random variables (or dimensions), the combination of variables follows table 4.2.

Table 4.2: Combination of random variables when increasing the number of dimensions (dims) for a stochastic event set.

Number of Dimensions (or Random Variables)	Random Variables Included
2 dims	<i>S Mag</i> , <i>P Mag</i>
3 dims	<i>S Mag</i> , <i>P Mag</i> , <i>T Mag</i>
4 dims	<i>S Mag</i> , <i>P Mag</i> , <i>T Mag</i> , <i>P Dur</i>
5 dims	<i>S Mag</i> , <i>P Mag</i> , <i>T Mag</i> , <i>P Dur</i> , <i>S Dur</i>
6 dims	<i>S Mag</i> , <i>P Mag</i> , <i>T Mag</i> , <i>P Dur</i> , <i>S Dur</i> , <i>P Lag</i>

The regular vine and copula pairs are chosen using the pyvinecopulib python package [Nagler and Vatter, 2023], which uses the absolute empirical Kendall's τ as a weight for Dißmann's algorithm (see section 2.4 for background on the theory). To simplify the regular vine and copula pairs, two additional steps are taken. Firstly, the independent pair copula is forced if the absolute empirical Kendall's τ is less than 0.05 [Eilander et al., 2023c]. Secondly, if all pair copulas associated with a random variable are independent, it is removed from the regular vine.

The result of this process can be found in chapter F, where they are compared to the correlation analysis provided in section E.1. The independence of certain variables (section E.1) resulted in 1 copula (2 dimensions) and 2 vine copulas (4 and 6 dimensions) being fitted.

4.4.2. Generating Stochastic Event Sets

Events can be generated from the vine copula (see section 2.4) using the pyvinecopulib python package [Nagler and Vatter, 2023]. For a given Random Number Generator (RNG) seed, and a fixed number of samples, stochastic event sets are generated in the univariate ranked space of each random variable present in the model.

For each combination of random variables, 10,000 events are generated using seed number 5. A training dataset is also created in the smallest, and the largest number of dimensions. This is done by generating 500 events using seed number 30. In both cases, if all pair copulas associated with a variable are statistically independent, it is removed from the regular vine. For this variable, the same number of events are generated from a uniform distribution between 0 and 1 with the same seed.

This results in 5 test event sets and 2 training event sets which are generated based on a copula and 2 vine copulas (chapter F). The copula generated test event sets in 2 and 3 dimensions. Since it is the lowest number of dimensions it is also used to generate a training event set. The vine copula fitted to 4 random variables is used to generate test stochastic event sets for 4 and 5 dimensions. Finally, the vine copula fitted to 6 random variables generated training and test event sets in 6 dimensions. This is summarized in table 4.3.

Table 4.3: Event sets generated from the various dependence structures.

Dependence Model Used	Stochastic Event Sets Generated
Copula fitted to: S Mag and P Mag	10,000 events for S Mag and P Mag
	500 events for S Mag and P Mag
	10,000 events for S Mag, P Mag, and T Mag (independence)
Vine copula fitted to: S Mag, P Mag, T Mag, and P Dur	10,000 events for S Mag, P Mag, T Mag, and P Dur
	10,000 events for S Mag, P Mag, T Mag, P Dur, and S Dur (independence)
Vine copula fitted to: S Mag, P Mag, T Mag, P Dur, S Dur, and P Lag	10,000 events for S Mag, P Mag, T Mag, P Dur, S Dur, and P Lag
	500 events for S Mag, P Mag, T Mag, P Dur, S Dur, and P Lag

These datasets are still in a ranked space (i.e. quantiles) and need to be transformed (section 4.4.2.1). Moreover, for datasets with less than 6 random variables, the missing variables need to be defined with constants (section 4.4.2.2)

4.4.2.1. Defining Non Extreme Marginal Probability Distribution Functions

A Cumulative Distribution Function (CDF) needs to be defined for all 6 random variables. The quantiles can be combined with the inverse CDF to obtain the real values of the variables. Currently, only one CDF is defined, which is S Mag (fig. 4.6). For non-extreme random variables, POT cannot be used. Therefore, to define their PDFs and corresponding CDFs, the fitter package [Cokelaer et al., 2024] is used. The package uses 80 continuous distributions from scipy [Virtanen et al., 2020] and fits parameters according to the Empirical Probability Distribution Function (EPDF) of a variable by maximizing the likelihood. Metrics can be used to select the best distribution. These include SSE, AIC, BIC, Kolmogorov-Smirnov (KS) statistic, and KS p-value.

These marginal PDFs are defined in chapter G and fig. 4.6. A summary containing the names of the distributions is shown in table 4.4.

Table 4.4: Distributions used to define the marginal PDFs and CDFs for the different random variables.

Random Variable	Distribution
S Mag	Exponential
P Mag	Exponential
T Mag	Empirical: HH Tides
P Dur	Truncated Normal
S Dur	Truncated Gumbel
P Lag	Truncated Normal

4.4.2.2. Constants in Event Sets

If one of the random variables is not defined for the stochastic event sets, a constant is assumed, which is the median of the historical event set for each random variable. These constants are shown in table 4.5.

Table 4.5: Constants associated with events that are in lower dimensions. For random variables associated with a given dimensionality see table 4.2. Black cells represent a random variable.

	2 dims	3 dims	4dims	5 dims	6 dims
T Mag [m]	0.85				
P Dur [hr]	14	14			
S Dur [tidal cycles]	4	4	4		
P Lag [hr]	-5	-5	-5	-5	
Discharge Magnitude [$\text{m}^3 \text{s}^{-1}$]	200	200	200	200	200

4.4.3. Defining Boundary Conditions for the SFINCS Model

Each event in the stochastic event set has a set of random variables associated with the different drivers. These random variables can be used as parameters to define a time series for the different boundary conditions of the SFINCS model. These are used to compute the economic damages associated with the event (fig. 4.2).

Non-tidal drivers are defined by a Gaussian distribution. Only three parameters are needed to correctly define one: the peak, the standard deviation, and the location. The peak and standard deviation are determined by the magnitude and duration of the event respectively. From a relative point of view, the dominant driver has a location of zero. The location of the other drivers (excluding tidal) are therefore given by their time lag.

The downstream water level is defined by three components: the MSL, tidal beating, and the skew surge. These are linearly superposed. MSL was identified as a constant with a magnitude of 0.2 m (chapter B). For a given T Mag, a random historical beating from the HH tide empirical distribution is sampled (with a fixed seed). To match the temporal resolution of the SFINCS model, linear interpolation is used for simplicity. The timing of the HH tide corresponds with that of the skew surge. Skew surge is defined as a Gaussian. Since it used a quantile to define its duration, this quantile is used to place the asymptote of the Gaussian. In section 4.3.2, this quantile was identified and is equivalent to 0.2 m.

Figure 4.8 visually shows how these parameters are combined to create the time series of boundary conditions for the downstream water level and precipitation.

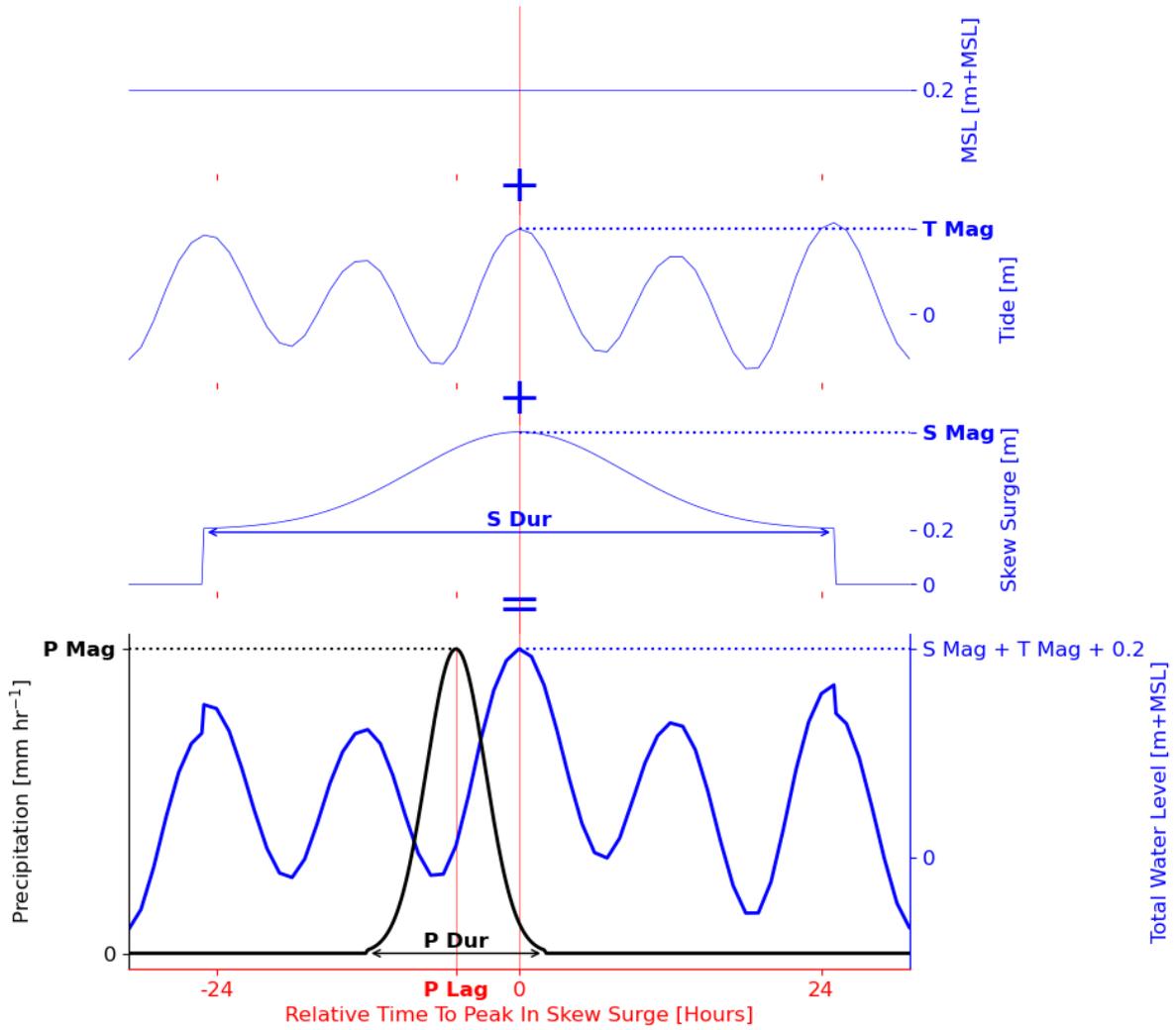


Figure 4.8: Boundary condition schematization, showing that the total water level (blue) depends on three components: the *MSL*, the tide, and the skew surge. These are linearly superposed. The *MSL* is assumed constant. The tide is solely defined by the *T Mag*. The skew surge is defined by its magnitude (*S Mag*) and duration (*S Dur*). The precipitation (black) is defined by its magnitude (*P Mag*), duration (*P Dur*), and time lag with skew surge maximum (*P Lag*).

4.4.4. Event Sets

This subsection provides an example of an event set, for other event sets consult chapter H.

Figure 4.9 shows the test event set generated by the copula for the *S Mag* and *P Mag*. Figure 4.9a shows the event set generated in the ranked space (grey dots) alongside the observed values (blue dots). This ranked space is transformed to the variable space in fig. 4.9b using the marginal *PDFs* (table 4.4). The 4 other parameters used to define the boundary conditions are kept as constants and use table 4.5 for their values.

The *POT* applied to *S Mag* and *P Mag* showed the univariate 100-year return value was approximately 0.9 m (fig. 4.6b) and 27 mm hr^{-1} (section D.1.1) respectively. For skew surge, the overestimation of the *RP* of the largest event causes only a small number of generated surge magnitudes to be greater than the largest historical event. This was a known limitation of the approach to fit tropical and extra-tropical events simultaneously. For precipitation, using an exponential marginal *PDF* on non-extreme and extreme events causes the tail of the distribution to have a larger density when compared to the tail of the *POT* exponential only fitted to extremes (see section D.1.1). This causes larger event magnitudes to be generated. This reduces the effect of ERA5 underestimating extreme precipitation totals (Lavers et al., 2022) but is not significant enough to reach the precipitation return values provided by NOAA (fig. 3.2).

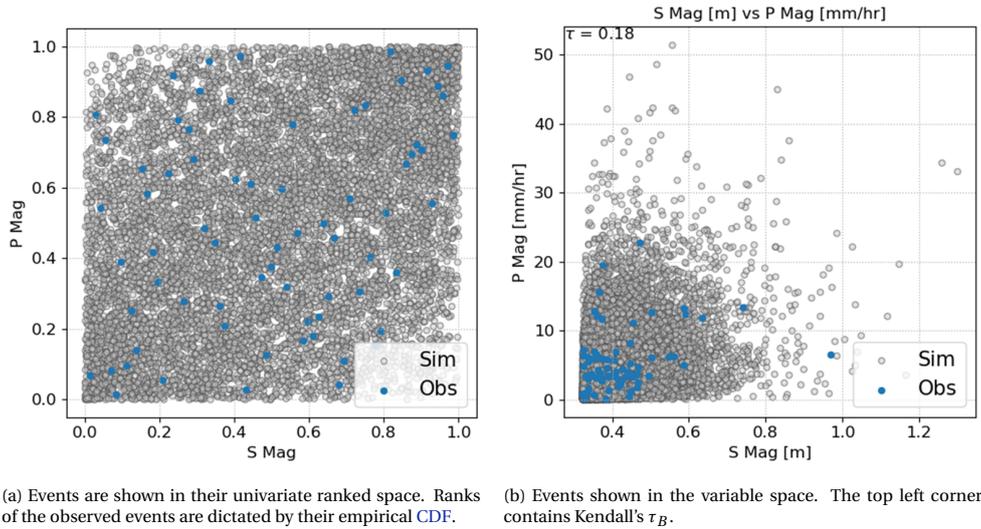


Figure 4.9: Test Dataset in two dimensions (10,000 samples) for P Mag and S Mag. Shows the observed events (historical event set) and events generated by the copula (Gaussian), which were obtained by using the `pyvinecopulib` python package [Nagler and Vatter, 2023]. The generated quantiles (fig. 4.9a) can be combined with the distributions defined in chapter G and fig. 4.6b to obtain fig. 4.9b.

4.5. Selection of Simulations and Surrogate Model

This section presents the methodology used to perform the current state-of-the-art approach (section 4.5.1), and the proposed approach (section 4.5.2), to reduce the computational cost associated with computing the economic damages for the entire stochastic event sets. This requires combining a sampling algorithm and a surrogate model. From here on we call the current approach the a priori approach as it selects simulations based on prior knowledge of the stochastic event set. We also call the proposed approach the a posteriori approach as it uses the knowledge gained from new observations to select samples.

4.5.1. A Priori Selection of Simulations and Surrogate Model

This section presents the a priori approach for sampling and modeling the economic damage surface associated with a stochastic event set. Jane et al. [2022] show combining a Maximum Dissimilarity Algorithm (MDA) with linear scatter interpolation provides the surrogate model with the best accuracy. This will therefore be used to represent the state-of-the-art approach (fig. 4.10). To reduce the computational cost, this approach is only be applied to 2 dimensions (fig. 4.9) and assumes constants (section 4.4.2.2) for other random variables. Since the current approach samples from the input, it can be assumed that all outputs (i.e. different FIAT models) are valid if a sufficient number of samples are simulated. The method used to compare both approaches is shown in section 4.5.3 and section 4.6.1.

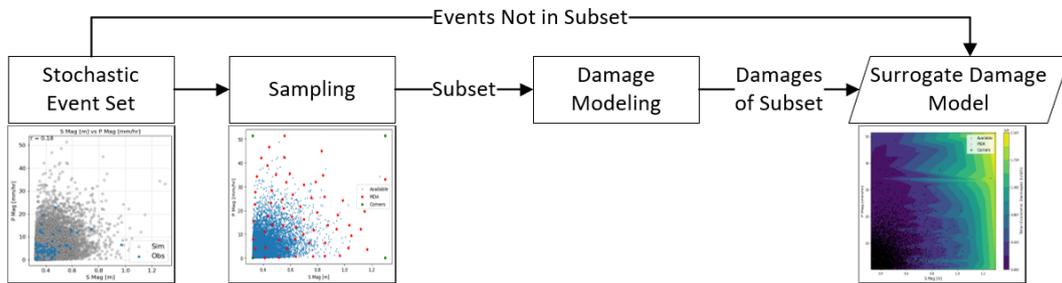


Figure 4.10: Current state-of-the-art approach for surrogate damage modeling compound floods.

4.5.1.1. Sampling with a Maximum Dissimilarity Algorithm (MDA)

Two different subsets of samples are generated. Firstly, 2^2 samples representing all the corners of the input space are simulated. These simulations are not part of the stochastic event set and prevent extrapolation.

Secondly, a **MDA** is applied to the stochastic event set. To prevent bias towards a variable, the dataset is normalized using a min-max scaler [Pedregosa et al., 2011]. Here it is assumed that $8^{d_{\text{driver}}} - 2^{d_{\text{driver}}} = 60$ **MDA** samples is sufficient for an understanding of the output space. The number of samples is based on the method used by Eilander et al. [2023c] where they use a structured grid with $8^{d_{\text{driver}}}$ simulations. The **MDA** is initialized by providing the event associated with the largest magnitude of the dominant driver. Euclidean distance is used as a metric of distance.

4.5.1.2. Surrogate Damage Modeling with Linear Scatter Interpolation

Linear scatter interpolation [Jane et al., 2022] is used to obtain economic damages for un-simulated samples from the stochastic event set. After the corners of the input space are simulated, interpolation is used recursively after each simulated event. This allows for an investigation of the accuracy of the approach (section 4.5.3), and how it improves with additional simulations. The computational cost of each method is assessed by timing the individual **SFINCS** and **FIAT** simulation times. It is assumed that the selection of simulations as well as the interpolation takes no additional computational cost.

4.5.2. A Posteriori Selection of Simulations and Surrogate Model

In this section, a description of the workflow used to select samples a posteriori with a corresponding surrogate model is given (fig. 4.11). To prevent initially sampling randomly, the Treed Gaussian Process (**TGP**) requires an initialization where the damages of a subset are simulated (section 4.5.2.1). In section 4.5.2.2, a description of the assumptions and products of the **TGP** are given. Section 4.5.2.3 shows how a stopping criterion is defined, and section 4.5.2.4 explains why this stopping criterion means the **TGP** should be defined on more than one output.

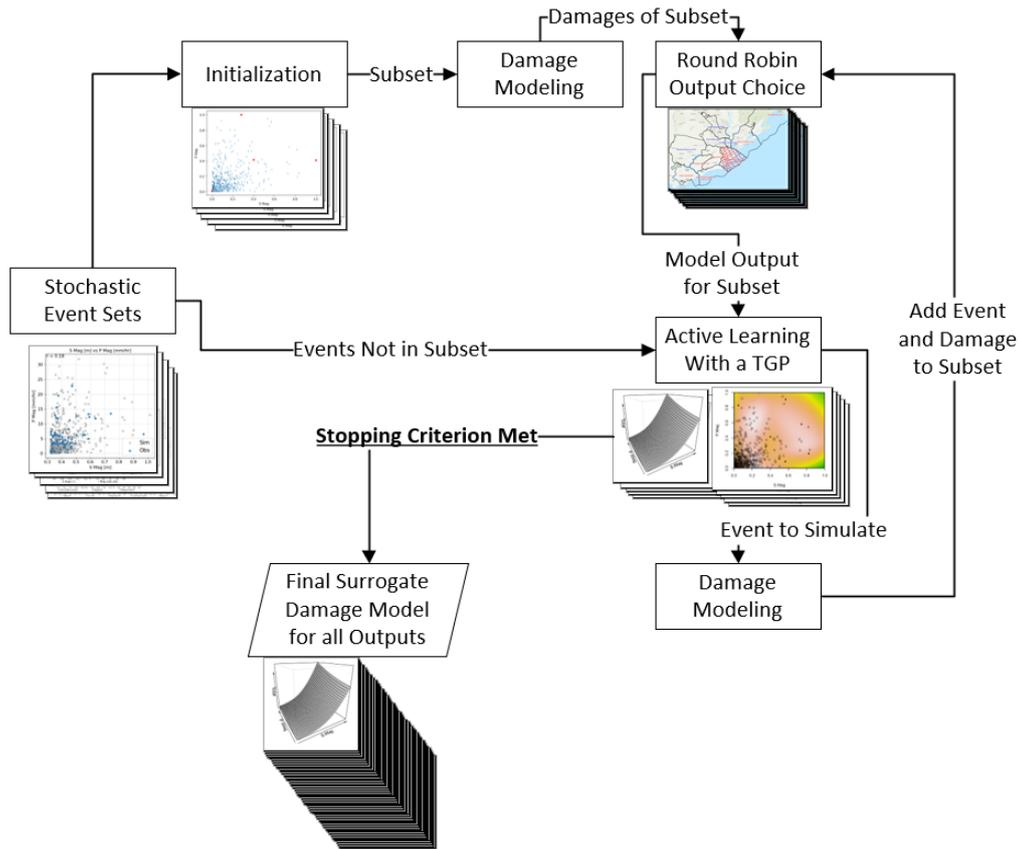


Figure 4.11: Workflow used to dynamically select samples a posteriori from a stochastic event set. The example shown in the workflow is the sub-county **FIAT** model (11 outputs) if it were applied to 2, 3, 4, 5, and 6 dimensions.

The a posteriori approach is used on the complete and classified **FIAT** models for the test stochastic event sets in 2, 3, 4, 5 and 6 dimensions. It is also deployed on the sub-county **FIAT** model in 2 dimensions. To choose a

stopping criterion, the a posteriori approach is also used on the training event sets in 2 and 6 dimensions for the complete FIAT model.

4.5.2.1. Initialization with a Maximum Dissimilarity Algorithm (MDA)

The initialization of the TGP is important for two distinct reasons. Firstly, it prevents the confidence interval of the TGP from being uniform over the entire input space resulting in random samples being simulated (fig. 2.3). Secondly, stochastic event sets that have a large number of dimensions are expected to require more samples to reach a stable output surface. Therefore, a larger initialization reduces the overall computational cost of using a TGP to reach a stable output.

To initialize the TGP, a MDA is used [Hendrickx et al., 2023]. It is known that the largest ALM statistic will be where the model needs to extrapolate [MacKay, 1992]. It is therefore more efficient to sample these without the TGP. For a structured grid, the number of vertices is equal to 2^d . If the MDA is initialized by a vertex of a grid and uses Euclidean distance as a distance metric, it will sample all of the 2^d vertices (fig. 2.2).

Here it is assumed that similar behavior occurs for an unstructured grid and the MDA generates a subset of $N_{sub} = 2^d$ events. Its exponential behavior makes it unfit for event sets with a large number of dimensions. After normalizing the stochastic event set with a min-max scaler [Pedregosa et al., 2011], the MDA is initialized by giving the event that represents the largest magnitude of the dominant driver and uses Euclidean distance as its distance metric. After $N = N_{sub}$ events are simulated, the output is also normalized using a min-max scaler. The simulated events become the sampled events, and the remaining events become the available events.

4.5.2.2. Active Learning with a Treed Gaussian Process (TGP)

A TGP allows to fit a surrogate model on the output, in this case, the economic damages of compound flood events. This model not only provides a regularized estimate for the entire stochastic event set, it also allows for the use of acquisition functions such as the ALM statistic [MacKay, 1992]. This statistic can be used to select samples based on the highest model uncertainty, allowing for a more efficient search.

For implementation, the TGP R package is used [Gramacy and Lee, 2009]. The sampled events and the package defaults for the Bayesian Treed Gaussian Process with a Linear Limiting Model (BTGPLLM) (i.e. squared exponential kernel, bflat prior, MCMC of 9000 samples with a burn-in of 2000 samples) were used to fit the surrogate model. From here on the BTGPLLM will be equivalent to TGP in the text. The posterior predictive surface of the surrogate model creates products such as the mean, 5, and 95 quantiles for both the sampled and available events. Moreover, the ALM statistic for the available events is computed. The ALM is used as an acquisition function where the available event corresponding to its maximum value is simulated, its damage is normalized and added to the sampled events. These steps are followed until a stopping criterion (section 4.5.2.3) is met. Each time the TGP is used, the computational time is measured.

The mean posterior surface can be used to compute the EAD and the 100-year RP (section 4.6.1). Similarly, the 5 and 95 quantiles can be used to show the modeling uncertainty in both metrics.

4.5.2.3. Defining a Stopping Criterion

A stopping criterion is essential in reducing the computational cost associated with performing a compound flood risk assessment. A TGP allows for the computation of metrics each time it is fitted when an additional sample is added to the subset. However, the literature (section 2.5) was unable to give a suitable metric to use, this subsection will describe the method to identify one.

For all events contained in the training event sets, the total economic damages from the complete FIAT model are all simulated (section 4.4.2 and section H.1). This allows for an estimate of the ground truth, which is the risk estimate calculated by simulating the economic damages of all events in the training event set. Then using the workflow presented in fig. 4.11, the posterior predictive surface of the TGP can be tested against the ground truth after each TGP iteration. To this end, different metrics are computed: the RMSE of the sampled and available events, the mean and maximum ALM of the available events, and the Coefficient Of Variation (COV) for a rolling window of size 10 of the EAD (see section 4.6 to understand how risk is modeled). The RMSE of the available events gives an indication of the accuracy of the current surface compared to the ground truth. Other metrics that follow its pattern should be used to define the stopping criterion.

Applying this methodology to the training event sets in 2 and 6 dimensions (see table 4.2 for random variables included) resulted in the stopping criterion being defined as 2 consecutive TGP models for an arbitrary output which results in an ALM mean smaller than 0.1 (chapter I).

4.5.2.4. Sampling from Multiple Outputs

The implementation of a stopping criterion means that there is high confidence in the sampled events to represent the current output. However, to the inverse of the a priori selection of simulations (where sampling the input for a large number of events gives good confidence on any output), it is unknown how well these sampled events will perform on an unseen output (such as a sub-county, but it could also be a social vulnerability) as the simulated events were chosen based on the features of a particular output.

To prevent this, fig. 4.11 shows how the TGP can sample from multiple outputs. After the initialization, the TGP is fitted to different outputs in a round-robin schedule. For example, suppose the classified model (table 4.1) is used. In that case, the TGP is first fitted to the inland economic damages, and the event corresponding to the highest ALM is simulated. Then, the TGP is fitted to the coastal economic damages, and the highest ALM is simulated. This process is repeated until one of the outputs reaches the stopping criterion, where the current output is removed from the round-robin schedule, and the TGP continues sampling from the other output until it also reaches the stopping criterion. In the case of the sub-county model, the round-robin schedule contains 11 outputs.

4.5.3. Comparing Both Approaches

To compare the computational cost, accuracy, and estimate of risk for both approaches, the test and training event set in two dimensions will be used. The large size of the test event set means the ground truth cannot be simulated within a reasonable amount of time. Nonetheless, a large event size provides a larger diversity in the stochastic events, bringing larger differences between both approaches. Therefore, both surrogate models will be generated on the test event set for the complete FIAT model. This will calculate their computational time. Then, the accuracy and estimate of risk of the surrogate models will be compared with the ground truth of the training event set for the complete FIAT model. Accuracy will be measured with Root Mean Square Error (RMSE). Section 4.6 will explain how risk will be compared.

4.6. Risk Modeling Under Different Approaches

This section gives an overview of how risk is modeled once an estimate of all economic damages is available for a stochastic event set (section 4.6.1). This is achievable by simulating or by using a surrogate damage model to estimate all economic damages associated with events in an event set (section 4.5). This section also explains the statistical tests performed to understand the effect a method or number of dimensions has on the estimate of risk (section 4.6.2). Finally, section 4.6.3 shows how to identify the drivers contributing to the events which cause economic damages with a RP larger than 100 years.

4.6.1. Defining Risk

The estimate of risk of an event is dependent on two components, its probability and its economic damages. For an estimate of the probability, the method used by Couasnon et al. [2022]; Eilander et al. [2023c] is followed. Each event in the stochastic event set (of size M) is assumed to occur at a frequency f_{oc} per year (eq. (2.6)). Then, all events are ranked according to the magnitude of their economic damage, creating an empirical CDF. The RP of each event is calculated using eq. (4.1). The RP of an event therefore depends on the type of output. This is because further upstream the 100-year RP will be caused by a different event than downstream.

$$RP_{\text{event}} = \frac{f_{oc}(M+1)}{\text{rank}_{\text{event}}(\text{economic damages})} \quad (4.1)$$

From this point onwards, the curve that plots the RP on the x-axis and the economic damages on the y-axis will be called the risk curve. The risk curve is integrated using the trapezoidal rule [Virtanen et al., 2020] to obtain an estimate of the EAD [Eilander et al., 2023c].

4.6.2. Measuring Differences in Risk Estimates

The estimate for a risk curve (RC_i , where i represents risk curves for different estimates of the same event set) with corresponding empirical CDFs ($ECDF_i$) and EAD (EAD_i) can be compared between different methods, ground truth, and the number of dimensions using statistical tests such as the Kolmogorov-Smirnov (KS) and Mann-Whitney U rank (MWU) tests.

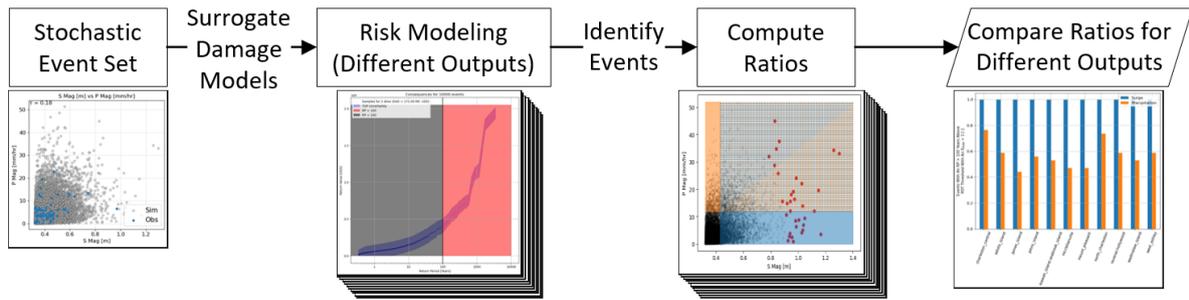
A two-sample KS test [Virtanen et al., 2020] can be used to compare the empirical CDFs of economic damages. A KS test is a non-parametric test that determines whether the two empirical CDFs ($ECDF_1$ and $ECDF_2$) are drawn from the same parent distribution. It calculates the maximum distance in cumulative probability (eq. (4.2)) and performs a hypothesis test. The null hypothesis assumes that both empirical CDFs are drawn from the same parent distribution. Therefore, for small p-values (in this case 0.05), the null hypothesis can be rejected showing there is a significant difference in the parent distribution, meaning the method or number of dimensions causes significant differences in the estimate of the risk curve. To compute this p-value, $ECDF_1$ and $ECDF_2$ were used as inputs for the `scipy.stats.ks_2samp` function [Virtanen et al., 2020]. Defaults are not changed.

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)| \quad (4.2)$$

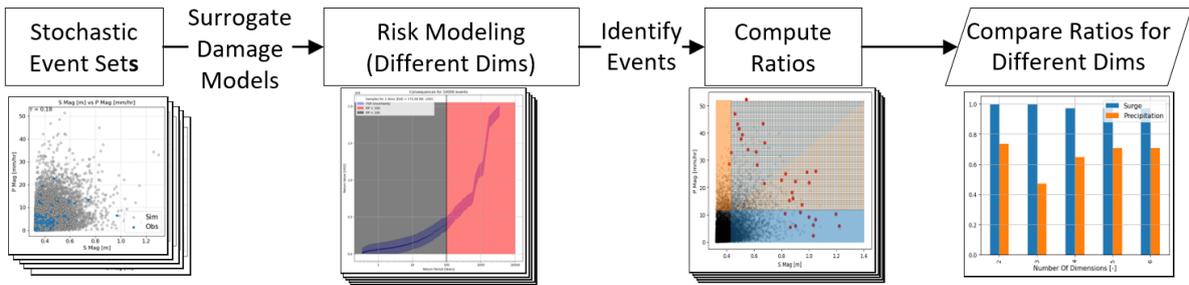
A MWU test [Virtanen et al., 2020] can be used to compare the EAD obtained from two different risk curves. It is a non-parametric test that tries to understand if there is a noticeable change in the location (EAD_1 , and EAD_2) for the two risk curves (RC_1 and RC_2). To obtain different estimations of the EAD_i for a risk curve (RC_i), an empirical bootstrap is used, which assumes each event in RC_i has an equal probability. It samples with replacement until M events are collected (i.e. the size of the event set). This process can be repeated multiple times, creating different realizations of the risk curve ($RC_{i,j}$, where j represents a bootstrap realization) and by extension, corresponding $EAD_{i,j}$. This process is performed for two risk curves that need to be compared (RC_1 and RC_2) 10,000 times. This creates 10,000 realizations of EAD_1 and EAD_2 which are used as inputs for the `scipy.stats.mannwhitneyu` function [Virtanen et al., 2020], where all the defaults are used. Using the p-value, a hypothesis test can be performed, where the null hypothesis assumes both RC_1 and RC_2 have the same EAD. Suppose the p-value is smaller than a certain threshold (in this case 0.05). In that case, the alternate hypothesis can be accepted, showing there is statistical significance that RC_1 and RC_2 have a noticeable change in location.

4.6.3. Identifying the Driver Dominating Large Return Periods

The dominance of a driver will be investigated by only looking at the magnitudes of surges and precipitation for the events that lead to a RP larger than 100 years in economic damages. A driver is found to contribute to the economic damages of an event if the driver magnitudes associated with stochastic events are greater than the POT threshold for an extreme rate of 1 event per year. For S Mag and P Mag this was found to be 0.43 m and 11.8 mm hr⁻¹ respectively. For each driver, a ratio can be created where the number of events that have their magnitudes exceed this threshold are compared to all events that cause damages with a RP larger than 100 years. The driver which has the largest ratio is the driver dominating the economic damages. This workflow is shown in fig. 4.12. The ratio is used to analyze the effect of multiple outputs/locations for the same stochastic event set (fig. 4.12a) and different stochastic event sets for the same output/location (fig. 4.12b) to understand how the events which cause a large RP are affected by these.



(a) Workflow used to compare different outputs/locations for the same stochastic event set.



(b) Workflow used to compare the effect of additional dimensions (dims, or random variables) for the same output/location.

Figure 4.12: Workflow used to identify the driver dominating the larger RPs. Once the ratios for P_{Mag} and S_{Mag} are computed for an output/location and stochastic event set, they can be compared with others.

5

Results

This chapter presents the main results required to reach the conclusions of chapter 7. Section 5.1 compares the results of the a priori and a posteriori methods. This investigates their accuracy (section 5.1.3), and cost (section 5.1.4). Section 5.2 shows and compares the results of including additional random variables when using the a posteriori method. First, the computational cost is shown (section 5.2). Second, the risk estimates are shown (section 5.2.2). Then, section 5.3 shows the results associated with identifying the flood driver dominating the economic damages.

5.1. Comparing a Priori and a Posteriori Methods

This section compares the a priori and a posteriori methods on the complete FIAT model for the test and training stochastic event sets which only include 2 random variables. First, a description of the sampling and surrogate modeling of each approach on the test event set is given (section 5.1.1, section 5.1.2). Second, in section 5.1.3, the surrogate models of both approaches obtained from the test event set have their accuracy assessed by comparing with the ground truth of the training event set (chapter 1). This is done by using different metrics: RMSE, KS p-value, and the estimate of the EAD. Finally, the computational cost of both approaches on the test event set is investigated (section 5.1.4).

5.1.1. Description of a Priori Sampling and Surrogate Modeling

Figure 5.1 shows the surrogate model generated for the a priori approach on the test event set in 2 dimensions, which is when S Mag and P Mag are random variables. The surrogate model uses linear scatter interpolation on 60 MDA samples and the 4 corners of the input space, and can be visualized with the color map of economic damages. Black dots represent available samples of the test event set in 2 dimensions (fig. 4.9) which use the interpolation technique to obtain economic damages. Red and green dots represent the subset of MDA samples and the corners of the input space respectively. Their economic damages were simulated with fig. 4.2.

The MDA is successfully able to generate regularly spaced grid points without being affected by the distribution of points within the input space. This facilitates the use of interpolation techniques. The surrogate model shows the economic damages increase for both S Mag and P Mag, but increases at a faster rate for S Mag. The surrogate model shows contour levels that display saw tooth profiles. This is caused by the piecewise linear interpolation of scatter interpolation, as it uses triangular surfaces to model the dependent variable. In locations with no available samples, this can cause sharp differences in contour levels, which may not represent reality. One example of this is located near a P Mag of 35 mm hr^{-1} . An argument could be made that the scatter interpolation is dependent on the event set, as the areas with poor interpolation occur because of a lack of available samples.

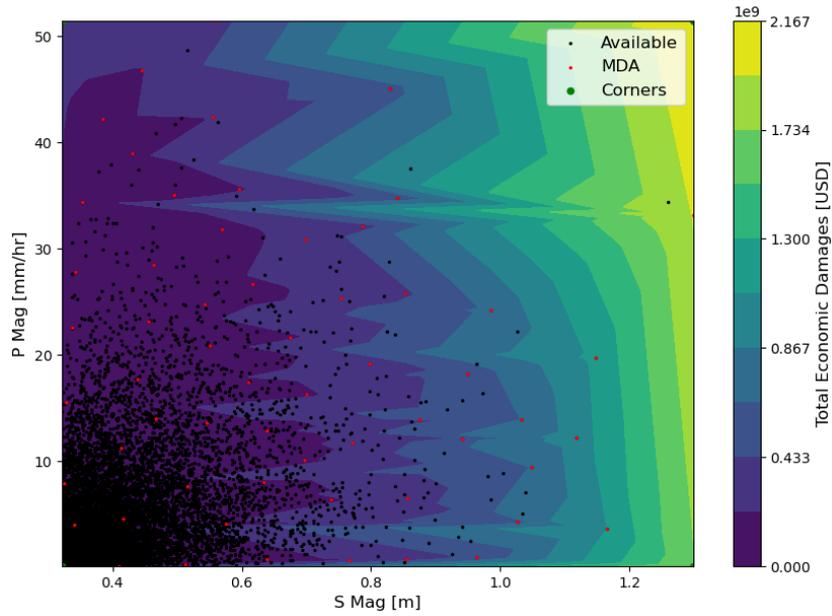
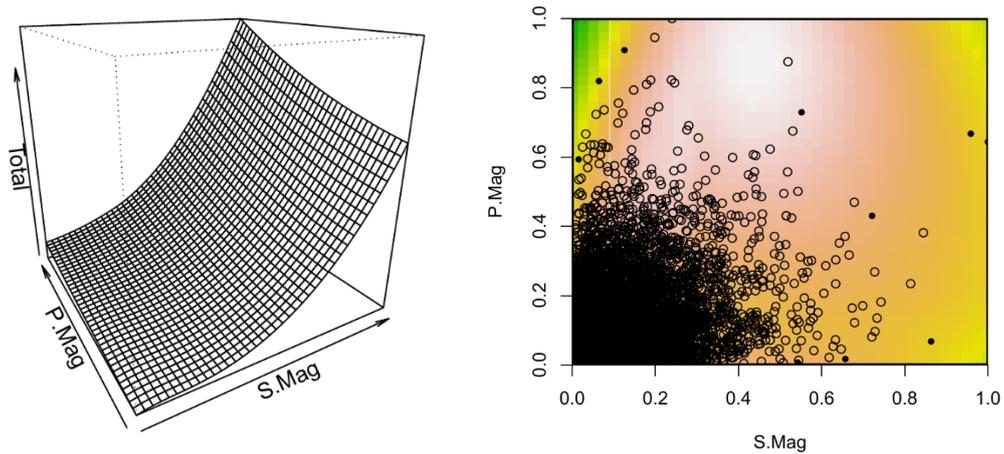


Figure 5.1: Surrogate model for the a priori approach. The color map indicates the total economic damages for the entire input space after interpolating with linear scatter interpolation. Scatter interpolation uses the MDA samples and the corners to interpolate to the rest of the event set.

5.1.2. Description of a Posteriori Sampling and Surrogate Modeling

Figure 5.2a shows the surrogate model generated by the a posteriori approach on the test event set in two dimensions for total economic damages. The surrogate model shows similar features as fig. 5.1, but is smooth and non-linear as it does not fit points deterministically. This is expected, as the posterior predictive surface uses hyperparameters to weigh prior and data-driven information to be regularized [Bishop, 2006; Gramacy and Lee, 2009].

Figure 5.2b shows the available events (open circles) and the sampled events (closed circles) in the normalized (min-max) space. The color map shows the ALM statistic associated with the input space. Shades of white/red indicate the highest value of the ALM statistic, which is where the TGP will sample next.



(a) Surrogate model for the a posteriori approach. Total represents the total economic damages.

(b) Available (open), sampled events (closed), and the color map represents the associated ALM statistic.

Figure 5.2: Surrogate model and associated uncertainty for the a posteriori approach after reaching the stopping criterion (14 total samples) for the complete FIAT model. Input and output are normalized. Plots generated by Gramacy [2023].

After 14 samples, fig. 5.2b identifies events with extreme **P Mags** to cause the largest uncertainty. In fig. 5.2b 11 of the sampled points can be seen. A quick comparison with fig. 5.1 shows three noticeable differences. Firstly, a smaller number of samples are required to reach the stopping criterion. Secondly, points which are similar in the input may be dissimilar in the output leading them to be sampled. This is the case for the two samples which have a surge magnitude greater than 1.2 meters. Finally, not all samples of the convex polygon need to be sampled to obtain the surrogate model shown in fig. 5.2a, as the **TGP** can extrapolate. This leads to the event with the largest **P Mag** not being sampled (which is the case for fig. 5.1).

5.1.3. Comparing the Accuracy of Both Approaches

Since not all 10,000 events of the test event set in two dimensions were simulated, the accuracy of both surrogate models can only be compared by using the training event set, where all 500 events were simulated (chapter I). Figure 5.3 shows the **RMSE** with the training event set for three different surrogate models: a priori method applied on test (fig. 5.1), a posteriori method applied on test (fig. 5.2a), and a posteriori method applied on training (fig. I.1).

Figure 5.3 shows that for the same output, but an unseen event set, the a posteriori approach outperforms the a priori approach in two ways. Firstly, for an arbitrary number of samples, the **RMSE** is smaller. Secondly, it reaches its smallest **RMSE** after 14 samples instead of 50. When comparing both approaches, the a posteriori approach improves the accuracy by a factor of 8.

Applying the a posteriori approach to the training event set without a stopping criterion shows that the accuracy of this approach would be improved if more samples were collected. This would mean using a more strict stopping criterion. The a posteriori approach also shows a smaller **RMSE** for any number of events for the training event set when compared to the test event set.

The significantly smaller **RMSE** for the training dataset shows the **TGP** is not only dependent on the output it samples from but also the event set. This is logical, as the **TGP** on the test set is more likely to sample events that have significantly larger magnitudes than the training set. The generation of 10,000 stochastic events leads to more extreme events than the generation of 500 stochastic events. Moreover, the **TGP** can only sample from events in the test event set. If the training set has a sample that is rarely generated and is not contained in the test set, it cannot be used to fit the surrogate model, leading to larger errors.

The **RMSE** is also expected to decrease with additional samples for the a priori approach as the spacing between grid points would decrease. This would lead to a more refined surrogate model. However fig. 5.3 shows that a plateau is reached after 40 samples, meaning a significantly large number of samples would be required to reach a similar accuracy of the a posteriori approach.

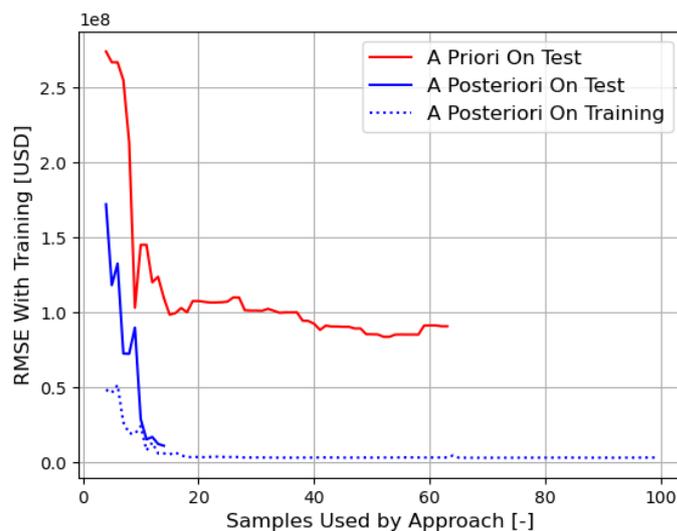


Figure 5.3: Comparing the **RMSE** of using the a priori approach and the a posterior approach on an unseen event set where each available event was simulated.

A comparison can also be made with the empirical CDF of economic damages and the EAD estimated by both approaches. Figure 5.4 shows the empirical CDF of economic damages of the ground truth (chapter I) when compared to the a priori approach (fig. 5.4a) and the a posteriori approach (fig. 5.4b). For each approach, a KS test is applied, and the statistic location can be seen with a dotted vertical line. Figure 5.9 shows the EAD for the ground truth, the a posteriori approach accompanied by the uncertainty bands, and the a priori approach.

Figure 5.4 shows the a priori approach and a posteriori approach have two different responses to the training event set when compared to the ground truth. The result of the KS two sample test means that both approaches generate estimates of the empirical CDF which are statistically different than the ground truth. This is because the p-value of each test is 0. On one hand, fig. 5.4a consistently overestimates the damages associated with different return periods. This leads to an overestimation of 96 million USD in the EAD when compared to the ground truth (fig. 5.5). On the other hand, for lower probabilities, the a posteriori approach initially underestimates the damages. However, for larger probabilities, it then overestimates the damages. This under than overestimation of damages results in an EAD that underestimates the ground truth by 5 million USD (fig. 5.5). This results in the a posteriori approach improving the estimate in EAD by a factor of 20. Finally, the magnitude of the uncertainty bars for the a posteriori approach in fig. 5.5 can be compared with the training event set, where these were noticeably smaller for the same stopping criterion (fig. I.1b in chapter I). This shows the stopping criterion, and by extension, the surrogate model are dependent on the event set they are trained on.

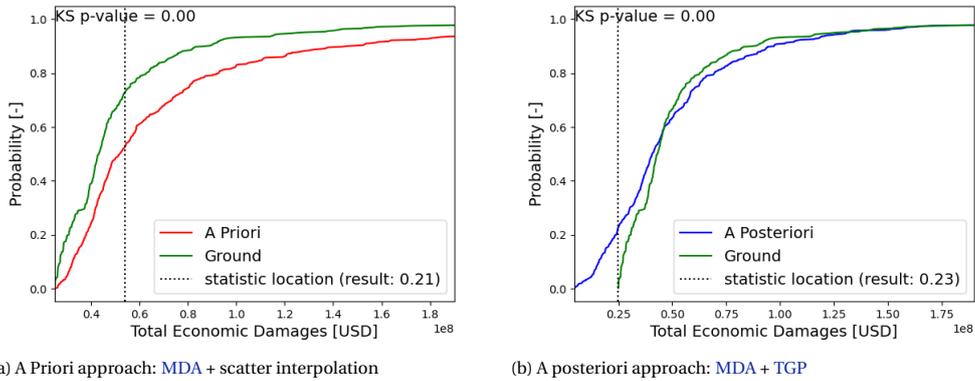


Figure 5.4: Test statistics and associated p-values associated with two-sample KS test for empirical CDFs of the different approaches when compared to the ground truth. If the p-value is smaller than 0.05, the null hypothesis (empirical CDFs come from the same parent distribution) is rejected

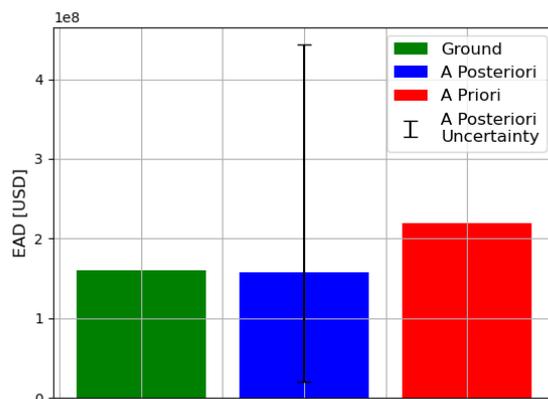


Figure 5.5: Comparing the EAD estimates of the a priori and a posteriori approach with the ground truth. A posteriori underestimates EAD by 5 million USD. A priori overestimates EAD by 96 million USD.

5.1.4. Comparing the Computational Cost of Both Approaches

It has been shown that the a posteriori outperforms the a priori approach in the number of samples it needs to fit the surrogate model, its accuracy, and its estimate of the EAD. One point of concern is the computational cost associated with the posterior predictive surface of the TGP. Figure 5.6 shows the computational cost associated with the a priori and a posteriori approach applied on a singular output. Each approach is subdivided into the main components of the workflow which contribute to the computational cost. For both approaches, SFINCS and FIAT are required to obtain economic damages. Their relative weights are dependent on the number of samples required by the approach. The a posteriori approach requires further computational time to compute the posterior predictive surface.

For a singular output, in the same dimensionality, the TGP outperforms the MDA by approximately 70 minutes. This results in the a posteriori approach reducing the computational cost by a factor of 4 when compared to the a priori approach. For a small number of samples, the cost of using a TGP is small, therefore, the reduction in the number of simulations (50) has a significant effect on the overall computational cost. This agrees with Hendrickx et al. [2023].

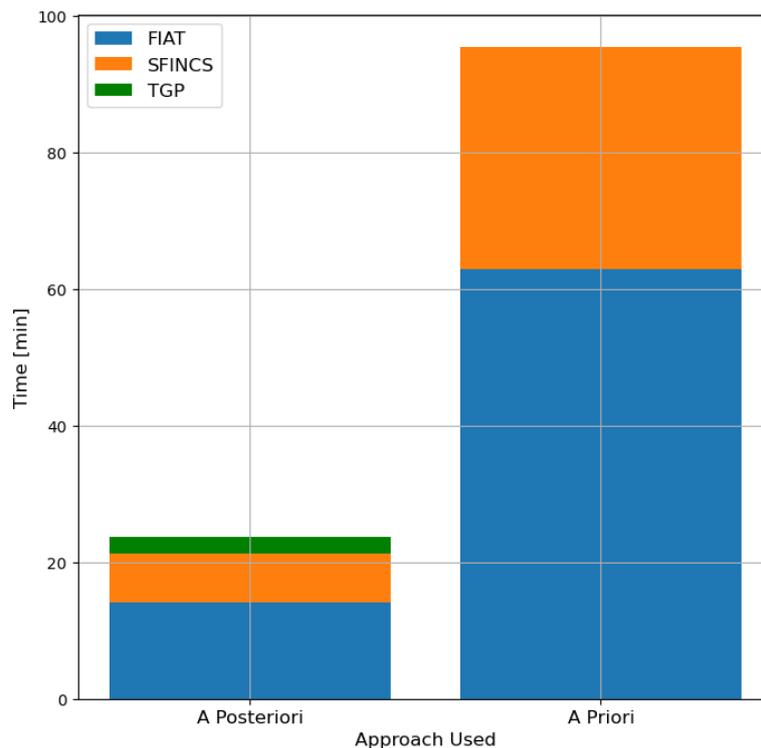


Figure 5.6: Computational cost associated with both sampling and surrogate modeling approaches.

5.2. A Posteriori Approach on Multivariate Input Spaces in Different Dimensions and Outputs

With this significant reduction in computational cost, there are more computational resources available, making it easier to include the effect of additional random variables in compound flood risk assessments. This reduction in computation cost is in part caused by the a posteriori approach only looking at a singular output, which is not always desired. For example, Eilander et al. [2023c] investigates social and economic damages. Therefore, the cost (section 5.2.1) and estimate of risk (section 5.2.2) when using the classified and sub-county FIAT models for different numbers of random variables are investigated.

5.2.1. Computational Cost of Using the a Posteriori Approach

The number of samples required to reach the stopping criterion (section 5.2.1.1) dictates the computational cost (section 5.2.1.2) of the a posteriori approach in two ways. Firstly, each additional sample requires the

cost of running the coupled **SFINCS** and **FIAT** models to obtain the economic damages. The cost of this operation is dependent on the dimensionality of the input space as the total duration of the event increases the simulation duration of **SFINCS** (chapter J). The duration of an event increases when including precipitation duration, surge duration, and precipitation lag as random variables. Secondly, the cost of computing the posterior predictive surface depends on the number of samples [Gramacy and Lee, 2009].

5.2.1.1. Samples Required to Reach the Stopping Criterion

Table 5.1 shows the total number of samples the a posteriori approach needs to reach the stopping criterion for different numbers of random variables and outputs. Numbers in bold show the total number of samples required for a **FIAT** model and the number of random variables. Three observations can be made regarding table 5.1.

Firstly, when increasing the number of random variables, the outputs for the complete and classified model show that more samples are required to reach the stopping criterion. There are exceptions to this observation. For example, the classified model, with the same number of dimensions, needs a smaller number of samples to reach the stopping criterion (38). In 5 dimensions, the total and coastal outputs need fewer samples than 4 dimensions to reach the stopping criterion.

Secondly, four cases (in more than two dimensions) reach the stopping criterion directly after using the **MDA**. In 5 and 6 dimensions, the **MDA** sampled 32 and 64 samples respectively before first using the **TGP**. The surface needs to be computed at least twice for an output to reach the stopping criterion. This is the case for the total and coastal (classified model) outputs. In the case of the classified model, two additional samples are needed as the surface first needs to be computed for the inland output, because the minimum number of samples to reach the stopping criterion for the first output follows eq. (5.1).

Finally, the larger the number of outputs, the more samples are required to reach the stopping criterion. This can be observed for any number of random variables for the complete and classified models. For 2 random variables, using the sub-county **FIAT** model and its 11 outputs significantly increases the number of samples that need to be simulated. This is in part caused by eq. (5.1).

$$\text{Min}_{\text{samples}} \text{ first stop} = \text{MDA}_{\text{samples}} + \text{Num}_{\text{outputs}} - 1 + \text{Order}_{\text{output}} \quad (5.1)$$

Table 5.1: Total samples required to reach stopping criterion for different dimensions (see table 4.2) and number of outputs (see section 4.1.2). For multiple outputs, the outputs are ordered top-down in the order they are sampled from (i.e. for the classified model, inland is sampled first). **Bold** numbers represent the output dictating the total number of samples required to reach the final stopping criterion.

		Number of Dimensions [-]				
FIAT Model	Output	2	3	4	5	6
Complete	Total	14	24	36	33	65
Classified	Inland	17	38	43	59	224
	Coastal	13	30	37	35	67
Sub-County	Charleston Central	36				
	Edisto Island	27				
	James Island	28				
	Johns Island	29				
	Kiawah Island	30				
	Seabrook Island	31				
	McClellanville	31				
	Mount Pleasant	32				
	North Charleston	33				
	Ravenel Hollywood	34				
	Wadmalaw Island	24				
West Ashley	35					

5.2.1.2. Computational Time Required to Reach the Stopping Criterion

Figure 5.7 shows the computational time associated with using the a priori and a posteriori approaches for different numbers of random variables (dimensions) (see table 4.2) and number of outputs (see section 4.1.2). For the a posteriori approach the cost is related to the stopping criterion being met for the final output. For each approach, number of outputs and dimensions, the computational time is subdivided into three components. The cost of using SFINCS, FIAT, and the posterior predictive surface (TGP). A horizontal red line is provided as a reference for the computational time of the a priori approach.

Figure 5.7 follows trends identified in table 5.1. It shows the non-linear increase in computational time associated with calculating the posterior predictive surface (TGP) when additional samples are included. For a large number of samples, this causes the TGP to dominate the computational cost of the method. This can be seen with the classified model in 6 dimensions, where the TGP amounts to 75% of the computational time. This is caused by the inland output requiring 224 samples to reach the stopping criterion (table 5.1). For a small number of samples, the FIAT model is the component that causes the largest computational cost. In larger dimensions, the inclusion of P Dur, S Dur, P Lag means the duration of the SFINCS model increases (see chapter J), which reduces the gap in computational time between the SFINCS and FIAT components.

Section 5.1.4 showed that the a posteriori approach outperformed the a priori approach on a singular output and the same number of dimensions. It is expected that increasing the number of random variables will increase the estimate of risk, but how does the cost increase? Figure 5.7 shows that for a singular output, only 6 dimensions exceeds the computational time of the a priori approach, as it requires 126 minutes. The cost associated with the TGP is limited as the stopping criterion is often met quickly after the MDA initialization. The increase in cost is mostly caused by SFINCS (see chapter J). The MDA initialization also results in 5 dimensions having a smaller computational time than 4 dimensions.

The classified model can also be used to compare the effect of including additional outputs on the computational cost. Sampling from multiple outputs provides a higher confidence in the estimate of risk at different locations, as the response of these locations to the economic damages will change, requiring more samples. Because the a priori approach is assumed independent of the output, the computational cost remains the same. In 2 dimensions, the cost associated with the TGP increases marginally as 3 additional samples are needed to reach the stopping criterion (table 5.1). Therefore, in the same dimensionality and on 2 outputs, the a posteriori approach still outperforms the a priori approach (fig. 5.7). For the classified model, apart from 2 dimensions, only 4 dimensions has a lower computational time than the a priori approach. While it required 5 additional samples when compared to 3 dimensions (table 5.1), its computational time is smaller. This is because of the exponential behavior of the MDA initialization, which means the TGP had to be computed a smaller number of times. In 6 dimensions, the inland output only reaches the stopping criterion after 224 samples (section 5.2.1.1), which causes the overall cost to be 16 times larger than the a priori approach (section 5.2.1.2).

Finally, for the sub-county model and its 11 outputs, the a posteriori approach still outperforms the a priori approach. However, the difference in computational time is smaller when compared to the complete and classified model. The main reason for this is the stopping criterion (eq. (5.1)), which requires the TGP to be fitted at least twice to each output.

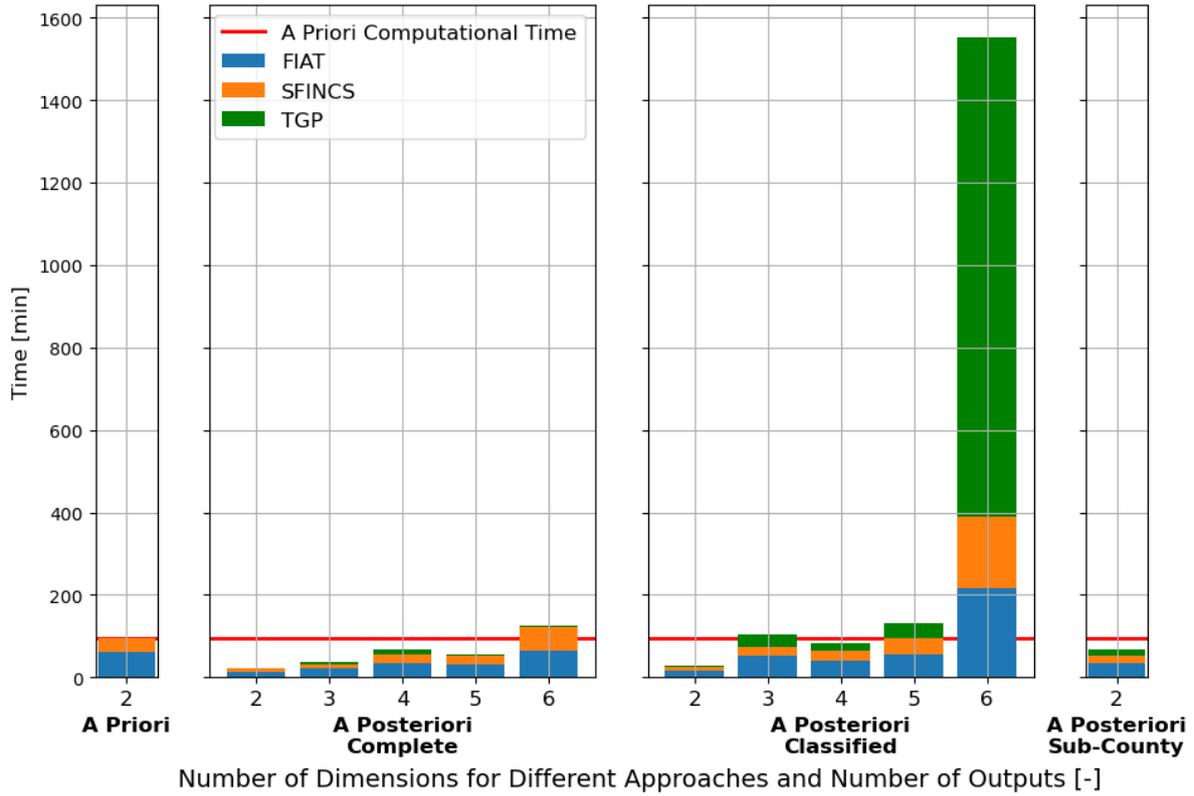


Figure 5.7: Computational time (in minutes) required to reach stopping criterion for different dimensions (see table 4.2), approach, and number of outputs (see section 4.1.2).

5.2.2. Significance of Additional Dimensions on Risk Curves and EAD

This subsection presents the outcomes of adding random variables to the estimate of risk when using the complete FIAT model.

Figure 5.8 shows the effect of including additional random variables in the multivariate input space (see table 4.2) on the risk curve for the total economic damages. The logarithmic behavior of the x-axis of fig. 5.8 makes it easier to see the differences in the economic damages at large RPs. The samples represent the mean response of the TGP to their event set. Since the TGP is probabilistic, the uncertainty associated with each return value can also be plotted. fig. 5.8 shows the uncertainty bands associated with each dimensionality is constant. The main reason for this is the stopping criterion is universal.

When considering the TGP uncertainty, 2 dimensions seem to be a clear outlier as it consistently underestimates the return values associated with economic damages that have a RP larger than 10 years. When considering the mean response, the risk curve shows noticeable differences in behavior between the different dimensions for all RPs. In 2 dimensions, the return values associated with RPs larger than a year seem to underestimate the return value for larger dimensions. The main reason for this is the omission of the T Mag as a random variable. The linear summation of the tidal beating with the skew surge causes a non-linear increase in damages (damage fraction curves are non-linear (fig. 4.3)). However, the tide cannot be the only explanation for this, as it underestimates return values associated with smaller return periods when compared to 5 and 6 dimensions. In 4 dimensions, by making the P Dur probabilistic, events with a small duration decrease the P Mag which co-occurs with the peak total water level, and the total precipitation volume. This decreases the damage associated with the events for RPs of 1 - 10 years. The inclusion of S Dur as a random variable has a significant effect on the EAD. This is primarily caused by its response to lower RPs (smaller than a year), which are significantly larger than for other dimensions [Couasnon et al., 2022; Eilander et al., 2023c]. While S Dur does not increase the peak water level, if the duration is larger than the median value, it increases the volume of water that overflows the flood protection measures. It also increases the magnitude of the total water level that co-occurs with the precipitation event. In 6 dimensions, the P Lag further decreases the number

of compound events, which reduces the EAD concerning five dimensions. The estimate in EAD between the lowest and highest number of dimensions is 172 and 219 Million USD respectively.

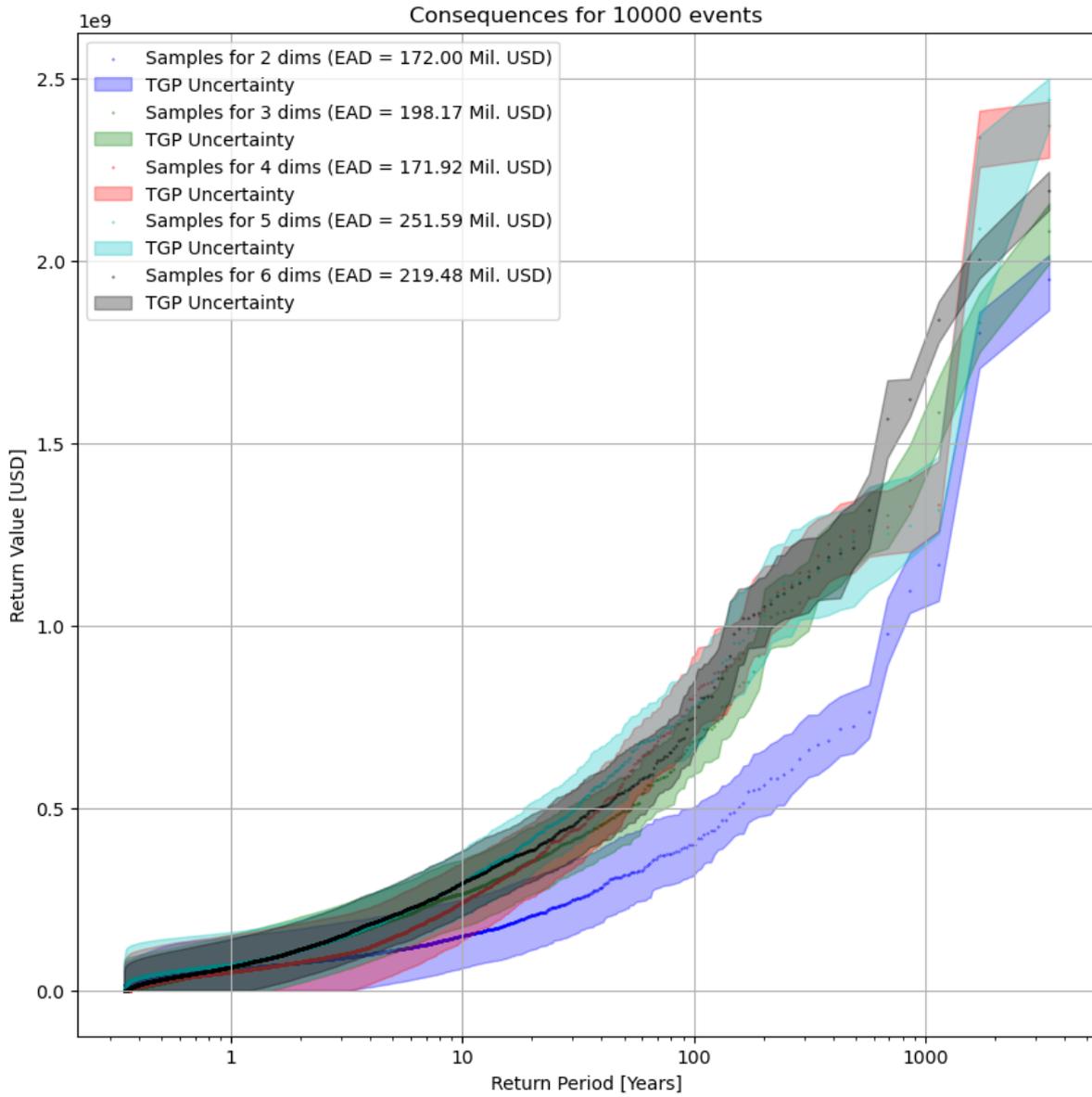


Figure 5.8: Risk curves associated with different dimensionalities of the multivariate input space for total economic damages. Legend includes the EAD estimate.

The risk curves associated with the inland and coastal damages were also computed for each test stochastic event set. These show a similar behavior to fig. 5.8 (fig. K.1, fig. K.4). The linear summation of the EAD for inland and coastal outputs leads to similar estimates in the EAD as reported in fig. 5.8 (fig. 5.9). Moreover, the error bars can explain the differences, however their magnitude, and range make them unsuitable to use. Figure 5.9 can also be used to show which location contributes the most to the EAD.

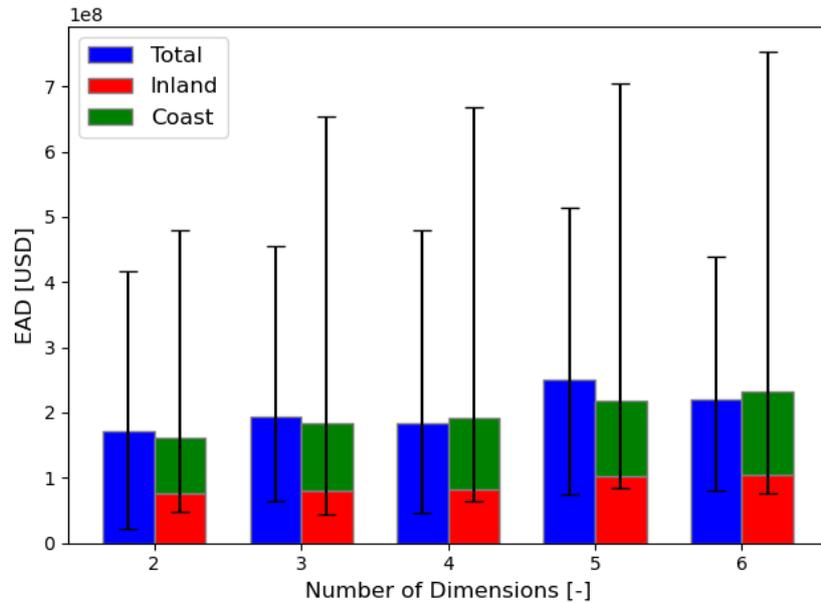


Figure 5.9: EAD estimates for the complete and classified model in different dimensions. Error bars are obtained from the TGP uncertainty of the risk curves. In the case of the classified model, these are linearly superposed.

Figure 5.10 shows the empirical CDFs related to the mean risk curves shown in fig. 5.8 for each dimensionality (see table 4.2). Each subplot represents a different combination of dimensions when applying the two-sample KS test. If the p-value is smaller than 0.05, the null hypothesis (empirical CDFs come from the same parent distribution) is rejected. The y-axis is synonymous with the RP of events. The large extreme rate used to model the historical event set (chapter D) means a large proportion of the empirical CDF only shows the effect of small RPs. All combinations have a p-value of zero. This shows that none of the empirical CDFs show statistical significance, they therefore do not come from the same parent distribution. In other words, each dimension leads to a significantly different response to the total economic damages. This is also the case for inland and coastal (fig. K.2, K.5), showing that for this case study, additional random variables such as time lag still have a noticeable effect on damages even if a driver is expected to dominate the response (see section 5.3).

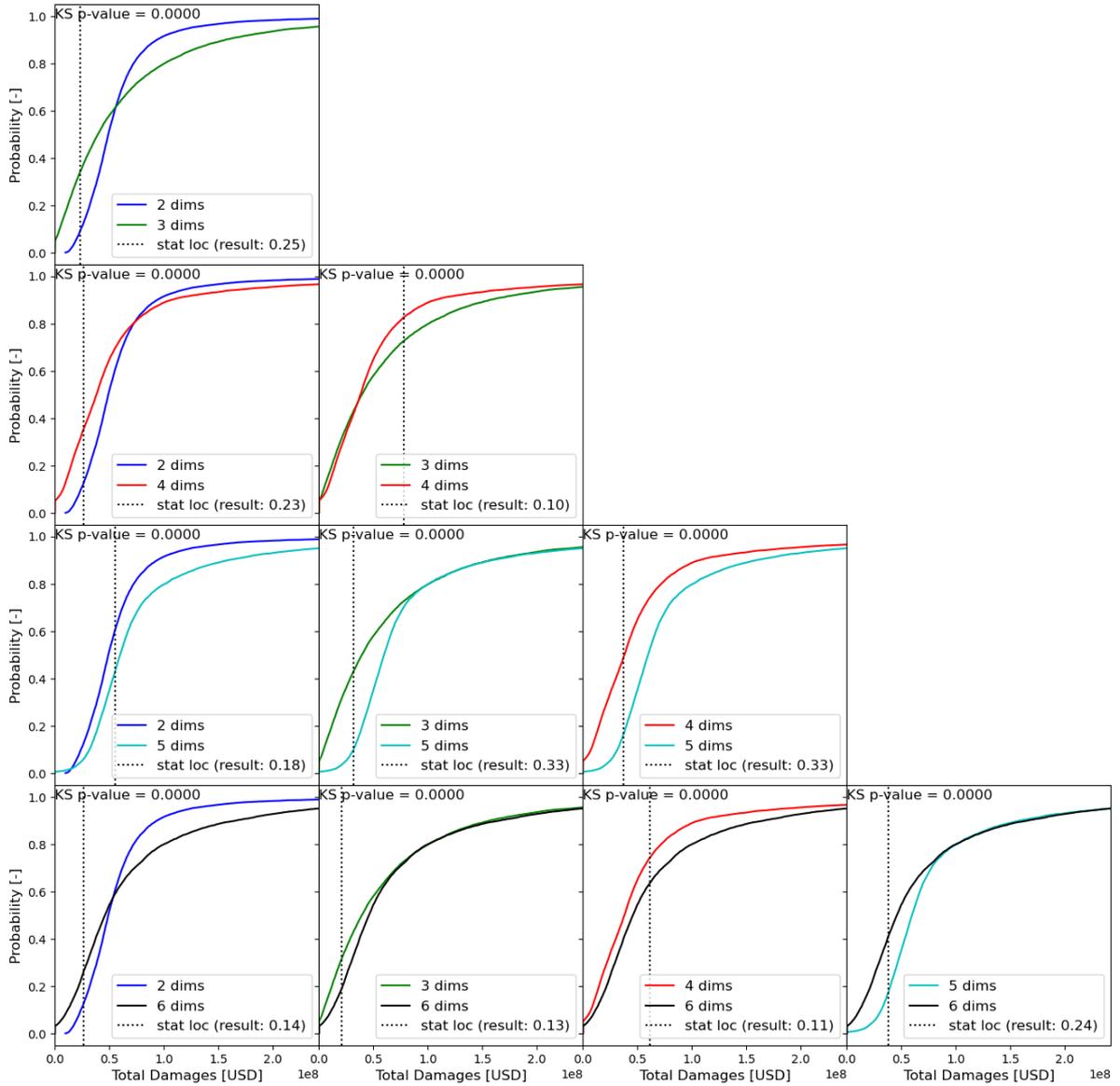


Figure 5.10: Test statistics and associated p-values associated with two-sample KS test for the empirical CDFs of different dimensionalities when sampling from the total economic damages. If the p-value is smaller than 0.05, the null hypothesis (empirical CDFs come from the same parent distribution) is rejected.

Figure 5.11 shows the EAD distribution obtained by performing an empirical bootstrap on the empirical CDFs in fig. 5.10. If the p-values are smaller than 0.05, the null hypothesis (empirical CDFs have the same location (EAD)) is rejected. Only one combination has a p-value that is greater than 0.05, which is two and four dimensions. Therefore, for this combination of dimensions, there is statistical evidence that the two empirical CDFs have the same location. This conflicts with previous findings. Firstly, the empirical CDFs are different with strong statistical evidence (fig. 5.10). Secondly, the risk curves are visually different (fig. 5.8). The reason for this outcome is related to the return values corresponding to small RPs in four dimensions (fig. 5.10), where approximately 5 - 10 % of the events have economic damages equal to zero. Since the EAD is strongly influenced by these smaller RPs, the large return values in four dimensions have a small effect on increasing the EAD [Couason et al., 2022; Eilander et al., 2023c].

The MWU rank test on the inland and coastal EADs show no significance for all variables (fig. K.3, fig. K.6), the empirical CDF for coastal and inland in four dimensions show strong similarities with the empirical CDF of total damages (fig. 5.10). This shows that the significance identified for 2 and 4 dimensions is either output-specific or is caused by small differences in the empirical CDF generated by the mean responses of the TGP.

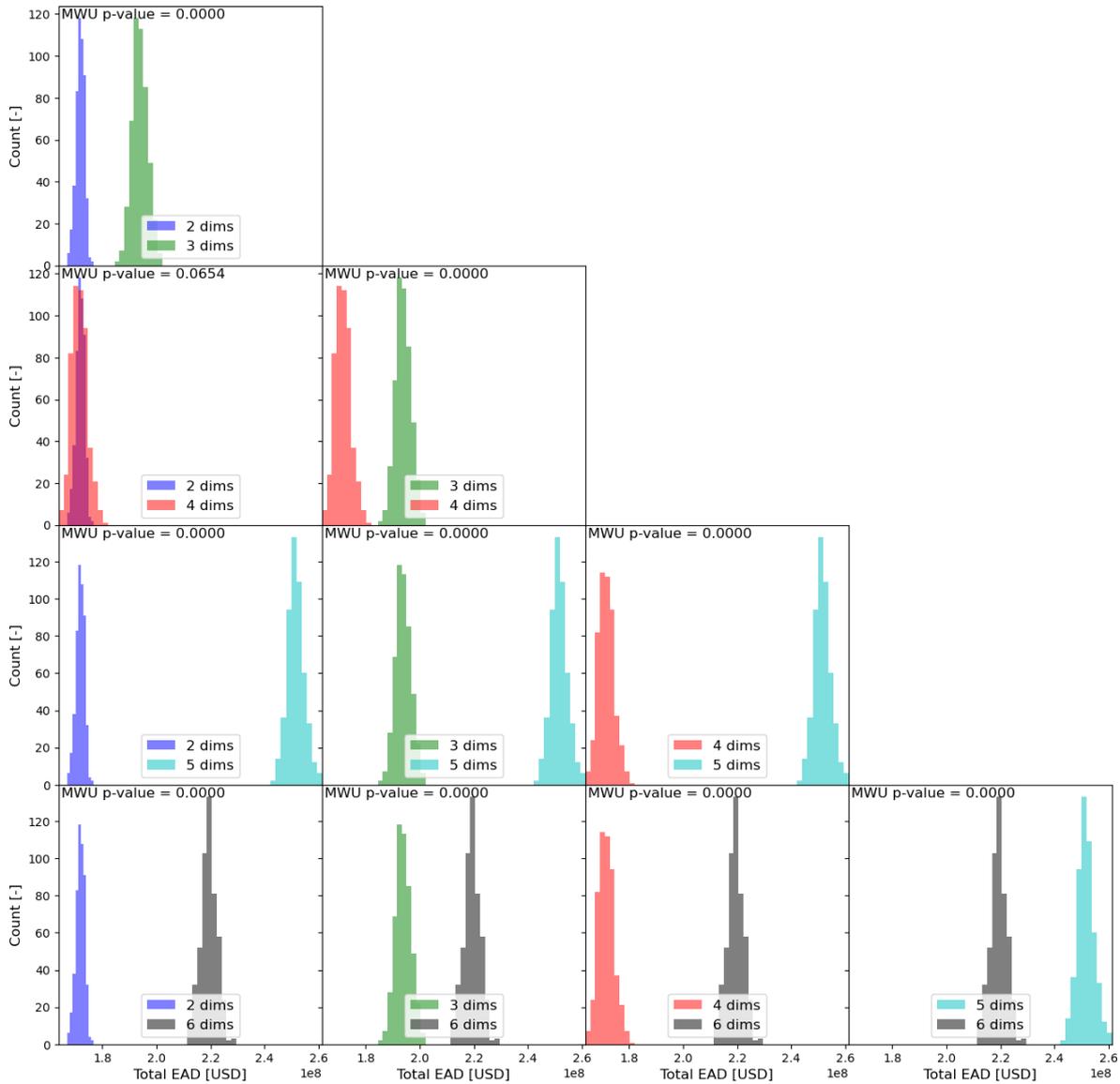


Figure 5.11: P-values associated with MWU test for empirical CDFs of different dimensionalities (see table 4.2) when sampling from the total economic damages. If the p-values are smaller than 0.05, the null hypothesis (empirical CDFs have the same location (EAD)) is rejected. The empirical bootstrap was repeated 500 times.

5.3. Identifying the Driver Dominating the Large Return Periods

While the sensitivity analysis performed on the addition of random variables to the multivariate input space can give insight into the effect of these variables on the response of damages, it does not explicitly say what type of events are dominating the economic damages. This can be investigated by looking at the ratio described in section 4.6.3. This is applied to the sub-county model to investigate the effect of the location of the outputs on this ratio (section 5.3.1). Then it is applied to the total, inland, and coastal damages when increasing the number of random variables in multivariate input space (section 5.3.2).

5.3.1. Different Sub-County's for Two Random Variables

Figure 5.12 shows the ratio of events for surge and precipitation that have magnitudes greater than their respective POT thresholds (extreme rate of 1). These are compared to all events that cause economic damages greater than the 100-year RP. The ratio for the surge is consistently 1 for all sub-counties. This shows in two dimensions that extreme skew surge is always present in the largest economic damages. On the other hand, this is not the case for the ratio representing precipitation, which varies based on the sub-county. Four sub-

counties have a ratio which is greater than 0.6: Charleston Central, North Charleston, Ravenel Hollywood, and West Ashley.

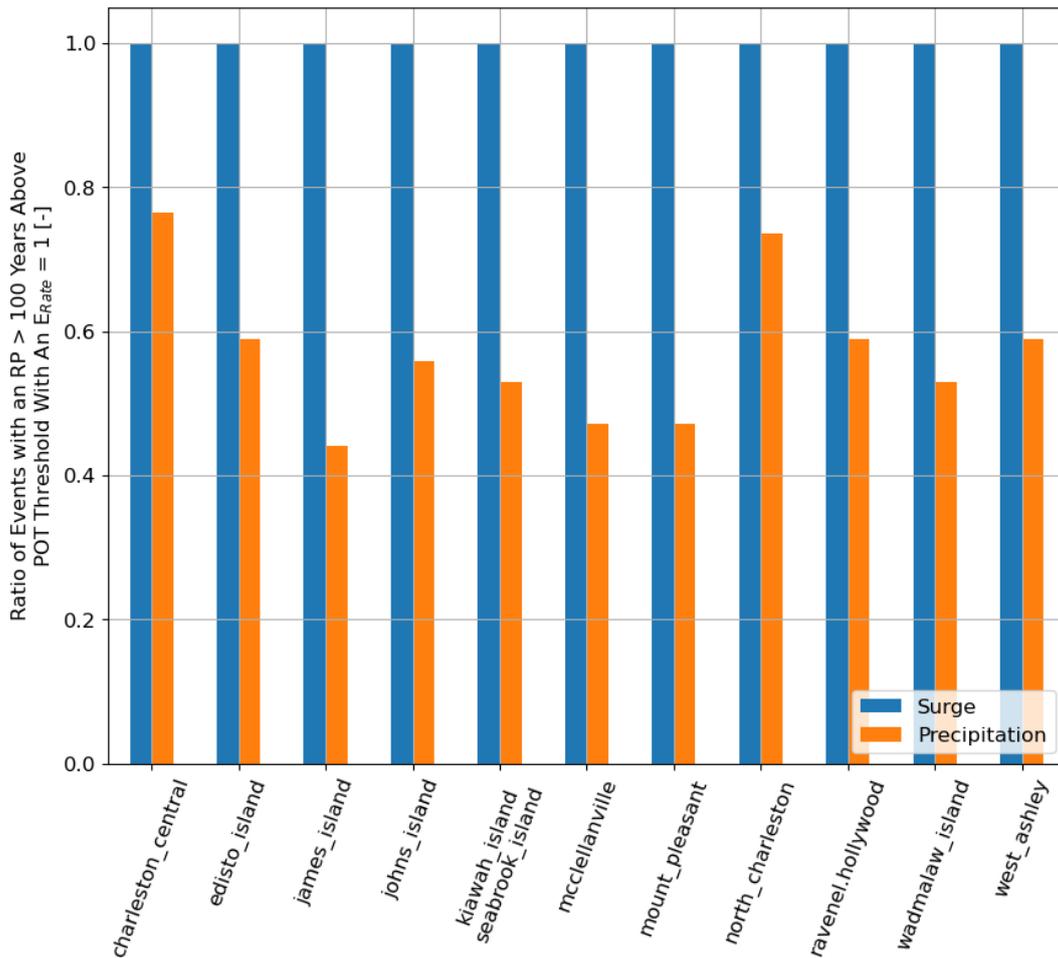


Figure 5.12: Ratio of driver magnitudes above their POT threshold (for an extreme rate of one) for events with a return period larger than 100 years for all sub counties.

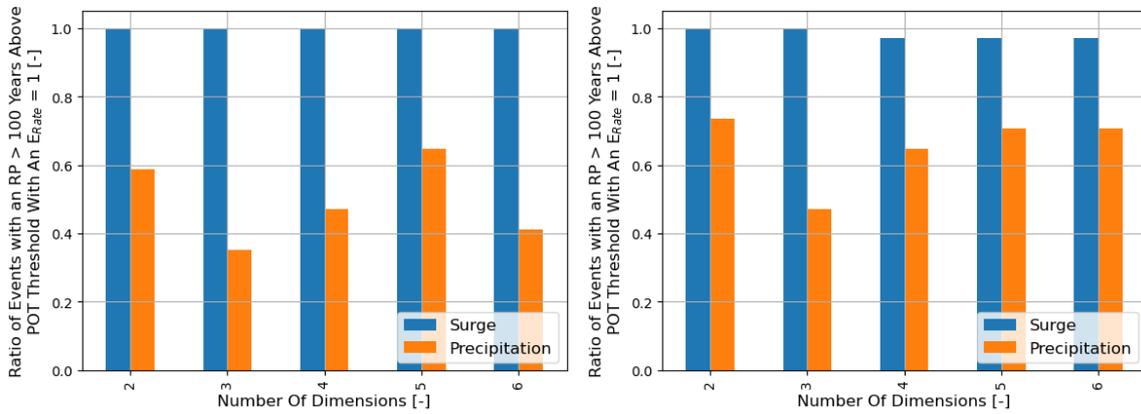
5.3.2. Different Outputs and Different Number of Random Variables

Recall that each dimensionality has a predefined combination of random variables (table 4.2). Figure 5.13 shows the effect of additional dimensions on the relative influence of drivers on extreme events. Similarities with fig. 5.12 can be observed in two dimensions. The ratio for the surge is always close to 1, and the ratio for precipitation depends on the location of the output. In more than 2 dimensions, fig. 5.13 shows that including additional random variables has little effect on the ratio for the surge. However, it affects the precipitation ratio, which does not show the same trends between the different outputs.

When comparing fig. 5.13a, and fig. 5.13b the precipitation ratio is larger for the inland output than the total output for all random variables. Moreover, when comparing fig. 5.13a, and fig. 5.13c the precipitation ratio for the total output for all random variables is larger or equal to the precipitation ratio for the coast output.

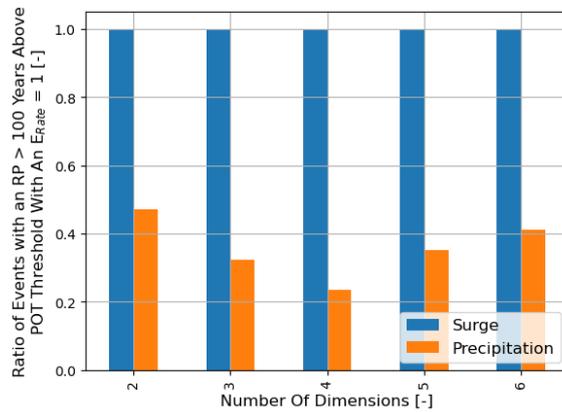
For fig. 5.13a, fig. 5.13b, and fig. 5.13c, the inclusion of the tidal magnitude as a random variable decreases the precipitation ratio. The inclusion of the precipitation duration further increases the ratio of events that have an extreme precipitation magnitude for fig. 5.13a and fig. 5.13b. This is not the case for fig. 5.13c, where the ratio for precipitation decreases. Figure 5.13a, 5.13b, and 5.13c show an increase in the precipitation ratio when including skew surge duration. The inclusion of precipitation lag has conflicting results for all three outputs. In the case of fig. 5.13a it decreases the precipitation ratio. For fig. 5.13b it has no influence, and fig. 5.13c it increases the ratio.

Figure 5.13b still shows the ratio of the surge is larger than precipitation for any number of dimensions. However, for 4, 5, and 6 dimensions, the ratio is smaller than one. This shows that an extreme precipitation event must have co-occurred with a non-extreme skew surge and caused significant damage. This was only captured when precipitation duration was included as a random variable.



(a) Complete model, total output.

(b) Classified model, inland output.



(c) Classified model, coastal output.

Figure 5.13: Ratio of driver magnitudes above their POT threshold (for an extreme rate of one) for events with a RP larger than 100 years for different dimensions and outputs.

6

Discussion

The results shown in chapter 5 allow for a comparison between (i) an a posteriori and a priori approach for the same event set with 2 random variables, (ii) the sensitivity of the compound flood risk estimates to additional random variables, (iii) the effect of using multiple outputs to increase the estimate of compound flood risk for different locations, (iv) which compound flood drivers cause the largest total economic damages.

The random variables included in the multivariate input space were based on the recommendations of [Eilander et al. \[2023c\]](#) and an initial sensitivity analysis. This resulted in 6 random variables being used to define the historical event set. These were Skew surge Magnitude (*S Mag*), Precipitation Magnitude (*P Mag*), Tidal Magnitude (*T Mag*), Precipitation Duration (*P Dur*), Skew surge Duration (*S Dur*), and Precipitation Lag (*P Lag*). Peak Over Threshold (*POT*) was applied on the *S Mag*, and co-occurring events were identified by using a window of a fixed length, which was based on the average duration of hurricanes. To define the dependence structure, (vine) copulas were used as they create stochastic event sets, where each event represented a set of local boundary conditions reducing the numerical model extents [[Couasnon et al., 2022](#); [Eilander et al., 2023c](#); [Gori et al., 2020](#)]. Points of novelty were (i) the use of a Treed Gaussian Process (*TGP*) to dynamically select samples and generate a surrogate model to reduce the computational cost associated with compound flood risk assessments, (ii) the definition of a stopping criterion based on the needs of such assessments (chapter I) and (iii) sampling from different outputs to decrease the uncertainty associated with their response.

6.1. Limitations of the Study

Some elements of the presented research are dependent on the case study. Moreover, because of the relatively short timeline, some assumptions were made to simplify the steps taken to generate stochastic event sets. Before concluding, the underlying limitations are listed below. These can be split into three distinct subsections. Some limitations are caused by simplifying assumptions to reduce the time required to perform a complete compound flood risk assessment (section 6.1.1). Others are currently affecting state-of-the-art compound flood risk assessments (section 6.1.2). Finally, using a *TGP* has brought a new set of limitations (section 6.1.3).

6.1.1. General Limitations

1. Single case study location

The methodology is not only applicable to the case study of Charleston. However, different locations will be affected by a different combination of compound flood drivers. This will change the number and correlation of random variables. For case studies with a large number of random variables, the vine copula may become a limitation, as it is difficult to find the correct regular vine which is associated with the smallest error. Moreover, it has also been shown that the computational cost increases with additional dimensions. This is caused by longer boundary conditions in *SFINCS* (chapter J), and the computational cost of the *TGP* increases with the number of sampled events. For case studies with the same set of random variables, the conclusions on the sensitivity of the compound flood risk assessments to certain random variables will vary, as the correlation and associated dependence are dependent on the case study location.

2. Focusing on economic risk

To simplify the analysis, it was decided to only focus on the effect of compound floods on economic damages. However, damages can also be social [Diermanse et al., 2023; Eilander et al., 2023c], and there is no guarantee that economic and social damages follow the same patterns with respect to the stochastic event sets. This is caused by zoning patterns, creating a disparity between the spatial patterns of exposure and vulnerability to economic and social consequences.

3. Focus on the current climate

The joint probability is likely to change with climate change, which would change the generation of stochastic events. Moreover, if the trend noticed in fig. B.1 is extrapolated, it is expected that sea level rise will continue to increase, further increasing the effect of the downstream boundary condition on compound floods. Both of these are not taken into account. This means the estimate of risk and EAD should change over time.

4. Omission of tropical cyclones

Seasonality of events is not taken into account, which should normally be the case for the case study location. There is therefore an underestimation of return values associated with Return Periods (RPs) larger than 5-10 years [Parker et al., 2023], which will underestimate risk [Nederhoff et al., 2024].

5. Using a dominant driver for event identification

Identifying a dominant driver for the historical event set restricts the analysis to damages caused by extremes of this driver. There is therefore an omission of events that lead to extreme consequences. This limitation was minimized by (i) increasing the extreme rate of the POT on the dominant driver, (ii) decoupling the skew surge and the tidal driver, and (iii) defining the P Mag with a non-extreme PDF (see section D.1.1). These lead to smaller total water levels at small RPs, and significantly larger precipitation magnitudes at large RPs.

6. Correlation and dependence associated with time lags

Previous research has shown that certain time lags increase the dependence between driver magnitudes [Khanal et al., 2019]. To accommodate this approach, the historical event set should be split into two for this particular time lag. However, this would require a significant amount of events to ensure the vine copula can be safely fitted to both.

7. Using Dißmann's algorithm to fit regular vines

Dißmann's algorithm [Dißmann et al., 2013] was used to fit the regular vines in this research. It is uncertain if this algorithm provides the best possible regular vine. It was chosen as it was unknown how many random variables would have their dependence modeled, which could have caused the use of a brute force approach to be unsuitable. A brute force approach ensures the best regular vine is chosen according to a certain metric, but has only been used on up to 8 random variables [Morales-Nápoles et al., 2023], and can be computationally expensive for 7 and 8 random variables. However, for 6 random variables, the brute force approach would have been computationally feasible.

6.1.2. Current Compound Flood Risk Assessments

1. Spatial uniformity of compound flood drivers

While this research does attempt to include additional random variables, it assumes spatial uniformity for precipitation, skew surge, and tide. Taking this into account would bring some additional issues as it would (i) increase the number of random variables to correctly model the compound flood drivers, (ii) require accurate and long time records of drivers at different locations, (iii) require a parametric model to correctly model historic events to include the parameters in the dependence structure. Nonetheless, by assuming spatial uniformity, there is an overestimation of economic damages associated with events.

2. Using global models to obtain the historical time record of a driver

Previous research has shown that ERA5 underpredicts extreme precipitation totals [Lavers et al., 2022]. Using ERA5 to retrieve the historical time record of precipitation therefore leads to an underpredicting of damages associated with extreme precipitation events. Chapter 2 showed machine learning tools could be used to solve this issue.

3. Super-Fast INundation of CoastS (SFINCS)

The use of **SFINCS** significantly decreases the computational cost associated with the hydrodynamic simulation of an event. However, its numerous assumptions cause certain physical phenomena to be neglected, increasing the uncertainty associated with the hazard map obtained for stochastic events. Nonetheless, some of this uncertainty is caused by the global datasets [Bates et al., 2021] used to calibrate and validate the **SFINCS** model, which would also be present in more complex hydrodynamic models. Moreover, the use of **SFINCS** only investigates overflow and overtopping as potential failure mechanisms. It does not consider other levee failures that can occur at high water levels such as piping.

4. Using stochastic event sets

When comparing events simulated by the vine copula in 6 dimensions (fig. H.6), **S Mags** and **P Mags** do not correspond with fig. 4.9 even though the same seed was used, and the corresponding pair copula represents a node of the vine copula. The only test event sets that are the same are (i) 2 and 3 dimensions, (ii) 4 and 5 dimensions. If variations do exist, they should only affect the tail of the risk curve (i.e. high **RPs**). This is because the same vine copula and marginal **PDFs** are used. For a large number of samples, it is expected that the kernel density estimate (**KDE**) should roughly remain the same. Therefore, the effect on **EAD**, and empirical **CDF** of economic damages, which are dependent on the overall dataset, and not individual simulations, should be limited. If significant differences are observed, this should be caused by the extra dimensions and not the specific realization of the event set. This will have a noticeable effect on the identification of drivers causing events with a **RP** larger than 100 years.

6.1.3. Use of a Treed Gaussian Process (TGP)

1. Simplistic boundary condition representation of events

To simplify the generation of local boundary conditions, a Gaussian shape is assumed as it is only dependent on three variables, the location (mean), peak magnitude, and the duration (standard deviation). The small number of parameters in a Gaussian prevents the artificial boundary condition from accurately representing the historical events. For example, a Gaussian is unable to show multiple peaks in magnitude and is symmetric.

2. Threshold-based stopping criterion

The stopping criterion uses the **ALM** mean to decide whether enough samples have been collected to get a sufficient understanding of the response of the economic damages to the input space. However, the **ALM** mean depends on the **MCMC** to compute the hyperparameters related to the **TGP**. This process requires a large number of iterations. Therefore, if the posterior predictive surface was computed multiple times for the same sampled events, it could lead to a different realization of the **TGP** mean and **ALM** statistics for the available samples. For a threshold-based stopping criterion, an inaccurate **ALM** mean could lead to early stopping (less accurate representation of the economic damages) or more samples (than necessary) being drawn from the input space (increasing the computational cost).

3. Sampling historical tides randomly

For artificial events, the tidal beating related to a given Higher High (**HH**) tide magnitude is sampled randomly from the time record. While the magnitude of the event is correct, the significant differences in beating may lead the damages associated with the same event to change. This can only be explained by the **TGP** as noise, increasing the uncertainty associated with the event set.

6.2. Sampling and Surrogate Modeling

Section 5.1 shows the a priori and a posteriori approaches are dependent on the event set used. This is one of the advantages of the structured grid used by [Couasnon et al., 2022; Diermanse et al., 2023; Eilander et al., 2023c] as the simulations and surrogate model are independent of not only the output but also the event set. The a priori and a posteriori methods could be applied to a structured grid, but it is expected this will increase the number of samples required to reach an accurate estimation for all the dependent variables. This is because a larger number of samples will be located in areas of the input space where stochastic events will not be generated. This is caused by the dependence structure and marginal probability distribution functions.

6.2.1. A Priori Sampling and Interpolation

A priori sampling is independent of the output, as the number of samples and interpolation technique is chosen without knowledge of the response of the dependent variable. The sampled events can be used to

interpolate in different outputs leading to no increase in computational cost. However, there is no certainty that the same number of samples or interpolation technique provides the same level of accuracy for the different outputs.

The a priori approach was chosen based on the results of [Jane et al. \[2022\]](#) which showed that a [MDA](#) combined with a scatter interpolation outperformed other sampling and surrogate model techniques. [Camus et al. \[2011\]](#), recommended using a [MDA](#) as it provided a good subset of points on which to use interpolation. This was observed in section 5.1.1. However, the use of dissimilarity prevents the [MDA](#) from sampling two points close to one another, which can lead to poor interpolation and loss of accuracy in locations where significant differences in the output are expected. Additionally, while [Jane et al. \[2022\]](#) recommends the use of linear scatter interpolation technique, it is unknown if this interpolation provides the best representation of the dependent variables for a stochastic event set. Section 5.1.1 showed saw tooth contour levels in the interpolation of the dependent variable providing an irregular response. In certain cases, this leads to sharp differences in the dependent variable at locations in the input where this was not expected. This may lead to inaccurate estimations of the dependent variable.

Therefore, the inaccuracy described in section 5.1.3 is caused by both the interpolation technique, which is not flexible enough to correctly represent the output, and the [MDA](#) not selecting the samples that bring the largest expected information gain.

6.2.2. Applicability of a Treed Gaussian Process (TGP)

A [TGP](#) is more computationally expensive to use than a [GP](#) because of the additional hyperparameters needed to define its hierarchical model (eq. (2.18)).

Nonetheless, a [TGP](#) was used over a [GP](#) as it was believed that the binary partitions would bring about a smaller computational cost and a better understanding of the uncertainty associated with the input space. Moreover, it could lead to a better fitting model in the context of flooding, as levees and other flood protection measures can suddenly be overtopped, which would require a discontinuity in the response of the output [[Gramacy and Lee, 2009](#)]. A binary partition could therefore indicate at which magnitude of a driver a flood protection measure is no longer viable. Moreover, if [P Lag](#) is included, the effect of compound floods should only be relevant for a range of time lags, which a binary split could show when this is the case.

However, for a partition to be statistically significant, it requires a large number of sampled events. The use of a stopping criterion to minimize the computational cost therefore reduces the likelihood of observing a partition. This can be minimized by using the [TGP](#) on multiple outputs, as the number of samples needed to reach the stopping criterion increases (eq. (5.1)). Nonetheless, the outcomes of the training dataset showed if a partition occurs it is not always stable (chapter I).

In this research, the [TGP](#) was used on 26 distinct outputs, and only partitioned on 2 of them, showing the additional computational cost may not be appropriate for (i) this case study (ii) a stopping criterion. The outputs that were partitioned were North Charleston (sub-county model in 2 dimensions) and Inland in 6 dimensions. In both cases, only the precipitation magnitude caused the binary split.

In the case of North Charleston (chapter L), when conditionalizing on the minimum of [S Mag](#), the binary split does not seem to occur at the value of [P Mag](#) where damages start occurring, demonstrating that if binary splits did occur for the other outputs, they may have not brought the desired information of the state of the current flood protection measures.

6.2.2.1. Early Stopping

Using an a posteriori method allows for an estimation of a stopping criterion, which should reduce the computational cost associated with the hydrodynamic and impact models, as it gives more confidence in the number of samples needed to model the response of an output.

One observation made in section 5.2.1.1 was related to the [TGP](#) stopping after the initialization provided by the [MDA](#). This was also observed with the training dataset (chapter I). However, for the training dataset, it was expected that this was caused by the small size of the event set. If it also occurs for the test set, it could be caused by (i) the stopping criterion not being robust enough leading to an early stop, (ii) the incremental addition of the [P Dur](#), [S Dur](#), and [P Lag](#), which reduces the number and severity of compound flood events. This demonstrates the importance of correctly defining the vine copula, as it dictates the number and severity of compound floods in the stochastic event set.

An early stop does not cause the TGP to be a poor surrogate model as it is regularized (chapter 2), but as seen with the training dataset (chapter I), it does lead to the possible accuracy. Moreover, fig. 5.9 showed that this can result in small differences in the estimate of the EAD. It can also lead to differences in the economic empirical CDF and EAD distribution generated by these as these are based on the mean response of the TGP and do not take into account its uncertainty.

6.2.2.2. Multiple Outputs

While early stopping is a limitation of this research, consistent trends are still noticeable in the results. One of these is the additional number of samples required to reach the stopping criterion when there are multiple outputs. Part of this can be attributed to eq. (5.1). Since a round-robin schedule is used, the larger the number of outputs, the more samples are needed to remove outputs from the schedule. However, this is not always the case. This shows that using multiple outputs requires additional samples to better understand the response of the dependent variable in different contexts. For the sub-county model, Charleston Central requires the largest number of samples, which is caused by it being the most affected by compound flooding (section 5.3.1). The TGP therefore requires more information on different combinations of events to correctly model the output. Using multiple outputs therefore increases the understanding of compound flooding, which reduces the likelihood of an early stop, which comes at an additional computational cost.

By having multiple outputs, the TGP is forced to sample from different random variables to understand the local response to economic damages, which will vary depending on the location (distance from the open coast) and characteristics (i.e. flood protection measures) of the output. This is a desired effect, as it provides a better understanding of flood risk for different areas of the SFINCS and FIAT model boundaries. It therefore allows governing bodies to use a utilitarian framework to know where to spend their resources.

The use of multiple outputs shows that a TGP should not only apply to the case study of Charleston. In 2 dimensions, it is known that the estimate in EAD provided by the TGP is accurate (section 5.1.3). Moreover, the classified model in 2 dimensions reaches a similar quantification of EAD when combining the inland and coastal output (fig. 5.9). When comparing the driver ratios for surge and precipitation between the inland (fig. 5.13b) and the coast (fig. 5.13c) in 2 dimensions, it can be seen the response of the locations are significantly different. The inland location is more affected by precipitation, than the coast. Nonetheless, the TGP can still accurately model both locations and provide an accurate estimate of EAD.

6.2.3. Comparing the Computational Cost of Both Approaches

The results shown in section 5.1.4 and section 5.2.1.2 are dependent on the processor and the hydrodynamic model used. One of the advantages of SFINCS is its speed. If a more complex hydrodynamic model was used to solve the flood hazard of an event (i.e. Delft3D), it is expected that this would be the component dominating the cost for the number of samples shown in section 5.2.1.1.

Section 5.1 showed that the a posteriori outperformed the a priori approach on the same event sets for the number of samples it needs to fit the surrogate model, its accuracy, and cost. With this significant reduction in computational cost, it could be interesting to increase the number of random variables, decrease the number of assumptions used in compound flood risk assessments, and obtain a better understanding of the true response of the economic damages to compound flood drivers. With six random variables, the longer boundary conditions in terms of time combined with the larger number of sampled events lead to an increase in the total duration of 30 minutes for a singular output (section 5.2.1.2). This outcome might be dependent on the MCMC, which leads the TGP to an early stop. However, fig. 5.9 showed that the estimate in EAD was consistent when comparing the outcome of the complete and classified models for any number of random variables. Moreover, chapter I (more specifically fig. I.2) showed that the training event set would have stopped after 66 total samples if a stopping criterion was used. Therefore, there is some confidence in the computational times reported in fig. 5.7.

section 5.2.1.2 exposed one potential issue of using a TGP and a stopping criterion. It is unknown what the cost of modeling an event set will be. For 6 dimensions and the classified model, this leads to a large computational cost. This may have been caused by small amounts of noise introduced by the schematization / parametrization of the boundary conditions (section 4.4.3). Ishibashi and Hino [2021] also experienced this and used the argument that in certain active learning contexts, less information can provide better accuracy than more information. To prevent large computational times, a stopping criterion may be placed on the number of samples for the a posteriori approach, which is similar to the a priori approach. This may remove

the benefit of using an a posteriori approach. This is because the surrogate model can provide metrics that can monitor if enough samples have been collected to measure the accuracy of the dependent variables of an event set (chapter I).

Finally, for the sub-county model, the difference in computational cost between both approaches further decreases, but it is still advantageous to use a TGP. Moreover, the TGP can give estimates of the confidence for each output, providing a more robust measure of EAD for multiple outputs.

6.3. Sensitivity Analysis of Additional Random Variables

When comparing the risk curves shown in fig. 5.8, a clear outlier was identified. This was also observed for the inland and cost outputs (fig. K.1, and fig. K.4 respectively). For higher RPs, the risk curve associated with the non-tidal driver magnitudes underpredicts the return values of economic damages when compared to higher dimensions. This remains the case when the modeling uncertainty of the TGP is taken into account. Including the other driver magnitudes (only T Mag in this case) reduces the difference with the higher dimensions. This may be caused by the proximity of Charleston County to the open coast, which makes it susceptible to damage from the coastal water level.

The sensitivity analysis in section 5.2.2 showed that each additional variable had a noticeable effect on the different metrics such as the empirical CDF of economic damages and EAD. Without taking into account the limitations, a conclusion could be that additional variables improve the understanding of compound flood risk assessments, and justify the increase in computational cost. However, the limitations influence the results.

Firstly, the statistical tests were dependent on the realization of the TGP mean for an event set, and do not take into account the model uncertainty of the TGP. Depending on the realization, it can lead to different estimates of the metrics and associated test outcomes. This means that a range of samples and associated computational costs for each FIAT model and dimensionality exist. For each of these exists a risk curve, with an associated confidence interval, empirical CDF, and EAD. If early stopping occurs, it could cause noticeable differences between different realizations of the method for the same event set, dimensionality, and output.

Secondly, because of the large event set size, combined with a large extreme rate, this leads 70% of the event set to have a RP smaller than 2 years. The EAD and empirical CDF are therefore dominated by the response of the surrogate model in areas with small economic damages. However, the TGP samples from areas with the largest expected information gain, which is often located in areas with significant differences in economic damages concerning the input (i.e. large RPs). When using a stopping criterion, it may prevent the TGP from understanding the damages associated with these lower RPs, leading to noticeable differences in EAD and empirical CDF.

Finally, another reason for the small p-values related to the empirical CDFs could be related to the type of test conducted. On one hand, the size of the event set dictates the significance of the KS-statistic. On another hand, the KS test only takes into account the largest difference in probability between two empirical CDFs. It does not investigate the mean difference. The estimates of the empirical CDF may give large values of the KS statistic, which is caused by a large difference in the lower RPs. However, in the larger RPs, the empirical CDFs may be similar as shown in the risk curves (fig. 5.8). The KS test does not account for this.

These outcomes did not affect the results of the training dataset (chapter I). This is related to the size of the event set. Because of the probabilistic nature of an event set, a smaller number of events leads to less diverse pairs of generated events. This has two effects. Firstly, the MDA initialization will lead to a larger proportion of the available samples to be simulated. Secondly, the smaller event space leads to an easier MCMC computation for the hyperparameters. Both of these will decrease the variety between realizations of the TGP mean.

6.4. Identifying the Driver Dominating the Large Return Periods

The identification of the dominant driver is important, as it can indicate to governing bodies what flood protection measures should be improved. Section 5.3.1 and section 5.3.2 both showed that the driver ratio for S Mag was always close to 1. This may be caused by the choice of threshold which was too small for S Mag. Nonetheless, other factors can cause this such as the proximity of the sub-counties to the open coast and the estuary. It is therefore expected that large S Mags will cause significant damages. Moreover, in section 3.4.1,

when discussing the previous validation of the [SFINCS](#) model, there were some unanswered questions on the number of false positive inundated homes. If this false positive number is high, it is expected that the effect of the skew surge driver on economic damages will increase.

The classified model assumed the sub-counties could be clustered according to their proximity to the open coast. This would allow the output space for the different clusters to exhibit different behaviors without increasing the computational cost significantly. The four sub-counties used for inland were: Charleston Central, North Charleston, Ravenel Hollywood, and West Ashley. These sub-counties are associated with the four largest precipitation ratios (section [5.3.1](#)) demonstrating that this assumption was justified. Thus, in two dimensions, the observations made on the sub-county model should be comparable to the classified model and should reduce the computational cost of using the methodology. This required some form of prior knowledge to make this decision. Moreover, greater confidence is gained by sampling from multiple outputs and should give insight into the location of flood protection measures. For the same event set, the different sub-counties see noticeable differences in the types of events that can cause damages with a [RP](#) larger than 100 years. This is not only caused by the location of the sub-county but also its flood protection measures concerning the different drivers. This analysis therefore disagrees with the use of design events for different [RPs](#) [[Jane et al., 2022](#); [Moftakhari et al., 2019](#)]. One reason for the noticeable difference is that this analysis took into account pluvial instead of fluvial drivers. Another reason could be because the analysis was focused on risk and not hazard.

Section [5.3.2](#) also showed the dominant driver was always skew surge. Moreover, some general observations in the precipitation and surge ratios were caused by the proximity of the sub-counties to the open coast. However, as mentioned in section [6.1.1](#) the ratio for precipitation was affected by differences in the stochastic event sets generated. Nonetheless, incrementally adding random variables had different effects on the number of extreme precipitation events. Including the [T Mag](#) increased the number of random variables that can cause extreme water levels, which increases the relative frequency and magnitude of extreme water levels. For example, a spring tide coinciding with a surge event that has a [RP](#) of 2 years can cause damages with a [RP](#) of 100 years. This decreases the relative damage of extreme precipitation events. Including [P Dur](#) increased the precipitation totals of extreme precipitation events and the co-occurring [P Mag](#) with the total water level event. This is because of the correlation between [P Mag](#) and [P Dur](#) (section [E.1](#)). Including [S Dur](#) increases the overflow volume and the co-occurring skew surge magnitude with the precipitation event. Finally, including [P Lag](#) decreases the number of events that are co-occurring, as the marginal probability density function (chapter [G](#)) has a higher likelihood of generating events that have a significant time lag.

7

Conclusion and Recommendations

7.1. Conclusions

In conclusion, by answering the following sub-questions, a Treed Gaussian Process (TGP) has shown that it can improve the quantification of compound flood risk in different ways.

1. **To what extent can a Treed Gaussian Process improve the estimate of compound flood risk while limiting the increase in computational cost?**

The results demonstrate that if the state-of-the-art approach and the TGP are applied on the same event set in 2 dimensions, the TGP improves the accuracy in RMSE by a factor of 8, improves the estimate in EAD by a factor of 20, and reduces the computational cost by a factor of 4. While the cost reduction is significant, this is dependent on a stopping criterion, which may not provide the best representation of the risk curves. For multiple outputs, the cost of using a TGP increases, but for Charleston, these are always smaller than the current approach for the same number of dimensions.

2. **To what extent can a Treed Gaussian Process give insight into the dominant driver when drivers are defined by multiple random variables?**

A TGP reduces the cost associated with compound flood risk assessments facilitating the inclusion of additional random variables. Including the duration and lag of drivers, causes events with damages with a RP larger than 100 years to change. This changes the magnitude of drivers when events co-occur which changes the relative influence of a dominant driver. For the case study of Charleston, skew surge was always dominant where almost all events with a RP larger than 100 years had an extreme S Mag. However, the relative influence of precipitation changed depending on the number of random variables and the output.

3. **To what extent is the value of risk, its associated uncertainty, and the computational cost of a Treed Gaussian Process sensitive to the number and choice of random variables when modeling compound flood events?**

By using a stopping criterion with an ALM mean of 0.1 on the TGP, the uncertainty associated with the surrogate model of economic damages was approximately constant for different combinations of random variables. For the case study of Charleston, the computational cost was mostly dominated by SFINCS and DELFT-FIAT, which in certain cases was dependent on the initialization of the TGP with the MDA. This led the computational cost to increase when increasing the number of random variables. For a singular output, the cost increased by 100 minutes between the lowest and highest number of dimensions. When taking into account the uncertainty of the TGP the driver magnitudes provide a similar risk curve to the risk curve of all random variables. However, the sensitivity analysis demonstrated that sequentially including additional random variables always had significant effects on the risk curves and corresponding estimates of EAD. This may have been affected by the MCMC and the stopping criterion, which did not provide the best representation of the economic damages. Nonetheless, the estimate of EAD for 2 and 6 random variables was found to be significantly different and in-

creased from 172 to 219 Million USD for Charleston County. This shows the importance of including the tide, driver duration, and lag as random variables.

4. How are the results associated with the case study applicable to other locations?

The focus of this research was on improving the selection of simulations and surrogate modeling related to compound flood risk assessments. Individual outcomes such as the identification of a dominant driver, or the sensitivity analysis cannot be assumed to translate to other locations. This is because these are dependent on characteristics that are intrinsic to the case study. These include flood protection measures and the dependence between compound flood drivers. Nonetheless, by applying the TGP on different outputs, and the number of random variables, this research shows the outcomes related to the cost and accuracy of the surrogate model can be applied to other locations. The outputs had significant differences in the response of the dependent variables, and the TGP was still able to model these. In different locations, additional compound flood drivers may have to be included, such as fluvial discharge. In these locations, it is unknown how non-surge, tide, or pluvial drivers will affect the cost and accuracy.

7.2. Recommendations

1. Uncertainty in estimates from the Treed Gaussian Process (TGP)

The uncertainty associated with the TGP is dependent on the stopping criterion used. To prevent large increases in computational cost for marginal improvements in accuracy, it is recommended to have the threshold on the ALM mean at 0.1 (for a normalized output). If the accuracy associated with this level of uncertainty is insufficient, it is recommended to decrease the threshold, to allow the TGP to sample more, providing it more information on the relationship between input and output. However, if this is too low, the stopping criterion may never be met. Moreover, if different realizations of the TGP occur, it is recommended to increase the number of iterations used by the MCMC.

2. Minimizing the computational cost associated with a Treed Gaussian Process (TGP)

When using SFINCS, the TGP dominates the computational cost for a large number of samples. Therefore, an argument could be made to minimize the number of times the posterior predictive surface needs to be calculated. This can be done in three ways. Firstly, the number of iterations used for the MCMC of hyperparameters should be increased to increase the accuracy of the ALM statistic. This would increase the accuracy and consistency of the number of samples needed to reach the stopping criterion. Secondly, the number of outputs should be minimized in order to reduce the effect of Equation 5.1. Finally, the TGP could be trained every x samples instead of every sample. This would decrease the cost associated with the TGP, but would increase the cost of SFINCS and FIAT as more samples would be drawn. It would also increase the cost associated with multiple outputs. An argument could also be made to increase the number of MDA samples used for the initialization of the TGP. However, this could lead to the TGP only being used for the surrogate modeling, and not the sample choice. The latter is advantageous in reducing the number of samples [Hendrickx et al., 2023] as the TGP chooses the sample that brings the largest expected information [MacKay, 1992]. Increasing the number of MDA samples may lead to more samples being drawn, which could increase the computational cost.

3. Inclusion of driver duration and lag as random variables

The analysis demonstrated that the duration and lag of a compound flood driver had a statistically significant effect on the EAD and the risk curve. When performing a compound flood risk assessment, these should be made random variables. However, in areas affected by multiple drivers, it may become difficult to successfully define the dependence structure between the different random variables. For these locations, it is recommended to take them as constants but assume a value that will produce a conservative estimate. This is because the results for Charleston show that only taking into account the driver magnitudes can be sufficient if the uncertainty of the TGP is acknowledged. This reduces the number of random variables, and also reduces the computational cost associated with using a brute-force approach to fit the regular vine, providing a more accurate way of modeling the dependence structure.

4. Number and spatial extent of outputs when using a Treed Gaussian Process (TGP)

This research showed that increasing the number of outputs increases the understanding of compound flood risk in different locations, as the relative effect of compound flood drivers varies based on the characteristics of the case study. To minimize the cost associated with the method, it is recommended to reduce the number of outputs. However, this requires some prior knowledge of where the transition zones are located. This knowledge could be improved with a preliminary sensitivity analysis. This research only focuses on economic damages. While social damages and economic damages may be correlated, their responses to compound flood events are expected to vary [Eilander et al., 2023c]. It is therefore recommended to include different outputs that vary in location, and type of damage to get the best representation of compound flood risk for a case study location.

5. Improve the quantification of extreme precipitation events

The historical precipitation magnitudes captured by ERA5 underestimate the return values provided by NOAA. This is a known drawback of using ERA5 [Lavers et al., 2022]. This underestimation of historical events leads to an underestimation of precipitation magnitudes when they are generated by the vine copula, which underestimates the damages of an event. Different methodologies are available to correct this. These methods map rainfall from ERA5 and local gauges and apply a transformation [Holthuijzen et al., 2022; Sun et al., 2022]. Therefore, data from local gauges can be extrapolated to the time record of ERA5.

6. Increasing size and number of training datasets

A large uncertainty related to the potential early stopping of the TGP on the test datasets could be resolved by comparing their outcomes with the training datasets. The accuracy of these comparisons would increase if they could be done for each dimension, and if their size was increased to allow for a more diverse set of event pairs. However, this would require significant computational resources.

7. Seasonality of events

It has been shown by previous research that the inclusion of seasonality [Couasnon et al., 2022] is vital in correctly quantifying the return values associated with tropical cyclones [Nederhoff et al., 2024]. The reason they were not included in this research is because of the short time record. There are two solutions that are often used in research. On the one hand, COAST-RP [Dullaart et al., 2022] has been used to gather a better understanding of the effect of tropical cyclones on extreme total water levels and has been used in compound flood risk assessments [Eilander et al., 2023c]. However, this does not include information on the precipitation magnitudes occurring during tropical cyclones. On the other hand, the joint probability modeling optimal sampling (JPM-OS) has been used by Diermanse et al. [2023] to simulate tropical cyclone tracks (similar to [Bakker et al., 2022]). Each track has parameters associated with it, which dictate the meteorological forcing, and by extension the magnitude of the storm surge and precipitation.

8. Optimal input reduction

Past research has used input reduction to reduce the number of samples that needed to be simulated. Sampling a posteriori has been shown to reduce the computational cost associated with higher dimensions, however, it can be significantly larger than lower dimensions if the surrogate model requires many samples. Similar to the concept of effective discharge [Wolman and Miller, 1960], it would be interesting to identify which combination of driver duration and lags can be assumed constant to arrive at a similar estimate of the risk curve and corresponding EAD in higher dimensions. This would require validation by performing a sensitivity analysis on different event sets and different constants in a lower dimensionality, which could be computationally expensive with a large number of random variables.

9. Compare with a Gaussian Process (GP)

The analysis showed that the partitions of the TGP only occurred for two of the outputs. Moreover, these partitions did not bring any additional information, in terms of flood risk assessment, and were not stable. Because of its additional computational cost, it would be interesting to compare the estimate and the cost of using a GP with a TGP. If no partitions occur, the GP is expected to provide the same estimate, but for a lower cost. However, partitions can occur before the stopping criterion. The TGP allows for heteroskedasticity, which makes it easier to only sample from interesting locations in the input space, leading to a smaller number of samples to simulate.

Appendices

A

Supplementary Information For Damage Modeling

A.1. Hazard Model Governing Equations

The following set of equations (eq. (A.1), eq. (A.2), eq. (A.3)) are used by [SFINCS](#) to compute the water levels and flow rates at each time step and sub-grid cell. For hazard modeling, only the maximum water depth for the different sub-grid cells in the model domain (fig. 3.3) has to be retained.

$$\zeta_{m,n}^{t+\Delta t} = \zeta_{m,n}^t + \left(\frac{(q_{x,m-1,n}^{t+\Delta t} - q_{x,m,n}^{t+\Delta t})}{\Delta x} + \frac{(q_{y,m,n-1}^{t+\Delta t} - q_{y,m,n}^{t+\Delta t})}{\Delta y} + S_{m,n} \right) \Delta t \quad (\text{A.1})$$

$$q_{x,m,n}^{t+\Delta t} = \frac{\left[\theta q_{x,m,n}^t + \frac{1-\theta}{2} (q_{x,m+1,n}^t + q_{x,m-1,n}^t) \right] - \left(g h_{x,m,n}^t \frac{(\zeta_{m+1,n}^t - \zeta_{m,n}^t)}{\Delta x} + adv_x - \frac{\tau_{w,x}}{\rho_w} \right) \Delta t}{(1 + g \Delta t n^2 q_{x,m,n}^t / h_{x,m,n}^t)^{7/3}} \quad (\text{A.2})$$

$$q_{y,m,n}^{t+\Delta t} = \frac{\left[\theta q_{y,m,n}^t + \frac{1-\theta}{2} (q_{y,m,n+1}^t + q_{y,m,n-1}^t) \right] - \left(g h_{y,m,n}^t \frac{(\zeta_{m,n+1}^t - \zeta_{m,n}^t)}{\Delta y} + adv_y - \frac{\tau_{w,y}}{\rho_w} \right) \Delta t}{(1 + g \Delta t n^2 q_{y,m,n}^t / h_{y,m,n}^t)^{7/3}} \quad (\text{A.3})$$

In the [SFINCS](#) model for Charleston used by [Diermanse et al. \[2023\]](#), the advection term adv was set to zero. Furthermore, from the shape and geometric constraints of the water bodies of the case study, it is further assumed that wind setup is negligible (wind stress, τ_w is always zero). $S_{m,n}$ is the source term and can represent precipitation, infiltration, or discharge from a point source [[Leijnse et al., 2021](#)]. ζ represents the water level, and q is the discharge per meter width. For other variables, see [Leijnse et al. \[2021\]](#).

A.2. Removing Permanently Flooded Buildings From Impact Model

The DELFT-Fast Impact Assessment Tool ([FIAT](#)) model defined for Charleston gives estimates of total economic damages that are too large for a median Higher High (HH) tide. After visualizing which buildings contributed to this damage, it was realized that the local bathymetry gives incorrect estimations of the ground floor height of buildings along the waterfront. Since a median HH tide occurs frequently, it will make the estimate of the Expected Annual Damages (EAD) dependent on the extreme rate of the historical event set. To solve this, a model ([Figure 4.2](#)) with the following boundary conditions is used: linearly superpose a median HH tide with a Mean Sea Level (MSL) of 0.2 m, a discharge of $200 \text{ m}^3 \text{ hr}^{-1}$, and no precipitation. The output attributes an economic damage value for all building footprints (object IDs) contained in the model. The object IDs of buildings causing damages were then removed from the DELFT-[FIAT](#) model.

B

Time Records of Drivers

This appendix presents observations made on the time record of the different compound flood drivers when preprocessing them. Firstly, section B.1 shows how the signal of the non-tidal residual was made stationary. Secondly, section B.2 investigates the daily inequality observed in the tidal time record. Thirdly, section B.3 shows the effect of the dam management on the upstream discharge observations. Finally, section B.4 shows the extent to which ERA5 underpredicts precipitation extremes.

B.1. Stationarity of the Non-Tidal Residual

After removing the tidal signal from the total water level to obtain the non-tidal residual, a trend caused by a change in the Mean Sea Level (MSL) was present. To apply Peak Over Threshold (POT), the time record needs to be stationary, and the trend needs to be removed. This was done using a one-year moving average to minimize the bias introduced when applying POT [Arns et al., 2013]. This trend can be observed in fig. B.1, which shows the value of MSL over the time record of available observations steadily increases. There are periods of seasonality that create a cyclic behavior. Since 2010, the rate of the MSL increase has increased when compared to previous years.

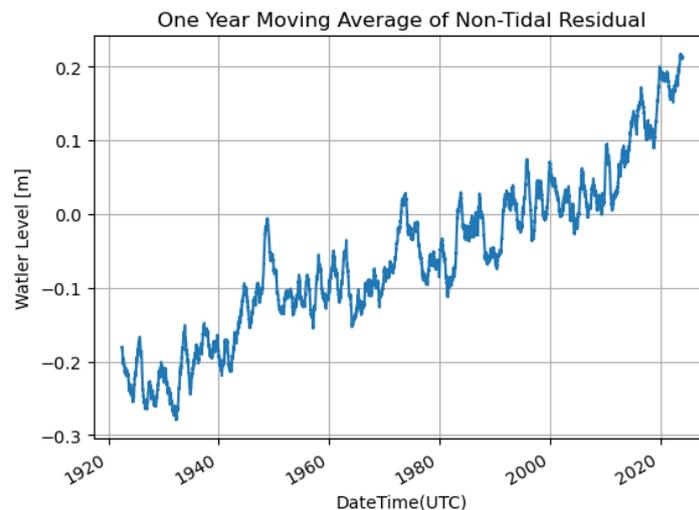


Figure B.1: 1-year moving average of non-tidal residual for available time record at the NOAA gauge in Charleston.

To better characterize the magnitude of the non-tidal residual, skew surge [Williams et al., 2016] is used. This means observations go from an hourly resolution to a 12-hour resolution (Charleston has a semi-diurnal tidal regime [Bosboom and Stive, 2021]). Therefore, the duration of skew surge events will occur in 12-hour increments.

The boundary conditions for the downstream water level are obtained by linearly superposing the tide, **MSL**, and the skew surge. From fig. B.1, the **MSL** does not increase at a constant rate but fluctuates. This results in the **MSL** at the end of the time record not having the largest observed value. This could be caused by atmospheric patterns or tidal constituents that have a time period larger than a year. A decision therefore has to be made when defining the magnitude of **MSL**. Overall, fig. B.1 shows a positive trend, therefore, as a conservative estimate, the maximum observed **MSL** during the coinciding record of all drivers is used. This occurred in 2019 (the record is between 1998 and 2023), resulting in an **MSL** of 0.2 meters.

A one-year moving average does not remove monthly seasonality. Figure B.2 shows the seasonality of Monthly Mean Sea Level (**MMSL**). Each month contains a box plot. The orange line represents the mean, while the upper and lower limits of the box represent the 75th and 25th quantile respectively. The whiskers show the minimum/maximum of: the addition/subtraction of 1.5 times the interquartile range and the 75th/25th quantile, or the maxima/minima for that month. Larger water levels can be expected in August, September, October, and November, which are caused by the hurricane season (chapter 3) and smaller pressures associated with warmer temperatures. In December, January, February, March, and April, it can be expected that lower water levels will occur, caused by larger pressures associated with colder temperatures. The largest upper quantile occurs in September and causes a **MMSL** of 0.2 meters.

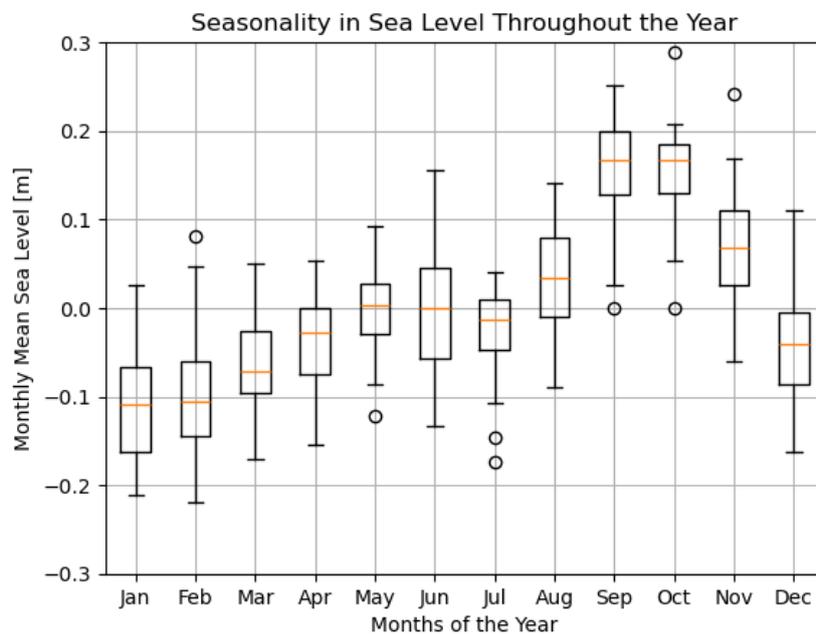


Figure B.2: **MMSL** seasonality of non-tidal residual at the **NOAA** gauge in Charleston after removing a one-year moving average.

B.2. Daily Inequality Caused by the Beating of Tidal Constituents

Figure B.3 shows the frequency of daily inequalities in high tides observed at the **NOAA** Charleston gauge. The daily inequality has a maximum of 0.53 meters with the largest density occurring between 0.2 and 0.25 meters (fig. B.3). Since the **POT** is applied to the skew surge (and not the total water level), it shows the importance of only investigating events that occur during Higher High (**HH**) tides. This ensures the water levels that co-occur with an extreme skew surge can be considered extreme. For example, for a large daily inequality, if the peak of a storm surge event occurs during a lower high tide, it may lead to a smaller total water level than the tail of the event occurring during a **HH** tide. While this can happen in reality, the parametrization of events (section 4.4.3) may introduce artificial noise (or uncertainties) when using a Treed Gaussian Process (**TGP**) (section 4.5.2).

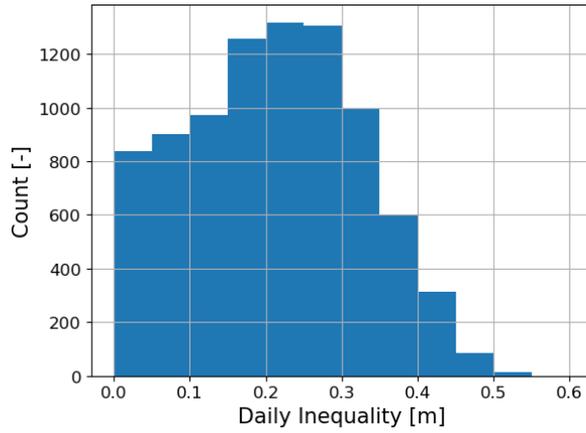
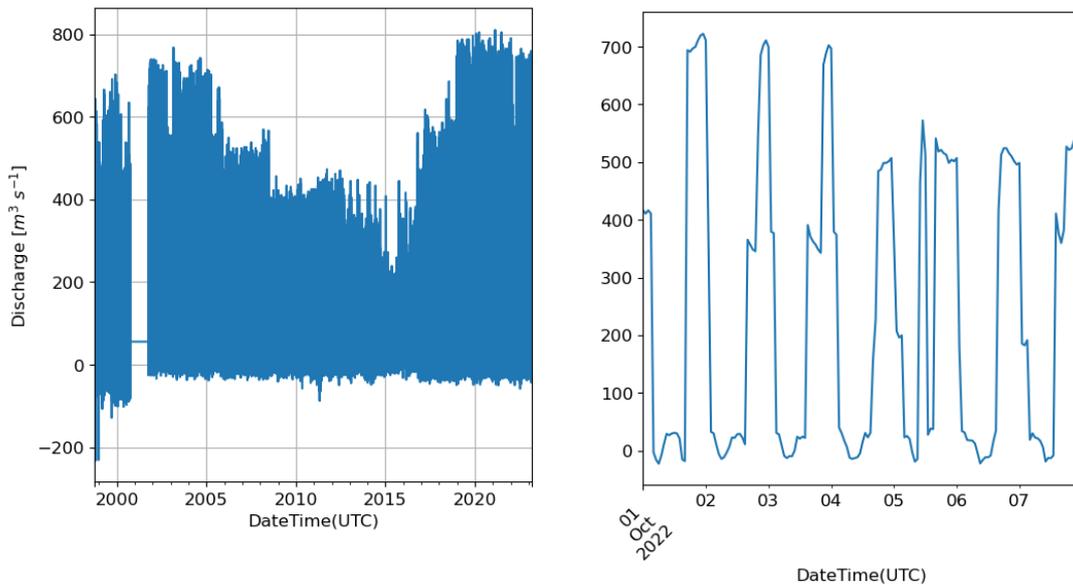


Figure B.3: Daily inequality histogram observed at the NOAA gauge in Charleston.

B.3. Dam Management of Upstream Discharge

The time record of the upstream discharge is non-stationary (fig. B.4), making it difficult to apply POT. The non-stationarity for water level is caused by a regular change in the datum (MSL), making it relatively easy to correct. In this case, the non-stationarity is caused by dam management, which causes hydropeaking. Hydropeaking occurs when the hydrograph shows peaks that occur during dam releases, which follow electricity demand. This is difficult to correct as it causes the discharge not to follow its natural variability, dictated by the properties of its upstream catchment (catchment properties decide the base flow, properties related to runoff and lag between peak precipitation in the catchment and peak fluvial discharge [Hendry et al., 2019]). The results of chapter C meant this non-stationarity was not corrected.

As shown in fig. B.4a, these management periods can last on the scale of several years. In the right plot of fig. B.4b, it can be seen that for the most recent period of management, discharge is often released once a day with a pre-defined magnitude and duration. This is further supported by fig. B.5. Figure B.5a shows that releases often occur around 5:00 - 8:00 or 12:00 - 14:00 local time. The Figure B.5b shows that these releases are frequently associated with a daily maximum of 400 or 700 meters cubed per second.



(a) Time record of discharge showing the different periods of dam management as well as the missing data (2001). (b) Time series for a random week in the current period of dam management.

Figure B.4: Discharge time record is affected by hydropeaking.

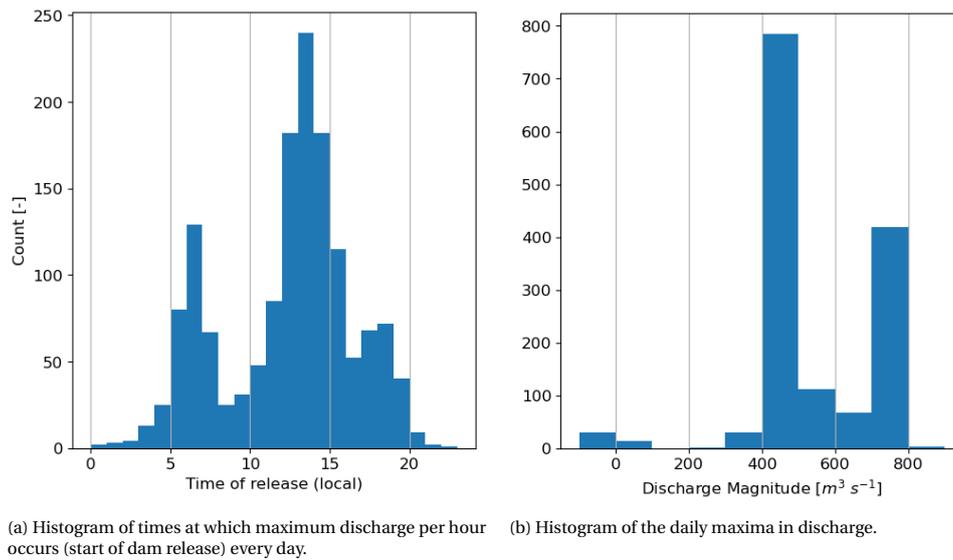


Figure B.5: Hydropeaking statistics for the current period of dam management (January 2019 - March 2023).

B.4. ERA5 Precipitation

In previous research, ERA5 has been shown to capture the location and shape of extreme precipitation events but cannot capture extreme precipitation totals [Lavers et al., 2022]. This is also the case for Charleston. fig. 3.2 predicts 2.1 to 2.2 inches per hour (equivalent to 53-56 $mm hr^{-1}$) for a 2-year Return Period (RP) for the coordinate at which precipitation data is collected. For the time record of coinciding data, ERA5 reanalysis has a maximum precipitation magnitude of 23 $mm hr^{-1}$. Since local rain gauges are not available for the required time record, machine learning techniques can be used to correct ERA5 data [Sun et al., 2022]. This step would have to be taken for an accurate compound flood risk assessment. However, performing this methodology is outside the scope of this research, and the underestimation of precipitation magnitudes will be considered a limitation.

C

Sensitivity Analysis of Three Compound Flood Drivers

This appendix describes the method and corresponding results of the sensitivity analysis performed to obtain the driver dominating the total economic damages.

For precipitation and skew surge, a Gaussian shape is assumed. For Gaussian's, a heuristic exists, where the standard deviation is equal to the duration of events. This duration is taken as 12 hours for both drivers. Moreover, no time lag between precipitation and skew surge is assumed. Thus, only the magnitude is needed to correctly define the Gaussian and the boundary condition. For discharge, a constant magnitude is assumed to simplify the modeling of the boundary condition. Finally, a median Higher High (HH) tide is assumed to coincide with the skew surge. This is done to reduce the relative importance of the downstream boundary condition.

To decrease the potential dominance of the coastal water level, the skew surge time series is linearly superposed with a median HH tide and maximum observed Mean Sea Level (MSL). Table C.1 shows the combinations of events used to test the sensitivity of the economic damages to the three compound flood drivers: precipitation, discharge (Cooper River), and skew surge.

Table C.1: The 8 simulations used to conduct a sensitivity analysis to determine which compound flood driver contributes the most to the total economic damages of Charleston County. For S Mag, a constant magnitude is assumed. For precipitation, a Gaussian shape is assumed, where its peak (P Mag) coincides with the HH tide (no time lag). For discharge, a constant magnitude (Q Mag) is assumed. The duration of all three drivers is 12 hours. Since the simulation is only 12 hours long, an artificial tidal signal is created which assumes a sinusoidal shape with a tidal period of 12.42 hours (M2 constituent), which has a magnitude of the HH tide. Finally, MSL is assumed to be constant.

Simulation	Coastal Water Level [m]			P Mag [mm hr ⁻¹]	Q Mag [m ³ s ⁻¹]
	S Mag [m]	HH Tide [m]	MSL [m]		
1	0	0.85	0.2	0	200
2	0	0.85	0.2	0	800
3	0.97	0.85	0.2	0	200
4	0.97	0.85	0.2	0	800
5	0	0.85	0.2	23	200
6	0	0.85	0.2	23	800
7	0.97	0.85	0.2	23	200
8	0.97	0.85	0.2	23	800

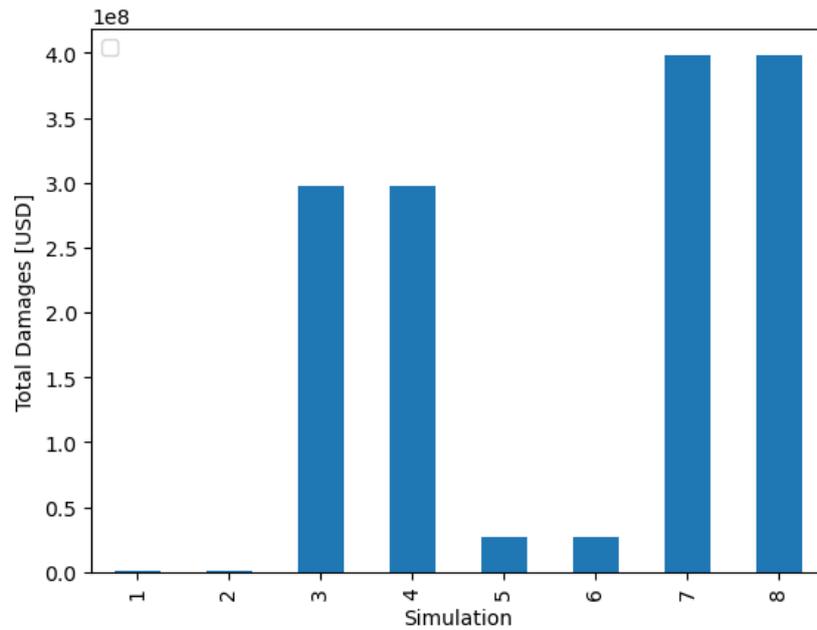


Figure C.1: Using fig. 4.2 and simulations shown in table C.1 to perform a sensitivity analysis of the compound flood driver magnitudes.

The sensitivity analysis (Figure C.1) shows the economic damages associated with the different simulations described in table C.1. It shows that discharge does not affect economic damage regardless of the combination of compound flood drivers. This can be seen when comparing simulation 2 with 1, 4 with 3, 6 with 5, and 8 with 7. This is probably caused by the proximity of Charleston County to the open coast. The backwater effect means the coastal water level dominates the magnitude of the water level. If the Fast Impact Assessment Tool (FIAT) model was also defined for the county upstream (Berkeley), the backwater effect would decrease the dominance of the coastal water level and would show the effect of the so-called transition zone [Gori et al., 2020; Jane et al., 2022], probably resulting in discharge causing economic damages. Therefore, for the extent of the FIAT model, the sensitivity analysis shows that fluvial discharge can be safely removed as a compound flood driver, reducing the number of drivers and dimensions.

This proximity to the open coast also demonstrates that the skew surge events dominate the economic damages. This can be seen when comparing simulations 3 and 4 with 5 and 6. Comparing simulations 3, 4, and 5, 6 with 7 and 8 shows that precipitation introduces a compound effect, as the economic damages increase when compared to simulations 3 and 4. Therefore, skew surge can be identified as the dominant driver, and precipitation should be considered a compound flood driver.

D

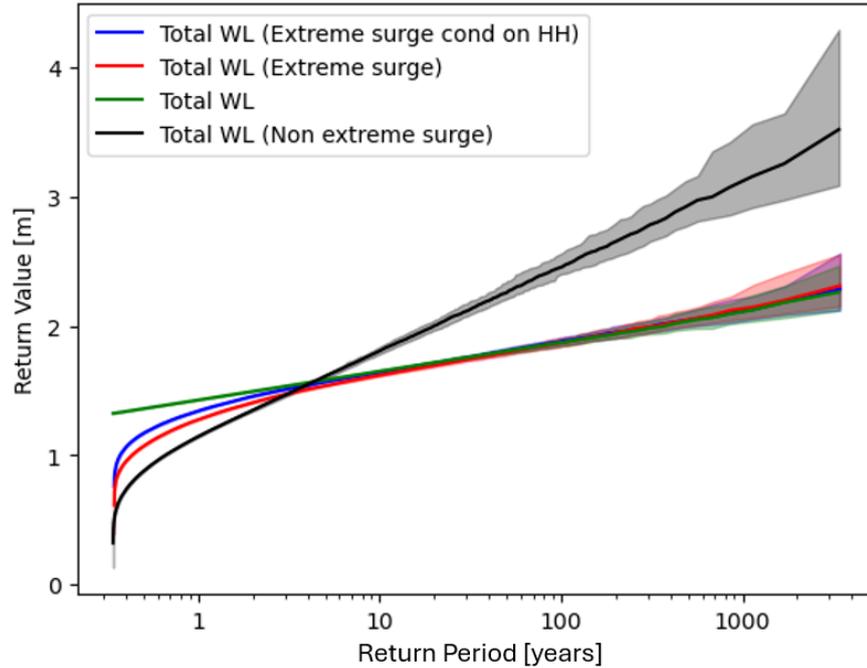
Sensitivity of Statistical Models to Peak Over Threshold (POT)

The sensitivity analysis on economic damages showed that skew surge was the dominant driver (chapter C). In this appendix, the methodology described in section 4.3 is applied. Firstly, in section D.1 the events used to quantify the skew surge extremes and co-occurring tides are defined (total water level extremes). Secondly, Peak Over Threshold (POT) is applied to precipitation to identify the precipitation extreme events for the multivariate sensitivity analysis (section D.2). Thirdly, a multivariate sensitivity analysis is performed (section D.3). Finally, a univariate sensitivity analysis is performed on the threshold used for the skew surge POT (section D.4). The results of these sections provide an answer for the definition of the skew surge, and the threshold to use to identify the historical observations of compound flooding for Charleston (section D.5).

D.1. Quantifying the Total Water Level

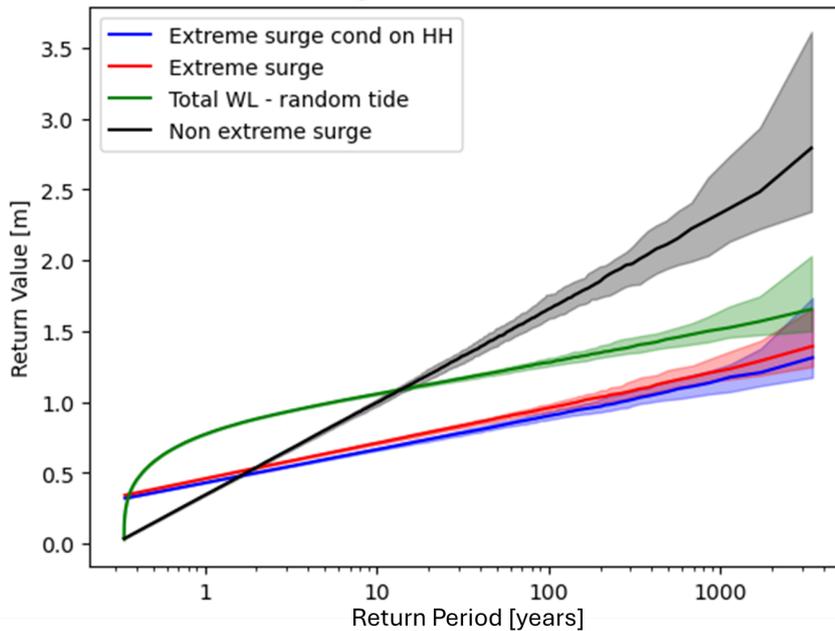
An informed decision had to be made on how to quantify the Tidal Magnitude (*T Mag*) that corresponds with skew surge events, as this would affect the magnitude of identified skew surge events and corresponding total water levels. A preference is given to performing the Extreme Value Analysis (EVA) on the skew surge as it is expected to have a higher correlation with other weather-related drivers (see section D.3.1 for a confirmation of this expectation). However, the large daily inequality (section B.2) means including all *T Mags* could cause a stochastic event to have the maximum total water level that does not co-occur with the maximum skew surge. This is not desired because of the parametrization of the boundary conditions, which may increase the uncertainty associated with a stochastic event set when using a Treed Gaussian Process (TGP). A sensitivity analysis was therefore performed in fig. D.1. Figure D.1 shows 80 different realizations of the empirical marginal probability distribution when generating 10,000 stochastic events for different definitions of skew surge (fig. D.1b) and total water level (fig. D.1a). Firstly, extreme surges conditional on Higher High (HH) tides are investigated (blue). Total water levels are obtained by linearly superposing a statistically independent HH tide [Williams et al., 2016]. Secondly, extreme surge conditional on any tides is investigated (red). Total water levels are obtained by linearly superposing a statistically independent higher tide [Williams et al., 2016]. Thirdly, extreme total water levels are investigated (green). Skew surge is obtained by removing a statistically independent high tide [Williams et al., 2016]. Finally, a non-extreme probability distribution (section 4.4.2.1) is fitted to the skew surge historical observations which were identified by the total water level EVA (black). Total water levels are obtained by linearly superposing a random high tide [Williams et al., 2016].

10,000 Simulated Total WL Return Values for an Extreme Rate of 2.91



(a) Effect of definition on total water level extremes.

10,000 Simulated Skew Surge Return Values for an Extreme Rate of 2.91



(b) Effect of definition on skew surge extremes.

Figure D.1: Investigating the definition of the total water level on the quantification of extreme events.

Differences in surge magnitudes can be quite significant, but the total water level is more important, as it is the boundary condition of the hydrodynamic model. After a Return Period (RP) of 10 years, all curves bar non-extreme surge are approximately equivalent. Before this RP, extreme surges underestimate the total water level. This is because the extreme water levels (green) are king tides co-occurring with small skew surges, which means the skew surge magnitude is conditional on the tide (small surge, extreme tides). This underestimation is smaller if the skew surge is conditional on HH tides (blue versus red). For RPs smaller

than 2-3 years, the difference becomes significant. This can be seen as an advantage as it means extreme precipitation events can co-occur with non-extreme total water levels (non-compound). This reduces the effect of using a dominant driver to obtain historical compound flood events. This underestimation is also not a problem if flood defenses are designed for a RP of 2 years.

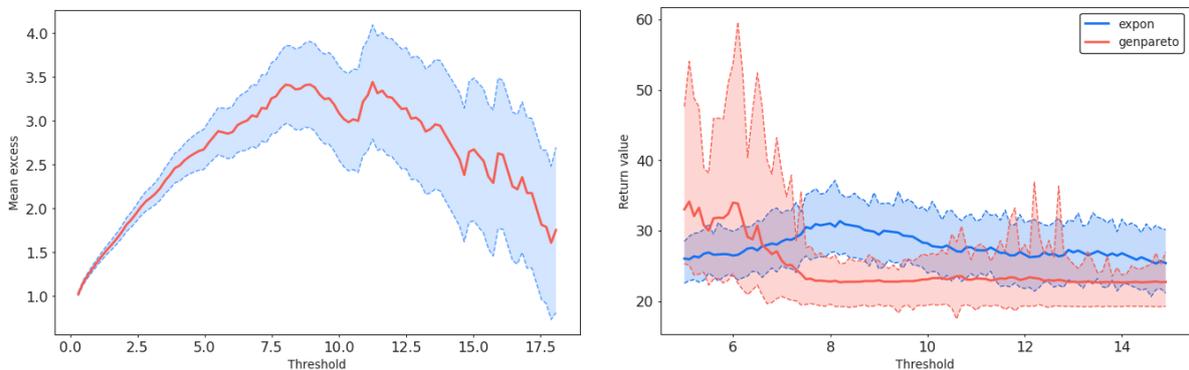
Combining this information with the expectation that skew surge provides a better correlation with precipitation (see section D.3.1), means skew surge which is conditional on HH tides is used as a random variable to apply POT.

D.1.1. Non-Extreme Marginal Probability Distribution Functions

The non-extreme surge shows the effect of representing extremes with a non-extreme probability distribution (fig. D.1). It underpredicts the return values associated with lower RPs as the density is dominated by non-extreme historical observations. It is also unable to correctly capture the tail of the historical observations and over-predicts the return values associated with large RPs. These observations are expected to also occur for other non-extreme marginal probability distribution functions for the magnitude of drivers.

D.2. Applying POT to Non-Dominant Drivers

Figure D.2 shows the plots used to determine a stable threshold for the non-dominant drivers. In the case of Charleston, this only had to be applied to precipitation. This was performed to compute the number of extreme precipitation events. Figure D.2a shows the mean excess rate when the threshold is varied. Figure D.2b shows the magnitude of the 100-year return value when varying the threshold for exponential and Generalized Pareto Distribution (GPD) curves. It shows the exponential curve predicts a larger magnitude of the return value, but is stable for a smaller range of thresholds. The return value for a 100-year RP is stable between 11 and 14 mm hr⁻¹, and the mean residual life is approximately linear between 11 and 14.5 mm hr⁻¹. It was therefore decided to use a threshold of 11.86 mm hr⁻¹ as a way to include more samples. This resulted in an extreme rate of 1.11 extreme events per year.

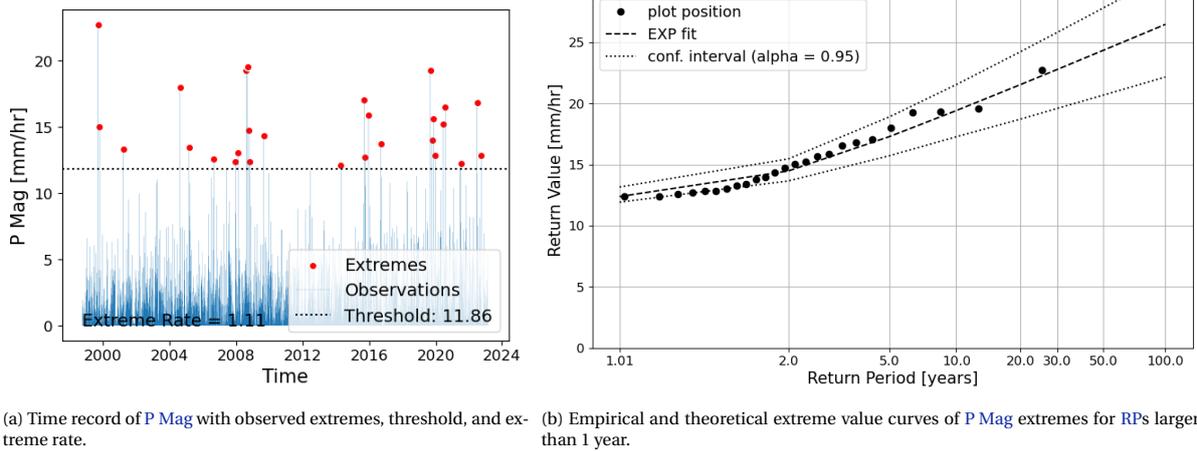


(a) Mean excess rate of different thresholds when applying POT to P Mag. The y-axis is the mean excess. The x-axis is the threshold. Both have mm hr⁻¹ as units.

(b) Return value stability for a RP of 100 years when applying POT to P Mag. The exponential extreme value distribution is 'expon' (blue) and the GPD is 'genpareto' (red). The 95% confidence intervals are created by bootstrapping 100 samples. The y-axis is the return value. The x-axis is the threshold. Both have mm hr⁻¹ as units.

Figure D.2: Univariate stability of POT on P Mag. Plots are created using the pyextremes python packages [Bocharov, 2023].

Figure D.3a shows the identified extremes (red) alongside the time record of the skew surge. Figure D.3b shows the exponential (smaller BIC value than the GPD) extreme value curve and the 25 events which have the largest skew surge magnitudes. It shows the 2 and 100-year return values are approximately 15 and 27 mm hour⁻¹ respectively. Since fig. D.3b is not used for the rest of the analysis, its accuracy with the historical observations is not important. However, a comparison of 2-year return values can be made with fig. 3.2 to show the extent to which ERA5 underpredicts extremes for this particular threshold and location. Figure 3.2 estimated the 2-year return value to be 53-56 mm hour⁻¹. Figure D.3b estimates this value to be 15 mm hour⁻¹. For a 2-year RP, ERA5 underpredicts the magnitude of events by a factor of approximately 3.5. Depending on the RP, this effect is enhanced or reduced by using a non-extreme probability distribution function (section D.1.1). For large RPs, the effect is expected to be reduced providing more meaningful extremes.



(a) Time record of $P\ Mag$ with observed extremes, threshold, and extreme rate.

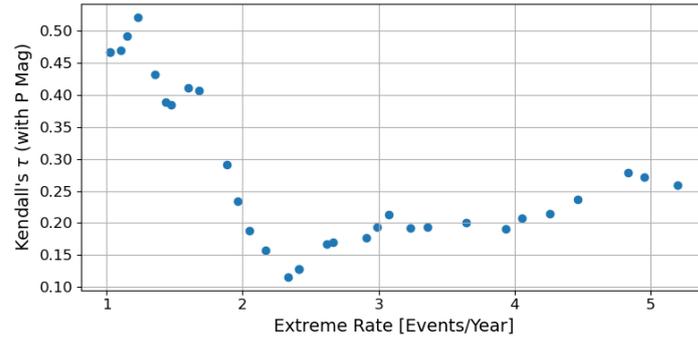
(b) Empirical and theoretical extreme value curves of $P\ Mag$ extremes for RPs larger than 1 year.

Figure D.3: Output of EVA of $P\ Mag$ using POT with a threshold of $11.86\ \text{mm}\ \text{hr}^{-1}$ (extreme rate of 1.11). Performed using the HydroMT package [Eilander et al., 2023a] and code obtained from Eilander et al. [2023c].

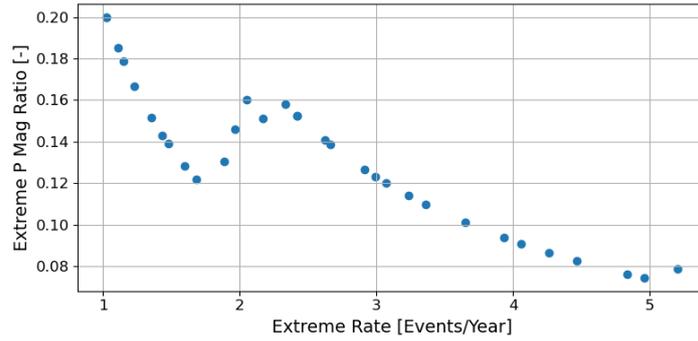
D.3. Multivariate Sensitivity of POT

For the multivariate sensitivity, stability in the correlation and the number of extreme precipitation extremes is desired. Moreover, a large number of extreme precipitation events are desired as it makes it easier to model the tail of its marginal probability distribution. This ensures that a particular instance of the threshold does not have a significant effect on the statistical models used to characterize dependence and the probability distribution function of the precipitation marginal.

Figure D.4 shows the effect of including more events in the skew surge POT by decreasing the threshold and increasing the extreme rate on two variables of interest: correlation with $P\ Mag$, and the relative number of $P\ Mag$ which can be considered extreme. Figure D.4 shows an extreme rate of 1 has the highest correlation and ratio of extreme precipitation events. The small number of events present in the historical event set could have caused this, as the combined record is 25 years, which means the extreme rate of 1 means only 25 events will be present in the historical event set. A small number of events also make it harder to fit dependence models and probability distribution functions to the marginals. It is therefore preferred to have a larger extreme rate to avoid this. However, as the extreme rate increases, both the correlation and ratio of extreme precipitation events decreases. In fig. D.4b, it can be visually seen when a new precipitation event is included, as there is a sudden increase in the ratio of extreme precipitation events. For example, at an extreme rate of 1, 5 events are included. It is only at an extreme rate of 1.7 that an additional extreme precipitation event is added. Between extreme rates of 2.4 and 5, no additional events are identified. It can also be visually observed that additional extreme precipitation events decrease the correlation in fig. D.4a. This is caused by extreme precipitation events coinciding with the smallest surges when they are initially introduced into the event set. This can be visualized in fig. D.5, where the vertical blue line, which represents the threshold of the skew surge, slowly shifts to the left when decreasing the magnitude of the threshold. It is also for this reason that the correlation increases or is stable when the precipitation ratio decreases, as the additional events that are added are probably small for (extreme) skew surge and precipitation.



(a) Sensitivity of Kendall's τ_B (between S Mag and P Mag).



(b) Sensitivity of extreme precipitation (P) ratio.

Figure D.4: Sensitivity of multivariate metrics when increasing the extreme rate for the POT applied to S Mag.

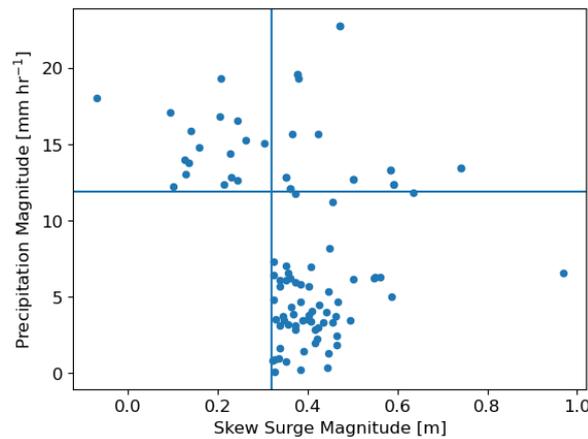


Figure D.5: Events identified by POT when applied on P Mag and S Mag. To identify event pairs, the definition of compound is used (window of 3 days before and after identified event, the largest magnitude is taken). The upper right quadrant shows events that are both extreme. Events to the right of the vertical line (represents the threshold used for POT of S Mag) are all identified by the POT of skew surge. Events above the horizontal line (represents the threshold used for POT of P Mag) are all identified by the POT of P Mag. The marginal PDF are not valid below their thresholds. Using the results of the preliminary sensitivity, it was decided to only use the events identified by the POT of S Mag, making the joint and the results dependent on extreme S Mag.

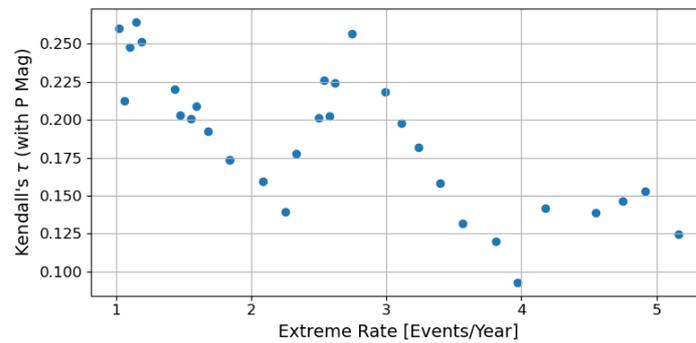
In summary, fig. D.4 shows that the correlation between the two compound flood drivers is stable when no additional precipitation extremes are added to the historical event set. Moreover, it is easier to fit statistical models when there is a sufficiently large number of events. Additionally, it is easier to define the tail behavior of the probability distribution for the precipitation magnitude marginal when larger magnitudes are included in the event set. Therefore a threshold that coincides with an extreme rate greater than 2.5 seems to be the preferred option as it maximizes the number of extreme precipitation events. This does mean that the non-extreme precipitation magnitudes will have more weight when defining the probability distribution functions

for the marginals, and could skew the dependence structure by making it less likely to generate coinciding extremes in both skew surge and precipitation. The univariate stability will now be investigated to make a more informed decision.

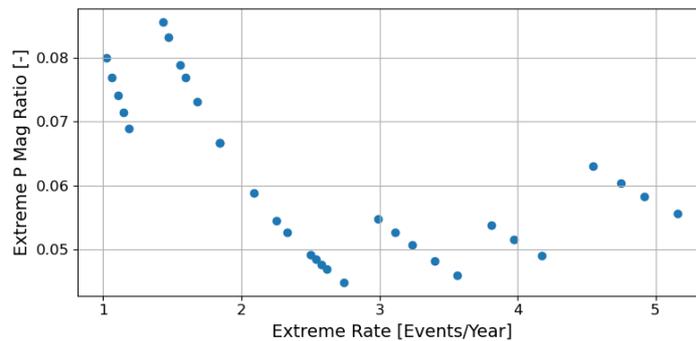
D.3.1. Multivariate Sensitivity When Applying POT to the Total Water Level

This subsection describes why *S Mag* and *T Mag* are linearly superposed instead of using the total water level. To this end a multivariate sensitivity analysis (similar to section D.3) was applied to the total water level. The results of this analysis are compared with section D.3.

When comparing fig. D.6 with fig. D.4, some clear differences in the sensitivity of the multivariate metrics can be observed. Firstly for fig. D.6a and fig. D.4a, Kendall's τ_B is smaller for the total water level for extreme rates between 1 and 2. Additionally, the total water level only reaches a period of stability for Kendall's τ_B after an extreme rate of 4, when compared to 2.5-3 for *S Mag*. Secondly, for fig. D.6b and fig. D.4b, the extreme precipitation ratio is smaller for all extreme rates for the total water level when compared to *S Mag*. This means using the total water level instead of the skew surge will identify less extreme *P Mag* events.



(a) Sensitivity of Kendall's τ_B (between the total water level and *P Mag*).



(b) Sensitivity of extreme precipitation (P) ratio.

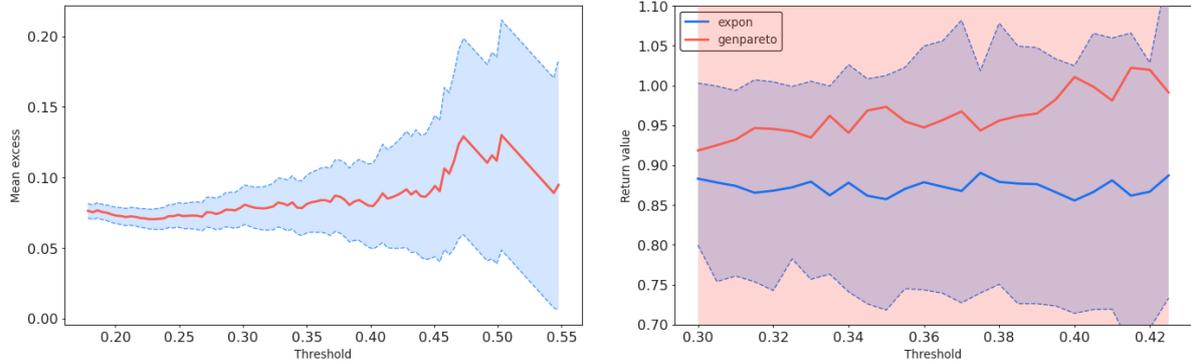
Figure D.6: Sensitivity of multivariate metrics when increasing the extreme rate for the POT applied to the total water level.

These observations are caused by the total water level extremes co-occurring with *P Mags* which are not always extreme. This is because the *T Mag* can be considered independent of the weather. This has two effects. On one hand, since the total water level is caused by the linear superposition of the tide and the skew surge, weather patterns may cause extreme *S Mags* which may co-occur with extreme *P Mags*, however these may co-occur with small *T Mags* resulting in a non extreme total water level. On another hand, larger *T Mags* may co-occur with a weather pattern which causes small *P Mags* and *S Mags*. As shown in section D.1, large *T Mags* are sufficient in generating extreme total water levels for the case study of Charleston. These extreme total water levels than coincide with non extreme *P Mags*.

Therefore to ensure the historical event set contains a large number of extreme *P Mag* events, and a good correlation with *P Mag*, it is preferred to use *S Mag* over the total water level to identify historical flood events.

D.4. Univariate Sensitivity of POT

Figure D.7a shows the plots used to determine a stable threshold for *S Mag* which was identified as the dominant driver. Figure D.7a shows the mean excess rate when the threshold is varied and is stable for thresholds ranging between 0.3 and 0.42 meters. Figure D.7b shows the magnitude of the 100-year return value when varying the threshold. For exponential and GPD curves. It shows the exponential curve predicts a smaller magnitude of the return value, but is stable for the range of thresholds identified by fig. D.7a. Since BIC is used as a metric to fit the extreme value distribution, the exponential distribution is likely to be preferred, which means any threshold between 0.3 and 0.42 meters is a good option.



(a) Mean excess rate of different thresholds when applying POT to *S Mag*. The y-axis is the mean excess. The x-axis is the threshold. Both have meters as units.

(b) Return value stability for a RP of 100 years when applying POT to *S Mag*. The exponential extreme value distribution is 'expon' (blue) and the GPD is 'genpareto' (red). The 95% confidence intervals are created by bootstrapping 100 samples. The y-axis is the return value. The x-axis is the threshold. Both have meters as units.

Figure D.7: Univariate stability of POT on *S Mag*. Plots are created using the pyextremes python packages [Bocharov, 2023]. For an easier comparison with Figure D.4: extreme rate = [1.35, 2.05, 2.42, 2.91, 3.93], and threshold [m] = [0.4, 0.36, 0.34, 0.32, 0.29].

D.5. Conclusion of Sensitivity Analysis

Using the results from the multivariate and univariate sensitivity analysis, an informed decision can now be made on the threshold used to define the POT of *S Mag* and the corresponding historical event set. To maximize the number of events, number of *P Mag* extremes, and the correlation between *S Mag* and *P Mag*, while minimizing the noise caused by non-extreme events when fitting different statistical models, an extreme rate of 2.91 (threshold of 0.32 m) was chosen.

E

Correlation Analysis

This appendix provides a correlation analysis of the historical event set (section E.1). It also investigates the effect of splitting the event set into two on the correlation between bivariate pairs of random variables (section E.2).

E.1. Historical Event Set

Figure E.1 shows the identified historical event set for Skew surge Magnitude (**S Mag**), Precipitation Magnitude (**P Mag**), Tidal Magnitude (**T Mag**), Skew surge Duration (**S Dur**), Precipitation Duration (**P Dur**), and Precipitation Lag (**P Lag**). These are plotted in the uniformly ranked space. This means the events which are associated with the minimum and maximum magnitude of a probabilistic variable have ranks $\frac{1}{N+1}$ and $\frac{N}{N+1}$ respectively, where N is the number of events in the historical event set. For **P Lag**, ranks of 0 and 1 correspond to -3 days and +3 days respectively. The correlation coefficient in the form of Kendall's τ_B is shown in the upper left corner of each bivariate plot. Table E.1 shows the p-values associated with each Kendall's τ_B . The null hypothesis uses Kendall's τ_B value of 0. Therefore, if the p-value is smaller than 0.05, the correlation is significant at the 5% level. The event set can be split into events identified by the **POT** applied on precipitation (fig. D.3) and other events. These extreme precipitation events and table E.1 can be used to show if the identified correlation (positive, negative, none) holds for extreme events.

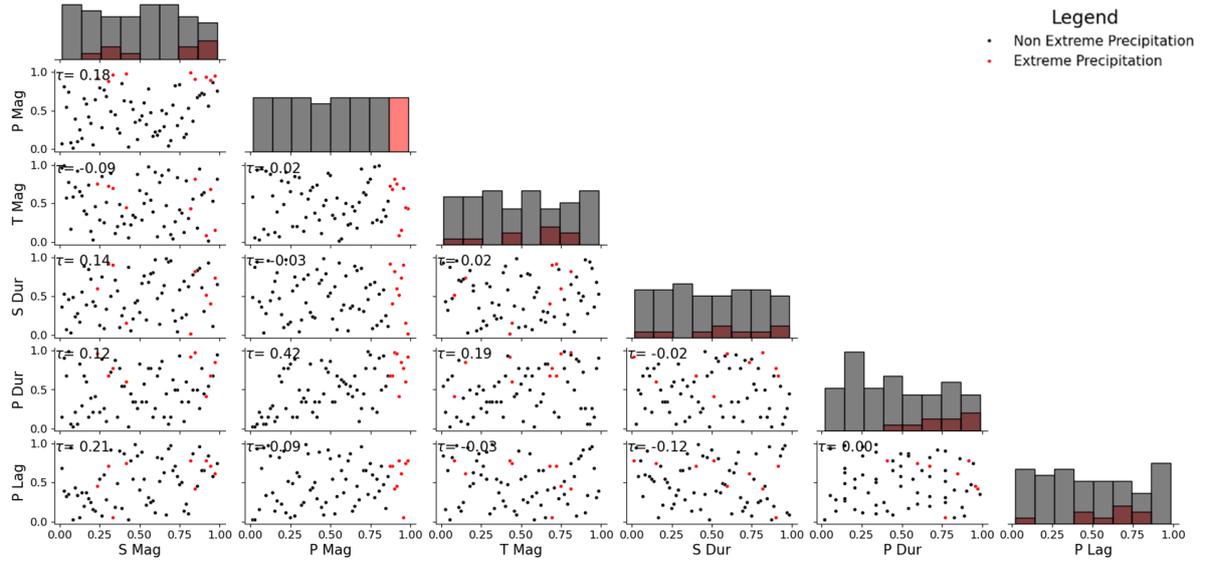


Figure E.1: Scatter matrix for the event set using the Seaborn python package [Waskom, 2021], which shows the correlation between different dimensions in the uniformly ranked space. Includes S Mag, P Mag, T Mag, S Dur, P Dur, and P Lag. Along the main diagonal, a histogram for each dimension is plotted. It shows the probability density for non-extreme (black) and extreme precipitation (red) events. If both histograms are combined, a uniform density is achieved. Kendall's τ_B for each combination can be found in the upper left corner of the respective subplot.

Table E.1: P-values associated with each Kendall's τ_B shown in Figure E.1. Shades of green show bivariate pairs of probabilistic variables that have significant correlation at the 5% level.

	S Mag	P Mag	T Mag	S Dur	P Dur
P Mag	0.027				
T Mag	0.27	0.84			
S Dur	0.09	0.69	0.80		
P Dur	0.14	$4.6e-07$	0.025	0.83	
P Lag	0.011	0.26	0.72	0.15	0.97

Table E.1 shows that only four pairs of variables are significant. Firstly, the highest confidence is associated with P Mag and P Dur. This positive correlation means that a large P Mag is often associated with a large P Dur. The lower tail for these variables in fig. E.1 shows a stronger dependence, showing the effect of including non-extreme P Mag events in the event set. Secondly, the next highest confidence is associated with P Lag and S Mag, showing a positive correlation. Negative lag is associated with smaller S Mags, and positive lag is associated with larger S Mags. This bivariate pair shows lower and upper tail independence in fig. E.1. The third pair which has a significant positive correlation is the T Mag and P Dur. This could be caused by the seasonality of the tidal beating, and underlying weather patterns causing larger durations of rain. Finally, S Mag and P Mag are also significant. This was expected as the threshold was chosen based on the value of Kendall's τ_B (section D.3). Figure D.4 showed that including a smaller number of events would increase the correlation, demonstrating the noise introduced by non-extreme events. There is also tail independence for this pair of variables in fig. E.1.

Table E.1 also shows that S Dur is independent of other variables. One potential reason for this is a small amount of noise was introduced to break the ties caused by the duration being defined to the nearest tidal cycle (i.e. 12 hours). Another outcome from table E.1 is the independence of P Lag and P Mag. Finally, the independence of S Mag and T Mag corresponds with the outcomes of Williams et al. [2016].

E.2. Splitting Event Set Based on Precipitation Lag

Chapter 2 showed that different values of lag caused differences in correlation [Khanal et al., 2019; Wahl et al., 2015]. Therefore, if large P Mags are associated with a specific P Lag, and small P Mags are associated with

large values of positive and negative P Lag, Kendall's τ_B will show independence. One way to resolve this would be to split the event set into two. However, this would be dependent on the number of events in either split. This method would result in different kernel density estimates (fig. E.4) as the bivariate pairs of random variables showing statistically significant correlation changes (table E.3 and table E.2 versus table E.1). However, it is unknown if doing this split results in better joint modeling of compound floods. It was therefore decided not to pursue this further, and assume the dependence structure associated with Figure E.1 provides a sufficiently accurate representation of the joint distribution.

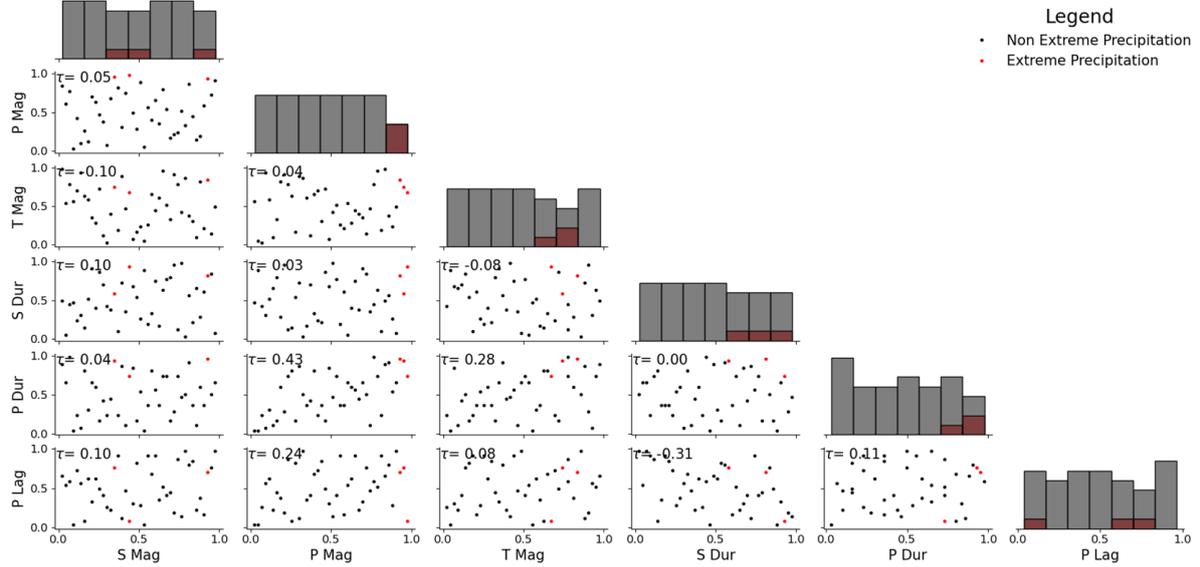


Figure E.2: Scatter matrix for the event set which has negative precipitation lag. Uses the Seaborn python package [Waskom, 2021], which shows the correlation between different dimensions in the uniformly ranked space. Includes S Mag, P Mag, T Mag, S Dur, P Dur, and P Lag. Along the main diagonal, a histogram for each dimension is plotted. It shows the probability density for non-extreme (black) and extreme precipitation (red) events. If both histograms are combined, a uniform density is achieved. Kendall's τ_B for each combination can be found in the upper left corner of the respective subplot. For precipitation lag, a rank of 0 and 1 indicates a lag of -3 days and 0 days respectively.

Table E.2: P-values associated with each Kendall's τ_B shown in Figure E.2. Shades of green show bivariate pairs of probabilistic variables that have significant correlation at the 5% level.

	S Mag	P Mag	T Mag	S Dur	P Dur
P Mag	0.65				
T Mag	0.34	0.68			
S Dur	0.36	0.79	0.47		
P Dur	0.74	$8.1e-5$	0.0091	0.97	
P Lag	0.35	0.028	0.43	0.0043	0.31

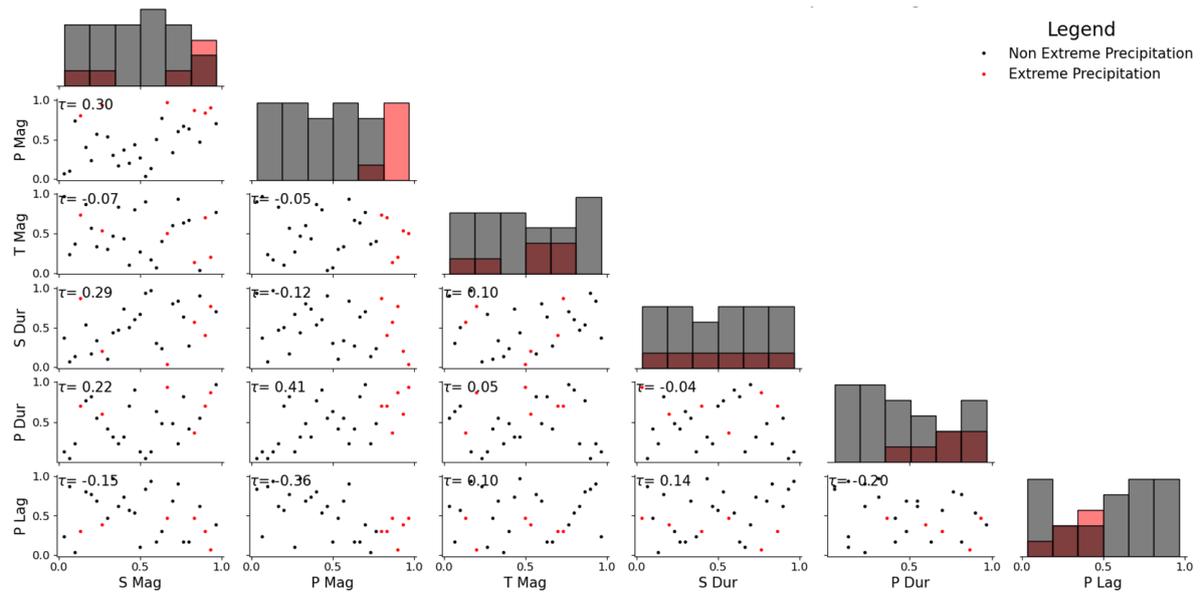


Figure E.3: Scatter matrix for the event set which has positive precipitation lag. Uses the Seaborn python package [Waskom, 2021], which shows the correlation between different dimensions in the uniformly ranked space. Includes S Mag, P Mag, T Mag, S Dur, P Dur, and P Lag. Along the main diagonal, a histogram for each dimension is plotted. It shows the probability density for non-extreme (black) and extreme precipitation (red) events. If both histograms are combined, a uniform density is achieved. Kendall's τ_B for each combination can be found in the upper left corner of the respective subplot. For precipitation lag, a rank of 0 and 1 indicates a lag of 0 days and +3 days respectively.

Table E.3: P-values associated with each Kendall's τ_B shown in Figure E.3. Shades of green show bivariate pairs of probabilistic variables that have significant correlation at the 5% level.

	S Mag	P Mag	T Mag	S Dur	P Dur
P Mag	0.025				
T Mag	0.59	0.72			
S Dur	0.03	0.36	0.47		
P Dur	0.09	0.0019	0.72	0.78	
P Lag	0.25	0.0068	0.47	0.29	0.14

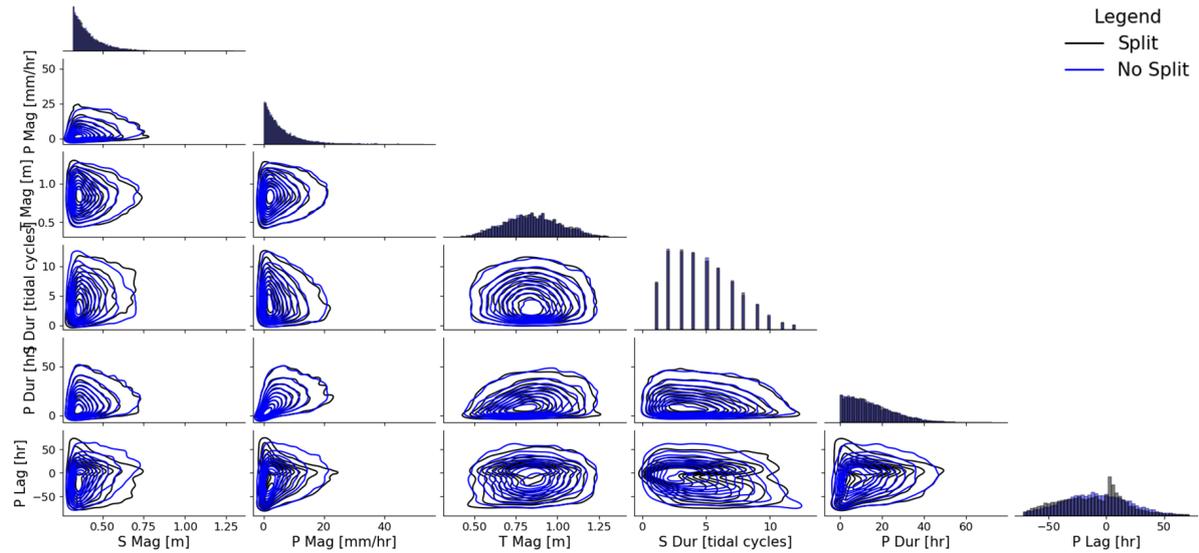


Figure E.4: Kernel density estimation (KDE) after generating 10,000 events for historical event sets with a split and no split. Uses the Seaborn python package [Waskom, 2021]. Includes *S Mag*, *P Mag*, *T Mag*, *S Dur*, *P Dur*, and *P Lag*. Along the main diagonal, the probability distributions of the marginals is shown. The only significant difference is for precipitation lag, where the marginal needs to be split to better characterize the event set. The split KDE shows a larger concentration of extreme magnitudes at a lag of zero hours (mostly influenced by the marginal, because surge magnitude has no correlation with precipitation lag, but exhibits a similar behaviour. The independence does cause the peak in KDE to be less pronounced).

F

Regular Vine And Pair Copulas

This appendix presents the outcome of the regular vine and pair copulas used when fitting the vine copulas with the pyvinecopulib python package [Nagler and Vatter, 2023] in 2, 3, 4, 5, and 6 dimensions. Table 4.2 shows which random variables are included for a given dimensionality.

Table E.1 shows the Bayesian Information Criterion (BIC) score associated with the different copulas when fitting them to the two dimensional historical event set, which contains Skew surge Magnitude (*S Mag*) and Precipitation Magnitude (*P Mag*) as random variables. In two dimensions, the Gaussian copula has the smallest BIC score. In section E.1, it was identified that the historical event set had no tail dependence, which matches the characteristics of a Gaussian copula.

Table E.1: BIC score associated with available copulas in the pyvinecopulib python package [Nagler and Vatter, 2023] when fitting the copulas to the historical observations of *S Mag* and *P Mag*. The smallest BIC score is highlighted in **bold**.

Copula Family	BIC Score
Independent	0
Gaussian	-3.37
Student	0.93
Clayton	-0.94
Gumbel	-3.28
Frank	-1.05
Joe	-2.67
BB1	0.74
BB6	0.99
BB7	0.32
BB8	1.34

In three dimensions (see table 4.2), the pair copula associated with *S Mag* and *P Mag* is the same as for two dimensions. Moreover, the pair copulas associated with Tidal Magnitude (*T Mag*) are found to be independent. This follows the outcomes of section E.1. To simplify the regular vine (chapter 4), *T Mag* was removed. To obtain stochastic events, quantiles are randomly sampled from a uniform distribution between 0 and 1 and are randomly associated with events generated by the copula in two dimensions.

Table E.2 shows the pair copulas associated with the different edges of each tree in the regular vine when fitting a vine copula to 4 random variables. The regular vine can be visualized in fig. F.1. Nodes are defined by each random variable of the historical event set for 4 dimensions (see table 4.2). Tree 2 and 3 have edges which are defined by conditional pair copulas. Tree 2 models the dependency between 2 different pair of variables given a third variable. Tree 3 models the dependency between a pair of variables given a third and fourth variable. A similar description can be given for table E.3, which shows the vine copula for 6 dimensions (see table 4.2). The regular vine shown in fig. E.2 shows there are 2 additional nodes, 2 additional pair copulas per tree, and 2 additional trees. Moreover, the node of *S Mag* in fig. E.2 has a degree of 3.

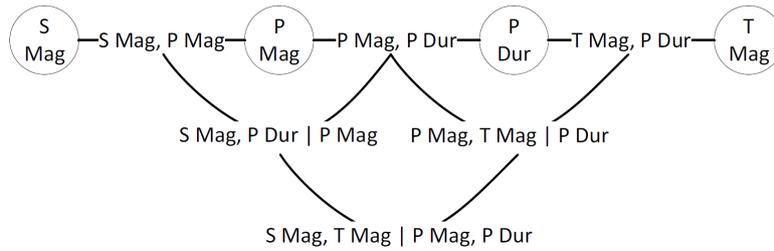


Figure E1: Regular vine fitted to 4 random variables.

In 4 dimensions (table E2), **T Mag** needs to be included in the vine copula. This follows the outcomes of section E.1 which showed that Precipitation Duration (**P Dur**) and **T Mag** had a significant correlation. A non-independent pair copula is also fitted to **P Mag** and **P Dur**. A rotated 180° Joe copula follows the lower tail dependence identified in section E.1.

Table E2: Representation of the vine copula in 4 dimensions [Nagler and Vatter, 2023]. Includes **S Mag**, **P Mag**, **T Mag**, and **P Dur**. Each edge represents a pair copula.

Tree	Edge	Copula Model
1	S Mag, P Mag P Mag, P Dur T Mag, P Dur	Gaussian Joe 180° Frank
2	S Mag, P Dur P Mag P Mag, T Mag P Dur	Independence Independence
3	S Mag, T Mag P Mag, P Dur	Independence

In five dimensions, the pair copulas associated with Skew surge Duration (**S Dur**) are all independent, which was one of the outcomes of section E.1. Again, to simplify the regular vine, **S Dur** is removed. The vine copula in four dimensions is used to generate events for **S Mag**, **P Mag**, **T Mag**, and **P Dur**. To generate co-occurring values for **S Dur**, quantiles are randomly sampled from a uniform distribution between 0 and 1.

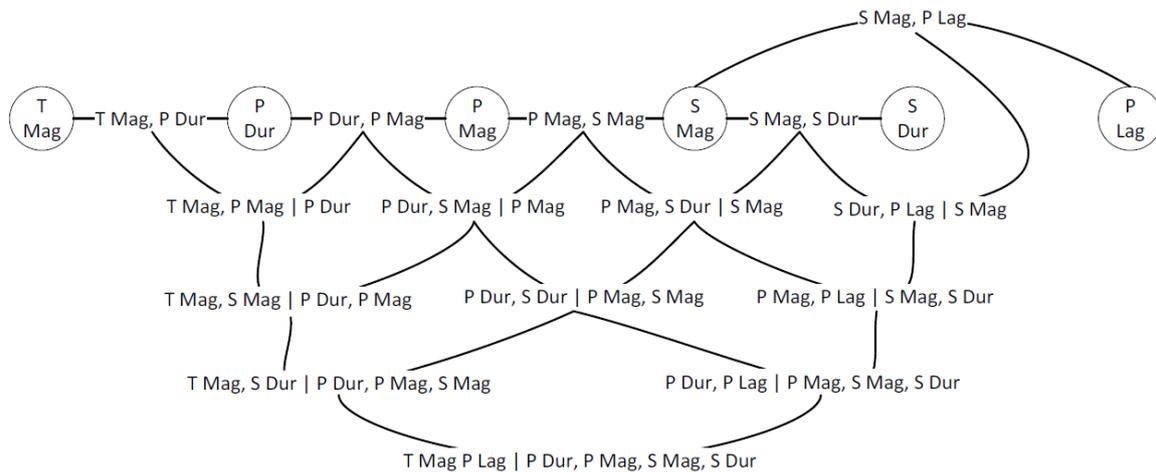


Figure E2: Regular vine fitted to 6 random variables.

In 6 dimensions, all variables need to be included (fig. E2 and table E3). While surge duration was found to be independent in section E.1, a non-independent copula is found to have a smaller BIC score than an independent copula on the second tree for the pair copula representing **S Dur, P Lag | S Mag** (table E3). This is also the case for the third tree where **P Mag, P Lag** are non-independent (table E3) even though this was the

case in section E.1. This is caused by the pair copula being conditioned on other variables, which changes the significance of their correlation.

Table F3: Representation of the vine copula in six dimensions [Nagler and Vatter, 2023]. Includes S Mag, P Mag, T Mag, P Dur, S Dur, and P Lag. Each edge represents a pair copula.

Tree	Edge	Copula Model
1	T Mag, P Dur P Dur, P Mag P Mag, S Mag S Dur, S Mag S Mag, P Lag	Frank Joe 180 ^o Gaussian Independence Frank
2	T Mag, P Mag P Dur P Dur, S Mag P Mag P Mag, S Dur S Mag S Dur, P Lag S Mag	Independence Independence Independence Joe 270 ^o
3	T Mag, S Mag P Dur, P Mag P Dur, S Dur P Mag, S Mag P Mag, P Lag S Mag, S Dur	Independence Independence Joe 180 ^o
4	T Mag, S Dur P Dur, P Mag, S Mag P Dur, P Lag P Mag, S Mag, S Dur	Independence Independence
5	T Mag, P Lag P Dur, P Mag, S Mag, S Dur	Independence

G

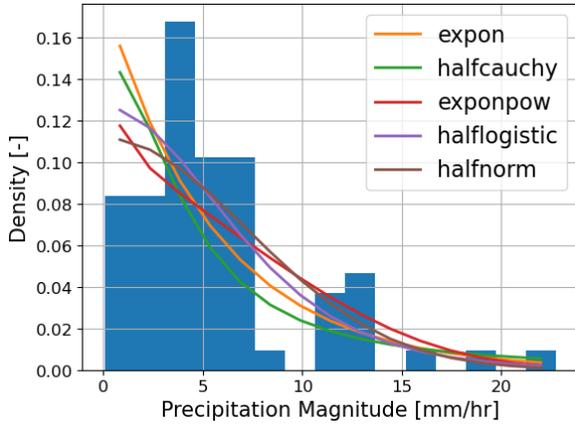
Defining Marginal Probability Distribution Functions (PDFs) for Non-Dominant Magnitudes

This appendix presents the marginal Probability Distribution Functions (PDFs) used to define the random variables that are not the dominant driver's magnitude. A short methodology is provided for the different types of random variables: magnitudes, durations, lag, and the tide.

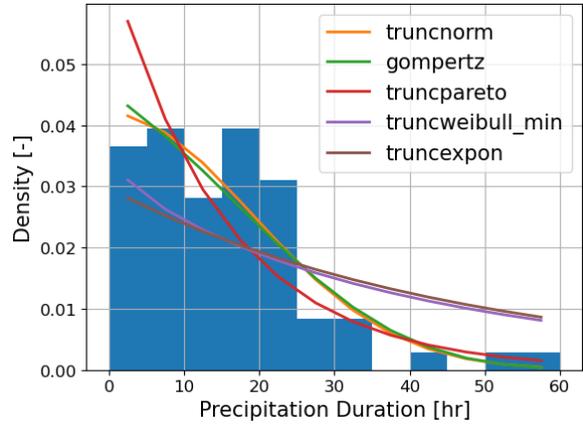
For non-tidal driver magnitudes, good tail behavior is desired to extrapolate plausible values. For this reason, the Bayesian Information Criterion (BIC) score is used to choose the best marginal as it favors simple models with steady tail behavior. Distributions that do not match the expected behavior are filtered out (i.e. uniform distribution). The result for Precipitation Magnitude (P Mag) can be found in fig. G.1a.

For duration and time lag, extrapolation is not a necessary feature. Therefore, only truncated distributions are used to fit the empirical distributions. In this case, the BIC may be biased towards overly simple models, favoring distributions that can underfit and misrepresent the data. Therefore, the Sum Square Error (SSE) is used as a metric to choose the best distribution. The results for Precipitation Duration (P Dur), Skew surge Duration (S Dur) and Precipitation Lag (P Lag) can be found in fig. G.1b, fig. G.1c, and fig. G.1d respectively.

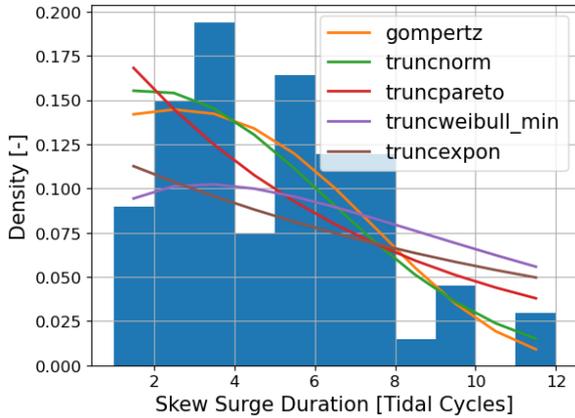
The Tidal Magnitude (T Mag) is bounded by the beating of its constituents. For sufficiently long time records, the majority of beating cycles should have occurred. Therefore, the fitter package is not required, as the Empirical Probability Distribution Function (EPDF) of all Higher High (HH) tides can be used instead. The corresponding histogram can be found in fig. G.2.



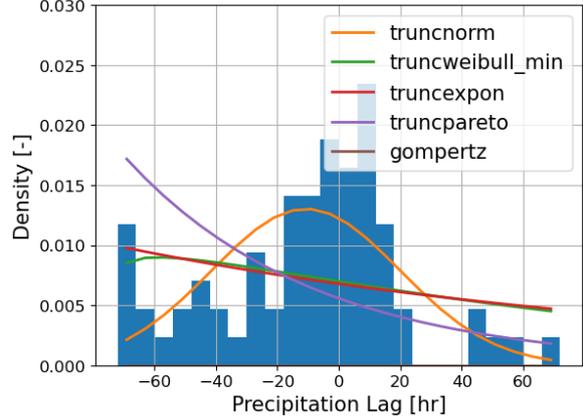
(a) The 5 PDFs which result in the smallest BIC score for the P Mag EPDF. The exponential ('expon') is the distribution that minimizes the BIC score. Visually, the fit performs relatively well. The peak in density is displaced to smaller magnitudes because of the parameters available for the exponential distribution. When generating events, it will overestimate the number of precipitation events which have a magnitude of zero.



(b) The 5 PDFs which result in the smallest SSE for the P Dur EPDF. BIC or AIC scores are not used as they prioritize the truncated exponential ('truncexpon'). The truncated normal is the best fit and also performs well visually. When generating events, it will overestimate the number of events with large P Dur.



(c) The 5 PDFs which result in the smallest SSE for the S Dur EPDF. The truncated gumbel ('gompertz') is the best fit and also performs well visually for a variable that is defined at coarse intervals.



(d) The 5 PDFs which result in the smallest SSE for the P Lag EPDF. The truncated normal ('truncnorm') is the best fit and also performs the best visually as it is the only one that captures the peak in density near 0 hours of lag. However, it is not flexible enough to capture the true location of the peak and underestimates the lag corresponding to peak density by approximately (10 - 15) hours.

Figure G.1: Empirical and theoretical marginal PDFs. Theoretical marginals are chosen based on a metric (BIC or the SSE) using the Fitter python package [Cokelaer et al., 2024].

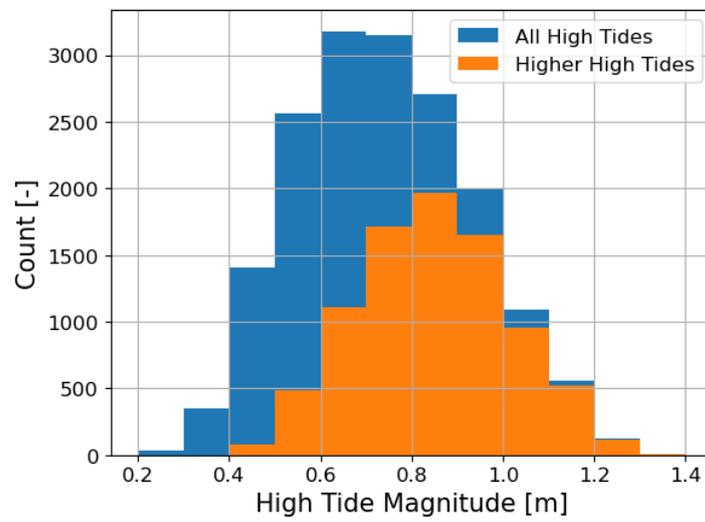


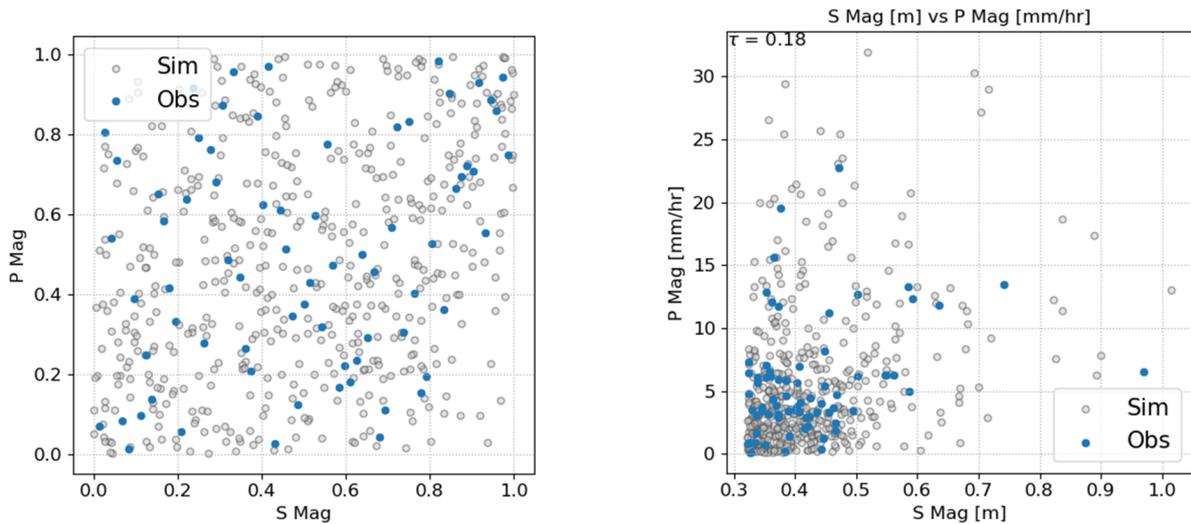
Figure G.2: EPDF for T Mag. Only HH tides are used because of how skew surge events were identified. Used instead of a theoretical marginal PDF as S Mag can be assumed independent of T Mag (Table E.1 and [Williams et al., 2016]). Moreover, because of the tidal beating, the 25 years of data should give a strong confidence in observing all possible T Mags.

H

Stochastic Event Sets

This appendix presents a graphical representation of the training and test stochastic event sets used in this research.

H.1. Training Sets



(a) Events are shown in their univariate ranked space. Ranks of the observed events are dictated by their empirical CDF.

(b) Events shown in the real space. The top left corner contains Kendall's τ_B .

Figure H.1: Training set in two dimensions (500 samples) for **P Mag** and **S Mag**. Shows the observed events (historical event set) and events simulated by the copula (Gaussian), which were obtained by using the `pyvinecopulib` python package [Nagler and Vatter, 2023]. The simulated quantiles (Figure H.1a) can be combined with the distributions defined in Appendix G and Figure 4.6b to obtain Figure H.1b.

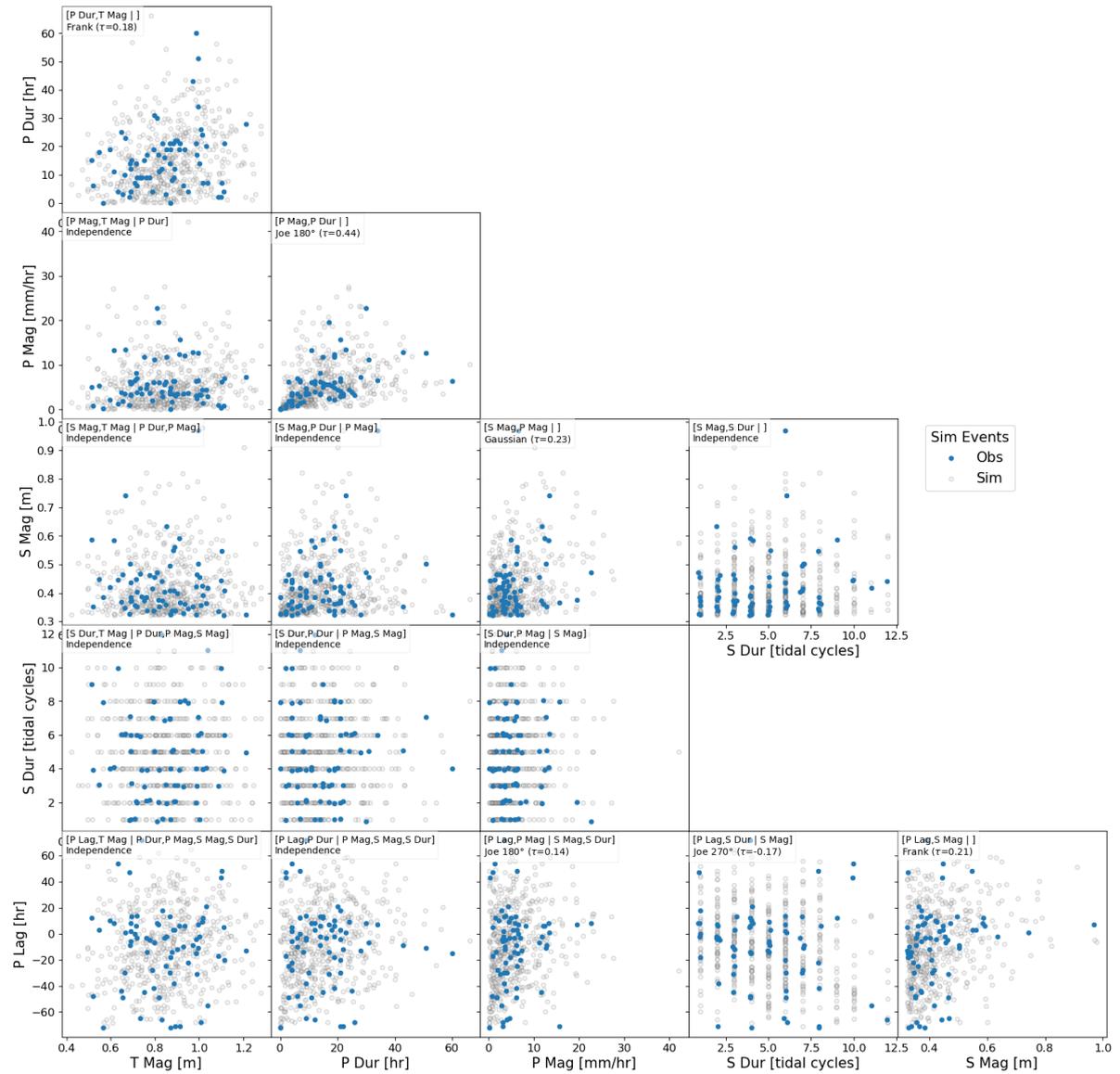


Figure H.2: Training set in six dimensions (500 samples) for P Mag, S Mag, T Mag, P Dur, S Dur, and P Lag. Shows the observed events (historical event set) and events simulated by the vine copula, which were obtained by using the pyvinecopulib python package [Nagler and Vatter, 2023]. The upper left corner of each subplot shows the edge of the vine copula, its corresponding pair copula, and empirical Kendall's τ . Code for the plot was obtained from Eilander et al. [2023c].

H.2. Test Sets

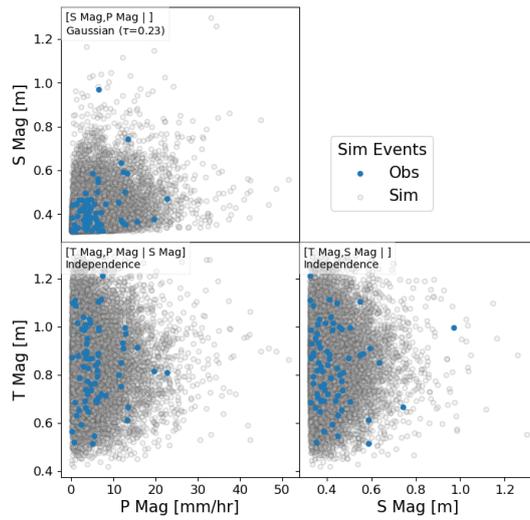


Figure H.3: Test Dataset in three dimensions (10,000 samples) for *P Mag*, *S Mag*, and *T Mag*. Shows the observed events (historical event set) and events simulated by the vine copula, which were obtained by using the `pyvinecopulib` python package [Nagler and Vatter, 2023]. The upper left corner of each subplot shows the edge of the vine copula, its corresponding pair copula, and empirical Kendall's τ . Code for the plot was obtained from Eilander et al. [2023c].

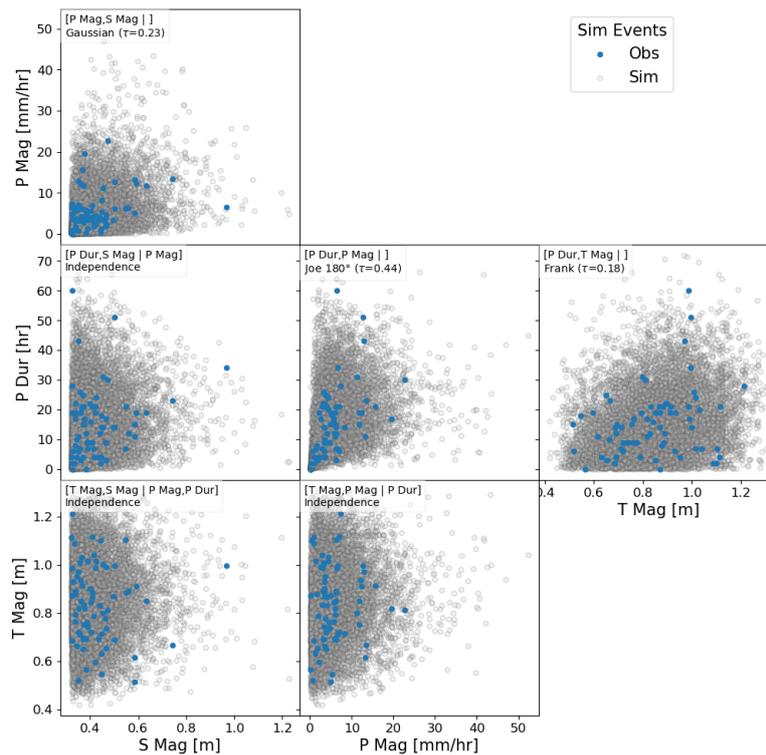


Figure H.4: Test Dataset in four dimensions (10,000 samples) for *P Mag*, *S Mag*, *T Mag*, and *P Dur*. Shows the observed events (historical event set) and events simulated by the vine copula, which were obtained by using the `pyvinecopulib` python package [Nagler and Vatter, 2023]. The upper left corner of each subplot shows the edge of the vine copula, its corresponding pair copula, and empirical Kendall's τ . Code for the plot was obtained from Eilander et al. [2023c].

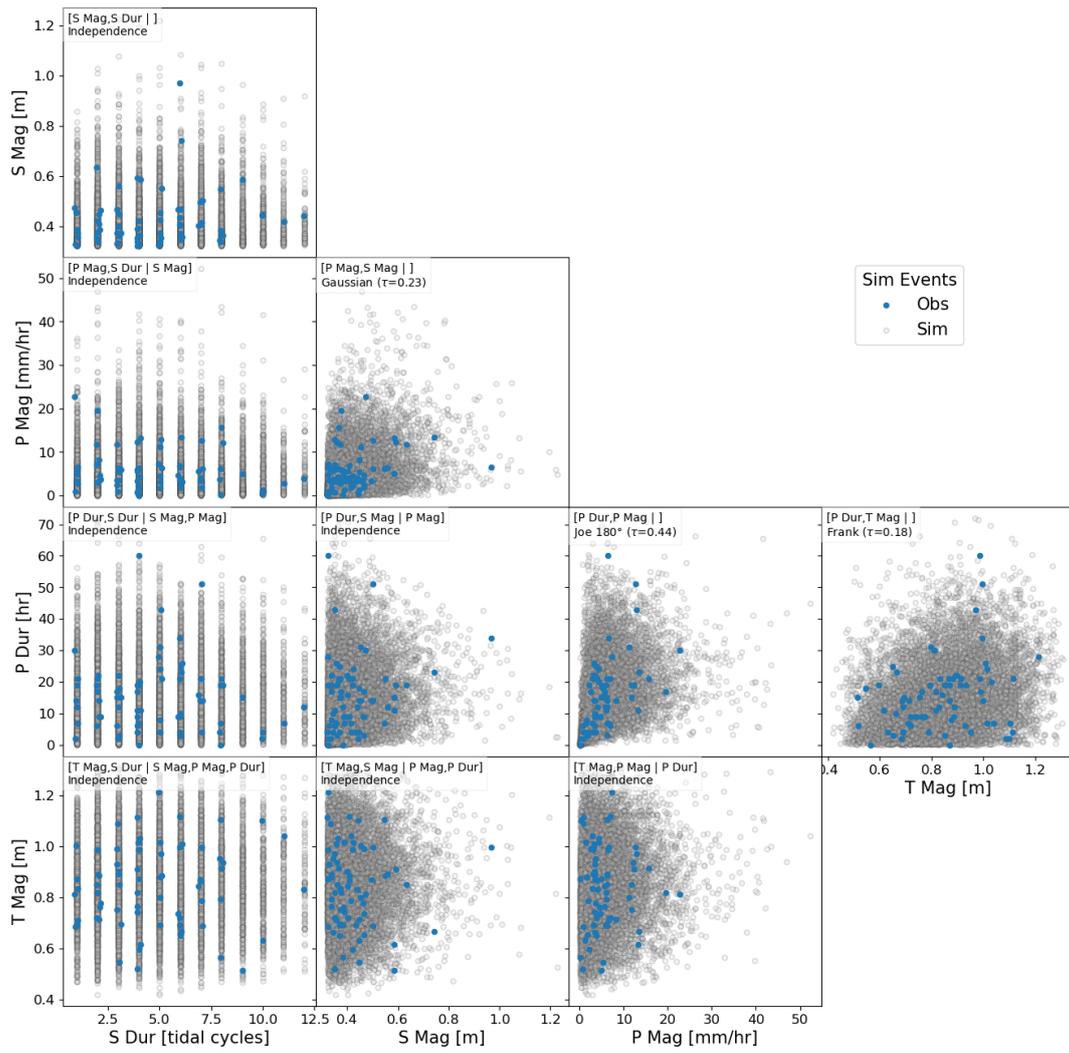


Figure H.5: Test Dataset in five dimensions (10,000 samples) for P Mag, S Mag, T Mag, P Dur, and S Dur. Shows the observed events (historical event set) and events simulated by the vine copula, which were obtained by using the pyvinecopulib python package [Nagler and Vatter, 2023]. The upper left corner of each subplot shows the edge of the vine copula, its corresponding pair copula, and empirical Kendall's τ . Code for the plot was obtained from Eilander et al. [2023c].

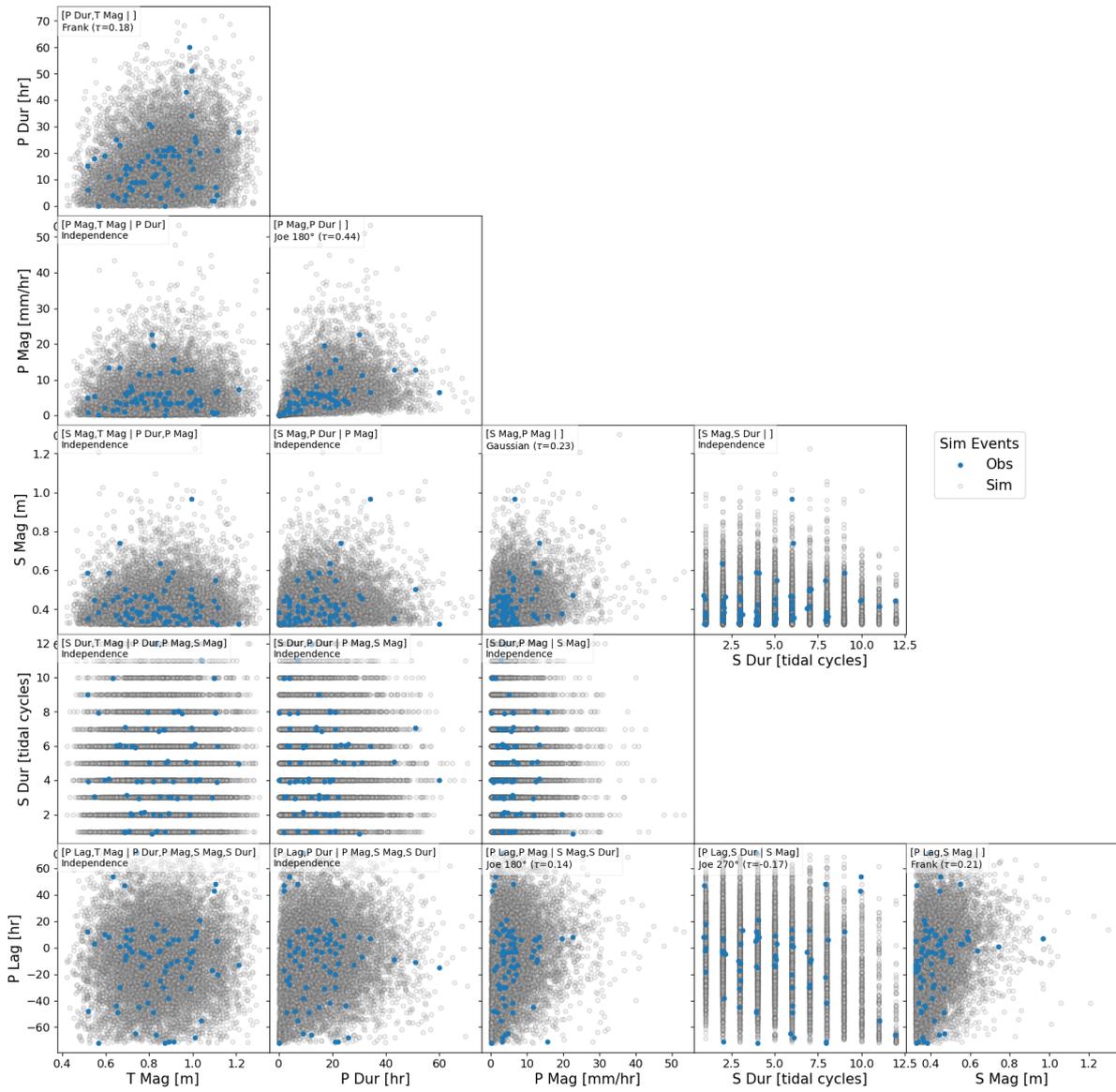


Figure H.6: Test Dataset in six dimensions (10,000 samples) for P Mag, S Mag, T Mag, P Dur, S Dur, and P Lag. Shows the observed events (historical event set) and events simulated by the vine copula, which were obtained by using the pyvinecopulib python package [Nagler and Vatter, 2023]. The upper left corner of each subplot shows the edge of the vine copula, its corresponding pair copula, and empirical Kendall's τ . Code for the plot was obtained from Eilander et al. [2023c].



Defining a Stopping Criterion for Compound Floods

This appendix presents the results, analysis, and the decision taken to define the stopping criterion used for the a posteriori approach. Both fig. I.1 and fig. I.2 are used. Since they both have the same format, a description is only given for fig. I.1, but it is valid for fig. I.2.

Figure I.1 presents the outcomes of fitting the Treed Gaussian Process (TGP) iteratively to the training event set in 2 dimensions (fig. H.1) using the methodology presented in section 4.5.2. Figure I.1a shows the Root Mean Square Error (RMSE) associated with the sampled events and the available events as more events are used to fit the TGP. This compares the estimate of economic damages estimated by the TGP with the economic damages simulated with the coupled SFINCS and FIAT models for all events in the stochastic event set. Figure I.1c shows the mean and maximum Active Learning Mackay (ALM) statistic of all available events in the stochastic event set as the number of samples used to fit the TGP increases. Figure I.1b shows the estimate of the TGP for the EAD mean with the associated uncertainty, and the ground truth as the number of samples used to fit the TGP increases (see section 4.6 to understand how risk is modeled). Figure I.1d shows the COV associated with the EAD mean as the number of samples increases. Figure I.1e shows the p-value associated with a two-sample KS test (section 4.6.2) which compares the empirical CDF of the ground truth with the empirical CDF estimated by the TGP as the number of samples used to fit the TGP increases (see section 4.6.2). Finally, all subplots include a vertical black line which indicates the number of total samples used to reach the stopping criterion. Figure I.1f shows the empirical CDFs associated with the ground truth and the TGP when the TGP uses the number of samples needed to reach the stopping criterion.

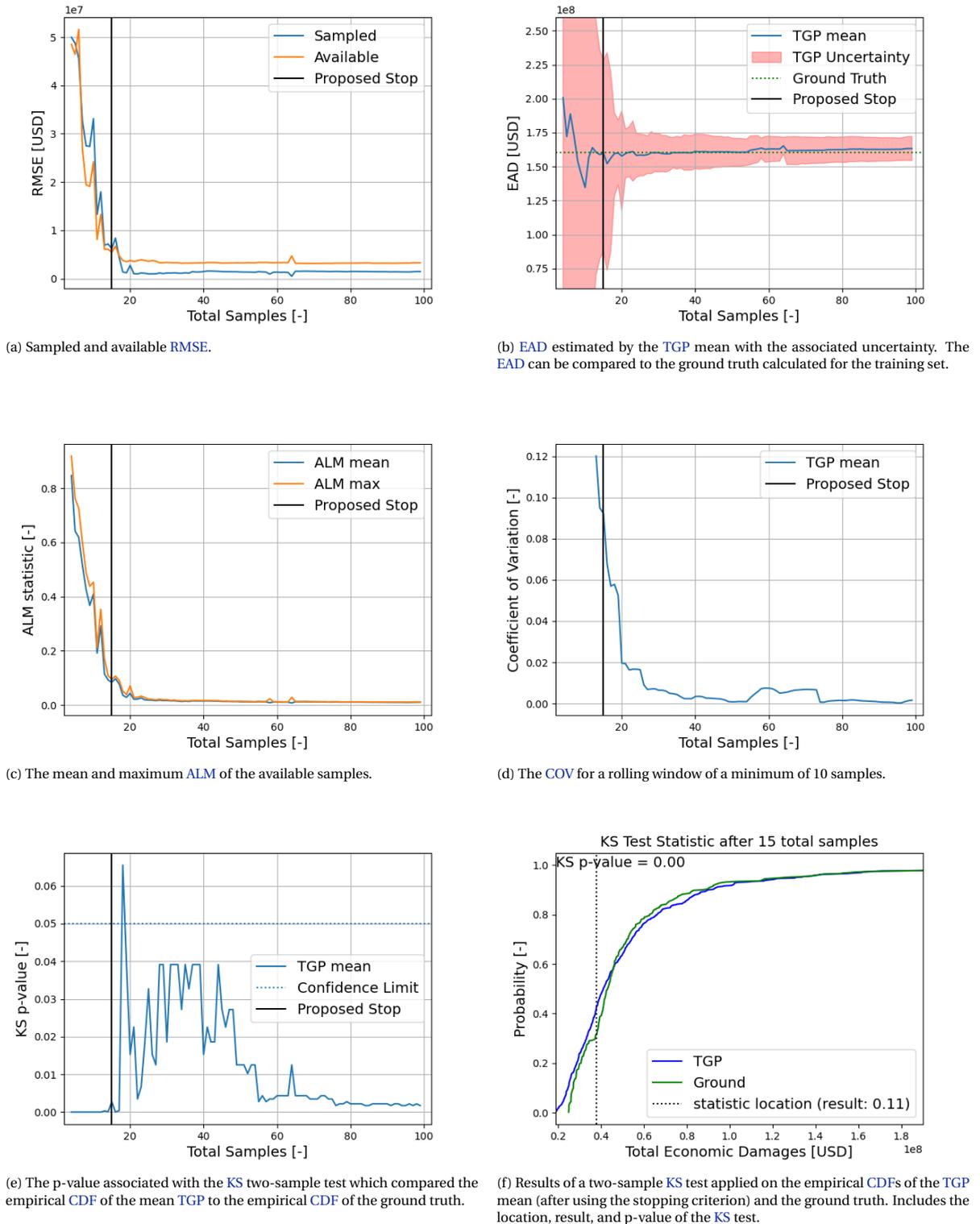


Figure I.1: Outcome of performing the analysis on a training dataset of 500 samples in two dimensions (fig. H.1). The TGP is first fitted to the sampled points after the economic damages of 4 samples are computed.

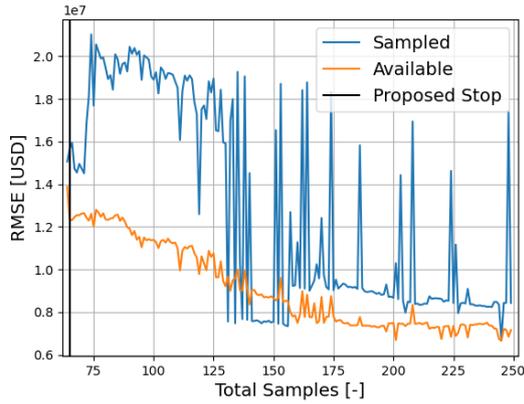
The first observation can be made with the available RMSE which is almost monotonically decreasing (the signal is rough, which results in relatively small increases in the loss) for 2 dimensions (fig. I.1a). Based on this limited amount of evidence, it can be assumed that the more samples the TGP uses, the better the fit

becomes, and the smaller the available loss becomes. However, to reduce the computational cost related to the posterior predictive surface (which is proportional to the number of samples cubed) and the damage modeling, a stopping criterion should be defined at which the loss is sufficiently small. Two different metrics could be used to define this stopping criterion: the **ALM** (mean or maximum) and the **COV** of the **EAD**.

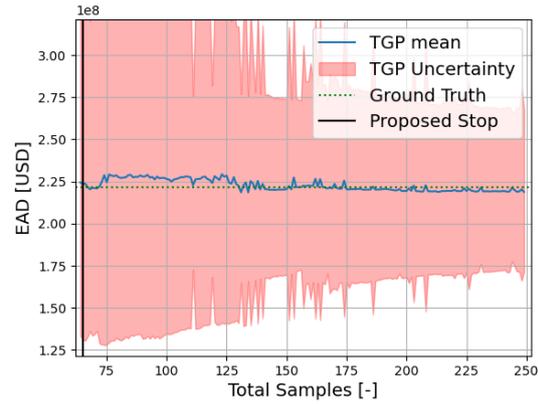
From fig. I.1c, the **ALM** mean seems to follow trends in the sampled loss shown in fig. I.1a. This is logical, as the **ALM** is equivalent to the quantile difference of the **TGP** which is a measure of its uncertainty. The **TGP** is uncertain when the mean is unable to fit the sampled data points exactly, which is caused by the Markov Chain Monte Carlo (**MCMC**) being unable to find a suitable set of hyperparameters. This results in a loss occurring for the sampled points. Another observation with the **ALM** mean can be made with the **TGP** uncertainty of the **EAD** where small magnitudes in the **ALM** mean correspond with a more accurate estimation of the **EAD** and a smaller corresponding uncertainty range (fig. I.1b). On the other hand, the **ALM** maximum seems to have non-harmonic peaks in its magnitude. These are most probably caused by a partition, which results in a small subset of samples having a large uncertainty. Therefore, these peaks do not correspond with a period of larger training loss or validation loss. If the **ALM** maximum was used to define the stopping criterion, it would most probably prevent the final model from having partitions, resulting in a model that does not provide the most accurate representation of the entire event set [Gramacy and Lee, 2009].

Figure I.1d shows the **COV** relating to the **EAD** shown in fig. I.1b. The main setback of using the **COV** as a stopping criterion is that 10 consecutive samples from the **TGP** need to be computed. Depending on the dimensionality, this will result in more samples having to be modeled to reach an understanding that stability has been reached. In six dimensions, small values of **COV** do not result in the best estimate of **EAD**. Moreover, for multiple outputs, using a **COV** would mean either computing the posterior predictive surface for all outputs after each sample is modeled, or would require $\text{num}_{\text{outputs}} \cdot \text{window}_{\text{size}}$ samples to be modeled before an understanding of stability has been reached. Both of these cases are computationally expensive. This results in the **COV** not being suitable to be defined as a stopping criterion.

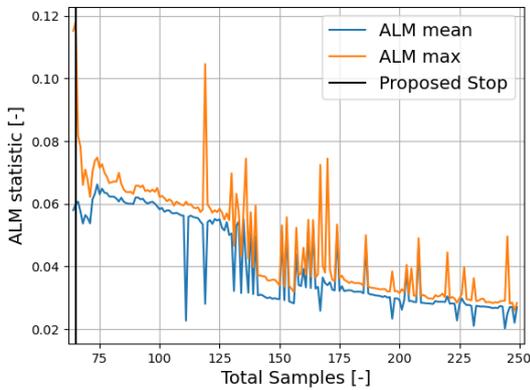
Similar observations can also be made in fig. I.2. Figure I.2a and fig. I.2c show there is a correlation between the sampled loss and the **ALM** mean. The **RMSE** and **ALM** statistic initially start at smaller magnitudes than fig. I.1. This is caused by the **MDA** initialization, which requires 64 samples in 6 dimensions. This translates to a more stable **EAD** and a smaller uncertainty band initially (fig. I.2b and fig. I.2d).



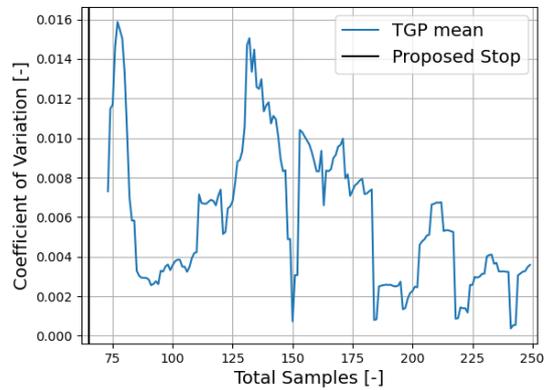
(a) Sampled and available RMSE loss.



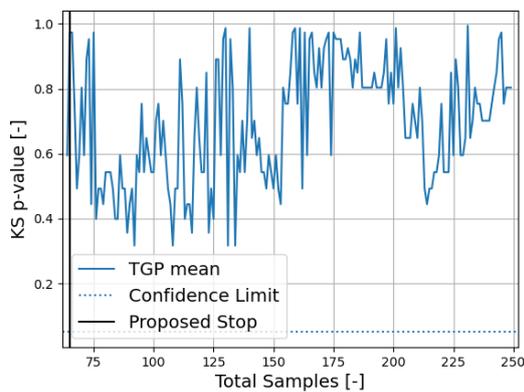
(b) EAD estimated by the TGP mean with the associated uncertainty. The EAD can be compared to the ground truth calculated for the training set.



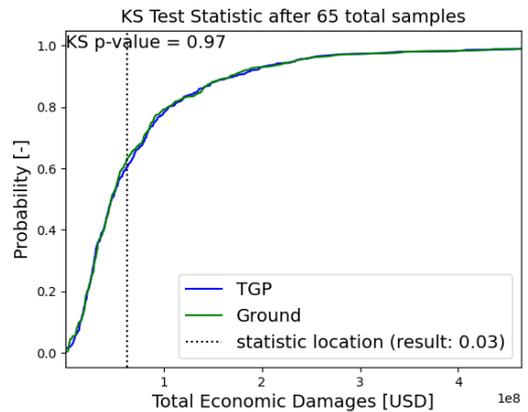
(c) The mean and maximum ALM of the available samples.



(d) The COV for a rolling window of a minimum of 10 samples.



(e) The p-value associated with the KS two-sample test which compared the empirical CDF of the mean TGP to the empirical CDF of the ground truth.



(f) Results of a two-sample KS test applied on the empirical CDFs of the TGP mean (after using the stopping criterion) and the ground truth. Includes the location, result, and p-value of the KS test.

Figure I.2: Outcome of performing the analysis on a training dataset of 500 samples in six dimensions (fig. H.2). The TGP is first fitted to the sampled points after the economic damages of 64 samples are computed.

It was therefore decided to use the ALM mean as a stopping criterion as it is the most suitable variable to use when compared to the ALM maximum and the COV. The choice of threshold dictates how accurate the posterior predictive surface becomes, and the number of samples that need to be modeled. Since the components

used in the damage modeling are deterministic for a set of boundary conditions, it can be assumed that the best surrogate model should have an uncertainty of zero (small **ALM** mean). However, the **TGP** will initially assume small differences in economic damages are caused by noise. Moreover, fig. I.1a shows that there is a plateau in the RMSE, showing a **TGP** is unable to fit stochastic event sets perfectly, no matter the number of samples it uses to calculate the posterior predictive surface. This means a small threshold could result in large or infinite computational costs for marginal gains in accuracy. Since a **TGP** is regularized, choosing a large threshold will result in a less accurate model, but not one that overfits the sampled points. In 6 dimensions, fig. I.2e shows that the empirical **CDF** of the Ground truth and the **TGP** mean results in a p-value that is larger than the 5% significance level for any number of samples (after using a **MDA**). This means, that the **TGP** gives an accurate estimate of the risk curve after only using samples chosen a priori. Therefore, the **ALM** mean corresponding to the initial fit of the **TGP** should result in a good stopping criterion. However, it is unknown if the magnitude of the **ALM** statistic is dependent on the size of the event set as the number of samples that need to be modeled could be proportional to the size of the event set.

In two dimensions, the size of the event set becomes less problematic, because there is less sparsity in the input space between the available samples. The irregularity of the empirical **CDF** of the ground truth (fig. I.1f) shows the p-value is never steadily above the 5% significance level (fig. I.1e). The irregularity of the ground truth could be caused by the magnitude of the total economic damages, which makes the modeled events susceptible to small changes in the boundary conditions. For example, sudden overflow, caused by small changes in the surge magnitude could result in a sudden increase in economic damages. The **TGP** can only correctly model this if it partitions (around sample 18), but none of the metrics correctly identify when a partition causes a significant p-value in the **KS** test. Nonetheless, the available **RMSE** (fig. I.1a) reaches a stable value after approximately 10 samples. This corresponds to an **ALM** mean of approximately 0.1. In addition, the **MCMC** causes different realizations of the **TGP** to have different **ALM** statistics for the same samples. This is because the **MCMC** converges if it is given enough samples, if not, it will give different estimates of the hyperparameters. To prevent early stopping, the stopping criterion is therefore defined as: 2 consecutive TGP models for an arbitrary output which result in an ALM mean smaller than 0.1.

J

Computational Cost For Components Of Damage Modeling

This appendix presents the computational cost associated with the components used in the damage modeling workflow (fig. 4.2). To this end, fig. J.1 and fig. J.2 present the cost associated with the **SFINCS** and **FIAT** models for Charleston when modeling all events in the training stochastic event set in 6 dimensions (fig. H.2).

Figure J.1 shows box plots for the computational time associated with the **SFINCS** and **FIAT** models. For each model, The orange line represents the mean, while the upper and lower limits of the box represent the 75th and 25th quantile respectively. The whiskers show the minimum/maximum of: the addition/subtraction of 1.5 times the interquartile range and the 75th/25th quantile. On one hand, the relatively large interquartile range for the **SFINCS** model shows that is dependent on the magnitude of the different random variables when they are used in the boundary condition schematization. On the other hand, the small distance between the whiskers of the box plot for the **FIAT** model shows that its computational time is independent of the magnitude of the random variables.

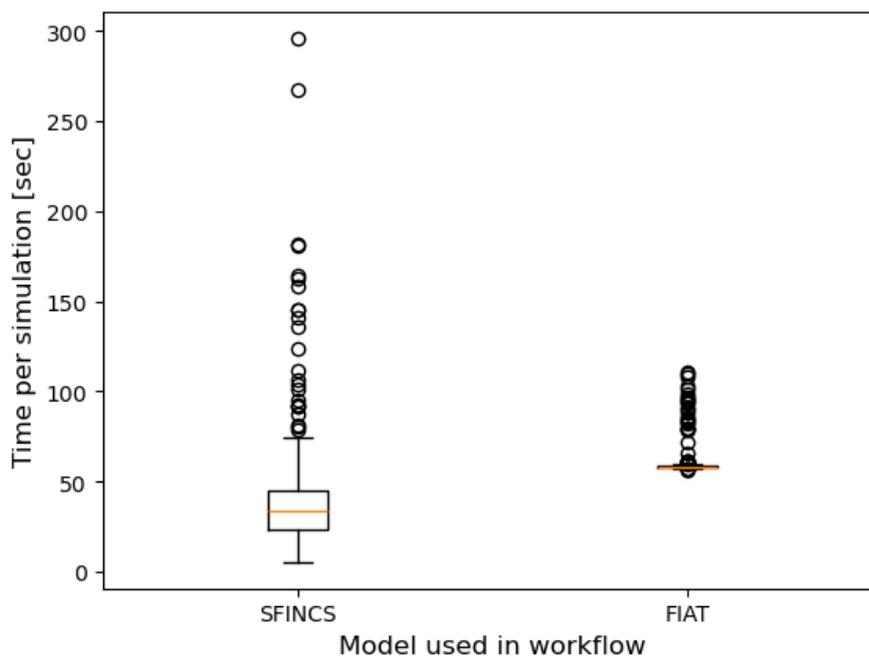
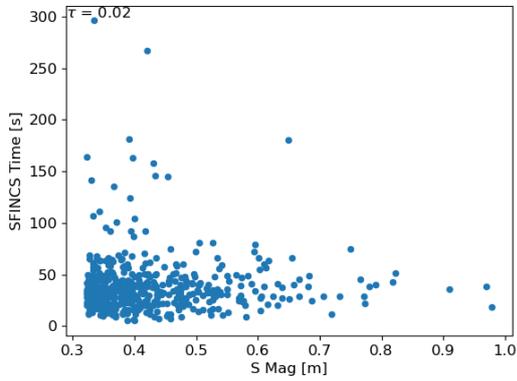
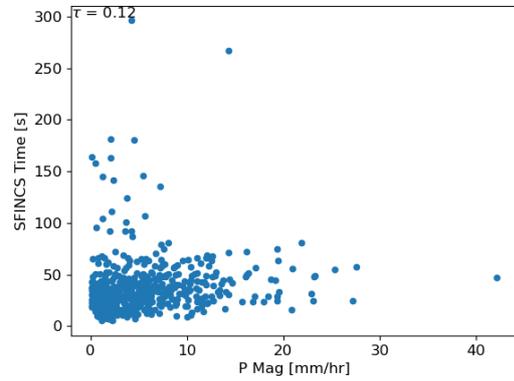


Figure J.1: Box plots of the computational cost of using **SFINCS** and **DELFT-FIAT** models for all the events in [Figure H.2](#).

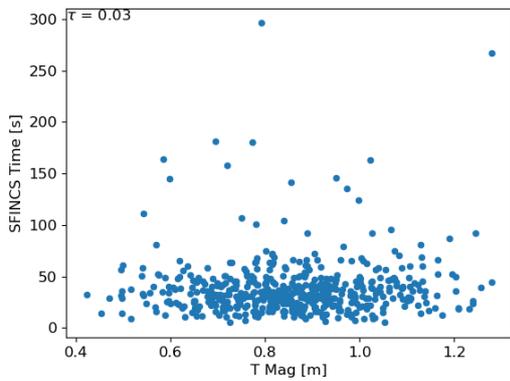
Figure J.2 shows the computational time for **SFINCS** when compared to the magnitude of the different random variables used in the boundary condition. Kendall's τ_B is used to identify any potential correlations. Simulation time seems to be independent of **S Mag** and **T Mag** (fig. J.2a and fig. J.2c respectively). Other variables show some correlation. The most significant is **S Dur** with a Kendall's τ_B of 0.65 (fig. J.2e).



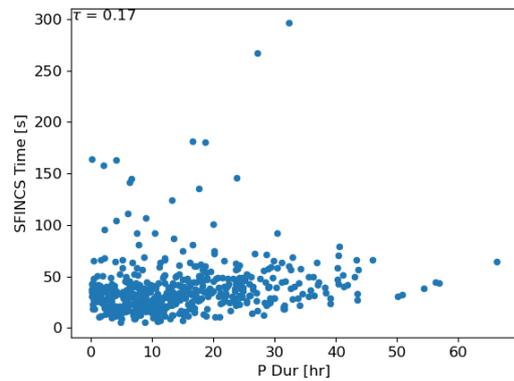
(a) Correlation with **S Mag**.



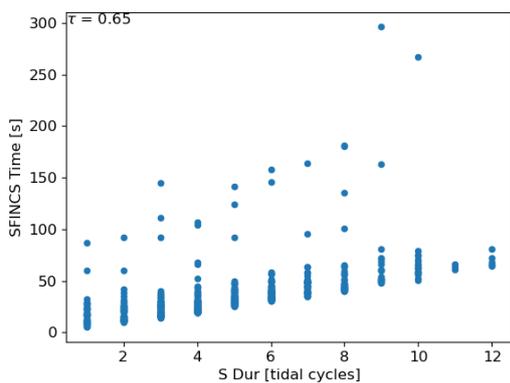
(b) Correlation with **P Mag**.



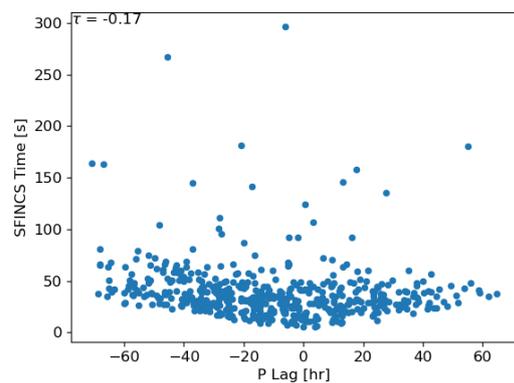
(c) Correlation with **T Mag**.



(d) Correlation with **P Dur**.

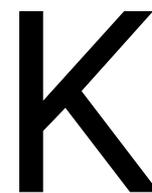


(e) Correlation with **S Dur**.



(f) Correlation with **P Lag**.

Figure J.2: Kendall's τ_B to test for correlation between the computational time of **SFINCS** and the different random variables.



Sensitivity of Additional Dimensions on Risk Curves and EAD for the Classified Model

This appendix presents the sensitivity of additional dimensions (or random variables) on the inland and coastal outputs of the classified [FIAT](#) model.

K.1. Inland

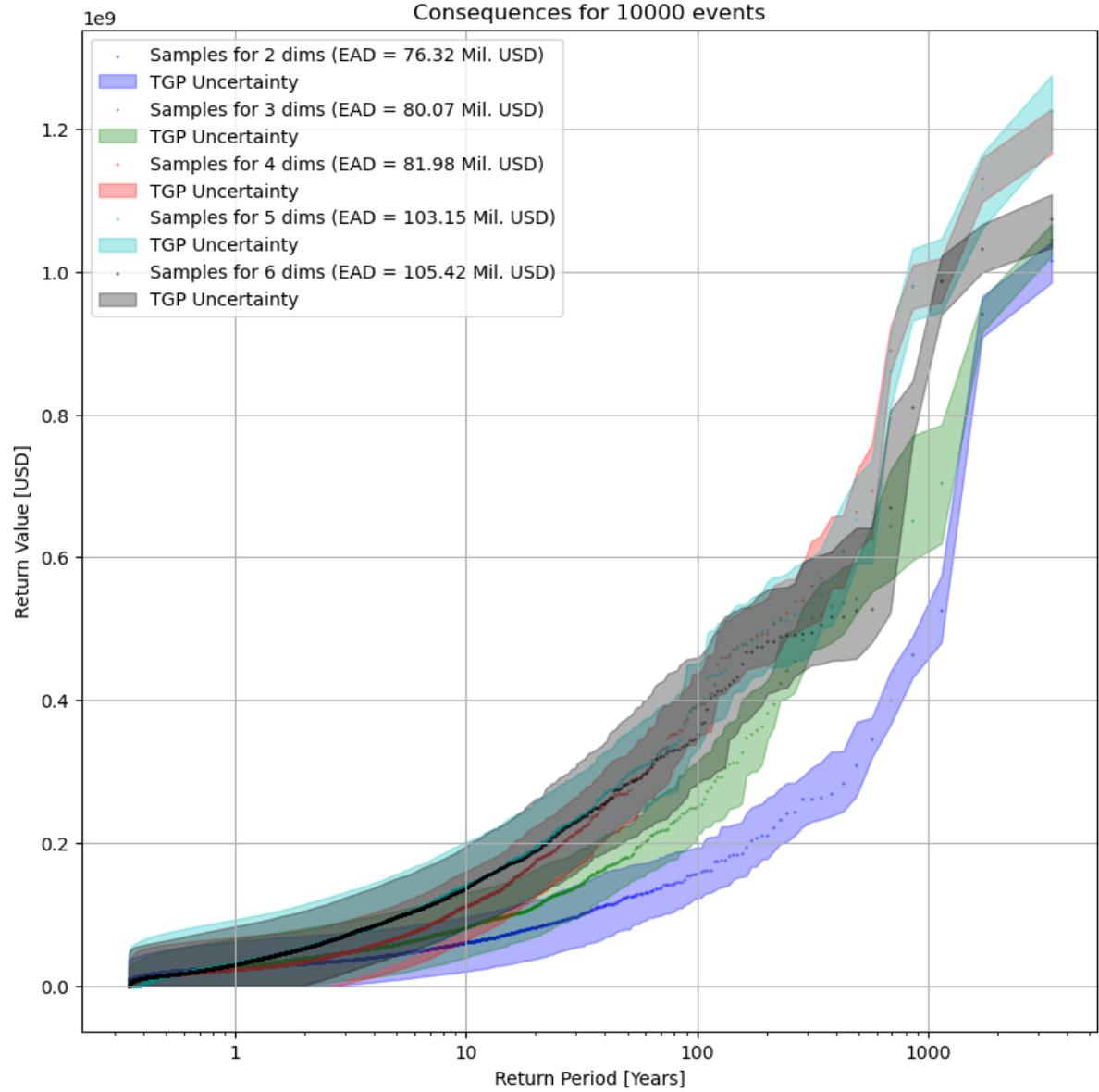


Figure K.1: Risk curves associated with different dimensionalities of the multivariate input space for inland economic damages. Legend includes the EAD estimate.

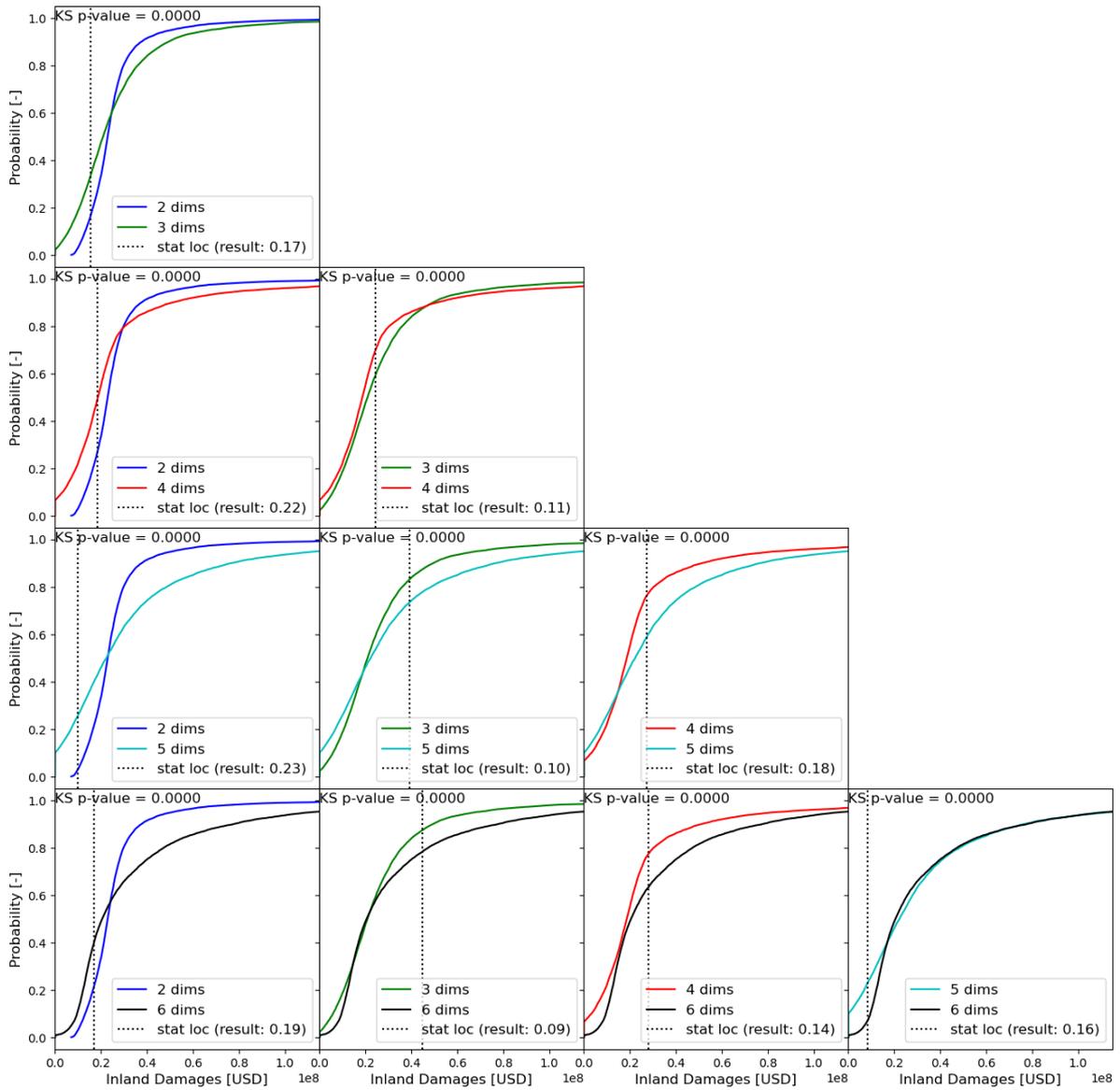


Figure K.2: Test statistics and associated p-values associated with two-sample KS test for empirical CDFs of different dimensionalities when sampling from the inland economic damages. If the p-value is smaller than 0.05, the null hypothesis (empirical CDFs come from the same parent distribution) is rejected.

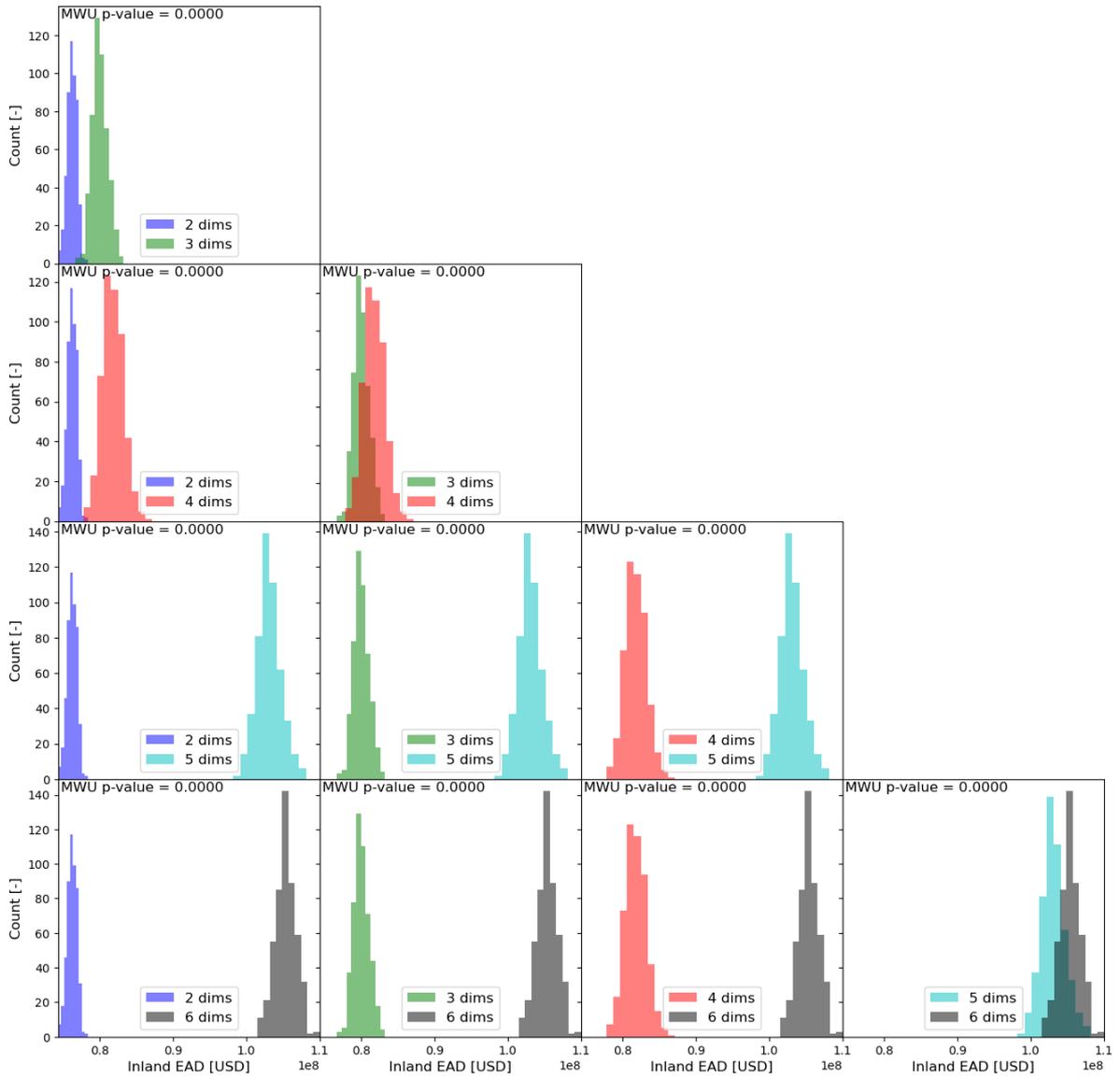


Figure K.3: P-values associated with MWU test for empirical CDFs of different dimensionalities when sampling from the inland economic damages. If the p-values are smaller than 0.05, the null hypothesis (empirical CDFs have the same location (EAD)) is rejected. The empirical bootstrap was repeated 500 times.

K.2. Coast

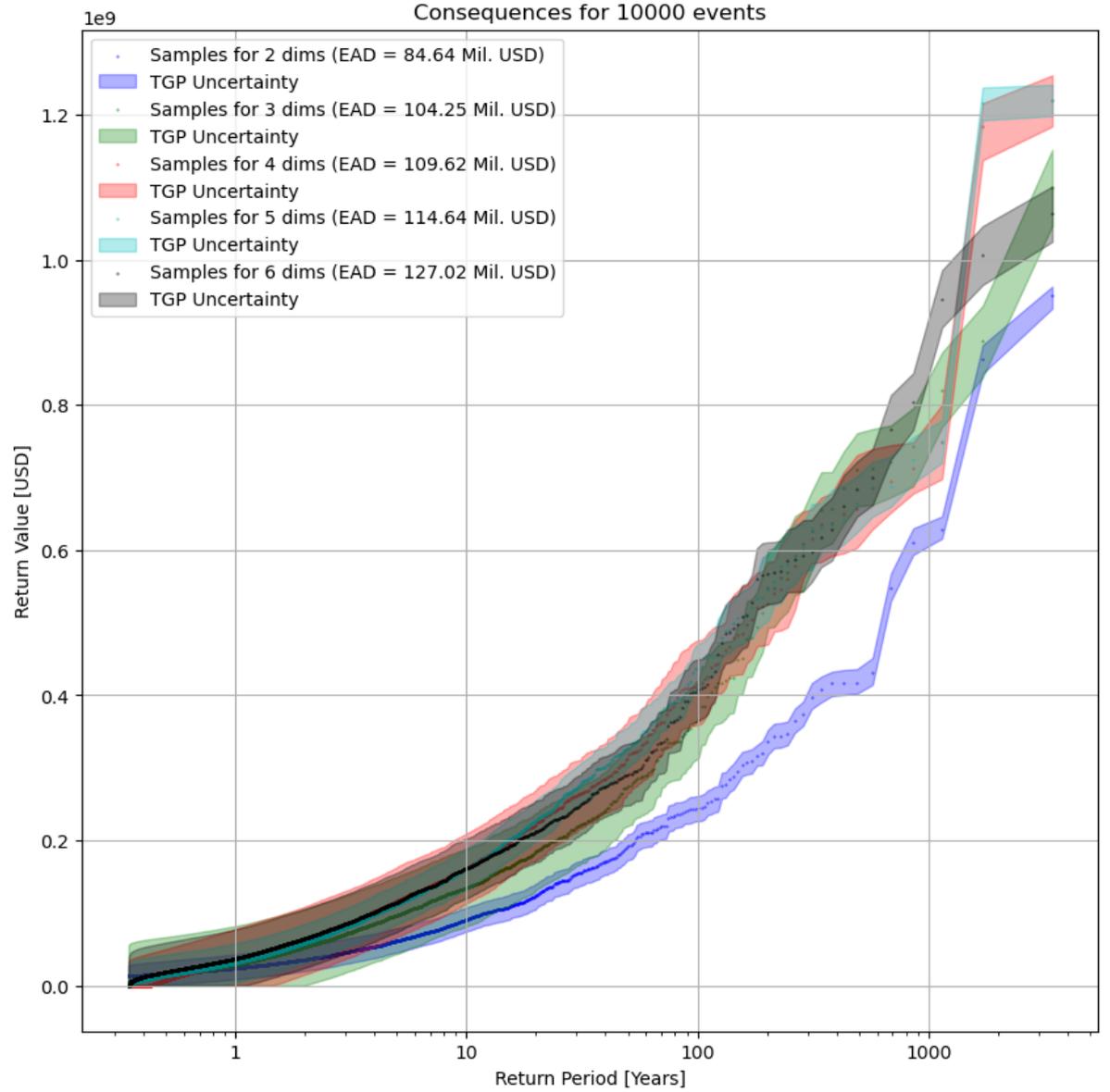


Figure K.4: Risk curves associated with different dimensionalities of the multivariate input space for coast economic damages. Legend includes the EAD estimate.

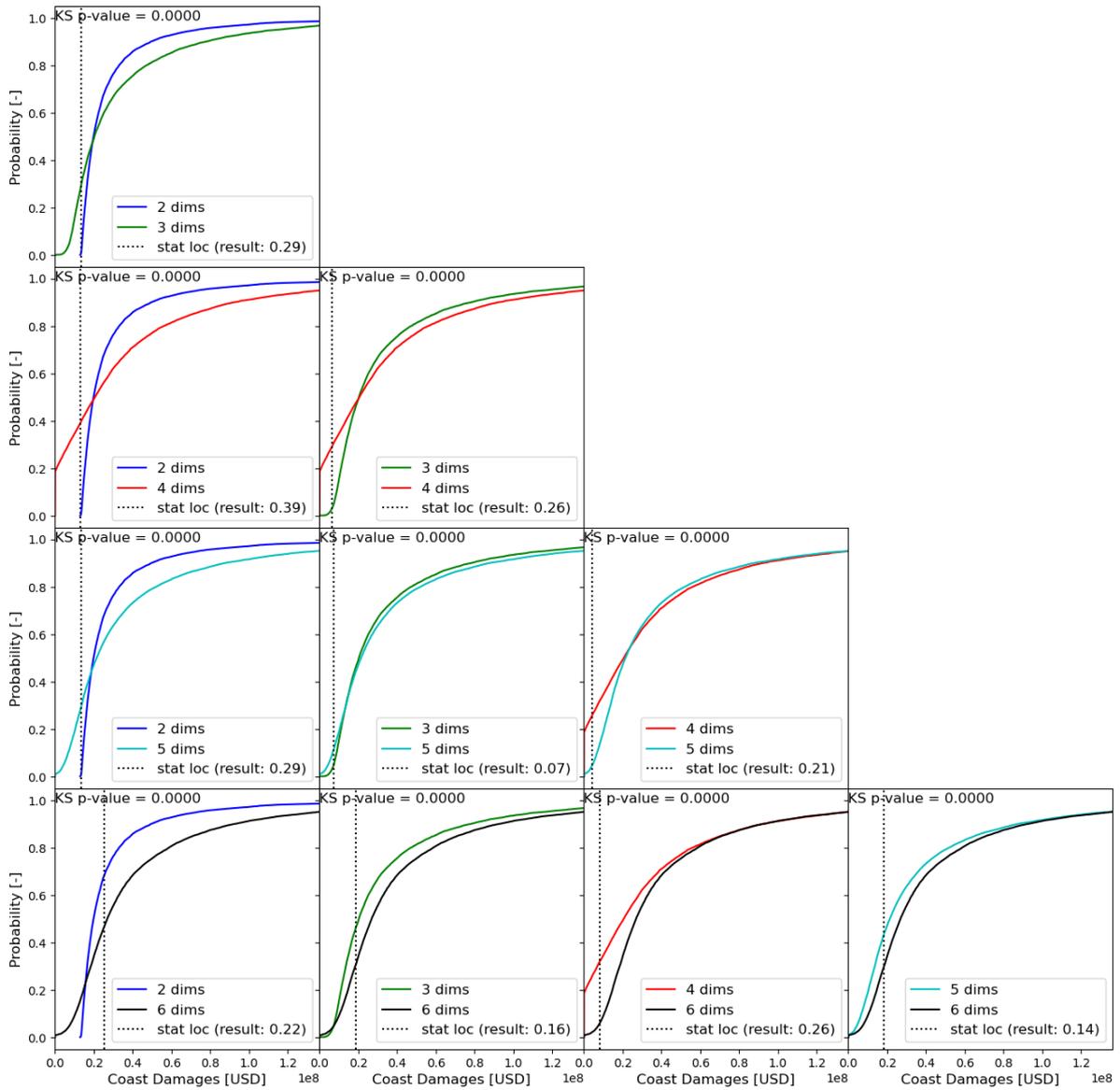


Figure K.5: Test statistics and associated p-values associated with two-sample KS test for empirical CDFs of different dimensionalities when sampling from the coast economic damages. If the p-value is smaller than 0.05, the null hypothesis (empirical CDFs come from the same parent distribution) is rejected.

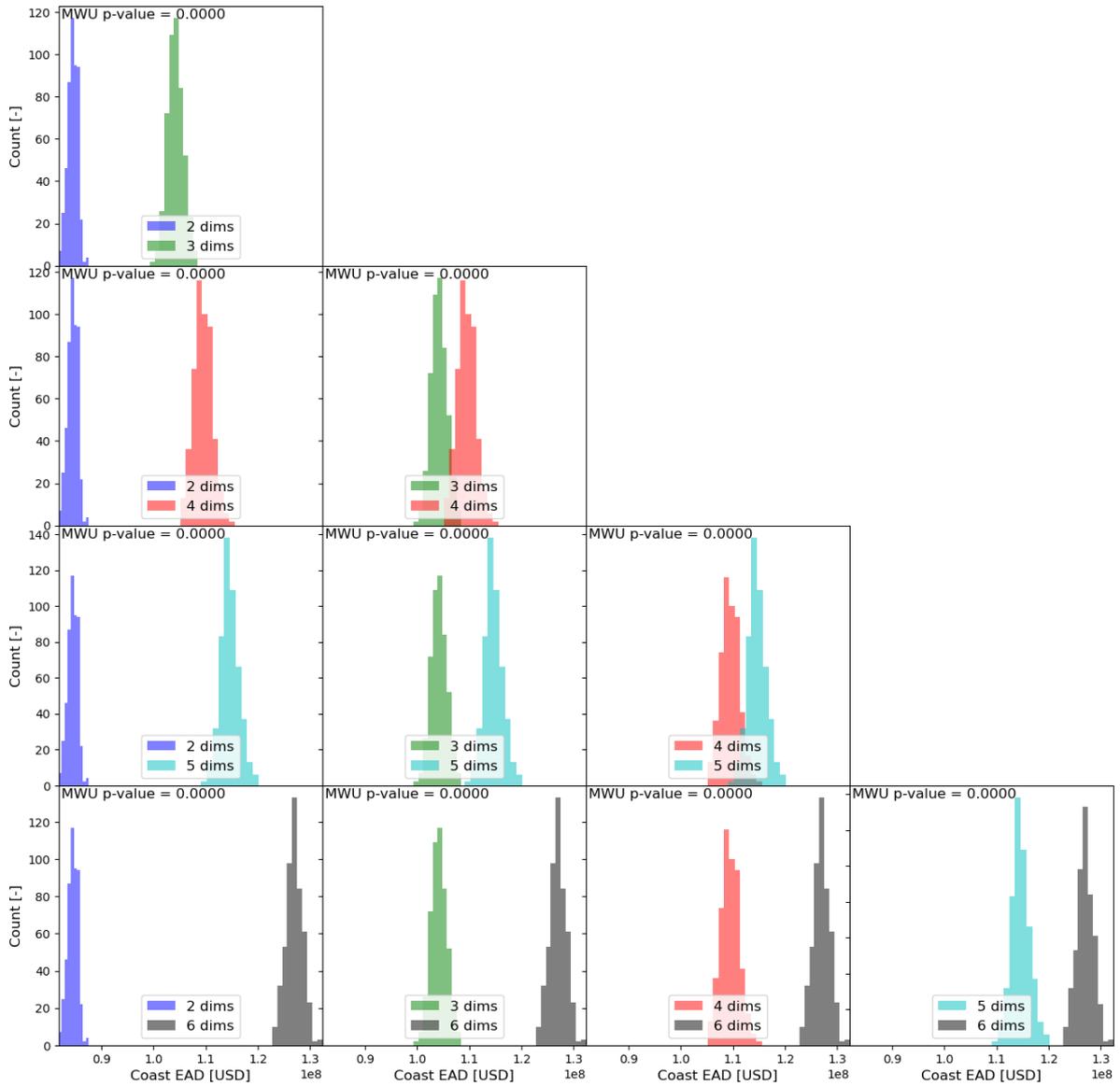
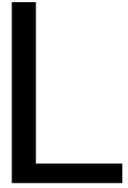


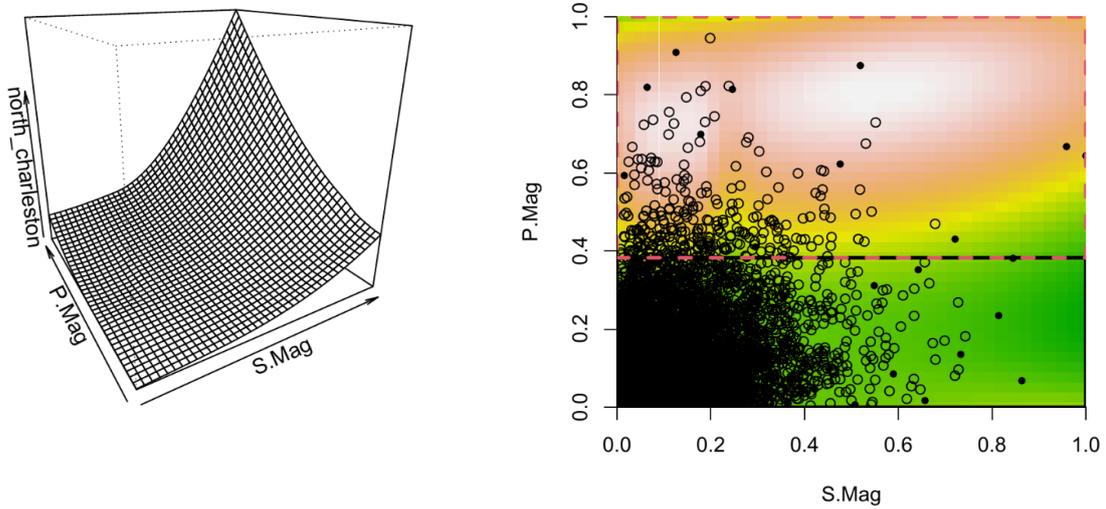
Figure K.6: P-values associated with MWU test for empirical CDFs of different dimensionalities when sampling from the coast economic damages. If the p-values are smaller than 0.05, the null hypothesis (empirical CDFs have the same location (EAD)) is rejected. The empirical bootstrap was repeated 500 times.

K.3. Discussion

Upon close inspection of fig. K.3 and K.6, an argument could be made for the visual overlap of the EAD distributions for the combination of certain variables. For example, the comparison between 5 and 6 dimensions for inland EAD shows a large portion of both tails overlap (fig. K.3). Moreover, there is a small difference of 3 - 5 Million USD in the mean EAD of the distributions. The similarity between these two distributions could show that P Lag has a negligible effect on the resources used in flood protection measures of communities located slightly upstream of the open coast.



Observed Binary Partition in a TGP



(a) Surrogate model for the a posteriori approach applied to North Charleston economic damages.

(b) Open and closed circles represent available and sampled events respectively. Normalized *ALM* statistics associated with the entire event space after using 33 samples. White represents the location with the largest *ALM* statistic. The available event the closest to this location will be added to the sampled events next. Partition is visible with dashed black and red lines.

Figure L.1: Surrogate model and associated uncertainty for the a posteriori approach after reaching the stopping criterion for North Charleston. Input and output are normalized.

Bibliography

- Heiko Apel, Oriol Martínez Trepas, Nguyen Nghia Hung, Do Thi Chinh, Bruno Merz, and Nguyen Viet Dung. Combined fluvial and pluvial urban flood hazard analysis: concept development and application to can tho city, mekong delta, vietnam. *Natural Hazards and Earth System Sciences*, 16(4):941–961, 2016. ISSN 1684-9981. doi: 10.5194/nhess-16-941-2016. URL <http://dx.doi.org/10.5194/nhess-16-941-2016>.
- A. Arns, T. Wahl, I.D. Haigh, J. Jensen, and C. Pattiaratchi. Estimating extreme water level probabilities: A comparison of the direct methods and recommendations for best practise. *Coastal Engineering*, 81:51–66, November 2013. ISSN 0378-3839. doi: 10.1016/j.coastaleng.2013.07.003. URL <http://dx.doi.org/10.1016/j.coastaleng.2013.07.003>.
- Mahmoud Ayyad, Muhammad R. Hajj, and Reza Marsooli. Artificial intelligence for hurricane storm surge hazard assessment. *Ocean Engineering*, 245, 2022. ISSN 0029-8018. doi: 10.1016/j.oceaneng.2021.110435. URL <http://dx.doi.org/10.1016/j.oceaneng.2021.110435>.
- Paul J.A. Baan and Frans Klijn. Flood risk perception and implications for flood risk management in the netherlands. *International Journal of River Basin Management*, 2(2):113–122, June 2004. ISSN 1814-2060. doi: 10.1080/15715124.2004.9635226. URL <http://dx.doi.org/10.1080/15715124.2004.9635226>.
- Tije M. Bakker, José A.A. Antolínez, Tim W.B. Leijnse, Stuart G. Pearson, and Alessio Giardino. Estimating tropical cyclone-induced wind, waves, and surge: A general methodology based on representative tracks. *Coastal Engineering*, 176:104154, September 2022. ISSN 0378-3839. doi: 10.1016/j.coastaleng.2022.104154. URL <http://dx.doi.org/10.1016/j.coastaleng.2022.104154>.
- Paul D. Bates, Matthew S. Horritt, and Timothy J. Fewtrell. A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. *Journal of Hydrology*, 387(1–2):33–45, June 2010. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2010.03.027. URL <http://dx.doi.org/10.1016/j.jhydrol.2010.03.027>.
- Paul D. Bates, Niall Quinn, Christopher Sampson, Andrew Smith, Oliver Wing, Jeison Sosa, James Savage, Gaia Olcese, Jeff Neal, Guy Schumann, Laura Giustarini, Gemma Coxon, Jeremy R. Porter, Mike F. Amodeo, Ziyang Chu, Sharai Lewis-Gruss, Neil B. Freeman, Trevor Houser, Michael Delgado, Ali Hamidi, Ian Bolliger, Kelly E. McCusker, Kerry Emanuel, Celso M. Ferreira, Arslaan Khalid, Ivan D. Haigh, Anaïs Couasnon, Robert E. Kopp, Solomon Hsiang, and Witold F. Krajewski. Combined modeling of us fluvial, pluvial, and coastal flood hazard under current and future climates. *Water Resources Research*, 57(2), February 2021. ISSN 1944-7973. doi: 10.1029/2020wr028673. URL <http://dx.doi.org/10.1029/2020WR028673>.
- R. Bentivoglio, E. Isufi, S. N. Jonkman, and R. Taormina. Rapid spatio-temporal flood modelling via hydraulics-based graph neural networks. *Hydrology and Earth System Sciences*, 27(23):4227–4246, 2023. doi: 10.5194/hess-27-4227-2023. URL <https://hess.copernicus.org/articles/27/4227/2023/>.
- S. Birkholz, M. Muro, P. Jeffrey, and H.M. Smith. Rethinking the relationship between flood risk perception and flood management. *Science of The Total Environment*, 478:12–20, April 2014. ISSN 0048-9697. doi: 10.1016/j.scitotenv.2014.01.061. URL <http://dx.doi.org/10.1016/j.scitotenv.2014.01.061>.
- Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- Georgii Bocharov. pyextremes, October 2023. URL <https://github.com/georgebv/pyextremes>.
- Judith Bosboom and Marcel Stive. *Coastal Dynamics*. TU Delft Open, 2021. doi: 10.5074/t.2021.001. Accessed on 02-May-2024.
- K. Breinl, U. Strasser, P. Bates, and S. Kienberger. A joint modelling framework for daily extremes of river discharge and precipitation in urban areas. *Journal of Flood Risk Management*, 10(1):97–114, 2015. ISSN 1753-318X. doi: 10.1111/jfr3.12150. URL <http://dx.doi.org/10.1111/jfr3.12150>.

- Christina R. Butler. Lowcountry at High Tide: A History of Flooding, Drainage, and Reclamation in Charleston, South Carolina. University of South Carolina Press, June 2020. ISBN 9781643360621. doi: 10.2307/j.ctvrk36b. URL <http://dx.doi.org/10.2307/j.ctvrk36b>.
- Paula Camus, Fernando J. Mendez, Raul Medina, and Antonio S. Cofiño. Analysis of clustering and selection algorithms for the study of multivariate wave climate. Coastal Engineering, 58(6):453–462, June 2011. ISSN 0378-3839. doi: 10.1016/j.coastaleng.2011.02.003. URL <http://dx.doi.org/10.1016/j.coastaleng.2011.02.003>.
- Wei-Bo Chen and Wen-Cheng Liu. Modeling flood inundation induced by river flow and storm surges over a river basin. Water, 6(10):3182–3199, 2014. ISSN 2073-4441. doi: 10.3390/w6103182. URL <http://dx.doi.org/10.3390/w6103182>.
- Thomas Cokelaer, Arseny Kravchenko, Ashutosh Varma, Brian L. Caio Eadi Stringari, Christian Brueffer, Eike Broda, Elmar Pruesse, Karthikeyan Singaravelan, Stefano Alberto Russo, and Zhengyi Li. cokelaer/fitter: v1.7.0, 2024. URL <https://zenodo.org/doi/10.5281/zenodo.10459943>.
- A. Couasnon, P. Scussolini, T. V. T. Tran, D. Eilander, S. Muis, H. Wang, J. Keesom, J. Dullaart, Y. Xuan, H. Q. Nguyen, H. C. Winsemius, and P. J. Ward. A flood risk framework capturing the seasonality of and dependence between rainfall and sea levels—an application to ho chi minh city, vietnam. Water Resources Research, 58(2), February 2022. ISSN 1944-7973. doi: 10.1029/2021wr030002. URL <http://dx.doi.org/10.1029/2021WR030002>.
- Anaïs Couasnon. The role of (in)dependence in flood risk assessments. PhD thesis, VU E-Publishing, 2023. URL <http://dx.doi.org/10.5463/thesis.176>.
- Anaïs Couasnon, Dirk Eilander, Sanne Muis, Ted I. E. Veldkamp, Ivan D. Haigh, Thomas Wahl, Hessel C. Winsemius, and Philip J. Ward. Measuring compound flood potential from river discharge and storm surge extremes at the global scale. Natural Hazards and Earth System Sciences, 20(2):489–504, February 2020. ISSN 1684-9981. doi: 10.5194/nhess-20-489-2020. URL <http://dx.doi.org/10.5194/nhess-20-489-2020>.
- Claudia Czado. Analyzing Dependent Data with Vine Copulas. Springer, 2019.
- Silvia De Angeli, Mirko D’Andrea, Giacomo Cazzola, Daniele Dolia, Enrico Duo, and Nicola Reborra. Coastal risk assessment framework: Comparison of modelled fluvial and marine inundation impacts, bocca di magra, ligurian coast, italy. Coastal Engineering, 134:229–240, 2018. ISSN 0378-3839. doi: 10.1016/j.coastaleng.2017.09.011. URL <http://dx.doi.org/10.1016/j.coastaleng.2017.09.011>.
- Deltares. FloodAdapt – an adaptation planning tool. <https://www.deltares.nl/en/software-and-data/products/floodadapt>, 2024. Accessed 08-May-2024.
- Ferdinand Diermanse, Kathryn Roscoe, Maarten van Ormondt, Tim Leijnse, Gundula Winter, and Panagiotis Athanasiou. Probabilistic compound flood hazard analysis for coastal risk assessment: A case study in charleston, south carolina. Shore & Beach, page 9–18, 2023. ISSN 0037-4237. doi: 10.34237/1009122. URL <http://dx.doi.org/10.34237/1009122>.
- J. Dißmann, E.C. Brechmann, C. Czado, and D. Kurowicka. Selecting and estimating regular vine copulae and application to financial returns. Computational Statistics & Data Analysis, 59:52–69, March 2013. ISSN 0167-9473. doi: 10.1016/j.csda.2012.08.010. URL <http://dx.doi.org/10.1016/j.csda.2012.08.010>.
- DSAIE. Data science and artificial intelligence for engineers. <https://interactivetextbooks.citg.tudelft.nl/dsaie/intro.html>, 2024. Accessed on 29-Apr-2024.
- Job Dullaart, S. (Sanne) Muis, Nadia Bloemendaal, Maria Chertova, Anaïs Couasnon, and J.C.J.H. (Jeroen) Aerts. Coast-rp: A global coastal dataset of storm tide return periods, 2022. URL https://data.4tu.nl/articles/_/13392314/2.
- Dirk Eilander, H el ene Boisgontier, Laur ene J. E. Bouaziz, Joost Buitink, Ana is Couasnon, Brendan Dalmijn, Mark Hegnauer, Tjalling de Jong, Sibren Loos, Indra Marth, and Willem van Verseveld. Hydromt: Automated and reproducible model building and analysis. Journal of Open Source Software, 8(83):4897, March 2023a. ISSN 2475-9066. doi: 10.21105/joss.04897. URL <http://dx.doi.org/10.21105/joss.04897>.

- Dirk Eilander, Anaïs Couasnon, Tim Leijnse, Hiroaki Ikeuchi, Dai Yamazaki, Sanne Muis, Job Dullaart, Arjen Haag, Hessel C. Winsemius, and Philip J. Ward. A globally applicable framework for compound flood hazard modeling. *Natural Hazards and Earth System Sciences*, 23(2):823–846, 2023b. ISSN 1684-9981. doi: 10.5194/nhess-23-823-2023. URL <http://dx.doi.org/10.5194/nhess-23-823-2023>.
- Dirk Eilander, Anaïs Couasnon, Frederiek C. Sperna Weiland, Willem Ligtoet, Arno Bouwman, Hessel C. Winsemius, and Philip J. Ward. Modeling compound flood risk and risk reduction using a globally applicable framework: a pilot in the sofala province of mozambique. *Natural Hazards and Earth System Sciences*, 23(6):2251–2272, 2023c. ISSN 1684-9981. doi: 10.5194/nhess-23-2251-2023. URL <http://dx.doi.org/10.5194/nhess-23-2251-2023>.
- Esri Federal Datasets. County Subdivisions, 2024. URL <https://www.arcgis.com/home/item.html?id=8e640fe423a941328039d57634c27aaa>. Accessed on 6-May-2024.
- Avantika Gori, Ning Lin, and Dazhi Xi. Tropical cyclone compound flood hazard assessment: From investigating drivers to quantifying extreme water levels. *Earth's Future*, 8(12), 2020. ISSN 2328-4277. doi: 10.1029/2020ef001660. URL <http://dx.doi.org/10.1029/2020EF001660>.
- Robert B. Gramacy. tgp: an r package for bayesian nonstationary, semiparametric nonlinear regression and design by treed gaussian process models. <https://cran.r-project.org/web/packages/tgp/vignettes/tgp.pdf>, 2023. Accessed on 30-Apr-2024.
- Robert B. Gramacy and Herbert K. H. Lee. Bayesian treed gaussian process models with an application to computer modeling, 2009. URL <https://doi.org/10.48550/arxiv.0710.4536>.
- E. J. Gumbel. The return period of flood flows. *The Annals of Mathematical Statistics*, 12(2):163–190, June 1941. ISSN 0003-4851. doi: 10.1214/aoms/1177731747. URL <http://dx.doi.org/10.1214/aoms/1177731747>.
- M. Hegnauer, J.J. Beersma, H.F.P. van den Boogaard, T.A. Buishand, and R.H. Passchier. Generator of rainfall and discharge extremes (grade) for the rhine and meuse basins. Technical report, Deltares, 2014. URL https://publications.deltares.nl/1209424_004_0018.pdf.
- Gijs G. Hendrickx, José A.A. Antolínez, and Peter M.J. Herman. Predicting the response of complex systems for coastal management. *Coastal Engineering*, 182, 2023. ISSN 0378-3839. doi: 10.1016/j.coastaleng.2023.104289. URL <http://dx.doi.org/10.1016/j.coastaleng.2023.104289>.
- A. Hendry, I. D. Haigh, R. J. Nicholls, H. Winter, R. Neal, T. Wahl, A. Joly-Laugel, and S. E. Darby. Assessing the characteristics and drivers of compound flooding events around the uk coast. *Hydrology and Earth System Sciences*, 23(7):3117–3139, 2019. doi: 10.5194/hess-23-3117-2019. URL <https://hess.copernicus.org/articles/23/3117/2019/>.
- Maike Holthuijzen, Brian Beckage, Patrick J. Clemins, Dave Higdon, and Jonathan M. Winter. Robust bias-correction of precipitation extremes using a novel hybrid empirical quantile-mapping method: Advantages of a linear correction for extremes. *Theoretical and Applied Climatology*, 149(1–2):863–882, May 2022. ISSN 1434-4483. doi: 10.1007/s00704-022-04035-2. URL <http://dx.doi.org/10.1007/s00704-022-04035-2>.
- Debra Holtz, Adam Markham, Kate Cell, and Brenda Ekwurzel. National landmarks at risk: How rising seas, floods, and wildfires are threatening the United States' most cherished historic sites. Union of Concerned Scientists, 2014. URL <https://www.jstor.org/stable/resrep00036>.
- J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.
- Hideaki Ishibashi and Hideitsu Hino. Stopping criterion for active learning based on error stability, 2021.
- R. A. Jane, V. Malagón-Santos, M. M. Rashid, L. Doebele, T. Wahl, S. R. Timmers, K. A. Serafin, L. Schmied, and C. Lindemer. A hybrid framework for rapidly locating transition zones: A comparison of event- and response-based return water levels in the suwannee river fl. *Water Resources Research*, 58(11), November 2022. ISSN 1944-7973. doi: 10.1029/2022wr032481. URL <http://dx.doi.org/10.1029/2022WR032481>.

- R. W. Kennard and L. A. Stone. Computer aided design of experiments. *Technometrics*, 11(1):137–148, 1969. ISSN 1537-2723. doi: 10.1080/00401706.1969.10490666. URL <http://dx.doi.org/10.1080/00401706.1969.10490666>.
- Sonu Khanal, Nina Ridder, Hylke de Vries, Wilco Terink, and Bart van den Hurk. Storm surge and extreme river discharge: A compound event analysis using ensemble impact modeling. *Frontiers in Earth Science*, 7, 2019. ISSN 2296-6463. doi: 10.3389/feart.2019.00224. URL <http://dx.doi.org/10.3389/feart.2019.00224>.
- Frans Klijn, Heidi Kreibich, Hans de Moel, and Edmund Penning-Rowsell. Adaptive flood risk management planning based on a comprehensive flood risk conceptualisation. *Mitigation and Adaptation Strategies for Global Change*, 20(6):845–864, March 2015. ISSN 1573-1596. doi: 10.1007/s11027-015-9638-z. URL <http://dx.doi.org/10.1007/s11027-015-9638-z>.
- Kristian Kumbier, Rafael C. Carvalho, Athanasios T. Vafeidis, and Colin D. Woodroffe. Investigating compound flooding in an estuary using hydrodynamic modelling: a case study from the shoalhaven river, australia. *Natural Hazards and Earth System Sciences*, 18(2):463–477, 2018. ISSN 1684-9981. doi: 10.5194/nhess-18-463-2018. URL <http://dx.doi.org/10.5194/nhess-18-463-2018>.
- David A. Lavers, Adrian Simmons, Freja Vamborg, and Mark J. Rodwell. An evaluation of era5 precipitation for climate monitoring. *Quarterly Journal of the Royal Meteorological Society*, 148(748):3152–3165, August 2022. ISSN 1477-870X. doi: 10.1002/qj.4351. URL <http://dx.doi.org/10.1002/qj.4351>.
- Tim Leijnse, Maarten van Ormondt, Kees Nederhoff, and Ap van Dongeren. Modeling compound flooding in coastal systems using a computationally efficient reduced-physics solver: Including fluvial, pluvial, tidal, wind- and wave-driven processes. *Coastal Engineering*, 163, 2021. ISSN 0378-3839. doi: 10.1016/j.coastaleng.2020.103796. URL <http://dx.doi.org/10.1016/j.coastaleng.2020.103796>.
- David J. C. MacKay. Information-Based Objective Functions for Active Data Selection. *Neural Computation*, 4(4):590–604, 07 1992. ISSN 0899-7667. doi: 10.1162/neco.1992.4.4.590. URL <https://doi.org/10.1162/neco.1992.4.4.590>.
- Hamed Moftakhari, Jochen E. Schubert, Amir AghaKouchak, Richard A. Matthew, and Brett F. Sanders. Linking statistical and hydrodynamic modeling for compound flood hazard assessment in tidal channels and estuaries. *Advances in Water Resources*, 128:28–38, 2019. ISSN 0309-1708. doi: 10.1016/j.advwatres.2019.04.009. URL <http://dx.doi.org/10.1016/j.advwatres.2019.04.009>.
- Hamed R. Moftakhari, Gianfausto Salvadori, Amir AghaKouchak, Brett F. Sanders, and Richard A. Matthew. Compounding effects of sea level rise and fluvial flooding. *Proceedings of the National Academy of Sciences*, 114(37):9785–9790, 2017. ISSN 1091-6490. doi: 10.1073/pnas.1620325114. URL <http://dx.doi.org/10.1073/pnas.1620325114>.
- Oswaldo Morales-Nápoles, Mojtaba Rajabi-Bahaabadi, Gina Alexandra Torres-Alves, and Cornelis Marcel Pieter 't Hart. Chimera: An atlas of regular vines on up to 8 nodes. *Scientific Data*, 10(1), May 2023. ISSN 2052-4463. doi: 10.1038/s41597-023-02252-6. URL <http://dx.doi.org/10.1038/s41597-023-02252-6>.
- Thomas Nagler and Thibault Vatter. *pyvinecopulib*, 2023. URL <https://zenodo.org/doi/10.5281/zenodo.10435751>.
- Kees Nederhoff, Tim W. B. Leijnse, Kai Parker, Jennifer Thomas, Andrea O'Neill, Maarten van Ormondt, Robert McCall, Li Erikson, Patrick L. Barnard, Amy Foxgrover, Wouter Klessens, Norberto C. Nadal-Caraballo, and Thomas Chris Massey. Tropical or extratropical cyclones: what drives the compound flood hazard, impact, and risk for the united states southeast atlantic coast? *Natural Hazards*, April 2024. ISSN 1573-0840. doi: 10.1007/s11069-024-06552-x. URL <http://dx.doi.org/10.1007/s11069-024-06552-x>.
- NOAA. National Hurricane Center and Central Pacific Hurricane Center. Saffir-Simpson Hurricane Wind Scale. <https://www.nhc.noaa.gov/aboutsshws.php>, Mar 2012. Accessed on 13-Mar-2024.
- NOAA. Tides & Currents. About Harmonic Constituents. https://tidesandcurrents.noaa.gov/about_harmonic_constituents.html, Jun 2019. Accessed on 13-Mar-2024.

- NOAA. National Hurricane Center and Central Pacific Hurricane Center. Tropical Cyclone Climatology. <https://www.nhc.noaa.gov/climo/>, Sep 2021. Accessed on 13-Mar-2024.
- Elisa Oliveri and Mario Santoro. Estimation of urban structural flood damages: the case study of palermo. *Urban Water*, 2(3):223–234, September 2000. ISSN 1462-0758. doi: 10.1016/s1462-0758(00)00062-5. URL [http://dx.doi.org/10.1016/S1462-0758\(00\)00062-5](http://dx.doi.org/10.1016/S1462-0758(00)00062-5).
- Kai Parker, Li Erikson, Jennifer Thomas, Kees Nederhoff, Patrick Barnard, and Sanne Muis. Relative contributions of water-level components to extreme water levels along the us southeast atlantic coast from a regional-scale water-level hindcast. *Natural Hazards*, 117(3):2219–2248, 2023. ISSN 1573-0840. doi: 10.1007/s11069-023-05939-6. URL <http://dx.doi.org/10.1007/s11069-023-05939-6>.
- F Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- R. C. Phillips, S. Samadi, D. B. Hitchcock, M. E. Meadows, and C. A. M. E. Wilson. The devil is in the tail dependence: An assessment of multivariate copula-based frameworks and dependence concepts for coastal compound flood dynamics. *Earth's Future*, 10(9), 2022. ISSN 2328-4277. doi: 10.1029/2022ef002705. URL <http://dx.doi.org/10.1029/2022EF002705>.
- Mehdi Pourmostafa and Parviz Ghadimi. Unsteady 2d and 3d navier-stokes solver with application of multi-grid scheme to pressure poisson fractional step on arbitrary unstructured grids in various applications with emphasis on ship motion. *Mathematical Problems in Engineering*, 2020:1–28, October 2020. ISSN 1024-123X. doi: 10.1155/2020/3248958. URL <http://dx.doi.org/10.1155/2020/3248958>.
- Simon J.D. Prince. *Understanding Deep Learning*. The MIT Press, 2023. URL <http://udlbook.com>.
- Francesco Serinaldi. Dismissing return periods! *Stochastic Environmental Research and Risk Assessment*, 29(4):1179–1189, July 2014. ISSN 1436-3259. doi: 10.1007/s00477-014-0916-1. URL <http://dx.doi.org/10.1007/s00477-014-0916-1>.
- Robert H. Stewart. *Introduction to Physical Oceanography*. Texas A & M University, september edition, 2008.
- He Sun, Tandong Yao, Fengge Su, Zhihua He, Guoqiang Tang, Ning Li, Bowen Zheng, Jingheng Huang, Fanchong Meng, Tinghai Ou, and Deliang Chen. Corrected era5 precipitation by machine learning significantly improved flow simulations for the third pole basins. *Journal of Hydrometeorology*, 23(10):1663–1679, October 2022. ISSN 1525-7541. doi: 10.1175/jhm-d-22-0015.1. URL <http://dx.doi.org/10.1175/JHM-D-22-0015.1>.
- United States Census Bureau, n.d. URL https://www2.census.gov/geo/maps/DC2020/DC20BLK/st45_sc/cousub/. Accessed on 6-May-2024.
- USGS. USGS Water Data for the Nation. <https://waterdata.usgs.gov/nwis>, Dec 2009. Accessed on 13-Mar-2024.
- Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, Ilhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2. URL <https://doi.org/10.1038/s41592-019-0686-2>.
- Thomas Wahl, Shaleen Jain, Jens Bender, Steven D. Meyers, and Mark E. Luther. Increasing risk of compound flooding from storm surge and rainfall for major us cities. *Nature Climate Change*, 5(12):1093–1097, 2015. ISSN 1758-6798. doi: 10.1038/nclimate2736. URL <http://dx.doi.org/10.1038/nclimate2736>.

- Philip J Ward, Anaïs Couasnon, Dirk Eilander, Ivan D Haigh, Alistair Hendry, Sanne Muis, Ted I E Veldkamp, Hessel C Winsemius, and Thomas Wahl. Dependence between high sea-level and high river discharge increases flood hazard in global deltas and estuaries. *Environmental Research Letters*, 13(8), 2018. ISSN 1748-9326. doi: 10.1088/1748-9326/aad400. URL <http://dx.doi.org/10.1088/1748-9326/aad400>.
- Michael L. Waskom. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021. doi: 10.21105/joss.03021. URL <https://doi.org/10.21105/joss.03021>.
- Tim Webster, Kevin McGuigan, Kate Collins, and Candace MacDonald. Integrated river and coastal hydrodynamic flood risk mapping of the lahave river estuary and town of bridgewater, nova scotia, canada. *Water*, 6(3):517–546, 2014. ISSN 2073-4441. doi: 10.3390/w6030517. URL <http://dx.doi.org/10.3390/w6030517>.
- Wikipedia. Hurricane Matthew. https://en.wikipedia.org/wiki/Hurricane_Matthew, Apr 2023. Accessed on 13-Mar-2024.
- Wikipedia. Hurricane Irma. https://en.wikipedia.org/wiki/Hurricane_Irma, Feb 2024. Accessed on 13-Mar-2024.
- Joanne Williams, Kevin J. Horsburgh, Jane A. Williams, and Robert N. F. Proctor. Tide and skew surge independence: New insights for flood risk. *Geophysical Research Letters*, 43(12):6410–6417, June 2016. ISSN 1944-8007. doi: 10.1002/2016gl069522. URL <http://dx.doi.org/10.1002/2016GL069522>.
- M. Gordon Wolman and John P. Miller. Magnitude and frequency of forces in geomorphic processes. *The Journal of Geology*, 68(1):54–74, January 1960. ISSN 1537-5269. doi: 10.1086/626637. URL <http://dx.doi.org/10.1086/626637>.